Published in final edited form as:

Am J Med Genet A. 2020 January ; 182(1): 250–254. doi:10.1002/ajmg.a.61411.

X-chromosome association studies of congenital heart defects

A. J. Agopian[1], **Thanh T. Hoang**[1], **Elizabeth Goldmuntz**[2,3], **Hakon Hakonarson**[3,4], **Fadi I. Musfee**[1], **Laura E. Mitchell**[1], **Pediatric Cardiac Genomics Consortium**[†]

[1]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, UTHealth School of Public Health, Houston, Texas

[2]Division of Cardiology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

[3]Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania

[4]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

Several genome-wide association studies (GWAS) have been undertaken to identify common variants associated with the risk of congenital heart defects (CHDs). GWAS are commonly used to identify disease-associated variants for conditions that appear to be genetically complex. Such studies have contributed to understanding of a range of conditions (e.g., diabetes), including other structural birth defects (e.g., cleft lip; Lupo, Mitchell & Jenkins, 2019; Tam et al., 2019). However, in the majority of GWAS of CHDs, X-linked variants have been excluded. Exclusion of data from the X-chromosome is a common practice in GWAS because the statistical methods for X-linked and autosomal variants differ and methods for X-linked variants have lagged behind those for autosomal variants. There are, however, methods for analyzing common variants across the X-chromosome that can be used to expand the scope of GWAS and could reveal novel CHD-related genes.

We have reported on GWAS and meta-analyses of two common types of CHDs, conotruncal heart defects (CTDs), and left ventricular outflow tract obstructions (LVOTOs), that are more common in males than females (Agopian et al., 2017). These analyses were, however,

Correspondence: Laura E. Mitchell, University of Texas Health Science Center at Houston, School of Public Health, 1200 Pressler Street, Houston, TX 77030. laura.e.mitchell@uth.tmc.edu.

[†]The Pediatric Cardiac Genomics Consortium Membership list is provided in Table S1.

CONFLICT OF INTEREST The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT Data from the Pediatric Cardiac Genomics Consortium are available through dbGAP (phs001194.v2.p2). Data from the Children's Hospital of Philadelphia may be requested from Dr. E. Goldmuntz.

SUPPORTING INFORMATION Additional supporting information may be found online in the Supporting Information section at the end of this article.

WEB RESOURCES

IMPUTE2: https://mathgen.stats.ox.ac.uk/impute/impute_v2.html
METAL: http://csg.sph.umich.edu/abecasis/metal/
OMIM: https://omim/org/
PIXLRT: https://www.niehs.nih.gov/research/resources/software/biostatistics/pixlrt/index.cfm
PLINK: https://www.cog-genomics.org/plink/1.9/
SHAPEIT2: http://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
XWAS: http://keinanlab.cb.bscb.cornell.edu/content/xwas

restricted to autosomal variants. Here, we present results from X-chromosome-wide analyses and meta-analyses conducted in the same five study populations (Table 1). Briefly, study participants were recruited under protocols approved by the Children's Hospital of Philadelphia (CHOP) or Pediatric Cardiac Genomics Consortium (PCGC) clinical centers. Adults participants provided informed consent for themselves and their participating minor children. Cases included individuals with a CTD (N = 1,123) or an LVOTO (N = 384) who were not diagnosed with an underlying syndrome. Four datasets were based on case-parent trios and a fifth was based on cases and controls.

