Contribution of rare transmitted and *de novo* variants among 2,871 congenital heart disease probands

Sheng Chih Jin[1]*, Jason Homsy[2.3]*, Samir Zaidi[1]*, Qiongshi Lu[4], Sarah Morton[5], Steven R. DePalma[2], Xue Zeng[1], Hongjian Qi[6], Weni Chang[7], Wei-Chien Hung[1], Michael C. Sierant[1], Shozeb Haider[8], Junhui Zhang[1], James Knight[9], Robert D. Bjornson[9], Christopher Castaldi[9], Irina R. Tikhonoa[9], Kaya Bilguvar[9], Shrikant M. Mane[9], Stephan J. Sanders[10], Seema Mital[11], Mark Russell[12], William Gaynor[13], John Deanfield[14], Alessandro Giardini[14], George A. Porter Jr.[15], Deepak Srivastava[16,17,18], Cecelia W. Lo[19], Yufeng Shen[20], W. Scott Watkins[21], Mark Yandell[21,22], H. Joseph Yost[21], Martin Tristani-Firouzi[23], Jane W. Newburger[24], Amy E. Roberts[24], Richard Kim[25], Hongyu Zhao[4], Jonathan R. Kaltman[26], Elizabeth Goldmuntz[27], Wendy K. Chung[28], Jonathan G. Seidman[2‡], Bruce D. Gelb[29‡], Christine E. Seidman[2,3,30†‡], Richard P. Lifton[1,31†‡], Martina Brueckner[1,32†‡]

[1]Department of Genetics; Yale University School of Medicine, New Haven, CT, USA
[2]Department of Genetics, Harvard Medical School, Boston, MA, USA
[3]Cardiovascular Division, Brigham and Women's Hospital, Boston, MA, USA
[4]Department of Biostatistics; Yale School of Public Health, New Haven, CT, USA
[5]Division of Newborn Medicine, Department of Medicine, Boston Children's Hospital, Boston, USA
[6]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA
[7]Department of Pediatrics, Columbia University Medical Center, New York, NY, USA
[8]Department of Computational Chemistry, University College London School of Pharmacy, WC1N1AX, UK
[9]Yale Center for Genome Analysis, Yale University, New Haven, CT, USA
[10]Department of Psychiatry, University of California San Francisco, San Francisco, CA, USA
[11]Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada
[12]Division of Pediatric Cardiology, University of Michigan, Ann Arbor, MI, USA.

[13]Department of Pediatric Cardiac Surgery, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
[14]Department of Cardiology, University College London and Great Ormond Street Hospital, London, UK
[15]Department of Pediatrics, University of Rochester Medical Center, The School of Medicine and Dentistry, Rochester, NY, USA
[16]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA
[17]Roddenberry Stem Cell Center at Gladstone, San Francisco, CA 94158, USA
[18]Departments of Pediatrics and Biochemistry & Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA
[19]Department of Developmental Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15201, USA
[20]Departments of Systems Biology and Biomedical Informatics, Columbia University Medical Center, New York, NY, USA
[21]Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah and School of Medicine, Salt Lake City, UT, USA
[22]USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT, USA
[23]Division of Pediatric Cardiology, University of Utah, Salt Lake City, UT, USA
[24]Department of Cardiology, Boston Children’s Hospital, Boston, MA, USA
[25]Pediatric Cardiac Surgery, Children’s Hospital of Los Angeles, Los Angeles, CA, USA
[26]Heart Development and Structural Diseases Branch, Division of Cardiovascular Sciences, NHLBI/NIH, Bethesda, MD, USA
[27]Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[28]Departments of Pediatrics and Medicine, Columbia University Medical Center, New York, NY, USA
[29]Mindich Child Health and Development Institute and Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[30]Howard Hughes Medical Institute, Harvard University, Boston, MA, USA
[31]Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, NY, USA
[32]Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA
*These authors contributed equally to this work. †These authors contributed equally to this work. ‡Corresponding author. E-mail: seidman@genetics.med.harvard.edu (J.G.S.);

68 bruce.gelb@mssm.edu (B.D.G); cseidman@genetics.med.harvard.edu (C.E.S.);
69 richard.lifton@rockefeller.edu (R.P.L); martina.brueckner@yale.edu (M.B.).
70

ABSTRACT

Congenital heart disease (CHD) is the leading cause of mortality from birth defects. Exome sequencing of a single cohort of 2,871 CHD probands including 2,645 parent-offspring trios implicated rare transmitted mutations in 1.8%, including a recessive founder mutation in *GDF1* accounting for ~5% of severe CHD in Ashkenazim, recessive genotypes in *MYH6* accounting for ~11% of Shone complex, and dominant *FLT4* mutations accounting for 2.3% of Tetralogy of Fallot. *De novo* mutations (DNMs) accounted for 8% of cases, including ~3% of isolated CHD patients and ~28% with both neurodevelopmental and extra-cardiac congenital anomalies. Seven genes surpassed thresholds for genome-wide significance and 19 genes not previously implicated in CHD had > 70% probability of being disease-related; DNMs in ~440 genes are inferred to contribute to CHD. There was striking overlap between genes with damaging DNMs in probands with CHD and autism.

INTRODUCTION

Congenital heart disease (CHD) affects approximately 1% of all live births and remains the leading cause of mortality from birth defects world-wide[1]. Despite dramatic improvement in surgical and medical care, these patients remain at risk for developing cardiac arrhythmias and heart failure. In addition, CHD patients are enriched for extra-cardiac congenital anomalies (EA), and neurodevelopmental deficits[2,3]. While aneuploidies and copy number variations (CNVs) collectively account for ~23% of patients with CHD[4-6], few individual causal genes have been implicated from these CNVs, thwarting a detailed understanding of disease mechanisms. Rare Mendelian, syndromic forms of CHD have allowed the identification of some CHD genes. Nonetheless, the genes underlying the large majority of sporadic cases of CHD have not been well defined. Next generation sequencing allows new approaches to identify rare mutations with large effect in CHD[7].

To this end, the NHLBI Pediatric Cardiac Genomics Consortium (PCGC) has collected more than 10,000 probands with CHD, including over 5,000 parent-offspring trios[8]. Previous exome sequencing of 1,228 parent-offspring trios from this cohort has shown that ~10% of cases are attributable to *de novo* mutations (DNMs) in an estimated > 400 target genes, including dramatic enrichment for damaging mutations in genes encoding chromatin modifiers[9-11]. Moreover, these studies have demonstrated a striking shared genetic etiology between CHD and neurodevelopmental disorders (NDD)[6,10]. Nonetheless, few new individual genes have been definitively implicated owing to the high locus heterogeneity, and transmitted variants have not to date been comprehensively studied.

Genetic studies of humans and mice predict a role for transmitted variants with large effect[12,13]. Their discovery in humans has likely been limited by variations to the Mendelian paradigm such as high locus heterogeneity, markedly impaired reproductive fitness of individuals harboring risk genotypes, variable expressivity of mutant genotypes, and incomplete penetrance of mutant genotypes[12]. Detection of disease loci in this setting requires large cohorts of well-phenotyped subjects coupled with comprehensive genomic data and robust statistical methods.

Here, we present the first comprehensive analysis of the combined impact of rare recessive and dominantly transmitted variants, and DNMs on CHD identified from analysis of whole exome sequencing (WES) of a single large CHD cohort. The results

implicate new genes in CHD pathogenesis and expand the phenotypic spectrum associated with previously implicated genes.

RESULTS

Cohort Characteristics and Sequencing

We studied 2,871 probands with CHD comprising 2,645 parent-offspring trios and 226 singletons (**Supplementary Excel S1**). These include 1,204 previously reported trios[10] and 1,441 new trios. Patients were recruited to the PCGC and the Pediatric Heart Network (PHN) programs; cardiac and extracardiac phenotypic data on all probands were collected as previously described[8] (**Supplementary Table S1**). The ethnicities and gender of probands are shown in **Supplementary Table S2**. We excluded patients with clinically ascertained trisomies and CHD-associated CNVs. The distribution of cardiac lesions, extra-cardiac manifestations, and syndromes clinically diagnosed prior to enrollment are documented in **Supplementary Tables S3a-c**.

Genomic DNA from all trios and singletons underwent exome sequencing (see **Methods**). In parallel, sequence data from 1,789 control trios comprising unaffected siblings of autism probands and their parents was analyzed[14]. Cases and controls showed similar sequencing metrics with sufficient depth of coverage to make confident calls of DNMs and transmitted variants in the vast majority of targeted bases (**Supplementary Table S4**). Variants were called using Genome Analysis Toolkit (GATK) Haplotype Caller[15,16] and annotated for allele frequencies in Exome Aggregation Consortium (ExAC)[17], 1,000 Genomes[18] and Exome Variant Server (EVS). DNMs were called as described in methods[9,10]. The likelihood of missense variants being deleterious was inferred using MetaSVM[19].

Recessive Genotypes Enriched in CHD

We used genotype data to perform principal component analysis (PCA) and assess the inbreeding coefficient of all probands. CHD cases were more frequently of non-European ancestry and had a higher inbreeding coefficient compared to controls (**Supplementary Figure S1**). These differences affect expected recessive genotype (RG) frequencies, complicating direct comparison of cases and controls. Accordingly, we implemented a binomial test to quantify the enrichment of damaging RGs in a specific gene or gene set in cases, independent of controls. This method estimates the expected frequency of rare damaging variants in each gene from the *de novo* probability, then

determines the probability of the observed RGs in each gene occurring by chance, adjusting for the effect of inbreeding (see **Methods** and **Supplementary Figures S2-S6**).

