



# Genomic analysis of 11,555 probands identifies 60 dominant congenital heart disease genes

Michael C. Sierant<sup>ab</sup>, Sheng Chih Jin<sup>b,c,d</sup>, Kaya Bilguvar<sup>a,e,f,g,h,i</sup>, Sarah U. Morton<sup>h,k,l</sup> , Weilai Dong<sup>a,b</sup> , Wei Jiang<sup>m</sup>, Ziyu Lu<sup>n</sup>, Boyang Li<sup>m</sup>, Francesc López-Giráldez<sup>e</sup> , Irina Tikhonova<sup>e</sup>, Xue Zeng<sup>a,b</sup>, Qiongshi Lu<sup>o</sup> , Jungmin Choi<sup>a,b,p</sup> , Junhui Zhang<sup>a</sup> , Carol Nelson-Williams<sup>a</sup>, James R. Knight<sup>e</sup>, Hongyu Zhao<sup>am</sup> , Junyue Cao<sup>n</sup>, Shrikant Mane<sup>e</sup>, Stanley C. Sedore<sup>qr</sup>, Peter J. Gruber<sup>s</sup>, Monkol Lek<sup>2</sup>, Elizabeth Goldmuntz<sup>t</sup>, John Deanfield<sup>u</sup>, Alessandro Giardini<sup>v</sup>, Seema Mital<sup>w</sup> , Mark Russell<sup>x</sup>, J. William Gaynor<sup>y</sup>, Eileen King<sup>z</sup>, Michael Wagner<sup>aa,bb</sup>, Deepak Srivastava<sup>cc</sup>, Yufeng Shen<sup>dd,ee</sup>, Daniel Bernstein<sup>ff</sup> , George A. Porter Jr.<sup>g,gg</sup> , Jane W. Newburger<sup>hh,ii</sup>, Jonathan G. Seidman<sup>jj</sup> , Amy E. Roberts<sup>hh,ii</sup>, Mark Yandell<sup>kk</sup>, H. Joseph Yost<sup>kk,ll</sup> , Martin Tristani-Firouzi<sup>mm</sup>, Richard Kim<sup>nn</sup>, Wendy K. Chung<sup>oo,pp</sup>, Bruce D. Gelb<sup>qq,rr</sup>, Christine E. Seidman<sup>ss,tt</sup> , Martina Brueckner<sup>a,q,1</sup> , and Richard P. Lifton<sup>b,1</sup>

Affiliations are included on p. 11.

Contributed by Richard P. Lifton; received October 16, 2024; accepted February 18, 2025; reviewed by Jay Shendure and Matthew State

**Congenital heart disease (CHD) is a leading cause of infant mortality. We analyzed de novo mutations (DNMs) and very rare transmitted/unphased damaging variants in 248 prespecified genes in 11,555 CHD probands. The results identified 60 genes with a significant burden of heterozygous damaging variants. Variants in these genes accounted for CHD in 10.1% of probands with similar contributions from de novo and transmitted variants in parent–offspring trios that showed incomplete penetrance. DNMs in these genes accounted for 58% of the signal from DNMs. Thirty-three genes were linked to a single CHD subtype while 12 genes were associated with 2 to 4 subtypes. Seven genes were only associated with isolated CHD, while 37 were associated with 1 or more extracardiac abnormalities. Genes selectively expressed in the cardiomyocyte lineage were associated with isolated CHD, while those widely expressed in the brain were also associated with neurodevelopmental delay (NDD). Missense variants introducing or removing cysteines in epidermal growth factor (EGF)-like domains of NOTCH1 were enriched in tetralogy of Fallot and conotruncal defects, unlike the broader CHD spectrum seen with loss of function variants. Transmitted damaging missense variants in *MYH6* were enriched in multiple CHD phenotypes and account for ~1% of all probands. Probands with characteristic mutations causing syndromic CHD were frequently not diagnosed clinically, often due to missing cardinal phenotypes. CHD genes that were positively or negatively associated with development of NDD suggest clinical value of genetic testing. These findings expand the understanding of CHD genetics and support the use of molecular diagnostics in CHD.**

congenital heart disease | molecular inversion probes | exome sequencing | human genetics | genomics

Congenital heart disease (CHD) is one of the most common congenital malformations, affecting 1.0 to 1.8% of all live births (1, 2). One-third of patients require intervention within the first year of life, and while ~90% will now survive to adulthood, many will have lifelong comorbidities including neurodevelopmental disabilities, arrhythmia, and heart failure (3, 4).

