



# Genetic and functional insights into the fractal structure of the heart

Hannah V Meyer<sup>1,+</sup>, , Timothy JW Dawes<sup>2,3,+</sup>, Marta Serrani<sup>4,9,+</sup>, Wenjia Bai<sup>5</sup>, Pawel Tokarczuk<sup>2</sup>, Jiashen Cai<sup>6,7</sup>, Antonio de Marvao<sup>2</sup>, Albert Henry<sup>12,15</sup>, R Thomas Lumbers<sup>12,13,14</sup>, Jakob Gierden<sup>16,17</sup>, Thomas Thumberger<sup>17</sup>, Joachim Wittbrodt<sup>17</sup>, James S Ware<sup>2,3,18</sup>, Daniel Rueckert<sup>5</sup>, Paul M Matthews<sup>8</sup>, Sanjay K Prasad<sup>3</sup>, Maria L Costantino<sup>9,\*</sup>, Stuart A Cook<sup>2,3,6,10,\*</sup>, Ewan Birney<sup>11,\*</sup>, and Declan P O'Regan<sup>2,\*</sup>, 

<sup>1</sup>Cold Spring Harbor Laboratory, Simons Center for Quantitative Biology, USA

<sup>2</sup>MRC London Institute of Medical Sciences, Imperial College London, UK

<sup>3</sup>National Heart and Lung Institute, Imperial College London, UK

<sup>4</sup>Department of Chemical Engineering and Biotechnology, University of Cambridge, UK

<sup>5</sup>Department of Computing, Imperial College London, UK

<sup>6</sup>Duke–National University of Singapore, Singapore

<sup>7</sup>Department of Internal Medicine, Singapore Health Services, Singapore

<sup>8</sup>Department of Brain Sciences, Department of Medicine and UK Dementia Research Institute at Imperial College London, UK

<sup>9</sup>Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Italy

<sup>10</sup>National Heart Centre Singapore, Singapore

<sup>11</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), UK

<sup>12</sup>Institute of Health Informatics, University College London, UK

<sup>13</sup>Health Data Research UK, London, UK

<sup>14</sup>Barts Heart Centre, St. Bartholomew's Hospital, London, UK

<sup>15</sup>Institute of Cardiovascular Science, University College London, UK


<sup>16</sup>Department of Pediatric Cardiology, University Hospital Heidelberg, Heidelberg, Germany

<sup>17</sup>Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany

<sup>18</sup>Cardiovascular Magnetic Resonance Unit, Royal Brompton Hospital, London, UK

+These authors contributed equally to this work

\*These authors jointly supervised this work

Corresponding authors

## ABSTRACT

The inner surfaces of the human heart are covered by a complex network of muscular strands that is thought to be a vestige of embryonic development.<sup>1,2</sup> The function of these trabeculae in adults and their genetic architecture are unknown. To investigate this we performed a genome-wide association study using fractal analysis of trabecular morphology as an image-derived phenotype in 18,096 UK Biobank participants. We identified 16 significant loci containing genes associated with haemodynamic phenotypes and regulation of cytoskeletal arborisation.<sup>3,4</sup> Using biomechanical simulations and human observational data, we demonstrate that trabecular morphology is an important determinant of cardiac performance. Through genetic association studies with cardiac disease phenotypes and Mendelian randomisation, we find a causal relationship between trabecular morphology and cardiovascular disease risk. These findings suggest an unexpected role for myocardial trabeculae in the function of the adult heart, identify conserved pathways that regulate structural complexity, and reveal their influence on susceptibility to disease.

## Main

The chambers of the mature human heart have a complex inner surface whose function is unknown. Unlike the smooth endothelium of the great vessels, the endocardial surfaces of both ventricles are lined by a fenestrated network of muscular trabeculae which extend into the cavity. Their embryological development is driven by highly-conserved signalling pathways involving the endocardium-myocardium and extra-cellular matrix that regulate myocardial proliferation during cardiac morphogenesis.<sup>2,5-9</sup>

Cell lineage tracing suggests that trabeculae have a molecular and developmental identity which is distinct from the compact myocardium.<sup>10</sup> The high surface area of trabeculae enables nutrient and oxygen diffusion from blood pool to myocardium before the coronary circulation is established.<sup>1</sup> Trabeculae are also vital to formation of the conduction system.<sup>11</sup> Theoretical analyses have proposed that their complex structure may contribute to efficient intra-ventricular flow patterns.<sup>12–14</sup> While hypertrabeculation is observed as a feature of some genetically-characterised cardiomyopathies,<sup>15</sup> the physiological function of trabeculae in adult hearts, their genetic architecture, and potential role in common disease have not been determined.

The distinguishing trait of trabeculae is their branching morphology and the degree of such biological complexity in the heart can be quantified by fractal dimension (FD) analysis of cardiac magnetic resonance (CMR) imaging.<sup>8</sup> In a replicated genome-wide association study (GWAS), using FD as an image-derived phenotype, we identify loci linked with trabecular morphology. Knockout models of loci-associated genes showed a marked decrease in trabecular complexity. Using biomechanical modelling and human observational data, we find a causal relationship between myocardial trabeculation and ventricular performance, with Mendelian randomisation showing that reduced trabecular complexity is causally associated with the risk of heart failure.

## Data overview

UK Biobank is a prospective cohort study collecting deep genetic and phenotypic data on approximately 500,000 individuals from across the United Kingdom, aged between 40 and 69 at recruitment.<sup>16</sup> Of these, 100,000 participants are being recalled for enhanced phenotyping which includes CMR imaging.<sup>17</sup> Non-invasive data on a range of haemodynamic parameters are also collected at the time of imaging. Following automated image-quality control<sup>18</sup> and exclusion of subjects with missing covariates, 18,096 unrelated participants that formed a well mixed population of European ethnicity (Extended Data Fig 1a) were used for discovery (Supplementary Table 1 and Fig. 1d). A separate UK Biobank dataset of 6,536 participants and a further independent cohort of 1,129 healthy adults (UK Digital Heart study) were used for validation (Extended Data Fig 1b).<sup>19</sup> Disease associations were assessed in 510 patients with dilated cardiomyopathy (DCM) (Supplementary Table 1) of which 307 also had CMR imaging, as well as in summary GWAS data for heart failure of mixed aetiology from 47,309 cases and 930,014 controls, across 26 studies of European ancestry from the HERMES consortium.<sup>20</sup>

## Fractal analysis of trabeculation

We used a fully convolutional network for automated left ventricular segmentation and volumetry of CMR images.<sup>21</sup> Using edge detection of the endocardium we derived a scale invariant FD ratio for each slice, where a higher value indicates a greater degree of surface complexity (Fig. 1).<sup>22</sup> To account for variations in cardiac size and for consistent anatomical comparisons within and between populations we interpolated the data to 9 slices (see Extended Data Fig 2) which were equally divided into basal, mid-ventricular and apical thirds. An identical analytic pipeline was performed in the validation cohorts. We also showed that fractal analysis could be performed on other imaging modalities (see Extended Data Fig 3a). In addition, we used motion analysis to determine spatial components of myocardial strain (Extended Data Fig 3b).

## Genome-wide association analysis

We first explored variation in FD across slices using principal components analysis and noted multiple modes of variation. (Extended Data Fig 3c). We then used individual slice level GWAS to test association for different modes of variation, followed by meta-analysis across the slices to capture any additional global associations.

We performed a linear model for genetic association of 14,134,301 genetic variants on each of the 9 interpolated slice FD measures (Fig. 2a, Supplementary Data 1) of 18,096 individuals using anthropometric variables and genetic principal components as covariates. These genome-wide association studies showed low inflation and many individual loci passing the commonly used genome-wide association threshold of  $5 \times 10^{-8}$  after p-value adjustment for multiple testing by the effective number of tests ( $T_{eff} = 6.6$ ; see Extended Data Fig 4a,b; Supplementary Information Table 2). Figure 2b shows the resulting 16 independent loci from the meta-analysis of the per-slice GWAS summary statistics and the individual slice(s) which the loci are associated with. Four loci were only discovered using this joint meta-analysis approach (Fig. 2b, orange circles); the remaining 12 loci show patterns of association that extend over multiple adjacent slices with varying effect sizes from base to apex (Extended Data Fig 5a). We conducted two additional, analogous association studies including either end-diastolic volume or myocardial strain as covariates. Both studies led to the discovery of the same loci, indicating that FD associations are independent of ventricular size and strain (Extended Data Fig 4c,d).

To replicate our findings we analysed the genetic associations of the discovered loci with trabeculation-derived FD measurements in two separate cohorts: CMR images of 6,536 UK Biobank participants (released after the initial discovery GWAS) and 1,129 healthy volunteers from the UK Digital Heart Study. We applied the same image analysis pipeline and conducted an equivalent genetic association study on genetic variants in the 16 loci associated in the discovery cohort. In the larger UK Biobank replication cohort, eight of the loci replicated the results observed in the discovery cohort (Supplementary Information Tables 3 and 4). In the smaller, healthy volunteer replication cohort fewer associations passed the Bonferroni-adjusted p-value threshold (threshold  $p_{Bonferroni} = 0.003$ ; 2 variants, Supplementary Information Table 5). In both replication

studies, the estimates of effect direction were highly concordant with the original discovery effect sizes (UK Biobank: 97% and UK Digital Heart: 91% of comparisons concordant) and showed correlation of the effect size estimates ( $r^2 = 0.87$  and  $r^2 = 0.50$ , respectively; [Extended Data Fig 5b,c](#)). Permutation tests generating empirical concordance distributions show that the observed concordances are unlikely to be observed by chance ( $p_{\text{empirical}} < 10^{-5}$ ).

### Associations of discovered loci

We systematically analysed the 16 discovered loci with the rich genetic resources of other studies, drawing from both the extensive GWAS Catalog,<sup>23</sup> and more recent phenome-wide associations (PheWAS) from UK Biobank.<sup>24</sup> [Supplementary Table 2](#) summarises our findings (for details on loci see [Supplementary Information Table 6](#)).

Ten of the 16 loci are also associated with at least one component of heart function, such as pulse rate, QRS duration, left ventricular structure and function (for details see [Supplementary Data 2](#) and [Supplementary Information Table 8](#)). We compared our loci to the extensive GTEx catalog<sup>25</sup> of gene expression quantitative trait loci (eQTL; [Supplementary Table 2](#), [Supplementary Data 3](#) and [Extended Data Fig 6a](#)). Nine of the 16 loci showed an overlap with a GTEx locus; in eight cases at least one of the eQTL tissues was either cardiac tissue or skeletal muscle; in one case the only significant tissue was transformed fibroblasts. A particularly strongly annotated association is on chromosome 8, in a region of open chromatin that is an eQTL for the *MTSS1* gene ([Fig. 2c](#)). This locus is also associated with a variety of cardiac structure and function phenotypes ([Supplementary Table 2](#), rs35006907), and the lead genetic variant located is in a region of open chromatin in heart tissues (ENSR00000868700, ENSEMBL regulatory build, Ensembl release 99<sup>26</sup>). Representative myocardial borders associated with this locus are depicted in [Extended Data Fig 3d](#).

As well as previously reported associations, we were interested in the functional annotations of our GWAS results. As the per-slice GWAS ([Fig. 2b](#)) suggested regionally-driven signals, we conducted genome-wide associations of FD in basal (1-3), mid-ventricular (4-6) and apical (7-9) slices. We analysed all genetic variants of these association results for enrichment in regulatory and functional annotations. The strongest associations of the genetic loci were to open-chromatin regions in fetal heart tissue, particularly in the mid and apical regions ([Extended Data Fig 6b](#)).

Overall the discovered loci are mainly linked with either molecular or physiological cardiac phenotypes. Some loci are likely developmental, such as the locus on chromosome 8 associated with *MTSS1*, affecting many aspects of cardiac function whereas other loci have more specific associations. Amongst the well-annotated loci electrophysiological, haemodynamic and structural traits are common themes, for example rs17608766 which is associated with QRS duration, blood pressure, cardiac anatomy and eQTLs to three genes in skeletal muscle.

### Knockout models

To gain further confidence in the role of the trabeculae-associated genes *GOSR2* and *MTSS1*, we assessed *in vivo* CRISPR-Cas9-mediated gene knock-outs (KO) in medaka (*Oryzias latipes*). Crispant embryos were phenotypically evaluated at 4 days post fertilization when significant steps of cardiovascular development are complete. Two batches (replicates) of crispants were initially classified into three main categories ([Fig. 3a](#)). A significant proportion of embryos were dead after *mtss1* KO. In viable embryos, we found retardation of development with a range of severe, sub-lethal to moderate phenotypes for both *gosr2* and *mtss1* crispants. Features observed on the level of the cardiovascular system were further described using qualitative phenotypic terms including morphological abnormalities, atrioventricular (AV) block, reverse heart looping, and haemorrhage or coagulation ([Fig. 3b](#)). [Fig. 3c](#) shows an example of a moderately affected *mtss1* crispant embryo. To specifically address the endocardial structure, entire heart volumes of *mtss1* crispant and control embryos (*myl7::EGFP* reporter line) were further analyzed at high resolution using light-sheet microscopy. Surface rendering revealed a marked reduction of trabeculation in the *mtss1* crispant compared to the control ([Fig. 3d](#)).

### Cardiac function

To understand the influence of myocardial trabeculation on cardiac function, we used a biomechanical simulation of the heart in a haemodynamic circuit. This allowed us to vary trabecular morphology selectively and observe its effects on ventricular performance in comparison to equivalent observational data in humans.

A visualisation of cardiac mechanics during systole and diastole is provided by plotting a closed loop describing the relationship between left ventricular pressure and left ventricular volume at multiple time points during a complete cardiac cycle ([Fig. 4a](#)).<sup>27,28</sup> To understand how trabeculae influence cardiac function we therefore assessed the relationship between FD and pressure-volume parameters of the left ventricle both in human populations and *in silico* models. We performed this in the UK Biobank participants by analysing non-invasive estimates of central pressures combined with volumetric CMR data. In parallel, we developed a cardiovascular simulation, using finite element analysis of the left ventricle in a haemodynamic circuit. In this simulation, we selectively varied trabecular complexity, under the same initial loading and boundary conditions, to observe the consequent effect on stroke work and contractility. In UK Biobank participants, increasing FD was associated with higher stroke volume, stroke work and cardiac index (standardised  $\beta = 0.52, 0.67, 0.12$ ), findings which were concordant with

the biomechanical simulation across a range of filling pressures (Fig. 4a). Together these results suggest a causal relationship between trabecular complexity and ventricular performance (Extended Data Fig 7a-c, Supplementary Information Table 10). Trabeculae also give rise to the ventricular conduction system during embryonic heart development,<sup>11</sup> and we found a positive correlation in UK Biobank of QRS duration with FD (Extended Data Fig 3e).

## Disease association

Finally, we explored the relationship of trabeculae-associated loci with cardiovascular disease using broad genetic correlation analyses and disease-specific locus and phenotype analyses. We first applied cross-trait LD score regression to screen for genetic correlations between trabecular complexity and 732 traits available on LDhub (Supplementary Data 4).<sup>29</sup> The strongest positive and negative genetic correlations were with hypertension phenotypes and diagnosed vascular or heart problems (Extended Data Fig 7e-g), respectively.

We then analysed CMR images in patients diagnosed with DCM, a disease of the myocardium that may progress to heart failure. We observed that these patients show higher trabecular FD than controls especially towards the base and apex of the left ventricle (Fig. 4b, linear mixed model  $-\log_{10}(p) = 2846$ ).

In a logistic regression association analysis between trabeculation-associated loci and DCM, we find two loci with genetic association ( $p_{\text{empirical}} < 0.05$ ) even in this more modestly sized patient cohort (510 DCM cases). We then used summary case-control GWAS data with a clinical diagnosis of heart failure (HF) of any aetiology from the HERMES Consortium to directly explore the associations of trabeculation-linked loci. We found that two of the loci are also associated with HF at a Bonferroni-adjusted significance level ( $p < 0.003$ , Supplementary Information Table 11).

For both DCM and HF, we find a negative correlation with trabeculation, i.e. loci associated with decreasing trabeculation are associated with increased susceptibility to disease - with the locus around *GOSR2* (Extended Data Fig 7d) showing a strong association in both cohorts. To test the hypothesis that trabecular morphology is causally-related to heart failure, we used a two-sample Mendelian randomisation (MR) framework,<sup>30</sup> with the discovered independent loci as instrumental variables.<sup>31</sup> We used FD as our exposure variable and HF or DCM as outcomes. We tested a number of MR techniques, each addressing different assumptions (for details refer to Supplementary Note 1.1: Mendelian Randomisation) and found parameter estimates that support a causal relationship between trabecular morphology and both HF (Fig. 4c; Supplementary Information Table 12 and Extended Data Fig 8a), and DCM (Fig. 4d; Supplementary Information Table 12 and Extended Data Fig 8b). In both populations an increase in trabeculation leads to decreased risk of disease. The directionality of the MR associations, with trabeculation causally upstream of HF and DCM, was confirmed by MR Steiger test<sup>32</sup>, Supplementary Information Table 13). Using MR Egger, we detected weak pleiotropic effects for MR on HF; for MR on DCM, none were observed (Supplementary Information Table 15). Furthermore, estimates of the F-statistic indicate no weak instrument bias (Supplementary Information Table 14 and Supplementary Note 1.1: Mendelian Randomisation, Limitations).

## Discussion

Myocardial trabeculae were first described by the early human anatomists,<sup>33</sup> and although they are remarkably well-conserved in vertebrate evolution,<sup>34</sup> beyond a role in facilitating oxygenation of the developing fetal heart their function in adults has remained an enigma.<sup>35</sup> Deep learning image analysis enabled us to perform the first reported GWAS of trabecular morphology - using fractal dimension to quantify their characteristic geometric complexity. We found associations with trabecular complexity in loci related to cardiac function and electrocardiographic phenotypes, gene expression variation in cardiac tissues and cardiac development chromatin annotation that were independent of biophysical variables, ventricular volume and myocardial strain.

Two discovered loci (*MTSS1*, *GOSR2*) point to molecular pathways involved in cytoskeletal actin dynamics. Variants of *MTSS1* are known to be associated with myocardial geometry and cardiac function in mouse models and patient populations.<sup>36-38</sup> Interference with *mtss1* function in medaka was characterised by a range of phenotypes that included marked reduction in trabeculation. *MTSS1* is also highly expressed in cerebellar Purkinje cells where it regulates dendritic complexity by promoting the branching of actin filaments and inhibiting the formation of straight filaments.<sup>3</sup> Similarly, truncating mutations in *GOSR2* cause cytoskeletal fragmentation with reduced elaboration of neuronal dendritic arbors.<sup>4</sup> Dichotomous fractal branching greatly amplifies the surface area of tissues whether for information processing (neurons) or haemodynamic effects (heart),<sup>39</sup> suggesting these discovered loci may play a critical role in regulating arborisation traits across different organs.

