Genetic determinants of interventricular septal anatomy and the risk of ventricular septal defects and hypertrophic cardiomyopathy.

Mengyao Yu PhD[1,2]*, Andrew R. Harper MRCP DPhil[3,4,5]*, Matthew Aguirre AB[1,6], Maureen Pittman [7,8], Catherine Tcheandjieu DVM PhD[1,2,9], Dulguun Amgalan PhD[10,11], Christopher Grace PhD[3,4], Anuj Goel MBBS MSc[3,4], Martin Farrall FRCPath[3,4], Ke Xiao MS[12], Jesse Engreitz PhD[10,11], Katherine Pollard PhD[7,8,13], Hugh Watkins MD PhD[3,4], James R. Priest MD[1,2,13,14]

Affiliations:

1) Department of Pediatrics, Division of Pediatric Cardiology, Stanford University School of Medicine, Stanford, California, USA
2) Stanford Cardiovascular Institute, Stanford University, Stanford, California, USA
3) Radcliffe Department of Medicine, University of Oxford, Division of Cardiovascular Medicine, John Radcliffe Hospital, Oxford, UK.
4) Wellcome Centre for Human Genetics, Roosevelt Drive, Oxford, UK.
5) Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK
6) Department of Biomedical Data Science, Stanford Medical School
7) University of California, San Francisco, San Francisco, CA, USA
8) Gladstone Institute of Data Science & Biotechnology, San Francisco, CA, USA
9) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California, USA
10) Department of Genetics, Stanford University, Stanford, CA, USA
11) Basic Sciences and Engineering Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, CA, USA
12) College of Information & Computer Sciences at University of Massachusetts Amherst, Amherst, Massachusetts USA
13) Chan-Zuckerberg Biohub, San Francisco, California, USA
14) Current affiliation: Tenaya Therapeutics, South San Francisco, California, USA

*These authors contributed equally to this work

ABSTRACT

Background: The interventricular septum (IVS) plays a primary role in cardiovascular physiology and a large proportion of genetic risk remains unexplained for structural heart disease involving the IVS such as hypertrophic cardiomyopathy (HCM) and ventricular septal defects (VSD).

Objectives: We sought to develop a reproducible proxy of IVS structure from standard medical imaging, discover novel genetic determinants of IVS structure, and relate these loci to two rare diseases of the IVS.

Methods: We performed machine learning to estimate the cross-sectional area of the interventricular septum (IVS.csad) obtained from the 4-chamber view of cardiac MRI in 32,219 individuals from the UK Biobank. Using these extracted measurement of IVS.csad we performed phenome-wide association to relate this proxy measure to relevant clinical phenotypes, followed by genome-wide association studies and Mendelian Randomization.

Results: Automated measures of IVS.csad were highly accurate, and strongly correlated with anthropometric measures, blood pressure, and diagnostic codes related to cardiovascular physiology. A Single nucleotide polymorphism in the intron of *CDKN1A* was associated with IVS.csad (rs2376620, Beta 8.4 mm2, 95% confidence intervals (CI) 5.8 to 11.0, p=2.0e-10), and a common inversion incorporating *KANSL1* predicted to disrupt local chromatin structure was associated with an increase in IVS.csad (Beta 8.6 mm2, 95% CI 6.3-10.9, p=1.3e-13). Mendelian Randomization suggested that inheritance of a larger IVS.csad was causal for HCM (Beta 2.45

log odds ratio (OR) HCM per increase in SD of IVS.csad, standard error (SE) 0.48, pIVW = 2.8e-7) while inheritance of a smaller IVS.csad was causal for VSD (Beta -2.06 log odds ratio (OR) VSD per decrease in SD of IVS.csad, SE 0.75, pIVW = 0.006)

Conclusion: Automated derivation of the cross sectional area of the IVS from the 4-chamber view allowed discovery of loci mapping to genes related to cardiac development and Mendelian disease. Inheritance of a genetic liability for either large or small interventricular septum, appears to confer risk for HCM or VSD respectively, which suggests that a considerable proportion of risk for structural and congenital heart disease may be localized to the common genetic determinants of cardiovascular anatomy.

INTRODUCTION

Within the cardiac ventricles, the interventricular septum (IVS) separates deoxygenated pulmonary blood flow from oxygenated systemic blood flow and is a foundation of mammalian physiology[1]. The IVS is the site of two important diseases; ventricular septal defects (VSD) are among the most common forms of congenital heart disease, where a thin or abnormally developed IVS leaves a communication linking the lumens of the left and right ventricle which when untreated may lead to heart failure and pulmonary hypertension[2]. Alternatively, the increased thickness and altered geometry of the IVS in hypertrophic cardiomyopathy (HCM) may progress to the point of obstruction of the left-ventricular outflow tract.[3] While both VSD and HCM are primarily genetic in origin, Mendelian inheritance does not fully account for the risk of HCM[4] or for individual variability in disease amongst individuals carrying the same pathogenic variant. Monogenic disease causing VSD is very rare, and the genetic basis of VSD is simply not known in the majority of affected individuals[5].

The IVS has two primary developmental origins, with the membranous portion arising from mesenchymal lineages in the cardiac cushions[6] and the muscular portion developing from ingrowth of the primary heart tube[7]. A complex microanatomical orientation of cardiomyocyte fibers through the IVS links the function of the right and left ventricles[8]. After adaptation to postnatal circulation over the first months of life, the mature IVS has a complex relationship with normal physiology, with the thickness of the IVS increased in both exercise capacity[9] as well as in hypertension[10].

Here we derive a simple automated measure of the IVS derived from MRI imaging of the heart in the 4-chamber view, IVS cross-sectional area at diastole (IVS.csad). We show that IVS.csad is correlated with standard clinical measures and is useful as suitable proxy for the architechture of the IVS. We report genetic correlates of IVS.csad in 32,219 individuals derived from cardiac MRI of the UK Biobank. Using Mendelian Randomization, we describe new causal relationships linking smaller IVS.csad with risk for VSD and larger IVS.csad with risk for HCM.

METHODS

Ethics statement. The NHS National Research Ethics Service (ref: 11/NW/0382) granted ethical approval for the distribution of deidentified imaging, genetic, and medical record data from the UK Biobank (UKB) for any qualified researcher.

UK Biobank cohort and IVS cross sectional area measurement from cardiac MRI

We built a U-Net segmentation model of cine images of the 4-chamber (4Ch) view of the heart with pre-trained weights from VGG11, which was further trained on 60 hand-labeled 4Ch images, validated on 20 hand-labeled images, and tested on 20 hand-labeled images[11]. This approach yielded a validation dice score of 91.2% and a test dice score of 93.8% which were equivalent to differences observed between expert human annotators in an open-source deep learning framework upon which this work is based[12]. We then applied the segmentation model to 32,219 4Ch images from the UKB and generated masks for all 4 chambers including the left and right ventricles and the left and right atrium. Following generation of the masks from the trained U-Net segmentation model, we used another function to measure the interventricular septal areas. We first located the base of the atriums and apex of the ventricles along the medial axis of the chambers, and secondly we located the annuli of the atrioventricular valves using the intersecting lines of the atriums and the ventricles, then finally we use the contour lines of the masks of atriums and ventricles between the annulus lines and the base of the atriums or the apex of the ventricles to enclose the area of the interventricular septum. The metadata in the dicom measurements were used to convert from $pixels^2$ to mm^2. The frame representing end-diastole was obtained by selecting the image frame with the largest estimate of left ventricular volume provided by the U-Net segmentation algorithm.

To exclude outliers related to imaging error or methodological inaccuracy, automated measurements were plotted relative to body surface area with standard measures of quality control. We excluded measures +/- 3 standard deviations. Manual annotation of 50 randomly selected images spanning the cardiac cycle was performed by a clinician (JRP) blinded to automated measures or anthropometric characteristics. The percent difference between automated and manual measurements ([automated measure - manual measure]/manual measure) was examined for systematic relationships to anthropomorphic predictors (Age in years, body surface area, genetic sex). Body surface area was estimated from height and weight using the Haycock formula[13].

Phenome wide association study

Given that IVS.csad is not a standard clinical measurement, using previously described methods[14] we performed a phenome-wide association studies (PheWAS) to highlight clinical associations with 67 cardiovascular phenotypes aggregated from ICD-10 codes as phecodes[15]. For the reporting of PheWAS results, we excluded phenotypes with less than 50 individuals (for continuous traits) or less than 50 cases (for binary-coded traits), and controled the positive false discovery rate (pFDR). We performed PheWAS for the IVS.csad amongst individuals with MRI data (n=32,219).

GWAS, Burden testing, and Annotation

The UK Biobank data release available at the time of analysis included genotypes for 488,377 participants, obtained through either the custom UK Biobank Axiom array or the Affymetrix

Axiom Array. Genotypes were imputed to the TOPMed panel (Freeze5) at the Michigan imputation server. Only variants with minor allele frequency (MAF) greater than or equal to 0.01, and minor allele count (MAC) greater or equal to 5, and variants which have Hardy-Weinberg equilibrium exact test p-value greater than 1e-20 in the entire MRI dataset and an empirical-theoretical variance ratio (MaCH r^2) threshold above 0.3 were included. The main GWAS was conducted on the largest subset of participants with MRI data from the largest unrelated European-ancestry cohort defined using the variable in.white.British.ancestry.subset in the file ukb_sqc_v2.txt provided as part of the UKB data release (n = ~27,100 individuals with estimates derived from imaging data, age = 55.0±7.4). To replicate the findings, we separated the dataset into a discovery set of 22,124 participants and a replication set of 4,899 participants released at a later date. To further explore the findings from the European GWAS, we included three independent sets of other ethnic backgrounds with cardiac MRI who were not included in the discovery set, including African/Afro-Caribbean (AF, n = 222, age = 49.6±7.0), East Asian (EAS, n = 85, age = 49.2±5.5), and South Asian (SAS, n = 368, age = 52.1±7.9). Examination of those samples according the genetic principal components showed that many were mostly of non-European ancestry and were unrelated [Table S1]. Testing of single nucleotide variants and indels was performed using an using linear regression PLINK[16]2 (v2.00a2LM) additive model[17,18], including gender, age, BSA, and genetic principal components 1-4. For genic and regional CNV burden tests, associations were performed using linear regression PLINK2 (v2.00a2LM) additive model[17,18], including gender, age, BSA, Principal components 1-10 and the length and total number of CNV per individual as covariates as previously described[19]. Locuszoom or custom python scripts were used to generate regional association plots[20]. Trans-

ethnic meta-analysis was performed using the METAL software using the standard error analysis scheme.

