

RESEARCH

Open Access



Genome-wide association studies highlight novel risk loci for septal defects and left-sided congenital heart defects

Martin Broberg^{1,2}, Minna Ampuja¹, Samuel Jones², Tiina Ojala³, Otto Rahkonen³, Riikka Kivelä^{1,4,5}, James Priest⁶, FinnGen, Aarno Palotie², Hanna M. Ollila^{2,7,8,9} and Emmi Helle^{1,3,10,11*}

Abstract

Background Congenital heart defects (CHD) are structural defects of the heart affecting approximately 1% of newborns. They exhibit low penetrance and non-Mendelian patterns of inheritance as varied and complex traits. While genetic factors are known to play an important role in the development of CHD, the specific genetics remain unknown for the majority of patients. To elucidate the underlying genetic risk, we performed a genome wide association study (GWAS) of CHDs in general and specific CHD subgroups using the FinnGen Release 10 (R10) ($N > 393,000$), followed by functional fine-mapping through eQTL and co-localization analyses using the GTEx database.

Results We discovered three genome-wide significant loci associated with general CHD. Two of them were located in chromosome 17: 17q21.32 (rs2316327, intronic: *LRRC37A2*, Odds ratio (OR) [95% Confidence Interval (CI)] = 1.17[1.12–1.23], $p = 1.5 \times 10^{-9}$) and 17q25.3 (rs1293973611, nearest: *BAHCC1*, OR[95%CI] = 4.48[2.80–7.17], $p = 7.0 \times 10^{-10}$), respectively, and in addition to general CHD, the rs1293973611 locus was associated with the septal defect subtype. The third locus was in band 1p21.2 (rs35046143, nearest: *PALMD*, OR[95%CI] = 1.15[1.09–1.21], $p = 7.1 \times 10^{-9}$), and it was associated with general CHD and left-sided lesions. In the subgroup analysis, two additional loci were associated with septal defects (rs75230966 and rs6824295), and one with left-sided lesions (rs1305393195). In the eQTL analysis the variants rs2316327 (general CHD), and rs75230966 (septal defects) both located in 17q21.32 (with a LD r^2 of 0.41) were both predicted to significantly associate with the expression of *WNT9B* in the atrial appendage tissue category. This effect was further confirmed by co-localization analysis, which also implicated *WNT3* expression in the atrial appendage. A meta-analysis of general CHD together with the UK Biobank (combined $N = 881,678$) provided a different genome-wide significant locus in *LRRC37A2*; rs16941382 (OR[95%CI] = 1.15[1.11–1.20], $p = 1.5 \times 10^{-9}$) which is in significant LD with rs2316327.

Conclusions Our results of general CHD and different CHD subcategories identified a complex risk locus on chromosome 17 near *BAHCC1* and *LRRC37A2*, interacting with the genes *WNT9B*, *WNT3* and *MYL4*, may constitute potential novel CHD risk associated loci, warranting future experimental tests to determine their role.

Keywords Genetics, Congenital heart disease, Genome-wide association study, Single nucleotide polymorphisms

*Correspondence:

Emmi Helle
emmi.helle@helsinki.fi

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Congenital heart defects (CHD) are globally the most common birth defect. CHD include a wide range of structural malformations of the heart and great vessels ranging from mild defects, such as bicuspid aortic valve (BAV), which might not cause any symptoms, to critical defects, such as transposition of the great arteries (TGA) and hypoplastic left heart syndrome (HLHS), which are generally lethal without surgical interventions. CHD have their origin in the first trimester of pregnancy [1, 2], and may occur as an element of a syndrome or as an isolated defect. Although monogenic cases have been reported, most isolated CHD are thought to be complex traits with reduced penetrance and variable disease severity and phenotype [3]. In addition to genetic factors, environmental exposures such as maternal obesity, diabetes, and certain medications increase the risk for CHD in the offspring [4–7].

Next-generation sequencing studies have identified that the haploinsufficiency of NOTCH- and VEGF-signaling pathway genes associate with left ventricular outflow tract obstruction (LVOTO) defects and tetralogy of Fallot (TOF), respectively [8–10]. Damaging variants in transcription factors, such as *NKX2-5*, *GATA4*, and *TBX5* have been associated with various CHD, and familial septal defects in particular [11–13]. Excess of rare inherited variants have been identified in isolated CHD subjects [14, 15]. In most isolated CHD cases, however, the genetic etiology remains to be identified.

As genome-wide sequence data are available for large populations, genome-wide association studies (GWAS) have been used as an additional tool to identify genetic determinants of CHD [16, 17]. Previous studies have

identified genome-wide significant loci for all CHD [18] as well as for certain CHD subgroups, such as TOF [19], conotruncal heart defects and LVOTO [20], and coarctation of the aorta [21]. As genetic and phenotypic data for hundreds of thousands of individuals have become integrated into large biobank-based projects such as FinnGen, UK Biobank (UKBB), and Japan Biobank, the statistical strength for population-wide GWAS has increased (<https://finngen.gitbook.io/documentation/>). Here, we conduct a GWAS of general CHD and three CHD subcategories and identify three significant risk loci in the FinnGen and UK Biobank data.

Results

FinnGen GWAS highlights SNPs in chromosome 17 near *BAHCC1* and *LRRC37A2*, and in chromosome 1 near *PALMD* as significantly associated with CHDs

We detected three loci in the general CHD GWAS (Table 1) rs35046143 (chr1, near *PALMD*, bp 99,567,756), rs2316327 (chr17, intronic *LRRC37A2*, bp 47,001,132) and rs1293973611 (chr17, near *BAHCC1*, bp 81,382,139, FinnGen imputed INFO score 0.97) as depicted in Fig. 1 and Table 1. The SNP rs1293973611 was also found significant in a GWAS of septal defects (Odds ratio (OR) [95% Confidence Interval (CI)]=6.23[3.60–10.80], $p=1.2 \times 10^{-10}$). Furthermore, SNPs rs75230966 (chr 17, intronic *LRRC37A2*, bp 46,890,164) and rs6824295 (chr 4, near *STX18*, bp 4,612,553) also associated with septal defects. Both rs2316327 and rs75230966 are found in the chromosomal band 17q21.32 (Fig. 2). In FinnGen the calculated LD r^2 between two SNPs is 0.41. In analyzing left-sided lesions, the general CHD associated SNP rs35046143 in chr1 was significant, and an additional

