

ASSOCIATION STUDIES ARTICLE

A genome-wide association study of congenital cardiovascular left-sided lesions shows association with a locus on chromosome 20

Neil A. Hanchard^{1,2}, Shanker Swaminathan^{1,2}, Kristine Bucasas^{1,3}, Dieter Furthner⁶, Susan Fernbach¹, Mahshid S. Azamian¹, Xueqing Wang¹, Mark Lewin⁷, Jeffrey A. Towbin⁸, Lisa C.A. D'Alessandro⁴, Shaine A. Morris⁴, William Dreyer⁴, Susan Denfield⁴, Nancy A. Ayres⁴, Wayne J. Franklin⁴, Henri Justino⁴, M. Regina Lantin-Hermoso⁴, Elena C. Ocampo⁴, Alexia B. Santos⁴, Dhaval Parekh⁴, Douglas Moodie⁴, Aamir Jeewa⁴, Emily Lawrence⁴, Hugh D. Allen⁴, Daniel J. Penny⁴, Charles D. Fraser⁵, James R. Lupski^{1,2}, Mojisola Popoola², Lalita Wadhwa⁵, J. David Brook⁹, Frances A. Bu'Lock¹⁰, Shoumo Bhattacharya¹¹, Seema R. Lalani¹, Gloria A. Zender¹⁴, Sara M. Fitzgerald-Butt^{12,14,15}, Jessica Bowman^{12,15}, Don Corsmeier^{12,16}, Peter White^{12,16}, Kelsey Lecerf¹³, Gladys Zapata^{1,2}, Patricia Hernandez^{1,2}, Judith A. Goodship¹⁷, Vidu Garg^{12,14,15}, Bernard D. Keavney¹⁸, Suzanne M. Leal^{1,3}, Heather J. Cordell¹⁷, John W. Belmont^{1,2,†,*} and Kim L. McBride^{12,14,*}

¹Department of Molecular and Human Genetics, ²Department of Pediatrics, ³Center for Statistical Genetics, ⁴Division of Cardiology, Department of Pediatrics, and ⁵Department of Surgery, Baylor College of Medicine, Houston, TX, USA, ⁶Department of Paediatrics, Children's Hospital, Linz, Austria, ⁷Division of Cardiology, Seattle Children's Hospital, Seattle, WA, USA, ⁸Pediatric Cardiology, University of Tennessee Health Science Center, Memphis, TN, USA, ⁹School of Life Sciences, University of Nottingham, Nottingham, UK, ¹⁰East Midlands Congenital Heart Centre, Glenfield Hospital, Leicester, UK, ¹¹Radcliffe Department of Medicine, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, ¹²Department of Pediatrics and ¹³College of Medicine, Ohio State University, Columbus, OH, USA, ¹⁴Center for Cardiovascular and Pulmonary Research, ¹⁵The Heart Center, and ¹⁶Center for Microbial Pathogenesis, Nationwide Children's Hospital, Columbus, OH, USA, ¹⁷Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK and ¹⁸Institute of Cardiovascular Sciences, The University of Manchester, Manchester, UK

*To whom correspondence should be addressed. Tel: +1 6143555759; Fax: +1 6143555725; Email: kim.mcbride@nationwidechildrens.org or jbelmont@bcm.edu

[†]Present address: Illumina, Inc., San Diego, CA, USA.

Received: December 19, 2015. Revised and Accepted: February 26, 2016

© The Author 2016. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Abstract

Congenital heart defects involving left-sided lesions (LSLs) are relatively common birth defects with substantial morbidity and mortality. Previous studies have suggested a high heritability with a complex genetic architecture, such that only a few LSL loci have been identified. We performed a genome-wide case-control association study to address the role of common variants using a discovery cohort of 778 cases and 2756 controls. We identified a genome-wide significant association mapping to a 200 kb region on chromosome 20q11 [$P = 1.72 \times 10^{-8}$ for rs3746446; imputed Single Nucleotide Polymorphism (SNP) rs6088703 $P = 3.01 \times 10^{-9}$, odds ratio (OR) = 1.6 for both]. This result was supported by transmission disequilibrium analyses using a subset of 541 case families (lowest P in region = 4.51×10^{-5} , OR = 1.5). Replication in a cohort of 367 LSL cases and 5159 controls showed nominal association ($P = 0.03$ for rs3746446) resulting in $P = 9.49 \times 10^{-9}$ for rs3746446 upon meta-analysis of the combined cohorts. In addition, a group of seven SNPs on chromosome 1q21.3 met threshold for suggestive association (lowest $P = 9.35 \times 10^{-7}$ for rs12045807). Both regions include genes involved in cardiac development—*MYH7B/miR499A* on chromosome 20 and *CTSK*, *CTSS* and *ARNT* on chromosome 1. Genome-wide heritability analysis using case-control genotyped SNPs suggested that the mean heritability of LSLs attributable to common variants is moderately high (h_g^2 range = 0.26–0.34) and consistent with previous assertions. These results provide evidence for the role of common variation in LSLs, proffer new genes as potential biological candidates, and give further insight to the complex genetic architecture of congenital heart disease.

Introduction

Congenital heart disease (CHD) is one of the most common birth defects in the USA, and carries a large personal, familial and societal impact. With a birth prevalence of 5–8/1000 for clinically significant lesions, they account for one-quarter of all birth defects (1,2). Advances in the medical and surgical management of CHD have led to an increasing percentage of children who survive to adulthood (now ~85%), resulting in well over 1 million adults with CHD in the US alone (3,4). However, CHDs still cause substantial mortality and morbidity; they account for <4% of all pediatric hospitalizations, but represent over 15% of total costs (5). They are also a leading cause of infant mortality (6) and contribute adversely to long-term neurodevelopment (7,8). In addition, CHD is associated with shortened lifespan in adult survivors (9,10) and late complications of heart failure and arrhythmia are common (11).

CHD manifests along a wide spectrum of anatomical defects; this presents considerable difficulty investigating their etiologies. It is often not clear how to classify individuals, as the appropriate grouping for medical and surgical management may not be ideal for identification of causal mechanisms. In this study, we have used the classification adopted by the National Birth Defects Prevention Study in which lesions are classified by their presumed developmental mechanism(s) (12). The left-sided lesions (LSLs), which include aortic valve stenosis (AS), coarctation of the aorta (CoA), mitral valve stenosis, interrupted aortic arch type A (IAAA), hypoplastic left heart syndrome (HLHS) and Shone complex, are among the most common and most severe CHDs. Bicuspid aortic valve (BAV), which occurs in ~1% of the general population and is associated with late onset ascending aortic aneurysm, is more frequent in first-degree relatives of patients with severe LSLs and may, in some families, represent a *form fruste* of LSLs. LSLs are thought to arise as a consequence of altered flow in the embryonic cardiac outflow (13) or inflow (9) tract, but a lack of informative genetic models has hampered pathway and molecular analysis. Specific known causes of LSLs are diverse, and include environmental agents [e.g. maternal phenylketonuria (14)], chromosomal abnormalities [Turner and Jacobsen syndrome (MIM no. 147791)], genomic disorders caused by rare pathogenic copy number variants (CNVs) (15–18) and single-gene disorders [e.g. Kabuki syndrome (MIM no. 147920) and Holt–Oram syndrome (MIM no. 142900)]. In most of these patients, however, there are characteristic extracardiac defects or dysmorphisms, and they are thus classified as syndromic CHD.

Much less information is available concerning the etiologies of isolated or non-syndromic LSL.

Non-syndromic LSLs are generally considered to be complex diseases, with incidences varying by sex, maternal age, geography and ancestry (19). There is also indirect evidence for substantial, albeit complex, genetic effects; as many as 20% of individuals with LSLs have multiple affected family members, almost always with a concordant CHD (20–23) and epidemiological studies have demonstrated an increased risk in individuals with predominantly European ancestry (24). There is also a strong ‘familiality’ of LSL, with estimates of the relative risk to first-degree relatives ranging from 10 to 37 (25,26). These observations strongly suggest the action of ancestral risk-modifying variants that may have arisen in concert with ancient population subdivision and/or more recent ‘clan’ type variation (27) coupled with variable expression and incomplete penetrance. Indeed, complex segregation analysis has demonstrated high heritability of LSL, and a model of autosomal dominant inheritance with reduced penetrance is most consistent with the data from these uncommon families (25). While several loci have been identified by linkage analysis of multiplex families (23,28), only *NOTCH1* has been convincingly established as a disease gene (29,30).

