Title

Oligogenic Architecture of Rare Noncoding Variants Distinguishes 4 Congenital Heart Disease Phenotypes.

Journal

Circulation: Cardiovascular Genetics, 16(3)

Authors

Yu, Mengyao
Aguirre, Matthew
Jia, Meiwen
et al.

Publication Date

2023-06-01

DOI

10.1161/CIRCGEN.122.003968

Peer reviewed

HHS Public Access

Author manuscript

Circ Genom Precis Med. Author manuscript; available in PMC 2024 June 01.

Published in final edited form as:

Circ Genom Precis Med. 2023 June ; 16(3): 258–266. doi:10.1161/CIRCGEN.122.003968.

An Oligogenic Architecture of Rare Noncoding Variants Distinguishes Four Congenital Heart Disease Phenotypes

Mengyao Yu, PhD[1], **Matthew Aguirre, AB**[1,2], **Meiwen Jia, PhD**[3], **Ketrin Gjoni, BS**[4], **Aldo Cordova-Palomera, PhD**[1], **Chad Munger, BS**[5], **Dulguun Amgalan, PhD**[5], **X Rosa Ma, BS**[5], **Alexandre Pereira, MD, PhD**[6], **Catherine Tcheandjieu, DVM, PhD**[1,4], **Christine Seidman, MD**[6], **Jonathan Seidman, PhD**[6], **Martin Tristani-Firouzi, MD, PhD**[7], **Wendy Chung, MD, PhD**[8], **Elizabeth Goldmuntz, MD**[9], **Deepak Srivastava, MD**[4], **Ruth J.F. Loos, PhD**[10], **Nathalie Chami, PhD**[10], **Heather Cordell, PhD**[11], **Martina Dreßen, PhD**[12], **Bertram Mueller-Myhsok, PhD**[3], **Harald Lahm, PhD**[12], **Markus Krane, MD**[12,13], **Katherine S. Pollard, PhD**[4,14], **Jesse M. Engreitz, PhD**[5,15], **Sarah A. Gagliano Taliun, PhD**[16,17], **Bruce D. Gelb, MD**[18], **James R. Priest, MD**[1]

[1]Dept of Pediatrics, Stanford Univ School of Medicine;

[2]Dept of Biomedical Data Science, Stanford Univ, Stanford CA;

[3]Dept of Translational Research in Psychiatry, Max Planck Institute of Psychiatry Munich, Munich, Germany;

[4]Gladstone Institutes; Univ of California San Francisco, San Francisco CA;

[5]Dept of Genetics, Stanford Univ School of Medicine;

[6]Dept of Genetics, Harvard Univ, Cambridge MA;

[7]Dept of Pediatrics, Univ of Utah, Salt Lake City, UT;

[8]Dept of Pediatrics, Columbia Univ, New York, NY;

[9]Dept of Pediatrics, Univ of Pennsylvania, Philadelphia, PA;

[10]Icahn School of Medicine at Mount Sinai, New York, NY;

[11]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle Univ, International Centre for Life, Central Parkway, Newcastle upon Tyne, United Kingdom;

[12]Dept of Cardiovascular Surgery, Division of Experimental Surgery, Institute Insure (Institute for Translational Cardiac Surgery), German Heart Center Munich & Technical Univ of Munich, School of Medicine & Health, Munich, Germany;

[13]Dept of Cardiac Surgery, Yale School of Medicine, New Haven, CT;

Correspondence: James R. Priest MD, Stanford Univ School of Medicine, 171 Oyster Point Boulevard Suite 500, South San Francisco, CA 94080, Tel: 510-213-2043, jpriest@tenayatherapeutics.com.

Disclosures: Dr. Priest is a full-time employee and shareholder of Tenaya Therapeutics, the other authors have nothing to disclose.

Supplementary Materials:
Supplemental Methods
Supplemental Tables SI–VX
Supplemental Figures SI–XIX
References[44–57]

[14]Chan Zuckerberg Biohub, San Francisco;

[15]Basic Sciences and Engineering (BASE) Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, CA;

[16]Dept of Medicine & Dept of Neurosciences, Faculty of Medicine, Université de Montréal;

[17]Montreal Heart Institute, Montreal, Quebec, Canada;

[18]The Mindich Child Health & Development Institute at the Hess Center for Science & Medicine at Mount Sinai, New York, NY

Abstract

Background: Congenital heart disease (CHD) is highly heritable, but the power to identify inherited risk has been limited to analyses of common variants in small cohorts.

Methods: We performed re-imputation of four CHD cohorts (n=55,342) to the TOPMed reference panel (freeze 5), permitting meta-analysis of 14,784,017 variants including 6,035,962 rare variants of high imputation quality as validated by whole genome sequencing.

Results: Meta-analysis identified 16 novel loci, including 12 rare variants, which displayed moderate or large effect sizes (median odds ratio (OR) 3.02) for four separate CHD categories. Analyses of chromatin structure link 13 of the genome-wide significant loci to key genes in cardiac development; rs373447426 (minor allele frequency (MAF) 0.003, OR 3.37 for Conotruncal heart disease (CTD), p=1.49E-08) is predicted to disrupt chromatin structure for two nearby genes *BDH1* and *DLG1* involved in conotruncal development. A lead variant rs189203952 (MAF 0.01, OR 2.4 for left ventricular outflow tract obstruction, p=1.46e-08) is predicted to disrupt the binding sites of four transcription factors known to participate in cardiac development in the promoter of *SPAG9*. A tissue-specific model of chromatin conformation suggests that common variant rs78256848 (MAF 0.11, OR 1.4 for CTD, p=2.6e-08) physically interacts with *NCAM1* (p_{FDR}=1.86e-27), a neural adhesion molecule acting in cardiac development. Importantly, while each individual malformation displayed substantial heritability (observed h2 ranging from 0.26 for complex malformations to 0.37 for LVOTO) the risk for different CHD malformations appeared to be separate, without genetic correlation measured by LD score regression or regional colocalization.

Conclusions: We describe a set of rare noncoding variants conferring significant risk for individual heart malformations which are linked to genes governing cardiac development. These results illustrate that the oligogenic basis of CHD and significant heritability may be linked to rare variants outside protein-coding regions conferring substantial risk for individual categories of cardiac malformation.