Independent SNPs were identified by linkage disequilibrium (LD) r2 0.6 and were annotated to cardiac eQTL (GTEx v8: heart atrial appendage and heart left ventricle), CADD, RDB, and the GWAS catalog. GWAS of rare variant and gene-based tests were performed using score and SKATO as implemented in the RVTEST package (link). For the RV-GWAS, age, sex, BSA, Principal components 1-10 were used as covariates to calculate the association of 6 million (6,593,945) imputed/genotyped variants of IVS with minor allele frequency between 0.01 and 0.0005, minor allele count greater than 15, an imputation quality MachR2 greater than 0.8, and p-value for Hardy Weinberg Equilibrium greater than 1e-20. For gene-based tests, the selected variants were mapped to the UCSC GRCh38 refGene annotations, where every unique protein-coding gene (n = 22,933) listed was tested.

In silico analyses for contact prediction and enhancer activity.

To predict changes in 3D chromatin folding brought about by the chr17 inversion, we used Akita[21], a convolutional neural network model trained to predict Hi-C maps from about one megabase of DNA. We first validated that Akita can reproduce wildtype contact frequencies in the region of [GRCh38/hg38] chr17:45392804-46440468. We then used the inversion sequence as input to generate a mutant Hi-C prediction. Comparing this to the wildtype, we predicted changes in contact frequency between promoters, gene bodies, and candidate enhancer elements in the GeneHancer database[22]. To determine possible regulatory functions for single variants in GWAS loci, we examined all variants with P < 10-6. We intersected these variants with

188 predicted enhancers in 131 cell types identified by the activity-by-contact (ABC) Model, which
189 combines measurements of enhancer activity (based on ATAC-seq, DNase-seq, and H3K27ac
190 ChIP-seq) with estimates of enhancer-promoter 3D contact frequencies (based on Hi-C)[23,24]. We
191 examined DNase-seq and ATAC-seq data across a range of cardiovascular cell types from the
192 ENCODE Project. Finally, we examined sequence motif predictions for identified variants using
193 a database of transcription factor binding site motifs[25].
194
195 **LDSC and genetic correlation analyses**
196 To calculate genetic correlation between polygenic risk score of IVS cross sectional area annular
197 size and other related phenotypes, we obtained summary statistics for cardiac MRI-derived LV
198 measurements (left ventricular end-diastolic volume (LVEDV), left ventricular end-systolic
199 volume (LVESV), stroke volume (SV), the body-surface-area (BSA) indexed versions for
200 cardiovascular traits (LVEDVi, LVESVi, and SVi), and left ventricular ejection fraction
201 (LVEF)[26], atrial fibrillation (AF)[27], nonischemic cardiomyopathy (NICM)[28], heart failure[29], heart
202 failure using UK Biobank data[28], hypertension[30], PR Interval[31], Myocardial Infarction (MI) and
203 coronary artery disease (CAD)[32], heart rate[33], and IVS cross sectional area prolapse (MVP)[34].
204 Using these data we performed LD Score regression[35] based on the reformatted summary
205 statistics filtered to HapMap3.
206
207 **Mendelian Randomization**
208 We performed two-sample Mendelian randomization (MR) to test for causal relationships
209 between IVS.csad and risk of two structural diseases of the IVS, specifically VSD and HCM.
210 Summary statistics from a VSD GWAS that considered 191 individuals with VSD and 5,159

controls were retrieved.[26] Summary statistics from a multi-ancestry HCM GWAS that evaluated 2,780 HCM cases and 47,486 controls were retrieved[36]. An instrumental variable (IV), consisting of 22 variants was generated by first selecting variants with p-value < 5e-06 from a standardized trans-ethnic meta-analysis of IVS.csad and then, applying a linkage disequilibrium clumping procedure (r^2 of 0.01 across a window of 1000 kb) via TwoSampleMR (https://github.com/MRCIEU/TwoSampleMR/). The F statistic for this 22 SNP instrument was 15.1 and accounts for 1.04% SNP heritability.

We performed two-sample MR using four different methods, specifically inverse-variance weighted (IVW), weighted median, MR-Egger, and MR-PRESSO (mendelian randomisation pleiotropy residual sum and outlier) using MR-Base[37,38]. Each method assumes a different set of underlying assumptions; all variants included in the instrument are assumed to be valid in the IVW method, and a fixed effects IVW model assumes each variant confers the same mean effect, with no horizontal pleiotropy. A random effects IVW model, accounts for the possibility that variants included in the instrument yield different mean effects and can provide an unbiased estimate when horizontal pleiotropy is balanced[39]. The MR-Egger regression estimates the relationship of effects across instruments functioning as a sensitivity analysis for directional pleiotropy[40], weighted median offers a consistent estimate of effect size when a minimum of 50% of the weights come from valid IVs, and the MR-PRESSO model measures and robustly accounts for the presence of horizontal pleiotropy[38].

RESULTS

After exclusion of outliers we obtained automated estimates of the cross-sectional area of the interventricular septum during ventricular diastole (IVS.csad) for 31,587 individuals yielding a mean IVS.csad of 651 mm^2 (range 207 – 1108 mm^2, standard deviation 166 mm^2) [Fig.1A]. In comparison to blinded manual measurements, automated measures were on average 24 mm^2 or 3.2% smaller than manual measures of IVS.csad with no systematic relationship of measurement error to body surface area (BSA), genetic sex, or age across 50 randomly selected test images [Table S2].

Given that IVS.csad is not a standard clinical measure in MRI or echocardiography and is infrequently characterized in the literature[27] we sought to provide clinical context. Similar to the size of other cardiac structures, IVS.csad scales with BSA in a linear fashion [Fig.1B]. Amongst the 31,587 individuals with an automated estimate, IVS.csad was also correlated with genetic sex and body mass index (BMI) (Pearson's $r > 0.6$), moderately correlated with systolic and diastolic blood pressure (Pearson's $r > 0.3$), and mildly correlated with birthweight (Pearson's $r > 0.1$) [Fig.1C]. Additionally, in a PheWAS of IVS.csad for 67 cardiovascular phenotypes we observed that a larger IVS.csad was positively associated with PheCODES encompassing "abnormal heart sounds" ($p_{fdr} = 0.003$, Odds Ratio (OR) 2.1 per standard deviation (SD) increase in IVS.csad, ICD10 codes R00 - R01.3), "non-rheumatic aortic valve disease" ($p_{fdr} = 0.003$, OR 2.4 per SD IVS.csad, ICD10 I08.2, I35, I35.9), and "abnormal functional study of the cardiovascular system" ($p_{fdr} = 0.02$, OR 2.6 per SD IVS.csad, ICD10 R94.3[Fig.1D]. Importantly, in a random subset of 50 individuals selected for manual measurement, IVS.csad was correlated with

maximal interventricular septal thickness (Pearson's r = 0.62), a standard clinical measure of IVS structure[28].

Given that IVS.csad was strongly correlated with key clinical measures of cardiovascular structure and function, we performed a GWAS of IVS.csad as a proxy for interventricular septal mass in 26,844 individuals of European ancestry divided into discovery (n = 21,945) and replication (n = 4,899) subsets [Table S1]. Using the summary statistics, we performed LD-score correlation (LDSC) to relate the genetic determinants of IVS.csad to a set of common cardiovascular traits and diseases to provide further clinical context, which highlighted significant genetic overlap with left ventricular mass (correlation 0.81, $p = 1.6 \times 10^{-24}$) and the PR interval (correlation 0.26, $p = 6.2 \times 10^{-06}$) but did not show strong negative genetic correlations with cardiovascular traits selected for measurement [Fig.S1]. By LDSC, the total observed h2 for IVS.csad was high at 0.068 (SD 0.0106).

Within European ancestry individuals there were two loci which were strongly significant [Table 1, Fig.2A]. A lead variant on chromosome 6 (rs2376620, $p_{combined} = 2.00 \times 10^{-10}$, *CDKN1A* intron) which tags a haplotype of linked variants overlapping key open chromatin regions exlusive to ventricular myocardium within the introns of *CDKN1A* [Fig.2C], a canonical regulator of ventricular cardiomyocyte proliferation[41,42]. The other lead variant on chromosome 17 (rs62063281, $p_{combined} = 1.31 \times 10^{-13}$) appeared to display strong linkage across a large region of 690,200 base pairs in length which represents a duplication flanked inversion (hg38 chr17:45571611-46261810) across three genes and is common (minor allele frequency 0.18) within in European populations. Using the Akita model of 3D genome folding[21], we quantified

expected changes in chromatin contact frequency caused by the inversion which predicts a severe disturbance in the local chromatin landscape, including loss of contact between the *KANSL1* promoter and regions including the *MAPT* gene body, *KANSL1-AS1*, and several candidate regulatory enhancers predicted by the GeneHancer database [Fig.2B]. A burden-test of structural variants overlapping genes did not reveal further structural variants[19] [Fig. S2].

In the European-only GWAS analysis, a number of additional loci did not meet standard genome-wide significance but had strong pre-existing evidence for involvement in cardiovascular biology [Table S4] . Following the European-only analysis, we performed analyses within three additional ethnic strata (South Asian, East Asian, and African/Afro-Caribbean) and a trans-ethnic meta-analysis which confirmed the findings in *CDKN1A* across population strata and identified additional variants on chromosome 3 (rs62253176, $p=2.3 \times 10^{-08}$, *MITF* intron) and chromosome 1 (rs2092867, $p=4.4 \times 10^{-08}$, *NFIA* intron) which were strongly suggestive but were not meet standard significance threshold in the European only analysis [Figs.S3 & S4]. Additional variants of interest on chromosome 2 (rs1368960, $p_{combined} = 1.3 \times 10^{-07}$, *ASB1* intron) and chromosome 15 (rs11633294, $p_{combined} = 2.7 \times 10^{-07}$, *IGF1R* intron) displayed clear genomic consequences [Fig.S5] but were not genome-wide significant in European or Trans-ethnic meta-analyses. Additionally we performed both rare variant analyses (MAF 0.01 to 0.001) and gene-burden testing which were unrevealing of genetic associations meeting standard pre-specified significance thresholds [Figs.S6 & S7].