Given the known complex genetic architecture of LSLs, we hypothesized that susceptibility loci could be identified by population based approaches. When this work was initiated, only a few small candidate gene-association studies (31–33) and no genome-wide association studies (GWAS) had been performed. Therefore, our goal was to identify loci associated with LSLs using both case-control and case-parent trio genome-wide association study GWAS designs. We report an association on chromosome 20 significant at a genome-wide level ($P < 5 \times 10^{-8}$), and an additional suggestive locus on chromosome 1.

Results

Study cohort

Final study subject numbers and characteristics are noted in Table 1; details of their recruitment are given in the ‘Materials and Methods’ section. There is a greater proportion of males than females in our cohort, consistent with the known epidemiology of LSLs (24). Our reference group consisted of control genotype data for 3034 individuals of reported European ancestry from two US studies downloaded through the database of Genotypes and Phenotypes (dbGaP—see the ‘Materials and Methods’ section

for details). After quality control (QC), the final case–control data set consisted of 3534 samples (778 cases, 2756 controls) and 534 461 single nucleotide polymorphisms (SNPs) passing QC thresholds. For family-based analyses, we included data from the extended families of our LSL cohort cases (as available) genotyped on the same platform. The final data set (after SNP-, individual- and family-level QC) included 1538 individuals in 541 families, including 372 affected-offspring trios.

Genome-wide association

Evaluation of population ancestry using multidimensional scaling (MDS) (see the Materials and Methods' section) demonstrated that although the vast majority of inferred cases and controls were of presumed Northern European ancestry, several cases were of apparent Hispanic ancestry (Fig. 1A), consistent with the predominant case recruitment from the South-Western USA. We, therefore, chose to restrict our primary analyses to a more homogenous ethnic sub-sampling (Fig. 1B) defined as being within 2 standard

Table 1. Study subject characteristics

Discovery case cohort		Family-based cohort	
Diagnoses	N	Individuals	N
CoA	226	Total	1538
HLHS	216	Affected	576
AS	233	Families	
Other	103	Trios	372
Gender		Duos	141
Male	534	Multiplex	28
Female	244	Sites	
Sites		TCH	260
TCH	428	NCH	236
NCH	305	CHA	45
CHA	45	Total families	541
Total cases	778		

CoA, coarctation of the aorta; HLHS, hypoplastic left heart syndrome; AS, aortic stenosis; Other, includes mixed lesions, lesions with associated bicuspid aortic valve, and unclassified LSLs; TCH, Texas Children's Hospital; NCH, Nationwide Children's Hospital; CHA, Children's Hospital Austria.

deviations (SDs) of the mean values for each of the first two MDS components—186 cases and 80 controls exceeded this threshold. A final total of 3268 samples (592 cases and 2676 controls) were used for discovery in our case–control association study.

We tested for association under an allelic disease model using logistic regression with covariates of sex and the first two MDS components. This analysis revealed a cluster of four associated SNPs on chromosome 20q11.22 with *P*-values below genome-wide significance (minimum $P = 1.72 \times 10^{-8}$, OR = 1.59 for rs3746446), with six other SNPs at the same locus having suggestive evidence of association $P < 1.0 \times 10^{-5}$ (Fig. 2A, Table 2, Supplementary Material, Table S1). We also observed a candidate locus on chromosome 1q21.3 consisting of seven SNPs that surpassed our threshold for suggestive association, but did not extend to genome-wide significance (Fig. 2A, Supplementary Material, Fig. S1, Supplementary Material, Table S1). The combined genomic inflation factor for association was 1.03 without evidence of systematic bias on the resulting quantile–quantile (Q–Q) plot (Fig. 2B). The same analyses performed using all discovery individuals (regardless of ancestry) supported our primary results (Supplementary Material, Fig. S2).

Further analysis of chromosome 20 locus

Transmission disequilibrium test (TDT) analysis of all available trios (case + both parents) and duos (case + one parent) from our cohort provided support for the chromosome 20 case–control results independent of population stratification. Several of the case–control associated SNPs showed evidence of over-transmission of the minor allele to affected offspring (Table 2; Supplementary Material, Table S2), with two other near-by SNPs showing evidence of more extensive over-transmission (minimum $P = 4.5 \times 10^{-5}$, maximum OR = 1.53; Supplementary Material, Table S2). There were no SNPs from the TDT analysis reaching genome-wide significance (minimum $P = 1.0 \times 10^{-6}$).

In order to further fine map our putative association, we imputed SNPs across chromosome 20 using the 1000 genomes Phase 3 reference panels (see the 'Materials and Methods' section), and then performed association testing using the same model parameters as our primary case–control analysis (first two MDS components and sex as covariates). This analysis

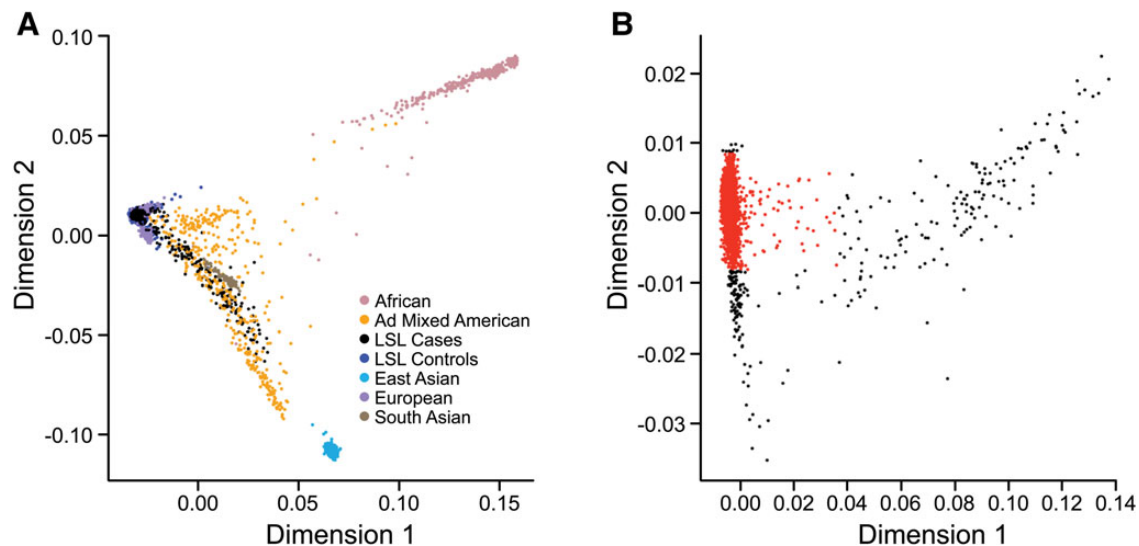


Figure 1. MDS plots of the first two principal components. (A) The LSL cohort in the context of global super populations from the 1000 genomes project. (B) The LSL cohort alone, with samples of close ancestry (within 2 SD of the first two MDS components) used for case–control association shown in red.