Recurrence of CHD in families (5, 6), Mendelian forms of CHD, and CHD with aneuploidy, recurring chromosomal deletions and duplications, established a genetic contribution to CHD (7). Whole exome sequencing (WES) in parent–offspring trios by the Pediatric Cardiac Genomics Consortium (PCGC) demonstrated that damaging de novo mutations (DNMs) contribute to CHD (8–11). DNMs were highly enriched for loss of function (LOF) mutations in genes intolerant to LOFs (high pLI), consistent with haploinsufficiency, with gain of function (GOF) mutations in a few genes such as *PTPN11* (12). The role of transmitted variants has been less well-studied (10, 12). Results suggest that DNMs in hundreds of genes contribute to CHD, with only seven genes implicated as having a significant burden of DNMs after the study of 2,645 trios (10). Genes involved in chromatin modification were highly enriched for DNMs (8–10).

The need for much larger cohorts is evident. Knowledge from prior results can help prioritize a subset of genes with increased likelihood of being CHD genes (13) that can be selectively sequenced using molecular inversion probe sequencing (MIPseq) at very low cost (14–16). A challenge of MIPseq has been achieving complete coverage of targeted bases (15).

## Significance

We identified 60 genes with significant burden of monoallelic damaging variants, accounting for 10.1% of probands, with equal contributions from de novo and transmitted variants. Mutations often produced variable congenital heart disease (CHD) and extracardiac (EC) phenotypes. Mutation frequency was variable and not explained by mutability. Probands with mutations associated with syndromic CHD were frequently not clinically diagnosed, often due to the absence of characteristic phenotypes. Genes with predominantly transmitted variants were enriched for isolated CHD and depleted for extracardiac features including neurodevelopmental disorders (NDD). There was wide variation in risk of NDD (e.g., 4% for *MYH6* mutation vs. 95% for *CHD7*). Results support increased use of molecular diagnosis to assess risk and allow early intervention to potentially improve outcomes.

Competing interest statement: M.Y. is a co-founder and sultan to Fabric Genomics. M.Y. has stock interests exceeding \$5,000 in value in Fabric Genomics.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: [martina.brueckner@yale.edu](mailto:martina.brueckner@yale.edu) or [rickl@rockefeller.edu](mailto:rickl@rockefeller.edu).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2420343122/-/DCSupplemental>.

Published March 24, 2025.

We report herein analyses of a targeted panel of 248 genes in 11,555 CHD probands from the PCGC and Pediatric Heart Network (PHN) cohorts who have extensive cardiac and extra-cardiac (EC) clinical phenotyping and have undergone WES or MIPseq, increasing the number of CHD probands fourfold from our most recent study (10). We identify 60 genes with significant enrichment of damaging variants that contribute to CHD in 10.1% of probands, with nearly equal contributions from de novo and transmitted variants. We show striking variable expressivity of phenotypes resulting from variants in many of these genes.

## Results

**A MIPseq Panel for Candidate CHD Genes.** From earlier analysis of WES in 1,213 parent–offspring trios (9) and prior reports of genes associated with CHD, we selected 248 CHD genes to target for further study (*SI Appendix, Fig. S1, Dataset S1, and Materials and Methods*). We synthesized and optimized 10,154 barcoded molecular inversion probes to capture, sequence, and annotate the entire coding sequence of each gene and flanking splice regions (*SI Appendix, Figs. S2–S4 and Dataset S2*). Analysis of rare variants called using a reference sample (17) and CHD samples with prior WES data (9) demonstrated very high sensitivity and precision of variant calls (*SI Appendix, Fig. S6*), at an all-in production cost of \$25.50 per sample compared to \$170 per sample for WES at the time of processing, reducing the cost of sequencing of these samples by 85%.