Study participants were genotyped using Illumina single nucleo-tide polymorphism (SNP) arrays (550, 610, 1M or 2.5M). Autosomal variants were imputed and standard quality control (QC) procedures were conducted for the autosomal variants (Agopian et al., 2017). For this study, we performed additional, X-specific imputations and QC procedures. Preimputation X-chromosome QC procedures (Konig, Loley, Erdmann, & Ziegler, 2014) were applied separately to each dataset using PLINK v1.9 (Chang et al., 2015). We excluded variants in the pseudoautosomal region of the X-chromosome and cases with discrepancies between reported and genotypic sex or an f-estimate between 0.31 and 0.80. In addition, heterozygous genotyping calls in males were set to missing. We also excluded trios with an X-chromosome Mendelian error rate >1%, subjects with an X-chromosome genotyping call rate <95% and SNPs with an X-chromosome Mendelian error rate >10%, minor allele frequency (MAF) <1%, or call rate <90%. Following these exclusions, we removed variants with a MAF <1% in either sex as well as variants with significantly different ($p < 10^{-7}$) missing rates in males and females.

After the preimputation QC exclusions, the three CHOP datasets were combined and X-chromosome SNPs present in all three datasets were used for imputation. Similarly, the two PCGC datasets were combined and SNPs present in both were used for imputation. Haplotypes were prephased using SHAPEIT v2.r644 (Delaneau, Zagury, & Marchini, 2013) and nonpseudoautosomal regions were imputed using IMPUTE2 v2.3.2 (Howie, Donnelly, & Marchini, 2009) with the 1000 Genomes Project Phase I integrated variant set v3 as the reference. Following imputation, we excluded SNPs if they were poorly imputed (info score <0.80), had a MAF <5% in either sex, or had a call rate <90%.

First, we conducted SNP-level analyses in each dataset (N = 5,853 genotyped and 108,456 imputed SNPs for CHOP; N = 14,667 genotyped and 140,763 imputed SNPs for PCGC). For these analyses, genotypes were coded to reflect the number of minor alleles (males: 0, 1; females: 0, 1, 2) and data for female cases were analyzed under an additive genetic model. The trio datasets were analyzed using the parent-informed X-chromosome likeli-hood ratio test (Wise, Shi, & Weinberg, 2015). We used the R package, PIXLRT, to generate separate test statistics for trios with male and female cases and to combine these estimates for an overall test. The case–control dataset was analyzed with the chromosome X-Wide Analysis toolset v2.0 (Gao et al., 2015) using the –strat-sex command to test for association separately by sex and Stouffer's method to combine results.

No genome-wide significant ($p < 5 \times 10^{-8}$) SNP-level associations were identified (Tables S2–S6). Moreover, evidence suggestive of association ($p < 10^{-5}$) was only observed in one

dataset: In the CHOP CTD case–control dataset, there was suggestive evidence of association for three intergenic SNPs (rs5908462, $p = 2.6 \times 10^{-6}$; rs5908494, $p = 2.6 \times 10^{-6}$; rs5908495, $p = 4.1 \times 10^{-6}$) that are in close proximity (~350 basepairs) to each other and approximately 22,000 basepairs (bp) from *SPANXN4* (OMIM: 300667), the nearest protein coding gene.

Next, we conducted SNP-level meta-analyses using the weighted *Z*-score method in METAL (Willer, Li, & Abecasis, 2010). Meta-analyses were performed separately for CTDs and LVOTOs and for the combined (CTDs + LVOTOs) datasets. The three intergenic SNPs with suggestive evidence of association in the CHOP case–control dataset were not suggestive of association in the CTD (or any other) meta-analysis. Further, no genome-wide significant associations were detected. The only suggestive association was for a single intergenic SNP (rs4826814, $p = 3.6 \times 10^{-6}$) in the CTD + LVOTO meta-analysis. This SNP lies approximately 94,000 bp from the nearest protein coding gene, *NLGN4X* (OMIM: 300427), which is associated with X-linked autism (OMIM: 300495) and Asperger syndrome (OMIM: 300497).