We first tested whether genes previously implicated in CHD in humans or in human orthologs of genes implicated in recessive CHD in mouse harbored more damaging RGs than expected using the polynomial model. To this end, we prospectively curated a set of 212 human CHD genes (H-CHD genes) from the Online Mendelian Inheritance in Man (OMIM) and published data[11], and human orthologs of 61 mouse CHD genes (M-CHD genes) identified in a forward recessive screen for structural CHD (**Supplementary Excel S2**)[13]. The human gene set comprised 104 genes causing dominant CHD, 85 recessive genes, 12 X-linked genes, and 11 genes with evidence for both dominant and recessive transmission. After accounting for 20 genes identified in both human and mouse studies, the combined set comprised 253 unique human genes (**Supplementary Excel S2**).

We identified rare (allele frequency < 0.001) likely loss-of-function (LoF; frameshift, nonsense, canonical splice site, and start loss), likely damaging missense variants (by MetaSVM; D-Mis), and non-frameshift insertion/deletion variants, and identified homozygous or compound heterozygous genotypes comprising these alleles. This identified 467 putative damaging RGs in CHD cases (**Supplementary Excel S3**) and 165 in controls (**Supplementary Excel S4**).

We then used the one-tailed binomial test to determine whether damaging RGs were enriched among 96 genes previously implicated in recessive CHD in humans (**Table 1a**). This gene set showed many more damaging RGs than expected (observed 29 vs. expected 6.7, enrichment = 4.4, $P = 8.0 \times 10^{-11}$; **Table 1a**, **Supplementary Figure S5b**, **Supplementary Table S5**). In contrast, this gene set showed zero RGs in controls (**Table 1b**). After addition of the 41 recessive genes unique to mouse, there were 34 damaging RGs in known recessive human or mouse genes compared to 11.1 expected (enrichment = 3.1, $P = 1.4 \times 10^{-8}$; **Table 1a**).

We next added 116 genes previously implicated in monoallelic CHD. This showed an additional 17 damaging RGs in 9 genes (cumulative total, observed 51 vs. expected 25.2, enrichment = 2.0, $P = 1.8 \times 10^{-6}$; **Table 1a**). Similar results were obtained from constructing an expected distribution of RG without using polynomial fits by independently modeling homozygous and compound heterozygous genotypes (see **Methods**, **Supplementary Table S6**, and **Supplementary Figures S7-S8**). These findings are further corroborated using a burden test-based approach[20,21] that also

Table 1a. Damaging recessive genotypes in known CHD genes in 2,871 CHD cases

Gene set (# genes)	Observed				Expected	Enrichment	P-value
	# homozygotes	# compound heterozygous	# unique genes	# recessive genotypes	# recessive genotypes		
All genes (18,989)	265	202	391	467	-	-	-
Recessive Known Human (96)	19	10	16	29	6.65	**4.36**	**8.0×10^{-11}**
Recessive Known Mouse or Human (137)	21	13	19	34	11.06	**3.07**	**1.4×10^{-8}**
Known Mouse or Human CHD (253)	28	23	28	51	25.15	**2.03**	**1.8×10^{-6}**

Table 1b. Damaging recessive genotypes in known CHD genes in 1,789 controls

Gene set (# genes)	Observed				Expected	Enrichment	P-value
	# homozygotes	# compound heterozygous	# unique genes	# recessive genotypes	# recessive genotypes		
All genes (18,989)	22	131	146	165	-	-	-
Recessive Known Human (96)	0	0	0	0	2.61	0	1
Recessive Known Mouse or Human (137)	1	1	2	2	4.47	0.45	0.94
Known Mouse or Human CHD (253)	2	3	5	5	10.18	0.49	0.98

The expected number of recessive genotypes was determined based on fitted values from the polynomial regression model using the damaging *de novo* probabilities. P-values were calculated using the one-tailed binomial probability. Values in bold are p-values exceeding the Bonferroni multiple testing cutoff = $0.05/(3\times2) = 8.3\times10^{-3}$

integrates proband phenotype information[22] (see **Methods** and **Supplementary Figure S9**). These findings collectively provide strong evidence that that RGs in known CHD genes contribute to CHD in 0.9% of cases in this cohort.

We examined the contribution of consanguinity to recessive genotypes in this cohort. 161 probands (5.6% of the total sequenced cohort) had homozygous segments spanning at least 0.35% of the genetic map, consistent with these probands being the offspring of the union of 3rd cousins or closer relationships (median 1.4% homozygosity for these probands, see **Methods**). We note that this group included 81 of 84 probands with a reported history of consanguinity, indicating that self-report was quite specific (96%) but not highly sensitive (50.3%) for consanguinity. 8.1% of these 161 consanguineous probands had damaging RGs in known recessive human genes (13 observed vs. 2.4 expected, 5.4-flold enrichment, $P = 1.3 \times 10^{-6}$; **Supplementary Table S7**). As expected, nearly all of these RGs (12) were homozygotes. In contrast, among the remaining 2710 probands, RGs were also highly significantly enriched (3.9-fold, 16 observed vs. 4.1 expected, $P = 5.3 \times 10^{-6}$), however RG's were found in only 0.6% of this group (**Supplementary Table S7**). Of the seven homozygotes in this group, five had inbreeding coefficients between 0.0015 and 0.0035, implying more distant parental relatedness, whereas two homozygotes and all nine compound heterozygotes in this group had inbreeding coefficients of zero. Thus, the probability that a RG contributes to CHD is strongly influenced by parental consanguinity. Similarly, 38% of RGs in known recessive CHD genes in our cohort were attributable to a *GDF1* founder mutation (see below). Significant enrichment for RGs in known CHD genes persists after removal of recurrences of *GDF1* homozygotes (**Supplementary Table S8**).

To search for genes not previously associated with recessive CHD, we identified genes with > 1 damaging RGs. We observed a total of 44 such genes compared to 26.4 expected (enrichment = 1.7; $P = 8.9 \times 10^{-5}$ by permutation; see **Methods**); in contrast, there was no significant enrichment for genes with > 1 synonymous RG (observed = 167; expected = 156.7, P = 0.15 by permutation). This excess persisted after removal of 5 genes (*GDF1, ATIC, DNAH5, DAW1, LRP1*) previously implicated in recessive CHD (enrichment = 1.6; $P = 10^{-3}$ by permutation). GO ontology of this set revealed significant enrichment of terms involved in muscle cell development (GO:0055001, enrichment = 29.5, FDR = 3.2×10^{-3}). Genes in this GO term include *KEL*, *MYH6*, *MYH11*, *NOTCH1*, and *RYR1* (**Supplementary Excel S3** and **S5**).

Founder Mutation in *GDF1* in Ashkenazim

We compared the observed number of damaging RGs in each gene to the expected distribution using the binomial test as described above. Quantile-quantile (Q-Q) plots showed that while the observed distribution of P-values closely followed the expected distribution for nearly all genes, two genes, *GDF1* and *MYH6*, departed from the expected distribution and surpassed thresholds for genome-wide significance ($P < 2.6x10^{-6} = 0.05/18,989$ tests) in cases (**Figure 1a**; **Supplementary Table S9**). Modeling of homozygous and compound heterozygous genotypes separately yielded similar results (**Supplementary Table S10**). In contrast, no genes approached genome-wide significance in controls (**Figure 1b**).

GDF1 had 11 damaging RGs in apparently unrelated subjects compared with 0.016 expected (enrichment = 692.6, one-tailed binomial $P = 3.6x10^{-28}$; **Supplementary Table S9**). All 11 genotypes were confirmed by Sanger sequencing (**Supplementary Figure S10**). Ten of the eleven cases harbored the identical homozygous p.M364T variant, suggesting a founder mutation. The other *GDF1* RG was p.364_364del/p.C227X. Consistent with a founder mutation, PCA showed that all ten p.M364T homozygotes clustered with Ashkenazim (**Supplementary Figure S11**), and the allele was absent among African, Asian, and Finnish European populations in ExAC.