Introduction

Congenital heart defects (CHD) comprise a heterogeneous group of malformations of the heart and great vessels, and are the most common cause of mortality during early childhood[1]. Occurring in 0.8–1% of live births, CHD appears to be increasing in prevalence worldwide[2] and survivors are at significantly increased risk of adult-onset cardiovascular disease[3,4,5], neuropsychiatric disease[6], and cancer[7].

CHD is observed to be highly heritable[8–12] but the power to identify inherited genetic risk has been primarily limited to analysis of common variants in small cohorts. Familial clustering of specific malformations indicates that up to 90% of the risk for CHD is attributable to heritable genetic variation[8,9] which includes common[10–12] and rare[13,14] variants linked to specific CHD phenotypes. Mendelian inheritance of deleterious protein coding variation in any individual gene accounts for less than 1% of the burden of CHD[15], and when Mendelian forms of CHD are combined they identify less than 10% of the risk for disease[16]. Therefore, important genetic risk factors for CHD remain to be discovered.

Relative to other types of cardiovascular disease, genome wide association studies (GWAS) of CHD have been underpowered due to the relatively small number of affected individuals included within single studies. Here we bring together individual level data from separate cohorts (Cordell Welcome Case Control (Cordell), Pediatric Cardiac Genomics Consortium (PCGC), German Heart Center Munich (DHM), and UK Biobank (UKB)) totaling 4,597 cases and 50,745 controls. To identify rare inherited variants[13,14] we performed high-quality re-imputation of data from four cohorts (n=55,342) to a large diverse reference panel, permitting meta-analysis of 14,784,017 variants including 6,035,962 rare variants.

Methods

For detailed description of the cohorts, analytical methods, and post-GWAS analyses please see the online supplemental material. For each cohort included in the analysis (described below), local IRB approval was obtained and informed consent performed for all participants as described in the primary reports. The analytic methods and summary statistics are available to other researchers for purposes of reproducing the results or replicating the procedure. Primary individual-level data from the UK Biobank dataset is available to any qualified researcher (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access) and primary individual-level data from the PCGC Cohort is available in dbGaP (Accession: phs000571.v1.p1).

Results

We studied four individual phenotypes: Conotruncal heart disease (CTD), left ventricular outflow tract obstruction (LVOTO), atrial septal defects (ASD), and other forms of complex heart malformations including transposition of the great arteries and related developmental malposition of the outflow tracts (Complex) [Table S.I]. The analysis was augmented by high quality re-imputation of common and rare variants from the TOPMed reference panel (Freeze 5)[17]. After evaluating available approaches for logistic regression for phenotypes with high heritability in small cohorts without related individuals [Figures S.I and S.II], we employed PLINK2 to perform logistic regression upon variants filtering to include rare alleles imputed to high quality with a minor allele count (MAC) of 20 or greater [Supplemental Methods]. Subsequently we performed meta-analysis of three cohorts for each phenotype which was well powered for discovery of common variants in all phenotypes and rare variants for the CTD, LVOTO and ASD phenotypes [Table S.II]. We validated the quality of rare-variant imputation using existing whole genome sequencing data, observing excellent sensitivity of 98.4% [Table S.III].

For each of the four CHD classes, the meta-analysis identified primarily rare variants with moderate or large effect sizes (median odds ratio (OR) 3.02 and median allele frequency 0.008 across 16 lead variants) in addition to a previously described common risk locus at 4p16 for ASD [Figure 1]. Overall, there was little heterogeneity (Cochrane's Q > 0.01) for each lead variants across studies [Table 1, Table S.IV], and diagnostic analyses of each component GWAS suggested an expected distribution of effect size estimates arising from either study or phenotype[18] [Figure S.III]. Six of the 16 observed rare variants were without local linkage partners [Figures S.IV–S.VII].

Overall, we observe many identified loci to display genomic interactions with genes involved in cardiac development across a variety of different embryonic cell types and transitional cell states [Figure 2, Table S.V, Figures S.VIII–S.XVI]. Among the lead and closely linked variants in strong linkage disequilibrium (LD) there were notable findings. For ASD, rs17608766 (MAF 0.14, OR 1.4, p=3.1e-08) is located in the 3' UTR of *GOSR2*, a SNAP receptor in the cis-Golgi. We have previously found that this variant is predicted to disrupt a KLF4 binding motif in an enhancer element predicted in the activity-by-contact (ABC) model to regulate the promoter of *GOSR2* in heart ventricle tissue and endothelial cells[19]. A lead CTD variant rs373447426 (MAF 0.005, OR 3.4, p=1.5e-08) resides in an intron of *RUBCN* and based on the Akita tool for computational analysis of chromatin conformation[20] the alternate allele is predicted to change genomic interactions at two relevant loci across a large region [Fig 2A]. The alternate allele, a deletion of A at 3:197,690,545, is predicted to increase contact at this locus with the promoter of *DLG1* which encodes membrane-associated protein involved in cardiac development in mice[21]. Simultaneously this variant also disrupts contact with the promoter of *BDH1;* duplications of *BDH1* have been previously associated with CTDs[22,23]. For the LVOTO phenotype, a lead variant rs189203952 (MAF 0.01, OR 2.4, p=1.46e-08) is predicted to disrupt the motifs of four transcription factor binding sites known to participate in cardiac development[24–27] and is located in a region of open chromatin 250 bp distal to the promoter of *SPAG9,* a cytoskeletal adaptor protein and mediator of jun-kinase signaling[28] [Fig 2B]. Three-dimensional chromatin data from mesodermal-related tissue suggests that rs78256848, which confers risk for CTD (MAF 0.11, OR 1.4, p=2.6e-08), physically interacts with *NCAM1* (p_{FDR}=1.86e-27), a neural adhesion molecule recently recognized to play a role in aspects of cardiac development[29] [Fig S.VI]. Re-analysis of the genomic-context surrounding the 4p16 risk locus for ASD[10] reveals that the lead variant rs1510798 (MAF 0.23, OR 1.36, p=4.6e-09) is linked with two variants predicted to disrupt motifs for AP-2 or NFAT factors in an enhancer active in mesendodermal precursors differentiated from human embryonic stem cells[30,31] [Fig 2C]. This enhancer is predicted by the ABC model to regulate both *STX18*, another SNAP receptor in the *cis*-Golgi that shares protein-protein interaction partners with GOSR2 (STX5, according to STRINGdb), and *MSX1*, a transcription factor that regulates second heart field and endocardial cushion development[32] [Fig 2C, Table S.VI].