Finally, we conducted gene-level analyses using the summary statistics from the three SNP-level meta-analyses. For these analyses, genes were defined by their transcription start–stop positions (GRCh37/hg19) plus 1 kb upstream and downstream. Gene test-statistics were calculated as the weighted mean of the statistic for the SNP with the lowest *p*-value and the average of the test statistics for all SNPs in the gene using the multi = SNP-wise option in MAGMA (de Leeuw, Mooij, Heskes, & Posthuma, 2015). No significant ($p < 7.1 \times 10^{-5}$, corrected for 706 genes) or suggestive ($p < 10^{-3}$) associations were detected (Tables S7–S9). Nineteen genes had association *p*-values <.01 in at least one of the meta-analyses (Table 2).

One gene, *SSR4* (OMIM: 300090), had an association *p*-value <.01 in two meta-analysis (LVOTO, $p = .002$; CTD + LVOTO, $p = .007$). Genetic variants in *SSR4* cause congenital disorder of gly-cosylation Type 1y (CDG1Y, OMIM: 300934): One of nine reported patients with CDG1Y had an unspecified cardiac anomaly (Ng et al., 2015). Of the genes with *p*-values <.01 in a single meta-analysis, two (*PIH1D3*, CTD meta-analysis $p = .003$; *CCDC22*, CTD + LVOTO meta-analysis $p = .003$) are associated with syndromes that include CHDs. Genetic variants in *PIH1D3* (OMIM: 300933) are one cause of primary ciliary dyskinesia (PCD, OMIM: 300991). Approximately, 50% of cases with PCD have situs inversus, situs ambiguous, or other laterality defects (Shapiro et al., 2014). Among 15 reported cases of *PIH1D*-associated PCD, approximately 50% were noted to have situs inversus, but information on cardiac phenotypes was not provided (Olcese et al., 2017; Paff et al., 2017). Genetic variants in *CCDC22* (OMIM:300859) are associated with Ritscher-Schinzel syndrome 2 (OMIM: 300963), which is characterized by intellectual disabilities and CHDs (septal defects; Kolanczyk et al., 2015; Voineagu et al., 2012). *LPAR4* also had an association *p*-value <.01 in one meta-analysis, and animal models suggest that *LPAR4* plays roles in in vascular development and is important for cardiogenesis (Sumida et al., 2010; Wang et al., 2012; Yukiura et al., 2011), regulating formation of the vascular network, as well as endothelial permeability, hematopoiesis, and lymphocyte migration (Yang et al., 2019).

In summary, our analyses suggest that, individually, common X-linked SNPs are unlikely to be strongly associated with either CTDs or LVOTOs. This is consistent with the results from three prior GWAS of CHDs that also evaluated X-linked SNPs. These studies were based on data from Europe and Australia and included cases with: septal, obstructive, and cyanotic CHDs (Cordell, Bentham, et al., 2013); tetralogy of Fallot (Cordell, Topf, et al., 2013); and ostium secundum atrial septal defects (Cordell, Bentham, et al., 2013). In our X-chromosome gene-level analyses, two of the genes with the lowest p-values (i.e., meta-$p < .01$) are associated with syndromes that include CHDs. Hence, although the statistical evidence linking these genes with CHDs is quite modest, our findings could help in the prioritization of potentially disease-related variants in CHD cases that are consistent with X-linked inheritance.

In combination with our prior GWAS of autosomal SNPs, the X-chromosome studies presented here provide a comprehensive assessment of common genomic variants in both CTDs and LVOTOs. However, because our sample sizes were relatively small and the power to detect X-linked variants is lower than that for autosomal variants (Chang et al., 2014), our analyses may have missed both SNP- and gene-level associations. For example, a case–control sample of ~8,522 cases and ~8,522 controls would be needed to achieve 80% power to detect a GWAS significant association ($p < 5 \times 10^{-8}$) with a variant of 5% MAF and an odds ratio of 1.5. Furthermore, since our analyses were restricted to common SNPs (MAF >5%), we cannot rule-out a potential role for rarer, X-linked variants. Additional studies addressing the potential role of X-linked genes in the etiology of CHDs are, therefore, warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P01HD070454); the National Heart, Lung, and Blood Institute (P50-HL74731), including the Pediatric Cardiac Genomics Consortium (U01-HL098188, U01HL131003, U01-HL098147, U01-HL098153, U01-HL098163, U01-HL098123, U01-HL098162, U01-HL-09003); the National Human Genome Research Institute number U54HG006504); the National Center for Research Resources (M01-RR-000240, RR024134; now the National Center for Advancing Translational Sciences, grant number UL1TR000003).