Multiple additional lines of evidence support this homozygous p.M364T genotype having a large effect on CHD risk among Ashkenazim. The p.M364T variant shows remarkable violation of Hardy Weinberg equilibrium among Ashkenazi CHD cases, with 10 homozygotes and only 1 heterozygote among 204 Ashkenazi cases defined by PCA ($P = 5.5x10^{-38}$, 1-df chi-square test with Yate's correction; **Supplementary Table S11a**). In contrast, among 302 Ashkenazi autism parental controls and 926 additional Ashkenazi adults from an independent cohort without CHD, there were no homozygotes and only 12 heterozygotes (carrier frequency = 1.0%), providing strong evidence of association of homozygosity for p.M364T with CHD among Ashkenazim (two-sided Fisher's Exact $P = 2.8x10^{-9}$, **Supplementary Table S11b**). Lastly, all homozygotes shared the p.M364T variant on a common haplotype background, supporting identity by descent from a shared ancestor (**Figure 2a**). The extent of the shared haplotype varied widely in length (0.4-5.9 Mb; **Figure 2a**), indicating the absence of recent shared ancestry, and the minimum shared haplotype among all affected subjects was only 234 kb. From this, the inferred coalescent time for the last shared ancestor is 50 generations

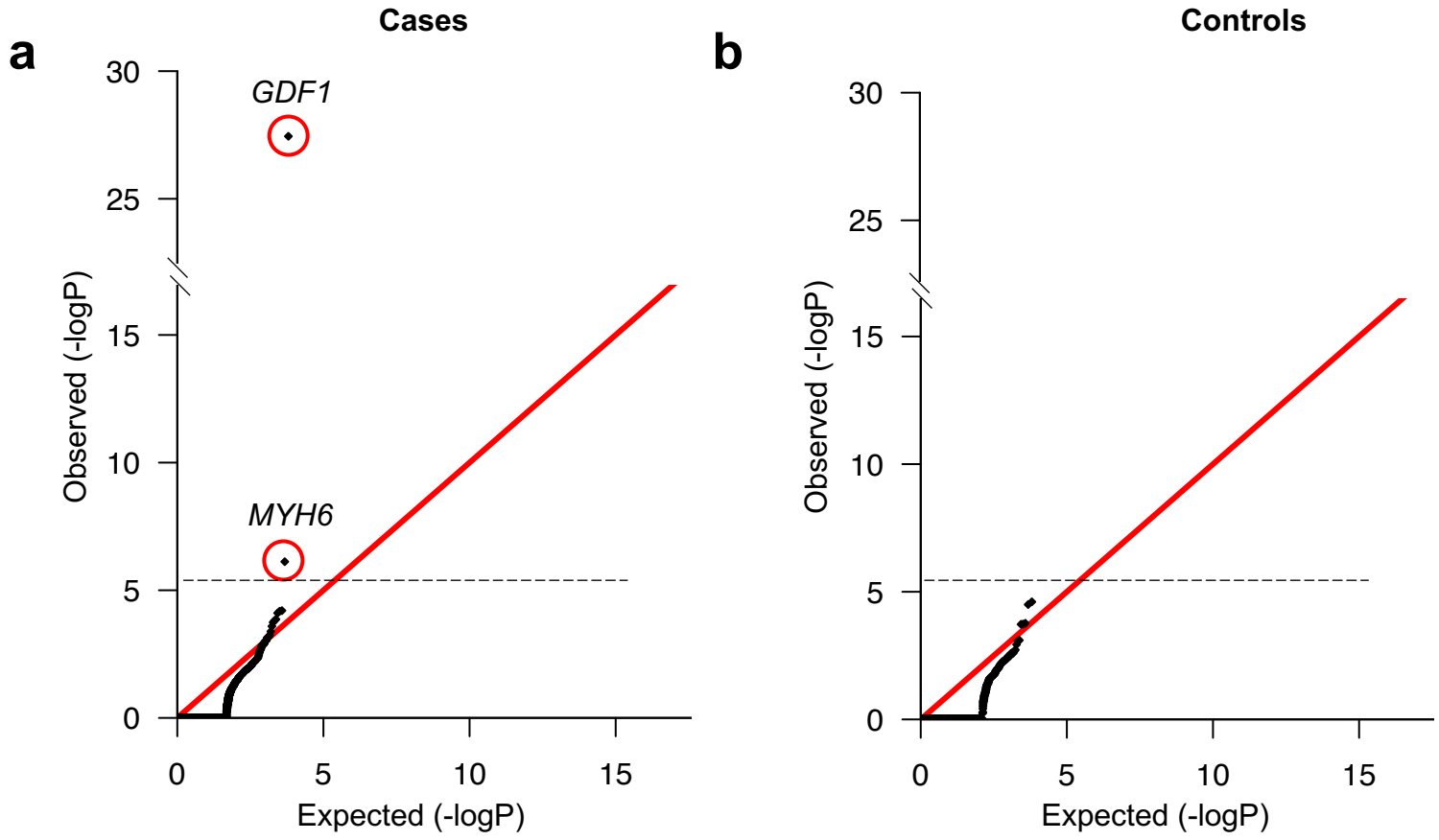
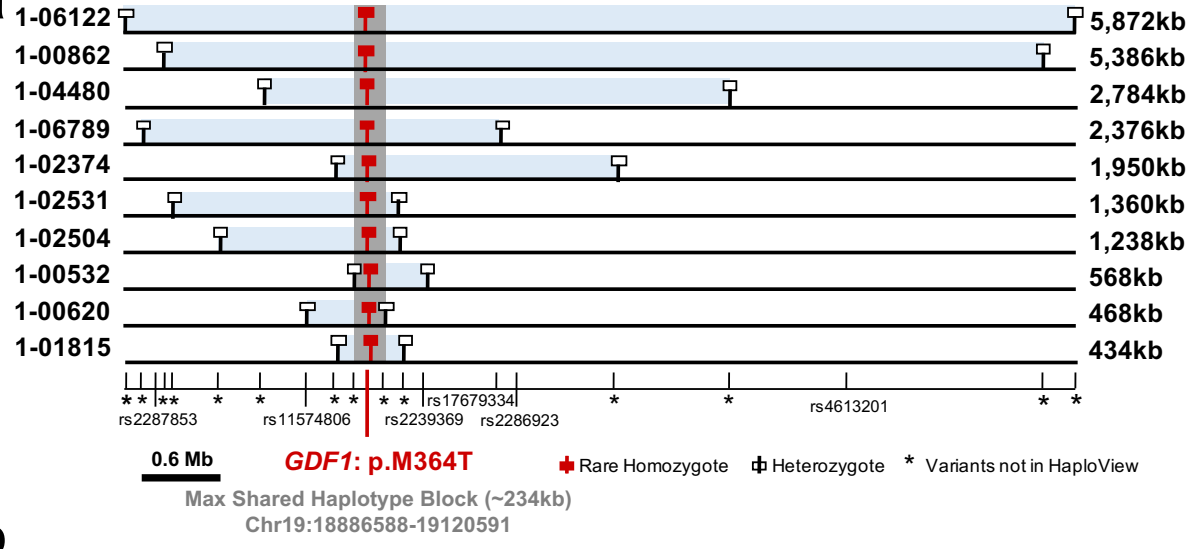


Figure 1. Quantile-quantile plots comparing observed versus expected P-values for recessive genotypes in each gene in cases and controls. Recessive genotypes (RGs) shown include LoF, D-Mis, and non frameshift insertion/deletions. The expected number of RGs in each gene was calculated from the total number of observed RGs as described in Methods. The significance of the difference between the observed and expected number of RGs was calculated using a one-sided binomial test. (**a**). Quantile-quantile (Q-Q) plot in cases. (**b**). Q-Q plot in controls. While the observed values closely conform to expected values in controls, two genes, *GDF1* and *MYH6*, show a significantly increased burden of RGs in cases and survive the multiple-testing correction threshold.

a

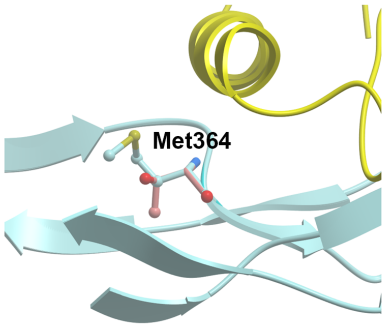


b

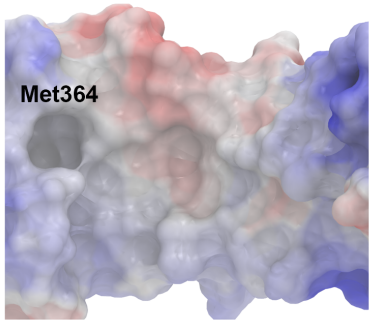
Detailed phenotypes for patients carrying homozygous *GDF1*-M364T mutation

ID	DORV	TGA	PS/PA	Other cardiac phenotypes	Extracardiac	NDD
1-06122	+	-	+	-		-
1-00862	-	L	+	VSD	Inguinal hernia	+
1-04480	-	L	+	-		+
1-06789	-	-	+	TOF		-
1-02374	-	D	+	-		NA
1-02531	+	D	+	TAPVR		-
1-02504	-	-	+	TOF		-
1-00532	-	L	-	-		+
1-00620	-	D	-	-	Inguinal hernia	-
1-01815	-	D	+	-		NA

c



d



e

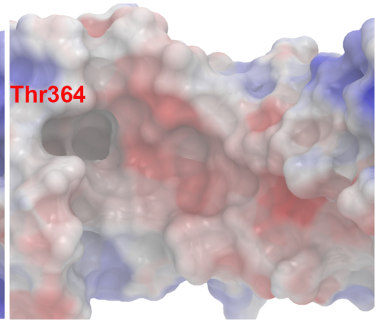


Figure 2. Phenotypes and shared haplotypes among homozygotes for *GDF1*-p.M364T. (**a**). Extent of homozygous SNPs flanking homozygous *GDF1*-p.M364T genotypes. A 5.9 Mb segment of chromosome 19 extending across the location of the homozygous *GDF1*-p.M364T variant (denoted by red square) in each unrelated subject is depicted. At the bottom, tick marks indicate location of all SNPs found by exome sequencing among Ashkenazim in cases . Known SNPs are shown via their rs identifiers. Allele frequencies of novel SNPs are indicated by asterisks. The closest heterozygous SNP to either side of the *GDF1*-p.M364T in each subject is shown as a white square; all SNPs between these two heterozygous SNPs, encompassed by the light blue bar, are homozygous for the same allele seen in other subjects, consistent with the p.M364T variant being identical by descent among all subjects. The length of each homozygous segment is indicated at the right of the panel. The maximum length of the homozygous segment shared by all subjects is 234 kb (shown as grey vertical bar), consistent with the mutation having been introduced into a shared ancestor many generations ago. (**b**). Cardiac and extracardiac phenotypes of *GDF1*- p.M364T homozygotes. (**c**). Ribbon diagram of part of GDF1 homodimer containing p.M364. The hydrophobic helix from one subunit (yellow) sits above p.M364 on the other subunit (blue). (**d**). Space filling model of the segment of GDF1 containing the wild-type p.M364 showing surface electrostatic charge (blue=positive, red=negative). (**e**). Surface electrostatic charge of the segment containing mutant p.T364. Compared to wild-type, the mutant peptide shows a more negatively charged cavity.

as calculated by the DMLE+2.3 software[23] (95% CI: 45 to 63 generations; **Supplementary Figure S12**).