**Assembly of a Cohort of 11,555 CHD Probands.** MIPseq was performed on 5,929 PCGC probands (*Materials and Methods and Dataset S3*); >95% of targeted bases had  $\geq 10$  independent reads (*SI Appendix, Fig. S7*); diverse very rare variants were validated by Sanger sequencing (*SI Appendix, Table S1*). Additional WES increased totals to 3,887 CHD trios and 1,739 singletons, which combined with MIPseq probands yielded 11,555 CHD probands (*SI Appendix, Table S2 and Dataset S3*). Cardiac, structural EC, and neurodevelopmental (NDD; neurodevelopmental delay, learning disability, intellectual disability, or autism, based on parental/adult proband report of physician diagnosis at age  $\geq 1$  y) phenotypes were obtained at the time of recruitment (*SI Appendix, Table S3 and Materials and Methods*). The presence of structural EC and neurodevelopmental disorders (NDD) across different CHD subtypes departed from the null distribution (chi-square  $P < 2.2 \times 10^{-16}$ ). For example, 56% of probands with hypoplastic left heart syndrome (HLHS) had EC features, while ECs were found in only 28% of probands with left ventricular outflow tract abnormalities (LVO) (*SI Appendix, Fig. S8*).

**Pathogenic Variants in Known and New CHD Genes.** De novo and transmitted variants in trios and unphased variants in singleton probands were called and annotated for impact on encoded proteins and allele frequency (10) (*Materials and Methods and Dataset S4*). The primary analytic strategy was a meta-analysis that combined i) the significance of the burden of damaging DNMs in the 248 panel genes in probands compared to expectation; ii) the significance of the frequency of very rare damaging variants [MAF  $< 10^{-5}$  in both the BRAVO Freeze 3a WGS database (18) and Exome Sequencing Project v25 (ESP) (19) WES database] in these genes that were transmitted to probands in trios or found in singleton probands (referred to as transmitted and unphased variants, TUVs) compared to their frequency among up to 135,743 individuals in the gnomAD database (20) (*Materials and Methods*).

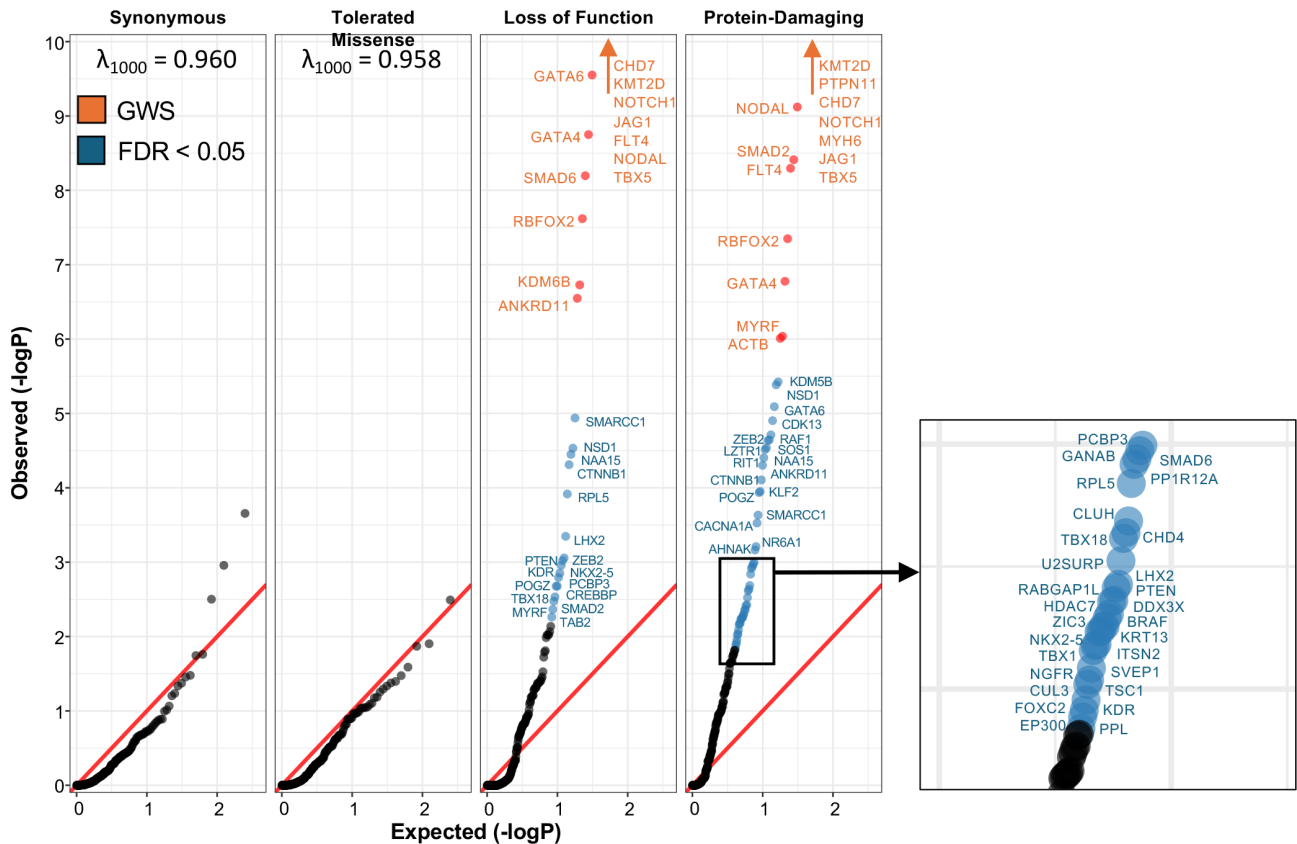
DNM's showed no enrichment of synonymous or tolerated missense variants ("T-mis"), but highly significant enrichment for both LOF (1.49-fold enriched,  $P = 4.63 \times 10^{-19}$ ) and damaging missense (D-mis) DNMs (1.35-fold enriched,  $P = 1.24 \times 10^{-13}$ ) (*SI Appendix, Table S4A*). Across all  $\sim 19,000$  genes, the combined excess LOF and D-mis DNMs implicated damaging DNMs in 9.1% of CHD probands (*SI Appendix, Table S4A*). This enrichment is dramatically increased when only the 248 panel genes are considered, with de novo LOFs enriched 14.3-fold,  $P < 10^{-117}$ , and de novo D-mis variants enriched 6.1-fold,  $P < 10^{-58}$  (*SI Appendix, Table S4B*). These damaging DNMs can explain 6.6% of all probands, comprising 72% of the DNM signal. The strong enrichment in this 248 gene set is not the result of ascertainment bias, because exclusion of the 1,213 trios used to identify this gene set, leaving only the 2,686 additionally sequenced trios, yielded highly concordant estimates of enrichment of LOF and D-mis DNMs and percentage of probands explained in this panel (LOF enrichment 14.5-fold, D-mis 5.6; 6.3% of probands explained; *SI Appendix, Table S5B*).