Common genetic variation is known to underpin susceptibility to a variety of cardiovascular diseases, including those determined by extremes in anatomical size.[29,30,31] We sought to relate

the genetic determinants governing normal variation in the IVS to two rare forms of structural heart disease manifesting as either a small or insufficient interventricular septum (VSD) or as a large interventricular septum (HCM). Using an instrumental variable of 22 variants derived from the IVS.csad GWAS [Table S4], we performed two-sample Mendelian Randomization for VSD (191 VSD cases and 5,159 controls) and HCM (2,780 HCM cases and 47,486 controls) as outcome phenotypes. We observed evidence suggesting a causal effect where inheritance of a smaller IVS.csad corresponded to an increase in risk for VSD (beta coefficient: 2.06 log odds ratio (OR) VSD per decrease in SD of IVS.csad, standard error (SE) 0.75, $p_{IVW} = 0.006$) while inheritance of a larger IVS.csad corresponded to an increase in risk for HCM (beta coefficient: 2.45 log OR HCM per increase in SD of IVS.csad, SE: 0.48, $p_{IVW} = 2.8e\text{-}7$) [Figure 3].

When evaluating IVS.csad as the exposure and VSD as the outcome, sensitivity analyses suggested the observed causal effect was uniform across the variants included in the instrumental variable (Q value = 15.8, df=20, p-value=0.73). The MR-Egger test of horizontal pleiotropy was insignificant with broad standard errors around the slope and intercept (beta=-0.62, SE=2.43; intercept=-0.06). The median weighted method, which provides a causal estimate assuming at least 50% of the weight comes from valid instrumental variables, was consistent with the IVW result (beta=-2.40; se=1.05; p=0.02). The MR-PRESSO global test for outliers indicates there was no evidence of horizontal pleiotropy for the outcome of VSD (p=0.78).

Sensitivity analyses evaluating IVS.csad as the exposure and HCM as the outcome suggested there was evidence of heterogeneity within the instrument (Q value = 84.2, df=20, p-value=7.58E-10, i2 = 0.76) and the MR-PRESSO global test for outliers indicated there was

horizontal pleiotropy (p = 0.001). The unadjusted MR-PRESSO result (beta coefficient: 2.45, SE: 0.48, $p_{MR\text{-}PRESSO}$ = 4.36E-05) is similar is magnitude to the IVW result, but considering the observed horizontal pleiotropy could be artificially inflated. Performing the outlier-corrected MR-PRESSO test (beta coefficient: 1.62 log OR HCM per increase in SD of IVS.csad, SE: 0.42, $p_{MR\text{-}PRESSO}$ = 0.001) shows an effect size comparable to the median (weighted) results (1.47 log OR HCM per increase in SD of IVS.csad, SE 0.47, $p_{median\text{-}weighted}$ =0.002). The estimate of effect size derived from the MR-Egger test is internally consistent (beta=1.82 log OR HCM per increase in SD of IVS.csad, SE=1.67, p=0.29, intercept=-0.03).

Because left-ventricular mass and IVS thickness are known to increase in the setting of chronic hypertension[10], a phenomenon that we observed in our clinical correlates of automated measures of IVS.csad [Fig.1B], we performed an additional sensitivity analysis to rule out confounding for the causal effect observed between IVS.csad and HCM. For the outcome of HCM we repeated the two-sample MR for IVS.csad conditioned upon diastolic blood pressure, which did not significantly change the positive and statistically significant causal estimate for IVS.csad (beta 3.08 log OR HCM per increase in SD of IVS.csad, SE 0.48, p_{IVW} = 1.05e-05) [Table S5].

DISCUSSION

Here we report upon the genetic architecture of the cross-section of the interventricular septum derived from a standard view in cardiac imaging. The measured cross-sectional area of the IVS in diastole is correlated with genetic sex and body surface area along with a clinically accepted measure of the interventricular septal thickness[43] and genetically correlated with left ventricular mass suggesting that our 2-dimensional automated measure of IVS.csad is a good proxy for the mass of this portion of cardiac anatomy. Genetic associations for IVS.csad are centered upon genetic loci with previously established roles in cardiac development and mendelian forms of cardiac malformations. Additionally, we show a causal relationship between inheritance of a larger interventricular septum with risk of HCM and a smaller interventricular septum with risk of ventricular septal defects.

Haploinsufficiency or truncating mutations in *KANSL1* are causal for Koolen-de Vries syndrome which includes ventricular septal defects amongst a variety of other phenotypes[44,45]. We found that a variant associated with IVS.csad, rs62063281, tags a common inversion on chromosome 17, referred to as CPX_17_4670 on the gnomAD CNV browser[46], that encompasses *KANSL1* and *KANSL1-AS1* along with four other protein coding genes, and was associated with an increase in IVS.csad (beta= 8.60 mm^2, p=1.31e-13). The variant rs62063281 which tags the inversion appears to be a strong tissue-specific splice QTL (Normalized expression -1.5, p = 1.7e-88) for *KANSL1*, In addition to disruption of the interaction of the *KANSL1* promoter and predicted regulatory elements, genes that contact *KANSL1-AS1* (including *ARL17B, LRRC37A, NSFP1, LRRC37A2, NSF*) in the reference configuration are also expected to lose this interaction as a result of the common inversion, which potentially may additionally result in disrupted

361 expression of these other genes. As a group histone, modifying genes including the WDR5-
362 MLL1 complex of which KANSL1 is a component are well-recognized to be involved in cardiac
363 development and diverse forms of congenital heart disease[5,47]. Together these data suggest that
364 common alterations to dosage and transcription of *KANSL1* impact the development and growth
365 of the interventricular septum.
366
367 The *CDKN1A* locus has been specifically identified as a target of HIF-1a within the developing
368 ventricular septum[48] and has been experimentally described as a key mediator of the
369 hypertrophic response within postnatal life[41,49]. The variant rs2376620 sits within an intron of
370 *CDKN1A,* a locus well known to play a role in cardiomyocyte growth[42] recently identified in
371 heart failure[29]. Together these data suggest a specific role for *CDKN1A* in normal development
372 of the interventricular septum along with physiological function throughout postnatal life.
373 Among the other identified in the trans-ethnic analysis, the variant rs62253176 occurs within an
374 intron of the transcription factor *MITF*, an important mediator of beta-adrenergic induced
375 hypertrophy via the renin-angiotensin system expressed in cardiomyocytes[50]. Importantly, *MITF*
376 binds the promotor of canonical cardiac development transcription factor *GATA4* and
377 knockdowns in human embryonic stem cell derived cardiomyocytes specifically reduce the
378 expression of genes in the sarcomere[51,52]. The role of *NFIA* in cardiac development and function
379 is less well defined though it is associated with a variety of electrocardiographic traits[53,54] and the
380 variant rs2092867 occurs on a haplotype associated with a recently reported novel trait related to
381 the presence of cardiac trabeculae derived from the same UK Biobank imaging dataset[55].
382

383 It is well established that rare causal variants confer large effects and produce extreme
384 phenotypes. However, many individuals possessing an extreme phenotype do not yield such
385 causal variants when subject to clinical genetic testing[56]. Instead, there is increasing evidence
386 that such extreme phenotypes may also be attributable to the aggregate burden of common
387 variants associated with the disease or a causal trait[57]. As such, extremes in polygenic risk may
388 explain variability in cardiovascular anatomy and represent an important disease risk factor for a
389 variety of different diseases. For instance, we have previously shown that polygenic risk scores
390 influence ascending aorta size (PRS_{Aorta}); increases in PRS_{Aorta} are associated with risk of
391 developing thoracic aortic aneurysm and dissection[58], and decreases in PRS_{Aorta} are associated
392 with left ventricular outflow tract congenital heart disease[59].
393
394 Using two-sample Mendelian randomization, we show that inheritance of a smaller IVS.csad
395 results in an increase in ventricular septal defect risk, and conversely, inheritance of a larger
396 IVS.csad confers increased risk in HCM. Identifying risk of ventricular septal defects is
397 increased through the inheritance of a smaller interventricular septum, as measured by IVS.csad,
398 supports the need for systematic inquiry into the relationship between normal variation in the
399 size and shape of cardiovascular anatomy and congenital heart disease, for which the majority of
400 genetic risk remains unexplained. For HCM, rare pathogenic variants within core sarcomere
401 genes account for 40% of HCM cases, but with unexplained phenotypic heterogeneity. Recent
402 data suggest common genetic variants also underpin HCM risk[36]. The analyses presented here
403 suggest that a proportion of HCM risk is attributable to normal variation in the size and mass of
404 the interventricular septum as measured by IVS.csad [Fig.S2].
405

406 There are limitations to each aspect of our study. While we have created an automated measure
407 of the interventricular septum which is simple in conception and shown correlation with a variety
408 of traits, the measure lacks phenotypic context provided by many years of use in standard clinical
409 imaging. The cross-sectional area of the interventricular septum is a reductive two-dimensional
410 measure of a complex three-dimensional structure which changes dynamically over the course of
411 the cardiac cycle[60]. More complex measures of the septum are likely to provide nuanced
412 correlations with clinical phenotypes, patient outcomes, as well as substrate for discovery of
413 novel genetic determinants of structure and function[55]. Our analyses were limited by the small
414 proportion of individuals of African/Afro-Caribbean, South and East Asian descent, which
415 reduced the power to detect genetic signal within these groups for this trait and underscores the
416 urgent need to diversify genetic studies of all human diseases and traits[53]. While the inclusion of
417 structural variation in our association studies is novel and our finding is supported by the
418 association of *KANSL1* with Mendelian diasease and mouse models of cardiac development,
419 there are no cohorts available for external confirmation which include both CNV calls and
420 readily available bespoke measures of cardiac anatomy. Although our Mendelian randomization
421 analyses yielded clear and consistent results relating the heritable architecture of septal
422 architecture to rare cardiovascular diseases, they are limited in scope by the relatively
423 underpowered outcome GWAS for VSD and the clinical utility of these findings require further
424 contextualization in larger cohorts.
425

Overall, the data and analyses presented here illustrate the power of combining genetics with phenotyping of cardiac imaging accelerated by machine learning to identify new loci related to cardiovascular development and pathology. The study of cardiovascular anatomy measured as a continuous trait offers the potential to disentangle complex risk factors and reveal previously unrecognized heritability for rare forms of structural heart disease. While the risk is primarily thought to be inherited, the genetic architecture of congenital heart disease has been difficult to establish. In addition other evidence[59], the relationship between a small IVS to increased risk of VSD may suggest a that a proportion of genetic risk for other forms of congenital heart disease can be localized to heritable extremes in size of cardiovascular anatomy which arise primarily from common genetic variation.