Also consistent with this genotype causing CHD and not merely being in linkage disequilibrium with another causal variant, the phenotype of the 10 p.M364T homozygotes is very similar to that previously described in patients with different recessive *GDF1* mutations[24,25,24]. Like prior cases, all *GDF1* p.M364T homozygotes had D- or L-transposition of the great arteries and pulmonary stenosis/atresia (or both D/L-TGA and PS/PA) (**Figure 2b**). GDF1 belongs to the transforming growth factor-beta (TGF-ß) superfamily. Previous studies of *Gdf1* in mouse showed a critical role in left-right asymmetry in embryonic development and in neural development[25-27]. GDF1 acts as a homodimer with two-fold inverted symmetry (**Figure 2c** and **Supplementary Figure S13**). The interaction surface between monomers comprises a hydrophophic α-helix (residues 316-327) from one monomer and a shallow cavity formed by hydrophobic residues from the adjacent monomer; this interaction occurs reciprocally (*i.e.*, twice per dimer). Met^{364} lies at the floor of the hydrophobic cavity (**Figure 2d-e**). The p.M364T substitution introduces the polar side chain of threonine into the hydrophobic cavity; in p.M364T homozygotes, we infer that both interaction surfaces between monomers are destabilized, impairing dimer formation and downstream signaling (**Figure 2c**). This mechanism is consistent with the observed recessive transmission.

Collectively, these findings identify a recessive *GDF1* founder mutation accounting for ~5% of severe CHD among Ashkenazim studied, including 18% of those with TGA (7 of 38), and 31% with TGA plus PS/PA (5 of 16). This finding has direct clinical implications for assessing risk of CHD among Ashkenazim.

Recessive *MYH6* Genotypes in Shone Complex

MYH6 encodes the alpha cardiac heavy chain, a developmentally regulated protein that is most highly expressed in the embryonic heart. Dominant mutations in *MYH6* have been shown to cause atrial septal defect[28] and cardiomyopathy[29,30]. We identified seven rare damaging RGs in *MYH6* compared with 0.482 expected (enrichment = 14.5, $P = 7.6 \times 10^{-7}$; **Supplementary Table S9**). These included diverse and very rare likely LoF alleles and D-Mis variants comprising five compound heterozygotes and two homozygotes, all validated by Sanger sequencing (**Table 2**, **Supplementary Table S9**, and **Supplementary Figure S14**). Five probands had left ventricular obstruction, including four meeting diagnostic criteria for Shone complex[31],

Table 2. Recessive *MYH6* genotypes associated with Shone complex and valvular disease.

ID	AA Change	ExAC Ethnic Specific Freq	Shone complex	Detailed Cardiac Phenotype	Cardiac Function	Extracardiac	NDD	Age at follow-up
1-00051	p.K1932X/p.A1891T	3.0×10^{-5}/ 0	+	LSVC, abn MV, sub AS, valve AS, CoA	LV diastolic dysfunction		+ (LD)	22
1-01407	p.E98K	3.0×10^{-4}		mitral atresia, DORV, CoA	mild RV systolic dysfunction	Hypothyroid	+ (LD)	16
1-04847	p.R1899H/p.N598fs	0/ 0	+	parachute MV, BAV, CoA	NL		-	16
1-05009	p.A1327V/p.L388F	2.7×10^{-3}/ 0		TA, PA	dilated, hyper-trabeculated LV		NA	0
1-06399	p.G585S/p.I512T	2.0×10^{-4}/ 3.0×10^{-5}	+	mitral stenosis, VSD, BAV, hypoplastic transv. Ao	NL		NA	0.08
1-06876	p.I1068T/ c.3979-2A>C	1.5×10^{-5}/ 2.0×10^{-5}	+	LSVC, abn mitral valve, valve AS, CoA	dilated LV		-	22
1-07343	p.R1610C	3.0×10^{-5}		ASD/VSD	NA		NA	NA

Abbreviations: ASD- Atrial septal defect, AS- Aortic stenosis, BAV- Bicuspid aortic valve, CoA- Coarctation of the aorta, DORV- Double outlet right ventricle. MV- mitral valve, PA-Pulmonary atresia, TA-Tricuspid atresia, VSD-Ventricular septal defect. Extracardiac manifestations refer to CHD probands displaying additional abnormalities not pertaining to the heart. NDD- neurodevelopmental disabilities (+, -), LD-Learning Disability, NA – NDD status not attained as proband < age 1.

having mitral valve and aortic valve obstruction in addition to aortic arch obstruction (**Table 2**). These *MYH6* genotypes accounted for 11% of the 37 patients with Shone complex in the sequenced cohort (enrichment = 57.45, two-sided Fisher's exact $P=6.7x10^{-5}$).

Recessive Genotypes Enriched in Patients with Laterality Defects

Among the major clinical CHD subgroups (laterality defects, left ventricular obstruction, conotruncal defects and others, outlined in **Supplementary Table S3a**), only laterality defects (heterotaxy and D-TGA) were enriched for damaging RGs in the known human and mouse CHD gene set. These included 21 damaging RGs in 13 genes compared to 4.8 expected RGs (enrichment = 4.4, $P = 8.5x10^{-9}$; **Supplementary Table S12**). Significant enrichment was observed even after removing *GDF1* RGs (enrichment = 3.2, $P = 1.2x10^{-4}$). These genes included eight that have been previously implicated in laterality defects (*ARMC4*, *BBS10*, *DAW1*, *DNAAF1*, *DNAH5*, *DYNC2H1*, *GDF1*, and *PKD1L1*) and five that have not (*ATIC*, *COL1A1*, *COL5A2*, *DGCR2*, and *MYH6*).

We performed GO ontology analysis of all 82 genes with LoF RGs observed in patients with any cardiac phenotype. This identified significant terms related to cilia structure and regulation, a predominant mechanism in laterality determination (**Supplementary Excel S6**). Genes in these GO terms included *DNAI2*, *ARMC4*, *DNAH5*, and *DNAAF1* (proband phenotypes in **Supplementary Excel S3**). Although all four of these genes have been associated with human primary ciliary dyskinesia and situs inversus totalis, and *Armc4* mutation leads to a range of CHD in mouse, only *DNAH5* has been previously associated with human CHD[32].

Heterozygous LoF Mutations in *FLT4* in Tetralogy of Fallot

We next compared the frequency of rare (MAF ≤ 10^{-5}) heterozygous LoF variants in 115 known dominant CHD genes in cases and controls to the number expected by chance using the binomial test (**Supplementary Excel S7-S8**). We found no significant enrichment in cases (enrichment = 1.1, P = 0.32; **Supplementary Table S13a**) or controls (enrichment = 0.7, P = 1; **Supplementary Table S13b**). Analysis of heterozygous LoF variants in all 212 known human CHD genes also showed no enrichment.

To search for novel haploinsufficient CHD genes, we compared the observed distribution of rare heterozygous LoFs in each gene to the values expected by chance

across the genome (see **Methods**). Q-Q plots (**Supplementary Figure S15**) showed that the observed distribution closely conformed to the expected in cases and controls with the exception of one gene, *FLT4,* which harbored eight different rare transmitted LoFs in cases, (enrichment = 15.5, P = 7.6×10^{-8}, **Supplementary Table S14**). In addition, there were two *de novo FLT4* LoF mutations, yielding a combined p-value of 9.8×10^{-10} (p-values combined by the Fisher's method, **Figure 3**). All *FLT4* LoF variants were confirmed by Sanger sequencing (**Supplementary Figure S16**).

FLT4 was highly intolerant to LoF mutation in the ExAC database (pLI = 1) and only one LoF allele in *FLT4* was identified among 3,578 parental controls. There was no significant variation in coverage of coding bases across *FLT4* exons between cases and controls. Examination of the pedigrees of the ten cases with *FLT4* LoFs revealed four other affected family members with CHD and the *FLT4* LoF mutation found in the proband (**Figure 3a**). Interestingly, however, *FLT4* mutations showed incomplete penetrance, as only 4 of the 10 mutation-carrier relatives had CHD by report (estimated 40% penetrance).

FLT4 is a VEGF receptor known to be expressed in lymphatics and the vasculature and missense mutations affecting the kinase domain have been associated with hereditary lymphedema (**Fig. 3b**)[33]. Strongly supporting a pathogenic role for these *FLT4* mutations, the phenotype of mutation carriers was highly homogeneous, with 9 of 10 probands and 3 of their 4 affected relatives having a diagnosis of tetralogy of Fallot (TOF) (**Figure 3**). Among the 426 probands with TOF in our cohort, *FLT4* mutations were found in 2.3% of TOF cases and were enriched 95.2-fold among TOF probands; this enrichment was highly unlikely to occur by chance (P = 1.9×10^{-12}; **Supplementary Table S15**).