Observational data in UK Biobank showed that trabecular complexity was associated with increasing stroke work - and biomechanical simulations provided concordant data showing that trabeculae have a load independent effect on left ventricular diastolic filling, contractility and systemic blood pressure. The architecture of trabeculae, at the interface between intra-cardiac flow and the compact myocardium, may therefore be important in explaining individual variation in cardiac efficiency. Furthermore, we found that trabecular morphology in humans was associated with intra-ventricular conduction - a discovery that implicates these complex structures in cardiac electrophysiology as well as mechanical function.<sup>40</sup>

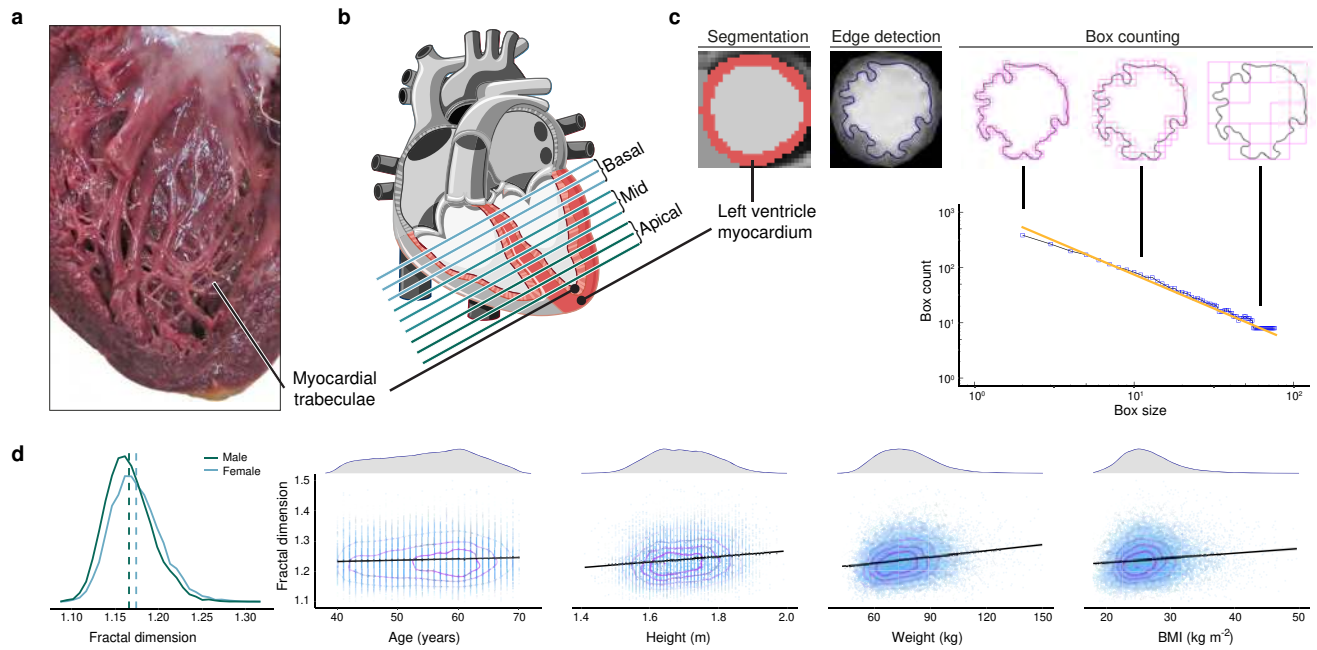
Our MR analyses support a causal role for trabecular morphology in both mixed aetiology heart failure and DCM. Taken with the observation of higher FD in disease phenotypes and our computational modelling of trabecular function, these findings suggest that trabeculae maintain cardiac performance in both healthy and failing hearts by increasing contractility and stroke work. We also found a number of loci that overlap with well-established cardiac genes (*TTN*, *TNNT2*), linked to sarcomeric function and cardiac morphogenesis, that are related to a spectrum of hyper-trabeculation phenotypes.<sup>41–43</sup> This suggests that genes linked to primary cardiomyopathies highlight molecular pathways that are important for trabecular formation and cardiac function more generally.<sup>44</sup>

The triangulation of theoretical models, observational data and genomics<sup>45</sup> is persuasive evidence that trabeculae are not simply vestigial features of development but are unexpected determinants of cardiac performance in adult hearts. Understanding the pathways which regulate the development of such complex biological structures provides a foundation for exploring new causal mechanisms in common cardiovascular diseases.

## References

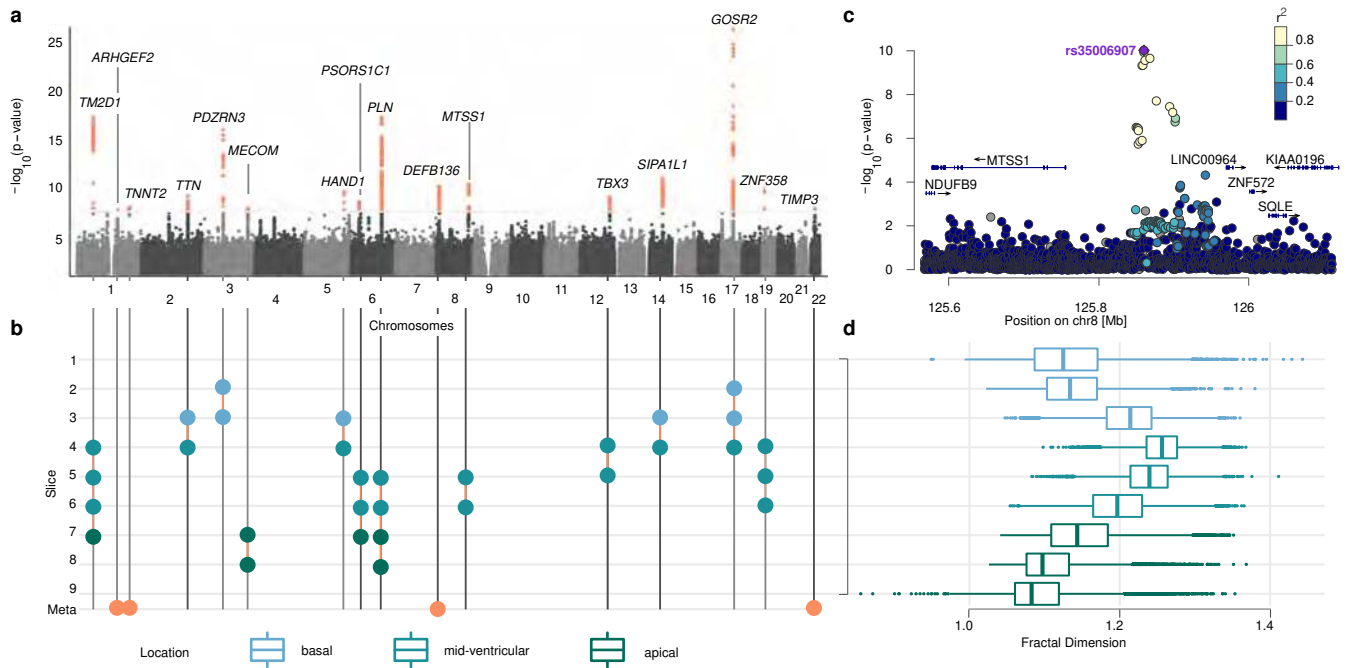
1. Sedmera, D. & McQuinn, T. Embryogenesis of heart muscle. *Heart Fail Clin* **4**, 235–245 (2008).
2. Sizarov, A. *et al.* Formation of the building plan of the human heart: morphogenesis, growth, and differentiation. *Circulation* **123**, 1125–35 (2011).
3. Kawabata Galbraith, K. *et al.* MTSS1 regulation of actin-nucleating formin DAAM1 in dendritic filopodia determines final dendritic configuration of purkinje cells. *Cell Rep* **24**, 95–106.e9 (2018).
4. Prashberger, R. *et al.* Mutations in Membrin/GOSR2 reveal stringent secretory pathway demands of dendritic growth and synaptic integrity. *Cell Rep* **21**, 97–109 (2017).
5. Chen, X. *et al.* Knockout of SRC-1 and SRC-3 in mice decreases cardiomyocyte proliferation and causes a noncompaction cardiomyopathy phenotype. *Int J Biol Sci* **11**, 1056–72 (2015).
6. Luxán, G., D’Amato, G. & de la Pompa, J. L. *Intercellular Signaling in Cardiac Development and Disease: The NOTCH pathway*, 103–114 (Springer Japan, Tokyo, 2016).
7. Han, P. *et al.* Coordinating cardiomyocyte interactions to direct ventricular chamber morphogenesis. *Nature* **534**, 700–4 (2016).
8. Captur, G. *et al.* Morphogenesis of myocardial trabeculae in the mouse embryo. *J Anat* **229**, 314–25 (2016).
9. Miquerol, L. *et al.* Biphasic development of the mammalian ventricular conduction system. *Circ Res* **107**, 153–161 (2010).
10. del Monte-Nieto, G. *et al.* Control of cardiac jelly dynamics by NOTCH1 and NRG1 defines the building plan for trabeculation. *Nature* **557**, 439–445 (2018).
11. van Weerd, J. H. & Christoffels, V. M. The formation and function of the cardiac conduction system. *Development* **143**, 197–210 (2016).
12. Vedula, V., Seo, J.-H., Lardo, A. C. & Mittal, R. Effect of trabeculae and papillary muscles on the hemodynamics of the left ventricle. *Theor Comput Fluid Dyn* **30**, 3–21 (2016).
13. Sacco, F. *et al.* Left ventricular trabeculations decrease the wall shear stress and increase the intra-ventricular pressure drop in CFD simulations. *Front Physiol* **9**, 458 (2018).
14. Paun, B., Bijmens, B. & Butakoff, C. Relationship between the left ventricular size and the amount of trabeculations. *Int J Numer Method Biomed Eng* **34**, e2939 (2018).
15. van Waning, J. I. *et al.* Genetics, clinical features, and long-term outcome of noncompaction cardiomyopathy. *J Am Coll Cardiol* **71**, 711–722 (2018).
16. Bycroft, C. *et al.* The UK biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
17. Petersen, S. E. *et al.* UK biobank’s cardiovascular magnetic resonance protocol. *J Cardiovasc Magn Reson* **18**, 8 (2016).
18. Tarroni, G. *et al.* Learning-based quality control for cardiac MR images. *IEEE Trans Med Imaging* **38**, 1127–1138 (2019).
19. Bai, W. *et al.* A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. *Med Image Anal* **26**, 133–45 (2015).
20. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun* **11**, 163 (2020).
21. Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* **20**, 65 (2018).
22. Captur, G. *et al.* Fractal analysis of myocardial trabeculations in 2547 study participants: Multi-ethnic study of atherosclerosis. *Radiology* **277**, 707–15 (2015).
23. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896–D901 (2017).
24. 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* **12**, 1–10 (2015).
25. GTEx portal. <https://gtexportal.org/> (2018).
26. Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T. & Flicek, P. R. The Ensembl regulatory build. *Genome Biol* **16**, 56 (2015).

27. Burkhoff, D., Mirsky, I. & Suga, H. Assessment of systolic and diastolic ventricular properties via pressure-volume analysis: a guide for clinical, translational, and basic researchers. *Am J Physiol Heart Circ Physiol* **289**, H501–12 (2005).
28. Seemann, F. *et al.* Noninvasive quantification of pressure-volume loops from brachial pressure and cardiovascular magnetic resonance. *Circ Cardiovasc Imaging* **12**, e008493 (2019).
29. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
30. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
31. Smith, G. D. *et al.* Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology. *PLoS Med* **4**, e352 (2007).
32. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).
33. Keele, K. *Leonardo Da Vinci's Elements of the Science of Man* (Academic Press, 2014).
34. Jensen, B., Wang, T., Christoffels, V. M. & Moorman, A. F. Evolution and development of the building plan of the vertebrate heart. *Biochim Biophys Acta* **1833**, 783–94 (2013).
35. Brutsaert, D. L. Cardiac endothelial-myocardial signaling: its role in cardiac growth, contractile performance, and rhythmicity. *Physiol Rev* **83**, 59–115 (2003).
36. Morley, M. P. *et al.* Cardioprotective effects of MTSS1 enhancer variants. *Circulation* **139**, 2073–2076 (2019).
37. Wild, P. S. *et al.* Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J Clin Invest* **127**, 1798–1812 (2017).
38. 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* **140**, 1318–1330 (2019).
39. Kenkel, N. C. & Walker, D. J. Fractals in the biological sciences. *Coenososes* **11**, 77–100 (1996).
40. Olejníčková, V., Šaňková, B., Sedmera, D. & Janáček, J. Trabecular architecture determines impulse propagation through the early embryonic mouse heart. *Front Physiol* **9**, 1876 (2019).
41. Ingles, J. *et al.* Evaluating the clinical validity of hypertrophic cardiomyopathy genes. *Circ Genom Precis Med* **12**, e002460 (2019).
42. Schafer, S. *et al.* Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat Genet* **49**, 46–53 (2017).
43. Miszalski-Jamka, K. *et al.* Novel genetic triggers and genotype-phenotype correlations in patients with left ventricular noncompaction. *Circ Cardiovasc Genet* **10**, e001763 (2017).
44. Tayal, U., Prasad, S. & Cook, S. A. Genetics and genomics of dilated cardiomyopathy and systolic heart failure. *Genome Med* **9**, 20 (2017).
45. Munafo, M. R. & Davey Smith, G. Robust research needs many lines of evidence. *Nature* **553**, 399–401 (2018).
46. UK Biobank - Neale Lab. <http://www.nealelab.is/uk-biobank/> (2018).

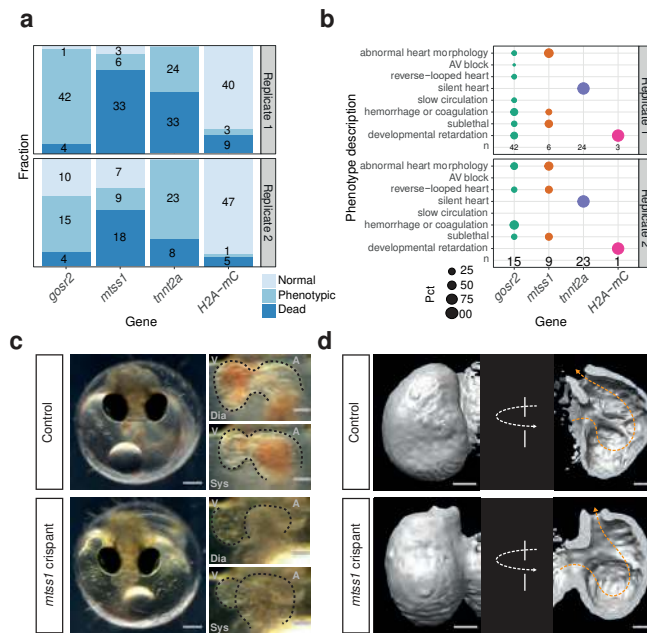


**Figure 1. Trabeculation phenotypes and covariates.** a) Macroscopic cut pathological section of the left ventricle demonstrating the branching network of muscular trabeculae lining the endocardial surface. b) Diagram of the heart illustrating the positioning of sections acquired during cardiac magnetic resonance (CMR) imaging for the assessment of trabecular complexity. c) Deep learning image segmentation was used for anatomical annotation of each pixel in the CMR dataset and to define an outer region of interest for subsequent fractal analysis. A binary mask was taken of the image followed by edge detection of the trabeculae. Box-counting across a range of sizes generated a log-log plot from which the gradient of a least-squares linear regression defined the fractal dimension. d) Distribution of fractal dimension and its relation to covariates used in the association study (n=18,096).

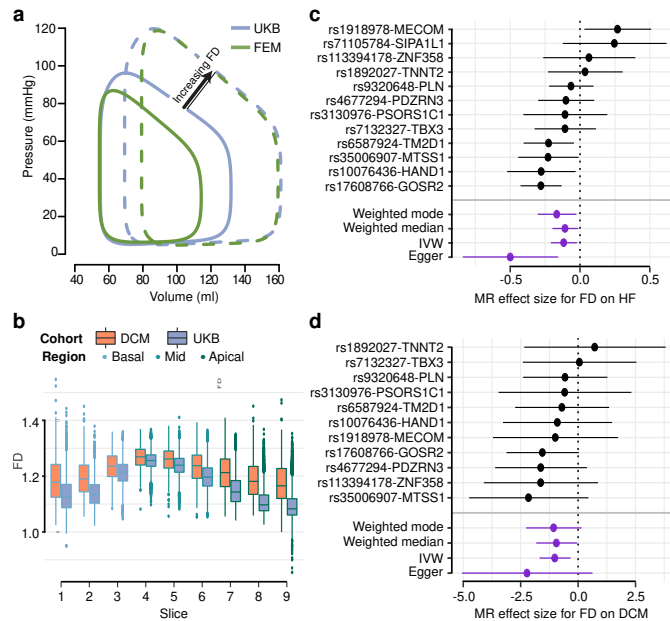




**Figure 2. Genetic associations of left ventricular trabeculation.** a) Manhattan plot (number of variants = 14,134,301) of meta-analysis p-values, depicted on  $-\log_{10}$  scale, uncorrected for multiple comparisons. Meta-analysis p-values estimated based on transformation of univariate signed t-statistic and  $\chi^2$  distribution with 9 degrees of freedom. Loci passing the genome-wide significance threshold  $5 \times 10^{-8}$  highlighted in orange (top). b) Diagram showing the slices driving the genetic association signal (compare [Extended Data Fig 4](#)): circles indicate a locus being associated (panel a) with respective slice and region (panel d). Loci marked in orange circles have no individual association  $p_{adj} = p \times T_{eff} < 5 \times 10^{-8}$  (where  $T_{eff} = 6.6$  is the effect number of independent phenotypic tests) and were only discovered in the meta-analysis. Loci are labeled by their nearest protein coding gene. c) Locus zoom of the locus on chromosome 8, associated with slices 5 and 6. d) Box-plot of FD measurements per slice, colour-coded by cardiac region. The lower and upper hinges in the boxplot correspond to the 25th and 75th percentiles (IQR), the horizontal line in the boxplot the median. The lower/upper whisker extends from the hinge to the smallest/largest value no further than  $1.5 \times IQR$ . Association (a) and phenotype (d) sample size: n=18,096.



**Figure 3. Knock-out of *mtss1* leads to reduction of cardiac trabeculation in medaka.** a) Counts (numbers) and corresponding proportion (bars) of normal, phenotypic and dead embryos after CRISPR-Cas9-mediated KO of *gosr2*, *mtss1*, and *mnt2a* (positive control), and H2A-mCherry (H2A-mC, injection control) at 4 DPF. b) Percentages of cardiovascular phenotypes, sublethal phenotypes, and developmental retardation. c) A moderately affected *mtss1* crispant in comparison to a control embryo at 4 DPF; overview of injected embryos (left), magnifications of the heart region (right) captured in end-diastolic (Dia) and end-systolic (Sys) phase, respectively; scale bars: 200  $\mu\text{m}$  (whole embryos) and 50  $\mu\text{m}$  (hearts), atrium (A), ventricle (V). d) Surface rendering of light-sheet microscopy recordings of control-injected embryo at 6 DPF and *mtss1* crispant at 7 DPF, images cropped to the ventricle; frontal view (left), and 180° rotated (right, cut open to visualize the endocardial surface); direction of blood flow (orange line), scale bars: 50  $\mu\text{m}$ .



**Figure 4. Relationship between trabecular complexity and cardiac function and disease** a) Variation in pressure-volume relationship with respect to trabecular fractal dimension (FD) in UK Biobank participants (UKB) and *in silico* biomechanical modelling (FEM) showing a positive association with left ventricular volumes and stroke work. b) Per-slice distribution of FD in the UK Biobank cohort (n=18,096) and dilated cardiomyopathy (DCM; n=510) patients. Boxplot: lower and upper hinges are 25th and 75th percentiles (IQR), the horizontal line at median; ower/upper whisker from hinge to the smallest/largest value no further than  $1.5 \times$  IQR. c). and d). Forest plots for FD effect on HF and DCM estimated by four MR methods. The contribution of each genetic variant to the overall estimate (black; estimated by Wald ratio) and their combined effect as a single genetic instrument (purple; estimated by indicated method) are shown for the four tested MR methods. Center values mark effect size point estimates, error bars the 95% confidence intervals. FD effect size estimates from uni-variate GWAS results on n=18,096 samples. HF samples sizes:  $n_{cases} = 47,309$ ,  $n_{controls} = 930,014$ ; DCM sample sizes:  $n_{cases} = 510$ ,  $n_{controls} = 1,136$ .

## Acknowledgments

The research was supported by the Medical Research Council, UK (MC-A651-53301); British Heart Foundation (NH/17/1/32725, RG/19/6/34387, RE/18/4/34215); Wellcome Trust (107469/Z/15/Z); National Institute of Environmental Health Sciences (R01 ES029917-02); Heidelberg University; the Simons Center for Quantitative Biology at Cold Spring Harbor Laboratory; and the National Institute for Health Research Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. P.M.M. also has been in receipt of generous personal and research support from the Edmond J Safra Foundation and Lily Safra, an NIHR Senior Investigator's Award, the Medical Research Council and the UK Dementia Research Institute. R.T.L. is supported by a UKRI Health Data Research Rutherford Fellowship (MR/S003754/1). A.H. is supported by a BHF PhD Studentship. J.G. is supported by a Research Center for Molecular Medicine (HRCMM) Career Development Fellowship, the MD/PhD program of the Medical Faculty Heidelberg, the Deutsche Herzstiftung e.V. (S/02/17), and by an Add-On Fellowship for Interdisciplinary Science of the Joachim Herz Stiftung.

Research funding for cohorts used in Mendelian Randomisation: NIHR Cardiovascular Biomedical Research Unit of Royal Brompton and Harefield NHS Foundation Trust (DCM cohort); Funding information for HERMES participating studies is detailed in Shah et al. 2020 (<https://doi.org/10.1038/s41467-019-13690-5>); data aggregation and downstream bioinformatics were funded through grants from the MRC Proximity to Discovery scheme, the NIHR UCLH Biomedical Research Centre, and the EU/EFPIA Innovative Medicines Initiative 2 Joint Undertaking BigData@Heart grant no. 116074.

The authors would like to thank Dr Hideaki Suzuki, previously of the Department of Medicine, Imperial College London, for his work on pre-processing the image data, and Prof Roberto Fumero, at the Politecnico di Milano, Italy, for advice on the finite element modelling. The authors thank Virginie Uhlmann (EMBL-EBI) for advice on radial image registration. We also thank L. Schertel and C. Baader for sgRNA production and the Wittbrodt Laboratory for critical discussion and support. We also acknowledge Ben Statton and Marjola Thanaj at Imperial College London for assisting with data pre-processing.

## Competing interests

The authors declare no competing interests.