Table 1 Lead SNPs detected in FinnGen. Specific data for all lead SNPs across the three CHD categories with significant findings. Conotruncal defects results are not included as they did not include any genome-wide significant SNPs. NFEESignifies our results compared to Non-Finnish Non-Estonian Europeans, based on the GnomAD 2.1 reference panel containing 5421 individuals. MAF signifies minor allele frequency

Trait	Lead SNP	GRCh38 Position	Ref Allele	Alt Allele	Odds Ratio [95% Confidence Interval]	p	Finnish Enrichment vs NFEESignifies our results compared to Non-Finnish Non-Estonian Europeans, based on the GnomAD 2.1 reference panel containing 5421 individuals.	MAF
General CHD								
	rs35046143	1:99,567,756	TA	A	1.15[1.09–1.21]	7.1×10^{-9}	NA GnomAD 2.1	0.48
	rs2316327	17:47,001,132	G	A	1.17[1.12–1.23]	1.5×10^{-9}	1.6	0.29
	rs1293973611	17:81,382,139	C	T	4.48[2.80–7.17]	7.0×10^{-10}	NA GnomAD 2.1	9.3×10^{-4}
Septal defects								
	rs6824295	4:4,612,553	C	T	1.23[1.15–1.33]	1.2×10^{-8}	1	0.24
	rs75230966	17:46,890,164	G	A	1.27[1.17–1.38]	7.9×10^{-9}	1.3	0.16
	rs1293973611	17:81,382,139	C	T	6.23[3.60–10.80]	1.2×10^{-10}	NA GnomAD 2.1	9.3×10^{-4}
Left-sided lesions								
	rs35046143	1:99,567,756	TA	A	1.17[1.11–1.24]	3.2×10^{-8}	NA GnomAD 2.1	0.48
	rs1305393195	3:3,995,049	T	G	11.02[10.15–11.97]	6.2×10^{-9}	NA GnomAD 2.1	2.7×10^{-4}

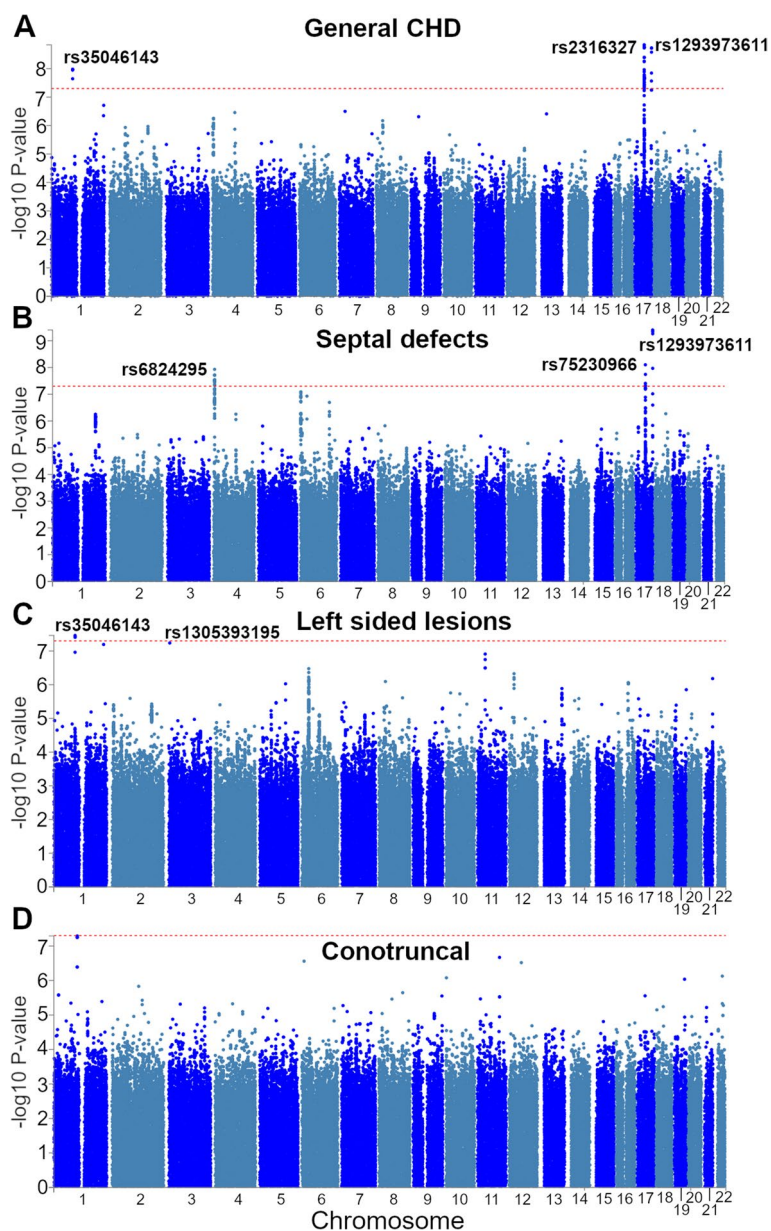


Fig. 1 Lead SNPs associating with CHD identified in FinnGen R10. Manhattan plots of the different CHD categories: **A** general CHD, **B** septal defects, **C** left-sided lesions and **D** conotruncal defects, with genome-wide significant lead SNPs indicated, meaning the SNPs within a region with the highest association signal, as the other SNPs may only show association due to being in linkage disequilibrium (LD) with the lead signal

SNP rs1305393195 (chr 3, near *PNTPI1*, bp 3,995,049, FinnGen imputed INFO score 0.93) was found to be genome-wide significant ($p=6.2 \times 10^{-9}$, Table 1). These signals disappeared when we ran a sensitivity test using only cases with aortic valve disease under 50 years of age (results not shown). The overall quantile–quantile p distribution plot showed no evidence of inflation (Supplemental Fig. 1). No genome-wide significant associations were found for conotruncal defects. As a sensitivity

analysis we performed a GWAS for general CHD but removing patients with Atrial Septal Defects (ASD) (ICD-10 codes: Q21.10, Q21.11), and lost all significant signals.