To further assess the biological relevance to human cardiac morphogenesis of the variants and genes identified, we examined the expression of identified genes in relevant tissues from previously published single-cell RNAseq data from developing human heart[33]. Three identified Conotruncal genes (NCAM1, MSC, and DLG1) were observed to be clearly

expressed in mesenchymal cells in the pulmonary valve and MSC expression was observed specifically in fibroblasts in the pulmonary outflow tract [Figure S.XVII]. Genes identified in meta-analysis of ASD were observed to be expressed in Endothelial cells in the left atrium (CALM1, SEMA6D, GOSR2), cardiac fibroblasts in the left atrium (FBLN5), proepicardial origin cells in the left atrium (CALM1, SEMA6D, FBLN5), and proepicardial origin cells in the right atrium (SEMA6D) [Figure S.XVIII]. Analysis of the aortic valve and aorta for LVOTO genes revealed that SPAG9 was strongly expressed in Aortic fibroblasts along with mesenchymal and valvular interstitial cells in the aortic valve [Figure S.XIX]. Together the single-cell data provides additional evidence reinforcing a role for the identified genes in specific anatomical forms of CHD.

For each of the four CHD meta-analyses, which were limited to unrelated individuals, we also estimated both the observed scale and liability scale heritability using LD score regression[34]. Despite relatively small sample sizes, reliable estimates of substantial heritability and total liability were obtained for all four phenotypes [Table S.VI], with observed scale heritability of LVOTO (heritability $h2_{obs}$ =0.371, standard error (se) 0.0845), CTD ($h2_{obs}$ =0.3384, se 0.0794), ASD ($h2_{obs}$ =0.3265, se 0.1425), and Complex malformations ($h2_{obs}$=0.2629, se 0.2198). Liability scale heritability was high for both LVOTO ($h2_{lb}$=0.47, se 0.11) and ASD ($h2_{lb}$=0.6183, se 0.27). We also estimated genetic correlations between the CHD phenotypes and cardiac comorbidities[3], which are common in long-term survivors of CHDs. Importantly, none of the CHD phenotypes displayed significant genetic correlations with one another [Table S.VII.A]. Apart from a mild positive genetic correlation between CTD and coronary artery disease (rg 0.423, p=0.0001), there was a notable absence of genetic correlation for any of the CHD phenotypes with 38 cardiovascular diseases and phenotypes including arrhythmia and heart failure [Table S.VII.B].

With discovered loci across all four CHD phenotypes below the allele frequency thresholds included in LDSC, we additionally performed Bayesian multi-colocalization analyses[35] to specifically and systematically analyze the relationship of uncommon and rare variants between the studies at discovered loci. Regional colocalization analysis confirmed a notable absence of overlap between phenotypes within local linkage structures across 1-megabase regions surrounding the lead variants detected by the four meta-analyses [Table S.VIII]. Overall, the absence of common large-scale genetic correlation and local rare-variant colocalization may suggest largely distinct genetic architectures for different CHD malformations, which are not shared with other forms of cardiovascular phenotypes or disease.

Discussion

In summary, these findings support a set of oligogenic genetic architectures unique to different CHD phenotypes, which include risk centered around uncommon or rare genetic variation with larger effect sizes. While the approach to our meta-analysis is limited by sample size and the use of external controls from the UK Biobank for the PCGC GWAS, our estimates of heritability derived from LDSC are greater than many common forms of cardiovascular disease and consistent with longstanding clinical and epidemiological

observations of familial recurrence of cardiac malformations[8,36]. Notably, these heritability estimates are computed from population-based linkage structures of common genetic variation, which omits the heritability contributed by rare variants reported here and may therefore be conservative underestimates. The absence of regional colocalization between CHD phenotypes at discovered loci further suggests that the genetic risk for one malformation may not confer similar risk for other malformations.

Recent studies of population isolates have identified rare inherited coding variants underlying CHD[14,16,37], while our analysis revealed noncoding variants which are often more difficult to interpret. The analysis of noncoding variation impacting cardiac malformations is further complicated by developmental timing: genetic effects are likely to be exerted in a brief window of three weeks during human embryonic development. These transitional cell types and intermediate tissues of cardiogenesis[38] may not have corresponding or analogous eQTL or chromatin conformation data derived from adult tissues[39]. The genomic consequences for specific variants, suggested by analyses of chromatin structure from early cardiac development, models of three-dimensional genomic conformations, and the dynamic relationship of enhancer-chromatin interactions are centered around genes experimentally related to cardiovascular development. For a number of our findings, analyses of single-cell data from human heart support the expression of the identified genes in tissues and cell-types directly relevant to the specific cardiac malformations. Together, these findings are consistent with a presumed mechanism disrupting formation of the heart.

These findings are not without limitations. While we included only variants with high-quality imputation and made careful choices in our analytical schema to control error, our study of CHD remains underpowered relative to much larger studies of continuous cardiovascular traits that included rare variants[40,41]. Our analysis included data from three separate centers and meta-analysis demonstrated variant effects concordant between cohorts, however external replication of our findings in newly gathered and carefully phenotyped cohorts of people affected with CHD is necessary to better understand the generalizability of the reported associations and improve the power to detect rare genetic variation associated with cardiac malformations. The findings might further be supported by identity-by-state analyses of rare-variant containing haplotypes which are not feasible given the limitations of the included cohort studies. Many of the loci require further investigation to confirm the transcriptional and genomic mechanisms suggested by *in silico* analyses to relate variants to specific genes, a number of which have not been previously implicated in cardiac development[42]. Additionally, although our initial assessment of the expression of these genes in human cardiac tissue is encouraging, not all of the CHD phenotypes tested are represented in this single-cell dataset, and not all genes show clear expression patterns, while some of our genes are likely also expressed in other parts of the developing heart. We acknowledge that further experimental work is necessary to validate and dissect the mechanism by which these genetic loci disrupt cardiac development and contribute to the risk of CHD.