Next, the prevalence of very rare TUVs in each panel gene was compared between probands and gnomAD WES & WGS controls, after adjustment for per-base sequence coverage (*Materials and Methods*). The frequencies of synonymous and T-mis variants showed no significant difference between CHD probands and gnomAD controls, while LOF and D-mis variants were highly significantly enriched in CHD probands (1.48-fold enriched,  $P < 10^{-21}$  and 1.11-fold enriched,  $P < 10^{-10}$ , respectively) (*SI Appendix, Table S6A*). For comparison, we assessed LOF and D-mis variants in 28 MIPseq panel genes that were not LOF-intolerant ( $pLI < 0.9$ ), lacked protein-altering DNMs in the initial 1,213 CHD probands (10), and were not previously implicated in CHD. These 28 genes showed no enrichment in any variant functional class (*SI Appendix, Table S6B*) (*Materials and Methods*). The remaining 220 MIPseq genes showed highly significant enrichment for damaging, but not synonymous or T-mis variants (*SI Appendix, Table S6C*).

Results from the DNM and case–control tests were combined for each panel gene by meta-analysis and significance was adjusted for multiple-testing correction using the joint-local FDR (JL-FDR) method (*Materials and Methods*) (21). Quantile–quantile plots showed that synonymous and T-mis variants adhered to the expected distribution; no genes had JL-FDR  $< 0.05$ . In contrast,  $P$ -values for LOF and protein-damaging variants showed marked departure from expectation (Fig. 1).

Sixty genes had a significant excess of damaging variants. For 48 of these genes, the case–control data strengthened the significance from DNM alone, and for 14 of these genes, the case–control data were critical to the association. Thirteen of these 60 genes have not previously been implicated in CHD, three have only been previously reported in animal models or from biochemical studies in vitro, and 17 have been previously reported anecdotally in human CHD or without supporting statistical evidence (*SI Appendix, Table S7*). Significance of some genes was attributable to LOF variants (e.g., *JAG1, NODAL*), D-mis variants (e.g., *PTPN11, MYH6*), or to contributions from both (e.g., *NOTCH1, KMT2D*) (*SI Appendix, Table S8*). Variants in *GDF1* were predominantly biallelic with a founder effect (10) and were eliminated from analysis of these monoallelic variants. Data for the remaining 188 panel are shown in *Dataset S6*.