Funding information Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant/Award Number: P01HD070454; National Center for Advancing Translational Sciences, Grant/Award Number: UL1TR000003; National Center for Research Resources, Grant/Award Numbers: M01-RR-000240, RR024134; National Heart, Lung, and Blood Institute, Grant/Award Numbers: P50-HL74731, U01-HL-09003, U01-HL098123, U01-HL098147, U01-HL098153, U01-HL098162, U01-HL098163, U01-HL098188, U01HL131003; National Human Genome Research Institute, Grant/Award Number: U54HG006504

REFERENCES

Agopian AJ, Goldmuntz E, Hakonarson H, Sewda A, Taylor D, Mitchell LE, & Pediatric Cardiac Genomics C (2017). Genome-wide association studies and meta-analyses for congenital heart defects. Circulation. Cardiovascular Genetics, 10(3), e001449 10.1161/CIRCGENETICS.116.001449 [PubMed: 28468790]

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, & Lee JJ (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience, 4, 7 10.1186/s13742-015-0047-8 [PubMed: 25722852]

Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ, …Keinan A (2014). Accounting for eXentricities: Analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoim-mune diseases. PLoS One, 9(12), e113684 10.1371/journal.pone.0113684 [PubMed: 25479423]

Cordell HJ, Bentham J, Topf A, Zelenika D, Heath S, Mamasoula C, … Keavney BD (2013). Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. Nature Genetics, 45(7), 822–824. 10.1038/ng.2637 [PubMed: 23708191]

Cordell HJ, Topf A, Mamasoula C, Postma AV, Bentham J, Zelenika D, … Goodship JA (2013). Genome-wide association study identifies loci on 12q24 and 13q32 associated with tetralogy of Fallot. Human Molecular Genetics, 22(7), 1473–1481. 10.1093/hmg/dds552 [PubMed: 23297363]

de Leeuw CA, Mooij JM, Heskes T, & Posthuma D (2015). MAGMA: Generalized gene-set analysis of GWAS data. PLoS Compu-tational Biology, 11(4), e1004219 10.1371/journal.pcbi.1004219

Delaneau O, Zagury JF, & Marchini J (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nature Methods, 10(1), 5–6. 10.1038/nmeth.2307 [PubMed: 23269371]

Gao F, Chang D, Biddanda A, Ma L, Guo Y, Zhou Z, & Keinan A (2015). XWAS: A software toolset for genetic data analysis and association studies of the X chromosome. The Journal of Heredity, 106(5), 666–671. 10.1093/jhered/esv059 [PubMed: 26268243]

Howie BN, Donnelly P, & Marchini J (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genetics, 5(6), e1000529 10.1371/journal.pgen.1000529 [PubMed: 19543373]

Kolanczyk M, Krawitz P, Hecht J, Hupalowska A, Miaczynska M, Marschner K, … Horn D (2015). Missense variant in CCDC22 causes X-linked recessive intellectual disability with features of Ritscher-Schinzel/3C syndrome. European Journal of Human Genetics, 23(5), 633–638. 10.1038/ejhg.2014.109 [PubMed: 24916641]

Konig IR, Loley C, Erdmann J, & Ziegler A (2014). How to include chromosome X in your genome-wide association study. Genetic Epidemiology, 38(2), 97–103. 10.1002/gepi.21782 [PubMed: 24408308]

Lupo PJ, Mitchell LE, & Jenkins MM (2019). Genome-wide association studies of structural birth defects: a review and commentary. Birth Defects Research, 111, 1329–1342. [PubMed: 31654503]

