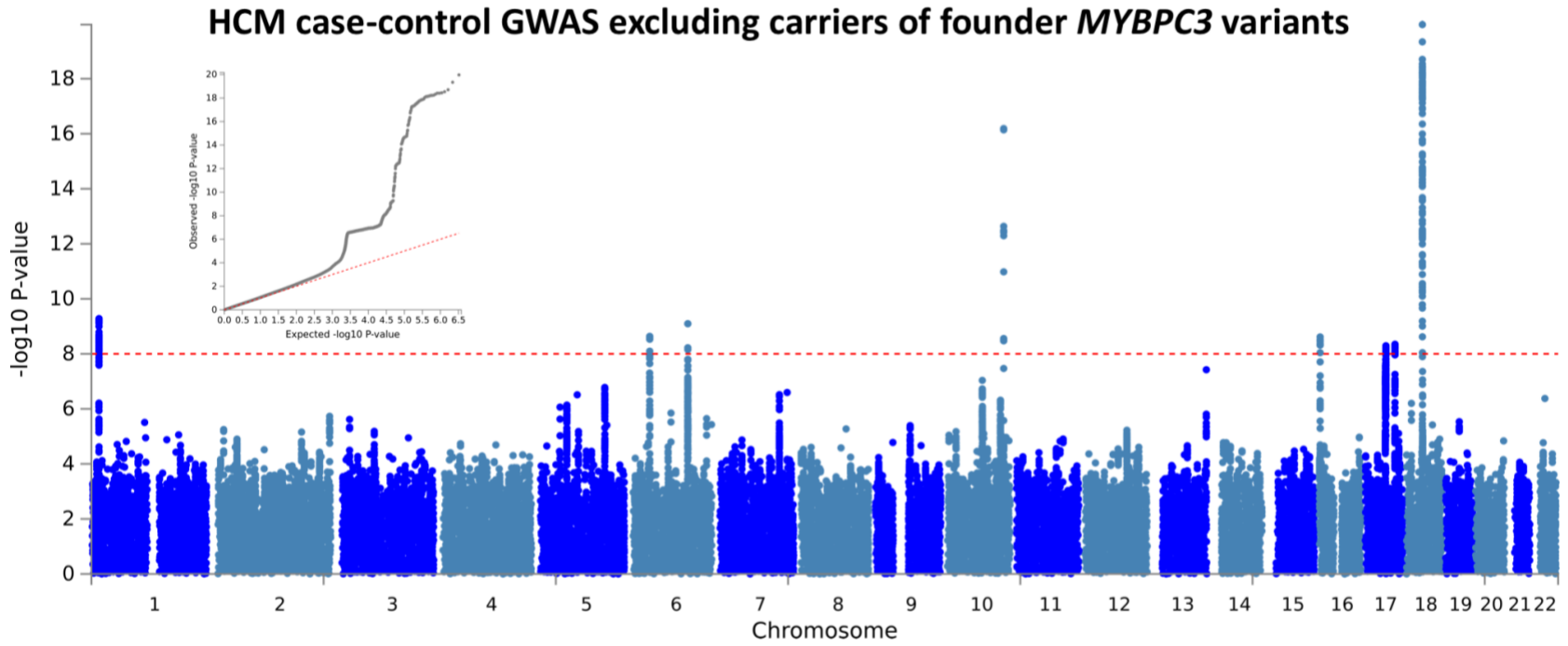
Supplementary Material

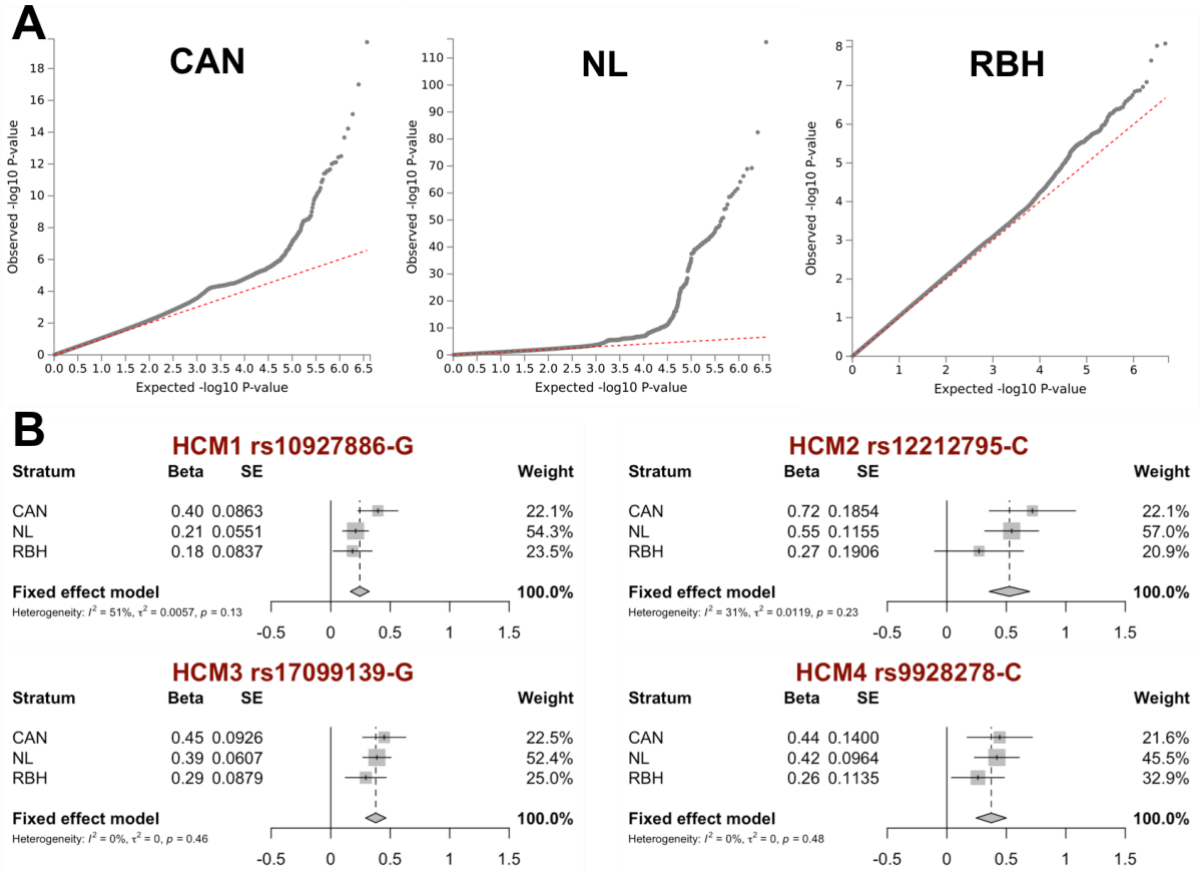
Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect

Tadros et al.

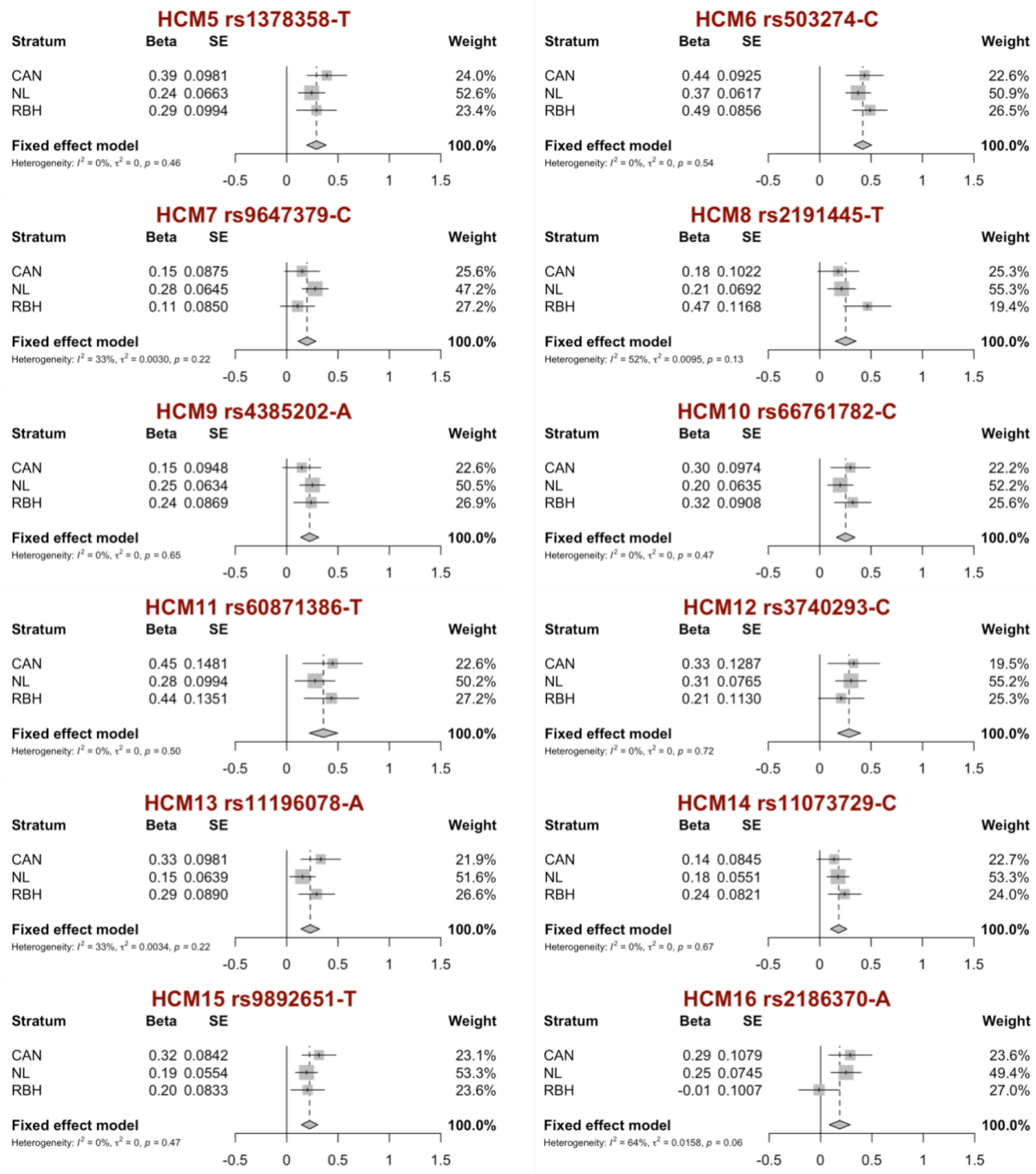
Supplementary Figure 1: Summary results of the HCM GWAS excluding HCM cases carrying *MYBPC3* founder variants, shown as a Manhattan plot. For the sake of this analysis, a *founder variant* was defined as a rare variant classified as pathogenic or likely pathogenic that was observed 10 or more times in the unrelated case cohort combining all strata. Sample size: 1445 cases and 6628 controls. Statistical analysis consisted of a fixed-effects meta-analysis of 3 case-control HCM GWAS using a frequentist test. Red dashed line shows the significance threshold of $P=1x10^{-8}$. Quantile-quantile (QQ) plot shown as insert. Note the broad signal around *MYBPC3* on chromosome 11 found in the primary analysis (**Figure 2, Panel A** in main text) disappears in this analysis excluding HCM cases with founder variants. This analysis identifies a total of 8 signals reaching statistical significance (6 are also identified in the primary single trait analysis, and the other 2 are identified in the multi-trait analysis; **Figure 2** in main text).