ACKNOWLEDGEMENTS

The authors wish to thank Dan Bernstein, James Pirruccello, and Euan Ashley for helpful commentary on the manuscript.

FUNDING

NIH NHGRI R35HG011324 (to JME), Gordon and Betty Moore and BASE Initiative at the Lucile Packard Children's Hospital at Stanford University (to JME), American Heart Association / Children's Heart Foundation Congenital Heart Defect Research Award (to DA), Stanford Maternal and Child Health Research Institute (to DA). JRP has been supported by the Stanford University Department of Pediatrics, National Heart Lung and Blood Institute (R00 HL130523), and Chan-Zuckerberg Biohub.

REFERENCES

1. Koshiba-Takeuchi K, Mori AD, Kaynak BL, Cebra-Thomas J, Sukonnik T, Georges RO, Latham S, Beck L, Henkelman RM, Black BL, Olson EN, Wade J, Takeuchi JK, Nemer M, Gilbert SF, Bruneau BG. Reptilian heart development and the molecular basis of cardiac chamber evolution. *Nature* [Internet]. 2009 [cited 2020 Oct 3];461:95–98. Available from: /pmc/articles/PMC2753965/?report=abstract

2. Penny DJ, Vick GW. Ventricular septal defect [Internet]. In: The Lancet. Lancet; 2011 [cited 2020 Sep 28]. p. 1103–1112.Available from: https://pubmed.ncbi.nlm.nih.gov/21349577/

3. Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, Naidu SS, Nishimura RA, Ommen SR, Rakowski H, Seidman CE, Towbin JA, Udelson JE, Yancy CW. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: Executive summary: A report of the American College of cardiology foundation/American heart association task force on practice guidelines. Circulation. 2011;124:2761–2796.

4. Canepa M, Fumagalli C, Tini G, Vincent-Tompkins J, Day SM, Ashley EA, Mazzarotto F, Ware JS, Michels M, Jacoby D, Ho CY, Olivotto I. Temporal Trend of Age at Diagnosis in Hypertrophic Cardiomyopathy: An Analysis of the International Sarcomeric Human Cardiomyopathy Registry. *Circ Heart Fail* [Internet]. 2020 [cited 2020 Sep 28];13:e007230. Available from: https://www.ahajournals.org/doi/10.1161/CIRCHEARTFAILURE.120.007230

5. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, Hung W-C, Haider S, Zhang J, Knight J, Bjornson RD, Castaldi C, Tikhonoa IR,

Bilguvar K, Mane SM, Sanders SJ, Mital S, Russell MW, Gaynor JW, Deanfield J, Giardini A, Porter GA, Srivastava D, Lo CW, Shen Y, Watkins WS, Yandell M, Yost HJ, Tristani-Firouzi M, Newburger JW, Roberts AE, Kim R, Zhao H, Kaltman JR, Goldmuntz E, Chung WK, Seidman JG, Gelb BD, Seidman CE, Lifton RP, Brueckner M. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet* [Internet]. 2017 [cited 2019 Apr 29];49:1593–1601. Available from: http://www.nature.com/articles/ng.3970

6. Camenisch TD, Runyan RB, Markwald RR. Molecular Regulation of Cushion Morphogenesis. In: Heart Development and Regeneration. Elsevier Inc.; 2010. p. 363–387.

7. Anderson RH, Webb S, Brown NA, Lamers W, Moorman A. Development of the heart: (2) Septation of the atriums and ventricles. *Heart* [Internet]. 2003 [cited 2020 Sep 21];89:949–958. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1767797/

8. Saleh S, Liakopoulos OJ, Buckberg GD. The septal motor of biventricular function [Internet]. Eur. J. Cardio-thoracic Surg. 2006 [cited 2020 Sep 28];29. Available from: https://pubmed.ncbi.nlm.nih.gov/16564701/

9. Gerling S, Pollinger T, Michel H, Dechant MJ, Melter M, Krutsch W. Z-score values of left ventricular dimensions in adolescent elite male soccer players. *Eur J Pediatr* [Internet]. 2020 [cited 2020 Sep 28];Available from: https://pubmed.ncbi.nlm.nih.gov/32705342/

10. Eliakim-Raz N, Prokupetz A, Gordon B, Shochat T, Grossman A. Interventricular Septum and Posterior Wall Thickness Are Associated With Higher Systolic Blood Pressure. *J Clin*

Hypertens [Internet]. 2016 [cited 2020 Sep 28];18:703–706. Available from: https://pubmed.ncbi.nlm.nih.gov/26607051/

11. Yu M, Tcheandjieu C, Georges A, Xiao K, Tejeda H, Dina C, Le Tourneau T, Fiterau I, Judy R, Tsao N, Amgalan D, Munger CJ, Engreitz JM, Damrauer S, Bouatia-Naji N, Priest JR. Computational estimates of mitral annular diameter in systole and diastole cardiac cycle reveal novel genetic determinants of valve function and disease [Internet]. medRxiv. 2020 [cited 2021 Mar 12];2020.12.02.20242206. Available from: https://doi.org/10.1101/2020.12.02.20242206

12. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, Zemrak F, Fung K, Paiva JM, Carapella V, Kim YJ, Suzuki H, Kainz B, Matthews PM, Petersen SE, Piechnik SK, Neubauer S, Glocker B, Rueckert D. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* [Internet]. 2018 [cited 2019 Jul 18];20:65. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30217194

13. Haycock GB, Schwartz GJ, Wisotsky DH. Geometric method for measuring body surface area: A height-weight formula validated in infants, children, and adults. *J Pediatr* [Internet]. 1978 [cited 2021 Mar 12];93:62–66. Available from: https://pubmed.ncbi.nlm.nih.gov/650346/

14. Cordova-Palomera A, Tcheandjieu C, Fries J, Varma P, Chen V, Fiterau M, Xiao K, Tejeda H, Keavney B, Cordell H, Tanigawa Y, Venkataraman G, Rivas M, Re C, Ashley E, Priest J. Cardiac imaging of aortic valve area from 26,142 UK Biobank participants reveal novel genetic associations and shared genetic comorbidity with multiple disease phenotypes. *medRxiv* [Internet]. 2020 [cited 2020 Sep 15];2020.04.09.20060012.

Available from: https://doi.org/10.1101/2020.04.09.20060012

15. Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, Zhao J, Carroll R, Bastarache L, Denny JC, Theodoratou E, Wei WQ. Mapping ICD-10 and ICD-10-CM codes to phecodes: Workflow development and initial evaluation. *J Med Internet Res* [Internet]. 2019 [cited 2020 Sep 22];21. Available from: https://pubmed.ncbi.nlm.nih.gov/31553307/

16. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* [Internet]. 2010 [cited 2019 Sep 18];26:2190–1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20616382

17. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* [Internet]. 2007;81:559–75. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929707613524

18. PLINK 2.0 [Internet]. [cited 2019 Sep 18];Available from: https://www.cog-genomics.org/plink/2.0/

19. Aguirre M, Rivas MA, Priest J. Phenome-wide Burden of Copy-Number Variation in the UK Biobank. *Am J Hum Genet* [Internet]. 2019 [cited 2019 Sep 9];105:373–383. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31353025

20. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ, Frishman D. LocusZoom: Regional visualization of genome-wide association scan results [Internet]. In: Bioinformatics. Oxford University Press; 2011 [cited 2021 Jan 20]. p. 2336–2337.Available from: https://pubmed.ncbi.nlm.nih.gov/20634204/

21. Fudenberg G, Kelley DR, Pollard KS. Predicting 3D genome folding from DNA sequence

with Akita. *Nat Methods* [Internet]. 2020 [cited 2021 Mar 22];17:1111–1117. Available from: https://pubmed.ncbi.nlm.nih.gov/33046897/

22. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, Lancet D, Cohen D. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* [Internet]. 2017 [cited 2021 Mar 22];2017. Available from: https://pubmed.ncbi.nlm.nih.gov/28605766/

23. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, Nguyen TH, Kane M, Perez EM, Durand NC, Lareau CA, Stamenova EK, Aiden EL, Lander ES, Engreitz JM. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat. Genet. 2019;51.

24. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Natri HM, Weeks EM, Munson G, Kane M, Kang HY, Cui A, Ray JP, Eisenhaure TM, Mualim K, Collins RL, Dey K, Price AL, Epstein CB, Kundaje A, Xavier RJ, Daly MJ, Huang H, Finucane HK, Hacohen N, Lander ES, Engreitz JM. Genome-wide maps of enhancer regulation connect risk variants to disease genes. *bioRxiv* [Internet]. 2020 [cited 2020 Dec 1];2020.09.01.278093. Available from: https://doi.org/10.1101/2020.09.01.278093

25. Kulakovskiy I V., Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, Medvedeva YA, Magana-Mora A, Bajic VB, Papatsenko DA, Kolpakov FA, Makeev VJ. HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res*. 2018;46.