As another test of significance of these 60 genes, we performed the transmission disequilibrium test (TDT) (22) on very rare damaging variants in the 3,887 trios (*Materials and Methods*). Damaging mutations were overtransmitted in the entire 248 gene panel (*SI Appendix, Table S9A*), with nearly the entire signal



**Fig. 1.** Quantile–quantile plots of meta-analysis using DNMs and TUVs in 248 genes in 11,555 probands. For indicated variant classes, Q–Q plot is shown. Genes surpassing genome-wide significance ( $P$ -value  $< 2.6 \times 10^{-6}$ ) and those with FDR  $< 0.05$  are shown in red and blue, respectively. The y-axis has been truncated at  $-\log P = 10$ . Red arrows denote genes that have  $P < 10^{-10}$ . The genomic inflation factor,  $\lambda$ , adjusted per 1,000 samples is shown for syn and T-Mis variants. In Damaging variant results, the boxed *Inset* shows genes with observed  $-\log P$  from 1.5 to 3.0.

coming from the 60 significant genes (*SI Appendix, Table S9 B and C*). 78% of LOF variants (118/152) and 59% of D-mis variants were transmitted to probands (chi-square  $P = 9.5 \times 10^{-12}$  and  $2.6 \times 10^{-6}$ , respectively). Significance remains after excluding the two most transmitted genes, *MYH6* and *NOTCH1* (*SI Appendix, Table S9D*). Results for genes with the greatest over transmission are shown in *SI Appendix, Table S9E*. TDT results for all 60 significant genes and the 248 targeted genes are shown in *Dataset S7*.

Damaging DNMs in the 60 significant genes can account for CHD in 5.2% of all probands (*SI Appendix, Table S10A*), comprising 58% of the total signal from damaging DNMs. After removal of these 60 genes, a highly significant residual signal for damaging DNMs persists among the remaining 188 genes in the targeted set ( $P = 3.8 \times 10^{-19}$ ) accounting for another 1.4% of probands (*SI Appendix, Table S10B*). Overtransmission of damaging variants in these same genes can explain another 5.4% of probands (*SI Appendix, Table S10C*), with little evidence of significant residual signal in the remaining 188 genes (*SI Appendix, Table S10D*). By chance, 0.3% of probands are expected to have variants in two genes in the 60 gene set, and were observed in 0.4% of probands, not significantly different from expectation. In total, 10.1% of probands had one or more damaging variants in the 60 gene set.

**Association of Mutated Genes with Specific CHD Phenotypes.** The relatively large number of damaging variants in the 60 significant genes provides the opportunity to assess the variability of CHD phenotypes arising from variants in each. Adjusting for multiple comparisons (*Materials and Methods*), we found that damaging

variants in 33 genes, including *GATA6*, *FLT4*, *SMAD2*, and *KLF2* were significantly associated with a single cardiac phenotype (Table 1). Twelve other genes showed striking variability in their associated CHDs. These included *NOTCH1*, which was associated with CTD, TOF, LVO, and HLHS; *KMT2D* with CTD, LVO, and HLHS; and *PTPN11* with CTD, ASD, and AVC (Table 1). The remaining 15 genes had smaller numbers of damaging variants that were not significantly associated with any phenotypic subset.

Damaging DNMs were found most frequently in patients with ASD (7.9% of probands) and HLHS (7.3%), and least frequently in those with laterality (LAT) defects (2.1%) and TOF (3.6%). Damaging transmitted monoallelic variants were most frequent in ASD (7.9%) and TOF (7.6%) and were not significantly over-transmitted in LAT and AVC (*SI Appendix, Table S11*).

**Association of CHD Genes with EC Phenotypes.** The distribution of EC phenotypes resulting from damaging variants was also not random ( $P < 10^{-4}$  by Monte Carlo simulation). Variants in each of the 60 genes were tested for association with isolated CHD, structural EC, and NDD (*SI Appendix, Table S1B*). Some genes were associated only with one of these groups (e.g., *MYH6* with isolated CHD; *MYRF* with CHD plus EC; *GANAB* with CHD plus NDD; *RAF1* with CHD plus EC and NDD). Others were associated with multiple phenotypes (e.g., *TBX5*, *NOTCH1*, *KMT2D*), demonstrating variable expressivity. 74% (56/76) of probands with damaging DNMs in NDD-associated genes and 9% (3/35) with damaging DNMs in non-NDD-associated genes had NDD.