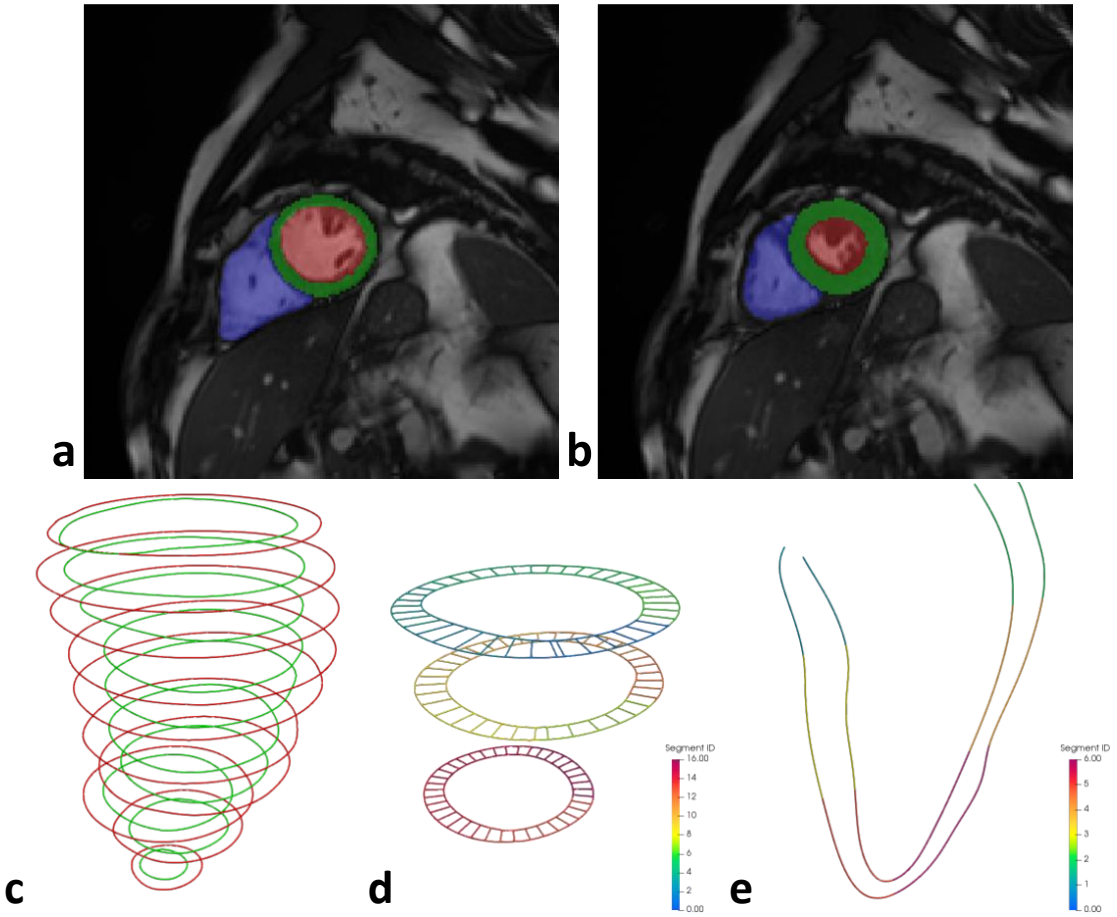
Supplementary Figure 2: HCM GWAS results per stratum (CAN, Canada; NL, Netherlands; RBH, Royal Brompton Hospital). **A**) Quantile-quantile plots per stratum. Note that the CAN and NL GWAS show deviation from the identify line partly due to SNPs tagging founder mutations in *TNNT2* (p.W287X in French Canadians) and *MYBPC3* (c.551dupT in French Canadians, and c.2373dupG, p.R943X and others in the Netherlands, **Supplementary Table 2**). Such SNPs are excluded from the meta-analysis given their low heterogeneity P (see Methods). **B**) Forest plots of the 16 HCM loci reaching statistical significance ($P<1x10^{-8}$) in the single trait meta-analysis (loci HCM1-HCM6) and multi-trait analysis (loci HCM7-HCM16). Lead SNP (rs identifier) and risk increasing allele are shown on top of each plot. The regression coefficient (Beta; marker) of the risk allele and standard error (SE, bar) obtained using a frequentist test are shown for each of the 3 strata (CAN, Canadian; NL, Dutch; RBH, Royal Brompton and Harefield) as well as the fixed effect model meta-analysis. The weight of each stratum is shown at the right side of the plot. Heterogeneity measures (I^2, τ^2 and Q statistic P-value) are shown at the bottom of each plot. Plots are constructed using the *forest* function of the R package *meta*.



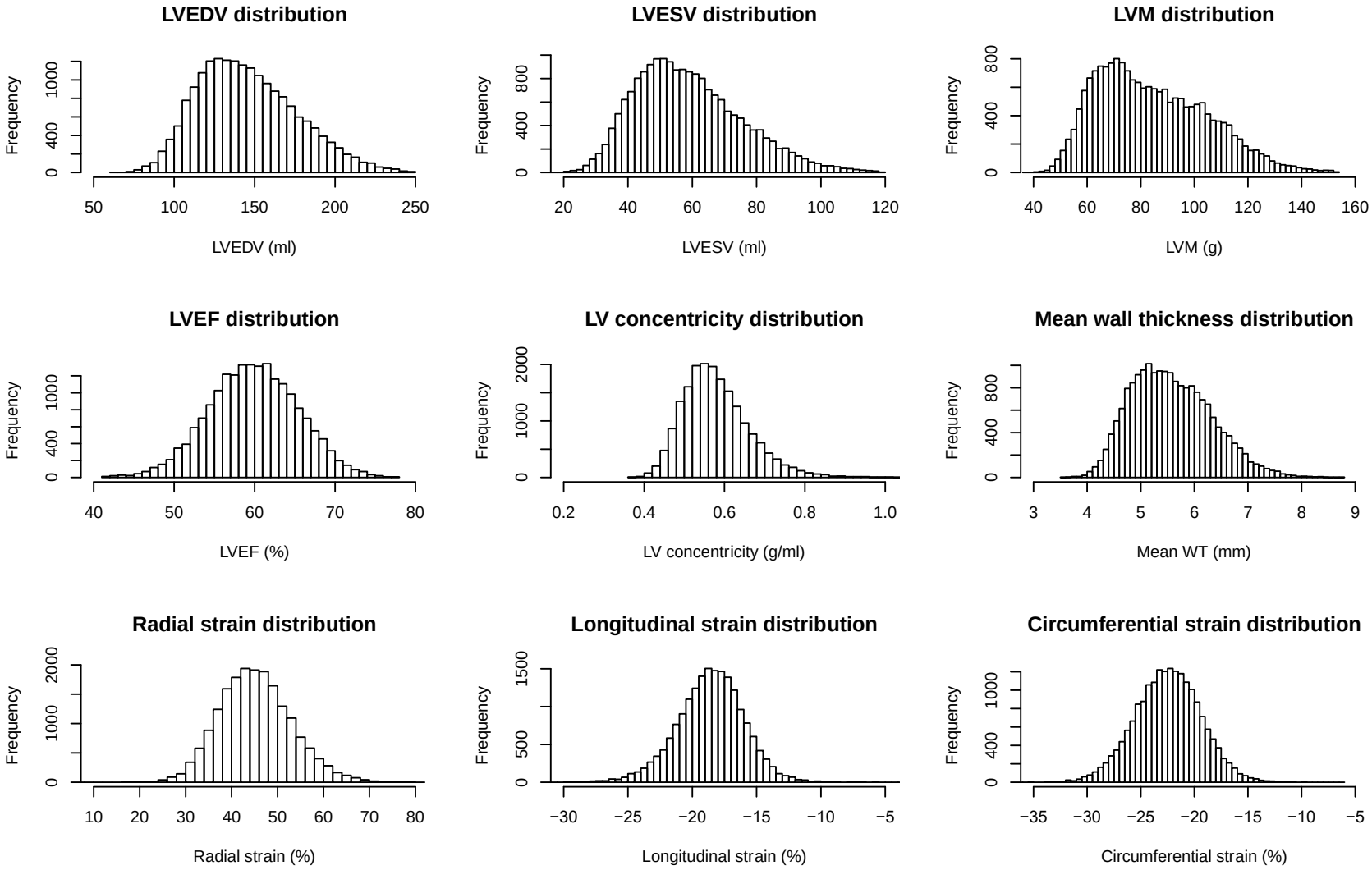
Supplementary Figure 2 (continued)