## Author contributions

H.V.M. and T.J.W.D. performed the formal analysis and co-wrote the manuscript; M.S. and M.L.C. performed the *in silico* modelling; T.J.W.D and A.M. collected and analysed image data; R.T.L, A.H., J.S.W. and S.K.P collected and analysed the clinical data; W.B., P.T., J.C. and D.R. developed the computational phenotyping; J.G., T.T., and J.W. detailed the experimental strategy for the medaka validation; J.G. and T.T. designed and performed CRISPR-Cas9 knock-out experiments, and conducted phenotypic analysis under the guidance of J.W.; J.G. acquired LSM recordings, and analyzed and plotted the medaka knock-out and imaging data.; P.M.M., E.B., S.A.C. and D.P.O. provided critical interpretation of the results; E.B., S.A.C. and D.P.O. conceived the study, managed the project and revised the manuscript. All authors reviewed the final manuscript.

## Data availability

The genetic and phenotypic UK Biobank data are available upon application to the UK Biobank as detailed here: <https://bbams.ndph.ox.ac.uk/ams/>

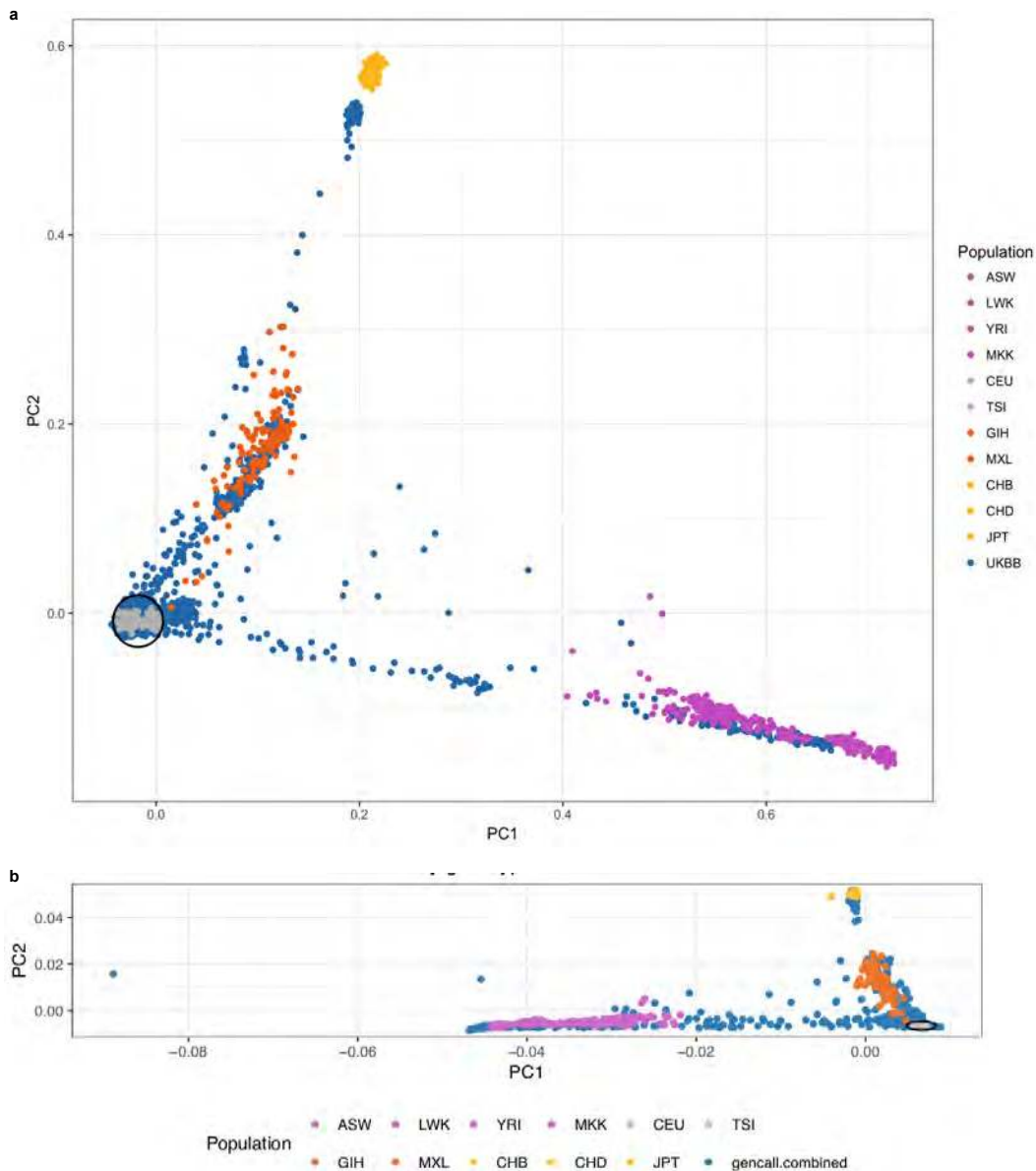
## Code availability

The analysis code is freely available on GitHub (10.5281/zenodo.3698268).

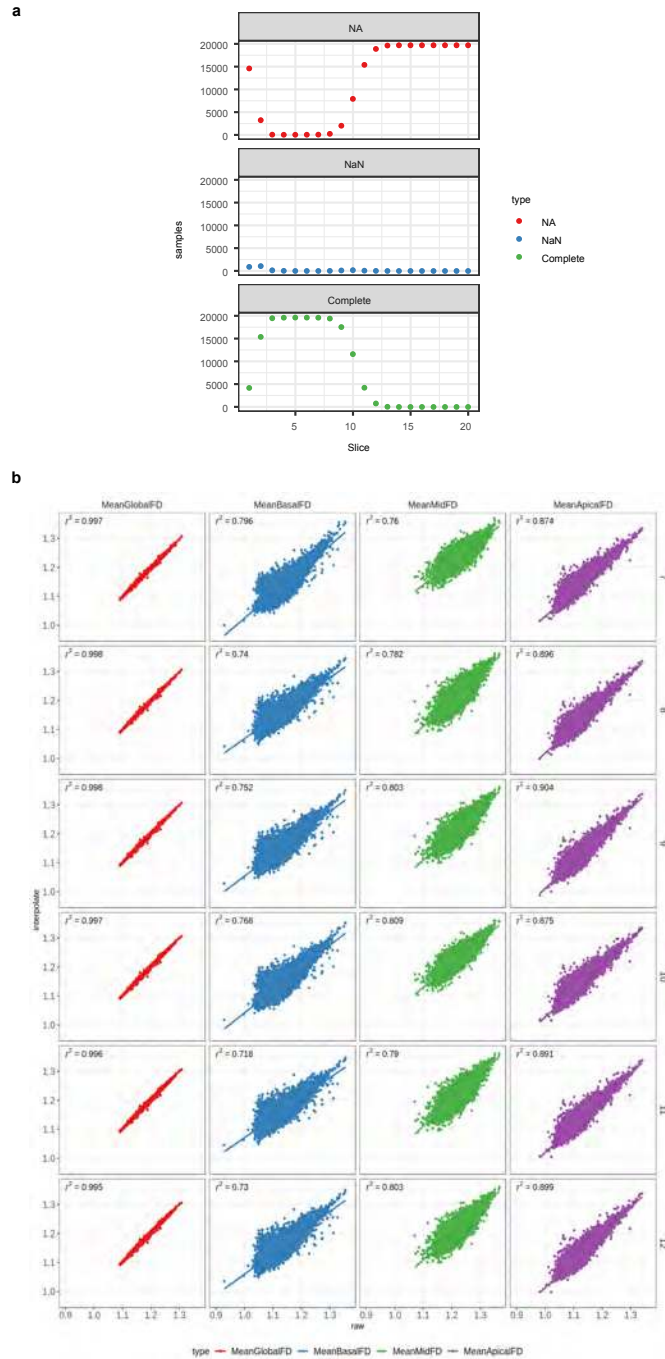
## Extended Figure and table legends

**Supplementary Table 1. Participant characteristics.** Characteristics of all participants that passed quality control in the UK Biobank (UKBB) discovery and replication cohort, the UK Digital heart cohort (UKDH) used as a second replication cohort and the dilated cardiomyopathy (DCM) patients used for disease association. Measurements are depicted in mean  $\pm$  standard deviation. End-systolic volume, end-diastolic volume and mass from CMR are indexed to body surface area. BP, Blood pressure; FD, fractal dimension.

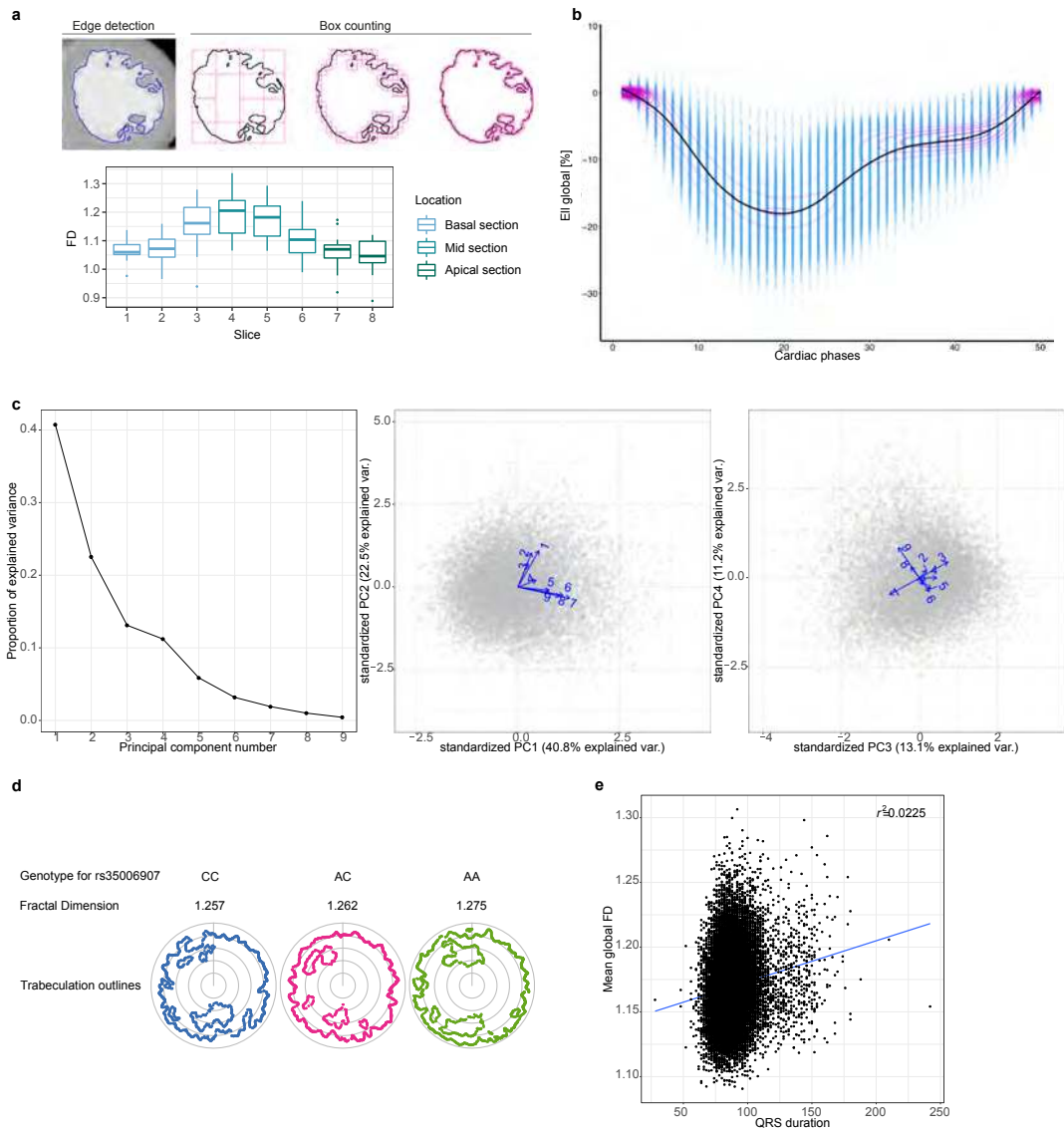
**Supplementary Table 2. Annotations of trabeculation-associated loci.** Overview of the 16 independent loci discovered in the trabeculation GWAS. The genetic variant with the lowest p-value per locus is shown. Annotations: PheWAS: phase 2 PheWAS described by<sup>46</sup> (phenotype reference ID in parentheses; derived from open Targets Genetics v0.3.2), GWAS: GWAS catalogue<sup>23</sup>, eQTLs and tissues: GTEx catalogue v7<sup>25</sup>. Chromosomes (CHR), base pair positions (BP) and nearest protein coding gene (nPCG): GRCh37 (Ensembl GRCh37 Release 95). No entry indicated by -. BP, Blood pressure; LV, left ventricle; AA, atrial appendage.



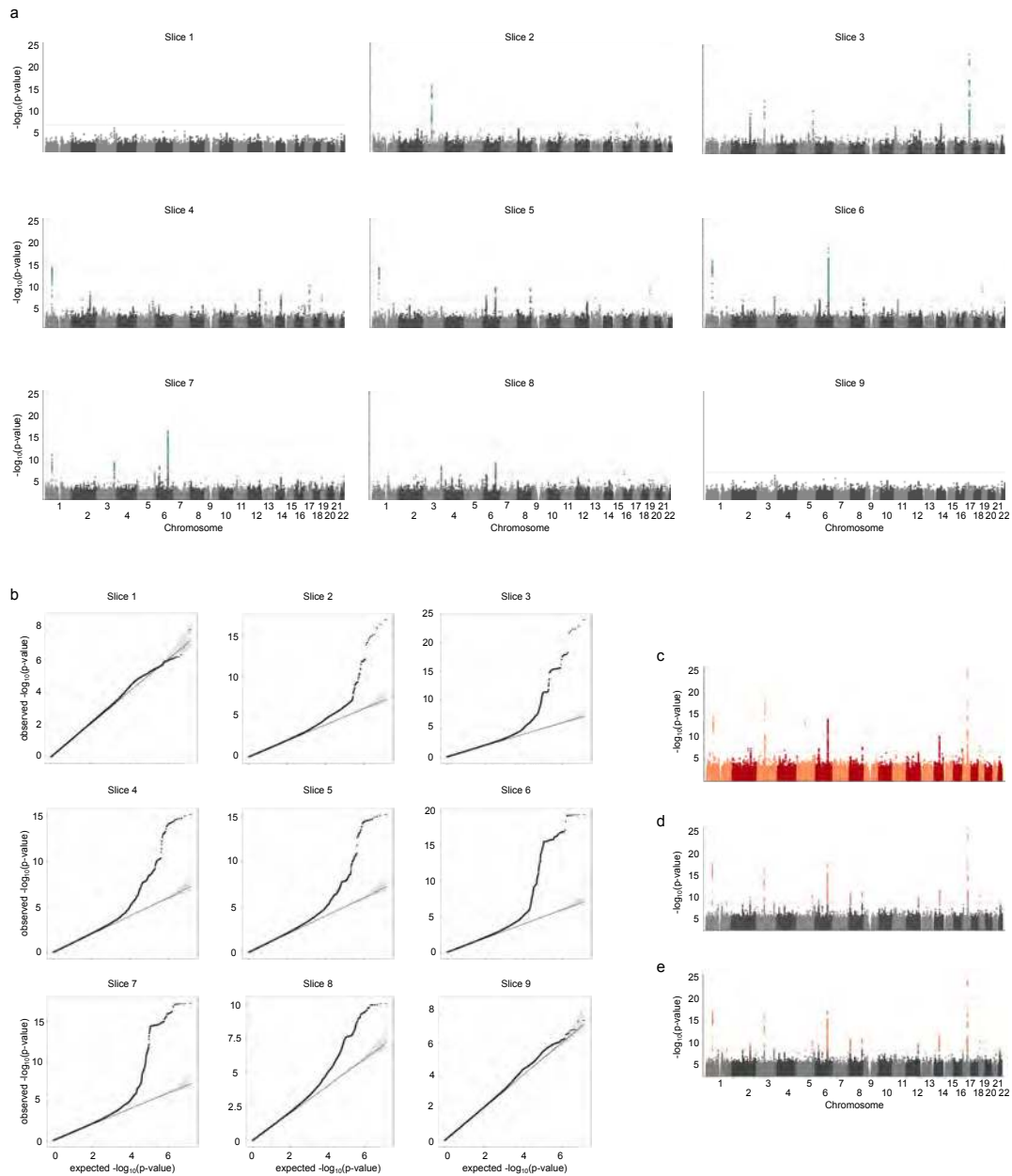
**Extended Data Figure 1. Participant ethnicities in discovery and replication.** Principal components 1 and 2 of the principal component analysis on the combined genotypes of the HapMap III datasets (n=1184) and either a. the discovery cohort UK Biobank (n=19,262; 159,243 independent genetic variants) and b. the UK Digital Heart study (n=2,985; 149,707 independent genetic variants). UK Biobank (a) or UK Digital Heart cohort (b) are depicted in blue, HapMap individuals colored by their ethnicity. Cohort individuals within 1.5 standard deviations distance from the center of the European HapMap individuals (grey) are selected for further analyses.



**Extended Data Figure 2. FD phenotypes.** The upper panels show the distribution of CMR image slices where FD was successfully measured. Missing FD measurements per slice can either arise because a slice was not measured (NA) or the FD estimation failed due to poor image quality or estimated FD failing quality control (NaN). a. depicts the distribution in the UK Biobank samples ( $n=19,761$ ), b. the distribution in the UK Digital Heart samples ( $n=1,901$ ). The lower panels show the correlation between FD summary measures derived from the observed FD slice measurements and interpolated FD measurements per sample. Interpolated FD measurements per sample were derived by using a Gaussian kernel local fit to a different numbers of slice templates, allowing for direct slice comparisons across individuals. Different numbers of slices for interpolation were tested (rows). Columns show different summary measures, either mean FD across all measured slices or mean FD per slice region. c. Linear model of measured  $\sim$  interpolated ( $r^2$ ) of the summary measures between measured and interpolated FD in the UK Biobank samples ( $n=19,761$ ), d. linear model in the UK Digital Heart samples ( $n=1,901$ ).

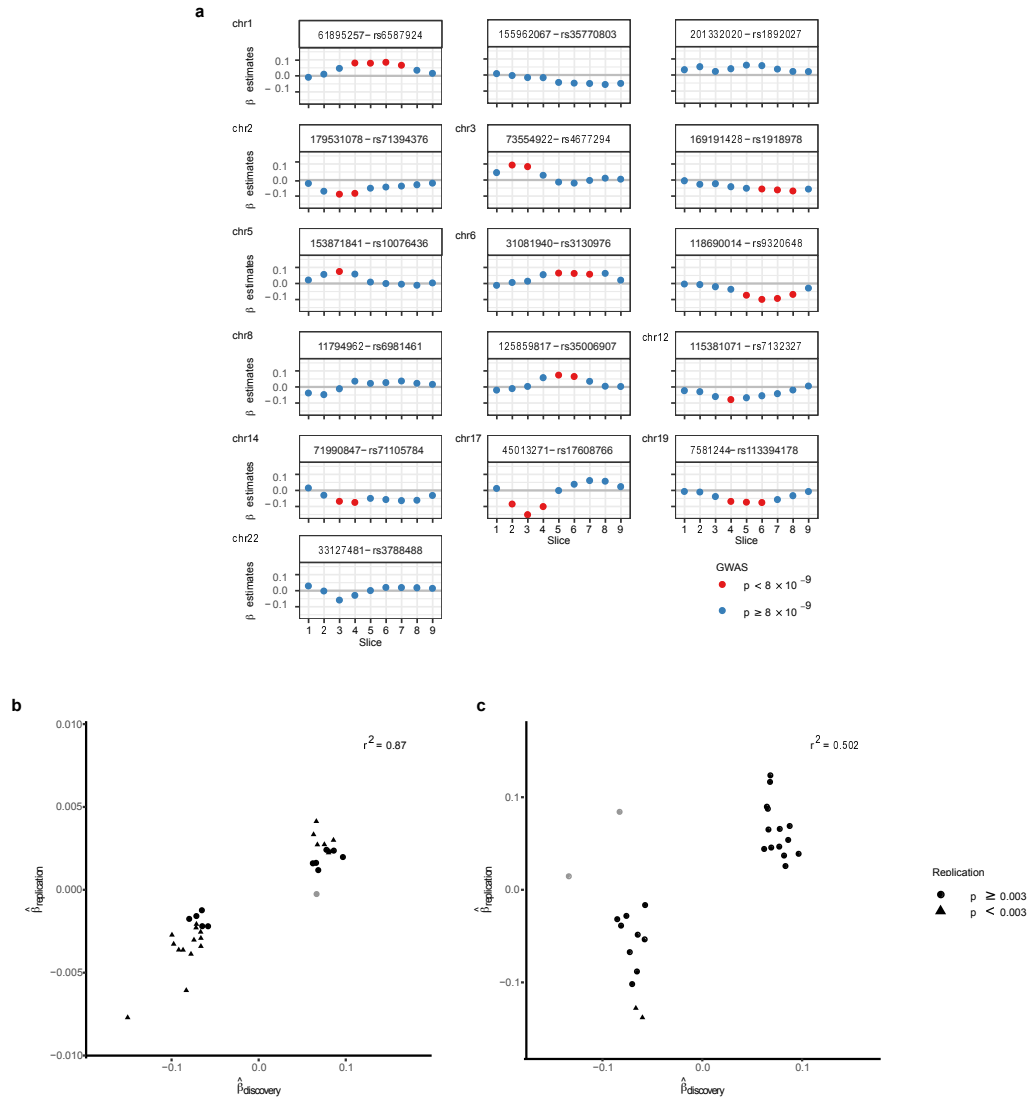


**Extended Data Figure 3. Phenotypes acquisition and processing.** a. Fractal dimension analysis on cardiac CT images. Fractal dimension was calculated using the same method as for CMR, but with manual regions of interest, in a set of gated cardiac computed tomography (CT) images. a. Analogous processing as described in Fig. 1c, using edge detection of the trabeculae and subsequent box-counting across a range of sizes. b. Analogous to Fig. 2a, box-plots of the FD measurements for 20 individuals per slice, colour-coded by cardiac region. The lower and upper hinges in the boxplot correspond to the 25th and 75th percentiles (IQR), the horizontal line in the boxplot the median. The lower/upper whisker extends from the hinge to the smallest/largest value no further than  $1.5 * IQR$ . b. Myocardial strain. Global longitudinal Lagrangian strain at each cardiac phase for all UK Biobank participants with CMR imaging ( $n=26,893$ ). Individual data points shown with a smoothed mean and density contours. c. Principal component analysis of trabeculation phenotypes. Principal component analysis of FD measurements across all 9 slices in the 18,096 individuals of the UK Biobank discovery cohort. Proportion of variance explained of each principal component (left). Biplot of individuals' first and second/third and fourth principal components (grey points) and the corresponding loadings for FD of slices 1-9 as vectors (middle and right). d. Genotype, FD and trabeculation outlines for rs35006907. Representative, registered, trabecular outlines at slice 5 representing the median FD for individuals with homozygous major (blue), heterozygous (pink) and homozygous minor (green) genotype for rs35006907. Pearson correlation of global FD and QRS duration ( $n=18,096$ ). QRS duration phenotype from UKB ID: qrs\_duration\_f12340\_2\_0. The Pearson correlation coefficient is indicated in the upper right corner.