26. Pirruccello JP, Bick A, Wang M, Chaffin M, Friedman S, Yao J, Guo X, Venkatesh BA,

Taylor KD, Post WS, Rich S, Lima JAC, Rotter JI, Philippakis A, Lubitz SA, Ellinor PT, Khera A V., Kathiresan S, Aragam KG. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat Commun* [Internet]. 2020 [cited 2020 Sep 14];11. Available from: https://pubmed.ncbi.nlm.nih.gov/32382064/

27. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, Almgren P, Alonso A, Anderson CD, Aragam KG, Arking DE, Barnard J, Bartz TM, Benjamin EJ, Bihlmeyer NA, Bis JC, Bloom HL, Boerwinkle E, Bottinger EB, Brody JA, Calkins H, Campbell A, Cappola TP, Carlquist J, Chasman DI, Chen LY, Chen YDI, Choi EK, Choi SH, Christophersen IE, Chung MK, Cole JW, Conen D, Cook J, Crijns HJ, Cutler MJ, Damrauer SM, Daniels BR, Darbar D, Delgado G, Denny JC, Dichgans M, Dörr M, Dudink EA, Dudley SC, Esa N, Esko T, Eskola M, Fatkin D, Felix SB, Ford I, Franco OH, Geelhoed B, Grewal RP, Gudnason V, Guo X, Gupta N, Gustafsson S, Gutmann R, Hamsten A, Harris TB, Hayward C, Heckbert SR, Hernesniemi J, Hocking LJ, Hofman A, Horimoto ARVR, Huang J, Huang PL, Huffman J, Ingelsson E, Ipek EG, Ito K, Jimenez-Conde J, Johnson R, Jukema JW, Kääb S, Kähönen M, Kamatani Y, Kane JP, Kastrati A, Kathiresan S, Katschnig-Winter P, Kavousi M, Kessler T, Kietselaer BL, Kirchhof P, Kleber ME, Knight S, Krieger JE, Kubo M, Launer LJ, Laurikka J, Lehtimäki T, Leineweber K, Lemaitre RN, Li M, Lim HE, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet* [Internet]. 2018 [cited 2020 Sep 14];50:1225–1233. Available from: https://pubmed.ncbi.nlm.nih.gov/29892015/

28. Aragam KG, Chaffin M, Levinson RT, McDermott G, Choi SH, Shoemaker MB, Haas ME, Weng LC, Lindsay ME, Smith JG, Newton-Cheh C, Roden DM, London B, Wells

QS, Ellinor PT, Kathiresan S, Lubitz SA. Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. *Circulation* [Internet]. 2019 [cited 2020 Sep 14];139:489–501. Available from: https://pubmed.ncbi.nlm.nih.gov/30586722/

29. Shah S, Henry A, Roselli C, Lin H, Sveinbjörnsson G, Fatemifar G, Hedman ÅK, Wilk JB, Morley MP, Chaffin MD, Helgadottir A, Verweij N, Dehghan A, Almgren P, Andersson C, Aragam KG, Ärnlöv J, Backman JD, Biggs ML, Bloom HL, Brandimarto J, Brown MR, Buckbinder L, Carey DJ, Chasman DI, Chen X, Chen X, Chung J, Chutkow W, Cook JP, Delgado GE, Denaxas S, Doney AS, Dörr M, Dudley SC, Dunn ME, Engström G, Esko T, Felix SB, Finan C, Ford I, Ghanbari M, Ghasemi S, Giedraitis V, Giulianini F, Gottdiener JS, Gross S, Guðbjartsson DF, Gutmann R, Haggerty CM, van der Harst P, Hyde CL, Ingelsson E, Jukema JW, Kavousi M, Khaw KT, Kleber ME, Køber L, Koekemoer A, Langenberg C, Lind L, Lindgren CM, London B, Lotta LA, Lovering RC, Luan J, Magnusson P, Mahajan A, Margulies KB, März W, Melander O, Mordi IR, Morgan T, Morris AD, Morris AP, Morrison AC, Nagle MW, Nelson CP, Niessner A, Niiranen T, O'Donoghue ML, Owens AT, Palmer CNA, Parry HM, Perola M, Portilla-Fernandez E, Psaty BM, Abecasis G, Backman J, Bai X, Balasubramanian S, Banerjee N, Baras A, Barnard L, Beechert C, Blumenfeld A, Cantor M, Chai Y, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun* [Internet]. 2020 [cited 2020 Sep 14];11. Available from: https://pubmed.ncbi.nlm.nih.gov/31919418/

30. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, Belbin GM, Bien SA, Cheng I, Cullina S, Hodonsky CJ, Hu Y, Huckins LM, Jeff J, Justice AE, Kocarnik JM, Lim U, Lin BM, Lu Y, Nelson

SC, Park SSL, Poisner H, Preuss MH, Richard MA, Schurmann C, Setiawan VW, Sockell A, Vahi K, Verbanck M, Vishnu A, Walker RW, Young KL, Zubair N, Acuña-Alonso V, Ambite JL, Barnes KC, Boerwinkle E, Bottinger EP, Bustamante CD, Caberto C, Canizales-Quinteros S, Conomos MP, Deelman E, Do R, Doheny K, Fernández-Rhodes L, Fornage M, Hailu B, Heiss G, Henn BM, Hindorff LA, Jackson RD, Laurie CA, Laurie CC, Li Y, Lin DY, Moreno-Estrada A, Nadkarni G, Norman PJ, Pooler LC, Reiner AP, Romm J, Sabatti C, Sandoval K, Sheng X, Stahl EA, Stram DO, Thornton TA, Wassel CL, Wilkens LR, Winkler CA, Yoneyama S, Buyske S, Haiman CA, Kooperberg C, Le Marchand L, Loos RJF, Matise TC, North KE, Peters U, Kenny EE, Carlson CS. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* [Internet]. 2019 [cited 2020 Sep 14];570:514–518. Available from: https://pubmed.ncbi.nlm.nih.gov/31217584/

31. Ntalla I, Weng LC, Cartwright JH, Hall AW, Sveinbjornsson G, Tucker NR, Choi SH, Chaffin MD, Roselli C, Barnes MR, Mifsud B, Warren HR, Hayward C, Marten J, Cranley JJ, Concas MP, Gasparini P, Boutin T, Kolcic I, Polasek O, Rudan I, Araujo NM, Lima-Costa MF, Ribeiro ALP, Souza RP, Tarazona-Santos E, Giedraitis V, Ingelsson E, Mahajan A, Morris AP, Del Greco M F, Foco L, Gögele M, Hicks AA, Cook JP, Lind L, Lindgren CM, Sundström J, Nelson CP, Riaz MB, Samani NJ, Sinagra G, Ulivi S, Kähönen M, Mishra PP, Mononen N, Nikus K, Caulfield MJ, Dominiczak A, Padmanabhan S, Montasser ME, O'Connell JR, Ryan K, Shuldiner AR, Aeschbacher S, Conen D, Risch L, Thériault S, Hutri-Kähönen N, Lehtimäki T, Lyytikäinen LP, Raitakari OT, Barnes CLK, Campbell H, Joshi PK, Wilson JF, Isaacs A, Kors JA, van Duijn CM, Huang PL, Gudnason V, Harris TB, Launer LJ, Smith A V., Bottinger EP, Loos RJF,

Nadkarni GN, Preuss MH, Correa A, Mei H, Wilson J, Meitinger T, Müller-Nurasyid M, Peters A, Waldenberger M, Mangino M, Spector TD, Rienstra M, van de Vegte YJ, van der Harst P, Verweij N, Kääb S, Schramm K, Sinner MF, Strauch K, Cutler MJ, Fatkin D, London B, et al. Multi-ancestry GWAS of the electrocardiographic PR interval identifies 202 loci underlying cardiac conduction. *Nat Commun* [Internet]. 2020 [cited 2020 Sep 14];11. Available from: https://pubmed.ncbi.nlm.nih.gov/32439900/

32. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang S-J, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen L-P, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith A V, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han B-G, Huang J, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki M-L, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon F-U-R, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* [Internet]. 2015 [cited 2019 Apr 29];47:1121–1130. Available from: http://www.nature.com/articles/ng.3396

656 33. den Hoed M, Eijgelsheim M, Esko T, Brundel BJJM, Peal DS, Evans DM, Nolte IM,
657 Segrè A V, Holm H, Handsaker RE, Westra H-J, Johnson T, Isaacs A, Yang J, Lundby A,
658 Zhao JH, Kim YJ, Go MJ, Almgren P, Bochud M, Boucher G, Cornelis MC, Gudbjartsson
659 D, Hadley D, van der Harst P, Hayward C, den Heijer M, Igl W, Jackson AU, Kutalik Z,
660 Luan J, Kemp JP, Kristiansson K, Ladenvall C, Lorentzon M, Montasser ME, Njajou OT,
661 O'Reilly PF, Padmanabhan S, St. Pourcain B, Rankinen T, Salo P, Tanaka T, Timpson NJ,
662 Vitart V, Waite L, Wheeler W, Zhang W, Draisma HHM, Feitosa MF, Kerr KF, Lind PA,
663 Mihailov E, Onland-Moret NC, Song C, Weedon MN, Xie W, Yengo L, Absher D, Albert
664 CM, Alonso A, Arking DE, de Bakker PIW, Balkau B, Barlassina C, Benaglio P, Bis JC,
665 Bouatia-Naji N, Brage S, Chanock SJ, Chines PS, Chung M, Darbar D, Dina C, Dörr M,
666 Elliott P, Felix SB, Fischer K, Fuchsberger C, de Geus EJC, Goyette P, Gudnason V,
667 Harris TB, Hartikainen A-L, Havulinna AS, Heckbert SR, Hicks AA, Hofman A,
668 Holewijn S, Hoogstra-Berends F, Hottenga J-J, Jensen MK, Johansson Å, Junttila J, Kääb
669 S, Kanon B, Ketkar S, Khaw K-T, et al. Identification of heart rate–associated loci and
670 their effects on cardiac conduction and rhythm disorders. *Nat Genet* [Internet]. 2013 [cited
671 2018 Jun 20];45:621–631. Available from:
672 http://www.ncbi.nlm.nih.gov/pubmed/23583979
673 34. Toomer KA, Yu M, Fulmer D, Guo L, Moore KS, Moore R, Ka'la DD, Glover J, Peterson
674 N, Ramos-Ortiz S, Drohan A, Catching BJ, Stairley R, Wessels A, Lipschutz JH, Delling
675 FN, Jeunemaitre X, Dina C, Collins RL, Brand H, Talkowski ME, Del Monte F,
676 Mukherjee R, Awgulewitsch A, Body S, Hardiman G, Starr Hazard E, Da Silveira WA,
677 Wang B, Leyne M, Durst R, Markwald RR, Le Scouarnec S, Hagege A, Le Tourneau T,
678 Kohl P, Rog-Zielinska EA, Ellinor PT, Levine RA, Milan DJ, Schott JJ, Bouatia-Naji N,