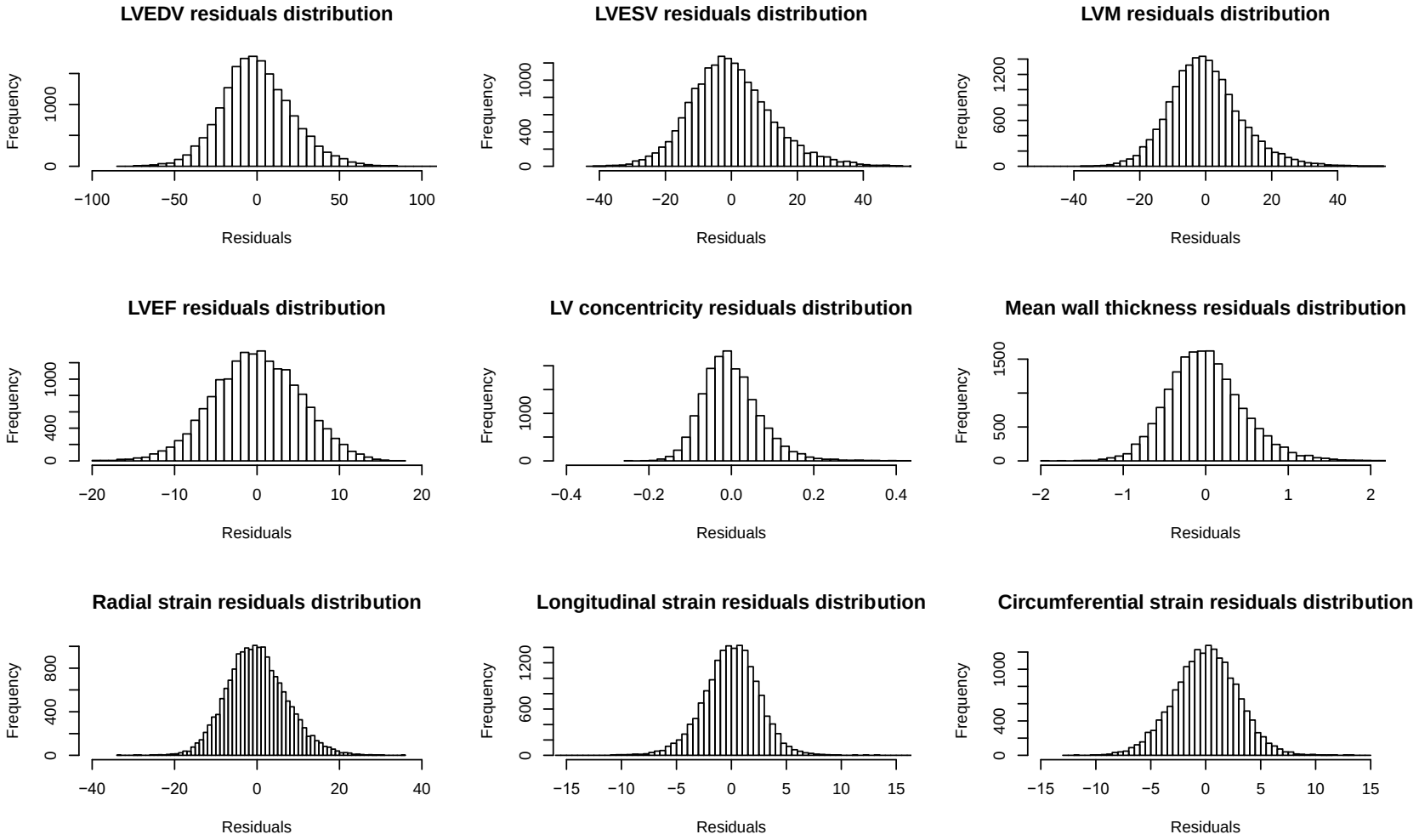
Supplementary Figure 3: Automated cardiac magnetic resonance (CMR) image analysis pipeline. Left ventricular (LV) volumes are derived from short-axis image segmentation shown as red shades at end-diastole (**a**) and end-systole (**b**). Myocardial wall thickness is measured using the distance between LV epicardial contour (red) and endocardial contour (green) at end-diastole (**c**). Three short-axis image slices are selected (**d**), including a basal slice at 75% LV location, a mid-cavity slice at 50% LV location and an apical slice at 25% LV location. The endocardial and epicardial contours are divided into 16 AHA segments, coded here by different colors. Motion tracking is performed on these short-axis image slices, warping the contours to each time frame across a cardiac cycle. Circumferential and radial strains are calculated using the change of length of the line segments shown. Long-axis 4 chamber view contours at end-diastole (**e**), divided into 6 AHA segments (coded by different colors). Longitudinal strain is calculated using the change of length of the line segments shown. Original CMR scans access through application 18545 reproduced by kind permission of UK Biobank.



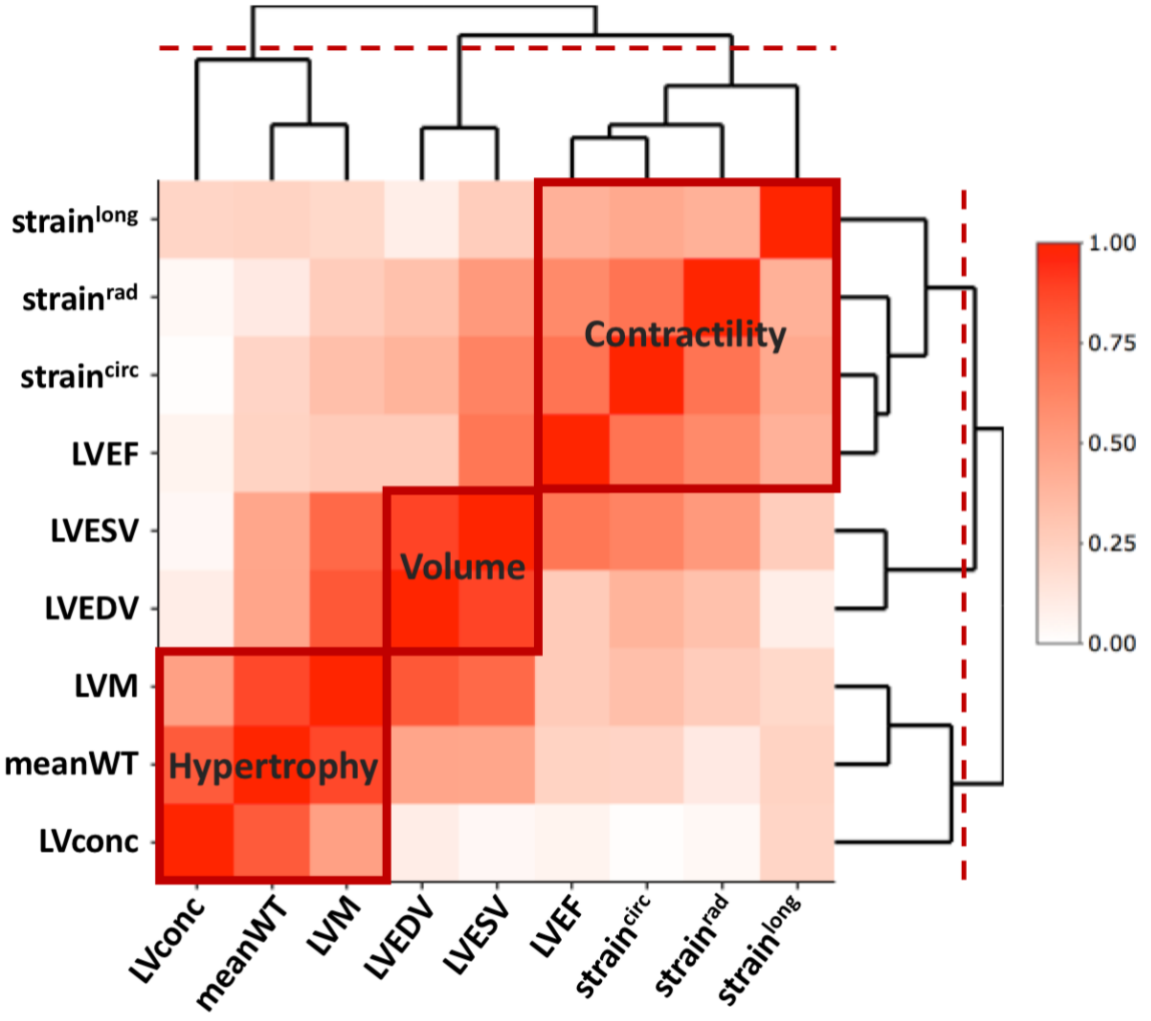
Supplementary Figure 4: Distribution of raw measures of the 9 left ventricular (LV) traits in the included UK Biobank CMR cohort (N=19,260). LVEDV, LV end-diastolic volume; LVEF, LV ejection fraction; LVESV, LV end-systolic volume; LVM, LV mass.