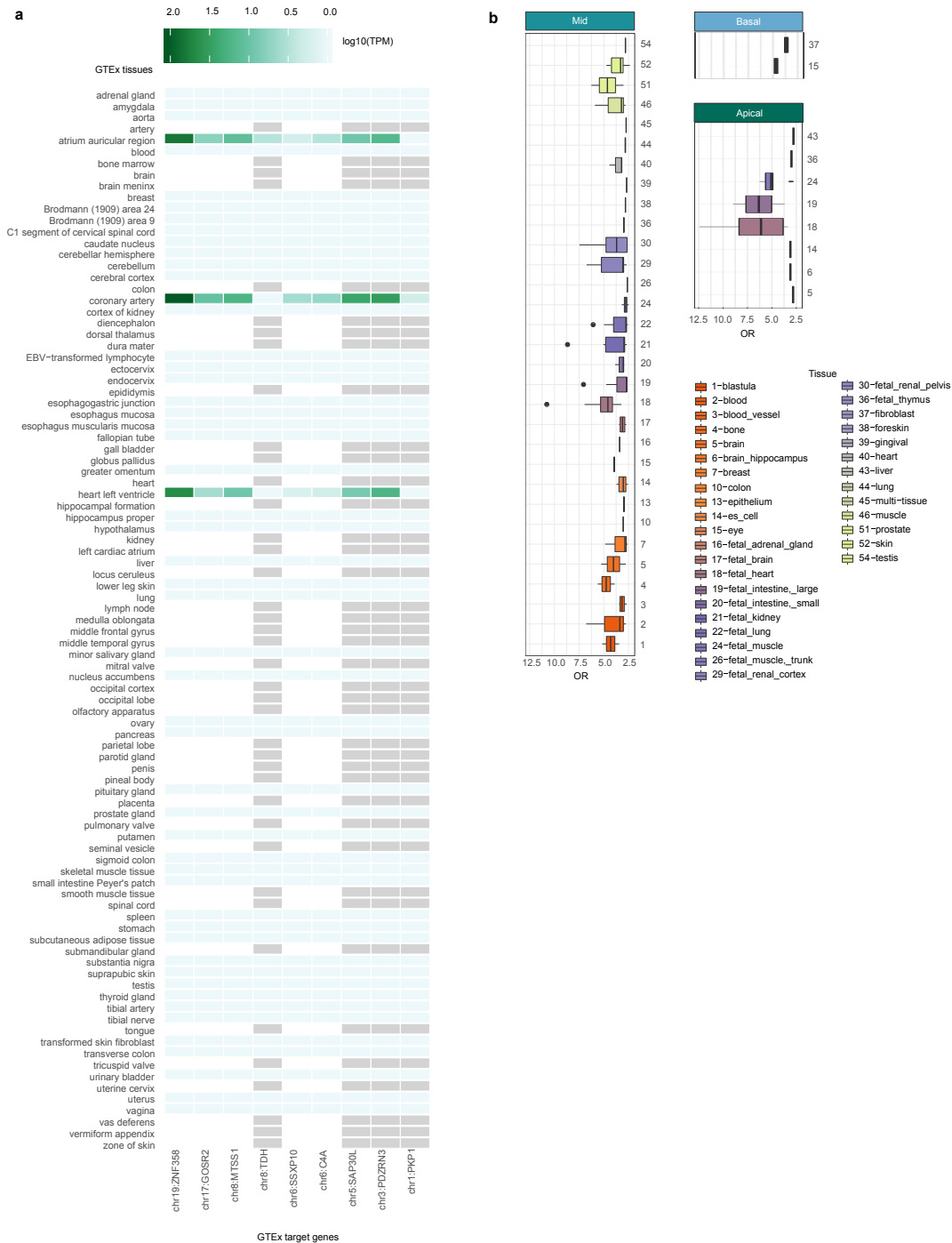


**Extended Data Figure 4. Per-slice FD GWAS and inclusion of additional covariates.** a. Manhattan plots and b. quantile-quantile plots of the independently conducted, nine univariate GWAS on the per-slice FD measurements for 18,096 samples. In the manhattan plots (a), the p-values (derived from linear association  $t$  statistic) were multiplied by the effective number of independent phenotypic tests  $T_{eff} = 6.6$  and  $\min(p_{adj}, 1)$  reported. In the qq-plots, the unadjusted p-values are plotted against equally spaced values in  $[0, 1]$  of the same sample size (expected p-values). The diagonal line starts at the origin and has slope one. The genomic control  $\lambda$  values for each qqplot are: 1.0557, 1.0436, 1.0496, 1.0557, 1.0649, 1.0679, 1.0679, 1.0618, 1.0436.  $\lambda$  were generated with LD score regression, for details see [Supplementary Information Table 2](#). c, d. Manhattan plot based on meta-analysis GWAS (sample size  $n=18,096$ ) with end-diastolic volume of the left ventricle (c.) or myocardial strain (d.) as co-variate. e. Manhattan plot based on meta-analysis GWAS (same as [Fig. 2a](#); shown for comparison). Other co-variates and analysis parameters (as described in methods) were kept the same in a-c. P-values are meta-analysis p-values, not adjusted for multiple testing derived from the transformation of the univariate signed t-statistics (associations on 14,134,301 genetic variants at 16 independent loci from 18,096 samples) and  $\chi^2$  distribution with 9 degrees of freedom. In a. and c., the horizontal grey line is drawn at the level of genome-wide significance:  $p = 5 \times 10^{-8}$ .

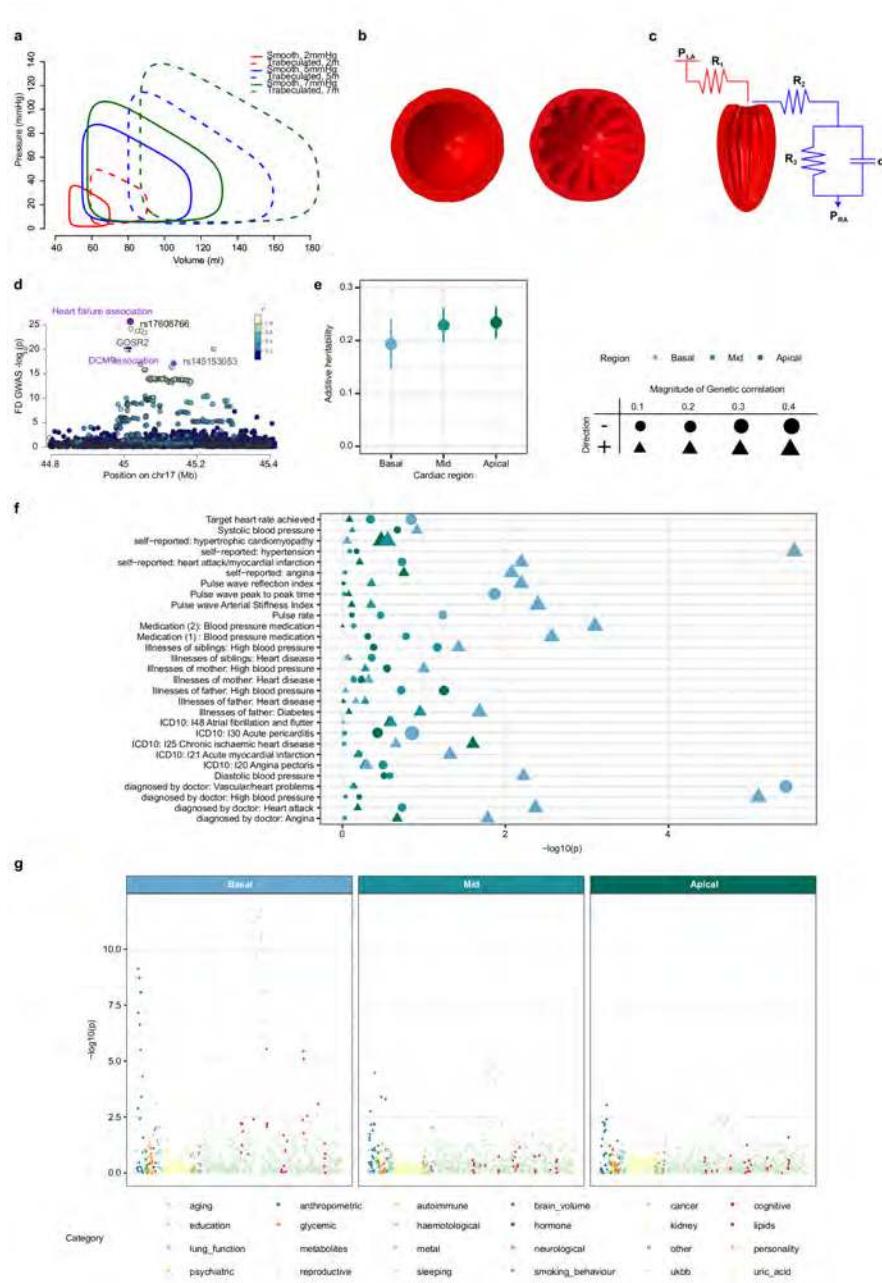




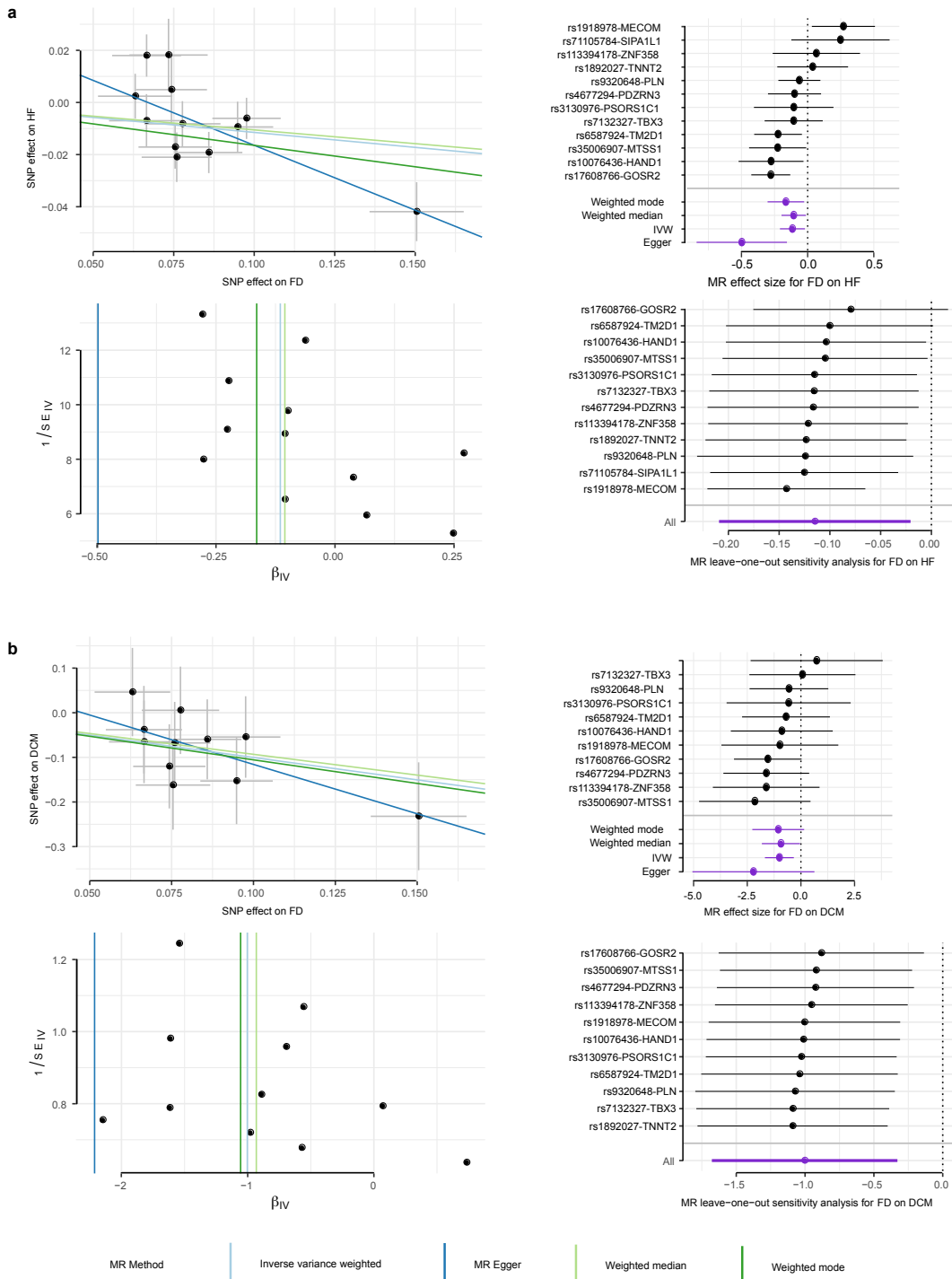
**Extended Data Figure 5. GWAS effect size estimates and replication.** a. Effect size distribution of loci with genetic variant associations of  $p_{ad\ just} = 5 \times 10^{-8}$  in any uni-variate per-slice FD GWAS (sample size  $n=18,096$ ). P-values derived from linear association  $t$  statistic. Distribution shown for each locus (indicated by chromosomal position and lead genetic variant in subplot title) across all slices and effect size colour-coded by p-value of the association. Variants with no  $p_{ad\ just} < 5 \times 10^{-8}$  in the univariate per-slice FD GWAS (all blue) were discovered in the multi-trait meta-analyses. b,c. Effect size estimate concordance in discovery and replication cohorts. For each of the nine uni-variate, per-slice FD GWAS, the effect size estimates of the genetic variants with the smallest p-value for each of the independent loci in the discovery cohort ( $n=18,096$ ) were selected. For some variants, associations passing the GWAS threshold of  $p_{ad\ just} < 5 \times 10^{-8}$  were discovered in more than one of the nine uni-variate GWAS FD slices; for these variants all effect size estimates were selected. Estimates were plotted against the corresponding slice-variant associations in the replication GWAS (b: UK Biobank replication,  $n=6,356$ ; c: UK Digital Heart cohort,  $n=1,029$ ). Non-concordant estimate pairs are depicted in light grey. Effect size estimates passing the Bonferroni-adjusted validation p-value threshold of  $p < \frac{0.05}{16} = 0.003$  are depicted as triangles.  $r^2$  for linear model of  $\hat{\beta}_{\text{discovery}} \sim \hat{\beta}_{\text{replication}}$



**Extended Data Figure 6. Annotation of trabeculation associated loci.** a. Gene expression of GTEx associated genes and tissues. Gene expression in  $\log_{10}$  transcripts per million (TPM) for genes whose expression is associated with trabeculation loci (via GTEx look-up, see [Supplementary Information Table 9](#)). Gene expression values and tissues were downloaded from <https://www.ebi.ac.uk/gxa/home> by querying: gene name AND tissue AND species, i.e. GTEx gene AND heart component AND Homo sapiens. Light grey tiles indicate NA gene expression values for gene/tissue pair. b. Enrichment of trabeculation associated variants in DNaseI Hypersensitive sites for all available tissues in GARFIELD. GARFIELD was used to compute the functional enrichment (odds ratio, OR) of genetic variants associated with the trabeculation phenotypes (GWAS:  $n=18,096$ ,  $p$ -values derived from linear association  $t$  statistic) at  $p < 10^{-6}$  for open chromatin regions. The results across all available studies per tissue are depicted in boxplots. Lower/upper hinges: 25th and 75th percentiles (IQR); horizontal line: median; lower/upper whisker extends from the hinge to the smallest/largest value no further than  $1.5 \times$  IQR.



**Extended Data Figure 7. Biomechanical model, genetic correlation and disease associations.** a. Left ventricular pressure-volume loops from finite-element modelling across range of atrial pressures. Solid and dashed lines indicate smooth and trabeculated ventricles, respectively. Left atrial pressures: 2mmHg (red), 5mmHg (blue) and 7mmHg (green). b. Mid short axis cross sections of the finite element model of the left ventricle, looking towards the apex, at different trabecular complexities. c. The ventricular model was in series with pre-load (red) and after-load circuits (blue) defining left atrial pressure ( $P_{LA}$ ), right atrial pressure ( $P_{RA}$ ), inflow resistance ( $R_1$ ), aortic resistance ( $R_2$ ), peripheral resistance ( $R_3$ ) and vascular capacitance ( $c$ ). Initial parameters were calibrated to approximate UK Biobank observations and the reference model was a trabeculated left ventricle with a  $P_{LA}$  of 5 mmHg. d. FD association p-values (depicted on  $-\log_{10}$  scale, uncorrected for multiple comparisons; estimated by transformation of univariate signed t-statistic with  $\chi^2$  distribution with 9 degrees of freedom; univariate GWAS with  $n=18,096$  samples) for the chr17 *GOSR2* locus and variants associated with mixed aetiology heart failure (HF;  $n_{cases} = 47,309$ ,  $n_{controls} = 930,014$ ) and DCM ( $n_{cases} = 510$ ,  $n_{controls} = 1,136$ ) highlighted in purple. Summary statistics of basal, mid and apical trabeculation GWAS were analysed for genetic correlation with all available summary statistics on LDhub (e-g). e. Additive heritability estimates  $h^2$  for regional summary statistics based on 1,208,036 genetic variants. f. Genetic correlation p-values (based on LD score regression correlation of 1,208,036 genetic variants) by region summarised in LDhub categories (color-code). Heart and cardiovascular phenotypes (see [Supplementary Information Table 16](#) and y-axis in panel g) are depicted in dark red. g. Association p-values of heart and cardiovascular phenotypes with corresponding estimate of genetic correlation (encoded by size). P-values are derived from cross-trait correlation analysis and block-jackknife approach for standard error estimation, see [Supplementary Information Table 16](#); depicted on  $-\log_{10}$  scale, uncorrected for multiple comparisons.