Conditional analysis for rs2316327 and rs75230966

To assess if rs2316327 and rs75230966 represent independent risk variants, conditional analysis were conducted. The conditional analyses resulted in a

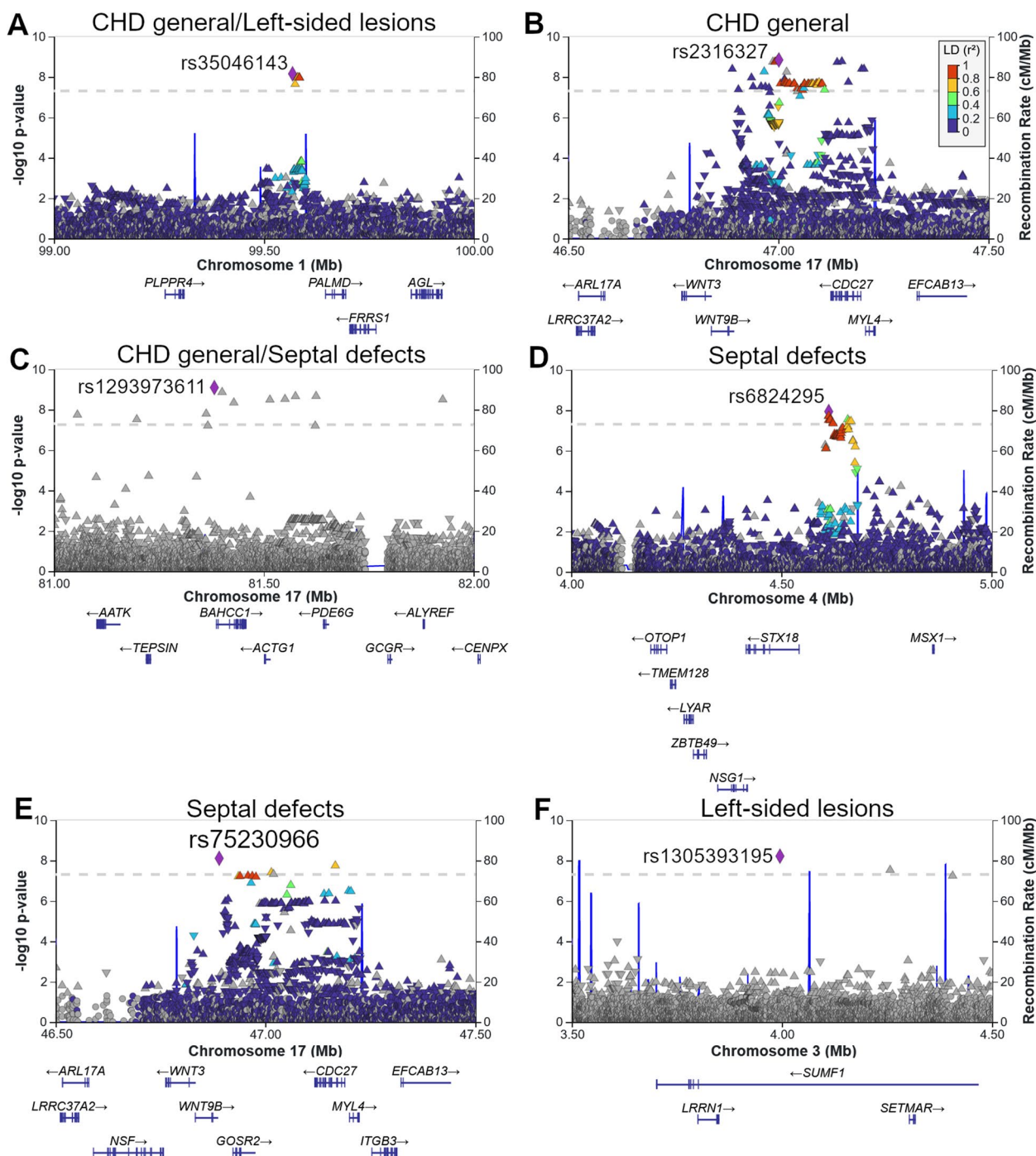


Fig. 2 The lead SNPs are located on chromosomes 1, 3, 4 and 17. LocusZoom plots of **A** rs35046143 (general CHD and left-sided lesions), **B** rs2316327 (general CHD), **C** rs1293973611 (general CHD and septal defects), **D** rs6824295 (septal defects), **E** rs75230966 (septal defects), **F** rs1305393195 (left-sided lesions) and their nearby genes and SNPs. The LD indicated by colors in the plots was computed using the 1000 genomes project as reference panel [22]

p -value=0.02 for rs2316327 for CHD general conditioned on rs75230966 as a covariate. Vice-versa, a p -value=0.007 for rs75230966 for Septal defects when conditioned on rs2316327.

CHD associated SNPs in 17q21.32 are associated with WNT9B and WNT3 expression

To understand the possible functional consequences of the identified lead SNPs on gene expression we computed

expression quantitative trait locus (eQTL) analysis focusing on heart tissues. We computed eQTLs for all lead SNPs with nearby genes ± 500 kbp from the SNP position, using the GTEx data and discovered a potential effect by rs2316327 and the nearby rs75230966 on the expression of *WNT9B* (eQTL normalized effect size (NES)=0.18 $p=2.8 \times 10^{-3}$ and NES=0.2 $p=1.2 \times 10^{-2}$, respectively), in atrial appendage tissue. A co-localization analysis using LocusFocus [23], which uses a 1Mbp window for finding the co-localization of a SNP signal and eQTL data from GTEx V8, further demonstrated significant association for rs2316327 with *WNT3* (Simple Sum (SS) $p=5.1 \times 10^{-3}$) and *WNT9B* (SS $p=3.9 \times 10^{-8}$) in atrial appendage tissue. For rs75230966 the co-localization analysis in atrial appendage tissue demonstrated significant association with *WNT3* (SS $p=5.1 \times 10^{-3}$) and *WNT9B* (SS $p=1.3 \times 10^{-6}$). There were no significant GTEx and LocusFocus results for the other lead SNPs in this study.

Meta-analysis with UKB supports association of 17q21.32 locus with CHD

We then examined the association of the lead SNPs with cardiac phenotypes in the UK Biobank data. While the lead variants did not associate in the UKB data, a meta-analysis of both FinnGen and UKB (total $N=881,678$) for general CHD discovered an association at the same 17q21 locus with rs2316327 but a different lead variant rs16941382 (chromosome 17, bp 46,966,142, intronic in *LRRC37A2*) as genome-wide significant lead SNP (meta-analysis OR[95%CI]=1.15[1.11–1.20], $p=1.5 \times 10^{-9}$, Table 2). The variant rs16941382 demonstrates significant association with CHD in the FinnGen data (OR[95%CI]=1.17[1.11–1.24], $p=5.5 \times 10^{-9}$) and is in significant LD with rs2316327 ($r^2=0.67$), thus likely reflecting the same signal. Furthermore, eQTL analysis and colocalization showed an association with *MYL4* expression: NES 0.1, $p=2.3 \times 10^{-3}$, and *WNT9B*: NES 0.16, $p=1.3 \times 10^{-2}$, both in atrial appendage in agreement with the FinnGen only results.