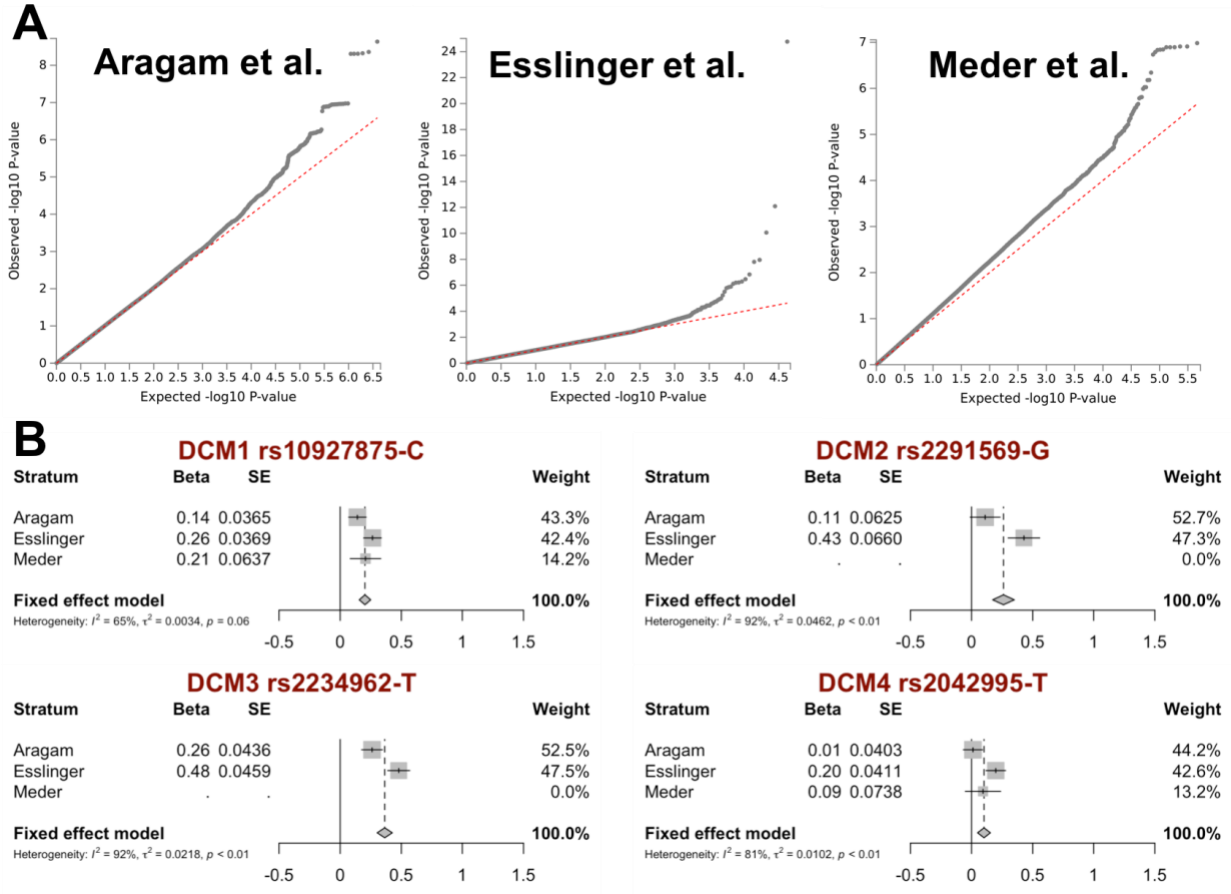
Supplementary Figure 5: Distribution of residuals from the regression models including covariates (age, sex, height, weight, and mean arterial pressure) for the 9 left ventricular (LV) traits in the included UK Biobank CMR cohort (N=19,260). LVEDV, LV end-diastolic volume; LVEF, LV ejection fraction; LVESV, LV end-systolic volume; LVM, LV mass.



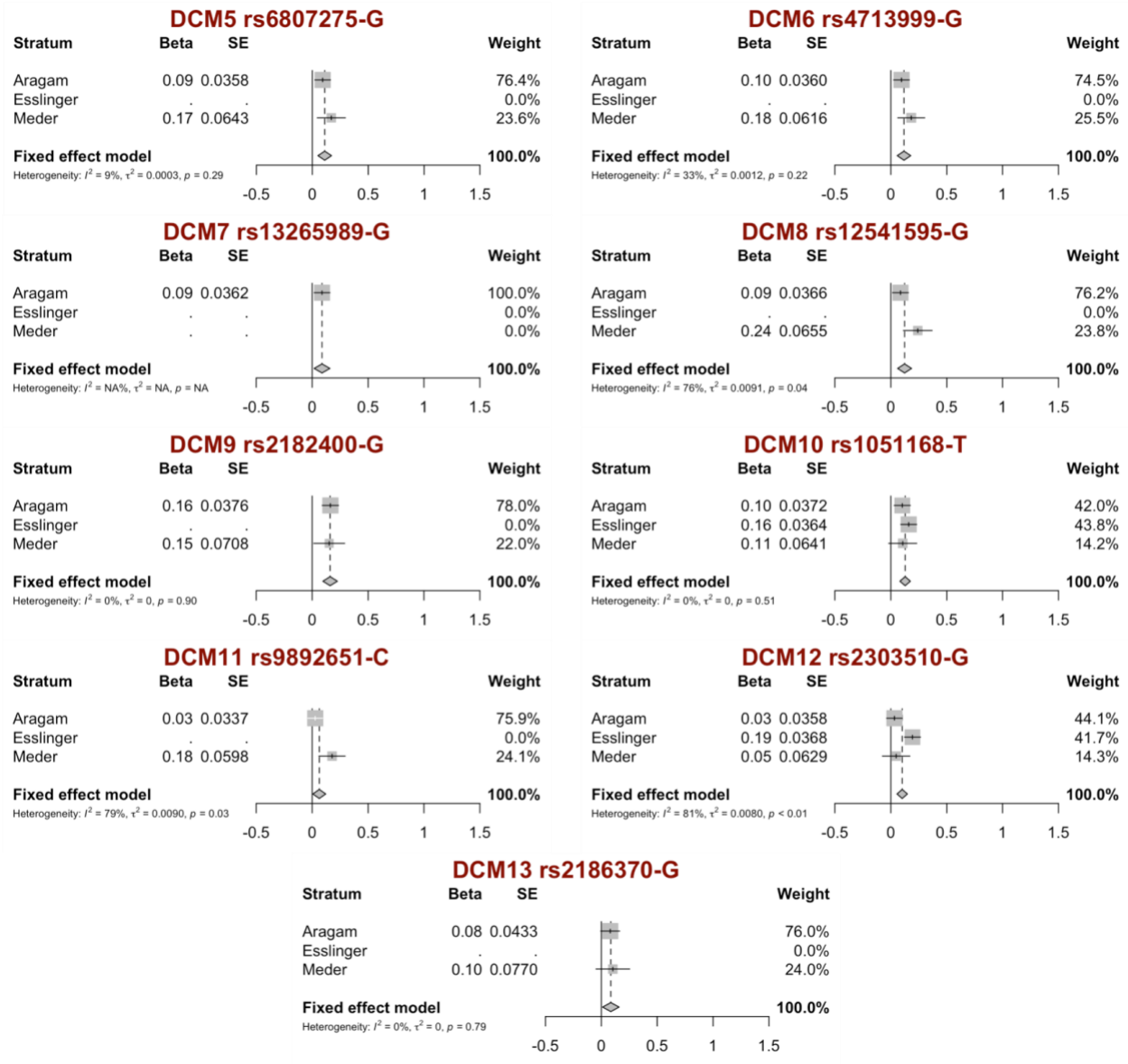
Supplementary Figure 6: Phenotypic correlation heatmap of 9 left ventricular (LV) traits derived from cardiac magnetic resonance. Absolute values of Pearson's correlation coefficients are represented as a color scale ranging from white to red (see scale in legend). Dendrograms constructed using Euclidean distance and complete hierarchical clustering. A dendrogram cut at a height of 1.5 (dashed red line) results in 3 LV phenotype clusters (thick dark red squares), corresponding to contractility, volume and hypertrophy markers. Pearson correlation coefficients are shown in **Supplementary Table 9**. LVconc, LV concentricity (LVM/LVEDV); LVEDV, LV end-diastolic volume; LVEF, LV ejection fraction; LVESV, LV end-systolic volume; LVM, LV mass; meanWT, mean LV wall thickness; $strain^{circ}$, global circumferential LV strain; $strain^{long}$, global longitudinal LV strain; $strain^{rad}$, global radial LV strain.