**Extended Data Figure 8. MR analysis of trabeculation on heart failure (HF) and dilated cardiomyopathy (DCM).** a. MR on HF with HF effect size estimates based on  $n_{cases}=47,309$  and  $n_{controls} = 930,014$  in HERMES study. b. MR on DCM with DCM effect size estimates based on  $n_{cases} = 1,136$  and  $n_{controls} = 510$ . For all panels in a. and b. FD effect size estimates from uni-variate GWAS results on  $n=18,096$  samples. Scatter plots (upper left) depict the genetic variant-exposure effect versus the genetic variant-outcome effect. Center values show effect size estimate on FD and DCM, error bars indicate standard error of association test (t-statistic for FD, logistic regression for HF). Forest plots (upper right) show the contribution of each genetic variant to the overall estimate (black; estimated by Wald ratio) and combined as a single genetic instrument (purple; estimated by indicated method) for the four tested MR methods (see legend). Funnel plots (lower left) depict the instrument strength against the causal effect of each instrument as a single IV. Vertical lines indicate the average estimated effect for the tested MR methods. Strong instruments are close to the estimated average effect, while weak instruments spread evenly on both sides. Leave-one-out plots (lower right) show the results of MR analysis (IVW only) where each genetic variant is sequentially excluded and can indicate if there are any single variants that drive the MR results. In right panels, center values mark effect size point estimates, error bars the 95% confidence intervals.

**Extended Data Table 1. Participant characteristics.**

Characteristics	UKBB (discovery)	UKBB (replication)	UKDH	DCM
Participants	18,096	6,536	1,129	510
Age (years)	55 ± 7	55 ± 7	41 ± 13	54 ± 15
Body surface area (m <sup>2</sup> )	1.90 ± 0.21	1.90 ± 0.22	1.85 ± 0.20	1.99 ± 0.28
Sex (m/f)	9402/8694	3378/3158	621/508	357/153
<b>Haemodynamics</b>				
Systolic BP (mmHg)	137 ± 18	136 ± 18	120 ± 14	122 ± 31
Diastolic BP (mmHg)	81 ± 10	82 ± 10	79 ± 9	73 ± 20
Heart rate (beats/minute)	62 ± 10	63 ± 11	62 ± 21	75 ± 24
Cardiac output (l/min)	5.5 ± 1.3	5.4 ± 1.3	6.0 ± 2.0	7.0 ± 3
<b>Left ventricular Volumetry</b>				
Ejection fraction (%)	59.5 ± 6.0	59.6 ± 6.3	65.3 ± 6.9	40 ± 17
End-systolic volume (ml/m <sup>2</sup> )	31.9 ± 8.6	31.6 ± 8.9	28.2 ± 7.9	79 ± 43
End-diastolic volume (ml/m <sup>2</sup> )	78.3 ± 13.9	77.7 ± 14.0	81.2 ± 14.4	127 ± 52
Mass (g/m <sup>2</sup> )	45.2 ± 8.5	44.9 ± 8.3	62.8 ± 15.3	91 ± 39
Mean global FD	1.169 ± 0.028	1.169 ± 0.029	1.215 ± 0.038	1.218* ± 0.075

Extended Data Table 2. Annotations of trabeculation-associated loci.

CHR	BP	SNP	nPCG	PheWAS	GWAS	eQTLs	Tissues
1	61895257	rs6587924	<i>TM2D1</i>	-	-	-	-
1	155962067	rs35770803	<i>ARHGEF2</i>	-	-	-	-
1	201332020	rs1892027	<i>TNNT2</i>	-	-	<i>PKP1</i>	Skeletal muscle, adipose
2	179531078	rs71394376	<i>TTN</i>	-	-	-	-
3	73554922	rs4677294	<i>PDZRN3</i>	-	-	<i>PDZRN3</i> , <i>PDZRN3-ASI</i>	LV, artery, AA, aorta
3	169191428	rs1918978	<i>MECOM</i>	-	-	-	-
5	153871841	rs10076436	<i>HAND1</i>	-	-	<i>SAP30L</i>	Transformed fibroblasts
6	31081940	rs3130976	<i>PSORS1C1</i>	84 phenotypes	Nephropathy, asthma, CLE	adult 18 genes	37 tissues including LV and AA
6	118690014	rs9320648	<i>PLN</i>	Pulse rate, diastolic BP (4195, 102)	-	<i>SSXP10</i> , <i>CEP85L</i> , <i>SLC35F1</i>	13 tissues including AA
8	11794962	rs6981461	<i>DEFB136</i>	-	C-reactive protein	28 genes	35 tissues including LV and AA
8	125859817	rs35006907	<i>MTSS1</i>	-	Ejection fraction, fractional shortening, LV internal dimension in systole and diastole, relative wall thickness, LV internal dimension, atrial fibrillation	<i>MTSS1</i> , <i>LINC00964</i>	LV, lung, AA
12	115381071	rs7132327	<i>TBX3</i>	-	Global electrical heterogeneity phenotypes, QRS complex, QRS duration, PR segment, PR interval	-	-
14	71990847	rs71105784	<i>SIPA1L1</i>	-	QRS complex, QRS duration, mitral valve prolapse	-	-
17	45013271	rs17608766	<i>GOSR2</i>	Systolic BP, hypertension (4080, 6150_4, 6150_100, 20002_1065, 103)	Systolic BP, QRS duration, pulse pressure, BP, aortic root size, atrial fibrillation	<i>RPRML</i> , <i>GOSR2</i> , <i>CDC27</i> , <i>RP11-63A1.2</i> , <i>RP11-156P1.3</i>	Skeletal muscle, testis, adrenal gland
19	7581244	rs113394178	<i>ZNF358</i>	-	-	<i>ZNF358</i>	AA
22	33127481	rs3788488	<i>TIMP3</i>	-	-	-	-

# Genetic and functional insights into the fractal structure of the heart

Hannah V Meyer<sup>1,11,+</sup>, Timothy JW Dawes<sup>2,3,+</sup>, Marta Serrani<sup>4,9,+</sup>, Wenjia Bai<sup>5</sup>, Pawel Tokarczuk<sup>2</sup>, Jiashen Cai<sup>6,7</sup>, Antonio de Marvao<sup>2</sup>, Albert Henry<sup>12,15</sup>, R Thomas Lumbers<sup>12,13,14</sup>, Jakob Gierten<sup>16,17</sup>, Thomas Thumberger<sup>17</sup>, Joachim Wittbrodt<sup>17</sup>, James S Ware<sup>2,3,18</sup>, Daniel Rueckert<sup>5</sup>, Paul M Matthews<sup>8</sup>, Sanjay K Prasad<sup>3,18</sup>, Maria L Costantino<sup>9,\*</sup>, Stuart A Cook<sup>2,3,6,10,\*</sup>, Ewan Birney<sup>11,\*</sup>, and Declan P O'Regan<sup>2,\*,</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Simons Center for Quantitative Biology, USA

<sup>2</sup>MRC London Institute of Medical Sciences, Imperial College London, UK

<sup>3</sup>National Heart and Lung Institute, Imperial College London, UK

<sup>4</sup>Department of Chemical Engineering and Biotechnology, University of Cambridge, UK

<sup>5</sup>Department of Computing, Imperial College London, UK

<sup>6</sup>Duke–National University of Singapore, Singapore

<sup>7</sup>Department of Internal Medicine, Singapore Health Services, Singapore

<sup>8</sup>Department of Brain Sciences, Department of Medicine and UK Dementia Research Institute at Imperial College London, UK

<sup>9</sup>Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Italy

<sup>10</sup>National Heart Centre Singapore, Singapore

<sup>11</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), UK

<sup>12</sup>Institute of Health Informatics, University College London, UK

<sup>13</sup>Health Data Research UK, London, UK

<sup>14</sup>Barts Heart Centre, St. Bartholomew's Hospital, London, UK

<sup>15</sup>Institute of Cardiovascular Science, University College London, UK

<sup>16</sup>Department of Pediatric Cardiology, University Hospital Heidelberg, Heidelberg, Germany

<sup>17</sup>Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany

<sup>18</sup>Cardiovascular Research Centre, Royal Brompton & Harefield NHS Foundation Trust, London, UK

<sup>+</sup>These authors contributed equally to this work

<sup>\*</sup>These authors jointly supervised this work

✉ Corresponding authors

## Contents

<b>1</b>	<b>Supplementary Note</b>	<b>1</b>
1.1	Mendelian randomisation . . . . .	1
	MR assumptions and biases • MR methods • Limitations in this study	
<b>2</b>	<b>Supplementary Methods</b>	<b>3</b>
	<b>References</b>	<b>10</b>
2.1	Supplementary Tables . . . . .	14
2.2	Supplementary Figures . . . . .	22

## 1 Supplementary Note

### 1.1 Mendelian randomisation

Mendelian randomization (MR) studies can be thought of as randomised control trials where genetic variants are used as instrumental variables (IV) to infer the effect of an exposure variable on the outcome variable. The randomization is described as Mendelian based on Mendel's laws of inheritance, i.e. the selection of alleles that an individual receives for a given genetic variant occurs at random during meiosis. This has two consequences, first the alleles are expected to be random with respect to confounders and second, they are causally upstream of the exposure traits. The effect of the exposure on outcome can then be inferred as the ratio of the genetic effect on the outcome over the genetic effect on the exposure.

In principle, there are two main types of MR analysis based on the study cohorts.<sup>47</sup> In one-sample MR studies both the genetic variant-exposure and -outcome effect size estimates are obtained from the same cohort. In two-sample MR studies, these effect size estimates are obtained in independent cohorts. In times of large biobanks where thousands of individuals are measured for hundreds of traits, intermediate set-ups exist, where there is sample overlap between the cohort in which the exposure association was measured and the cohort of the outcome association.

### 1.1.1 MR assumptions and biases

The basis of MR studies is ‘vertical’ pleiotropy i.e. the genetic variants under investigation are associated with both traits considered because one trait is causal to the other trait. In addition, Mendelian randomisation studies make the following main assumptions:

- the instrument is associated with the exposure (IV1 assumption),
- the instrument only influences the outcome through exposure and not through any other pathway (IV2 assumption),
- the instrument is not associated with confounders (IV3 assumption).

However, there is a second type of pleiotropy, ‘horizontal’ pleiotropy, where the genetic variant is associated with the outcome through confounders or a pathway other than the exposure. ‘Horizontal’ pleiotropy is a violation of the IV2 and IV3 assumption and has to be addressed when conducting MR analysis as it can lead to incorrect inference of the causal effects.<sup>48,49</sup> In addition, one has to further consider the nature of ‘horizontal’ pleiotropy: if the mean effect of the ‘horizontal’ pleiotropy is zero, it is considered balanced and will not effect the effect size estimate of the causal effect (but can effect its standard error). If the mean effect is unequal to zero, it is considered directional and the extend of pleiotropy can be estimated (see MR Egger below).

If there is sample overlap between the exposure and outcome cohort, the overlap can lead to the correlation of the uncertainty in the genetic variant-exposure association and the genetic variant-outcome association. This can cause a bias of the causal effect estimates towards the confounded observational association (especially for weak instruments due to Winner’s curse<sup>50</sup>). IVs strongly correlated with the exposure (in practice defined as F statistic > 10) are less prone to this bias even in cohorts with overlapping samples.<sup>51</sup>

### 1.1.2 MR methods

A number of methods exist that can address violations to the IV assumptions outlined above. In the following, a selection of four methods used in this study are described in brief (for a comprehensive overview see<sup>48-50</sup>):

- **Inverse-variance weighted (IVW) linear regression.** If all IV are valid, IVW can be used to obtain an unbiased causal estimate of the exposure on the outcome. In IVW, the effect of the exposure on the outcome is estimated by linear regression of the effect size estimates of the genetic variant-exposure association on effect size estimates of the genetic variant-outcome association. The contribution of each IV to the overall effect is weighted by the inverse of the variance of the genetic variant-outcome effect and the intercept of the linear regression is constrained to pass through zero (no horizontal pleiotropy/balanced horizontal pleiotropy).<sup>52</sup>
- **MR Egger.** MR Egger works similar to IVW with the exception that the intercept is not restrained to pass through zero i.e. it allows to adjust for (unbalanced) pleiotropic effects. However, the effect estimates are only unbiased if the genetic variant-exposure associations and the pleiotropic effects are not correlated, i.e. the instrument strength is independent of direct effect (InSIDE assumption).<sup>53</sup>
- **Weighted median-based estimator.** The median-based estimator makes the assumption that the majority of IV are valid instruments. It is based on the ordered effect size estimate ratios for all IV-exposure to IV-outcome associations, weighted by standardised, inverse of the standard error of the Wald ratio estimated by the delta method (analogues to IVW). The weighted median-based estimator allows for unbalanced pleiotropy of the IVs and unlike MR Egger does not rely on the InSIDE assumption.<sup>54</sup>
- **Weighted mode-based estimator.** The mode-based estimator works based on cluster selection. It first clusters the IV into groups based on the similarity of the effect size estimates, and then selects the effect size estimate from the cluster with the largest number of IVs. The mode-based causal effect estimate is valid if the IVs in the largest cluster are valid.<sup>55,56</sup>

In addition to these methods, there are a number of statistics that can be used to evaluate the validity of a given method. Both the IVW and MR Egger make the no measurement error (NOME) assumption, i.e. they consider the variance of the



genetic variant-exposure association as negligible.<sup>52</sup> In IVW, the presence of large measurement error (violation of NOME) can lead to weak instrument bias. The strength of the instruments can be assessed with the F-statistic and as above for overlapping sample sizes, IVs which strongly correlated with the exposure (in practice defined as F statistic > 10) are less prone to this bias.<sup>57</sup> However, the F statistic is only a proxy for the true, but unknown parameter of interest, the F parameter. In addition to computing the cohort F statistic, one can also estimate a lower bound on the true F parameter as described in [58, Appendix A3]. For MR Egger and NOME violation, the causal effect size estimates can suffer from regression dilution bias which attenuates the causal effect estimate towards the null,<sup>59</sup> The MR Egger  $I^2$  statistic can be calculated to evaluate the magnitude of dilution bias, e.g. an  $I^2$  of 90% will lead to a 10% underestimation of the effect size.<sup>59</sup> The mode-based estimator can be implemented with or without the assumption of NOME.<sup>55,56</sup>

MR analysis can be used to infer whether there is a causality between exposure and outcome and which direction the effect takes i.e. exposure-outcome or outcome-exposure. With the MR Steiger test, the directionality can be assessed based in the absolute correlations of the genetic variants with the exposure and outcome (with Steiger's Z-test for correlated correlations within a population). Based on the Z statistic and the associated Steiger p-value one can then test at a predefined level  $\alpha$  if one accepts the causal association for the model.<sup>60</sup>

### 1.1.3 Limitations in this study

We conducted MR analysis on the effect of FD on heart failure and dilated cardiomyopathy in the HERMES (The Heart Failure Molecular Epidemiology for Therapeutic Targets) and a DCM cohort, respectively (Figure 4c,d). To address potential biases in our analysis arising from potential violations in the instrumental variable assumptions we have conducted formal analysis to check for weak instrument bias, pleiotropy and dilution bias. We find evidence for weak pleiotropy effects in the MR on heart rate, but not on DCM (Supplementary Table 12). We only found very weak dilution bias (around 1.7%). We confirmed the direction of effect with the Steiger directionality test. To our knowledge, there is no sample overlap between individuals of the DCM study and the UK Biobank. However, there is a potential, minimal sample overlap between UK Biobank FD GWAS (18,096 samples) and cases in HERMES analysis. HERMES analysed a total of 47,309 heart failure cases, of which 6,504 were derived from UK Biobank (see Supplementary Tables document, sheet HERMES cohort characteristics). Of those heart failure cases in the UK Biobank a maximum of 185 samples had their MR taken, i.e. the potential overlap between samples of the FD GWAS and the cases in the HERMES Heart failure meta-analysis is 185/47,309=0.4%. We estimated the effect this minimal overlap could have (potential for weak instrument bias) by computing the F statistics and lower bound estimation of the F parameter (Supplementary Table 11) and find that there is no sufficient evidence for weak instrument bias.

## 2 Supplementary Methods

All analysis in R conducted with R version > 3.6.0. All analyses in this study can be found here:

<https://github.com/ImperialCollegeLondon/fractalgenetics/>.

### Phenotyping

**Participants:** For UK Biobank, approximately 500,000 community-dwelling participants aged 40–69 years were recruited across the United Kingdom between 2006 and 2010.<sup>61</sup> Baseline summary characteristics of the cohort can be viewed on the UK Biobank data showcase (<http://www.ukbiobank.ac.uk>). Since 2014, a subset of participants are being recalled for CMR and a total of 19,701 consecutive CMR datasets were available at the start of this study, which were processed for our discovery cohort (May 2018 data release) and 7,192 for our validation cohort (December 2018 data release). 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 40616. The second validation cohort was drawn from the UK Digital Heart Project - a single-centre prospective study recruiting 2000 healthy volunteers by advertisement between February 2011 and July 2016 at the MRC London Institute of Medical Sciences. All subjects provided written informed consent for participation in the study, which was also approved by the National Research Ethics Service (09/H0707/69).

**Assessment centre:** For both populations an equivalent CMR protocol was followed to assess LV structure and function using conventional two-dimensional retrospectively-gated cine imaging on a 1.5T magnet.<sup>62,63</sup> A contiguous stack of images in the LV short-axis plane from base to apex was used for volumetric analysis and trabecular phenotyping. Images were curated on open-source databases.<sup>64,65</sup> Participants also underwent a resting 12 lead electrocardiogram which was automatically analysed using proprietary software (CardioSoft, GE Healthcare).

**Image analysis:** Segmentation of the short-axis cine images was performed using a fully convolutional network, a type of deep learning neural network, which predicts a pixelwise image segmentation by applying a number of convolutional filters onto each input image. The accuracy of image annotation using this algorithm is equivalent to expert human readers.<sup>66</sup> Label maps

were derived for all images in the cardiac cycle and LV volumes and mass were calculated according to standard guidelines,<sup>67</sup> and then indexed to body surface area (BSA) which was computed according to the Mosteller formula:  $\sqrt{Weight \times Height}/60$ , with weight in kg and height in cm. Lagrangian strain was calculated automatically using a validated 2D B-spline free-form deformation registration technique to estimate motion between consecutive cine frames.<sup>68</sup>

Fractal dimension (FD) - a scale-invariant measure of trabecular complexity - was derived using a fully-automated algorithm executed from Matlab (Mathworks, Natick, MA) using custom-written code (autoFD, in GitHub repository at [automated-fractal-analysis](#)). Short-axis CMR images were pre-processed with bicubic interpolation to 0.25 mm x 0.25 mm pixels to enable consistent analysis between subjects acquired at different native resolutions. For each slice, a region of interest was defined within the midwall of the LV myocardium between the automated endocardial and epicardial segmentations. Subsequent image processing consisted of bias-field correction using histogram stretching, applying a region-based level-set algorithm and then binarization of the blood pool and myocardium.<sup>69</sup> The trabecular borders were then detected using a Sobel filter and FD was calculated using a standard box-counting method in which the target image is overlain by a grid of known box size and the number of boxes containing non-zero image pixels is recorded. This process is repeated with box sizes between two pixels and 45% of the image size. Fractal dimension is defined as the negative gradient of an ordinary least-squares fit line to the logarithm of box size and box count. The FD values from all slices were interpolated using a Gaussian kernel local fit to a nine-slice template to allow comparison across subjects (see [Extended Data Fig 2](#) and [fractal-analysis-processing](#) in the GitHub repository). The feasibility of extending fractal analysis to cardiac computed tomography (CT) was assessed on a publicly-available dataset of CT images in 20 individuals.<sup>70</sup>

For visual comparison of trabecular borders across individuals and genotypes, images were aligned by co-registering each ventricular segmentation to a common coordinate space and applying the same transformation to the corresponding greyscale image. For each slice, the trabecular and outer myocardial borders were extracted and their common center computed. The pixel positions of the edges were converted to radial coordinates  $(\theta, r)$ , with pre-defined step size of  $\frac{\pi}{2000}$ . The outer myocardial border was then registered to a circle with radius one. The radial translocation at each angle required for the outer border registration was then applied to trabecular outline, making the outlines comparable across individuals (in GitHub repository at [fractal-analysis-processing](#) and [UK-Biobank-segmentations](#)).

Pressure-volume loop analysis enables a detailed interpretation of cardiac physiology and ventricular work, but conventionally requires invasive catheterisation to obtain absolute pressure measurements which is not possible for population studies. Recently, a model-based framework combining non-invasive pressure measurements with CMR volumetry has been validated in a porcine model.<sup>71</sup> Here we take advantage of consecutive CMR imaging and peripheral pulse-wave analysis (Vicorder, Wuerzburg, Germany) for dynamic volumetric analysis and central pressure estimation respectively, to non-invasively model left ventricular pressure-volume relationships throughout the cardiac cycle. Peak systolic pressure and maximal aortic distension on axial cine imaging were assumed to be synchronous allowing LV volume at peak-systolic pressure to be assessed. The indexed volume difference between end-diastole and end-systole over a single cycle was defined as stroke volume index, and over a minute as cardiac index. Indexed systemic vascular resistance was defined as the difference between mean arterial pressure and central venous pressure divided by cardiac index. In the absence of invasive catheter data, a value of 5mmHg for central venous pressure was assumed. LV diastolic pressures were assumed to be normal and rise during diastole from 4mmHg to 8mmHg at end-diastole.<sup>72</sup> To understand the association of FD with pressure-volume dynamics, the mean ventricular FD value was associated with pressure and volume measurements during diastole and systole using linear regression, while controlling for contractility, systemic vascular resistance, trabecular mass and heart rate. Using this model, pressure-volume loops at comparable FD to the finite element model were plotted (in [pv-loops](#) of GitHub repository).