Ng BG, Raymond K, Kircher M, Buckingham KJ, Wood T, Shendure J, … Freeze HH (2015). Expanding the molecular and clinical phenotype of SSR4-CDG. Human Mutation, 36(11), 1048–1051. 10.1002/humu.22856 [PubMed: 26264460]

Olcese C, Patel MP, Shoemark A, Kiviluoto S, Legendre M, Williams HJ, … Mitchison HM (2017). X-linked primary ciliary dyskinesia due to mutations in the cytoplasmic axonemal dynein assembly factor PIH1D3. Nature Communications, 8, 14279 10.1038/ncomms14279

Paff T, Loges NT, Aprea I, Wu K, Bakey Z, Haarman EG, … Micha D (2017). Mutations in PIH1D3 cause X-linked primary ciliary dyskinesia with outer and inner dynein arm defects. American Journal of Human Genetics, 100(1), 160–168. 10.1016/j.ajhg.2016.11.019 [PubMed: 28041644]

Shapiro AJ, Davis SD, Ferkol T, Dell SD, Rosenfeld M, Olivier KN, … Genetic Disorders of Mucociliary Clearance Consortium. (2014). Laterality defects other than situs inversus totalis in primary ciliary dyskinesia: Insights into situs ambiguus and het-erotaxy. Chest, 146(5), 1176–1186. 10.1378/chest.13-1704 [PubMed: 24577564]

Sumida H, Noguchi K, Kihara Y, Abe M, Yanagida K, Hamano F, … Ishii S (2010). LPA4 regulates blood and lymphatic vessel formation during mouse embryogenesis. Blood, 116(23), 5060–5070. 10.1182/blood-2010-03-272443 [PubMed: 20713964]

Tam V, Patel N, Turcotte M, Bosse Y, Pare G, & Meyre D (2019). Benefits and limitations of genome-wide association studies. Nature Reviews. Genetics, 20(8), 467–484. 10.1038/s41576-019-0127-1

Voineagu I, Huang L, Winden K, Lazaro M, Haan E, Nelson J, … Geschwind D (2012). CCDC22: A novel candidate gene for syn-dromic X-linked intellectual disability. Molecular Psychiatry, 17(1), 4–7. 10.1038/mp.2011.95 [PubMed: 21826058]

Wang F, Hou J, Han B, Nie Y, Cong X, Hu S, & Chen X (2012). Developmental changes in lysophospholipid receptor expression in rodent heart from near-term fetus to adult. Molecular Biology Reports, 39(9), 9075–9084. 10.1007/s11033-012-1778-6 [PubMed: 22740131]

Willer CJ, Li Y, & Abecasis GR (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26 (17), 2190–2191. 10.1093/bioinformatics/btq340 [PubMed: 20616382]

Wise AS, Shi M, & Weinberg CR (2015). Learning about the X from our parents. Frontiers in Genetics, 6, 15 10.3389/fgene.2015.00015 [PubMed: 25713581]

Yang L, Kraemer M, Fang XF, Angel PM, Drake RR, Morris AJ, & Smyth SS (2019). LPA receptor 4 deficiency attenuates experimental atherosclerosis. Journal of Lipid Research, 60(5), 972–980. 10.1194/jlr.M091066 [PubMed: 30796085]

Yukiura H, Hama K, Nakanaga K, Tanaka M, Asaoka Y, Okudaira S, … Aoki J (2011). Autotaxin regulates vascular development via multiple lysophosphatidic acid (LPA) receptors in zebrafish. Journal of Bio-logical Chemistry, 286(51), 43972–43983. 10.1074/jbc.M111.301093

TABLE 1

Summary of the conotruncal heart defect and left ventricular outflow tract datasets