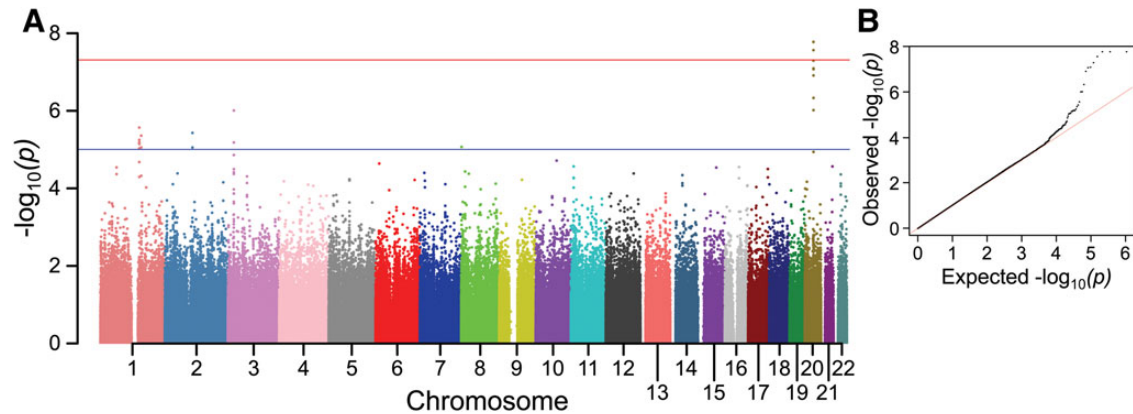


Figure 2. Case-control association results. Manhattan plot (A) and QQ plot (B) of genotyped SNPs from the primary case-control association are shown.

Table 2. SNP association at chromosome 20 locus

CHR	SNP	BP	Minor allele	State	MAF	Discovery cases (N = 789)	dbGaP controls (N = 2750)	Replication cases (N = 368)	WTCCC2 controls (N = 5159)	Association				
										Discovery OR	P	TDT families OR ^a	P ^a	Replication OR
20	rs6088662	33547633	G	Imputed	0.24	0.18	0.22	0.19	1.57	2.8×10^{-8}	–	–	1.21	0.032
20	rs6120777	33560172	A	Genotyped	0.21	0.15	0.18	0.16	1.59	5.2×10^{-8}	1.368	0.009	1.18	0.080
20	rs6088667	33566722	G	Genotyped	0.24	0.17	0.21	0.18	1.58	2.8×10^{-8}	1.357	0.008	1.23	0.033
20	rs3746446	33574765	G	Genotyped	0.24	0.17	0.21	0.18	1.59	1.7×10^{-8}	1.336	0.012	1.23	0.034
20	rs3746444	33578251	C	Genotyped	0.24	0.17	0.21	0.18	1.59	1.7×10^{-8}	1.322	0.016	1.23	0.036
20	rs6088678	33607551	T	Genotyped	0.24	0.17	0.21	0.18	1.59	1.7×10^{-8}	1.336	0.012	1.23	0.033
20	rs6120804	33623701	A	Imputed	0.24	0.17	0.21	0.18	1.59	1.2×10^{-8}	–	–	1.21	0.038
20	rs6088691	33633758	T	Imputed	0.24	0.17	0.21	0.18	1.60	7.8×10^{-9}	–	–	1.21	0.036
20	rs6088703	33676049	A	Imputed	0.24	0.17	0.21	0.18	1.63	3.0×10^{-9}	–	–	1.21	0.050
20	rs3746429	33703607	A	Genotyped	0.21	0.16	0.19	0.16	1.52	9.8×10^{-7}	1.231	0.075	1.21	0.043
20	rs6120849	33730387	T	Genotyped	0.27	0.20	0.24	0.21	1.52	1.3×10^{-7}	1.287	0.019	1.20	0.050
20	rs6060270	33736814	T	Genotyped	0.27	0.20	0.24	0.21	1.53	8.7×10^{-8}	1.287	0.019	1.19	0.041
20	rs6088732	33744134	T	Genotyped	0.27	0.20	0.24	0.21	1.53	8.2×10^{-8}	1.284	0.020	1.20	0.048
20	rs6088735	33745676	T	Imputed	0.27	0.20	0.24	0.21	1.53	7.1×10^{-8}	–	–	1.19	0.040
20	rs6060278	33753262	C	Imputed	0.27	0.20	0.24	0.21	1.52	1.3×10^{-7}	–	–	1.20	0.038
20	rs6060300	33780970	C	Genotyped	0.23	0.17	0.21	0.18	1.51	4.7×10^{-7}	1.287	0.022	1.15	0.133

CHR, chromosome; BP, base pair position (hg19); OR, odds ratio; TDT, transmission disequilibrium test.

^aCombined Tdt analysis from PLINK. 'State' refers to the discovery cohort; imputed variants in the replication cohort are italicized.

demonstrated a cluster of 36 associated SNPs with strong linkage disequilibrium (LD) extending across ~200 kb at our original chromosome 20 locus. This locus was anchored by an imputed SNP (rs6088703) with a P-value of 3.0×10^{-9} (Fig. 3; Table 2).

Next, we sought to replicate our association by interrogating existing GWA data from a case-control study of multiple CHD phenotypes undertaken in persons of European ancestry (34). A total of 367 CHD cases from that study also had LSLs. We, thus, looked for association with our top chromosome 20 SNPs using those 367 cases and the full set of 5170 controls used in the original publication (from the Wellcome Trust Case Control Consortium—WTCCC2). We observed a nominally statistically significant association ($P < 0.05$; Table 2) with several of the overlapping SNPs and a meta-analysis of the results from the two cohorts using METAL (35) was statistically significant (combined $P = 9.5 \times 10^{-9}$ for rs3746446).

Heritability analysis

Previous studies have estimated the heritability of LSL using family data only. The availability of genome-wide genotype

data allowed us to reassess the contribution of common variants to the genetic determinants of these malformations (36,37). We used genotyped SNPs from both cases and controls to estimate the heritability (h_g^2) of LSL on the liability scale. Using individuals of close European ancestry (1 SD of MDS components 1 and 2; $N = 2806$) case h_g^2 was 0.335 (SE 0.079) (Fig. 4A); using all available subjects regardless of ethnicity ($N = 3409$), case h_g^2 was 0.263 (SE 0.062) (Fig. 4B). Variance per chromosome using both the more restricted sample set (Fig. 4C) and the full data set (Fig. 4D) showed that chromosomes 2, 8, 1, 12 and 20 (in order) were responsible for the five highest variances observed.

Functional annotation

Annotation and potential functional consequences of associated SNPs on chromosomes 20 were explored visually in the University of California Santa Clara Genome Browser and investigated using Ensembl Variant Effect Predictor (38), RegulomeDB (39), GWAVA (40) and Genevar (41). We focused our analyses on eight genotyped and imputed SNPs associated with LSLs (rs6088662,

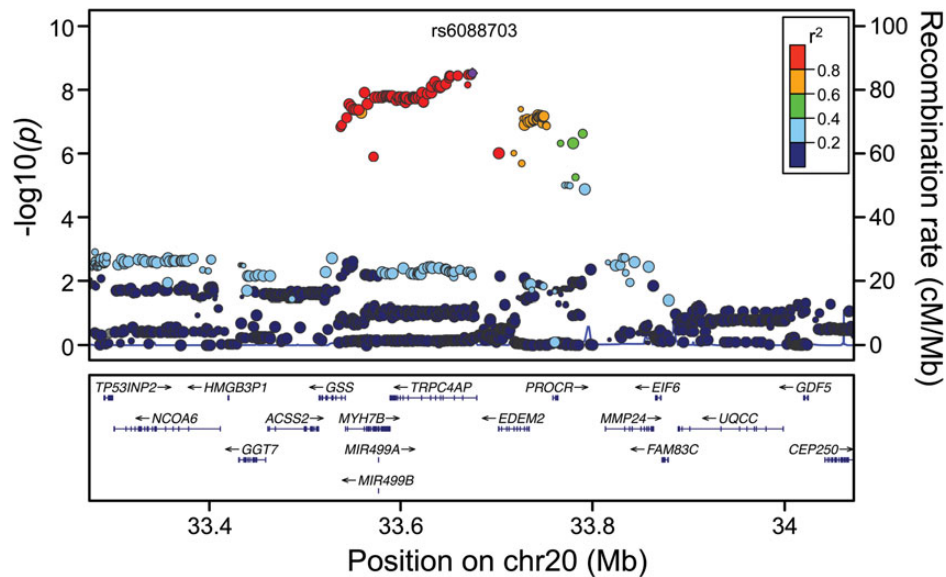


Figure 3. The LocusZoom plot of association at chromosome 20 locus. Both imputed and genotyped SNPs are shown.

rs6120777, rs6120804, rs6088691, rs3746429, rs6060270, rs6088735 and rs6060278) on chromosome 20.