## Genotyping

### Discovery cohort

UK Biobank genotypes release version 3 (in GRCh37 coordinates) were used in the genetic association studies, for computing genetic principal components and for the Mendelian randomisation analyses. First, only unrelated or distantly related individuals were selected for further analysis based on the estimated identity of descent (IBD) provided by UK Biobank (via 'ukbgene\_rel\_key.enc'). Selection of individuals from families is optimised to retain as many unrelated individuals as possible in the study i.e. in family trios only parents would be considered for analysis. For computation of genetic principal components, genotypes were formatted into plink binary format,<sup>73</sup> LD pruned (flag '-indep-pairwise 50kb 1 0.8') and a minor allele frequency (MAF) threshold of 1% applied. The filtered dataset was used with flashpca version 2,<sup>74</sup> to compute the first 50 principal components of the UK Biobank genotypes. Population substructures arising due to different ethnic origins of samples were examined by comparing the UK Biobank genotypes to genotypes from the HapMap Phase III study,<sup>75</sup> for four ethnic populations (with subpopulations, [Extended Data Fig 1a](#)). Download of reference data sets, fusion of UK Biobank and Hapmap data sets and PCA selection was done as described in plinkQC.<sup>76</sup> Individuals that clustered with the main cluster of the PC1/PC2 plot, which was also the position of the main cluster of HapMap III individuals of European ancestry were kept for further analyses,

as is standard in GWAS analysis to model a well mixed population. The genotypes of the remaining unrelated individuals were filtered for genetic variants that passed a MAF threshold of 0.1% and an imputation INFO threshold of  $> 0.4$ . For association analysis this was achieved by providing a variant ID list with variant MAF  $> 0.001$  to BGENIE. After genotyping and phenotyping quality control, the discovery cohort was composed of 18,096 individuals and 14,134,301 genetic variants. All relevant analysis in [UK-Biobank](#) and [ancestry](#) in the GitHub repository.

### **Replication cohorts**

**UK Biobank replication cohort.** The UK Biobank released an additional 7,192 CMR images on in December 2018. We applied the same genotyping and phenotyping quality control as described above for the UK Biobank discovery cohort and obtained a replication cohort of 6,536 individuals and 14,069,398 genetic variants.

**UK Digital Heart cohort.** UK Digital Heart project genotyping and genotype calling were carried out at the Genotyping and Microarray facility at the Wellcome Sanger Institute, UK and Duke-NUS Medical School, Singapore. Genotypes were assessed in five batches using Illumina HumanOmniExpress-12v1-1 (Sanger, two batches), Illumina HumanOmniExpress-24v1-0 (Duke-NUS, two batches) and Illumina HumanOmniExpress- 24v1-1 chips (Duke-NUS). Genotypes were called via the GenCall software.<sup>77</sup> For batches run on the same platform, genotype signals were combined and called in a single analysis, leading to three independent genotype batches. In order to avoid batch effects in genotype calling based on the probe sequences, probes targeting the same genotypes were checked for the concordance of the capture sequence. Genotyping probes common to all three platforms were selected and a common genotype dataset generated using PLINK v1.9,<sup>73</sup> with plinkQC<sup>76</sup> applied to assess the quality of the genotyping on a per-individual and per-marker level. In summary, the per-individual quality control included the identification of individuals with discordant sex information, missing genotype rates (more than 3% of genotypes not called) and heterozygosity rate outliers (three standard deviations outside of the mean heterozygosity rate). Population substructures arising due to different ethnic origins of samples were examined by comparing the sample genotypes to genotypes from the HapMap Phase III study,<sup>75</sup> for four ethnic populations (with subpopulations, [Extended Data Fig 1b](#)). Individuals that clustered with the main cluster of the PC1/PC2 plot, which was also the position of the main cluster of HapMap III individuals of European ancestry were kept for further analyses, as is standard in GWAS analysis to model a well mixed population. The per-marker quality control included filtering of genotypes with missing call rate in more than 1% of the samples and genotypes which significantly deviate from Hardy-Weinberg equilibrium (HWE,  $p < 0.001$ ). After removing samples and genotypes that failed quality control, we confirmed that any pattern of missing genotype information was not batch-specific. To analyse these patterns, each pair-wise combination of batches was treated as a case-control set-up and the differential missingness of genotypes computed. All variants with significant differential missingness ( $p < 10^{-5}$ ) were removed from the dataset. After quality control, genotypes were phased and imputed to the combined 1000 Genomes<sup>78</sup> and UK10K<sup>79</sup> reference panel using SHAPEIT (version 2.r727)<sup>80</sup> and IMPUTE2 (version 2.3.0).<sup>81</sup> The window size for phasing was set to 2Mb, and the number of conditioning states per genotype to 200. The imputation interval was set to 3Mb, with a buffer region of 250kb on either side of the analysis interval. The effective population size was set to 20,000 and the number of reference haplotypes to 1,000. For other non-specified parameters the default values were used. Analysis of UK Biobank cohort in [UK-Biobank-ancestry](#) and [UK-BioBank-phenotypes](#), of UK Digital Heart cohort in [digital-heart-genotypes](#) and [digital-heart-phenotypes](#) in the GitHub repository).

### **LD reference**

We generated a LD reference for downstream filtering of association results based on the quality controlled genotypes described above (non related, European, MAF  $> 0.001$ , INFO  $> 0.4$ ). For each variant we computed all variants with  $r^2 > 0.05$  in a LD window of 250 kb using the following plink version 2 command: 'plink2 -bfile input -indep-pairwise 50kb 1 0.8'.

### **Association analysis**

Analysis are in [UK-Biobank-association](#) for UK Biobank and in [digital-heart-association](#) for the UK Digital Heart cohort on GitHub.

**Univariate association:** Genetic association of trabecular FD for samples passing genotype and phenotype quality control (UK Biobank discovery: 18,096, UK Biobank replication: 6,536, UK Digital Heart project: 1,129) were conducted using BGENIE v1.3 (<https://jmarchini.org/bgenie>).<sup>82</sup> In the UK Biobank discovery cohort, we conducted univariate GWAS on the interpolated FD measurements per slice (9 independent GWAS) and the mean FD measurements per ventricular region (3 independent GWAS of basal, mid-ventricular and apical FD), by fitting an additive model of association at each variant based on the genotype dosage of the imputed genotypes. In addition, we included sex, age, height, weight, BMI and principal components of the genotypes as co-variates in the model. To test for the effect of ventricular size, we conducted analogous GWAS in the discovery cohort, where we additionally included left ventricular end-diastolic volume and longitudinal strain as a co-variates. The univariate GWAS were adjusted for multiple-hypothesis testing by estimating the effective number of

phenotype-association tests conducted<sup>83</sup>. The effective number of tests is estimated based on the eigenvalues  $\mathbf{u}$  of the empirical trait-by-trait correlation matrix  $\mathbf{C}$ . The effective number of tests is  $T_{eff} = \frac{\sum_{i=1}^n \sqrt{u_i^2}}{\sum_{i=1}^n u_i}$ , where  $n$  is the number of GWAS conducted i.e.  $n = 9$  and  $n = 3$  for per-slice and per-region GWAS, respectively. The p-values of these GWAS were multiplied by the effective number of tests and the  $\min(p_{adjust}, 1)$  reported. Associations were considered significant if  $p_{adjust} < 5 \times 10^{-8}$ . The association tests in both replication cohorts used the same analysis setting and parameters as in the discovery cohort.

**Multi-trait meta-analysis:** We used an approximate, multi-trait meta-analysis<sup>84</sup> based on the univariate signed  $t$ -statistics of the 14,134,301 genetic variants for the nine per-slice FD GWAS. The multivariate test statistic is computed as  $t_{meta} = \mathbf{t}_i^T \mathbf{V}^{-1} \mathbf{t}_i$ , where  $\mathbf{t}_i$  is the vector of the signed  $t$ -values of variant  $i$  for the nine FD measurements,  $\mathbf{t}_i^T$  is a transpose of vector  $\mathbf{t}_i$ ,  $\mathbf{V}^{-1}$  is the inverse of the trait-by-trait correlation matrix  $\mathbf{V}$ .  $\mathbf{V}_{i,j}$  for each trait-trait pair is the correlation over the 14,134,301 estimated signed  $t$ -values of the two traits.  $t_{meta}$  is approximately chi-square distributed with 9 degrees of freedom and tests the null hypothesis that the genetic variant tested does not affect any of the nine traits. We used the LD reference panel described above to clump all SNPs with  $p < 5 \times 10^{-8}$  into one SNP per locus, keeping the SNP with the lowest  $p$ -value.

**Validation analysis:** The effect size estimates of the FD GWAS of the discovery cohort (UK Biobank) and the corresponding associations of the validation cohorts (UK Biobank validation cohort with 6,536 individuals, UK Digital Heart project with 1,129 individuals) were tested for concordance. For each locus in the discovery cohort that showed association with  $p_{adjust} < 5 \times 10^{-8}$ , the genetic variant with the lowest  $p$ -value was selected. This selection was done for each univariate per-slice GWAS. The same loci-slice associations were selected in the validation cohort. Variant rs71394376 is not present in UK Digital Heart genotypes and concordance was only tested at 15 out of 16 associated loci. The effect size estimated for the selected genetic variants in the discovery and validation data set were then compared for concordance, i.e. same effect size direction. To evaluate if the observed concordance was likely to arise by chance alone, an empirical  $p$ -value for concordance was estimated by randomly selecting slice-variant associations with a slice distribution as observed in the discovery associations. In each random selection step, the concordance of the observed and randomly selected slice effect size estimates was computed. The number of times the random concordance was greater or equal than the observed concordance was divided by the number of random selections (10,000) to yield the empirical  $p$ -value.

### Variant annotations

Variant annotations from previous studies were retrieved from the GWAS catalogue [85, accessed 28 January 2019] for the GWAS annotations, Open targets genetics v0.3.2 (<https://genetics.opentargets.org/>)<sup>86</sup> for the nearest protein coding gene and PheWAS annotations, the ENSEMBL regulatory build<sup>87</sup> and GTEx v7 (<https://gtexportal.org/home/>, accessed 28 January 2019) for the expression quantitative trait loci annotations. Annotations were reported when they passed the platform-specific significant thresholds (0.05 FDR on GTEx) or the commonly used GWAS threshold of  $5 \times 10^{-8}$ . For variants annotated in GTEx, we downloaded the baseline gene expression counts of the associated genes and tissues from <https://www.ebi.ac.uk/gxa/home> by querying: gene name AND tissue AND species, e.g. GOSR2 AND heart component AND Homo sapiens. Cluster plots of genotype calls for all genotyped lead variants and their LD proxies were generated with *ScatterShot* (<http://mccarthy.well.ox.ac.uk/static/software/scattershot>; Supplementary Table 14 and Supplementary Fig 1).

### Functional enrichment analysis

We used GARFIELD version 2<sup>88</sup> for functional enrichment analyses of genetic variants with multi-trait GWAS  $p$ -value  $< 10^{-6}$ . GARFIELD accounts for LD structure and local gene density and derives functional enrichment scores (odds ratios) by fitting a logistic regression model. The GARFIELD software package and pre-computed data for samples of European ancestry (LD and annotation data, minor allele frequencies of genetic variants and their distances to nearest transcription start site) were downloaded from <https://www.ebi.ac.uk/birney-srv/GARFIELD>. GARFIELD was run as described in the user manual at <https://www.ebi.ac.uk/birney-srv/GARFIELD/documentation-v2/GARFIELD-v2.pdf>. Annotation results (.perm GARFIELD output file) were filtered for input GWAS threshold ( $P_{Thres} < 10^{-6}$ ) and significance of enrichment ( $EmpPval < 10^{-3}$ ).

### Genomewide trait and disease correlation

We analysed the genetic correlation of the basal, mid and apical trabeculation phenotypes with all available traits on LDhub, a centralized database of summary-level GWAS results for 732 traits (accessed August 2019, <http://ldsc.broadinstitute.org/ldhub/>)<sup>89</sup>. LDhub uses cross-trait LD score regression to estimate the genetic correlation between two traits  $z_1$  and  $z_2$ . It uses the amount of genetic variation tagged by a genetic variant  $j$  (its LD score  $l_j$ ) and regresses the LD scores against the product of the two traits  $z_1 z_2$ . The slope of this regression estimates the genetic covariance between traits  $z_1$  and  $z_2$ . The normalisation of the genetic covariance by the square root of the product of the traits' additive SNP heritabilities  $h_1^2$  and  $h_2^2$

yields the genetic correlation between the traits.<sup>90</sup> Prior to upload to LDhub, we reduced the trabeculation summary statistics to the number of variants included on LDhub (1,208,036 variants, as described in the LDhub test center). The additive heritability estimates were extracted from the -h2.log file, the genetic correlation results from the rg.results.csv file. Results were depicted in a PheWAS manhattan plot. The remainder of the analysis was focused on the genetic correlation of trabeculation phenotypes on cardiac and cardiovascular traits, selected by manual inspection of all LDhub traits (selected traits in [Supplementary Table 13](#)).

### Trabeculation phenotypes and associated loci in dilated cardiomyopathy

The dilated cardiomyopathy (DCM) cohort were prospectively recruited to the NIHR Biobank at the Royal Brompton Hospital, London. All patients underwent cardiac phenotyping with either CMR or transthoracic echocardiography, with DCM diagnosed based on evidence of left ventricular dilatation and systolic impairment with reference to age, gender and body surface area adjusted nomograms. All participants gave written informed consent and the study was approved by the relevant regional research ethics committees. Demographic and clinical characteristics are provided in [Extended Data Table 1](#).

We conducted a logistic regression of loci associated with trabecular phenotypes in the DCM patients and healthy volunteers of the UK Digital Heart Project (1,136 individuals as described in [Genotyping replication cohort](#)). We tested all genetic variants and variants in LD ( $r^2 > 0.1$ ) at the 16 loci associated with trabeculation (1,015 variants). We excluded any samples that did not cluster with the European samples of the HapMap consortium and filtered for unrelated individuals (see [Genotyping replication cohort](#); *plinkQC*<sup>76</sup>). We prioritised retaining DCM cases compared to controls in pairs of related individuals. We filtered any genetic variants that showed differential genotype missingness (p-value < 0.01 in cases versus controls using PLINK 1.9<sup>73</sup> '-test-missing'). 510 DCM cases and 1,134 healthy volunteers and 1,005 variants passed the quality control. They were analysed in a logistic regression model with allele dosage encoding for genetic variant effect, sex and age as covariates and an adaptive permutation approach for the genetic variant effect to obtain empirical p-values (using sample label swapping described here <http://zzz.bwh.harvard.edu/plink/perm.shtml>): 'plink -logistic perm'. In addition, we jointly modeled the per-slice FD phenotypes of DCM patients (307 patients) and UK Biobank discovery cohort as the response variable in a linear mixed model with study (DCM or UK Biobank) as fixed effect and slice (1-9) as random effect (R package *nlme*, v3.1-140.). Analysis can be found in [digital-heart-association](#) on GitHub.

### Trabeculae associated loci in heart failure

GWAS summary data were provided by the Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES) Consortium for 47,309 cases of heart failure and 930,014 controls of European ancestry from 26 cohorts (comprising 29 distinct datasets, cohort characteristics in [Supplementary Data 5](#), obtained from [Supplementary Table 17](#)<sup>91</sup>), with either a population-based or case-control study design. A detailed description of the study is reported elsewhere.<sup>91</sup> In brief, cases were defined by the clinical diagnosis of heart failure and no restrictions on aetiology or morpho-functional phenotypes were made. Following genotyping using high-density arrays and genotype and pre-imputation quality control, study-level genotype data were imputed using reference panels from the 1000 Genomes Project (60%), Haplotype Reference Consortium (35%) or population-specific reference data (5%). Association testing for single nucleotide polymorphisms was then performed using logistic regression including age, sex and principal components as covariates, and assuming additive genetic effects. Meta-analysis of GWAS estimates was performed using fixed-effect inverse variance weighted analysis, implemented in METAL (released March 25, 2011). The linkage disequilibrium score regression (LDSC) intercept, implemented using LDSC v1.0.022, was estimated at 1.0069, suggesting no inflation of the test statistic due to cryptic population structure. The summary statistics for the HERMES consortium are publicly available at: <http://www.broadcvdi.org/>.

### Mendelian randomisation analyses

Mendelian randomisation (MR) analysis was performed using all independent, genetic loci (see Association analysis: Multi-trait meta-analysis) with per-slice, univariate GWAS adjusted p-value  $p_{adj} = p \times T_{eff} < 5 \times 10^{-8}$ . For variants associated with FD in multiple slices, only the slice association with the lowest p-value was used. All variants available in the outcome studies were considered. All MR analyses were conducted using the R-package *TwoSampleMR* (<https://github.com/MRCIEU/TwoSampleMR>, version 0.5.1).<sup>56</sup> Two-sample MR studies with the FD associations described above as the exposure variables and the association of those loci with mixed aetiology heart failure (HERMES consortium, see above and summary of cohort characteristics in [Supplementary Data 5](#)) and DCM (DCM cohort, see above) as outcomes were conducted (harmonised, ie effect allele matched input data for MR in [Supplementary Data 6](#)). The effect size of trabeculation on these traits was estimated with *TwoSampleMR* functions for weighted median, weighted mode, inverse variance weighting and MR-Egger. In addition, the Steiger test for directionality,<sup>60</sup> leave-one-out sensitive analysis to determine the influence of single genetic variants on the overall effect, MR pleiotropy analyses and  $I^2$  analysis for assessing bias in MR-Egger analysis were conducted.<sup>59</sup> For details refer to [Supplementary Note 1.1: Mendelian randomisation](#).

## Finite element modelling

We used a biomechanical model to assess the causative effect of varying trabecular morphology on cardiovascular physiology. To achieve this we compared ventricular behaviour with different degrees of trabecular complexity in the non-compact layer while keeping the total ventricular mass constant. A geometric model of the left ventricle was represented by a symmetric truncated ellipsoid in series with a pre-load and after-load circuit. The simulation was calibrated to approximate median physiological variables observed in the UK Biobank population. Trabeculae were modelled as cylindrical strands orientated in the endocardial long-axis.<sup>92,93</sup> Fractal dimension was calculated from cross-sections of the left ventricular model using the same methodology as for clinical imaging. We recorded the consequent effect of changing trabecular morphology on ventricular volume, contractility and blood pressure at steady state.

The ventricular model replicated *ex vivo* fibre orientations,<sup>94</sup> and accounted for both active and passive material properties of the myocardium using a hyperelastic anisotropic constitutive framework. Specifically, the strain energy function  $\psi$  selected to model the cardiac tissue is  $\psi = C_{10}(\bar{I}_1 - 3) + \frac{k_1}{2k_2} [\exp(k_2(\bar{I}_4 - 1)^2) - 1]$ , where  $\bar{I}_1$  and  $\bar{I}_4$  are the first and fourth invariant of the modified Cauchy-Green tensor  $\bar{\mathbf{C}}$  and  $C_{10}$ ,  $k_1$  and  $k_2$  are material parameters. Boundary conditions for the simulations included constraints on rigid motion and displacements of the ventricular base with implementation of a pre-load (defining left atrial pressure and inflow resistance) and after-load circuit (defining right atrial pressure, aortic/peripheral resistance, and capacitance). The models were discretized to allow finite element analyses with eight-node hexahedral elements into more than  $10^4$  elements using Ansys Meshing (ANSYS Inc., Canonsburg, PA, USA). The property of differential fibre orientation within the ventricular wall, in which the sheets of muscle fibres describe a consistent helical pattern, was preserved by considering ventricular dynamics in nine separate myofibrillar sheets,<sup>94,95</sup> with linearly varying fibre orientation from +80 to -80 degrees with respect to the long-axis. Myocyte contraction in systole was simulated by changing the myocyte stiffness parameters to emulate the observed force/time data from cardiac fibres in response to intracellular calcium variation.<sup>96</sup> Torsional motion was modelled by myocyte contraction patterns creating a realistic counter-clockwise apical rotation with respect to the base.<sup>97-99</sup> The finite element problem was solved by means of the commercial code Abaqus (Abaqus 6.14, SIMULIA, Dassault Systemes). Simulations were performed under the assumption of quasi-static processes, so neglecting any inertia effects. Pressure-volume data are reported during steady state after a minimum of five simulated cardiac cycles. For details about the model parameters please refer to [Supplementary Table 15](#) and [Supplementary Table 16](#) and [Finite-element-modelling](#) on GitHub.