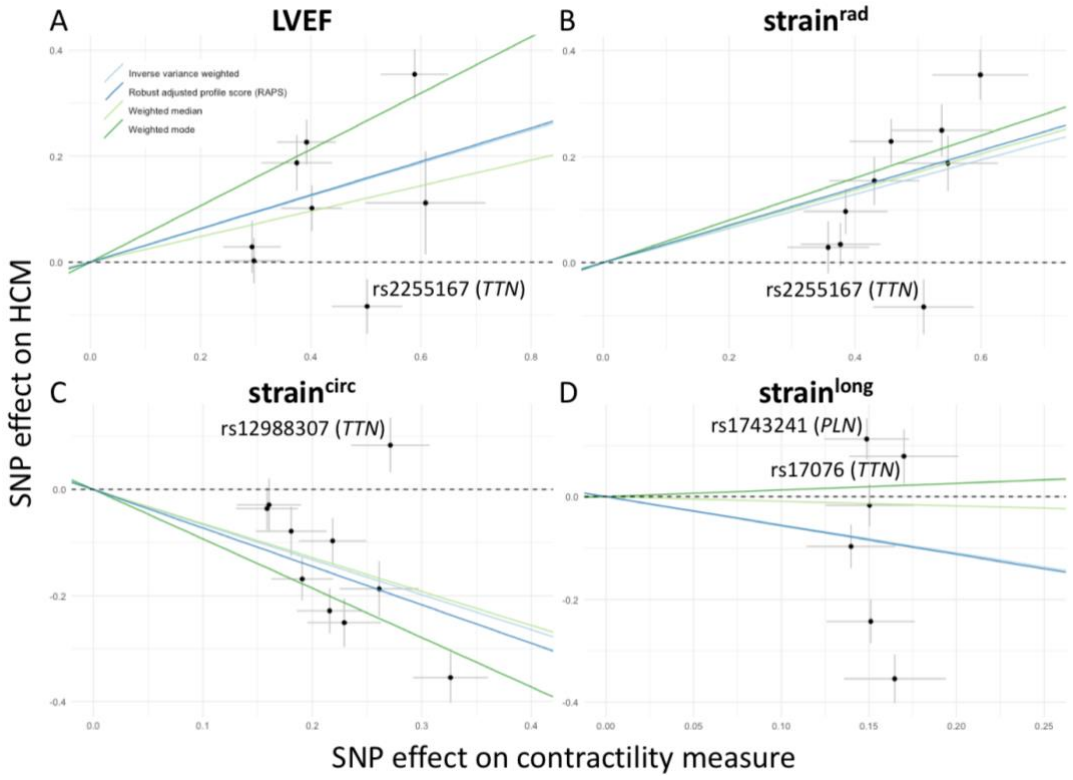
Supplementary Figure 7: DCM GWAS results per study identified by the first author (described in **Supplementary Table 5**). **A**) Quantile-quantile plots per study. **B**) Forest plots of the 13 DCM loci reaching statistical significance ($P<1x10^{-8}$) in the single trait meta-analysis (loci DCM1-DCM3) and multi-trait analysis (loci DCM4-DCM13). Lead SNP (rs identifier) and risk increasing allele are shown on top of each plot. The regression coefficient (Beta; marker) of the risk allele and standard error (SE, bar) reported in each published study are shown for each of the 3 published studies as well as the fixed effect model meta-analysis. The weight of each stratum is shown at the right side of the plot. Heterogeneity measures (I^2, τ^2 and Q statistic P-value) are shown at the bottom of each plot. Plots are constructed using the *forest* function of the R package *meta*.



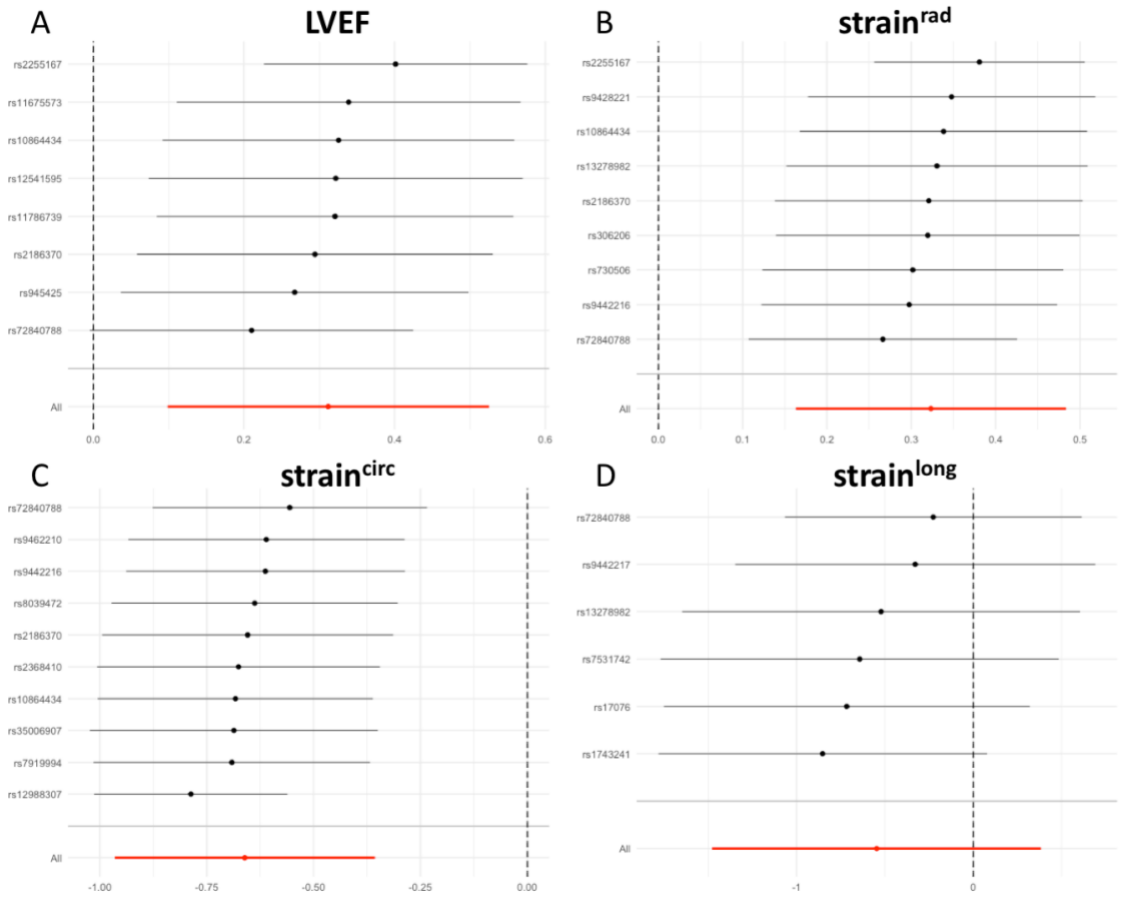
Supplementary Figure 7 (continued)



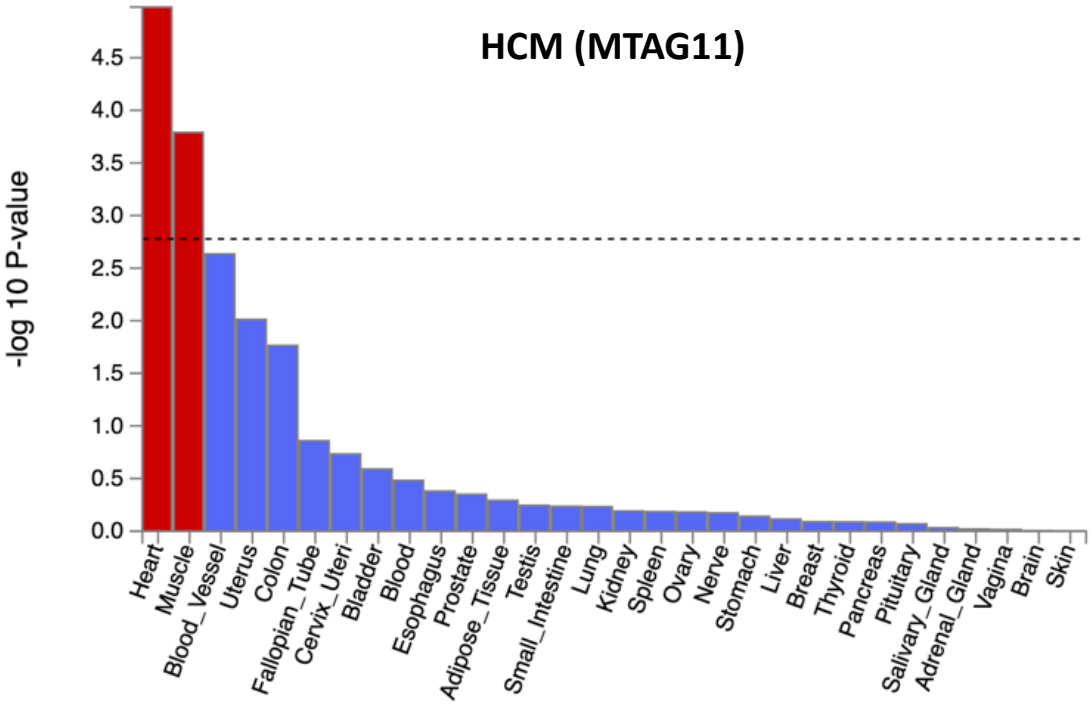
Supplementary Figure 8: Effect plots of two-sample Mendelian randomization (MR) analysis for contractility vs. HCM effects. Four left ventricular contractility measures are used as exposure variables: left ventricular ejection fraction (LVEF; **Panel A**), as well as global strain in the radial ($strain^{rad}$; **Panel B**), circumferential ($strain^{circ}$; **Panel C**) and longitudinal ($strain^{long}$; **Panel D**) directions. Independent instrument SNPs are selected based on their association at $P<5x10^{-8}$ with the corresponding contractility measure in the multi-trait analysis (MTAG9, see Methods). X-axis represents the effect size of the contractility measure for the trait increasing allele. Y-axis represents the effect size of the same allele in the HCM case-control meta-analysis. Each marker represents a single SNP, where the center and bars represent the regression coefficient and standard errors, respectively. Colored lines represent the fits using different MR methods (**legend** in **Panel A**). Analysis suggests a causal relation between increased contractility and risk of HCM. The negative associations for $strain^{circ}$ (Panel C) and $strain^{long}$ (Panel D) are expected where increasingly negative values reflect increased contractility. SNPs with divergent effects are identified with rs number and gene most likely to drive the association. Note the *TTN* locus which consistently shows divergent effects, possibly due to pleiotropic effects on left ventricular structure and function (see main text).