De Novo Damaging Mutations Enriched in CHD Cases

DNMs were identified in 2,645 parent-offspring case trios and 1,789 control trios, totaling 2,990 DNMs in cases (**Supplementary Excel S9**) and 1,830 in controls (**Supplementary Excel S10**). All had strong statistical support and were verified by *in silico* visualization of aligned reads. The observed number closely fit the expected Poisson distribution in cases and controls (**Supplementary Figure S17**). While CHD cases showed no enrichment for *de novo* synonymous or MetaSVM-tolerated (T-Mis) mutations, there was highly significant enrichment of *de novo* damaging (LoF and D-Mis) mutations (enrichment = 1.4, P = 2.4×10^{-17}, **Supplementary Table S16**). From the

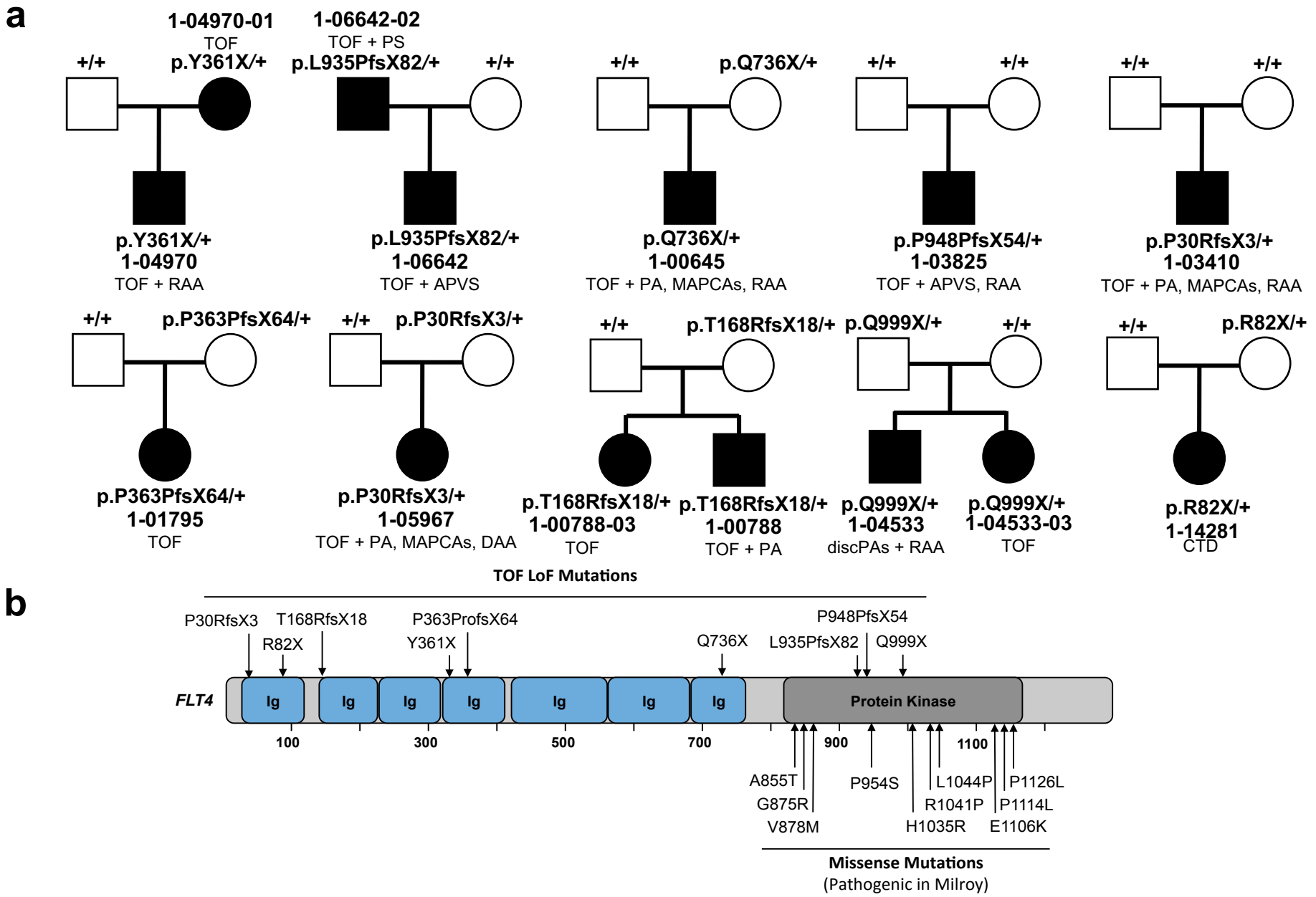


Figure 3. *FLT4* loss-of-function mutations in Tetralogy of Fallot (**a**). Pedigrees of 10 CHD kindreds with rare *FLT4* loss-of-function (LoF) mutations are shown. Subjects with and without CHD are shown as filled and unfilled symbols, respectively. Each kindred ID number is shown along with the *FLT4* genotype of each subject and CHD phenotype of affected subjects. (**b**) Diagram of FLT4 protein is shown with seven immunoglobulin domains (Ig) and a kinase domain. The top panel shows LoF mutations associated with Tetralogy-type CHD, whereas the bottom panel displays missense mutations associated with the Milroy disease (Hereditary Lymphedema).

difference in the fraction of patients with damaging DNMs compared to expectation, we inferred that DNMs contribute to CHD in ~8.3% of cases.

Expansion of the CHD cohort from previous analyses identified 66 genes with two or more damaging *(*LoF and D-Mis) DNMs (**Figure 4a**, **Supplementary Table S17-S18**), including eight genes not previously implicated in CHD: *AKAP12*, *ANK3*, *CTNNB1*, *FRYL*, *KDM5B*, *NAA15*, *POGZ*, and *PRRC2B*. Further, CHD cases were markedly enriched for damaging mutations in 104 known dominant human CHD genes (enrichment = 9.3, P = 5.5×10^{-65}; **Supplementary Table S19**). Similar findings were observed using an orthogonal method of disease gene discovery (see **Methods** and **Supplementary Figure S18**).

We previously implicated DNMs in chromatin modifiers in CHD[9,10]. We now find 89 damaging DNMs in 46 chromatin modifiers (including 58 LoFs; Figure 4b and **Supplementary Table S20**), (enrichment = 3.1, P = 8.7×10^{-20}; **Supplementary Table S21**). Seventeen of these genes have not previously been found mutated. Collectively, damaging DNMs in chromatin modifiers are inferred to contribute to 2.3% of cases (**Supplementary Table S21**). Gene Ontology enrichment analysis using all genes harboring at least one *de novo* LoF mutation genes expressed in the top quartile of the 14.5 mouse heart (HHE) gene set + LoF-intolerant gene set implicated the covalent chromatin modification gene set (GO:0016569; enrichment = 9.37, FDR q = 5.8×10^{-9}) and other terms including histone modification (**Supplementary Excel S11**).

De Novo Mutations Enriched in Isolated CHD

We evaluated the contribution of DNMs to patients with isolated CHD, CHD with extracardiac malformation, CHD with NDD, and CHD with both EA and NDD, focusing on HHE gene set[9] (**Supplementary Table S22a**). Unlike prior studies[9,10], we found a significant enrichment of damaging DNMs in isolated CHD cases (defined as CHD without any extracardiac congenital anomalies, clinically diagnosed syndrome or neurodevelopmental abnormality, and limited to patients over age 1 for whom NDD questionnaires were completed at entry into the study), contributing to ~3.1% of cases (1.5-fold enrichment, P = 8.5×10^{-4}; **Supplementary Table S22a**). We further estimated that damaging DNMs in the known CHD genes can account for ~50% (13/26) of the excess burden of mutations in isolated CHD. By comparison, DNMs contributed to 6% - 8% of probands with extracardiac features (either CHD + EA alone or CHD + NDD alone), and to a very high fraction of cases with CHD + EA + NDD, 28% of cases (95%