Phenome wide association testing in FinnGen suggest rs2316327, rs75230966, and rs35046143 are associated with a larger spectrum of cardiovascular endpoints

The FinnGen dataset includes longitudinal register data recording phenotypes and health events (including ICD-based diagnosis) of the study subjects obtained from Finnish national registers including data on all births and hospital care in Finland. Specifically, the Medical Birth Register includes data of all births in Finland since 1987 and the Care Register for Health Care (formerly Hospital Discharge Register) includes diagnosis and other data parameters of all hospital care in Finland since 1969. These registers are maintained by the Finnish Institute for Health and Welfare (THL) and have been shown to be reliable [24, 25]. This enables Phenome Wide Association Testing (PheWAS) to examine a single significant variant horizontally across other diseases and precomputed phenotypes to understand which other conditions are associated with the variant of interest. In Supplemental Fig. 2 we demonstrate the overall significant associations of the lead SNPs rs2316327 and rs75230966 with other binary cardiovascular endpoints in FinnGen similarly made using REGENIE. These endpoints include hypertension (cases=122,996 controls=289,117, $p=9.5 \times 10^{-9}$ and 3.9×10^{-18} respectively), use of antihypertensive medication (cases=222,561 controls=189,620, $p=8.3 \times 10^{-14}$ and 9.9×10^{-21} respectively), atrial fibrillation and flutter (cases=50,743 controls=210,652, $p=1.1 \times 10^{-9}$ and 2.3×10^{-17} respectively) and cardiac arrhythmias (cases=74,000 controls=239,778, $p=4.9 \times 10^{-9}$ and 0.5×10^{-11} respectively). The SNP rs35046143 that associated with left-sided lesions associated also with non-rheumatic valve disease (cases=22,653 controls=312,154, $p=3.7 \times 10^{-24}$), valvular heart disease excluding rheumatic fever (cases=22,818 controls=312,154, $p=7.2 \times 10^{-24}$), and operated calcific aortic valvular stenosis (cases=9,870 controls=402,311, $p=1.7 \times 10^{-26}$).

Table 2 Lead SNPs across cohorts. List of the general CHD lead SNP across the cohorts used in this study. NA signifies not available (SNP not present in summary statistics). OR [95% CI] signifies Odds ratio [95% confidence interval]

SNP	FinnGen		UKB		Meta-analysis	
	OR [95% CI]	p	OR[95%CI]	p	OR[95%CI]	p
rs35046143	1.15[1.10–1.21]	7.1×10^{-9}	NA	NA	NA	NA
rs2316327	1.17[1.12–1.23]	1.5×10^{-9}	1.08[1.00–1.17]	6.0×10^{-2}	1.14[1.09–1.20]	2.0×10^{-9}
rs1293973611	4.48[2.80–7.17]	7.0×10^{-10}	NA	NA	NA	NA
rs16941382	1.17[1.11–1.24]	5.5×10^{-9}	1.11[1.02–1.19]	2.1×10^{-2}	1.15[1.11–1.20]	1.5×10^{-9}

Discussion

Here we report the results of a GWAS of CHD traits in the FinnGen R10 database, where we detected three main regions associated with CHDs. Our aim was to elucidate the genetics behind CHD in general, but also explore specific sub-categories; conotruncal defects, septal defects and left-sided lesions. We discovered two genetic loci with the general CHD category GWAS in chromosome 17 (bands 17q21.32 and 17q25.3) and one in chromosome 1 (1p21.2). In addition, using GTEx we show that the loci are related to *WNT9B* and *WNT3* expression levels in the heart. Overall, our findings elucidate the genetic etiology behind CHD complementing the growing body of literature of these relatively poorly understood severe diseases. We focus in this study on the lead variants for associations, to avoid reporting potential false positives. Of the lead variants described, rs1305393195 had the lowest MAF at 2.7×10^{-4} , along with rs1293973611 having a MAF of 9.3×10^{-4} . This means these two SNPs should have a population-wide genotype count of over 106, but both are rare variants specific to the Finnish population. As neither of these two SNPs demonstrated any significant eQTLs nor neighboring SNPs with LD r^2 above 0.71, their association to their respective CHDs remains unclear. We did not discover any genome-wide significant lead SNPs for conotruncal defects.

The loci with the lead SNPs in chromosome 17 was also detected as significant in an analysis of the septal defect subcategory, while SNPs in the band 1p21.2 were found to significantly associate with left-sided lesions. The region in chromosome 17 (band 17q21.32), where the SNP rs2316327 associated with general CHD and rs75230966 and associated with septal defects are located, contains several genes linked to heart development, CHD, and other cardiac phenotypes; *MYL4*, *CDC27*, *GOSR2* and *WNT9B* [18, 26–28]. A recent GWAS study on patients with anomalies of thoracic arteries and veins identified three significant SNPs within the *GOSR2* locus in chromosome 17 close to rs11570508 [18]. Furthermore, the rs17608766 variant in *GOSR2* has been associated with a reduced aortic valve area [29] indicating importance of this locus in cardiac development. The same study indicated that the risk variants in the 17q21.32 influence the expression of *WNT3*, which is in accordance with our findings, and demonstrated that *WNT3* is expressed in cardiac progenitor cells [18].

According to eQTL and co-localization follow-up analyses of the genome-wide significant SNPs from FinnGen and a meta-analysis together with UKB (rs2316327, rs75230966 and rs16941382), we detected effects on primarily *WNT9B*. The SNP rs2316327 demonstrated an eQTL NES=0.18, $p=2.8 \times 10^{-3}$ for *WNT9B*. Wnt signaling is required for second heart field development, and

Wnt9b has shown to have an important role in endocardial endothelial cushion development in the developing heart in mice [26, 30]. Animal studies modeling cardiac development in mouse and zebrafish show that *Wnt9b* is expressed specifically in endocardial endothelial cells overlying the developing atrioventricular canal and outflow tract [31], and it has been suggested that congenital heart defects may arise due to defects in the Klf2-Wnt9b mechanotransduction pathway [26]. Thus, these data could suggest a pathogenic mechanism for our finding. Overall, the positive NES eQTL data of rs2316327, rs75230966 and rs16941382 on the expression of the *WNT9B* and significant co-localization results for these, suggests a positive effect of the gene on development of CHD in heart tissue. However, this connection will need to be tested experimentally. It is important to note that none of the eQTL multiple test corrections yielded significance.