Non-coding functional annotation of individual SNPs was analyzed in RegulomeDB. Evidence of likely binding effect and association with expression of a target gene was found for rs6088662 (score 1d) and rs3746429 (score 1f). GWAVA scores for the eight SNPs (Supplementary Material, Table S1) generally fell close to the median for GWAS SNPs that have been externally replicated (range, 0.03–0.61). The highest score was observed for rs3746429 (region score = 0.30, TSS score 0.26 and unmatched score 0.61), due to high-regional conservation (average Genomic Evolutionary Rate Profiling (GERP) score 3.34), histone modification and high-GC content (58.4%).

Haplotypes of SNPs were also annotated. Enhancer enrichment analysis was performed in HaploReg, using the 1000 Genomes Phase 1 European population for LD calculation and the Roadmap Epigenome Mapping Consortium data set for enhancer enrichment analysis. Statistically significant differences (uncorrected $P < 0.05$) from background (1000 Genomes Pilot) were found for several brain cell lines and a skeletal muscle line (Supplementary Material, Table S2). Four overlapping haplotype blocks containing MYH7B and TRPC4AP and four blocks containing EDEM2 were identified (Supplementary Material, Table S2). Multiple SNPs in LD with associated SNPs alter transcription factor binding motifs, including CTCF sites in MYH7B, histone enhancer marks and DNase sites.

Genevar cis-eQTL analysis was carried out using the data input from the GenCord individuals, consisting of all three tissues (fibroblast, lymphoblastoid cell line and T-cell) from umbilical cord samples (42) (1 Mb window P -value cut-off < 0.001). Using this data set, all eight SNPs were associated with expression of TRPC4AP (probe ILMN_2402805; Supplementary Material, Fig. S3). Two SNPs (rs6088662 and rs6120777) were also associated with expression of CEP250 in those tissues.

Post hoc power calculations

Calculations for *post hoc* power were performed in Quanto v1.1 (43). Based on an additive model, population frequency of LSLs = 0.001, sample size of cases = 780 and 4.5 unmatched controls per case, this study had > 0.80 power to identify risk alleles with

an odds ratio (OR) of 1.6 using SNPs with minor allele frequencies between 0.15 and 0.50.

Discussion

LSLs are relatively common birth defects with significant mortality and morbidity. Previous studies have suggested that a complex genetic architecture underlies these defects (25)—although anatomically distinct, they show co-segregation in families, have high heritability, exhibit variation in birth prevalence among different populations, and have been shown to be caused by chromosomal abnormalities, rare pathogenic CNVs (15–18), rare pathogenic variants in single genes (29,30,44,45) and a variety of environmental factors (46–48)—all hallmarks of complex genetic traits. We, thus, applied methodologies commonly used to dissect the contribution of common variation to a cohort of LSL patients using both case-control and family-based trio designs (the latter to address the issue of population stratification). We demonstrate a genome-wide significant association between LSLs and a region on chromosome 20, with evidence for replication; we also identified a candidate region on chromosome 1 with suggestive association just below genome-wide significance that was not replicated. These two regions represent novel loci that have not been identified in previous linkage, candidate gene associations or CNV studies.

The associated region on chromosome 20 spans a ~200 kb region (genomic coordinates 33 547 633–33 753 262; hg19) that includes five genes (Fig. 3). There are two compelling candidates for genes contributing to LSL pathogenesis within this region. MYH7B is an ancient duplicated gene (from MYH7, in which pathogenic variants lead to cardiomyopathy) that contains a micro RNA, miR-499A, in intron 19. Both are expressed in skeletal muscle, some regions of the brain, ocular muscles and heart (49). MYH7B displays frequent non-productive splicing resulting in uncoupling of expression of miR-499A; it is subject to decay and does not encode a functional protein, but miR-499A expression is preserved (50). Both are also known to be important in cardiomyocyte development, proliferation and maintenance, regulating expression of a number of developmentally important cardiac genes (51). Multiple MYH7B transcription start sites are

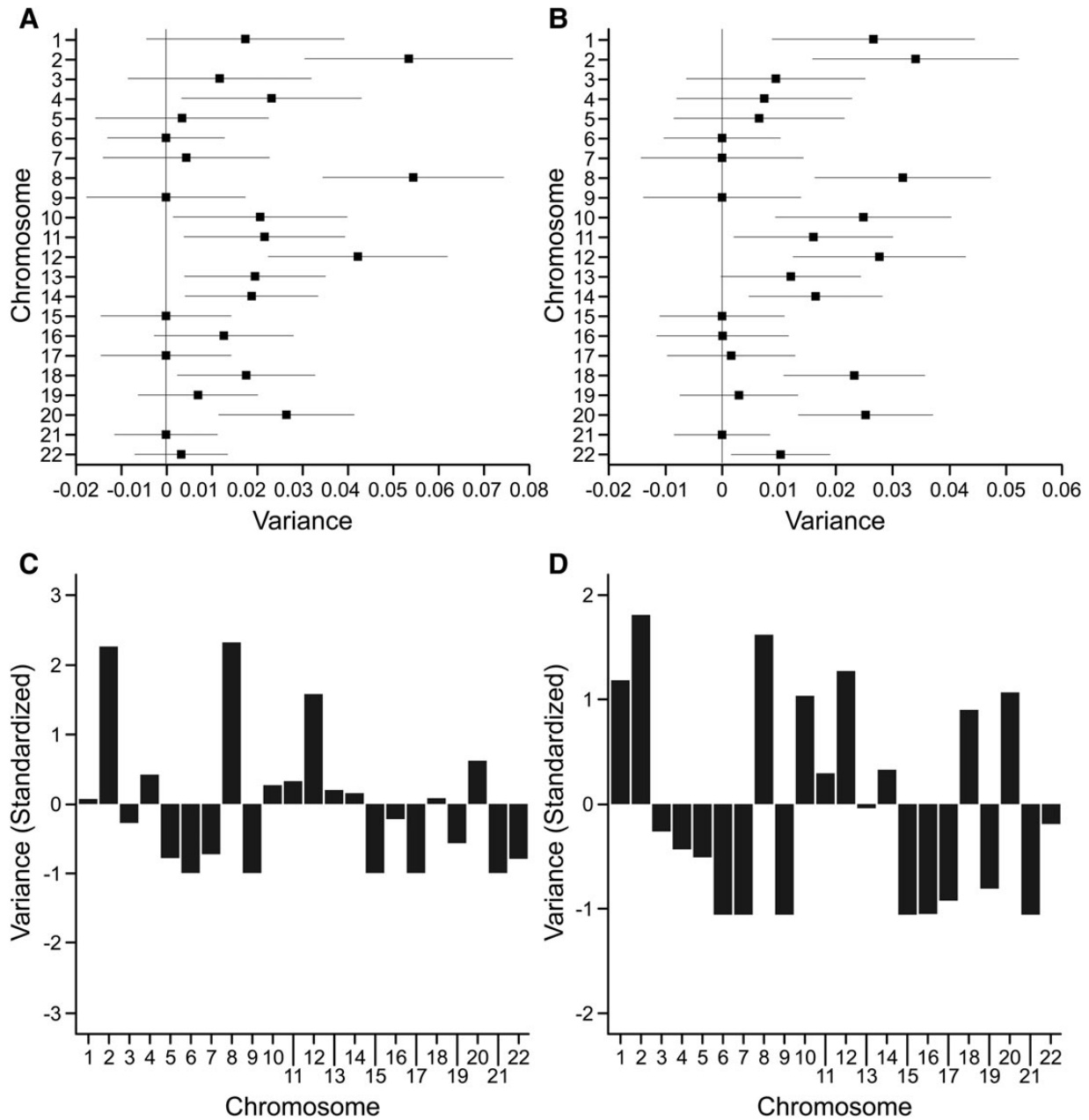


Figure 4. Genome-wide SNP heritability for LSLs. The figure shows absolute (upper) and normalized plots (lower) for an ancestry-restricted sample set (within 1 SD of MDS components 1 and 2) (A and C) and the full data set (B and D).

present in the heart, with a 6.2 kb promoter region upstream of *MYH7B* that is important in regulation of *MYH7B* transcription (52). Binding by myogenic regulatory factors to an E-box element and Eos to an Ikaros motif appear to regulate expression of *MYH7B*/*miR-499A* (51), and lead to modulation of downstream genes *MYH1* and *MYH7* via *SOX6* (52). *miR-499A* is also important in the pathogenesis of cardiac hypertrophy and cardiomyopathy (53). To date, there are no convincing genetic models for congenital AS, CoA or HLHS, illustrating the challenge of functional validation in this class of disease. The identified gene regions, therefore, stand as candidates to be investigated through model-organism knockout and gene-editing projects that include efforts to characterize developmental and cardiac phenotypes, such as the Knockout Mouse Phenotyping Program.