Slaugenhaupt SA, Norris RA. Primary cilia defects causing mitral valve prolapse. *Sci Transl Med* [Internet]. 2019 [cited 2020 Sep 14];11. Available from: https://pubmed.ncbi.nlm.nih.gov/31118289/

35. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, Lee P, Collier DA, Huang H, Pers TH, Agartz I, Agerbo E, Albus M, Alexander M, Amin F, Bacanu SA, Begemann M, Belliveau RA, Bene J, Bergen SE, Bevilacqua E, Bigdeli TB, Black DW, Bruggeman R, Buccola NG, Buckner RL, Byerley W, Cahn W, Cai G, Cairns MJ, Campion D, Cantor RM, Carr VJ, Carrera N, Catts S V., Chambert KD, Chan RCK, Chen RYL, Chen EYH, Cheng W, Cheung EFC, Chong SA, Cloninger CR, Cohen D, Cohen N, Cormican P, Craddock N, Crespo-Facorro B, Crowley JJ, Curtis D, Davidson M, Davis KL, Degenhardt F, Del Favero J, DeLisi LE, Demontis D, Dikeos D, Dinan T, Djurovic S, Donohoe G, Drapeau E, Duan J, Dudbridge F, Durmishi N, Eichhammer P, Eriksson J, Escott-Price V, Essioux L, Fanous AH, Farrell MS, Frank J, Franke L, Freedman R, Freimer NB, Friedl M, Friedman JI, Fromer M, Genovese G, Georgieva L, Gershon ES, Giegling I, Giusti-Rodríguez P, Godard S, Goldstein JI, Golimbet V, Gopal S, Gratten J, De Haan L, Hammer C, Hamshere ML, Hansen M, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* [Internet]. 2015 [cited 2020 Sep 14];47:291–295. Available from: https://pubmed.ncbi.nlm.nih.gov/25642630/

36. Harper AR, Goel A, Grace C, Thomson KL, Petersen SE, Xu X, Waring A, Ormondroyd E, Kramer CM, Ho CY, Neubauer S, Kolm P, Kwong R, Dolman SF, Desvigne-Nickens P, Dimarco JP, Geller N, Kim DY, Zhang C, Weintraub W, Abraham T, Anderson L,

Appelbaum E, Autore C, Berry C, Biagini E, Bradlow W, Bucciarelli-Ducci C, Chiribiri A, Choudhury L, Crean A, Dawson D, Desai MY, Elstein E, Flett A, Friedrich M, Heitner S, Helms A, Jacoby DL, Kim H, Kim B, Larose E, Mahmod M, Mahrholdt H, Maron M, McCann G, Michels M, Mohiddin S, Nagueh S, Newby D, Olivotto I, Owens A, Pierre-Mongeon F, Prasad S, Rimoldi O, Salerno M, Schulz-Menger J, Sherrid M, Swoboda P, van Rossum A, Weinsaft J, White J, Williamson E, Tadros R, Ware JS, Bezzina CR, Farrall M, Watkins H. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. *Nat Genet* [Internet]. 2021 [cited 2021 Mar 12];53:135–142. Available from: https://pubmed.ncbi.nlm.nih.gov/33495597/

37. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* [Internet]. 2018 [cited 2018 Jun 15];7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29846171

38. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* [Internet]. 2018 [cited 2021 Mar 12];50:693–698. Available from: https://www.nature.com/articles/s41588-018-0099-7

39. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.

40. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* [Internet]. 2017 [cited 2021 Mar 12];32:377–389. Available from: http://link.springer.com/10.1007/s10654-017-0255-x

41. Hauck L, Grothe D, Billia F. p21CIP1/WAF1-dependent inhibition of cardiac hypertrophy in response to Angiotensin II involves Akt/Myc and pRb signaling. *Peptides* [Internet]. 2016 [cited 2021 Jan 20];83:38–48. Available from: https://pubmed.ncbi.nlm.nih.gov/27486069/

42. Garnatz AS, Gao Z, Broman M, Martens S, Earley JU, Svensson EC. FOG-2 mediated recruitment of the NuRD complex regulates cardiomyocyte proliferation during heart development. *Dev Biol* [Internet]. 2014 [cited 2021 Mar 31];395:50–61. Available from: /pmc/articles/PMC4190033/

43. Toshima H, Koga Y, Uemura S, Zinnouchi J, Kimura N, Nakakura S. Echocardiographic Study on Hypertrophic Cardiomyopathy. *Jpn Heart J* [Internet]. 1976 [cited 2021 Mar 12];17:275–289. Available from: https://pubmed.ncbi.nlm.nih.gov/133253/

44. Zollino M, Marangi G, Ponzi E, Orteschi D, Ricciardi S, Lattante S, Murdolo M, Battaglia D, Contaldo I, Mercuri E, Stefanini MC, Caumes R, Edery P, Rossi M, Piccione M, Corsello G, Monica M Della, Scarano F, Priolo M, Gentile M, Zampino G, Vijzelaar R, Abdulrahman O, Rauch A, Oneda B, Deardorff MA, Saitta SC, Falk MJ, Dubbs H, Zackai E. Intragenic KANSL1 mutations and chromosome 17q21.31 deletions: Broadening the clinical spectrum and genotype-phenotype correlations in a large cohort of patients. *J Med Genet* [Internet]. 2015 [cited 2021 Jan 20];52:804–814. Available from: https://pubmed.ncbi.nlm.nih.gov/26424144/

45. Koolen DA, Pfundt R, Linda K, Beunders G, Veenstra-Knol HE, Conta EH, Fortuna AM,

Gillessen-Kaesbach G, Dugan S, Halbach S, Abdul-Rahman OA, Winesett HM, Chung WK, Dalton M, Dimova PS, Mattina T, Prescott K, Zhang HZ, Saal HM, Hehir-Kwa JY, Willemsen MH, Ockeloen CW, Jongmans MC, Van Der Aa N, Failla P, Barone C, Avola E, Brooks AS, Kant SG, Gerkes EH, Firth H V., Unap K, Bird LM, Masser-Frye D, Friedman JR, Sokunbi MA, Dixit A, Splitt M, Kukolich MK, McGaughran J, Coe BP, Flórez J, Nadif Kasr N, Brunner HG, Thompson EM, Gecz J, Romano C, Eichler EE, De Vries BBA. The Koolen-de Vries syndrome: A phenotypic comparison of patients with a 17q21.31 microdeletion versus a KANSL1 sequence variant. *Eur J Hum Genet* [Internet]. 2016 [cited 2021 Jan 20];24:652–659. Available from: https://pubmed.ncbi.nlm.nih.gov/26306646/

46. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, Khera A V., Lowther C, Gauthier LD, Wang H, Watts NA, Solomonson M, O'Donnell-Luria A, Baumann A, Munshi R, Walker M, Whelan CW, Huang Y, Brookings T, Sharpe T, Stone MR, Valkanas E, Fu J, Tiao G, Laricchia KM, Ruano-Rubio V, Stevens C, Gupta N, Cusick C, Margolin L, Alföldi J, Armean IM, Banks E, Bergelson L, Cibulskis K, Collins RL, Connolly KM, Covarrubias M, Cummings B, Daly MJ, Donnelly S, Farjoun Y, Ferriera S, Francioli L, Gabriel S, Gauthier LD, Gentry J, Gupta N, Jeandet T, Kaplan D, Karczewski KJ, Laricchia KM, Llanwarne C, Minikel E V., Munshi R, Neale BM, Novod S, O'Donnell-Luria AH, Petrillo N, Poterba T, Roazen D, Ruano-Rubio V, Saltzman A, Samocha KE, Schleicher M, Seed C, Solomonson M, Soto J, Tiao G, Tibbetts K, Tolonen C, Vittal C, Wade G, Wang A, Wang Q, Ware JS, Watts NA, Weisburd B, Whiffin N, Salinas CAA, Ahmad T, Albert CM, Ardissino D, Atzmon G, Barnard J, Beaugerie L, Benjamin EJ, Boehnke M, Bonnycastle LL, Bottinger EP, Bowden DW, Bown MJ,

Chambers JC, Chan JC, Chasman D, Cho J, Chung MK, Cohen B, et al. A structural variation reference for medical and population genetics. *Nature* [Internet]. 2020 [cited 2021 Jan 20];581:444–451. Available from: https://pubmed.ncbi.nlm.nih.gov/32461652/

47. Kulkarni SS, Khokha MK. WDR5 regulates left-right patterning via chromatin-dependent and-independent functions. *Dev* [Internet]. 2018 [cited 2021 Mar 31];145. Available from: https://dev.biologists.org/content/145/23/dev159889

48. Guimarães-Camboa N, Stowe J, Aneas I, Sakabe N, Cattaneo P, Henderson L, Kilberg MS, Johnson RS, Chen J, McCulloch AD, Nobrega MA, Evans SM, Zambon AC. HIF1α Represses Cell Stress Pathways to Allow Proliferation of Hypoxic Fetal Cardiomyocytes. *Dev Cell* [Internet]. 2015 [cited 2021 Jan 20];33:507–521. Available from: https://pubmed.ncbi.nlm.nih.gov/26028220/

49. Liu ZP, Olson EN. Suppression of proliferation and cardiomyocyte hypertrophy by CHAMP, a cardiac-specific RNA helicase. *Proc Natl Acad Sci U S A* [Internet]. 2002 [cited 2021 Jan 20];99:2043–2048. Available from: https://pubmed.ncbi.nlm.nih.gov/11854500/

50. Tshori S, Gilon D, Beeri R, Nechushtan H, Kaluzhny D, Pikarsky E, Razin E. Transcription factor MITF regulates cardiac growth and hypertrophy. *J Clin Invest* [Internet]. 2006 [cited 2021 Jan 20];116:2673–2681. Available from: https://pubmed.ncbi.nlm.nih.gov/16998588/

51. Fu K, Nakano H, Morselli M, Chen T, Pappoe H, Nakano A, Pellegrini M. A temporal transcriptome and methylome in human embryonic stem cell-derived cardiomyocytes identifies novel regulators of early cardiac development. *Epigenetics*. 2018;13:1013–1026.

52. Mehta G, Kumarasamy S, Wu J, Walsh A, Liu L, Williams K, Joe B, de la Serna IL.

MITF interacts with the SWI/SNF subunit, BRG1, to promote GATA4 expression in cardiac hypertrophy. *J Mol Cell Cardiol*. 2015;88:101–110.

53. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, Belbin GM, Bien SA, Cheng I, Cullina S, Hodonsky CJ, Hu Y, Huckins LM, Jeff J, Justice AE, Kocarnik JM, Lim U, Lin BM, Lu Y, Nelson SC, Park SSL, Poisner H, Preuss MH, Richard MA, Schurmann C, Setiawan VW, Sockell A, Vahi K, Verbanck M, Vishnu A, Walker RW, Young KL, Zubair N, Acuña-Alonso V, Ambite JL, Barnes KC, Boerwinkle E, Bottinger EP, Bustamante CD, Caberto C, Canizales-Quinteros S, Conomos MP, Deelman E, Do R, Doheny K, Fernández-Rhodes L, Fornage M, Hailu B, Heiss G, Henn BM, Hindorff LA, Jackson RD, Laurie CA, Laurie CC, Li Y, Lin DY, Moreno-Estrada A, Nadkarni G, Norman PJ, Pooler LC, Reiner AP, Romm J, Sabatti C, Sandoval K, Sheng X, Stahl EA, Stram DO, Thornton TA, Wassel CL, Wilkens LR, Winkler CA, Yoneyama S, Buyske S, Haiman CA, Kooperberg C, Le Marchand L, Loos RJF, Matise TC, North KE, Peters U, Kenny EE, Carlson CS. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570:514–518.

54. Evans DS, Avery CL, Nalls MA, Li G, Barnard J, Smith EN, Tanaka T, Butler AM, Buxbaum SG, Alonso A, Arking DE, Berenson GS, Bis JC, Buyske S, Carty CL, Chen W, Chung MK, Cummings SR, Deo R, Eaton CB, Fox ER, Heckbert SR, Heiss G, Hindorff LA, Hsueh WC, Isaacs A, Jamshidi Y, Kerr KF, Liu F, Liu Y, Lohman KK, Magnani JW, Maher JF, Mehra R, Meng YA, Musani SK, Newton-Cheh C, North KE, Psaty BM, Redline S, Rotter JI, Schnabel RB, Schork NJ, Shohet R V., Singleton AB, Smith JD, Soliman EZ, Srinivasan SR, Taylor HA, Van Wagoner DR, Wilson JG, Young T, Zhang

ZM, Zonderman AB, Evans MK, Ferrucci L, Murray SS, Tranah GJ, Whitsel EA, Reiner AP, Sotoodehnia N. Fine-mapping, novel loci identification, and SNP association transferability in a genome-wide association study of QRS duration in African Americans. *Hum Mol Genet*. 2016;25:4350–4368.

55. Meyer H V., Dawes TJW, Serrani M, Bai W, Tokarczuk P, Cai J, de Marvao A, Henry A, Lumbers RT, Gierten J, Thumberger T, Wittbrodt J, Ware JS, Rueckert D, Matthews PM, Prasad SK, Costantino ML, Cook SA, Birney E, O'Regan DP. Genetic and functional insights into the fractal structure of the heart. *Nature*. 2020;584:589–594.

56. Thomson KL, Ormondroyd E, Harper AR, Dent T, McGuire K, Baksi J, Blair E, Brennan P, Buchan R, Bueser T, Campbell C, Carr-White G, Cook S, Daniels M, Deevi SVV, Goodship J, Hayesmoore JBG, Henderson A, Lamb T, Prasad S, Rayner-Matthews P, Robert L, Sneddon L, Stark H, Walsh R, Ware JS, Farrall M, Watkins HC. Analysis of 51 proposed hypertrophic cardiomyopathy genes from genome sequencing data in sarcomere negative cases has negligible diagnostic yield. *Genet Med* [Internet]. 2019 [cited 2021 Jan 24];21:1576–1584. Available from: https://pubmed.ncbi.nlm.nih.gov/30531895/

57. Khera A V, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* [Internet]. 2018 [cited 2019 Apr 29];50:1219–1224. Available from: http://www.nature.com/articles/s41588-018-0183-z

58. Tcheandjieu C, Xiao K, Tejeda H, Lynch J, Ruotsalainen S, Bellomo T, Palnati M, Judy R, Kember R, Klarin D, Kember R, Verma S, Center RG, Program VMV, Project F, Palotie A, Daly M, Ritchie M, Rader D, Rivas MA, Assimes T, Tsao P, Damrauer S,

Priest J. High heritability of ascending aortic diameter and multi-ethnic prediction of thoracic aortic disease. *medRxiv* [Internet]. 2020 [cited 2020 Jul 21];14:2020.05.29.20102335. Available from: https://www.medrxiv.org/content/10.1101/2020.05.29.20102335v1

59. Tcheandjieu C, Zanetti D, Yu M, Priest JR. Inherited Extremes of Aortic Diameter Confer Risk for a Specific Class of Congenital Heart Disease [Internet]. Circ. Genomic Precis. Med. 2020 [cited 2021 Jan 20];13:724–727. Available from: https://www.ahajournals.org/doi/10.1161/CIRCGEN.120.003170

60. Bidviene J, Muraru D, Maffessanti F, Ereminiene E, Kovács A, Lakatos B, Vaskelyte J-J, Zaliunas R, Surkova E, Parati G, Badano LP. Regional shape, global function and mechanics in right ventricular volume and pressure overload conditions: a three-dimensional echocardiography study. *Int J Cardiovasc Imaging* [Internet]. 2021 [cited 2021 Jan 20];Available from: http://www.ncbi.nlm.nih.gov/pubmed/33389362

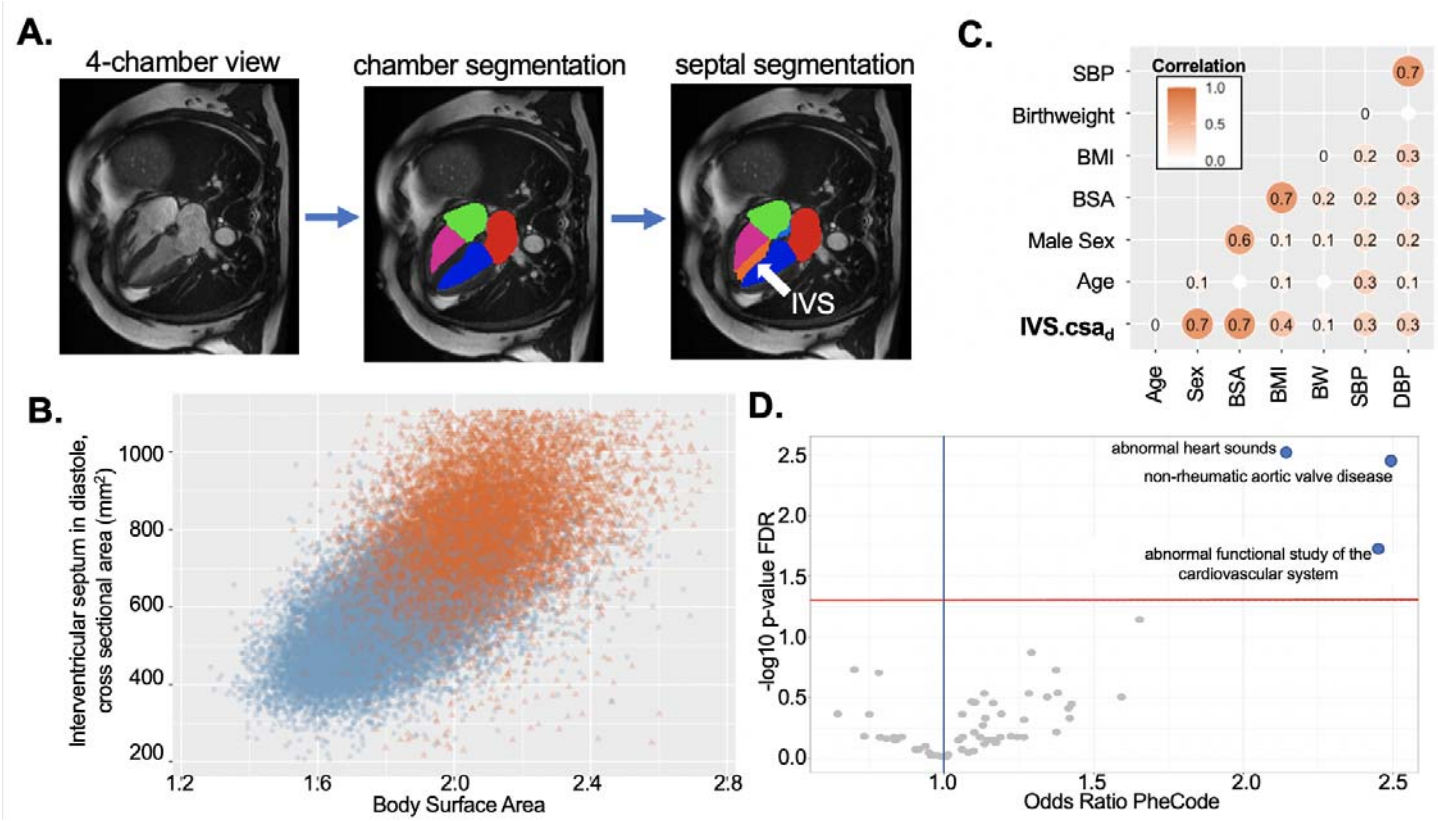

Figure 1. Cross-sectional area in diastole is a new proxy for structure and function of the interventricular septum. A. Schema of cardiovascular measurements. Still frame of a cardiac MRI in the four-chamber view (left panel) demonstrating segmentation of lumen of the cardiac chambers (middle panel) with the left ventricle indicated by the blue region, left atrium indicated by the red region, right atrium indicated by the green region, and right ventricle in magenta. Chamber segmentation is followed by segmentation of the interventricular septum (orange, indicated by white arrow). **B.** The distribution of interventricular septal cross-sectional area in diastole (IVS.csad). Values are plotted in mm^2 relative to body surface area (BSA), orange and blue points represent male and female participants respectively. In a simple linear model including Sex, Age, systolic and diastolic blood pressure and body mass index, BSA explains 44.8% of the observed variation in IVS.csad (284 mm^2 IVS.csad per m^2 BSA, $p < 2e\text{-}16$). **C.** Correlation plot of IVS.csad versus a standard set of clinical measures. IVS.csad is most strongly correlated with genetic sex and BSA, moderately correlated with BMI and blood pressure, and slightly correlated with birthweight. Notably there was no strong correlation with age. All non-zero Pearson's correlations reported are strongly statistically significant ($p < 5e\text{-}08$). **D.** Volcano plot representing the PheWAS of IVS.csad for 67 cardiovascular phenotypes grouped as PheCODES. The X-axis represents the association with the PheCODE per standard deviation change in the IVS.csad, and the absolute value of Y-axis represents the negative logarithm of the p-value adjusted with a standard false discovery rate. Across an unbiased assessment of clinical phenotypes, increased IVS.csad was positively associated with the clinical findings of abnormal heart sounds, non-rheumatic aortic valve disease, and abnormal functional studies of the cardiovascular system.