The GTEx analysis of the SNP rs16941382 identified in the meta-analysis of FinnGen and UKB indicated significance for *MYL4* expression. *MYL4* encodes the atrial light chain-1 (ALC-1) protein, which is expressed in the human heart during development [32]. After birth, the expression decreases in the ventricles but remains in the atria [33]. ALC-1 has a role in sarcomere assembly and fine tuning of cardiac contractility. The expression of *MYL4* has shown to abnormally persist in ventricular tissue of individuals with CHD [34], and the ventricular analog *MYL3* is replaced by re-expression of *MYL4* in failing and hypertrophied hearts, resembling a fetal remodeling pattern associated with ventricular dysfunction [33, 35]. *MYL4* loss-of-function variants have been associated with atrial cardiomyopathy and atrial arrhythmias such as atrial fibrillation (AF) [27, 36–38], demonstrating that although previously not associated with CHD directly, its role in cardiac development and function is important, and thus, could be a plausible candidate gene for CHD. Furthermore, we used the lead SNP positions in the Human Heart Genome Browser to search the ATACseq data for peaks indicating regulatory motifs [39], but no significant peaks were found.

Interestingly, in addition to CHD, the SNPs rs2316327 and rs75230966 were associated with hypertension, use of antihypertensive medications, atrial fibrillation and flutter, and other arrhythmias. Children and adults with CHD have increased risk for atrial fibrillation and flutter [40, 41], and adults with CHD have increased cardiovascular morbidity relative to the general population [42]. Although anatomical factors such as anomalous vessel anatomy and abnormal atrial hemodynamics, and disease related conditions in the heart such as progressive valvulopathy, residual shunts, and atrial scars from previous heart surgery are likely to be important predisposes

for atrial arrhythmias and other cardiovascular morbidity in CHD, it is tempting to speculate that shared genetic factors could potentially increase the risk for both developing CHD and later cardiovascular morbidity. In fact, atrial arrhythmias are a common phenotype in Holt-Oram syndrome, a developmental disorder that leads to heart and limb malformations. Holt-Oram syndrome is partially caused by variants in the *TBX5* gene [43, 44], and GWAS studies on AF have shown several significant loci with nearest genes known to be important for heart development and causal for CHD, such as *NKX2-5*, *GATA4*, *GJA1*, and *GJA5* [11, 45]. It has been also shown that women whose infants have congenital heart defects have an increased risk and earlier disease onset for cardiovascular morbidity, including atherosclerotic cardiovascular disease, ischemic heart disease, hospitalization due to cardiac diseases and cardiac transplantation [46]. Thus, these associations could indicate shared genetic risks.

The lead SNP findings did not replicate in the UKB data, but a nearby SNP showed genome-wide significance in the meta-analysis. This could be due to differences in the number and severity of included CHD cases between the two CHD cohorts as depicted in Supplemental Table 1. In addition, it is possible that the definitions varied somewhat, as different coding systems were used. The FinnGen samples have been gathered from Finnish biobanks including samples collected from hospitalized patients of all ages, whereas UKB samples are in average from older individuals. Indeed, there was a higher percentage of affected individuals as well as more severe phenotypes among the FinnGen individuals. Another explanation could be the different genetic backgrounds of these two cohorts.

Our approach has some limitations. The discovery of novel loci in GWAS studies does not prove functional causality, and validation studies should be conducted. In addition, ICD10-based studies are limited by the fact that mild CHD such as BAV and hemodynamically insignificant ASD are often underdiagnosed as they may be asymptomatic. However, removing the ASD patients (1215 individuals) from our study diminished the significant association signals, which is understandable since it meant a significant loss of power (a third) from the cases. Additionally, the conditional analyses identified the general CHD loci as independent from the septal defect loci. The loci on chromosome 1 was not identified as significant in the septal defect GWAS, further demonstrating that the results of the general CHD GWAS were not entirely dependent on septal defects such as ASD. As the region includes several genes linked to heart development, and SNPs in this region has been associated with anomalies of thoracic arteries and veins [18] and

reduced aortic valve area [29], it is likely that this locus has importance for cardiac development in general. Previous studies report that approximately 0.8–1.0 percent of the population are affected by CHD [47, 48], and of the FinnGen R10 individuals, 0.8% have at least one CHD diagnosis. While it is possible that some controls in FinnGen are affected with CHD, the registers where the diagnoses were acquired from have been validated and are concluded to be reliable [24], thus the case status can be anticipated to be defined correctly. Finally, our association results are from the isolated Finnish population, which is genetically unique. On the other hand, our study highlights the potential of FinnGen to identify disease associated high-impact variants that are very rare or absent in other populations.

Conclusions

In summary, our analysis of different CHD categories primarily using FinnGen identifies a complex risk locus for CHD on chromosome 17 for CHD in general and septal defects. This locus is associated with *WNT3*, *WNT9B* and *MYL4*. However, further experimental and clinical data is needed to determine the impact and roles of these genes on CHD onset. Additional data would be needed to establish the independencies and impact of the identified lead SNPs rs2316327 and rs75230966 on the onset of CHDs. These may be two independent contributing signals within a locus, or neither capturing the true loci, as the conditional analyses identified both as independent signals but both affecting the significance of each other. The associations between the identified loci and other cardiovascular conditions, such as AF and flutter and hypertension may suggest shared genetic etiology in developmental defects and adult cardiovascular morbidity.