The findings reported here may have important implications for both CHD research and clinical practice. Both research and clinical genetic testing for CHD have primarily focused

upon finding deleterious protein coding variation in genes related to cardiac development—all of the variants reported here are suggested by *in silico* analyses to be linked to cardiac developmental genes which may indicate the need for new approaches in modeling the mechanism of action for regulatory genetic variation in cardiac development. From a clinical perspective, the discovered non-coding genetic variation would not be detected by panel genetic testing of coding regions for a small number of genes previously implicated in CHD. The high heritability observed for these four CHD malformations could suggest that polygenic scoring may contribute to understanding the origins for different forms of CHD[43].

Overall, these data, particularly the significant risk associated with uncommon and rare genetic variants now accessible by high-quality imputation, may have implications for other heritable pediatric diseases and anatomical malformations for which discovery of genetic risk has remained elusive. In summary, alongside with *de novo* coding variation[16] and copy number variants[23], these results suggest the genetic architecture of CHD appears to include a substantial heritable proportion of rare variants largely specific to individual cardiac malformations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We are indebted to the patients affected with congenital heart disease who participated in all four cohort studies.

Sources of Funding:

This research was supported by grants to the Pediatric Cardiac Genomics Consortium (UM1-HL098147, UM1-HL128761, UM1-HL098123, UM1-HL128711, UM1-HL098162, U01-HL131003, U01-HL098153, U01-HL098163), and the National Center for Research Resources (U01-HL098153).

Nonstandard Abbreviations and Acronyms

ABC	activity by contact
ASD	atrial septal defect
ATAC-seq	Assay for Transposase-Accessible Chromatin with sequencing
CHD	congenital heart disease
CTD	Conotruncal heart disease
DHM	German Heart Center of Munich
FIMO	find individual motif occurences
GWAS	Genome wide association study
Lambda GC	Lambda Genomic Control
LDSC	Linkage disequilibrium score correlation

LVOTO	left ventricular outflow tract obstructive disease
MAC	minor allele count
MAF	minor allele frequency
OR	odds ratio
PCGC	pediatric cardiac genomics consortium
QQ Plot	quantile quantile plot
rg	regression intercept from LDSC
TOPMed	NHLBI Trans-omics for Precision Medicine
UKB	UK Biobank
WTCCC	Wellcome Trust Case Control Consortium

References:

1. Heron M Deaths: Leading Causes for 2015. Natl Vital Stat Rep [Internet]. 2017 [cited 2019 Sep 9];66:1–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29235984
2. Liu Y, Chen S, Zühlke L, Black GC, Choy M, Li N, Keavney BD. Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. Int J Epidemiol [Internet]. 2019 [cited 2019 Mar 4];Available from: http://www.ncbi.nlm.nih.gov/pubmed/30783674
3. Saha P, Potiny P, Rigdon J, Morello M, Tcheandjieu C, Romfh A, Fernandes SM, McElhinney DB, Bernstein D, Lui GK, et al. Substantial Cardiovascular Morbidity in Adults with Lower-Complexity Congenital Heart Disease. Circulation [Internet]. 2019 [cited 2019 Mar 8];Manuscript:CIRCULATIONAHA.118.037064. Available from: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.118.037064
4. Bokma JP, Zegstroo I, Kuijpers JM, Konings TC, van Kimmenade RRJ, van Melle JP, Kiès P, Mulder BJM, Bouma BJ. Factors associated with coronary artery disease and stroke in adults with congenital heart disease. Heart [Internet]. 2017 [cited 2017 Sep 26];heartjnl-2017-311620. Available from: http://heart.bmj.com/lookup/doi/10.1136/heartjnl-2017-311620
5. Labombarda F, Hamilton R, Shohoudi A, Aboulhosn J, Broberg CS, Chaix MA, Cohen S, Cook S, Dore A, Fernandes SM, et al. Increasing Prevalence of Atrial Fibrillation and Permanent Atrial Arrhythmias in Congenital Heart Disease. J Am Coll Cardiol [Internet]. 2017;70:857–865. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0735109717378579 [PubMed: 28797355]
6. Kasmi L, Bonnet D, Montreuil M, Kalfa D, Geronikola N, Bellinger DC, Calderon J. Neuropsychological and Psychiatric Outcomes in Dextro-Transposition of the Great Arteries across the Lifespan: A State-of-the-Art Review. Front Pediatr [Internet]. 2017 [cited 2017 Sep 26];5. Available from: http://journal.frontiersin.org/article/10.3389/fped.2017.00059/full
7. Diller G-P, Kempny A, Alonso-Gonzalez R, Swan L, Uebing A, Li W, Babu-Narayan S, Wort SJ, Dimopoulos K, Gatzoulis MA. Survival Prospects and Circumstances of Death in Contemporary Adult Congenital Heart Disease Patients Under Follow-Up at a Large Tertiary Centre. Circulation [Internet]. 2015 [cited 2019 Apr 12];132:2118–2125. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26369353 [PubMed: 26369353]
8. Øyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PKA, Melbye M. Recurrence of congenital heart defects in families. Circulation [Internet]. 2009;120:295–301. Available from: http://circ.ahajournals.org/content/120/4/295.long [PubMed: 19597048]