			CHOP CTD case/control			
	CHOP CTD trios (*N* = 476)	**PCGC CTD trios (*N* = 244)**	**Cases (*N* = 403)**	**Controls (*N* = 2,974)**	**CHOP LVOTO trios (*N* = 243)**	**PCGC LVOTO trios (*N* = 141)**
Sex						
Male	289 (60.7)	151 (61.9)	235 (58.3)	1,492 (50.2)	149 (61.3)	98 (69.5)
Female	187 (39.3)	93 (38.1)	168 (41.7)	1,482 (49.8)	94 (38.7)	43 (30.5)
Conotruncal heart defects						
Tetralogy of Fallot	193 (40.6)	73 (29.9)	133 (33.0)	–	–	–
D-transposition of the great arteries	94 (19.8)	52 (21.3)	79 (19.6)	–	–	–
Ventricular septal defects	90 (19.0)	34 (13.9)	108 (26.8)	–	–	–
Double outlet right ventricle	52 (10.9)	37 (15.2)	25 (6.2)	–	–	–
Isolated aortic arch anomalies	22 (4.6)	6 (2.5)	22 (5.5)	–	–	–
Truncus arteriosus	14 (3.0)	7 (2.9)	19 (4.7)	–	–	–
Interrupted aortic arch	5 (1.1)	6 (2.5)	10 (2.5)	–	–	–
Other	5 (1.1)	28 (11.5)	7 (1.7)	–	–	–
Missing	1 (0.2)	0 (0.0)	0 (0.0)	–	–	–
Left ventricular outflow tract defects						
Hypoplastic left heart syndrome	–	–	–	–	119 (49.0)	63 (44.7)
Coarctation of the aorta	–	–	–	–	69 (28.4)	48 (34.0)
Aortic stenosis	–	–	–	–	55 (22.6)	20 (14.2)
Other	–	–	–	–	0 (0.0)	10 (7.1)

Abbreviations: CHOP, Children's Hospital of Philadelphia; CTD, conotruncal defect; LVOTO, left ventricular outflow tract obstruction; PCGC, Pediatric Cardiac Genomics Consortium.

TABLE 2

Summary of genes with association *p*-values <.01 in at least one meta-analysis

Meta-analysis group	Gene symbol	Gene name	Chromosomal location	*p*-Value
CTD				
	RBM41	RNA binding motif protein 41	Xq22.3	.0012
	NUP62CL	Nucleoporin 62	Xq22.3	.0018
	PIH1D3	PIH1 domain containing 3	Xq22.3	.0028
	MORC4	MORC family CW-type zinc finger 4	Xq22.3	.0068
LVOTO				
	SSR4	Signal sequence receptor subunit 4	Xq28	.0018
	PHKA1	Phosphorylase kinase regulatory subunit alpha 1	Xq13.1	.0028
	MAGEA1	MAGE family member A1	Xq28	.0038
	ARMCX4	Armadillo repeat containing X-linked 4	Xq22.1	.0056
	LPAR4	Lysophosphatidic acid receptor 4	Xq21.1	.0064
	BMP15	Bone morphogenetic protein 15	Xp11.2	.0066
	ARMCX6	Armadillo repeat containing X-linked 6	Xq22.1	.0068
	P2RY10	P2Y receptor family member 10	Xq21.1	.0072
	ACOT9	Acyl-coA thioesterase 9	Xp22.1	.0084
CTD + LVOTO				
	CLCN4	Chloride voltage-gated channel 4	Xp22.2	.0026
	CCDC22	Coiled-coil domain containing 22	Xp11.23	.0030
	PPP1R3F	Protein phosphatase 1 regulatory subunit 3F	Xp11.23	.0032
	IDS	Iduronate 2-sulfatase	Xq28	.0052
	SSR4	Signal sequence receptor subunit 4	Xq28	.0066
	TSPAN6	Tetraspanin 6	Xq22.1	.0066
	TNMD	Tenomodulin	Xq22.1	.0070

Abbreviations: CTD, conotruncal defect; LVOTO, left ventricular outflow tract obstruction.