Although not reaching genome-wide significance in either the discovery or combined discovery-replication cohorts, the region on chromosome 1 is worth highlighting as a potential candidate region for future studies. This 300 kb region on chromosome 1 includes seven genes, of which *CTSS* and *CTSK*, which encode Cathepsins S and K, respectively, are attractive candidates as they have been implicated in the pathogenesis of acquired AS (54), heart failure secondary to hypertension (reviewed in 55), and dilated cardiomyopathy (56). Also in the region, *ARNT* (aryl hydrocarbon receptor nuclear translocator) is important in xenobiotic metabolism, involved in dioxin-related cardiac teratogenicity (57) and in development of cardiomyopathy (58). The associated variants in this region are of relatively low frequency [mean minor allele frequency (MAF) of 0.10]; thus, our ability to

confidently detect association at this locus is at the lower limit of power for our sample size; however, the candidate genes in this region warrant further scrutiny in larger LSL genetic studies.

Our study has several strengths. We confirmed the individual LSLs by review of echocardiograms, cardiac catheterization or surgical observation to ensure correct phenotyping, and did not include those with known syndromic diagnoses or those with large copy number losses (>1 Mb). Both case-control and trio-based analysis were performed, allowing us to maximize the number of subjects by including those for whom parental samples were unavailable, while also reducing the likelihood of false-positive association from population stratification. We were also able to replicate our observed association using data from a separate, albeit smaller, LSL cohort. The combined data set represents the largest genetic study of LSLs to date, including over 1000 individuals (discovery plus replication); despite this, our power to detect association is still modest, especially by current standards for complex traits such as LSL. Our analysis suggests that we have good power to identify association with an OR >1.5 at genome-wide significance, but are much less powered to identify more modest risk variants (with OR < 1.5 or MAF < 0.15) of the kind seen in the bulk of GWAS conducted for similar traits.

Previous analyses of multiplex families with LSLs have identified a number of linkage peaks; although none of these directly overlap the two suggested regions in our paper (23,28,29,59), a linkage peak on 20q11 (LOD = 2.0) centered on marker D20S107 (Chr20: 38 882 511–38 882 844), is ~4.5 Mb from the region identified here, providing some support for our association in that region. There have been six GWAS performed for CHDs to date (34,60–64), although only one directly addressed the burden of LSLs—Mitchell *et al.* (64) used a family-based approach to study LSLs. They found significant associations with the case genotype at 16q24.2 and with the maternal genotype at 10p11.23; neither of these loci were found at suggestive or genome-wide significance in the present study, likely reflecting both the broad complexity of the disease as well as methodological differences in study design. Neither study was able to replicate association at *ERBB4*—one of the few candidate genes previously shown to be associated with LSLs (33). Multiple CNV studies have also been performed for CHDs, usually involving small cohorts. A meta-analysis of CNVs found in CHDs catalogued nearly 70 recurrent events (65); however, no previously reported recurrent CHD associated CNV directly overlaps our identified regions. The closest, most relevant, recurrent CNV on 1q21.1 occurs ~3 Mb centromeric to our region at 1q21.3. The lack of consistently identified loci, despite multiple studies correlates with the known complexity and heterogeneity of LSLs, but may also reflect ‘genetic disharmony’ with distinct phenotypic classifications.

Finally, we present novel results revisiting the issue of heritability of LSL, now informed by genome-wide molecular data. These analyses support the previous conclusion that the overall heritability is high and also give a perspective on the contribution of common variants tagged by SNPs in our genome-wide data set. While recognizing that the current study is underpowered to give high-precision estimates of the contributions of individual chromosomes, the broad conclusion that there are additional loci that could be discovered in follow on studies with larger sample size is well-supported. While investigation of rare pathogenic variants with large effects is of great practical and theoretical importance, the genetic architecture of CHDs will not be completely accounted for without more thorough investigation of common variant contributions. LSL, like all severe cardiac malformations, reduce reproductive fitness. Alleles that have a large effect in increasing risk of LSL or actually cause LSL as part of a more

complex syndrome, by necessity, have low-allele frequency and are eliminated quickly in populations. Under a simplistic model of mutation selection balance the aggregate allele frequency should be very close to the newborn prevalence rate. It may, therefore, seem counterintuitive that common alleles could contribute to such a severe phenotype. However, it has been noted that when risk-increasing alleles have a weak effect like those typically detected in GWAS, then the associated selection coefficient is relatively modest and the frequency of risk alleles may be dominated by drift, migration, bottlenecks etc. (66,67). Because the proportion of the population harboring common risk alleles is large compared with those bearing highly penetrant rare variants, such common alleles can underlie a substantial fraction of cases. This is reminiscent of other complex traits for which both rare Mendelian and common variants have been implicated in disease risk; both in the same gene (*LDL-R* in hypercholesterolemia) and different genes (*Presenilin* and *APOE* in Alzheimer’s). While the congenital nature of LSLs represents a distinct difference from later-onset diseases, the phenotypic spectrum of LSLs (within the context of genetic modifiers of penetrance and expression) allows for this model to hold true, as would one where variants are pleiotropic in their manifestation and thus intermediate in their allele frequency.

We provide evidence for association of LSLs with a novel locus on chromosome 20q11, and proffer *MYH7B/miR-499A* as a candidate susceptibility gene in this region. Family-based analyses, replication and previous linkage data support this hypothesis. Our inability to replicate association at regions implicated by other studies of this phenotype, further bolster the notion that significant locus heterogeneity exists for LSLs. Consequently, future studies that embrace larger and more collaborative study samples will be of high value in ultimately understanding the allelic spectrum of this complex phenotype.

Materials and Methods

Description of cohorts

Probands, parents (and other relatives) along with individual cases (without parents) were enrolled from Texas Children’s Hospital in Houston, Texas; Children’s Hospital in Linz, Austria and Nationwide Children’s Hospital in Columbus, Ohio, under IRB approved protocols.

Inclusion criteria consisted of a diagnosis of congenital AS, CoA with or without BAV and/or ventricular septal defect, IAAA, Shone complex, mitral valve stenosis or atresia, or HLHS. Diagnosis was confirmed by echocardiography, cardiac catheterization or direct observation at surgery. Cases were excluded if they had multiple congenital extra-cardiac anomalies or had a CHD secondary to a known syndrome or single gene disorder.

Control data sets

Controls genotyped on Illumina Omni1-Quad platform were obtained from the dbGaP, with permission. The Omni1-Quad platform has a 97% SNP overlap with the Illumina OmniExpress platform on which the LSL families were genotyped. Genotype data from control individuals enrolled in the *high-density SNP association analysis of melanoma: case-control and outcomes investigation* (phs000187.v1.p1) and from the *GWAS of Parkinson disease: genes and environment* (phs000196.v2.p1) were used as our control reference group; both control groups were recruited from within the USA. SNP genotypes in the control set were subject to the same data QC (including MDS and ancestry stratification) and analyses as the cases. In addition, we compared allele

frequencies between cohorts by testing for allelic association between the two groups. We did not observe any systematic differences between these two control cohorts (Supplementary Material, Fig. S4).