9. Øyen N, Poulsen G, Wohlfahrt J, Boyd H a., Jensen PK a, Melbye M. Recurrence of discordant congenital heart defects in families. Circ Cardiovasc Genet. 2010;3:122–128. [PubMed: 20173214]
10. Cordell HJ, Bentham J, Topf A, Zelenika D, Heath S, Mamasoula C, Cosgrove C, Blue G, Granados-Riveron J, Setchfield K, et al. Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. Nat Genet [Internet]. 2013;45:822–4. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3793630&tool=pmcentrez&rendertype=abstract [PubMed: 23708191]
11. Zhao B, Lin Y, Xu J, Ni B, Da M, Ding C, Hu Y, Zhang K, Yang S, Wang X, et al. Replication of the 4p16 susceptibility locus in congenital heart disease in Han Chinese populations. PLoS One [Internet]. 2014 [cited 2019 Sep 12];9:e107411. Available from: http://dx.plos.org/10.1371/journal.pone.0107411 [PubMed: 25215500]
12. Córdova-Palomera A, Priest JR. Association between the 4p16 genomic locus and different types of congenital heart disease: results from adult survivors in the UK Biobank. Sci Rep. 2019;9.
13. Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadottir HT, Zanon C, Magnusson OT, Helgason A, Saemundsdottir J, Gylfason A, et al. A rare variant in MYH6 is associated with high risk of sick sinus syndrome. Nat Genet [Internet]. 2011;43:316–320. Available from: http://www.nature.com.laneproxy.stanford.edu/ng/journal/v43/n4/full/ng.781.html [PubMed: 21378987]
14. Bjornsson T, Thorolfsdottir RB, Sveinbjornsson G, Sulem P, Norddahl GL, Helgadottir A, Gretarsdottir S, Magnusdottir A, Danielsen R, Sigurdsson EL, et al. A rare missense mutation in MYH6 associates with non-syndromic coarctation of the aorta. Eur Heart J [Internet]. 2018 [cited 2019 Sep 13];39:3243–3249. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29590334 [PubMed: 29590334]
15. Rauch R, Hofbeck M, Zweier C, Koch A, Zink S, Trautmann U, Hoyer J, Kaulitz R, Singer H, Rauch A. Comprehensive genotype-phenotype analysis in 230 patients with tetralogy of Fallot. J Med Genet [Internet]. 2010;47:321–331. Available from: http://jmg.bmj.com/content/47/5/321.full [PubMed: 19948535]
16. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, et al. 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 [PubMed: 28991257]
17. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. bioRxiv [Internet]. 2019 [cited 2019 Sep 12];563866. Available from: https://www.biorxiv.org/content/10.1101/563866v1
18. O'Connor LJ. The distribution of common-variant effect sizes. Nat Genet 2021 538 [Internet]. 2021 [cited 2022 Feb 15];53:1243–1249. Available from: https://www.nature.com/articles/s41588-021-00901-3
19. Yu M, Tcheandjieu C, Georges A, Xiao K, Tejeda H, Dina C, Le Tourneau T, Fiterau M, Judy R, Tsao NL, et al. Computational estimates of annular diameter reveal genetic determinants of mitral valve function and disease. JCI insight [Internet]. 2022 [cited 2022 Feb 9];7. Available from: https://pubmed.ncbi.nlm.nih.gov/35132965/
20. 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/ [PubMed: 33046897]
21. Iizuka-Kogo A, Senda T, Akiyama T, Shimomura A, Nomura R, Hasegawa Y, Yamamura KI, Kogo H, Sawai N, Matsuzaki T. Requirement of DLG1 for cardiovascular development and tissue elongation during cochlear, enteric, and skeletal development: Possible role in convergent extension. PLoS One. 2015;10.
22. Soemedi R, Topf A, Wilson IJ, Darlay R, Rahman T, Glen E, Hall D, Huang N, Bentham J, Bhattacharya S, et al. Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. Hum Mol Genet [Internet]. 2012;21:1513–1520. Available from: http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddr589 [PubMed: 22199024]
23. Glessner JT, Bick AG, Ito K, Homsy JG, Rodriguez-Murillo L, Fromer M, Mazaika E, Vardarajan B, Italia M, Leipzig J, et al. Increased frequency of de novo copy number variants in congenital

heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. Circ Res. 2014;115.

24. Duim SN, Goumans M-J, Kruithof BPT. WT1 in Cardiac Development and Disease. Wilms Tumor [Internet]. 2016 [cited 2022 Jan 25];211–233. Available from: https://pubmed.ncbi.nlm.nih.gov/27512758/
25. Valentino T, Palmieri D, Vitiello M, Simeone A, Palma G, Arra C, Chieffi P, Chiariotti L, Fusco A, Fedele M. Embryonic defects and growth alteration in mice with homozygous disruption of the Patz1 gene. J Cell Physiol [Internet]. 2013 [cited 2022 Jan 25];228:646–653. Available from: https://pubmed.ncbi.nlm.nih.gov/22886576/ [PubMed: 22886576]
26. Doppler SA, Werner A, Barz M, Lahm H, Deutsch MA, Dreßen M, Schiemann M, Voss B, Gregoire S, Kuppusamy R, et al. Myeloid zinc finger 1 (Mzf1) differentially modulates murine cardiogenesis by interacting with an Nkx2.5 cardiac enhancer. PLoS One [Internet]. 2014 [cited 2022 Jan 25];9. Available from: https://pubmed.ncbi.nlm.nih.gov/25436607/
27. Stevens SJC, van Essen AJ, van Ravenswaaij CMA, Elias AF, Haven JA, Lelieveld SH, Pfundt R, Nillesen WM, Yntema HG, van Roozendaal K, et al. Truncating de novo mutations in the Krüppel-type zinc-finger gene ZNF148 in patients with corpus callosum defects, developmental delay, short stature, and dysmorphisms. Genome Med [Internet]. 2016 [cited 2022 Jan 25];8. Available from: https://pubmed.ncbi.nlm.nih.gov/27964749/
28. Dixon DM, Choi J, El-Ghazali A, Park SY, Roos KP, Jordan MC, Fishbein MC, Comai L, Reddy S. Loss of muscleblind-like 1 results in cardiac pathology and persistence of embryonic splice isoforms. Sci Rep [Internet]. 2015 [cited 2022 Jan 25];5. Available from: https://pubmed.ncbi.nlm.nih.gov/25761764/
29. Delgado C, Bu L, Zhang J, Liu FY, Sall J, Liang FX, Furley AJ, Fishman GI. Neural cell adhesion molecule is required for ventricular conduction system development. Dev. 2021;148.
30. Wang J, Ji W, Zhu D, Wang W, Chen Y, Zhang Z, Li F. Tfap2b mutation in mouse results in patent ductus arteriosus and renal malformation. J Surg Res. 2018;227.
31. Bushdid PB, Osinska H, Wacław RR, Mołkentin JD, Yutzey KE. NFATc3 and NFATc4 are required for cardiac development and mitochondrial function. Circ Res. 2003;92.
32. Chen YH, Ishii M, Sun J, Sucov HM, Maxson RE. Msx1 and Msx2 regulate survival of secondary heart field precursors and post-migratory proliferation of cardiac neural crest in the outflow tract. Dev Biol [Internet]. 2007 [cited 2022 Feb 15];308:421–437. Available from: https://pubmed.ncbi.nlm.nih.gov/17601530/ [PubMed: 17601530]
33. Cui Y, Zheng Y, Liu X, Yan L, Fan X, Yong J, Hu Y, Dong J, Li Q, Wu X, et al. Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. Cell Rep. 2019;26:1934–1950.e5. [PubMed: 30759401]
34. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, Duncan L, Perry JRB, Patterson N, Robinson EB, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet [Internet]. 2015 [cited 2019 Sep 19];47:1236–1241. Available from: http://www.nature.com/articles/ng.3406 [PubMed: 26414676]
35. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, Howson JMM. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun. 2021;12.
36. Nora JJ, Nora AH. Maternal transmission of congenital heart diseases: new recurrence risk figures and the questions of cytoplasmic inheritance and vulnerability to teratogens. Am J Cardiol [Internet]. 1987;59:459–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3812316 [PubMed: 3812316]
37. Helgadottir A, Thorleifsson G, Gretarsdottir S, Stefansson OA, Tragante V, Thorolfsdottir RB, Jonsdottir I, Bjornsson T, Steinthorsdottir V, Verweij N, et al. Genome-wide analysis yields new loci associating with aortic valve stenosis. Nat Commun [Internet]. 2018 [cited 2019 Sep 5];9:987. Available from: http://www.nature.com/articles/s41467-018-03252-6 [PubMed: 29511194]
38. Lin CJ, Lin CY, Chen CH, Zhou B, Chang CP. Partitioning the heart: mechanisms of cardiac septation and valve development. Development [Internet]. 2012;139:3277–3299. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.063495 [PubMed: 22912411]