Overall, probands with isolated CHD were least likely to have a damaging DNM (5.0%), and those with CHD plus NDD &

**Table 1. Damaging variant meta-analysis identifies genes significantly enriched CHD probands with one or more cardiac, EC, or neurodevelopmental phenotypes**

Damaging Meta-Analysis FDR Stratified by Cardiac Lesion										Damaging Meta-Analysis FDR Stratified by NDD/EC Status					
Gene	All	CTD	TOF	LAT	LVO	HLHS	ASD	AVC	OTH	Gene	All	Isolated	EC	NDD	NDD & EC
RIT1	3.90E-04	6.40E-07	3.80E-01	7.20E-01	7.20E-01	5.10E-01	7.20E-01	7.20E-01	7.20E-01	FLT4	1.20E-07	1.90E-05	2.70E-01	3.70E-01	3.60E-01
GATA6	6.20E-05	1.20E-06	5.60E-01	3.10E-01	6.50E-01	6.50E-01	2.70E-01	6.20E-01	5.00E-01	MYH6	5.40E-14	2.40E-05	8.20E-02	7.90E-02	3.40E-01
* CACNA1A	1.40E-03	7.70E-04	4.50E-01	3.40E-01	6.40E-01	3.50E-01	5.90E-01	4.00E-01	4.00E-01	GATA4	1.60E-06	6.30E-04	2.10E-01	3.80E-01	3.00E-01
CDK13	9.40E-05	5.80E-03	2.90E-01	4.30E-01	5.60E-01	8.20E-02	4.80E-01	7.20E-01	5.60E-01	* CLUH	7.60E-03	2.30E-03	3.30E-01	3.00E-01	9.40E-02
SOS1	2.30E-04	6.70E-03	6.40E-01	6.10E-01	1.90E-01	6.10E-01	7.20E-01	1.20E-01	2.20E-01	△ KLF2	7.00E-04	5.40E-03	4.90E-01	4.90E-01	4.90E-01
DDX3X	1.80E-02	1.00E-02	7.20E-01	7.20E-01	7.20E-01	7.20E-01	7.20E-01	2.00E-01	7.20E-01	▲ TBX18	1.00E-02	1.10E-02	3.50E-01	1.50E-01	4.90E-01
<u>KDM5B</u>	2.50E-05	1.30E-02	6.50E-01	2.40E-01	1.40E-01	4.00E-01	3.60E-01	5.60E-01	2.40E-01	NODAL	1.40E-08	1.50E-02	1.60E-01	1.50E-01	1.40E-01
□ FLT4	1.20E-07	2.50E-01	7.80E-16	6.30E-01	6.00E-01	4.50E-01	6.20E-01	7.20E-01	6.30E-01	▲ MYRF	6.40E-06	1.90E-01	1.40E-03	4.90E-01	2.20E-01
▲ KDR	4.20E-02	6.00E-01	2.50E-03	6.50E-01	4.20E-01	5.80E-01	6.00E-01	2.90E-01	5.30E-01	* PPL	4.70E-02	4.00E-01	4.20E-03	4.00E-01	8.40E-02
* CLUH	7.60E-03	5.80E-01	1.40E-02	5.30E-01	5.80E-01	2.90E-01	5.60E-01	6.30E-01	4.60E-01	▲ ZIC3	2.30E-02	1.20E-01	1.70E-02	4.90E-01	4.90E-01
TBX1	3.00E-02	1.40E-01	3.70E-02	6.50E-01	5.90E-01	6.20E-01	6.00E-01	7.20E-01	7.20E-01	▲ TSC1	3.50E-02	4.10E-01	2.20E-02	3.90E-01	1.70E-01
▲ SMAD2	6.50E-08	2.00E-01	3.50E-01	3.20E-09	4.30E-01	5.60E-01	5.20E-01	4.60E-01	7.20E-01	CDK13	9.40E-05	1.10E-01	3.40E-02	4.90E-01	2.40E-01
▲ ZIC3	2.30E-02	7.20E-01	5.00E-01	6.60E-05	4.60E-01	7.20E-01	7.20E-01	7.20E-01	7.20E-01	* KRT13	2.50E-02	4.00E-01	3.90E-02	3.70E-01	4.90E-01