Genotyping and quality control

Genotyping of the LSL cohort (cases) was performed on an Illumina iScan using the HumanOmniExpress-12 v1.0 BeadChip per the manufacturer's instructions. Extensive QC was performed on the resulting data. BeadChips with genotype call rates <98% were removed. Data cleaning also included checks of gender, Mendelian inheritance errors, inbreeding coefficient (F) for excess homozygosity (>2 SDs) and heterozygosity (<2 SDs), Hardy–Weinberg equilibrium (HWE), and assessments of relatedness in both PLINK v1.07 (68) (IBD analysis with PI >0.10) and KING (69) (both between and within family). Discrepancies in gender or inheritance that could not be resolved resulted in removal of that sample. Samples were also removed if there were Mendelian errors in >1% of SNPs for a given trio (including misattributed paternity), genotype call rates <98%, and large copy number variations (>1 Mb) based on the Illumina Genome Studio plugin CNVpartition (N = 20). Four identical pairs of samples (PI_HAT = 1) were identified and one member of each pair (with the lower call rate) was excluded. SNPs with MAF <5%, HWE $P < 10^{-5}$ in controls, or missingness by case/control status $P < 0.0001$, were also removed from the analysis. All SNPs were mapped to genome build hg19. Subsequent analyses focused on the autosomal chromosomes.

Genome-wide association analyses

For case–control analyses, probands from trio and duo families plus all singleton probands were used as potential cases. Genotypes from these individuals were merged with the combined dbGaP control data set in PLINK, and subject to additional data QC to detect SNP-strand flips and remove ambiguous SNPs (A > T, T > A, C > G and G > C changes; N = 2020). MDS was performed on the merged data set using PLINK. The resulting MDS components were used both to detect population outliers (N = 4) for removal and as ancestry-related covariates in later analyses. For comparison with 1000 genomes super-populations (70,71), study samples were merged with 1000 Genome Project samples (N = 2276) using 521 869 SNPs common to both genotyping platforms. Subsequent SNP and sample QC proceeded as outlined above.

Case–control genome-wide association was executed in PLINK using logistic regression under an allelic disease model. The first two MDS components and sex were used as covariates. SNPs were deemed to be genome-wide significant if they had an unadjusted $P < 5 \times 10^{-8}$. Manhattan and QQ plots were generated using the R package 'qqman' in R version 3.2.2 (72). LocusZoom Version 1.1 (73) was used to create regional association plots. Genome-wide inflation, based on median χ^2 , was calculated in PLINK.

TDT analyses

TDT was performed on trios and duos genotyped on the OmniExpress platform using PLINK. A final set of 601 138 SNPs was used for the analysis. We also performed family-based association testing using the same families using the SNP and variation suite PBAT add-on module Version 8 (Golden Helix); the results were consistent across the two analyses.

Fine mapping analysis (imputation)

Imputation of SNPs on chromosomes 1 and 20 was implemented for the case–control data set using a subset of SNPs with MAF

>0.05 and HWE P -value $> 10^{-3}$. SHAPEIT2 (74) was used to pre-phase the chromosomes and Impute2 (75) was used for imputation. As recommended, all of the available reference panels in Phase 3 of the 1000 genomes project were used to infer genotypes, and imputation was implemented in non-overlapping 5 Mb windows (with 250 kb buffer) across the genome using the recommended default parameters. PLINK v1.90b3v (76) was used to convert genotype probabilities of variants into hard genotype calls. Imputed SNPs with 'INFO' scores >0.9 were included, resulting in a total of 447 504 QC SNPs on chromosomes 1 and 20 available for subsequent association analyses.

Genome-wide heritability analysis

Genome-wide complex trait analysis version 1.24.4 (36,37) was used for the heritability estimate analysis. Relatedness between individuals that is not attributable to case–control status can inflate measures of heritability; therefore, extensive QC and inclusion of relevant covariates was implemented. First, the genetic relationship between all pairs of samples was estimated from all SNPs. One sample from each pair was removed if the estimated relatedness was >0.05 (approximately second cousins); 125 samples were removed from this step leaving 3409 samples (659 cases and 2750 controls) for the primary analysis. The estimated genetic relationship matrix was then fitted in a linear mixed model to estimate proportion of phenotypic variance accounted for by the SNPs via a restricted maximum likelihood analysis. This method corrects for ascertainment bias in which the proportion of cases in the sample is higher than the prevalence in the population, adjusts for the number of SNPs, and transforms the variance estimate from the observed 0–1 scale to the unobserved continuous liability scale. A population prevalence of 0.0016 was used for the transformation (24). Analyses were performed to estimate the heritability on the liability scale for SNPs on all chromosomes (autosomal) as well as on individual chromosomes by including the individual chromosomes in a joint analysis. Sex and the first two MDS components were included as covariates. In order to evaluate the effects of population stratification on the heritability estimate, we also performed the analyses in a homogeneous subset of the sample for which the first and second MDS component values were within 1 SD of the mean in the first MDS component and 1 SD of mean in the second MDS component considering all samples; for this, 2806 samples (508 cases and 2298 controls) were included in the analysis. Forest plots were generated using the R package 'forest-plot'. All plots were generated in R version 3.2.2.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

The authors acknowledge the many families and auxiliary staff who participated in the study.

Conflict of Interest statement. None of the authors has any competing conflict of interest. Non-competing interests include: J.R.L.: stock ownership in 23andMe, Inc. and Lasergen, Inc.; paid consultant for Regeneron Pharmaceuticals; co-inventor on the multiple USA and European patents related to molecular diagnostics. H.J.: —consultant for St Jude Medical, Medtronic and Janssen pharmaceutical.

Funding

This work was funded in part by National Institutes of Health (NIH) grants to J.W.B. (1U54 HD083092, 5R01 HD039056, 5R01 HL090506, 5R01 HL091771) and K.L.M. (5 R01 HL109758 and 1R21 HL106549-01); B.D.K., H.J.C., D.B. and S.B. are funded through the British Heart Foundation, Heart Research UK, and the European Union. J.D.B. and F.A.B. are funded by the British Heart Foundation. N.A.H. is funded by a Clinical Scientist Development Award from the Doris Duke Charitable Foundation (grant no.: 2013096). J.R.L. is supported in part by the US National Human Genome Research Institute (NHGRI)/National Heart Blood Lung Institute (NHLBI) jointly funded Baylor Hopkins Center for Mendelian Genomics (U54HG006542). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHGRI/NHLBI, NIH.