Methods

Study cohorts and phenotypes

FinnGen

FinnGen is a large-scale research project that aims to genotype 500,000 Finnish participants recruited from hospitals as well as prospective and retrospective epidemiological and disease-based cohorts. The participants are of Finnish ancestry. These data are combined with longitudinal registries that record phenotypes and health events (including ICD-based diagnosis) over the entire lifespan including the Care Register for Health Care (inpatient and outpatient), Causes of Death Registry, the National Infectious Diseases Registry, Cancer Registry, Primary Health Care Registry (outpatient) and Medication Reimbursement Registry. This study used data from FinnGen R10, which consists of >390,000 individuals (see next paragraph for specific numbers in

each CHD category that was analyzed). FinnGen utilizes custom ThermoFisher Axiom microarrays for capturing variants, and individuals are then imputed against a whole-genome sequenced Finnish reference panel [49] using IMPUTE2 version 2.3 [50]. Participants without Finnish ancestry based on principal component analysis (PCA) against 1000 genomes reference samples have been removed from the FinnGen study [49]. Approximately 73.6% of participants have a 3rd-degree relative or closer among the cohort. Furthermore, genetic duplicates, monozygotic twins (one of the pair kept) and other genetic outliers have been removed. At the variant level, FinnGen QC removes variants with high missingness (>2%), low Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) and minor allele count < 3. For determining ancestry, the FinnGen PCA analysis was performed on unrelated individuals within FinnGen (relationship is more than a 3rd degree relative) includes the following variant filtering: exclusion of variants with info score < 0.95, exclusion of variants with missingness > 0.02 (based on the GP; see conversion), exclusion of variants with MAF < 0.05 and LD pruning with window interval of 500 kb and 50 kb steps, and an r^2 filter of 0.1 (<https://finngen.gitbook.io/documentation/methods/phewas/quality-checks>) [49].

CHD subjects in FinnGen

In the FinnGen research project, the CHD categories used for this study consisted of general CHD (3506 cases, 392,436 controls), conotruncal defects (404 cases, 392,942), left-sided lesions (LVOTO and early aortic valve disease diagnosed under 50 years of age; 2382 cases, 392,503 controls) and septal defects (1955 cases, 392,428 controls) using FinnGen R10. General CHD and the CHD subcategories were defined according to ICD-10 codes as depicted in Supplemental Table 1. We used 50 years of age for early aortic valve disease as this is considered to be primarily associated with congenital heart defects, not deterioration over age [51]. We code variants in the X chromosome as 0,1,2 in females and 0,2 in males to account for the difference in number of X chromosomes in females and males.

From the FinnGen population we identified patients with a CHD diagnosis and from those excluded individuals with a diagnosis for syndromes that are known to include CHD phenotypes.

REGENIE GWAS analysis

The GWAS for CHD was performed using REGENIE on the FinnGen R10 cohort using Finnish participants where over 70% had at least one 3rd degree relative or closer within the cohort [49, 52]. We utilized REGENIE to compute association statistics. To account for population stratification and relatedness, Regenie directly

estimates the polygenic effects parameter β by using ridge regression, which corresponds to fitting a linear regression model with a L2 penalty to impose shrinkage [52]. The covariates used for the analysis included: sex, imputed age, age at death or end of follow up, 10 principal components and microarray genotyping batch. The default Regenie filtering of a minimum allele count > 5 was used, along with the firth regression option. A total of 16,960,158 SNPs across the autosomal and X chromosomes were included in the analyses.

We performed a validation GWAS of general CHD using UKB (cases = 2165, controls = 484 865). As UKB uses a slightly different ICD10 coding structure, we had to modify the codes used for defining general CHD, as well as utilize the General Practitioner Read Codes V2 and V3 (Supplemental Table 1). We performed the GWAS using the REGENIE software [52], correcting for sex, age, PC1-10, array used for SNP detection and assessment center. Additionally, we performed a GWAS meta-analysis using METAL (sample size based analysis) [53], combining the two GWAS datasets for a combined $N = 881,678$ (Table 3) and only considering SNPs that occur in both studies.

Calculating eQTL data and co-localization

To test whether the genome-wide significant SNPs from the FinnGen GWAS could have an effect on the expression of nearby genes, we utilized the eQTL calculator function of the GTEx portal database (www.gtexportal.org) on all lead SNPs in this study [54, 55]. We examined expression in the following tissues: ‘heart—atrial appendage’ and ‘heart – left ventricle’. To further test the GTEx result for the SNPs demonstrating significant ($p < 0.05$) eQTL, we used LocusFocus [23] to calculate Simple Sum colocalization p for the lead SNPs in heart tissues using GTEx V8.

Defining endpoint associations using PheWeb

In addition to CHD, FinnGen contains diagnosis codes and precomputed phenotypes. Therefore, instead of looking at each disease or GWAS alone, it is possible to examine a single significant variant horizontally across other diseases to understand which other diseases are

Table 3 Cohorts used for GWAS in this study

Cohort	Cases	Controls
FinnGen – General CHD	3506	392,436
UKB – General CHD	2156	484,865
Meta-analysis – General CHD	5662	877,301
FinnGen – conotruncal defects	404	392,942
FinnGen – left-sided lesions	2382	392,503
FinnGen – septal defects	1955	392,428

associated with this variant. The FinnGen results website uses the PheWeb design interface for displaying SNP-endpoint associations for a convenient overview of GWAS results [56]. This approach was used to identify other endpoints associated with the lead SNPs. The FinnGen R10 data will become available to the public similarly as previous releases.

Conditional analysis

We used the GCTA COJO software v1.94 [57] to detect independent signal selection (`-cojo-slc`) for each dataset (CHD general and Septal defects), where the respective lead SNPs were identified as independent signals. A conditional analysis (`-cojo-cond`) where we tested for both CHD general and Septal defects removed the respective lead SNP from the other.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10172-x>.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements

We would like to thank the FinnGen consortium for GWAS summary statistics. We want to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auriia Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biopankki), Helsinki Biobank (www.helsinki.fi/nginbiopankki.fi), Biobank Borealis of Northern Finland (<https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biopankki-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitointa) and Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmrifi.fi). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB.

FinnGen ethics statement

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected before the Finnish Biobank Act came into effect (in September 2013) and before start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the

Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

The FinnGen study is approved by the Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017–3, VRK/6909/2018–3, VRK/4415/2019–3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020, THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020 and Statistics Finland (permit numbers: TK-53–1041-17 and TK/143/07.03.00/2020 (earlier TK-53–90-20)).

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 7 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auriia Biobank AB17-5154 and amendment #1 (August 17 2020), Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1–2017, and Terveystalo Biobank STB 2018001.

Authors' contributions

MB, EH, HMO and JP designed the study; interpretation of data/results. MB performed GWAS, statistical analyses. MA, SJ, TO, OR and RK provided datasets for analysis, background data. All authors have read and contributed to writing the manuscript.