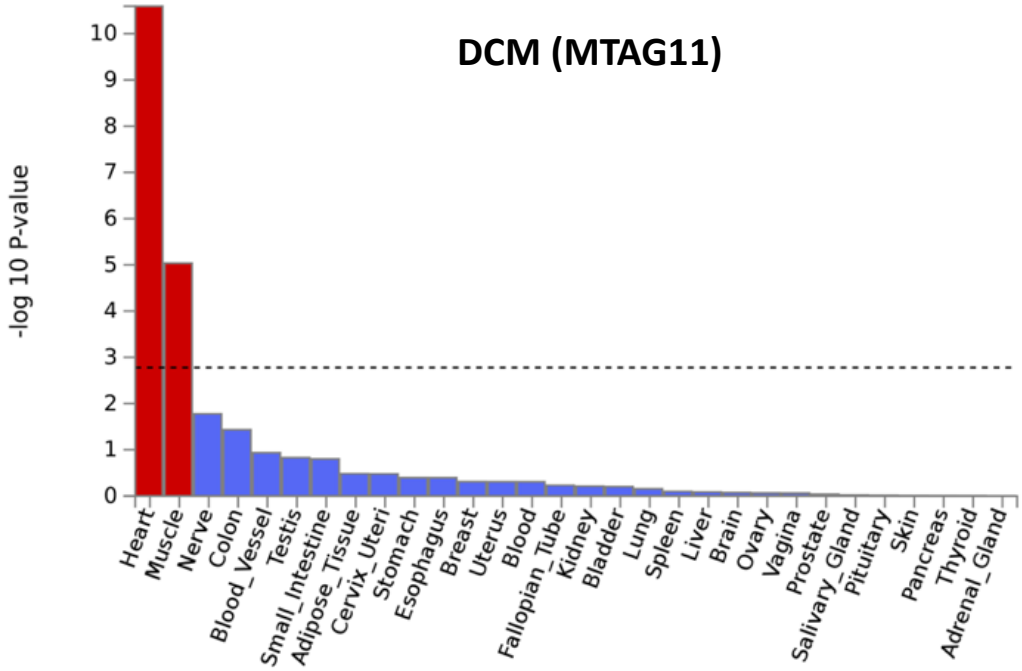


Supplementary Figure 9: Mendelian randomization leave-one-out sensitivity analysis using the inverse variance weighted (IVW) method, for 4 left ventricular contractility measures as exposure variables: left ventricular ejection fraction (LVEF; **Panel A**), as well as global strain in the radial ($strain^{rad}$; **Panel B**), circumferential ($strain^{circ}$; **Panel C**) and longitudinal ($strain^{long}$; **Panel D**) directions. Independent instrument SNPs are selected based on their association at $P<5x10^{-8}$ with the corresponding contractility measure in the multi-trait analysis (MTAG9, see methods). The outcome variable is HCM. The IVW MR method is used to estimate the causal effect of contractility on HCM (x-axis) leaving out one SNP at a time (y-axis) to assess whether the association is driven by a single SNP. The IVW effect estimate (center mark) is shown along with the standard error (bars). The overall effect estimate for each contractility measure is shown in the corresponding panel in red. The LVEF-HCM association (**A**) is significant overall and when excluding each SNP at a time, except when excluding rs72840788 (P=0.055). $Strain^{rad}$-HCM (**B**) and $strain^{circ}$-HCM (**C**) associations are significant overall and remain so in all leave-one-out analyses. $Strain^{long}$-HCM association (**D**) is not significant.



Supplementary Figure 10: MAGMA gene property analysis for tissue specificity for HCM and DCM using the multi-trait HCM (**A**) and DCM (**B**) summary statistics and the Genotype-Tissue Expression (GTEx) dataset v8. Tissues listed by decreasing order of -log(P). Gene property analysis is based on a regression model, and a one-sided test, implemented in FUMA using the default setting. The red bars represent tissues crossing the Bonferroni corrected P-value threshold (dashed line).





Supplementary Note

Diagnostic genetic testing and rare variant classification

Hypertrophic cardiomyopathy (HCM) cases underwent targeted sequencing of genes associated with HCM, as per local practice at the time of analysis. Rare variants detected through sequencing in each of the contributing cohorts of this study were centrally assessed for pathogenicity according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines[1], using an adapted version of the CardioClassifier resource.[2] Variants were subsequently classified as pathogenic, likely pathogenic, uncertain significance, likely benign or benign by expert assessment based on the combination of rules activated as described by Richards et al.[1]

The following ACMG/AMP rules were applied:

Population frequency:

- PM2: Applied if the filtering allele frequency in the Genomes Aggregation Database (gnomAD) exomes dataset (version 2.1) was below the calculated maximum tolerated allele frequency for HCM. The threshold applied was $4x10^{-5}$, based on a disease prevalence of 1 in 500, allelic heterogeneity of 0.02 and penetrance of 0.5 (calculated with https://www.cardiodb.org/allelefrequencyapp).[3]
- BS1: Applied if the filtering allele frequency in the gnomAD exomes dataset (version 2.1) was equal to or greater than $1x10^{-4}$.
- BA1: Applied if the filtering allele frequency in the gnomAD exomes dataset (version 2.1) was greater than 0.001.

General rules:

- PVS1: Truncating variants, i.e. frameshift, nonsense, splice donor and splice acceptor variants, in the following genes: *MYBPC3*, *TNNT2*, *PLN*, *FHL1*. Additionally, the ClinVar entries for splice region variants in *MYBPC3* outside of the canonical splice sites were assessed for published functional validation of disruption to splicing, with PVS1 applied for the following variants (ENST00000545968): c.1224-19G>A, c.1624+4A>T, c.3190+5G>A, c.927-9G>A.

- PP3/BP4: multiple lines of computation evidence support or refute a deleterious effect, as applied by CardioClassifier.
- PM4/BP3: a protein length change as a result of an in-frame deletion or insertion in a non-repeat region or within a region annotated by repeat masker.
- PS1: same amino acid change as a previously established pathogenic variant (multiple ClinVar submissions with no conflicting evidence).
- PM5: novel missense change at an amino acid residue where a different missense change has previously been established as pathogenic (multiple ClinVar submissions with no conflicting evidence).

Case-control analysis:

- PS4: Variants enriched in case cohorts compared to population controls. The HCM cohorts used were previously published clinical genetics cohorts from the Oxford Molecular Genetics Laboratory, UK and the Laboratory of Molecular Medicine, Boston, USA comprising up to 6,179 probands (enrichment was defined as a Fisher's exact test $P<1.79x10^{-6}$, after multiple testing correction as applied by CardioClassifier). PS4 was also applied for any additional variants enriched in the HCM cohort described in this study (present in ≥ 3 cases and $P<1.9x10^{-4}$ after Bonferroni multiple testing correction).
- PM1: This rule was applied based on pre-computed etiological fraction (EF) values for rare (PM2 rule activated), non-truncating variants as described by Walsh et al.[4] This approach defines the prior probability, as calculated through case-control cohort analysis, that a variant in a particular gene/protein region is pathogenic. The rules applied are PM1_strong (EF≥0.95), PM1_moderate (0.9≤EF<0.95) or PM1_supporting (0.8≤EF<0.9). For the HCM genes, the rules were applied based on the findings of Walsh et al[4], with the following rules applied: PM1_strong for *MYH7* (amino acid residues 167-931), *MYBPC3* (residues 485-502 and 1248-1266), *TNNI3* (residues 141-209), *TNNT2* (residues 79-179) and *TPM1* (all residues), PM1_moderate for *MYL3* (residues 143-180) and PM1_supporting for *CSRP3* (residues 44-71) and *MYBPC3* (remaining residues).

Rules that rely on manual curation of published evidence were not applied, i.e. co-segregation of variants with disease in family pedigrees (PP1/BS4), de novo inheritance with/without confirmed

paternity and maternity (PS2/PM6) and well established functional studies showing a deleterious or no effect (PS3/BS3). The results of this centralized variant classification are shown in **Supplementary Table 2**.

HCM GWAS – Dutch stratum (NL)

Unrelated Dutch HCM cases were included from the Amsterdam University Medical Center (location AMC, Amsterdam), Erasmus Medical Center (EMC, Rotterdam) and the University Medical Center Groningen (UMCG, Groningen). The control group consisted of a previously genotyped Dutch control population from project Mine (cohort NL4, as described in van Rheenen et al.[5]). All cases and controls underwent genome-wide array genotyping on an Illumina Infinium BeadChip. Cases from the AMC and all controls were genotyped on the Human OmniExpress array, while cases from EMC and UMCG were genotyped on the Illumina Global Screening Array (GSA). Cases were genotyped at the Genome analysis center at Helmholtz Zentrum München, Germany. All downstream analyses were performed on the Dutch SurfSara computer cluster. QC was performed using Plink v1.9 and in-house scripts. QC was first performed in each of the 3 genotyped cohorts separately (OmniExpress cases, GSA cases and OmniExpress controls). The 3 datasets were then merged, excluding non-overlapping single nucleotide polymorphisms (SNPs), and additional QC was performed. SNPs were excluded if ambiguous (A/T or C/G) or had missingness rates >0.05, Hardy-Weinberg equilibrium test $P<10^{-6}$ for controls or $P<10^{-10}$ for cases or minor allele frequency (MAF) <0.05. All SNPs were mapped on the positive strand of the GRCh37 build prior to merging. After merging the 3 cohorts, SNPs with divergent missingness rates between cases and controls ($P<10^{-4}$ using Fisher's exact test with Lancaster's mid-p correction), and those with missingness rates >0.02 in cases *or* controls were excluded. Samples with missingness >0.03, inbreeding coefficient $|F|\geq 0.1$, as well as those with unresolved genotype-phenotype sex mismatch were excluded. Identity by descent (IBD) analysis was performed using the --genome function in Plink. Samples with excessive rates of proportional IBD (PI-HAT)>0.05 were excluded. Pairs of related samples up to the 3rd degree (PI-HAT>0.125) were resolved by keeping only one individual per pair, favouring cases over controls and severely affected cases over less affected ones (based on maximal left ventricular wall thickness, LVWT). Genotypic principal components analysis (PCA) was then performed combining all cases and controls with samples from the 1000 Genomes Project. Non-European cases and controls were excluded based on inspection of scatter plots of the first 4 principal components. PCA was then performed again using the remaining

cases and controls and outliers were excluded, and PCA was again repeated to compute eigenvectors used as co-variates in the regression analysis. Genome-wide imputation was performed using Eagle2 phasing, Minimac3 and the Haplotype reference consortium (HRCr1.1) panel implemented on the Michigan Imputation Server v.1.0.2.[6] After imputation, only SNPs with MAF>0.01 and a Minimac R^2>0.5 were included in the association analysis.