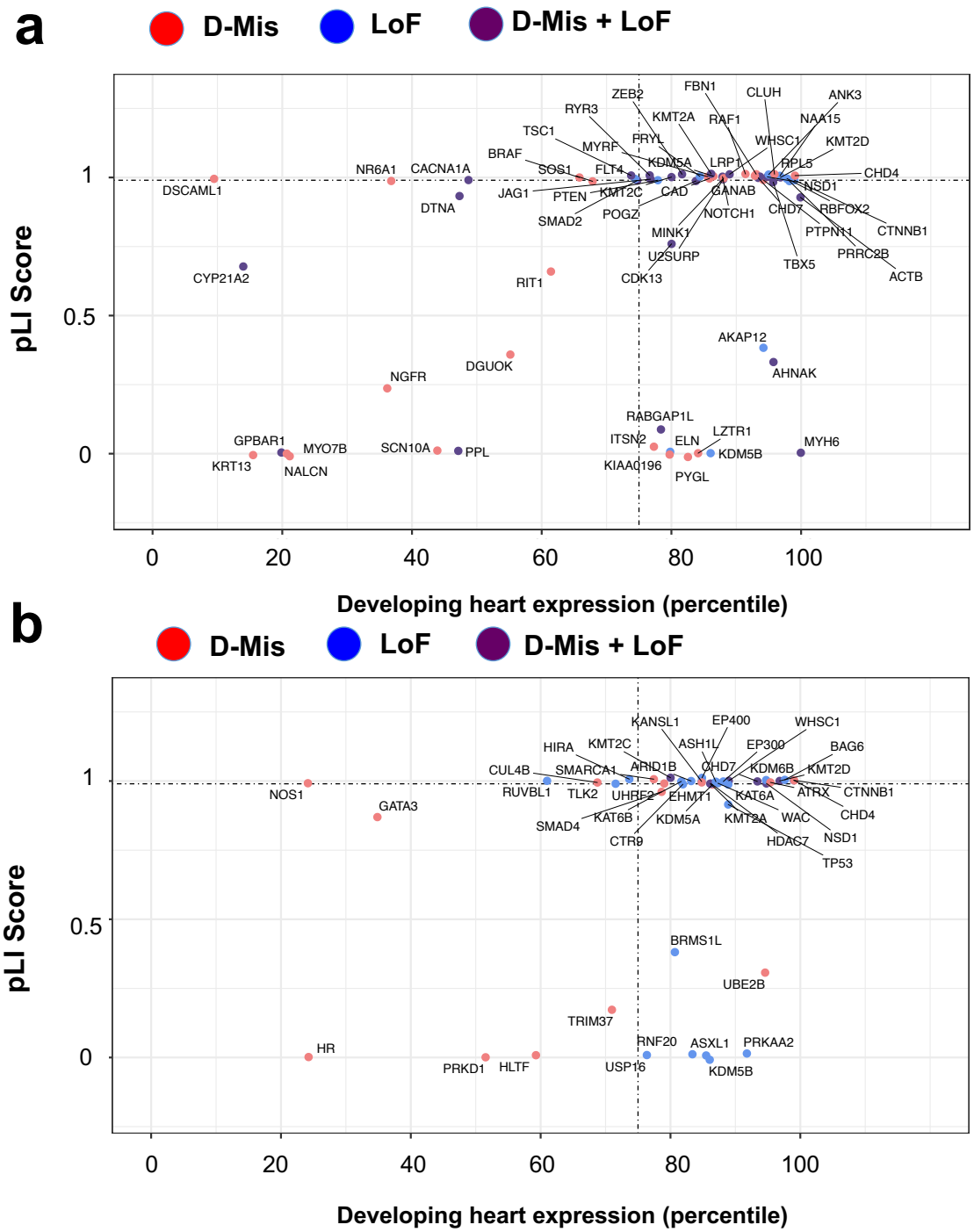


Figure 4. Genes with multiple damaging *de novo* mutations and in chromatin modification genes are enriched for high expression in developing heart and intolerance to loss-of-function mutation. (a) 66 genes with 2 or more damaging *de novo* mutations are plotted for percentile rank of heart expression in developing mouse heart at E14.5 (x axis, 0-100) and intolerance to loss-of-function (LoF) mutation (pLI) in the ExAC database (y axis, 0-1.0). Multihit genes are highly enriched (N=31) for genes that are highly expressed in developing heart and intolerant to loss-of-function mutation (pLI ≥ 0.99). (b) Enrichment of damaging mutations in chromatin modifiers in genes highly expressed in developing heart and intolerant to loss-of-function mutation.

CI: 19.6%-36.7%, 5-fold enrichment, P = 1.6×10^{-29}; **Supplementary Tables S22a-d** and **S23**).

***De novo* mutations are Enriched in Autism-Associated Genes.**

We previously showed unexpected overlap of genes harboring damaging DNMs between CHD and neurodevelopmental disorders[9,10]. With exome sequencing completed for two large cohorts of well-phenotyped autism patients comprising 4,778 parent-offspring trios[34,35], we compared the genes harboring damaging DNMs in our CHD cohort and these autism cohort trios[34,35]. There was highly significant overlap of these genes (**Supplementary Table 24a**), driven by genes with high expression in both developing heart and brain. This included 19 genes with LoF mutations in both cohorts (enrichment 5.2, $P < 10^{-6}$) and 48 genes with damaging mutations in both (enrichment 2.8, $P < 10^{-6}$; **Supplementary Table 24b**). Notably, 67% (21/31) of CHD patients with available neurodevelopmental data harboring LoF DNMs in the overlapping gene set had NDD, compared to 32.8% in the total cohort with certain NDD status; OR = 4.3; two-sided Fisher's $P = 1.4 \times 10^{-4}$; **Supplementary Table S25**). Three of the CHD cases with mutations in the overlapping gene set had an autism diagnosis at entry into the study. Notably, 14/35 genes with LoF DNMs in both the CHD and autism cohorts are chromatin modifiers (enrichment =14.7, $P < 10^{-6}$ by permutation; **Supplementary Table S25**). Most strikingly, 87% of all patients who were evaluated for NDD and have LoF DNMs in chromatin modifiers had an NDD diagnosis at study entry.

Meta-Analysis of Damaging *De Novo and* Loss-of-function Heterozygous Variants

We tested each gene for an excess of *de novo* and transmitted heterozygous variants using Fisher's method to combine the P-values from damaging *de novo* and LoF heterozygous LoF variants (degrees of freedom = 4)**.** Seven genes passed genome wide significance ($P < 2.6 \times 10^{-6}$ [0.05/18,989 genes]; **Table 3**). These included genes that were largely driven by damaging DNMs (*CHD7, KMT2D, PTPN11, and RBFOX2*) and three genes that predominantly had transmitted LoF mutations (*FLT4*, *SMAD6, and NOTCH1*) (**Table 3**). Of note, 12 of the top 25 (48%) genes are known CHD genes including *CHD7, KMT2D, PTPN11, NOTCH1, SMAD6, GATA6, ELN, PTEN, RPL5, NSD1, NODAL,* and *SOS1*. Those among the top 25 genes that have not previously been linked to CHD risk but are HHE and LoF-intolerant include *SMAD2* (a transducer of TGF-beta signaling), and *FRYL* (a transcriptional activator).

Table 3. Top 25 genes in the meta-analysis of damaging *de novo* mutations and loss-of-function heterozygous mutations in probands

Gene	Damaging *de novo*		LoF heterozygotes		Meta P-value	pLI	HHE Rank	Gene Set
	# Damaging	P-value	# LoF	P-value				
CHD7	**14**	**1.6×10^{-20}**	**0**	**1**	**7.5×10^{-19}**	**1**	**93.4**	**H-CHD/Chromatin**
KMT2D	**16**	**2.1×10^{-20}**	**1***	**0.86**	**8.5×10^{-19}**	**1**	**96.8**	**H-CHD/Chromatin**
PTPN11	**9**	**4.6×10^{-17}**	**0**	**1**	**1.8×10^{-15}**	**1**	**94.2**	**H-CHD**
FLT4	**2**	**5.2×10^{-4}**	**8**	**7.6×10^{-8}**	**9.8×10^{-10}**	**1**	**74.4**	**NA**
NOTCH1	**5**	**2.7×10^{-5}**	**6***	**1.8×10^{-4}**	**9.4×10^{-8}**	**1**	**87.9**	**H-CHD**
RBFOX2	**3**	**3.4×10^{-7}**	**1***	**0.18**	**1.1×10^{-6}**	**0.99**	**97.8**	**NA**
SMAD6	**1**	**0.012**	**8**	**6.0×10^{-6}**	**1.3×10^{-6}**	**0**	**78.3**	**M-CHD**
GATA6	4	2.4×10^{-7}	0	1	3.8×10^{-6}	N/A	94.8	H-CHD
ELN	2	1.3×10^{-4}	5*	8.7×10^{-3}	1.7×10^{-5}	0	79.8	H-CHD
CCDC154	0	1	7*	5.5×10^{-6}	7.2×10^{-5}	0.31	18.4	NA
SLCO1B3	0	1	9	6.6×10^{-6}	8.5×10^{-5}	0	11.7	NA
GPBAR1	2	2.6×10^{-5}	1	0.27	9.1×10^{-5}	0	19.9	NA
PTEN	2	6.0×10^{-5}	1	0.16	1.2×10^{-4}	0.98	77.9	H-CHD
RPL5	2	6.2×10^{-5}	1	0.16	1.3×10^{-4}	0.99	97.9	H-CHD
NSD1	5	1.0×10^{-5}	0	1	1.3×10^{-4}	1	94.8	H-CHD/Chromatin
SAMD11	2	1.8×10^{-4}	4*	0.06	1.4×10^{-4}	0	N/A	NA
C21ORF2	0	1	5	1.2×10^{-5}	1.5×10^{-4}	0.01	46.7	NA
NODAL	0	1	4	1.2×10^{-5}	1.5×10^{-4}	0.95	16.4	H-CHD
SMAD2	3	5.5×10^{-5}	1	0.24	1.6×10^{-4}	0.99	74.7	NA
H1FOO	0	1	4	1.6×10^{-5}	1.9×10^{-4}	0.4	10.3	NA
FRYL	2	2.8×10^{-3}	5*	8.3×10^{-3}	2.8×10^{-4}	1	84.4	NA
KDM5B	3	2.9×10^{-5}	2*	0.86	2.9×10^{-4}	0	86	Chromatin
POGZ	3	2.5×10^{-5}	0	1	2.9×10^{-4}	1	83.8	Chromatin
SOS1	3	2.6×10^{-5}	0	1	3.0×10^{-4}	1	67.9	H-CHD
TBX18	1	0.02	3	1.8×10^{-3}	3.0×10^{-4}	1	72.6	NA

Meta-analysis was performed by combining the p-values from damaging *de novo* mutations and loss-of-function (LoF) heterozygous mutations using the Fisher's method with 4 degrees of freedom. The top 25 genes are shown. Genes which are bolded surpass the Bonferroni multiple testing correction (2.6×10^{-6}, 0.05/18,989) for p-values tabulated by either *de novo,* heterozygous, or meta-analysis. H-CHD: Known human CHD genes. M-CHD: Known mouse CHD genes. Chromatin: Chromatin modification genes consists of 546 genes in GO:0016569.* denotes that at least one of the carriers has unknown transmission.