References

- Go, A.S., Mozaffarian, D., Roger, V.L., Benjamin, E.J., Berry, J.D., Borden, W.B., Bravata, D.M., Dai, S., Ford, E.S., Fox, C.S. et al. (2013) Heart disease and stroke statistics—2013 update: a report from the American Heart Association. *Circulation*, **127**, e6–e245.
- Hoffman, J.I. and Kaplan, S. (2002) The incidence of congenital heart disease. *J. Am. Coll. Cardiol.*, **39**, 1890–1900.
- Marelli, A.J., Ionescu-Ittu, R., Mackie, A.S., Guo, L., Dendukuri, N. and Kaouache, M. (2014) Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. *Circulation*, **130**, 749–756.
- Marelli, A.J., Mackie, A.S., Ionescu-Ittu, R., Rahme, E. and Pilote, L. (2007) Congenital heart disease in the general population: changing prevalence and age distribution. *Circulation*, **115**, 163–172.
- Simeone, R.M., Oster, M.E., Cassell, C.H., Armour, B.S., Gray, D.T. and Honein, M.A. (2014) Pediatric inpatient hospital resource use for congenital heart defects. *Birth Defects Res. A Clin. Mol. Teratol.*, **100**, 934–943.
- Heron, M. (2013) Deaths: leading causes for 2010. *Natl. Vital Stat. Rep.*, **62**, 1–96.
- Horner, T., Liberthson, R. and Jellinek, M.S. (2000) Psychosocial profile of adults with complex congenital heart disease. *Mayo Clin. Proc.*, **75**, 31–36.
- Marino, B.S., Lipkin, P.H., Newburger, J.W., Peacock, G., Gerdes, M., Gaynor, J.W., Mussatto, K.A., Uzark, K., Goldberg, C.S., Johnson, W.H. Jr. et al. (2012) Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. *Circulation*, **126**, 1143–1172.
- Feinstein, J.A., Benson, D.W., Dubin, A.M., Cohen, M.S., Maxey, D.M., Mahle, W.T., Pahl, E., Villafane, J., Bhatt, A.B., Peng, L.F. et al. (2012) Hypoplastic left heart syndrome: current considerations and expectations. *J. Am. Coll. Cardiol.*, **59**, S1–42.
- Oster, M.E., Lee, K.A., Honein, M.A., Riehle-Colarusso, T., Shin, M. and Correa, A. (2013) Temporal trends in survival among infants with critical congenital heart defects. *Pediatrics*, **131**, e1502–e1508.
- Verheugt, C.L., Uiterwaal, C.S., van der Velde, E.T., Meijboom, F.J., Pieper, P.G., van Dijk, A.P., Vliegen, H.W., Grobbee, D.E. and Mulder, B.J. (2010) Mortality in adult congenital heart disease. *Eur. Heart J.*, **31**, 1220–1229.
- Botto, L.D., Lin, A.E., Riehle-Colarusso, T., Malik, S. and Correa, A. and National Birth Defects Prevention Study. (2007) Seeking causes: classifying and evaluating congenital heart defects in etiologic studies. *Birth Defects Res. A Clin. Mol. Teratol.*, **79**, 714–727.
- Clark, E.B. (1996) Pathogenetic mechanisms of congenital cardiovascular malformations revisited. *Semin. Perinatol.*, **20**, 465–472.
- Lenke, R.R. and Levy, H.L. (1980) Maternal phenylketonuria and hyperphenylalaninemia. An international survey of the outcome of untreated and treated pregnancies. *N. Engl. J. Med.*, **303**, 1202–1208.
- Glessner, J.T., Bick, A.G., Ito, K., Homsy, J.G., Rodriguez-Murillo, L., Fromer, M., Mazaika, E., Vardarajan, B., Italia, M., Leipzig, J. et al. (2014) 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.*, **115**, 884–896.
- Soemedi, R., Wilson, I.J., Bentham, J., Darlay, R., Topf, A., Zelenika, D., Cosgrove, C., Setchfield, K., Thornborough, C., Granados-Riveron, J. et al. (2012) Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am. J. Hum. Genet.*, **91**, 489–501.
- Hitz, M.P., Lemieux-Perreault, L.P., Marshall, C., Feroz-Zada, Y., Davies, R., Yang, S.W., Lionel, A.C., D'Amours, G., Lemyre, E., Cullum, R. et al. (2012) Rare copy number variants contribute to congenital left-sided heart disease. *PLoS Genet.*, **8**, e1002903.
- Payne, A.R., Chang, S.W., Koenig, S.N., Zinn, A.R. and Garg, V. (2012) Submicroscopic chromosomal copy number variations identified in children with hypoplastic left heart syndrome. *Pediatr. Cardiol.*, **33**, 757–763.
- Ferencz, C., Loffredo, C.A., Corea-Vilasenor, A. and Wilson, P.D. (1997) Anderson, R. (ed.), In *Genetic and Environmental Risk Factors of Major Cardiovascular Malformations*. Futura Publishing Co. Inc., Armonk, NY, Vol. 5, pp. 165–225.
- Kerstjens-Frederikse, W.S., Du Marchie Sarvaas, G.J., Ruiters, J.S., Van Den Akker, P.C., Temmerman, A.M., Van Melle, J.P., Hofstra, R.M. and Berger, R.M. (2011) Left ventricular outflow tract obstruction: should cardiac screening be offered to first-degree relatives? *Heart*, **97**, 1228–1232.
- Lewin, M.B., McBride, K.L., Pignatelli, R., Fernbach, S., Combes, A., Menesses, A., Lam, W., Bezold, L.I., Kaplan, N., Towbin, J.A. et al. (2004) Echocardiographic evaluation of asymptomatic parental and sibling cardiovascular anomalies associated with congenital left ventricular outflow tract lesions. *Pediatrics*, **114**, 691–696.
- Hinton, R.B. Jr, Martin, L.J., Tabangin, M.E., Mazwi, M.L., Cripe, L.H. and Benson, D.W. (2007) Hypoplastic left heart syndrome is heritable. *J. Am. Coll. Cardiol.*, **50**, 1590–1595.
- McBride, K.L., Zender, G.A., Fitzgerald-Butt, S.M., Koehler, D., Menesses-Diaz, A., Fernbach, S., Lee, K., Towbin, J.A., Leal, S. and Belmont, J.W. (2009) Linkage analysis of left ventricular outflow tract malformations (aortic valve stenosis, coarctation of the aorta, and hypoplastic left heart syndrome). *Eur. J. Hum. Genet.*, **17**, 811–819.
- McBride, K.L., Marengo, L., Canfield, M., Langlois, P., Fixler, D. and Belmont, J.W. (2005) Epidemiology of noncomplex left ventricular outflow tract obstruction malformations (aortic valve stenosis, coarctation of the aorta, hypoplastic left heart syndrome) in Texas, 1999–2001. *Birth Defects Res. A Clin. Mol. Teratol.*, **73**, 555–561.
- McBride, K.L., Pignatelli, R., Lewin, M., Ho, T., Fernbach, S., Menesses, A., Lam, W., Leal, S.M., Kaplan, N., Schliekelman, P. et al. (2005) Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: segregation, multiplex relative risk, and heritability. *Am. J. Med. Genet. A*, **134A**, 180–186.