NAA15	3.30E-04	5.50E-01	1.70E-01	1.70E-03	5.00E-01	7.20E-01	4.10E-01	7.20E-01	7.20E-01	△ HDAC7	2.00E-02	1.80E-01	4.30E-02	4.90E-01	2.60E-01
LZTR1	2.80E-04	2.80E-01	6.30E-01	1.90E-01	2.00E-03	4.10E-01	5.80E-01	5.30E-01	5.10E-01	▲ GANAB	4.80E-03	2.90E-01	3.90E-01	2.80E-06	4.90E-01
□ SMAD6	4.00E-03	3.80E-01	1.90E-01	5.10E-01	7.70E-03	2.80E-01	3.70E-01	7.20E-01	6.40E-01	TBX1	3.00E-02	3.70E-01	4.10E-01	2.00E-02	2.40E-01
* AHNAK	2.50E-03	5.80E-02	6.10E-01	5.40E-01	2.10E-02	7.20E-01	7.20E-01	7.20E-01	3.60E-01	* AHNAK	2.50E-03	3.10E-01	3.80E-01	2.90E-02	3.00E-01
▲ SMARCC1	1.10E-03	3.20E-01	3.90E-01	1.20E-01	4.90E-02	5.10E-01	2.80E-01	7.20E-01	7.20E-01	□ SMAD6	4.00E-03	1.10E-01	1.60E-01	3.70E-02	3.30E-01
▲ MYRF	6.40E-06	5.40E-01	7.20E-01	5.30E-01	1.30E-01	7.90E-05	7.20E-01	7.20E-01	4.80E-01	▲ POGZ	8.30E-04	9.60E-02	1.10E-01	1.60E-01	4.40E-02
△ KLF2	7.00E-04	7.20E-01	7.20E-01	7.20E-01	4.20E-01	1.70E-04	3.60E-01	7.20E-01	7.20E-01	RIT1	3.90E-04	1.40E-01	2.30E-01	2.50E-01	2.80E-05
* RABGAP1L	1.60E-02	7.20E-01	7.20E-01	5.50E-01	7.20E-01	3.60E-04	5.50E-01	5.10E-01	7.20E-01	▲ PTEN	1.50E-02	3.60E-01	7.60E-02	2.80E-01	4.80E-02
* LHX2	1.30E-02	3.60E-01	7.20E-01	7.20E-01	5.70E-01	1.30E-03	7.20E-01	7.20E-01	7.20E-01	△ CTNBN1	5.60E-04	2.90E-01	5.80E-02	4.90E-01	1.20E-04
* PCBP3	3.30E-03	7.20E-01	4.90E-01	4.80E-01	7.20E-01	3.20E-03	7.20E-01	7.20E-01	7.20E-01	RAF1	1.40E-04	4.10E-01	6.30E-02	4.90E-01	1.20E-06
▲ GANAB	4.80E-03	3.90E-01	3.50E-01	6.40E-01	2.50E-01	2.60E-02	4.70E-01	3.80E-01	5.60E-01	PPP1R12A	5.60E-03	1.90E-01	8.90E-02	4.90E-01	3.50E-02
▲ PTEN	1.50E-02	5.30E-01	7.20E-01	7.20E-01	4.40E-01	7.20E-01	6.40E-04	7.20E-01	3.20E-01	▲ EP300	5.00E-02	3.90E-01	1.70E-01	4.90E-01	9.00E-04
NSD1	3.80E-05	9.90E-02	5.20E-01	3.20E-01	6.30E-01	3.50E-01	4.90E-03	6.40E-01	3.00E-01	DDX3X	1.80E-02	2.10E-01	4.90E-01	4.90E-01	1.80E-02
BRAF	2.10E-02	5.30E-01	7.20E-01	7.20E-01	5.70E-01	7.20E-01	2.90E-02	7.20E-02	4.20E-01	TBX5	3.10E-10	6.20E-06	1.20E-05	3.10E-01	4.90E-01
ANKRD11	4.60E-04	2.10E-01	5.00E-01	5.20E-01	7.50E-02	7.20E-01	7.20E-01	9.40E-04	7.20E-01	JAG1	2.80E-12	2.10E-04	1.60E-05	2.30E-01	2.90E-01
PPL	4.70E-02	2.10E-01	5.90E-01	6.10E-01	6.40E-01	6.50E-01	6.50E-01	1.10E-02	6.20E-01	GATA6	6.20E-05	8.60E-03	3.40E-04	3.40E-01	4.90E-01