39. Elorbany R, Popp JM, Rhodes K, Strober BJ, Qi G, Gilad Y, Battle A. Single-Cell Sequencing Reveals Lineage-Specific Dynamic Genetic Regulation of Gene 1 Expression During Human Cardiomyocyte Differentiation 2 3 4. bioRxiv. 2021;
40. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell [Internet]. 2016 [cited 2019 May 23];167:1415–1429.e19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27863252 [PubMed: 27863252]
41. Lin H, van Setten J, Smith AV, Bihlmeyer NA, Warren HR, Brody JA, Radmanesh F, Hall L, Grarup N, Müller-Nurasyid M, et al. Common and Rare Coding Genetic Variation Underlying the Electrocardiographic PR Interval. Circ Genomic Precis Med. 2018;11:e002037.
42. de Soysa TY, Ranade SS, Okawa S, Ravichandran S, Huang Y, Salunga HT, Schricker A, Del Sol A, Gifford CA, Srivastava D. Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. Nature [Internet]. 2019 [cited 2019 Sep 23];572:120–124. Available from: http://www.nature.com/articles/s41586-019-1414-x [PubMed: 31341279]
43. Richter F, Morton SU, Kim SW, Kitaygorodsky A, Wasson LK, Chen KM, Zhou J, Qi H, Patel N, DePalma SR, et al. Genomic analyses implicate noncoding de novo variants in congenital heart disease. Nat Genet [Internet]. 2020 [cited 2022 Dec 13];52:769–777. Available from: https://pubmed.ncbi.nlm.nih.gov/32601476/ [PubMed: 32601476]
44. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics [Internet]. 2010 [cited 2015 Aug 13];26:2867–73. Available from: http://bioinformatics.oxfordjournals.org/content/26/22/2867.long [PubMed: 20926424]
45. Hoang TT, Goldmuntz E, Roberts AE, Chung WK, Kline JK, Deanfield JE, Giardini A, Aleman A, Gelb BD, Mac Neal M, et al. The Congenital Heart Disease Genetic Network Study: Cohort description. PLoS One [Internet]. 2018 [cited 2019 Jan 23];13:e0191319. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29351346 [PubMed: 29351346]
46. 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 [PubMed: 20616382]
47. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, Benner C, O'Dushlaine C, Barber M, Boutkov B, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet 2021 537 [Internet]. 2021 [cited 2022 Jun 15];53:1097–1103. Available from: https://www.nature.com/articles/s41588-021-00870-7
48. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM, Corvin A, 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/ [PubMed: 25642630]
49. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat. Genet 2019;51. [PubMed: 30578418]
50. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F, et al. Genome-wide enhancer maps link risk variants to disease genes. Nature [Internet]. 2021 [cited 2021 Jun 22];593:238–243. Available from: https://www.nature.com/articles/s41586-021-03446-x [PubMed: 33828297]
51. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489.
52. Kulakovskiy IV, Boeva VA, Favorov AV, Makeev VJ. Deep and wide digging for binding motifs in ChIP-Seq data. Bioinformatics. 2010;26.
53. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, Medvedeva YA, Magana-Mora A, Bajic VB, Papatsenko DA, et al. 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.
54. Ward LD, Kellis M. HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 2016;44.

55. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun [Internet]. 2017 [cited 2020 Sep 14];8. Available from: /pmc/articles/PMC5705698/?report=abstract

56. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods [Internet]. 2017 [cited 2019 May 20];14:979–982. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28825705 [PubMed: 28825705]