Funding

Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). This work was funded by grants from the Finnish Cultural Foundation (EH, TO and MB), Finnish Foundation for Pediatric Research (EH, TO and MB), the Finnish Medical Foundation (EH), Sigrid Juselius Foundation (EH), Helsinki University (EH), and The Academy of Finland (EH).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study includes data from FinnGen which has been approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Stem Cells and Metabolism Research Program, Faculty of Medicine, University of Helsinki, 00014 Helsinki, Finland. ²Institute for Molecular Medicine Finland (FIMM), HILIFE, University of Helsinki, 00014 Helsinki, Finland. ³Department of Pediatric Cardiology, New Children's Hospital, Pediatric Research Center, Helsinki University Hospital, 00029 Helsinki, Finland. ⁴Wihuri Research Institute, 00290 Helsinki, Finland. ⁵Faculty of Sport and Health Sciences, University of Jyväskylä, 40014 Jyväskylä, Finland. ⁶School of Medicine, Stanford University, Stanford University, Stanford, CA 94305, USA. ⁷Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. ⁸Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA. ⁹Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA.

¹⁰Haartmaninkatu 8, Helsinki 00014, Finland. ¹¹Population Health Unit, Finnish Institute for Health and Welfare, Helsinki 00271, Finland.

Received: 4 August 2023 Accepted: 28 February 2024

Published online: 07 March 2024

References

- Liu Y, Chen S, Zühlke L, Black GC, Choy MK, Li N, et al. Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. *Int J Epidemiol*. 2019;48(2):455–63.
- Huisenga D, La Bastide-Van GS, Van Bergen A, Sweeney J, Hadders-Algra M. Developmental outcomes after early surgery for complex congenital heart disease: a systematic review and meta-analysis. *Dev Med Child Neurol*. 2021;63(1):29–46.
- Kerstjens-Frederikse WS, van de Laar IMBH, Vos YJ, Verhagen JMA, Berger RMF, Lichtenbelt KD, et al. Cardiovascular malformations caused by NOTCH1 mutations do not keep left: data on 428 probands with left-sided CHD and their families. *Genet Med*. 2016;18(9):914–23.
- Helle E, Priest JR. Maternal Obesity and Diabetes Mellitus as Risk Factors for Congenital Heart Disease in the Offspring. *J Am Heart Assoc*. 2020;9(8):e011541.
- Øyen N, Diaz LJ, Leirgul E, Boyd HA, Priest J, Mathiesen ER, et al. Prepregnancy Diabetes and Offspring Risk of Congenital Heart Disease: A Nationwide Cohort Study. *Circulation*. 2016;133(23):2243–53.
- Patomo E, Huybrechts KF, Bateman BT, Cohen JM, Desai RJ, Mogun H, et al. Lithium Use in Pregnancy and the Risk of Cardiac Malformations. *N Engl J Med*. 2017;376(23):2245–54.
- Persson M, Razaz N, Edstedt Bonamy AK, Villamor E, Cnattingius S. Maternal Overweight and Obesity and Risk of Congenital Heart Defects. *J Am Coll Cardiol*. 2019;73(1):44–53.
- Garg V. Insights into the genetic basis of congenital heart disease. *Cell Mol Life Sci*. 2006;63(10):1141–8.
- Helle E, Córdova-Palomera A, Ojala T, Saha P, Potiny P, Gustafsson S, et al. Loss of function, missense, and intronic variants in *NOTCH1* confer different risks for left ventricular outflow tract obstructive heart defects in two European cohorts. *Genet Epidemiol*. 2019;43(2):215–26.
- Reuter MS, Jobling R, Chaturvedi RR, Manshaei R, Costain G, Heung T, et al. Haploinsufficiency of vascular endothelial growth factor related signaling genes is associated with tetralogy of Fallot. *Genet Med*. 2019;21(4):1001–7.
- Roselli C, Rienstra M, Ellinor PT. Genetics of Atrial Fibrillation in 2020: GWAS, Genome Sequencing, Polygenic Risk, and Beyond. *Circ Res*. 2020;127(1):21–33.
- Elliott DA, Kirk EP, Yeoh T, Chandar S, McKenzie F, Taylor P, et al. Cardiac homeobox gene *NKX2-5* mutations and congenital heart disease. *J Am Coll Cardiol*. 2003;41(11):2072–6.
- Hirayama-Yamada K, Kamisago M, Akimoto K, Aotsuka H, Nakamura Y, Tomita H, et al. Phenotypes with *GATA4* or *NKX2.5* mutations in familial atrial septal defect. *Am J Med Genet A*. 2005;135A(1):47–52.
- Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet*. 2017;49(11):1593–601.
- Sifrim A, the INTERVAL Study, the UK10K Consortium, the Deciphering Developmental Disorders Study, Hitz MP, Wilsdon A, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet*. 2016;48(9):1060–5.
- Broberg M, Hästbacka J, Helle E. From Stem Cells to Populations—Using hiPSC, Next-Generation Sequencing, and GWAS to Explore the Genetic and Molecular Mechanisms of Congenital Heart Defects. *Genes*. 2021;12(6):921.
- Morton SU, Quiat D, Seidman JG, Seidman CE. Genomic frontiers in congenital heart disease. *Nat Rev Cardiol*. 2022;19(1):26–42.
- Lahm H, Jia M, Dreßen M, Wirth F, Puluca N, Gilsbach R, et al. Congenital heart disease risk loci identified by genome-wide association study in European patients. *J Clin Invest*. 2021;131(2):e141837.
- Cordell HJ, Topf A, Mamasoula C, Postma AV, Bentham J, Zelenika D, et al. Genome-wide association study identifies loci on 12q24 and 13q32 associated with Tetralogy of Fallot. *Hum Mol Genet*. 2013;22(7):1473–81.
- Agopian AJ, Goldmuntz E, Hakonarson H, Sewda A, Taylor D, Mitchell LE. Genome-Wide Association Studies and Meta-Analyses for Congenital Heart Defects. *Circ Cardiovasc Genet*. 2017;10(3):e001449.
- Björnsson T, Thorólfsson RB, Sveinbjörnsson G, Sulem P, Norddahl GL, Helgadóttir A, et al. A rare missense mutation in *MYH6* associates with non-syndromic coarctation of the aorta. *Eur Heart J*. 2018;39(34):3243–9.
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336–7.
- Panjwani N, Wang F, Mastromatteo S, Bao A, Wang C, He G, et al. LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. *PLoS Comput Biol*. 2020;16(10):e1008336.
- Aro S, Koskinen R, Keskimäki I. Reliability of hospital discharge data concerning diagnosis, treatments and accidents. *Duodecim Laaketieteellinen Aikakauskirja*. 1990;106(21):1443–50.
- Gissler M, Teperi J, Hemminki E, Meriläinen J. Data Quality after Restructuring a National Medical Registry. *Scand J Soc Med*. 1995;23(1):75–80.
- Goddard LM, Duchemin AL, Ramalingam H, Wu B, Chen M, Bamezai S, et al. Hemodynamic Forces Sculpt Developing Heart Valves through a *KLF2-WNT9B* Paracrine Signaling Axis. *Dev Cell*. 2017;43(3):274–289.e5.
- Orr Y, Leclair K, Jacobs S, Badawi N, Nicholson IA, Chard RB, et al. Early outcomes from a new regional programme for the surgical management of hypoplastic left heart syndrome: Outcomes after surgery for HLHS. *ANZ J Surg*. 2015;85(6):466–71.
- Sabour D, Machado RSR, Pinto JP, Rohani S, Sahito RGA, Hescheler J, et al. Parallel Genome-wide Profiling of Coding and Non-coding RNAs to Identify Novel Regulatory Elements in Embryonic and Maturated Heart. *Mol Ther - Nucleic Acids*. 2018;12:158–73.
- Córdova-Palomera A, Tcheandjieu C, Fries JA, Varma P, Chen VS, Fiterau M, et al. Cardiac Imaging of Aortic Valve Area From 34 287 UK Biobank Participants Reveals Novel Genetic Associations and Shared Genetic Comorbidity With Multiple Disease Phenotypes. *Circ Genomic Precis Med*. 2020;13(6):e003014.
- Klaus A, Saga Y, Taketo MM, Tzahor E, Birchmeier W. Distinct roles of *Wnt/β-catenin* and *Bmp* signaling during early cardiogenesis. *Proc Natl Acad Sci*. 2007;104(47):18531–6.
- Alfieri CM, Cheek J, Chakraborty S, Yutzey KE. *Wnt* signaling in heart valve development and osteogenic gene induction. *Dev Biol*. 2010;338(2):127–35.
- Barton PJR, Buckingham ME. The myosin alkali light chain proteins and their genes. *Biochem J*. 1985;231(2):249–61.
- Wang TY, Lee D, Fox-Talbot K, Arking DE, Chakravarti A, Halushka MK. Cardiomyocytes have mosaic patterns of protein expression. *Cardiovasc Pathol*. 2018;34:50–7.
- England J, Loughna S. Heavy and light roles: myosin in the morphogenesis of the heart. *Cell Mol Life Sci*. 2013;30(7):1221–39.
- Wang TY, Arking DE, Maleszewski JJ, Fox-Talbot K, Nieuwenhuis TO, Santhanam L, et al. Human cardiac myosin light chain 4 (*MYL4*) mosaic expression patterns vary by sex. *Sci Rep*. 2019;9(1):12681.
- Gudbjartsson DF, Holm H, Sulem P, Masson G, Oddsson A, Magnusson OTH, et al. A frameshift deletion in the sarcomere gene *MYL4* causes early-onset familial atrial fibrillation. *Eur Heart J*. 2017;38(1):27–34.
- Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddsson A, Gylfason A, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*. 2015;47(5):435–44.
- Peng G, Barro-Soria R, Sampson KJ, Larsson HP, Kass RS. Gating mechanisms underlying deactivation slowing by two *KCNQ1* atrial fibrillation mutations. *Sci Rep*. 2017;7(1):45911.
- Hocker JD, Poirion OB, Zhu F, Buchanan J, Zhang K, Chiou J, et al. Cardiac cell type-specific gene regulatory programs and disease risk association. *Sci Adv*. 2021;7(20):eabf1444.
- Labombarda F, Hamilton R, Shohoudi A, Aboulhosn J, Broberg CS, Chaix MA, et al. Increasing Prevalence of Atrial Fibrillation and Permanent Atrial Arrhythmias in Congenital Heart Disease. *J Am Coll Cardiol*. 2017;70(7):857–65.
- Mandalenakis Z, Rosengren A, Lappas G, Eriksson P, Gilljam T, Hansson PO, et al. Atrial Fibrillation Burden in Young Patients With Congenital Heart Disease. *Circulation*. 2018;137(9):928–37.
- Saha P, Potiny P, Rigdon J, Morello M, Tcheandjieu C, Romfh A, et al. Substantial Cardiovascular Morbidity in Adults With Lower-Complexity Congenital Heart Disease. *Circulation*. 2019;139(16):1889–99.

43. Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soultis J, et al. Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet.* 1997;15(1):30–5.
44. Vanlerberghe C, Jourdain AS, Ghomid J, Frenois F, Mezel A, Vaksman G, et al. Holt-Oram syndrome: clinical and molecular description of 78 patients with TBX5 variants. *Eur J Hum Genet.* 2019;27(3):360–8.
45. Pierpont ME, Brueckner M, Chung WK, Garg V, Lacro RV, McGuire AL, et al. Genetic Basis for Congenital Heart Disease: Revisited: A Scientific Statement From the American Heart Association. *Circulation.* 2018;138(21):e653–711.
46. Auger N, Potter BJ, Bilodeau-Bertrand M, Paradis G. Long-Term Risk of Cardiovascular Disease in Women Who Have Had Infants With Heart Defects. *Circulation.* 2018;137(22):2321–31.
47. Hoffman JIE, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol.* 2002;39(12):1890–900.
48. Reller MD, Strickland MJ, Riehle-Colarusso T, Mahle WT, Correa A. Prevalence of Congenital Heart Defects in Metropolitan Atlanta, 1998–2005. *J Pediatr.* 2008;153(6):807–13.
49. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature.* 2023;613(7944):508–18.
50. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955–9.
51. Roberts WC, Ko JM. Frequency by Decades of Unicuspid, Bicuspid, and Tricuspid Aortic Valves in Adults Having Isolated Aortic Valve Replacement for Aortic Stenosis, With or Without Associated Aortic Regurgitation. *Circulation.* 2005;111(7):920–5.
52. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet.* 2021;53(7):1097–103.
53. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190–1.
54. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
55. Stanfill AG, Cao X. Enhancing Research Through the Use of the Genotype-Tissue Expression (GTEx) Database. *Biol Res Nurs.* 2021;23(3):533–40.
56. Gagliano Taliun SA, VandeHaar P, Boughton AP, Welch RP, Taliun D, Schmidt EM, et al. Exploring and visualizing large-scale genetic associations by using PheWeb. *Nat Genet.* 2020;52(6):550–2.
57. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet.* 2011;88(1):76–82.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.