						Combined Discovery & Replication		Discovery set		Replication Set	
CHR	**POS**	**SNP**	**REF**	**RA**	**RA_FREQ**	**Beta [95%CI]**	**P-value**	**Beta [95%CI]**	**P-value**	**Beta [95%CI]**	**P-value**
17	45961419	rs62063281	A	G	0.22	8.60 [6.33, 10.88]	1.31×10^{-13}	7.72 [5.20, 10.2]	1.89×10^{-09}	12.39 [7.07, 17.71]	5.11×10^{-06}
6	36681816	rs2376620	A	G	0.16	8.38 [5.80, 10.96]	2.00×10^{-10}	7.72 [4.87, 10.5]	1.14×10^{-07}	11.29 [5.24, 17.34]	2.58×10^{-04}

Table 1. Genome wide associations from measurements of IVS.csad in European ancestry individuals.
RA: risk allele, FREQ: frequency

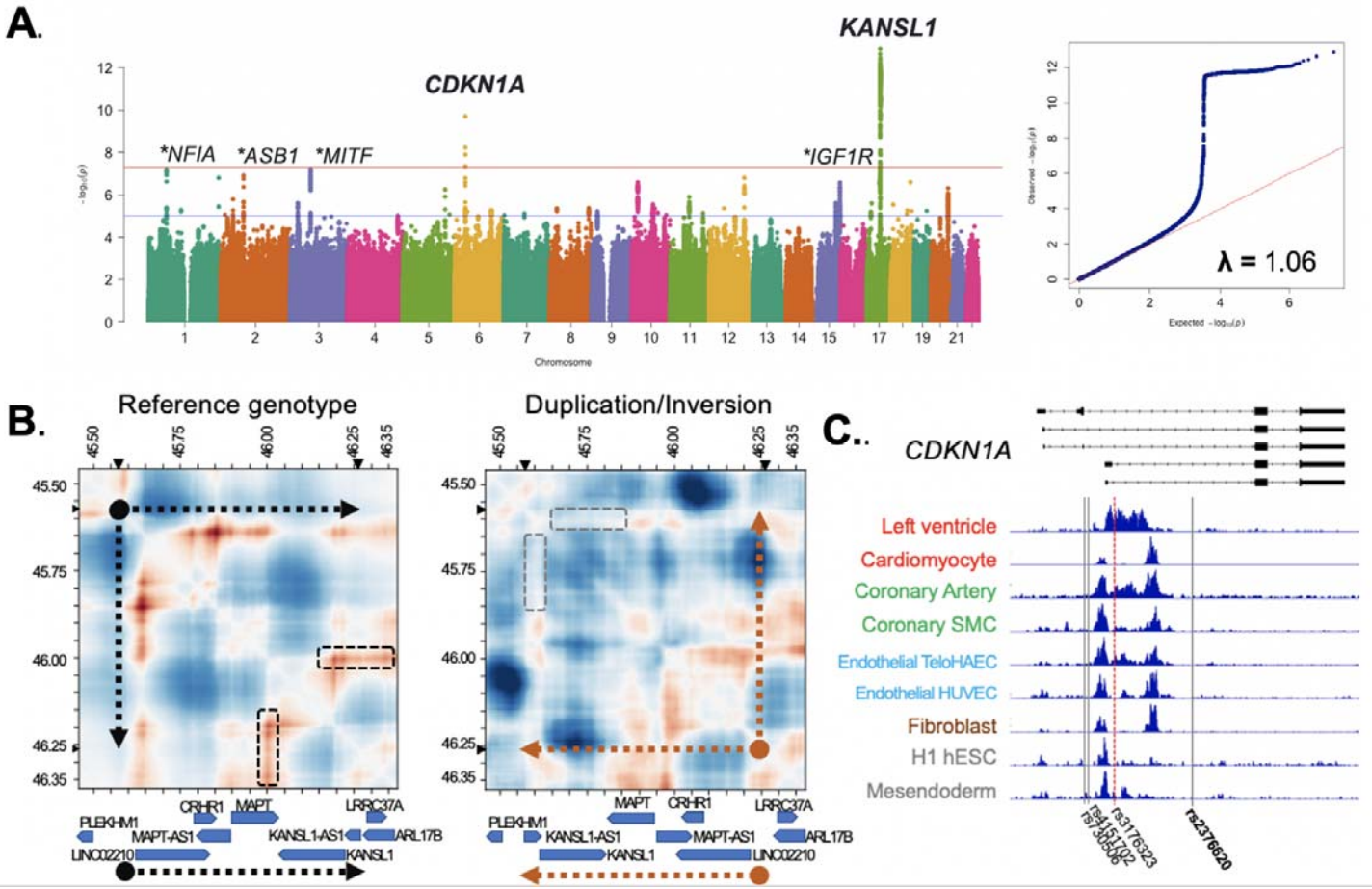

Figure 2. GWAS of IVS.csad identifies two loci meeting standard significance thresholds. A. Manhattan plot with two loci at KANSL1 and CDKN1A, along with loci below 5e-08 marked with a star Quantile-quantile plot indicates an absence of systematic inflation. **B.** Two-dimensional plots of the Akita model of genome folding with GRCh38 coordinates on the X and Y axis with relevant genes and strand orientation indicated by blue boxes. Predicted frequency of interaction between regions of DNA is indicated by color, where red(/blue) shows a higher(/lower) frequency of interaction relative to what is expected given linear genomic distance. The reference genotype predicts a strong chromatin interaction (black boxes) between the *KANSL1* promoter and regions including the *MAPT* gene body and *KANSL1-AS1.* The common duplication/inversion 17:45571611-46261810 (orientation indicated by the orange arrow) is predicted to completely disrupt the interaction (corresponding locations within the inversion indicated by the grey boxes). The lead variant rs62063281 is an eQTL for multiple genes across the locus and a strong splicing QTL for *MAPT* and *KANSL1* (-1.5 normalized expression, p=1.7e-88) within the left ventricle [Table S3]. **C.** Plot of tracks showing output of the activity-by-contact (ABC) which identifies *CDKN1A* as the gene target for the lead variant rs2376620, which is the marker for a haplotype inclusive of three other variants localized in open chromatin enhancer regions present most prominently in the left ventricle. Displayed for contrast are DNase-seq for cardiomyocyte, human umbilical vein endothelial cell (HUVEC), fibroblast of fetal lung, H1 human embryonic stem cells (hESCs), and H1-derived mesendodermal precursors; or ATAC-seq for heart left ventricle (LV), coronary artery, coronary smooth muscle cell (SMC), and telomerase immortalized human aortic endothelial cells (TeloHAEC).

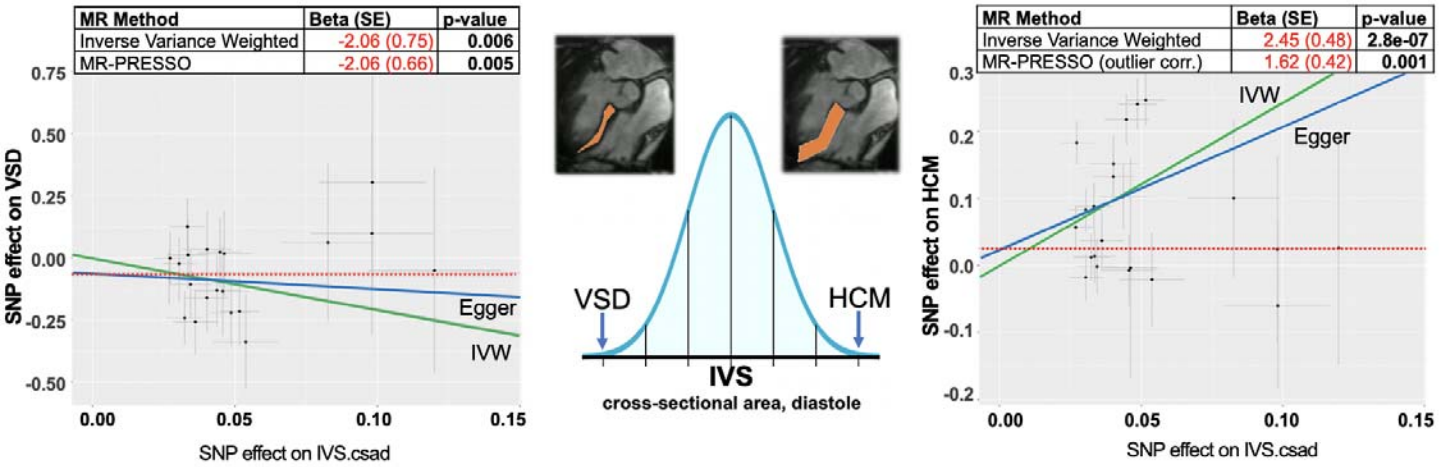

Figure 3. Mendelian Randomization suggests that inheritance of increased IVS.csad is causal for HCM while inheritance of decreased IVS.csad is causal for VSD. Twenty two loci were selected for use as instrumental variables for interventricular septum, cross-sectional area as the exposure for two-sample Mendelian Randomization analysis of ventricular septal defects (VSD) and hypertrophic cardiomyopathy (HCM). Regression lines for the MR-Egger and Inverse Variants Weighted (IVW) methods are displayed in blue and green respectively, with the intercept from the MR-Egger method in dotted red for both graphs. The causal relationships of IVS.csad and risks for VSD and HCM are not driven by a single variant when analyzed individually or in a “leave-one out” sensitivity analysis (Figure S6).

.