SMAD6, an inhibitor of BMP signaling, had 8 transmitted and one *de novo* LoF (Meta P = $1.3x10^{-6}$; **Table 3**). Patients with *SMAD6* mutations had TOF, hypoplastic left heart syndrome, coarctation and D-TGA. Zero LoF variants were noted in 7,156 control alleles (OR = Inf, one-sided Fisher's Exact P = $6.8x10^{-4}$). Case-control analysis comparing 2,063 CHD European cases (7 mutation carriers) with 30,216 ExAC non-Finnish European controls (5 carriers) also showed marked enrichment (OR = 20.5, two-sided Fisher's P = $2.7x10^{-6}$). One parent with a transmitted LOF *SMAD6* allele had a BAV, and the remainder did not have a history of CHD, and while parents did not have cardiac imaging, unrecognized severe CHD would be unlikely. Monoallelic *SMAD6* missense variants have been previously identified in three sporadic cases of bicuspid aortic valve and mitral valve disease[36], though the statistical relevance of these mutations to CHD was not established. Interestingly, LoF mutations in *SMAD6* with incomplete penetrance have also been implicated in midline craniosynostosis, with a common variant near *BMP2* modifying penetrance[37]. Our findings suggest that the phenotype resulting from *SMAD6* mutation is highly variable, dependent on additional genetic or environmental factors.

DISCUSSION

This study represents the largest genetic investigation of CHD, and the first analysis of the combined contribution of transmitted and *de novo* variants to CHD. By more than doubling the size of the studied cohort, many additional genes have been implicated. Our search for disease-associated transmitted variants and pathways was markedly enhanced by implementation of a gene burden analysis that estimates the expected number of recessive or dominant genotypes independent of control subjects and which accommodates variation in inbreeding and ethnic background. While the explicit extension of the expected frequency of DNMs to standing variation can be confounded by the impact of selection and drift on allele frequencies over subsequent generations, our analysis demonstrates that this approach is robust for estimating the expected frequency of rare transmitted variants, which are more likely to be newly introduced into the population. This approach will have application to many studies searching for the contribution of transmitted rare variants to human traits in the absence of family linkage data.

Rare transmitted genotypes accounted for at least 1.8% of CHD in this cohort. This number is an underestimate because it only takes into account the impact of rare mutations in known CHD genes plus additional genes that have reached genome-wide significance. Moreover, these genotypes are likely underrepresented in the general population owing to purifying selection, so the expected number is likely lower than the calculated frequency. Included in this group was a founder mutation in *GDF1* that accounted for 5% of all severe CHD in the Ashkenazi population studied, and an even higher percentage of Ashkenazi cases with TGA (18%), particularly those who also have PS/PA (31%). Genotyping for this specific variant, which has an allele frequency of ~0.6% in Ashkenazim, can immediately be used for diagnosis and population-based risk assessment.

It is interesting to note that enrichment of damaging RGs was most significant in patients with laterality defects. The epidemiology of laterality defects suggests a contribution by RGs: they have the highest recurrence risk within families of any CHD[12] , are more prevalent in populations with high consanguinity[38], and conversely have no enrichment of damaging DNMs[9,10].

We also found new recessive phenotypes arising from genes previously implicated in CHD caused by monoallelic mutations, including RGs in *MYH6* in 11% of Shone complex cases. Shone complex features multiple levels of left-ventricular outflow obstruction, posing a challenge for surgical management[39], and no clear genetic cause has yet been identified. *MYH6* compound heterozygous mutations have been identified in two patients with another left-ventricular obstructive lesion, hypoplastic left heart syndrome (HLHS) associated with reduced ventricular ejection fraction[40]. Further, a link was identified between monoalleleic variants in *MYH6* and decreased transplant-free survival in patients with HLHS[41]. This suggests that patients with Shone complex due to biallelic *MYH6* mutation may be at particular risk of developing ventricular dysfunction and their early identification may allow intervention leading to improved outcome. Other genes without previously described recessive phenotypes included *CHD7*, *COL1A1*, *COL5A2*, *FBN2*, *NOTCH1*, *NSD1*, and *TSC2,* as well as *DGCR2*, and *DAW1*, *LRP1*, and *MYH10*, which previously had been implicated only in mouse CHD.

Analysis of rare dominant variants strongly implicated LoF variants distributed throughout the type 3 VEGF receptor *FLT4* and showed that they predominantly result in TOF. Among 10 probands with *FLT4* LoFs, none had NDD and only 1 had EA, unlike 25% of all TOF probands in this study who have NDD and/or EA. *FLT4* LoF mutations

are distinct from reported heterozygous missense mutations in the *FLT4* kinase domain that impair kinase activity and result in defective lymphatic development[42]. It is, thus, clear that loss of FLT4 kinase function is not equivalent to a null allele. FLT4 functions that are kinase-independent may be preserved in the kinase-mutant alleles but lost in null alleles. Further studies of the expression and role of *FLT4* in the developing heart will be of interest.

Doubling the size of our sequenced cohort more than doubled the identified CHD risk genes. The current data set includes 66 genes with two or more damaging DNMs compared to 21 previously, and 19 with two or more LoF DNMs compared to five previously[10]. Among the most highly enriched gene sets in which 72%-85% of genes are expected to confer risk (i.e. 17 HHE genes with two or more LoF DNMs; 31 HHE + intolerant genes with two or more damaging mutations; 11 chromatin modifiers with two or more damaging DNMs), we identify 19 unique genes (*AKAP12*, *ANK3*, *CAD*, *CLUH*, *CTNNB1*, *FRYL*, *GANAB*, *KDM5A*, *KDM5B*, *KMT2C*, *MINK1*, *MYRF*, *NAA15*, *POGZ*, *PRRC2B*, *RBFOX2*, *RYR3*, *U2SURP*, and *WHSC1*) that have not been implicated in CHD prior to our studies of this CHD cohort. DNMs are highly enriched in cases with neurodevelopmental abnormalities and extra-cardiac structural manifestations, contributing to more than a quarter of these cases. Importantly, we report for the first time a significant contribution of DNMs to isolated CHD, occurring in 3.1% of cases. From the distribution of genes with multiple damaging DNMs, the estimate of the number of genes in which DNMs contribute to CHD in this cohort is 443 (95% CI = [154.1, 731.9]; **Supplementary Figure S19**; see **Methods**).

Pathway analysis identified distinct biological mechanisms underlying *de novo* and inherited CHD. As previously reported, chromatin modifiers represent the top terms associated with DNMs contributing to CHD. Eleven chromatin modifiers have two or more damaging DNMs, and we estimate from a maximum likelihood approach (see **Methods**) that ~38 genes in this pathway contribute to CHD (**Supplementary Figure S20**). The implication of LoF DNMs in writers, erasers and readers of many different specific chromatin marks underscores the importance of dosage sensitivity of these genes. Together these findings suggest that heart development depends on precise control of transcription mediated by changes in chromatin state in response to developmental signals[43-45]. After removing chromatin modifiers from GO term enrichment analysis, several terms broadly involved in developmental processes show enrichment (**Supplementary Excel S12**). Extension of pathway analysis to genes with damaging

RGs demonstrated enrichment of genes involved in cilia formation and function. These genes have long been known to play a critical role in establishment of the left-right body axis in early development, and mutations in this pathway play a frequent role in development of heterotaxy. Understanding the mechanisms underlying the effects of these biological pathways will be of great interest in determining mechanisms of normal and abnormal human development.

It will become important to link the genetic causes of CHD to patient outcomes. We report striking overlap of genes mutated in CHD and autism. Genes mutated in both are enriched > 2.7-fold among genes that are highly expressed in both developing heart and brain. In particular, patients in our cohort with LoF mutations in chromatin modifiers are at very high risk of NDD (87%). In contrast, none of our patients with *FLT4* mutation and TOF had NDD, which underscores the impact of specific mutations on risk of CHD patients having NDD/autism. These observations identify a set of CHD genes that may permit presymptomatic identification of patients with CHD who are at high risk for autism, and would be good candidates for early neurodevelopmental intervention[46]. Conversely, virtually all patients with LoF mutations in chromatin modifiers who have been ascertained for autism studies do not have CHD[35], indicative of variable expressivity of CHD, and raising the possibility that NDD phenotypes may also have variable expressivity.

Despite these advances, the pathogenesis of a large fraction of CHD cases remains unknown. Potential explanations include contributions from more common variants, structural variants that have eluded detection by WES, variants in the non-coding regions, polygenic inheritance and possible gene-environment interactions[6,37,47,48] . Monte Carlo simulations suggest that two or more damaging DNMs have been identified in ~10.5% of risk loci, and that sequencing 10,000 trios will yield 170.1 risk genes, predicting 38% saturation of all CHD risk genes acting by DNMs (**Supplementary Figure S21**). It is clear that loci suggested from human studies can be further substantiated at low cost by orthogonal approaches engineering mutations into model organisms and cells[49]. This study indicates that continued sequencing of large, well-phenotyped cohorts will provide an increasingly complete picture of the genetic underpinnings of CHD, allowing new insight into mechanisms governing human development, improved prediction of clinical outcome, and the opportunity to mitigate these risks.