## Medaka experiments

**Fish maintenance:** The wild-type Cab strain and a *myl7::EGFP* fluorescent cardiac reporter line were used. All fish are maintained in closed stocks at Heidelberg University. Medaka (*Oryzias latipes*) husbandry (permit number 35-9185.64/BH Wittbrodt) was performed according to local animal welfare standards (Tierschutzgesetz §11, Abs. 1, Nr. 1) and in accordance with European Union animal welfare guidelines.<sup>100</sup> The fish facility is under the supervision of the local representative of the animal welfare agency. Medaka was raised and maintained as described previously.<sup>101</sup> Unblinded analysis was performed using a sample size sufficient for descriptive analysis. All data were generated on embryos, whose sex is not discernible by external features.

**sgRNA target site selection:** *gosr2* (ENSORLG00000004536) sgRNAs, *mtss1* (ENSORLG00000004945) sgRNAs, and *tnnt2a* (ENSORLG00000024544) sgRNAs were designed with CCTop as described previously,<sup>102</sup> on the medaka genome in Ensembl release 95 (Japanese medaka HdrR assembly ASM223467v1, INSDC Assembly GCA\_002234675.1, Jul 2017). Target sites and oligonucleotides for sgRNA cloning in [Supplementary Tables 17](#) and [18](#). Cloning of sgRNA templates and in vitro transcription was performed as detailed in Stemmer et al. (2015).<sup>102</sup>

**In vitro transcription of mRNA:** The plasmids pCS2+(Cas9) and pCS2+(H2A-mCherry) were linearized using NotI, and mRNA in vitro transcription was performed using the mMessage mMachine SP6 Kit (ThermoFisher Scientific, AM1340).

**Microinjection and screening:** For the CRISPR-Cas9 experiments, medaka zygotes were injected with 150 ng/ $\mu$ l of Cas9 mRNA and 15 ng/ $\mu$ l of each sgRNA per gene as well as 10 ng/ $\mu$ l H2A-mCherry mRNA. Injected embryos were maintained at 22-28°C in embryo rearing medium (ERM, 17 mM NaCl, 40 mM KCl, 0.27 mM CaCl<sub>2</sub>, 0.66 mM MgSO<sub>4</sub>, 17 mM Hepes). One day post fertilization, embryos were selected for H2A-mCherry expression. Phenotypes of CRISPR-Cas9-mediated knock out of *gosr2*, *mtss1*, *tnnt2a*, and control injection (H2A-mCherry) were assessed at 4 DPF separately by two investigators.

**Imaging:** BF images were acquired using a Nikon SMZ18 equipped with a lumencor SOLA SE light source, a Nikon DS-Fi2 camera, and Software NIS-Elements v.4.20.

LSM was carried out on a 16x multiview selective plane illumination microscope (MuVi-SPIM).<sup>103</sup> Embryos were anesthetized with 200 mg/l tricaine and electromechanically decoupled with 30 mM 2,3-butanedione 2-monoxime (BDM). After complete inhibition of cardiac contraction, embryos were mounted with 1% low-melting agarose (LMA) containing 200 mg/l tricaine and 30 mM BDM into custom FEP tubes connected to a glass capillary. FEP-glass-capillaries were washed with

70% ethanol before usage. Entire heart volumes were acquired with 488 nm illumination, a 525/50 nm bandpass filter, and a z step size of 2  $\mu\text{m}$ . Surface rendering was performed with UCSF Chimera (1.11).

## References

47. Lawlor, D. A. Commentary: Two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* **45**, 908–15 (2016).
48. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* **23**, R89–R98 (2014).
49. Haycock, P. C. *et al.* Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* **103**, 965–978 (2016).
50. Holmes, M. V., Ala-Korpela, M. & Smith, G. D. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* **14**, 577–590 (2017).
51. Pierce, B. L. & Burgess, S. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. *Am J Epidemiol* **178**, 1177–1184 (2013).
52. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genet Epidemiol* **37**, 658–665 (2013).
53. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512–525 (2015).
54. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* **40**, 304–314 (2016).
55. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* **46**, 1985–1998 (2017).
56. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
57. Burgess, S., Thompson, S. G. & CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* **40**, 755–764 (2011).
58. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* **40**, 597–608 (2016).
59. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I<sup>2</sup> statistic. *Int J Epidemiol* **45**, 1961–1974 (2016).
60. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).
61. 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* **12**, 1–10 (2015).
62. Petersen, S. E. *et al.* UK biobank’s cardiovascular magnetic resonance protocol. *J Cardiovasc Magn Reson* **18**, 8 (2016).
63. de Marvao, A. *et al.* Population-based studies of myocardial hypertrophy: high resolution cardiovascular magnetic resonance atlases improve statistical power. *J Cardiovasc Magn Reson* **16**, 16 (2014).
64. Woodbridge, M., Fagiolo, G. & O’Regan, D. P. MRIdb: medical image management for biobank research. *J Digit Imaging* **26**, 886–90 (2013).
65. Marcus, D. S., Olsen, T. R., Ramaratnam, M. & Buckner, R. L. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* **5**, 11–34 (2007).
66. Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* **20**, 65 (2018).
67. Schulz-Menger, J. *et al.* Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing. *J Cardiovasc Magn Reson* **15**, 35 (2013).
68. Puyol-Antón, E. *et al.* Fully automated myocardial strain estimation from cine MRI using convolutional neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 1139–1143 (2018).
69. Li, C. *et al.* A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri. *IEEE Trans Image Process* **20**, 2007–16 (2011).



70. Zhuang, X. *et al.* Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. *Med Imag Anal* **58**, 101537 (2019).
71. Seemann, F. *et al.* Noninvasive quantification of pressure-volume loops from brachial pressure and cardiovascular magnetic resonance. *Circ Cardiovasc Imaging* **12**, e008493 (2019).
72. Bouchard, R. J., Gault, J. H. & Ross, J., J. Evaluation of pulmonary arterial end-diastolic pressure as an estimate of left ventricular end-diastolic pressure in patients with normal and abnormal left ventricular performance. *Circulation* **44**, 1072–9 (1971).
73. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
74. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 (2017).
75. The International HapMap Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467** (2010).
76. Meyer, H. V. R package: plinkQC. <https://cran.r-project.org/web/packages/plinkQC/> (2018).
77. Teo, Y. Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741–6 (2007).
78. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
79. UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
80. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5–6 (2013).
81. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–70 (2011).
82. Bycroft, C. *et al.* The UK biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
83. Galwey, N. W. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. *Genet Epidemiol* **33**, 559–68 (2009).
84. Bolormaa, S. *et al.* A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS genetics* **10**, e1004198 (2014).
85. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005–D1012 (2019).
86. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res* **47**, D1056–D1065 (2018).
87. Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T. & Flicek, P. R. The Ensembl regulatory build. *Genome Biol* **16**, 56 (2015).
88. Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet* **51**, 343–353 (2019).
89. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
90. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241 (2015).
91. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun* **11**, 163 (2020).
92. Gati, S. *et al.* Increased left ventricular trabeculation in highly trained athletes: do we need more stringent criteria for the diagnosis of left ventricular non-compaction in athletes? *Heart* **99**, 401–8 (2013).
93. Tamborini, G. *et al.* Incidence and characteristics of left ventricular false tendons and trabeculations in the normal and pathologic heart by second harmonic echocardiography. *J Am Soc Echocardiogr* **17**, 367–74 (2004).
94. Streeter, J., D. D., Spotnitz, H. M., Patel, D. P., Ross, J., J. & Sonnenblick, E. H. Fiber orientation in the canine left ventricle during diastole and systole. *Circ Res* **24**, 339–47 (1969).
95. Lombaert, H. *et al.* Human atlas of the cardiac fiber architecture: study on a healthy population. *IEEE Trans Med Imaging* **31**, 1436–47 (2012).

96. Backx, P. H., Gao, W. D., Azan-Backx, M. D. & Marban, E. The relationship between contractile force and intracellular [Ca<sup>2+</sup>] in intact rat cardiac trabeculae. *J Gen Physiol* **105**, 1–19 (1995).
97. Buckberg, G. *et al.* Ventricular torsion and untwisting: further insights into mechanics and timing interdependence: a viewpoint. *Echocardiography* **28**, 782–804 (2011).
98. Takeuchi, M., Otsuji, Y. & Lang, R. M. Evaluation of left ventricular function using left ventricular twist and torsion parameters. *Curr Cardiol Rep* **11**, 225–30 (2009).
99. Carreras, F. *et al.* Left ventricular torsion and longitudinal shortening: two fundamental components of myocardial mechanics assessed by tagged cine-MRI in normal subjects. *Int J Cardiovasc Imaging* **28**, 273–84 (2012).
100. Bert, B. *et al.* Considerations for a European animal welfare standard to evaluate adverse phenotypes in teleost fish. *EMBO J* **35**, 1151–4 (2016).
101. Gierten, J. *et al.* Automated high-throughput heartbeat quantification in medaka and zebrafish embryos under physiological conditions. *Sci Rep* **10**, 2046 (2020).
102. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* **10**, e0124633 (2015).
103. Krzic, U., Gunther, S., Saunders, T. E., Streichan, S. J. & Hufnagel, L. Multiview light-sheet microscope for rapid in toto imaging. *Nat Methods* **9**, 730–3 (2012).
104. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
105. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896–D901 (2017).
106. Stanescu, H. C. *et al.* Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. *N Engl J Med* **364**, 616–26 (2011).
107. Almaguera, B. *et al.* Identification of Four Novel Loci in Asthma in European American and African American Populations. *Am J Respir Crit Care Med* **195**, 456–463 (2017).
108. Kunz, M. *et al.* Genome-wide association study identifies new susceptibility loci for cutaneous lupus erythematosus. *Exp Dermatol* **24**, 510–5 (2015).
109. Ligthart, S. *et al.* Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. *BMC genomics* **17**, 443–443 (2016).
110. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390–400 (2018).
111. Wild, P. S. *et al.* Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J Clin Invest* **127**, 1798–1812 (2017).
112. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet* **50**, 1225–1233 (2018).
113. Tereshchenko, L. G. *et al.* Genome-Wide Associations of Global Electrical Heterogeneity ECG Phenotype: The ARIC (Atherosclerosis Risk in Communities) Study and CHS (Cardiovascular Health Study). *J Am Heart Assoc* **7** (2018).
114. van der Harst, P. *et al.* 52 Genetic Loci Influencing Myocardial Mass. *J Am Coll Cardiol* **68**, 1435–1448 (2016).
115. Evans, D. S. *et al.* Fine-mapping, novel loci identification, and SNP association transferability in a genome-wide association study of QRS duration in African Americans. *Hum Mol Genet* **25**, 4350–4368 (2016).
116. Verweij, N. *et al.* Genetic determinants of P wave duration and PR segment. *Circ Cardiovasc Genet* **7**, 475–81 (2014).
117. Pfeufer, A. *et al.* Genome-wide association study of PR interval. *Nat Genet* **42**, 153–9 (2010).
118. Dina, C. *et al.* Genetic association analyses highlight biological pathways underlying mitral valve prolapse. *Nat Genet* **47**, 1206–11 (2015).
119. Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–9 (2011).
120. Ehret, G. B. *et al.* The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet* **48**, 1171–1184 (2016).
121. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**, 1151–1161 (2016).

122. Wain, L. V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005–11 (2011).
123. Wain, L. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension (Dallas, Tex. : 1979)* (2017).
124. Sotoodehnia, N. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet* **42**, 1068–76 (2010).

## 2.1 Supplementary Tables

**Supplementary Table 1. Single-trait heritability estimates and genomic control** All estimates based on summary statistics ( $t$  statistic) of univariate association results of 14,134,301 genetic variants and 18,096 samples. Estimates are obtained via LD score regression. Heritability estimates ( $h_2$ ) are on the observed scale. Mean  $\chi^2$  is the mean  $\chi^2$  statistic.  $\lambda_{GC}$  is the  $\frac{\text{median}\chi^2}{0.4549}$ . Intercept is the LD Score regression intercept. Ratio is  $\frac{\text{Intercept}-1}{\text{Mean}\chi^2-1}$ . The intercept should be close to 1. The ratio measures the proportion of the inflation in the mean  $\chi^2$  that the LD Score regression intercept ascribes to causes other than polygenic heritability. The value of ratio should be close to zero, though in practice values of 10-20% are not uncommon. SE indicates the standard error. For details see: <https://github.com/bulik/ldsc/wiki/Heritability-and-Genetic-Correlation> and<sup>90,104</sup>

Slice	$h_2$	$h_2$ SE	Mean $\chi^2$	$\lambda_{GC}$	Intercept	Intercept SE	Ratio	Ratio SE
1	0.1339	0.0316	1.0481	1.0557	0.9995	0.0072	Ratio < 0	-
2	0.1482	0.0514	1.0538	1.0436	0.9989	0.0104	Ratio < 0	-
3	0.1732	0.0345	1.0688	1.0496	1.0065	0.0074	0.0949	0.1077
4	0.1546	0.0305	1.0641	1.0557	1.0092	0.0068	0.1438	0.1058
5	0.2125	0.0313	1.0817	1.0649	1.0055	0.0071	0.0676	0.0867
6	0.2187	0.032	1.0873	1.0679	1.0095	0.0069	0.1091	0.0786
7	0.2212	0.03	1.0823	1.0679	1.0036	0.0065	0.0442	0.079
8	0.2001	0.0287	1.0723	1.0618	1.0014	0.0066	0.019	0.091
9	0.1236	0.0265	1.0433	1.0436	0.9994	0.0063	Ratio < 0	-

**Supplementary Table 2. UK Biobank replication.** SNP with lowest meta-analysis p-value (uncorrected for multiple comparisons) per locus in the discovery cohort (SNP and P-value discovery, see [Supplementary Table 5](#)) and the p-values of the same SNP and SNP with lowest p-value in that locus ( $r^2 > 0.8$ ) for the replication cohort. Meta-analysis p-values were estimated based on the transformation of the univariate signed t-statistics (associations on 1,199 genetic variants at 16 independent loci from 6,536 samples) and  $\chi^2$  distribution with 9 degrees of freedom. Eight loci replicate with Bonferroni corrected threshold for the number of loci tested ( $n=16$ ):  $p < 0.05/16 = 0.003$ .

CHR	SNP	BP	P-value discovery	Replication	
				P-value SNP	P-value locus
1	rs6587924	61895257	1.57E-17	3.40E-05	1.56E-05
1	rs35770803	155962067	3.71E-08	4.96E-01	4.96E-01
1	rs1892027	201332020	2.01E-08	3.93E-04	3.92E-04
2	rs71394376	179531078	1.40E-09	2.15E-02	7.85E-03
3	rs4677294	73554922	2.61E-16	5.21E-02	1.44E-02
3	rs1918978	169191428	2.40E-08	8.57E-03	8.57E-03
5	rs10076436	153871841	4.90E-10	4.03E-02	3.53E-02
6	rs3130976	31081940	5.29E-09	1.38E-01	7.32E-02
6	rs9320648	118690014	1.67E-17	3.35E-02	3.35E-02
8	rs6981461	11794962	1.64E-10	1.73E-07	8.06E-08
8	rs35006907	125859817	9.55E-11	8.53E-05	6.59E-05
12	rs7132327	115381071	1.97E-09	1.21E-09	1.87E-10
14	rs71105784	71990847	2.63E-11	3.45E-04	1.86E-04
17	rs17608766	45013271	1.98E-26	4.04E-09	8.75E-10
19	rs113394178	7581244	3.81E-10	3.21E-04	1.30E-04
22	rs3788488	33127481	2.81E-08	4.29E-02	4.29E-02

**Supplementary Table 3. UK Biobank replication locus details.** SNP identifier and chromosomal position of SNP with lowest meta-analysis p-value (not adjusted for multiple comparisons) in locus used for replication; see [Supplementary Table 2](#)). Meta-analysis p-values were estimated based on the transformation of the univariate signed t-statistics (associations on 1,199 genetic variants at 16 independent loci from 6,536 samples) and  $\chi^2$  distribution with 9 degrees of freedom.

CHR	SNP	BP	Replication P-value locus
1	rs9436640	61895257	1.56E-05
1	rs35770803	155962067	4.96E-01
1	rs1104859	201332020	3.92E-04
2	rs2042995	179531078	7.85E-03
3	rs34210879	73554922	1.44E-02
3	rs1918978	169191428	8.57E-03
5	rs13185595	153871841	3.53E-02
6	rs548139121	31081940	7.32E-02
6	rs9320648	118690014	3.35E-02
8	rs60902764	11794962	8.06E-08
8	rs7461129	125859817	6.59E-05
12	rs61933462	115381071	1.87E-10
14	rs61991243	71990847	1.86E-04
17	rs11874	45013271	8.75E-10
19	rs644053	7581244	1.30E-04
22	rs3788488	33127481	4.29E-02

**Supplementary Table 4. UK Digital Heart study replication.** SNP identifier and chromosomal position of SNP with lowest meta-analysis p-value in discovery (not adjusted for multiple comparisons, see [Supplementary Table 5](#)) and the p-values of the same SNP for the UK Digital Heart study replication cohort. Meta-analysis p-values were estimated based on the transformation of the univariate signed t-statistics (associations on 1,015 genetic variants at 16 independent loci from 1,136 samples) and  $\chi^2$  distribution with 9 degrees of freedom. Two loci replicate at Bonferroni corrected threshold for the number of loci tested (n=16):  $p < 0.05/16 = 0.003$ . One SNP was not genotyped this replication cohort (indicated by NA).

CHR	SNP	BP	P-value discovery	P-value Replication
1	rs6587924	61895257	1.57E-17	6.04E-01
1	rs35770803	155962067	3.71E-08	2.73E-01
1	rs1892027	201332020	2.01E-08	2.97E-01
2	rs71394376	179531078	1.40E-09	NA
3	rs4677294	73554922	2.61E-16	9.16E-01
3	rs1918978	169191428	2.40E-08	9.11E-03
5	rs10076436	153871841	4.90E-10	1.51E-01
6	rs3130976	31081940	5.29E-09	3.29E-01
6	rs9320648	118690014	1.67E-17	7.12E-01
8	rs6981461	11794962	1.64E-10	2.43E-02
8	rs35006907	125859817	9.55E-11	6.50E-03
12	rs7132327	115381071	1.97E-09	3.21E-01
14	rs71105784	71990847	2.63E-11	1.57E-01
17	rs17608766	45013271	1.98E-26	6.74E-01
19	rs113394178	7581244	3.81E-10	3.80E-01
22	rs3788488	33127481	2.81E-08	5.17E-02

**Supplementary Table 5. Characteristics of trabeculation-associated loci.** Overview of the 16 independent loci discovered in the trabeculation GWAS. The genetic variant with the lowest meta-analysis p-value (not adjusted for multiple testing) per locus is shown. Chromosomes (CHR), base pair positions (BP), ID, Locus and Type based on GRCh37 (Ensembl GRCh37 Release 95). Allele frequency (AF), reference (A\_0) and alternative allele (A\_1) based on discovery cohort data. INFO denotes the imputation quality score of impute2<sup>81</sup>. Slice indicates the slices associated with this locus, where multi denotes association in multi-variate test only. No entry indicated by -. Meta-analysis p-values were estimated based on the transformation of the univariate signed t-statistics (associations on 14,134,301 genetic variants at 16 independent loci from 18,096 samples) and  $\chi^2$  distribution with 9 degrees of freedom.

CHR	SNP	BP	A_0	A_1	AF	INFO	P	Ensembl ID	Locus	Type	Slice
1	rs6587924	61895257	C	A	0.51	1.00	1.57E-17	ENSG00000162604	TM2D1	intron	4,5,6,7
1	rs35770803	155962067	G	GT	0.63	0.96	3.71E-08	ENSG00000224276	RP11-336K24.5	intron	multi
1	rs1892027	201332020	T	C	0.71	1.00	2.01E-08	ENSG00000118194	TNNT2	intron	multi
2	rs71394376	179531078	A	AATGT	0.21	1.00	1.40E-09	ENSG00000155657	TTN	intron	3,4
3	rs4677294	73554922	T	A	0.36	0.99	2.61E-16	ENSG00000121440	PDZRN3	intron	2,3
3	rs1918978	169191428	A	G	0.58	0.98	2.40E-08	ENSG00000085276	MECOM	intron	7,8
5	rs10076436	153871841	C	G	0.36	1.00	4.90E-10	intergenic	-	-	3,4
6	rs3130976	31081940	T	C	0.29	1.00	5.29E-09	intergenic	-	-	5,6,8
6	rs9320648	118690014	A	C	0.58	1.00	1.67E-17	intergenic	-	-	5,6,7,8
8	rs6981461	11794962	T	C	0.44	0.98	1.64E-10	intergenic	-	-	multi
8	rs35006907	125859817	C	A	0.31	0.99	9.55E-11	ENSG00000255080	RP11-1082L8.3	intron	5,6
12	rs7132327	115381071	T	C	0.27	1.00	1.97E-09	intergenic	-	-	4,5
14	rs71105784	71990847	C	CCTGT	0.25	1.00	2.63E-11	intergenic	-	-	3,4
17	rs17608766	45013271	T	C	0.15	1.00	1.98E-26	ENSG00000108433	GOSR2	3'UTR intron	2,3,4
19	rs113394178	7581244	C	A	0.61	0.96	3.81E-10	ENSG00000198816	ZNF358	intron	4,5,6
22	rs3788488	33127481	T	C	0.27	0.99	2.81E-08	ENSG00000100234	TIMP3	intron	multi

**Supplementary Table 6. Loci annotation in published GWAS.** Based on entries in the GWAS catalogue<sup>105</sup>. P-values are meta-analysis p-values, not adjusted for multiple testing derived from the transformation of the univariate signed t-statistics (associations on 14,134,301 genetic variants at 16 independent loci from 18,096 samples) and  $\chi^2$  distribution with 9 degrees of freedom. No entry indicated by -. BP, Blood pressure; LV, left ventricle/left ventricular; AA, atrial appendage; CLE, Cutaneous lupus erythematosus.