26. Oyen, N., Poulsen, G., Boyd, H.A., Wohlfahrt, J., Jensen, P.K. and Melbye, M. (2009) Recurrence of congenital heart defects in families. *Circulation*, **120**, 295–301.
27. Lupski, J.R., Belmont, J.W., Boerwinkle, E. and Gibbs, R.A. (2011) Clan genomics and the complex architecture of human disease. *Cell*, **147**, 32–43.
28. Hinton, R.B., Martin, L.J., Rame-Gowda, S., Tabangin, M.E., Cripe, L.H. and Benson, D.W. (2009) Hypoplastic left heart syndrome links to chromosomes 10q and 6q and is genetically related to bicuspid aortic valve. *J. Am. Coll. Cardiol.*, **53**, 1065–1071.
29. Garg, V., Muth, A.N., Ransom, J.F., Schluterman, M.K., Barnes, R., King, I.N., Grossfeld, P.D. and Srivastava, D. (2005) Mutations in *NOTCH1* cause aortic valve disease. *Nature*, **437**, 270–274.
30. McBride, K.L., Riley, M.F., Zender, G.A., Fitzgerald-Butt, S.M., Towbin, J.A., Belmont, J.W. and Cole, S.E. (2008) *NOTCH1* mutations in individuals with left ventricular outflow tract malformations reduce ligand-induced signaling. *Hum. Mol. Genet.*, **17**, 2886–2893.
31. Junker, R., Kotthoff, S., Vielhaber, H., Halimeh, S., Kosch, A., Koch, H.G., Kassenbohmer, R., Heineking, B. and Nowak-Gottl, U. (2001) Infant methylenetetrahydrofolate reductase 677TT genotype is a risk factor for congenital heart disease. *Cardiovasc. Res.*, **51**, 251–254.
32. McBride, K.L., Fernbach, S., Menesses, A., Molinari, L., Quay, E., Pignatelli, R., Towbin, J.A. and Belmont, J.W. (2004) A family-based association study of congenital left-sided heart malformations and 5,10 methylenetetrahydrofolate reductase. *Birth Defects Res. A Clin. Mol. Teratol.*, **70**, 825–830.
33. McBride, K.L., Zender, G.A., Fitzgerald-Butt, S.M., Seagraves, N.J., Fernbach, S.D., Zapata, G., Lewin, M., Towbin, J.A. and Belmont, J.W. (2011) Association of common variants in *ERBB4* with congenital left ventricular outflow tract obstruction defects. *Birth Defects Res. A Clin. Mol. Teratol.*, **91**, 162–168.
34. Cordell, H.J., Bentham, J., Topf, A., Zelenika, D., Heath, S., Mamasoula, C., Cosgrove, C., Blue, G., Granados-Riveron, J., Setchfield, K. et al. (2013) Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat. Genet.*, **45**, 822–824.
35. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
36. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
37. Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.
38. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, **26**, 2069–2070.
39. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S. et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genet. Res.*, **22**, 1790–1797.
40. Ritchie, G.R., Dunham, I., Zeggini, E. and Flicek, P. (2014) Functional annotation of noncoding sequence variants. *Nat. Methods*, **11**, 294–296.
41. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474–2476.
42. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M. et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, **325**, 1246–1250.
43. Gauderman, W. and Morrison, J. (2006) QUANTO 1.1: a computer program for power and sample size calculations for genetic-epidemiology studies. <http://biostats.usc.edu/Quanto.html>, July 2015, date last accessed.
44. Chang, S.W., Mislankar, M., Misra, C., Huang, N., Dajusta, D.G., Harrison, S.M., McBride, K.L., Baker, L.A. and Garg, V. (2013) Genetic abnormalities in *FOXP1* are associated with congenital heart defects. *Hum. Mutat.*, **34**, 1226–1230.
45. Riley, M.F., McBride, K.L. and Cole, S.E. (2011) *NOTCH1* missense alleles associated with left ventricular outflow tract defects exhibit impaired receptor processing and defective EMT. *Biochim. Biophys. Acta*, **1812**, 121–129.
46. Blue, G.M., Kirk, E.P., Sholler, G.F., Harvey, R.P. and Winlaw, D.S. (2012) Congenital heart disease: current knowledge about causes and inheritance. *Med. J. Aust.*, **197**, 155–159.
47. Csaky-Szunyogh, M., Vereczkey, A., Kosa, Z., Gerencser, B. and Czeizel, A.E. (2014) Risk factors in the origin of congenital left-ventricular outflow-tract obstruction defects of the heart: a population-based case-control study. *Pediatr. Cardiol.*, **35**, 108–120.
48. Liu, S., Joseph, K.S., Lisonkova, S., Rouleau, J., Van den Hof, M., Sauve, R. and Kramer, M.S. and Canadian Perinatal Surveillance System (Public Health Agency of Canada). (2013) Association between maternal chronic conditions and congenital heart defects: a population-based cohort study. *Circulation*, **128**, 583–589.
49. van Rooij, E., Quiat, D., Johnson, B.A., Sutherland, L.B., Qi, X., Richardson, J.A., Kelm, R.J. Jr and Olson, E.N. (2009) A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. *Dev. Cell*, **17**, 662–673.
50. Bell, M.L., Buvoli, M. and Leinwand, L.A. (2010) Uncoupling of expression of an intronic microRNA and its myosin host gene by exon skipping. *Mol. Cell Biol.*, **30**, 1937–1945.
51. Warkman, A.S., Whitman, S.A., Miller, M.K., Garriock, R.J., Schwach, C.M., Gregorio, C.C. and Krieg, P.A. (2012) Developmental expression and cardiac transcriptional regulation of *Myh7b*, a third myosin heavy chain in the vertebrate heart. *Cytoskeleton*, **69**, 324–335.
52. Yeung, F., Chung, E., Guess, M.G., Bell, M.L. and Leinwand, L.A. (2012) *Myh7b/miR-499* gene expression is transcriptionally regulated by MRFs and Eos. *Nucleic Acids Res.*, **40**, 7303–7318.
53. Matkovich, S.J., Hu, Y., Eschenbacher, W.H., Dorn, L.E. and Dorn, G.W. II. (2012) Direct and indirect involvement of microRNA-499 in clinical and experimental cardiomyopathy. *Circ. Res.*, **111**, 521–531.
54. Helske, S., Syvaranta, S., Lindstedt, K.A., Lappalainen, J., Oorni, K., Mayranpaa, M.I., Lommi, J., Turto, H., Werkkala, K., Kupari, M. et al. (2006) Increased expression of elastolytic cathepsins S, K, and V and their inhibitor cystatin C in stenotic aortic valves. *Arterioscler. Thromb. Vasc. Biol.*, **26**, 1791–1798.
55. Cheng, X.W., Shi, G.P., Kuzuya, M., Sasaki, T., Okumura, K. and Murohara, T. (2012) Role for cysteine protease cathepsins in heart disease: focus on biology and mechanisms with clinical implication. *Circulation*, **125**, 1551–1562.
56. Hua, Y., Xu, X., Shi, G.P., Chicco, A.J., Ren, J. and Nair, S. (2013) Cathepsin K knockout alleviates pressure overload-induced cardiac hypertrophy. *Hypertension*, **61**, 1184–1192.

57. Antkiewicz, D.S., Peterson, R.E. and Heideman, W. (2006) Blocking expression of AHR2 and ARNT1 in zebrafish larvae protects against cardiac toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol. Sci.*, **94**, 175–182.
58. Wu, R., Chang, H.C., Khechaduri, A., Chawla, K., Tran, M., Chai, X., Wagg, C., Ghanefar, M., Jiang, X., Bayeva, M. et al. (2014) Cardiac-specific ablation of ARNT leads to lipotoxicity and cardiomyopathy. *J. Clin. Invest.*, **124**, 4795–4806.
59. Martin, L.J., Ramachandran, V., Cripe, L.H., Hinton, R.B., Andelfinger, G., Tabangin, M., Shoener, K., Keddache, M. and Benson, D.W. (2007) Evidence in favor of linkage to human chromosomal regions 18q, 5q and 13q for bicuspid aortic valve and associated cardiovascular malformations. *Hum. Genet.*, **121**, 275–284.
60. Zhao, B., Lin, Y., Xu, J., Ni, B., Da, M., Ding, C., Hu, Y., Zhang, K., Yang, S., Wang, X. et al. (2014) Replication of the 4p16 susceptibility locus in congenital heart disease in Han Chinese populations. *PLoS One*, **9**, e107411.
61. Cordell, H.J., Topf, A., Mamasoula, C., Postma, A.V., Bentham, J., Zelenika, D., Heath, S., Blue, G., Cosgrove, C., Granados Riveron, J. et al. (2013) Genome-wide association study identifies loci on 12q24 and 13q32 associated with tetralogy of Fallot. *Hum. Mol. Genet.*, **22**, 1473–1481.
62. Hu, Z., Shi, Y., Mo, X., Xu, J., Zhao, B., Lin, Y., Yang, S., Xu, Z., Dai, J., Pan, S. et al. (2013) A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat. Genet.*, **45**, 818–821.
63. Lin, Y., Guo, X., Zhao, B., Liu, J., Da, M., Wen, Y., Hu, Y., Ni, B., Zhang, K., Yang, S. et al. (2015) Association analysis identifies new risk loci for congenital heart disease in Chinese populations. *Nat. Commun.*, **6**, 8082.
64. Mitchell, L.E., Agopian, A.J., Bhalla, A., Glessner, J.T., Kim, C.E., Swartz, M.D., Hakonarson, H. and Goldmuntz, E. (2015) Genome-wide association study of maternal and inherited effects on left-sided cardiac malformations. *Hum. Mol. Genet.*, **24**, 265–273.
65. Thorsson, T., Russell, W.W., El-Kashlan, N., Soemedi, R., Levine, J., Geisler, S.B., Ackley, T., Tomita-Mitchell, A., Rosenfeld, J.A., Topf, A. et al. (2015) Chromosomal imbalances in patients with congenital cardiac defects: a meta-analysis reveals novel potential critical regions involved in heart development. *Congenit. Heart Dis.*, **10**, 193–208.
66. Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.*, **69**, 124–137.
67. Pritchard, J.K. and Cox, N.J. (2002) The allelic architecture of human disease genes: common disease-common variant. . . or not? *Hum. Mol. Genet.*, **11**, 2417–2423.
68. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
69. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M. and Chen, W.M. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics*, **26**, 2867–2873.
70. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
71. Howie, B., Marchini, J. and Stephens, M. (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda)*, **1**, 457–470.
72. R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
73. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
74. Delaneau, O., Marchini, J. and Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179–181.
75. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
76. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.