57. Package “MASS.” 2022 [cited 2022 Oct 4];Available from: http://www.stats.ox.ac.uk/pub/MASS4/

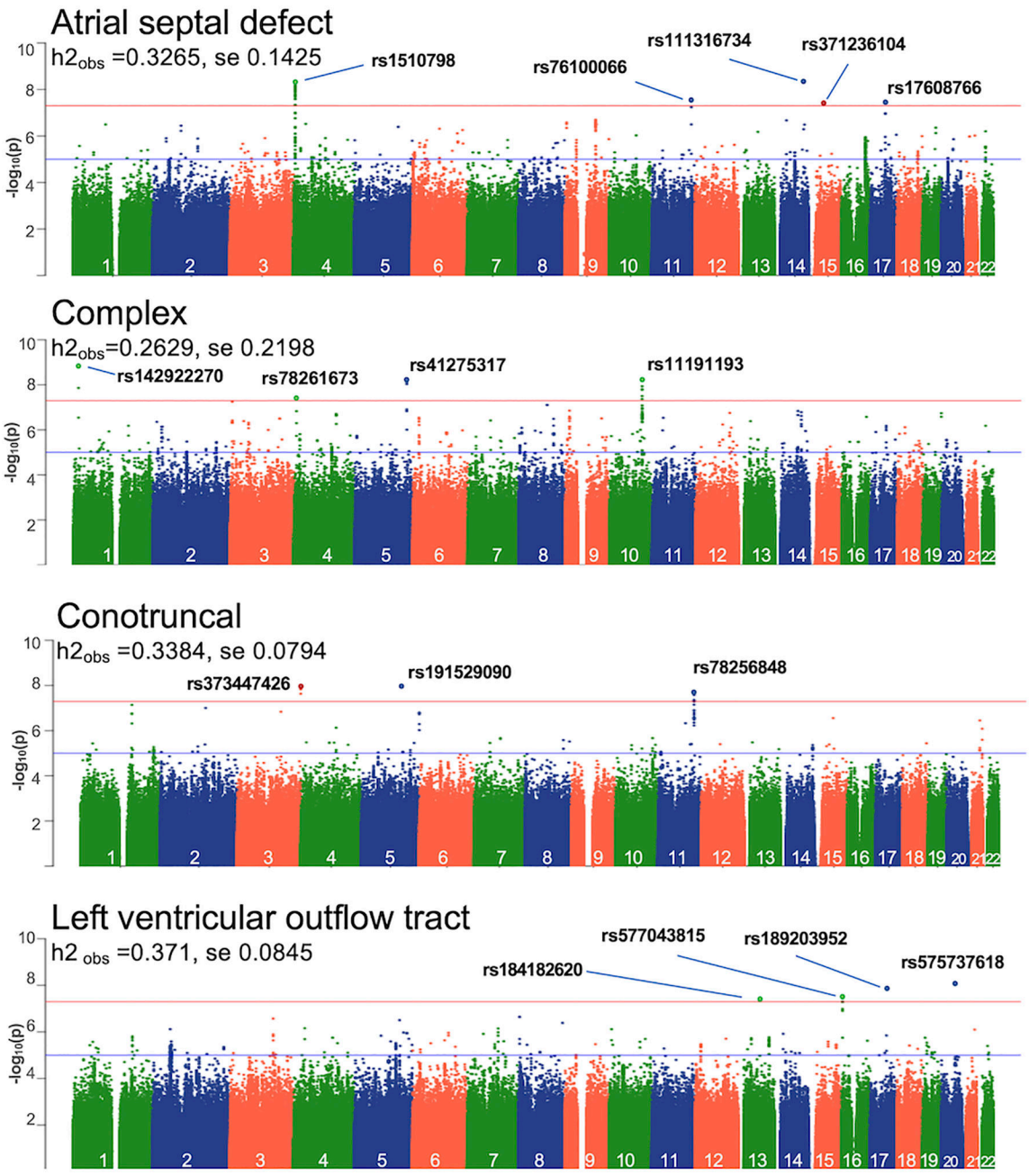

Figure 1.

Manhattan plots and component variants for meta-analysis of four congenital heart disease phenotypes. For each lead genetic variant, rsID and coordinate:variant identifiers are provided on the Manhattan plot. Comparing studies, allele frequencies are generally consistent despite genotyping on different platforms, and there is remarkable consistency in the direction and magnitude of risk estimates derived from each independent study. Detailed local linkage patterns for each hit are displayed in LocusZoom plots in figures S.IV through S.VII. All coordinates are from the GRCh38 genome build.

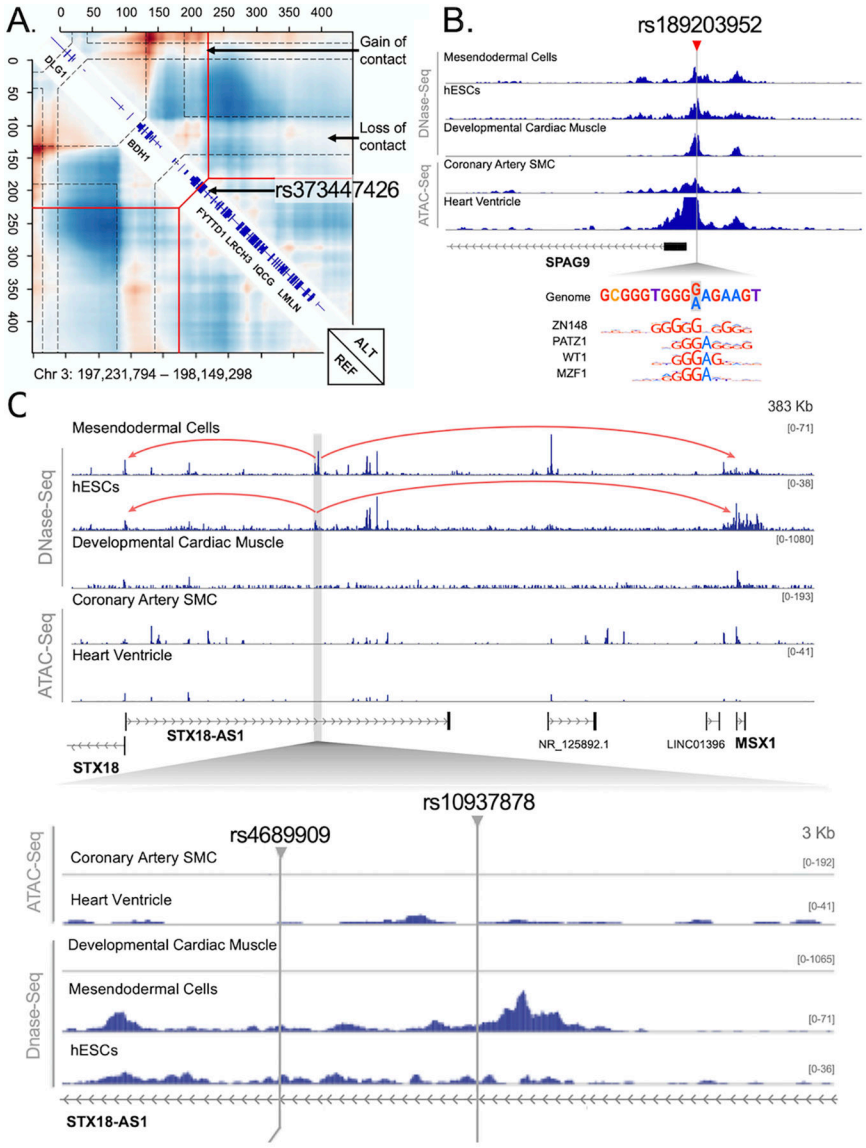

Figure 2.