CHR	BP	SNP	P	GWAS	Slice	Region	
6	31081940	rs3130976	5.29E-09	Nephropathy <sup>106</sup> , asthma <sup>107</sup> , CLE <sup>108</sup>	adult	5, 6, 7	mid/apical
8	11794962	rs6981461	1.64E-10	C-reactive protein <sup>109</sup>	multi-trait	-	-
8	125859817	rs35006907	9.55E-11	Ejection fraction, fractional shortening, LV internal dimension in systole and diastole, relative wall thickness <sup>110</sup> , LV internal dimension <sup>111</sup> , atrial fibrillation <sup>112</sup>	5, 6	mid	mid
12	115381071	rs7132327	1.97E-09	Global electrical heterogeneity phenotypes <sup>113</sup> , QRS complex <sup>114</sup> , QRS duration <sup>114,115</sup> , PR segment <sup>116</sup> , PR interval <sup>117</sup>	4	mid	mid
14	71990847	rs71105784	2.63E-11	QRS complex <sup>114</sup> , QRS duration <sup>114,115</sup> , mitral valve prolapse <sup>118</sup>	3, 4	basal/mid	basal/mid
17	45013271	rs17608766	1.98E-26	Systolic blood pressure <sup>119-123</sup> , QRS duration <sup>114,124</sup> , pulse and blood pressure <sup>123</sup> , aortic root size <sup>111</sup> , atrial fibrillation <sup>112</sup>	2, 3, 4	basal/mid	basal/mid

**Supplementary Table 7. Finite element modelling.** Steady-state measurements of the left ventricle obtained after 5 cardiac cycles, selectively varying only trabecular complexity, at constant total myocardial mass with identical boundary conditions.

Haemodynamics	Smooth model	Trabeculated model
Peak systolic pressure (mmHg)	93	123.6
End-systolic pressure (mmHg)	88	119.8
Diastolic pressure (mmHg)	50	65.9
Left ventricular volumes and function		
End-diastolic volume (ml)	114.8	160.4
End-systolic volume (ml)	54.8	80.1
Stroke volume (ml)	60	80.3
Stroke work (J)	0.62	1.09
Cardiac output (l/min)	3.6	4.8
Contractility (mmHg/ml)	1.61	1.50
Mass (g)	97.7	97.7

**Supplementary Table 8. Disease association of NFIA (chr1) and GOSR2 (chr17) loci.** rs17608766 and rs6587924 are associated with decreased trabeculation in the mid and basal regions of the heart (Supplementary Table 5 and 5). Regression of these variants or a variant in LD (LD proxy with LD  $r^2$  as in 1000 Genomes GBR population) on heart failure (HERMES consortium<sup>91</sup>) and dilated cardiomyopathy show an increased disease risk for these loci with increased beta (HERMES) and increased odds ratio (OR; marked with \*). Allele frequencies (AF), Odds ratios (OR), effect sizes (BETA) with respect to the A\_1 allele, i.e. frequency or effect of having an extra copy of A\_1. p-value is the p-value from the association test, N the number of individuals in the study, BF the Bonferroni-corrected p-value for the number of tests conducted. For UK Biobank discovery cohort p-values adjusted for the effective number of univariate associations  $T_{eff} = 6.6$ , for disease associations the number of loci tested:  $T_{eff} = 16$ . For details see Methods.

Locus	Cohort	Variant/LD proxy	LD $r^2$	A_0	A_1	AF	Beta/OR	SE	p-value	BF	N
chr1:61895257	UKB Discovery	rs6587924	-	C	T	0.5	-0.08	0.010	5.3E-15	3.5E-14	18,097
	HERMES	rs1997997	0.8	G	A	0.51	0.03	0.009	2.5E-03	4.0E-02	587,821
chr17:45013271	UKB Discovery	rs17608766	-	T	C	0.15	-0.07	0.011	1.7E-11	1.1E-10	18,097
	HERMES	rs17608766	-	T	C	0.14	0.04	0.011	2.2E-04	3.6E-03	955,712
	DCM	rs145153053	0.7	A	G	0.18	1.388	0.095	4.1E-04	6.6E-03	1,627

**Supplementary Table 9. Mendelian randomisation results for FD associations with mixed aetiology heart failure and dilated cardiomyopathy.** nsnp specifies the number of snps present in both the exposure (FD associations) and outcome study. b and se are the causal effect size estimate and standard error of the study and MR method, pval the p-value of the MR analysis (not adjusted for multiple testing) as specified in the 'methods' column. For MR Egger, the magnitude of dilution bias is  $I^2 = 0.98368$  and  $I^2 = 0.9830075$ , i.e. an approximately 1.6% and 1.7% underestimation of effect size is expected for DCM and mixed aetiology heart failure MR, respectively. F statistics and their lower bound can be found in Supplementary Table 11. For details on methods and statistics refer to Supplementary Note 1.1: Mendelian randomisation.

outcome	method	nsnp	b	se	pval
Heart failure - HERMES	MR Egger	12	-0.4978	0.174188655	0.017020598
Heart failure - HERMES	Inverse variance weighted	12	-0.11488	0.048131871	0.016994636
Heart failure - HERMES	Weighted median	12	-0.10535	0.047411945	0.026284417
Heart failure - HERMES	Weighted mode	12	-0.16451	0.07156832	0.042129831
Dilated cardiomyopathy	MR Egger	11	-2.21315	1.450045582	0.161287259
Dilated cardiomyopathy	Inverse variance weighted	11	-1.00316	0.343880362	0.003532274
Dilated cardiomyopathy	Weighted median	11	-0.93172	0.450643115	0.038683286
Dilated cardiomyopathy	Weighted mode	11	-1.05602	0.608401076	0.113262308

**Supplementary Table 10. Steiger directionality analysis for causality of trabecular complexity with heart failure and dilated cardiomyopathy.** snp r2.exposure and snp r2.outcome are the  $r^2$  estimates of the correlation of all instrumental variables with the exposure and outcome traits, respectively. snp r2.exposure differ dependent on how many genetic variants overlapped between the FD associations and the outcome study - correlation only estimated for variants present in outcome and exposure. correct specifies if the hypothesised direction of FD upstream of the outcome trait is true. The p-values (steiger pval) were estimated using the MR Steiger test. For details on MR Steiger refer to Mendelian randomisation.

outcome	snp r2.exposure	snp r2.outcome	correct causal direction	steiger pval
Heart failure - HERMES	0.035811917	5.39E-05	TRUE	1.97E-132
Dilated cardiomyopathy	0.03380195	0.007333759	TRUE	0.000119535

**Supplementary Table 11. F statistic of MR studies.** The F statistic depends on the sample size in the exposure study (N exposure), number of IVs (IV), and the proportion of variance in the risk factor explained by the IVs ( $r^2$ ). It is computed by  $\frac{r^2 \times (N \text{ exposure} - 1 - IV)}{(1 - r^2) \times IV}$ . As the F statistic is a cohort estimate of the unknown population F parameter a lower bound of the F parameter ( $F_{lower}$ ) was estimated according to [58, Appendix A3]. All lower bounds are  $\geq 10$ , so no strong weak instrument bias would be expected Mendelian randomisation.

outcome	IV	N exposure	N outcome	r2 exposure	F statistic	Lower bound F
Heart failure - HERMES	12	18097	589093	0.035811917	55.97306169	48.98868669
Dilated cardiomyopathy	11	18097	1628	0.03380195	57.5176894	50.1270644



**Supplementary Table 12. Pleiotropy assessment with MR-Egger for FD associations.** Pleiotropy assessments are shown for analysis of mixed aetiology heart failure and DCM as outcomes. egger intercept and se are the pleiotropy estimate and its standard error, pval the p-value of the intercept. All statistics were estimated by MR Egger. For details on MR Egger refer to [Mendelian randomisation](#).

outcome	egger intercept	se	pval
Heart failure - HERMES	0.033338577	0.014738401	0.047206158
Dilated cardiomyopathy	0.105545727	0.122876293	0.412655488

**Supplementary Table 13. Cardiac and cardiovascular LDhub traits.** Manual selection of all LDhub traits related to cardiac phenotypes and cardiovascular diseases. Used for genetic correlation analysis with trabeculation phenotypes depicted in [Extended Data Fig 7f,g](#).

Original LDhub label	Short names
Diastolic blood pressure_automated reading	Diastolic blood pressure
Systolic blood pressure_automated reading	Systolic blood pressure
Pulse rate	Pulse rate
Pulse wave reflection index	Pulse wave reflection index
Pulse wave peak to peak time	Pulse wave peak to peak time
Target heart rate achieved	Target heart rate achieved
Pulse wave Arterial Stiffness index	Pulse wave Arterial Stiffness Index
Non-cancer illness code_self-reported: hypertension	self-reported: hypertension
Non-cancer illness code_self-reported: angina	self-reported: angina
Non-cancer illness code_self-reported: heart attack/myocardial infarction	self-reported: heart attack/myocardial infarction
Non-cancer illness code_self-reported: hypertrophic cardiomyopathy (hcm / hocm)	self-reported: hypertrophic cardiomyopathy
Illnesses of father: Heart disease	Illnesses of father: Heart disease
Illnesses of father: High blood pressure	Illnesses of father: High blood pressure
Illnesses of father: Diabetes	Illnesses of father: Diabetes
Illnesses of mother: Heart disease	Illnesses of mother: Heart disease
Illnesses of mother: High blood pressure	Illnesses of mother: High blood pressure
Illnesses of siblings: Heart disease	Illnesses of siblings: Heart disease
Illnesses of siblings: High blood pressure	Illnesses of siblings: High blood pressure
Vascular/heart problems diagnosed by doctor: Heart attack	diagnosed by doctor: Heart attack
Vascular/heart problems diagnosed by doctor: None of the above	diagnosed by doctor: Vascular/heart problems
Vascular/heart problems diagnosed by doctor: Angina	diagnosed by doctor: Angina
Vascular/heart problems diagnosed by doctor: High blood pressure	diagnosed by doctor: High blood pressure
Medication for cholesterol_blood pressure_diabetes_or take exogenous hormones: Blood pressure medication	Medication (1) : Blood pressure medication
Medication for cholesterol_blood pressure or diabetes: Blood pressure medication	Medication (2): Blood pressure medication
Diagnoses - main ICD10: I20 Angina pectoris	ICD10: I20 Angina pectoris
Diagnoses - main ICD10: I21 Acute myocardial infarction	ICD10: I21 Acute myocardial infarction
Diagnoses - main ICD10: I25 Chronic ischaemic heart disease	ICD10: I25 Chronic ischaemic heart disease
Diagnoses - main ICD10: I30 Acute pericarditis	ICD10: I30 Acute pericarditis
Diagnoses - main ICD10: I48 Atrial fibrillation and flutter	ICD10: I48 Atrial fibrillation and flutter

**Supplementary Table 14. Trabeculation-associated variants and LD proxies.** Statistics and summaries for LD proxies of associated variants. Cluster plots of the variants or their LD proxies are depicted in [Supplementary Fig 1](#).

chr	rsID	BP	LD proxy			
			rID	BP	p-value	$r^2$
1	rs6587924	61895257	rs2207790	61897967	1.65E-15	0.84
1	rs35770803	155962067	rs12043212	155964698	6.08E-06	0.58
1	rs1892027	201332020	rs3729547	201334382	2.92E-08	0.97
2	rs71394376	179531078	rs2042995	179558366	1.35E-08	0.92
3	rs4677294	73554922	rs7647178	73565183	1.04E-13	0.86
3	rs1918978	169191428	rs7613621	169191186	3.81E-07	0.54
5	rs10076436	153871841	rs10054375	153871832	6.94E-10	1
6	rs3130976	31081940	rs3095298	31082932	1.21E-08	0.97
6	rs9320648	118690014	rs6569015	118686344	2.54E-14	0.96
8	rs6981461	11794962	rs60176945	11796674	1.03E-09	0.91
8	rs35006907	125859817	rs12542527	125876444	1.94E-08	0.86
12	rs7132327	115381071	rs10850409	115381740	6.47E-09	1
14	rs71105784	71990847	rs72728427	71769989	1.53E-10	1
17	rs17608766	45013271	-	-	-	genotyped
19	rs113394178	7581244	rs4804662	7575408	6.46E-02	0.26
22	rs3788488	33127481	rs2413146	33125992	1.55E-07	0.76

**Supplementary Table 15. Material parameters in the computational model strain energy function.**

Material parameter	Diastolic Value	Maximum systolic value
$C_{10}$ (kPa)	0.2	7.2
$k_1$ (kPa)	1	180
$k_2$ (-)	2	2

**Supplementary Table 16. Resistances, compliance and atrial pressures in the *in silico* model pre-load and after-load circuits.** Compare to [Extended Data Fig 7c](#).

$P_{LA}$ (mmHg)	$R_1$ (mmHg*min/l)	$R_2$ (mmHg*min/l)	$R_3$ (mmHg*min/l)	$C$ (l/mmHg)	$P_{RA}$ (mmHg)
5.25	0.2	0.375	18	0.0012	5

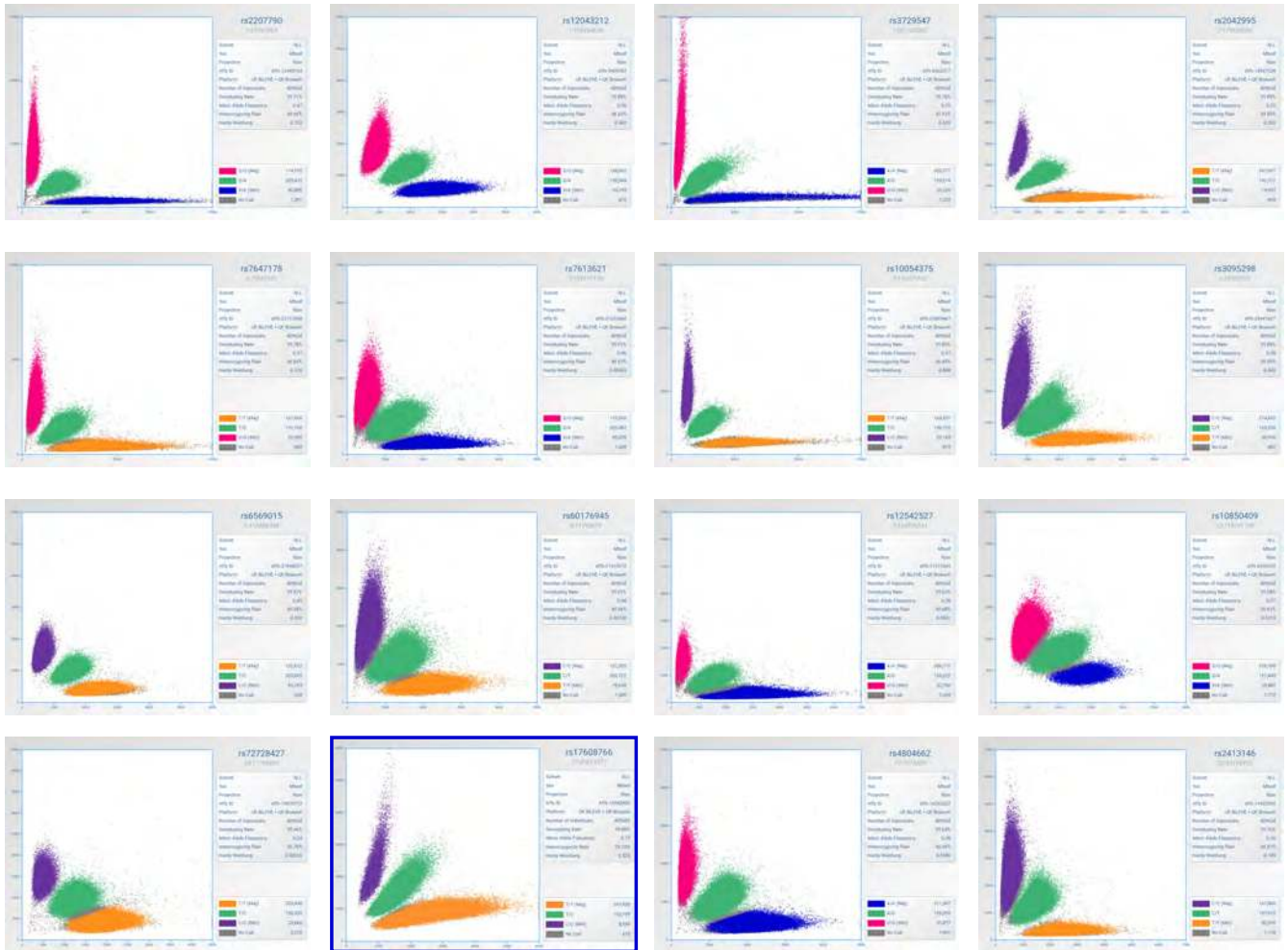
**Supplementary Table 17. CRISPR target sites for allelic series in medaka.** PAM sites specified in brackets.

Name	Sequence
gosr2_T1	CAGGAGGTTTCAGTGTCTGAT[GGG]
gosr2_T1	TTGGCGTTCTGGCGGCGATT[CGG]
gosr2_T3	ACGCTATAACCAGAGAGGCGC[AGG]
gosr2_T4	AACTGCAGAGTCTCGTCAAT[GGG]
mtss1_T1	TTGCGCTGTTTCAGTTGGGGC[TGG]
mtss1_T2	GATGAGGAGACGTCTATGTT[AGG]
mtss1_T3	GATCTGAAGGCCTCAGACTA[TGG]
tnnt2a_T1	GGACATCCATCGTAAGAGAA[TGG]
tnnt2a_T2	AGAGCGCCAAAAACGTCTTG[AGG]
tnnt2a_T3	CGAAAGAGCTCGTAAGGAAG[AGG]
tnnt2a_T4	TCAAGATAGACTTAAGTAAG[TGG]

**Supplementary Table 18. Oligonucleotides for CRISPR sgRNA cloning.** Oligonucleotides depicted in 5'-3' orientation.

Name	Sequence
gosr2_T1_F	TAggGGAGGTTTCAGTGTCTGAT
gosr2_T1_R	AAACATCAGACACTGAACCTCC
gosr2_T2_F	TAggGGCGTTCTGGCGGCGATT
gosr2_T2_R	AAACAATCGCCGCCAGAACGCC
gosr2_T3_F	TAggGCTATACCAGAGAGGCGC
gosr2_T3_R	AAACGCGCCTCTCTGGTATAGC
gosr2_T4_F	TAggCTGCAGAGTCTCGTCAAT
gosr2_T4_R	AAACATTGACGAGACTCTGCAG
mtss1_T1_F	TAggGCGCTGTTCAGTTGGGGC
mtss1_T1_R	AAACGCCCCAACTGAACAGCGC
mtss1_T2_F	TAGgTGAGGAGACGTCTATGTT
mtss1_T2_R	AAACAACATAGACGTCTCCTCA
mtss1_T3_F	TAGgTCTGAAGGCCTCAGACTA
mtss1_T3_R	AAACTAGTCTGAGGCCTTCAGA
tnnt2a_T1_F	TAGGACATCCATCGTAAGAGAA
tnnt2a_T1_R	AAACTTCTCTTACGATGGATGT
tnnt2a_T2_F	TAgGAGCGCCAAAAACGTCTTG
tnnt2a_T2_R	AAACCAAGACGTTTTTGGCGCT
tnnt2a_T3_F	TAgGAAAGAGCTCGTAAGGAAG
tnnt2a_T3_R	AAACCTTCCTTACGAGCTCTTT
tnnt2a_T4_F	TAggAAGATAGACTTAAGTAAG
tnnt2a_T4_R	AAACCTTACTTAAGTCTATCTT

## 2.2 Supplementary Figures



**Supplementary Figure 1. Cluster plots of trabeculation-associated variants and LD proxies.** Directly genotyped variants are indicated with a blue box. Cluster plots were generated via <http://mccarthy.well.ox.ac.uk/static/software/scattershot/>. Detailed information on LD proxies in [Supplementary Table 14](#).