HCM GWAS – Royal Brompton & Harefield Hospitals stratum (RBH)

Unrelated British HCM cases from the Royal Brompton & Harefield Hospitals NHS Trust Cardiovascular Research Biobank, and healthy controls from the UK Digital Heart Project[7] were included. The presence of HCM was excluded using cardiac magnetic resonance (CMR) imaging in all controls. Genotyping was performed in 3 batches at 2 centres (the Sanger Institute, London, and Duke-NUS in Singapore) using the Illumina Human OmniExpress beadchip. All downstream QC and data analyses were performed centrally. Data quality control was performed using Plink v1.9 and in-house scripts. We first mapped all SNPs to the positive strand of GRCh37 build. Pre-imputation SNP-level QC excluded SNPs with MAF<0.01, Hardy-Weinberg equilibrium test $P<10^{-7}$, missingness rate >0.05. Sample QC excluded samples with sex mismatch, heterozygosity rate >3 standard deviations (SD) from the mean, missingness rate >0.03. We aligned study genotypes with those from the HapMap III cohort, and used PCA to identify genotypically Caucasian samples to take forward for imputation, on the basis of cut-offs defined by the mean and SD of each HapMap ethnicity cohort. Batches were merged pre-imputation, excluding SNPs which failed QC checks in any batch. We also checked for differential missingness and excluded any SNPs with differential missingness test $P<1x10^{-7}$. We used the combined UK10K + 1000 Genomes Project dataset as the imputation reference panel. This has a proven record of accuracy for UK study cohorts, across a range of MAF.[8] We pre-phased study genotypes using SHAPEIT[9] (v2.r790) and imputed the genotype data using IMPUTE2[10] (v2.3.2). We finally performed a number of post-imputation QC steps per-SNP and per-sample, excluding SNPs with an INFO criterion score <0.4, MAF<1%, Hardy-Weinberg equilibrium $P<10^{-7}$. We then removed related individuals from the dataset, using a PI-HAT cut-off of 0.187. After the QC steps, we performed another principal components analysis to generate the eigenvectors used as covariates in the regression analysis.

HCM GWAS – Canadian stratum (CAN)

Unrelated Canadian HCM cases were included from the Montreal Heart Institute (MHI, Montreal) and the London Health Sciences Center (LHSC, London, Canada). All cases were genotyped on the Illumina GSA at the MHI Beaulieu-Saucier Pharmacogenomics Center (PGx). The control groups consisted of 1) samples from the MHI biobank previously genotyped[11] on the Illumina Multi-Ethnic Genotyping Array (MEGA), and 2) healthy, disease-free adults ascertained in London, Ontario who were genotyped on the Illumina GSA at PGx. Data analysis was performed on the Compute Canada cluster. QC was performed using Plink v1.9 and in-house scripts. QC was first performed in each of the 2 genotyped cohorts separately (GSA cases and controls, and MEGA controls). The 2 datasets were then merged, excluding non-overlapping SNPs, and additional stringent QC was performed. SNPs were excluded if ambiguous (A/T or C/G) or had missingness rates >0.05, Hardy-Weinberg equilibrium test $P<10^{-6}$ for MEGA or $P<10^{-10}$ for GSA or MAF <0.05. All SNPs were mapped on the positive strand of the GRCh37 build prior to merging. After merging the 2 sets, SNPs with divergent missingness rates between cases and controls ($P<10^{-3}$ using Fisher's exact test with Lancaster's mid-p correction), those with missingness rates >0.01 in cases *or* controls, and those with Hardy-Weinberg equilibrium test P<0.01 were excluded. Samples with missingness rate >0.02, inbreeding coefficient $|F|\geq 0.1$, as well as those with unresolved genotype-phenotype sex mismatches were excluded. IBD analysis was performed using the --genome function in Plink. Pairs of related samples up to the 3rd degree (PI-HAT>0.125) were resolved by keeping only one individual per pair, favouring cases over controls and severely affected cases over less affected ones (based on maximal LVWT). Genotypic PCA was then performed combining all cases and controls with samples from the 1000 Genomes Project. Non-European cases and controls were excluded based on inspection of scatter plots of the first 4 principal components. PCA was then performed again using the remaining cases and controls and outliers were excluded, and PCA was again repeated to compute principal components used as co-variates in the regression analysis. Genome-wide imputation was performed using Eagle2 phasing, Minimac4 and the HRCr1.1 panel implemented on the Michigan Imputation Server v1.2.1.[6] After imputation, only SNPs with MAF>0.01 and a Minimac R^2>0.5 were included in the association analysis.

Description of association studies in dilated cardiomyopathy included in the meta-analysis

A meta-analysis of 3 published case-control association studies[12-14] of dilated cardiomyopathy (DCM) was performed. The included studies are described below and in **Supplementary Table 5**.

The largest included study of Esslinger and colleagues[13] (2796 cases with sporadic DCM and 6877 controls) used the Illumina HumanExome BeadChip with only 116,855 mostly coding SNPs included in the analysis. The included cases had a diagnosis of DCM defined as a low left ventricular ejection fraction (LVEF; <45% by echocardiography or 2 SD below the age- and sex-adjusted average by CMR) AND enlarged left ventricular end-diastolic volume/diameter (>117% of value predicted from age and body surface area on echocardiography, or 2 SD from the age- and sex-adjusted mean by CMR), in the absence of significant coronary artery disease (CAD) or intrinsic valvular disease, documented myocarditis, systemic disease, sustained arterial hypertension, or congenital malformation, and in the absence of family history of DCM.[13]

The second largest study by Aragam and colleagues[12] included 1816 cases with *non-ischemic cardiomyopathy* and 388,326 controls from the UK Biobank (UKBB). *Non-ischemic cardiomyopathy* was defined as hospitalization for or death due to ICD-10 code for dilated cardiomyopathy or left ventricular failure (I42.0, I50.1); or hospitalization due to ICD-9 code for left heart failure (4281); excluding individuals with history of CAD (as defined in Aragam et al[12]), or history of hypertrophic cardiomyopathy during verbal interview with trained nurse, or hospitalization for or death due to ICD-10 code for hypertrophic cardiomyopathy (I42.1, I42.2). The diagnosis of DCM could not be accurately assessed in this study, but it is expected that most cases with *non-ischemic cardiomyopathy* would qualify for a diagnosis of DCM. The overlap of genetic loci identified by Aragam and colleagues in *non-ischemic cardiomyopathy* and those previously identified in *DCM* supports a strong overlap between the 2 phenotype definitions.

The third study was published by Meder and colleagues[14] and included 909 cases with DCM and 2120 controls. DCM was defined as reduced left ventricular systolic function with a LVEF < 45%, assessed by echocardiography or left ventricular angiography, in the absence of relevant CAD.

Left ventricular traits GWAS - automated image analysis

We segmented short and long axis cine images using a fully convolutional network[15], trained on manual annotations of 3975 subjects. See **Supplementary Figure 3** for further details. All image segmentations were manually quality controlled by an experienced cardiologist. The segmentation screenshots for short-axis and long-axis end-diastolic and end-systolic frames were visually inspected. Bad segmentations, images with insufficient coverage of the LV or missing anatomical structures were

discarded. For motion tracking, subjects with failed image registration or outlier peak global strain values (positive circumferential or longitudinal strain, or negative radial strain) were discarded.

LV end-diastolic (LVEDV) and end-systolic (LVESV) volumes and ejection fraction (LVEF, defined as [LVEDV-LVESV]/LVEDV) were derived from these segmentations. The LV myocardial mass (LVM) was calculated from the myocardial volume using a density of 1.05 g/mL. LV concentricity (LVconc) was defined as LVM/LVEDV. The LV wall thickness (WT) was measured at end-diastole, as illustrated in **Supplementary Figure 3**. The myocardium was divided into 16 AHA segments.[16] Maximum and mean wall thickness was derived for each AHA segment. Overall mean wall thickness (meanWT) per subject was calculated from the mean of each of the 16 mean wall thickness measurements.

Motion tracking was performed using non-rigid image registration[17] between successive timeframes, using the MIRTK toolkit.[18] Inter-frame displacement fields were composed to obtain the displacement with respect to a reference frame (the end-diastolic frame, or frame 0). To avoid drift effect due to accumulation of registration errors[19], motion tracking is performed twice – along the forward direction (tracking starting from frame 0 to frames 1, 2, 3,…) and backward direction (tracking from frame 0 to frames *T*-1, *T*-2, *T*-3, …) where *T* denotes the total number of frames per cardiac cycle). The average displacement field is calculated by weighted averaging of the forward and backward displacement field such that, for a frame at the start of the cardiac cycle, the forward displacement field will have a higher weight, whereas for a frame at towards the end of the cardiac cycle, the backward displacement field will have a higher weight. Four image slices were used for this: in the short axis, a basal slice at 75% LV location, a mid-cavity location and an apical slice at 25% LV location, and in the long axis, a 4-chamber view as previously described.[20,21] The myocardial contours on the three slices were divided into the 16 segments. Based on the displacement field from motion tracking, myocardial contours at the end-diastolic frame were warped onto each time frame of the cardiac cycle. Circumferential, radial and longitudinal strains were calculated for each time frame based on the change of length for each line segment[22], using the equation $E = \frac{\Delta L}{L}$, where *E* is the strain, ΔL is the change in length of the line segment, and *L* is the starting (end-diastolic) length of the line segment. Peak strain for each AHA segment, and *global* peak strain were calculated in radial ($\text{strain}^{\text{rad}}$), longitudinal ($\text{strain}^{\text{long}}$) and circumferential ($\text{strain}^{\text{circ}}$) directions.[22] Note that in contrast to $\text{strain}^{\text{rad}}$, both $\text{strain}^{\text{long}}$ and $\text{strain}^{\text{circ}}$ are negative values where more negative values reflect higher contractility. As such, we will often present results for $-\text{strain}^{\text{long}}$ and $-\text{strain}^{\text{circ}}$ to facilitate

interpretation of directionality.