ACKNOWLEDGMENTS

The authors are enormously grateful to the patients and families who participated in this research. We thank the following team members for outstanding contributions to patient recruitment: A.Julian, M.Mac Neal, Y.Mendez, T.Mendiz-Ramdeen, C.Mintz (Icahn School of Medicine at Mount Sinai); N.Cross (Yale School of Medicine); J.Ellashek and N.Tran (Children's Hospital of Los Angeles); B.McDonough, A.Monafo, J.Stryker (Harvard Medical School), K.Flack, L.Panesar, N.Taylor (University College London); E.Taillie (University of Rochester School of Medicine and Dentistry); S.Edman, J.Garbarini, J.Tusi, S.Woyciechowski, (Children's Hospital of Philadelphia); D.Awad, C.Breton, K.Celia, C.Duarte, D.Etwaru, N.Fishman, M.Kaspakoval, J.Kline, R.Korsin, A.Lanz, E.Marquez, D.Queen, A.Rodriguez, J.Rose, J.K.Sond, D.Warburton, A.Wilpers, and R.Yee (Columbia Medical School). We are grateful to Joseph Ekstein and Dor Yeshorim for provision of anonymized DNA samples. The authors thank Shiuan Wang for critical discussion.

This work was supported by the U01 HL098153 and Grant UL1TR000003 from the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, grants to the Pediatric Cardiac Genomics Consortium (U01-HL098188, U01-HL098147, U01-HL098153, U01-HL098163, U01-HL098123 and U01-HL098162), the NIH Centers for Mendelian Genomics (5U54HG006504), the Howard Hughes Medical Institute (RPL and CES) and the Simons Foundation (WKC). SCJ was supported by the James Hudson Brown-Alexander Brown Coxe Postdoctoral Fellowship at the Yale University School of Medicine. JH was supported by the John S. LaDue Fellowship at Harvard Medical School and is a recipient of the Alan Lerner Research Award at the Brigham and Women's Hospital. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Heart, Lung, and Blood Institute, the National Center for Research Resources or the NIH.

AUTHOR CONTRIBUTIONS

Study design: M.B., W.K.C., M.T.F., B.D.G., E.G., J.R.K., R.P.L., J.G.S., C.E.S.

Cohort ascertainment, phenotypic characterization and recruitment: M.B., W.C., W.K.C., J.D., A.G., B.D.G., E.G., W.G., J.H., R.K., S.M., J.W.N., G.A.P., A.E.R., M.R., C.E.S.

Exome sequencing production and validation: K.B., C.C., R.P.L., S.M.M., I.R.T., J.Z.

Exome sequencing analysis: M.B., R.D.B., S.R.DP., S.C.J., J.H., W.C.H., J.K., R.P.L., S.M., S.M.M., H.Q., C.E.S., J.G.S., M.C.S., S.J.S., Y.S., W.S.W., M.Y., S.Z., X.Z.

Statistical analysis: J.H., S.C.J., R.P.L., Q.L., S.M., C.E.S., S.W., M.Y., H.Z., S.Z.

S.H. performed the biophysical simulation for *GDF1*.

Writing and review of manuscript: M.B., W.K.C., M.T.F., B.D.G., E.G., J.H., S.C.J., J.R.K., C.W.L., R.P.L., Q.L., C.E.S., D.S., J.G.S., J.Y., S.Z.

Co-senior authors: M.B., R.P.L., C.E.S.

All authors read and approved the manuscript.

COMPETING FINANCIAL INTERESTS

None

REFERENCES

1. van der Linde, D. *et al.* Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol* **58**, 2241-7 (2011).
2. Egbe, A., Lee, S., Ho, D., Uppu, S. & Srivastava, S. Prevalence of congenital anomalies in newborns with congenital heart disease diagnosis. *Ann Pediatr Cardiol* **7**, 86-91 (2014).
3. Marino, B.S. *et al.* Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. *Circulation* **126**, 1143-72 (2012).
4. Soemedi, R. *et al.* Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am J Hum Genet* **91**, 489-501 (2012).
5. Glessner, J.T. *et al.* Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ Res* **115**, 884-96 (2014).
6. Zaidi, S. & Brueckner, M. Genetics and Genomics of Congenital Heart Disease. *Circ Res* **120**, 923-940 (2017).
7. Ng, S.B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-3 (2010).
8. Pediatric Cardiac Genomics, C. *et al.* The Congenital Heart Disease Genetic Network Study: rationale, design, and early results. *Circ Res* **112**, 698-706 (2013).
9. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220-3 (2013).
10. Homsy, J. *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262-6 (2015).
11. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet* **48**, 1060-5 (2016).
12. Oyen, N. *et al.* Recurrence of congenital heart defects in families. *Circulation* **120**, 295-301 (2009).
13. Li, Y. *et al.* Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature* (2015).
14. Krumm, N. *et al.* Excess of rare, inherited truncating mutations in autism. *Nat Genet* **47**, 582-8 (2015).
15. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
16. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 1-33 (2013).
17. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
18. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
19. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-37 (2015).

20. Hu, H. *et al.* VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genet Epidemiol* **37**, 622-34 (2013).
21. Yandell, M. *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Res* **21**, 1529-42 (2011).
22. Singleton, M.V. *et al.* Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet* **94**, 599-610 (2014).
23. Reeve, J.P. & Rannala, B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics* **18**, 894-5 (2002).
24. Kaasinen, E. *et al.* Recessively inherited right atrial isomerism caused by mutations in growth/differentiation factor 1 (GDF1). *Hum Mol Genet* **19**, 2747-53 (2010).
25. Lee, S.J. Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure. *Proc Natl Acad Sci U S A* **88**, 4250-4 (1991).
26. Rankin, C.T., Bunton, T., Lawler, A.M. & Lee, S.J. Regulation of left-right patterning in mice by growth/differentiation factor-1. *Nat Genet* **24**, 262-5 (2000).
27. Tanaka, C., Sakuma, R., Nakamura, T., Hamada, H. & Saijoh, Y. Long-range action of Nodal requires interaction with GDF1. *Genes Dev* **21**, 3272-82 (2007).
28. Ching, Y.H. *et al.* Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet* **37**, 423-8 (2005).
29. Hershberger, R.E. *et al.* Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ Cardiovasc Genet* **3**, 155-61 (2010).
30. Niimura, H. *et al.* Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation* **105**, 446-51 (2002).
31. Ikemba, C.M. *et al.* Mitral valve morphology and morbidity/mortality in Shone's complex. *Am J Cardiol* **95**, 541-3 (2005).
32. Harrison, M.J., Shapiro, A.J. & Kennedy, M.P. Congenital Heart Disease and Primary Ciliary Dyskinesia. *Paediatr Respir Rev* **18**, 25-32 (2016).
33. Karkkainen, M.J. *et al.* Missense mutations interfere with VEGFR-3 signalling in primary lymphoedema. *Nat Genet* **25**, 153-9 (2000).
34. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-15 (2014).
35. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-21 (2014).
36. Tan, H.L. *et al.* Nonsynonymous variants in the SMAD6 gene predispose to congenital cardiovascular malformation. *Hum Mutat* **33**, 720-7 (2012).
37. Timberlake, A.T. *et al.* Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. *Elife* **5**(2016).
38. Shieh, J.T., Bittles, A.H. & Hudgins, L. Consanguinity and the risk of congenital heart disease. *Am J Med Genet A* **158A**, 1236-41 (2012).
39. Krieger, E.V. & Fernandes, S.M. Heart failure caused by congenital left-sided lesions. *Heart Fail Clin* **10**, 155-65 (2014).

40. Theis, J.L. *et al.* Recessive MYH6 Mutations in Hypoplastic Left Heart With Reduced Ejection Fraction. *Circ Cardiovasc Genet* **8**, 564-71 (2015).
41. Tomita-Mitchell, A. *et al.* Impact of MYH6 variants in hypoplastic left heart syndrome. *Physiol Genomics* **48**, 912-921 (2016).
42. Kaipainen, A. *et al.* Expression of the fms-like tyrosine kinase 4 gene becomes restricted to lymphatic endothelium during development. *Proc Natl Acad Sci U S A* **92**, 3566-70 (1995).
43. Wamstad, J.A. *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206-20 (2012).
44. Paige, S.L. *et al.* A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221-32 (2012).
45. Ang, S.Y. *et al.* KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. *Development* **143**, 810-21 (2016).
46. Razzaghi, H., Oster, M. & Reefhuis, J. Long-term outcomes in children with congenital heart disease: National Health Interview Survey. *J Pediatr* **166**, 119-24 (2015).
47. Oyen, N. *et al.* Prepregnancy Diabetes and Offspring Risk of Congenital Heart Disease: A Nationwide Cohort Study. *Circulation* **133**, 2243-53 (2016).
48. Morishima, M., Yasui, H., Ando, M., Nakazawa, M. & Takao, A. Influence of genetic and maternal diabetes in the pathogenesis of visceroatrial heterotaxy in mice. *Teratology* **54**, 183-90 (1996).
49. Zhu, J.Y., Fu, Y., Nettleton, M., Richman, A. & Han, Z. High throughput in vivo functional validation of candidate congenital heart disease genes in Drosophila. *Elife* **6**(2017).