Key variants are predicted to disrupt genomic context proximal to genes related to cardiac development. **Panel A.** Predicted contact frequencies of the 1 megabase region surrounding variant rs373447426 (MAF 0.005, OR 3.4 for CTD, p=1.5e-08) for the reference (bottom left) and alternate (top right) alleles with the genes running down the diagonal. The structural similarity index for the alternate variant relative to the reference variant drops by 0.21 evidenced by regions of increased and decreased genomic contact noted directly with arrows. Regions with predicted gain and loss of genome contact are annotated with dashed lines within which sit the affected genes, *DLG1* and *BDH1*, respectively and the variant is annotated with a red line. In the heatmap, red denotes greater physical proximity between two regions than expected given the genomic distance and blue represents the opposite. Axes are genomic coordinates marked by 448 bins of 2048 bp. **Panel B.** Analysis of rs189203952 (MAF 0.01, OR 2.4 LVOTO, p=1.46e-08) shows significant disruption to four transcription

factor binding sites (Table S.VIX) within a region of open chromatin in the promoter of SPAG9 which is shared by smooth muscle cells, heart ventricle, developmental cardiac muscle, mesendoderm and human embryonic stem cells. **Panel C.** The activity-by-contact model reveals a long-range interaction between two variants rs10937878 and rs4689909 which are tightly linked (R^2 0.935) to the lead variant rs1510798 (MAF 0.23, OR 1.36 ASD, p=4.6e-09) in a region of open chromatin in mesendoderm and embryonic stem cells and disrupts the binding sites of two transcription factors well known to play a role in cardiac development NFATC4 and TFAP2B, with an interaction approximately 300kbp upstream to the promoter of *STX18*.

Table 1.

Meta-analysis across three studies reveals rare variants conferring significant risk for four congenital heart disease phenotypes.

						Meta-analysis			PCGC			Cordell			UK Biobank		
CHD	**rsid**	**chr**	**Position**	**A1**	**A2**	**A1Freq**	**OR**	**p-value**	**A1 Freq**	**OR**	**p-value**	**A1 Freq**	**OR**	**p-value**	**A1 Freq**	**OR**	**p-value**
ASD	rs1510798	4	4627439	T	G	0.226	1.36	4.60E-09	0.230	1.16	0.08	0.230	1.50	2.04E-06	0.240	1.51	1.40E-04
	rs76100066	11	123060309	T	C	0.018	2.23	3.04E-08	0.020	2.15	6.71E-04	0.020	2.82	1.84E-07	0.020	0.34	0.07
	rs111316734	14	90846774	C	G	0.004	4.54	4.23E-09	0.004	2.25	0.10	0.004	6.60	1.67E-06	0.004	5.01	5.82E-04
	rs371236104	15	45862310	C	G	0.003	6.50	4.38E-08	0.004	5.08	6.77E-04	n/a	n/a	n/a	0.003	8.44	1.33E-05
	rs17608766	17	46935905	C	T	0.151	1.41	3.11E-08	0.130	1.51	2.57E-05	0.150	1.48	9.16E-05	0.140	1.14	0.35
						Meta-analysis			**PCGC**			**Cordell**			**UK Biobank**		
CTD	rs373447426	3	197690545	T	TA	**0.003**	**3.37**	**1.49E-08**	0.005	2.97	1.26E-06	0.006	1.31E-05	0.36	0.003	6.00	3.31E-04
	rs191529090	5	123892360	A	G	**0.004**	**3.65**	**1.04E-08**	0.004	3.62	2.20E-07	0.004	0.05	0.45	0.005	3.93	0.002
	rs78256848	11	112415609	C	T	**0.899**	**1.41**	**2.63E-08**	0.890	1.42	9.23E-08	0.900	1.11	0.59	0.890	1.52	0.0098
						Meta-analysis			**PCGC**			**Cordell**			**UK Biobank**		
Complex	rs142922270	1	16211007	A	G	**0.006**	**3.87**	**1.37E-08**	0.007	2.62	0.018	0.006	5.01	5.36E-06	0.005	4.09	0.006
	rs78261673	4	7461701	T	A	**0.011**	**3.01**	**3.81E-08**	0.010	3.26	1.92E-04	0.010	2.46	0.007	0.008	3.47	0.001
	rs41275317	5	161890651	C	T	**0.013**	**3.05**	**6.72E-09**	0.010	2.00	0.03	0.009	4.31	1.24E-06	0.010	3.16	0.003
	rs11191193	10	102042651	G	A	**0.361**	**1.42**	**6.38E-09**	0.330	1.48	2.34E-05	0.360	1.33	0.005	0.350	1.44	0.003
						Meta-analysis			**PCGC**			**Cordell**			**UK Biobank**		
LVOTO	rs184182620	13	67948414	C	G	**0.003**	**3.93**	**4.61E-08**	0.003	3.07	0.001	0.003	5.89	4.75E-05	0.003	2.40	0.25
	rs577043815	16	3612650	G	C	**0.004**	**4.48**	**3.60E-08**	0.002	2.53	0.017	0.002	4.55	0.006	0.004	10.04	2.23E-06
	rs189203952	17	51121079	A	G	**0.011**	**2.42**	**1.46E-08**	0.010	2.67	1.33E-07	0.010	1.46	0.257	0.010	3.25	0.004
	rs575737618	20	43936594	A	G	**0.006**	**2.75**	**8.17E-09**	0.007	3.19	8.30E-08	0.006	2.64	0.005	0.010	1.60	0.306