Left ventricular traits GWAS - genotyping and imputation

Further details of the genotyping and QC strategy employed by UKBB were previously published.[23] Genotypes were called from 2 purpose-built arrays: the UK Biobank Axiom Array (825 927 markers) and UK BiLEVE Axiom Array (807 411 markers). These directly-called genotypes were imputed to more than 90 million variants using the haplotype reference consortium (HRC) and the UK10K reference panels. See UKBB website[24] for details of imputation and genotype quality control performed by a collaborative group headed by the Wellcome Trust Centre for Human Genetics. We excluded samples with outlying heterozygosity or missingness rates, as defined by UKBB and those with mismatches between the genotypic and recorded sex. We excluded SNPs failing UKBB protocols (filtered per batch by Hardy-Weinberg equilibrium and missingness)[23], those with imputation INFO score <0.3 or MAF <1%, or with missingness >0.1.

Locus Annotation

All loci associated with HCM, DCM and LV traits in the single trait or multi-trait analyses were annotated using lookup for 1) Proxy coding variants, 2) cis-expression quantitative trait loci (eQTL), 3) cis-splice quantitative trait loci (sQTL), and 4) contact with gene promoters using Hi-C. The data used for the eQTL/sQTL analyses described in the manuscript were obtained from the GTEx Portal in January 2020. Loci with mapping problems in GRCh37, defined as a presence of an assembly update/patch overlapping the lead SNP were not annotated.

1) *Proxy coding variants*: We performed a lookup for all protein-altering coding gene variants that are LD proxies with the lead SNP at each locus. LD was defined using an R^2>0.5 in the European subset of the 1000 Genomes Project. We used LDlink[25] v3.9 to look for proxies of lead SNPs. Protein-altering coding variants were defined as missense, splice site (within 2 nucleotide of the exon-intron boundary), in-frame deletions/insertions, frameshift, and stopgain. Results are shown in **Supplementary Table 15**.
2) *Cis-eQTL*: We assessed whether the lead SNP at each locus is a cis-eQTL by performing a lookup in the Genotype-Tissue Expression project (GTEx) dataset (version 8). The lookup was restricted to protein coding genes and to tissues that are significantly associated with HCM and DCM in the MAGMA tissue specificity analysis (i.e. heart and skeletal muscle). Only

variants with a significant variant-gene association are reported, as defined by GTEx. Specifically, Beta distribution-adjusted empirical p-values from FastQTL[26] were used to calculate q-values, and a false discovery rate (FDR) threshold of ≤0.05 was applied to identify genes with a significant eQTL (i.e. *eGenes*). To identify the list of all significant variant-gene pairs associated with eGenes, a genome-wide empirical p-value threshold, p_t, was defined as the empirical p-value of the gene closest to the 0.05 FDR threshold. p_t was then used to calculate a nominal p-value threshold for each gene based on the beta distribution model (from FastQTL) of the minimum p-value distribution $f(p_{min})$ obtained from the permutations for the gene. Specifically, the nominal threshold was calculated as $F^{-1}(p_t)$, where F^{-1} is the inverse cumulative distribution. For each gene, variants with a nominal p-value below the gene-level threshold were considered significant and included in the final list of variant-gene pairs. The results of the eQTL lookup are shown in **Supplementary Table 16**.

3) *Cis-sQTL*: We assessed whether the lead SNP at each locus is a cis-sQTL by performing a lookup in GTEx version 8. A similar methodology was used as the one described for the cis-eQTL lookup. These results are shown in **Supplementary Table 17**.
4) *Hi-C*: We assessed whether associated loci interact with gene promoter regions using tissue specific 3D chromatin interaction (Hi-C) mapping, incorporated in FUMA. For this, Hi-C data from human left ventricle was used.[27] Associated genetic loci were defined as the genetic region harbouring variants in LD with the lead SNP (R^2≥0.6 in the European population of the 1000 Genomes Project). The promoter region of genes was defined by 250bp upstream and 500bp downstream of the transcription start site. Only protein-coding genes were taken into account. Interactions with a FDR ≤ 10^{-6}, spanning from an associated locus ("region 1") to the promoter region ("region 2") of a gene were retained. These results are shown in **Supplementary Table 18**.

We assessed whether the transcription start site of any gene associated with Mendelian cardiomyopathy is within 500kb of the HCM, DCM or LV traits GWAS genetic loci (R^2≥0.6 in the European population of the 1000 Genomes Project). The list of Mendelian cardiomyopathy genes was retrieved from the Genomics England PanelApp version 3.1.1 (https://panelapp.genomicsengland.co.uk) on February 16th, 2020. We only included genes with level

of evidence green (diagnostic-grade) or amber (borderline) in the following 3 panels: 1) DCM - teen and adult, 2) HCM - teen and adult, and 3) Cardiomyopathies including childhood onset.

For each lead SNP associated with HCM, DCM or LV traits, we also performed a lookup of the effect in the single-trait analysis of the other phenotypes. If the lead SNP was not available, a proxy with the highest R^2 in the European subset of the 1000 Genomes Project dataset was used for the lookup, with a minimal R^2 of 0.5. Since the DCM GWAS meta-analysis included 2 datasets with non-imputed data including a large exome chip association study, the lookup in the DCM summary statistics was performed using the proxy with the largest sample size available with a R^2 of >0.5. The results of these cross-trait lookup are shown in **Figure 4** and **Supplementary Tables 4,7,8**.

References for Supplementary Material

1 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).

2 Whiffin, N. *et al.* CardioClassifier: disease- and gene-specific computational decision support for clinical genome interpretation. *Genet Med* **20**, 1246-1254, doi:10.1038/gim.2017.258 (2018).

3 Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* **19**, 1151-1158, doi:10.1038/gim.2017.26 (2017).

4 Walsh, R. *et al.* Quantitative approaches to variant classification increase the yield and precision of genetic testing in Mendelian diseases: the case of hypertrophic cardiomyopathy. *Genome Med* **11**, 5, doi:10.1186/s13073-019-0616-z (2019).

5 van Rheenen, W. *et al.* Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* **48**, 1043-1048, doi:10.1038/ng.3622 (2016).

6 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).

7 Schafer, S. *et al.* Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat Genet* **49**, 46-53, doi:10.1038/ng.3719 (2017).

8 Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* **6**, 8111, doi:10.1038/ncomms9111 (2015).

9 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181, doi:10.1038/nmeth.1785 (2011).

10 Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913, doi:10.1038/ng2088 (2007).

11 Low-Kam, C. *et al.* Variants at the APOE /C1/C2/C4 Locus Modulate Cholesterol Efflux Capacity Independently of High-Density Lipoprotein Cholesterol. *J Am Heart Assoc* **7**, e009545, doi:10.1161/JAHA.118.009545 (2018).

12 Aragam, K. G. *et al.* Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. *Circulation*, doi:10.1161/CIRCULATIONAHA.118.035774 (2018).
13 Esslinger, U. *et al.* Exome-wide association study reveals novel susceptibility genes to sporadic dilated cardiomyopathy. *PLoS One* **12**, e0172995, doi:10.1371/journal.pone.0172995 (2017).
14 Meder, B. *et al.* A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. *Eur Heart J* **35**, 1069-1077, doi:10.1093/eurheartj/eht251 (2014).
15 Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* **20**, 65, doi:10.1186/s12968-018-0471-x (2018).
16 Cerqueira, M. D. *et al.* Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* **105**, 539-542, doi:10.1161/hc0402.102975 (2002).
17 Rueckert, D. *et al.* Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* **18**, 712-721, doi:10.1109/42.796284 (1999).
18 Tadros, R. & Wilde, A. A. M. Revisiting the sensitivity of sodium channel blocker testing in Brugada syndrome using obligate transmittance. *Int J Cardiol* **245**, 183-184, doi:10.1016/j.ijcard.2017.07.006 (2017).
19 Tobon-Gomez, C. *et al.* Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. *Med Image Anal* **17**, 632-648, doi:10.1016/j.media.2013.03.008 (2013).
20 Taylor, R. J. *et al.* Myocardial strain measurement with feature-tracking cardiovascular magnetic resonance: normal values. *Eur Heart J Cardiovasc Imaging* **16**, 871-881, doi:10.1093/ehjci/jev006 (2015).
21 Schuster, A. *et al.* Cardiovascular magnetic resonance feature-tracking assessment of myocardial mechanics: Intervendor agreement and considerations regarding reproducibility. *Clin Radiol* **70**, 989-998, doi:10.1016/j.crad.2015.05.006 (2015).
22 Puyol-Anton, E. Fully automated myocardial strain estimation from cine MRI using convolutional neural networks. *IEEE International Symposium on Biomedical Imaging*, 1139–1143 (2018).
23 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
24 Lahrouchi, N. *et al.* Utility of Post-Mortem Genetic Testing in Cases of Sudden Arrhythmic Death Syndrome. *J Am Coll Cardiol* **69**, 2134-2145, doi:10.1016/j.jacc.2017.02.046 (2017).
25 Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557, doi:10.1093/bioinformatics/btv402 (2015).
26 Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479-1485, doi:10.1093/bioinformatics/btv722 (2016).
27 Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* **17**, 2042-2059, doi:10.1016/j.celrep.2016.10.061 (2016).