△ CTNBN1	5.60E-04	6.00E-01	2.10E-01	5.80E-01	7.20E-01	5.40E-01	4.80E-01	7.20E-01	5.00E-04	<u>SMARCC1</u>	1.10E-03	3.20E-02	4.70E-03	4.90E-01	4.90E-01
PPP1R12A	5.60E-03	7.20E-01	5.10E-01	5.20E-01	7.20E-01	4.40E-01	7.20E-01	3.70E-01	1.50E-03	▲ PTPN11	9.80E-26	2.10E-07	2.60E-08	5.30E-02	4.10E-12
▲ TBX18	1.00E-02	4.50E-03	7.20E-01	3.30E-01	4.30E-01	3.10E-01	7.20E-01	7.20E-01	2.80E-03	□ NOTCH1	8.60E-17	5.10E-05	4.90E-04	2.00E-03	1.70E-03
* CUL3	4.00E-02	5.30E-01	7.20E-01	4.80E-01	4.30E-01	2.50E-01	7.20E-01	7.20E-01	3.20E-02	▲ RPL5	6.40E-03	4.90E-01	2.70E-02	3.10E-03	4.90E-01
□ CHD7	1.40E-24	2.10E-17	1.70E-06	5.20E-01	3.20E-01	1.10E-01	2.30E-01	2.20E-03	2.90E-01	□ KMT2D	1.10E-32	1.90E-01	1.20E-21	3.60E-03	1.40E-08
JAG1	2.80E-12	2.30E-02	2.40E-12	4.90E-01	4.00E-02	7.80E-02	5.70E-01	7.20E-01	4.60E-01	□ CHD7	1.40E-24	1.80E-01	6.50E-23	3.90E-01	4.60E-11
□ NOTCH1	8.60E-17	4.10E-05	4.30E-10	5.40E-01	2.80E-05	1.20E-05	4.70E-01	4.70E-01	5.80E-01	▲ SMAD2	6.50E-08	3.20E-01	6.40E-05	2.80E-01	1.20E-02
□ KMT2D	1.10E-32	1.20E-10	5.90E-01	2.40E-01	7.40E-06	9.80E-25	5.70E-01	6.30E-01	2.70E-01	NSD1	3.80E-05	4.10E-01	1.20E-03	1.70E-01	3.80E-05
□ MYH6	5.40E-14	1.60E-02	2.10E-01	4.50E-01	1.10E-03	1.20E-08	1.10E-04	1.60E-01	2.20E-01	ZEB2	1.80E-04	3.70E-01	6.20E-03	4.90E-01	9.80E-03
□ RFX2	5.00E-07	3.50E-03	4.90E-01	7.20E-01	7.20E-01	2.50E-06	3.90E-01	7.20E-01	7.20E-01	▲ ACTB	1.10E-05	1.20E-01	1.40E-02	4.90E-01	7.70E-04
TBX5	3.10E-10	1.70E-02	7.20E-01	1.90E-01	5.20E-01	5.90E-01	2.60E-07	8.30E-01	1.40E-04	SOS1	2.30E-04	3.90E-01	3.00E-02	2.40E-01	9.50E-06
□ PTPN11	9.80E-26	1.80E-16	4.40E-01	3.70E-01	3.10E-01	6.00E-01	5.40E-05	9.40E-05	4.80E-06	▲ NAA15	3.30E-04	5.50E-02	4.10E-02	4.90E-01	2.30E-02
GATA4	1.60E-06	2.50E-04	3.90E-01	6.50E-01	6.50E-02	6.50E-01	1.70E-01	1.90E-05	6.40E-01	<u>CHD4</u>	8.80E-03	4.00E-01	4.60E-02	2.30E-01	2.60E-03
ZEB2	1.80E-04	1.90E-02	1.80E-01	6.10E-01	7.20E-01	7.20E-01	5.20E-01	4.70E-01	3.50E-02	<u>KDM5B</u>	2.50E-05	3.30E-01	3.20E-01	1.00E-03	2.50E-02
NODAL	1.40E-08	3.90E-03	5.70E-01	3.40E-14	7.20E-01	7.20E-01	7.20E-01	4.20E-01	5.50E-01	ANKRD11	4.60E-04	1.00E-01	4.90E-01	7.80E-03	7.00E-03
▲ ACTB	1.10E-05	1.50E-01	7.20E-01	3.10E-01	4.60E-02	7.20E-01	1.40E-01	7.20E-01	8.90E-03						

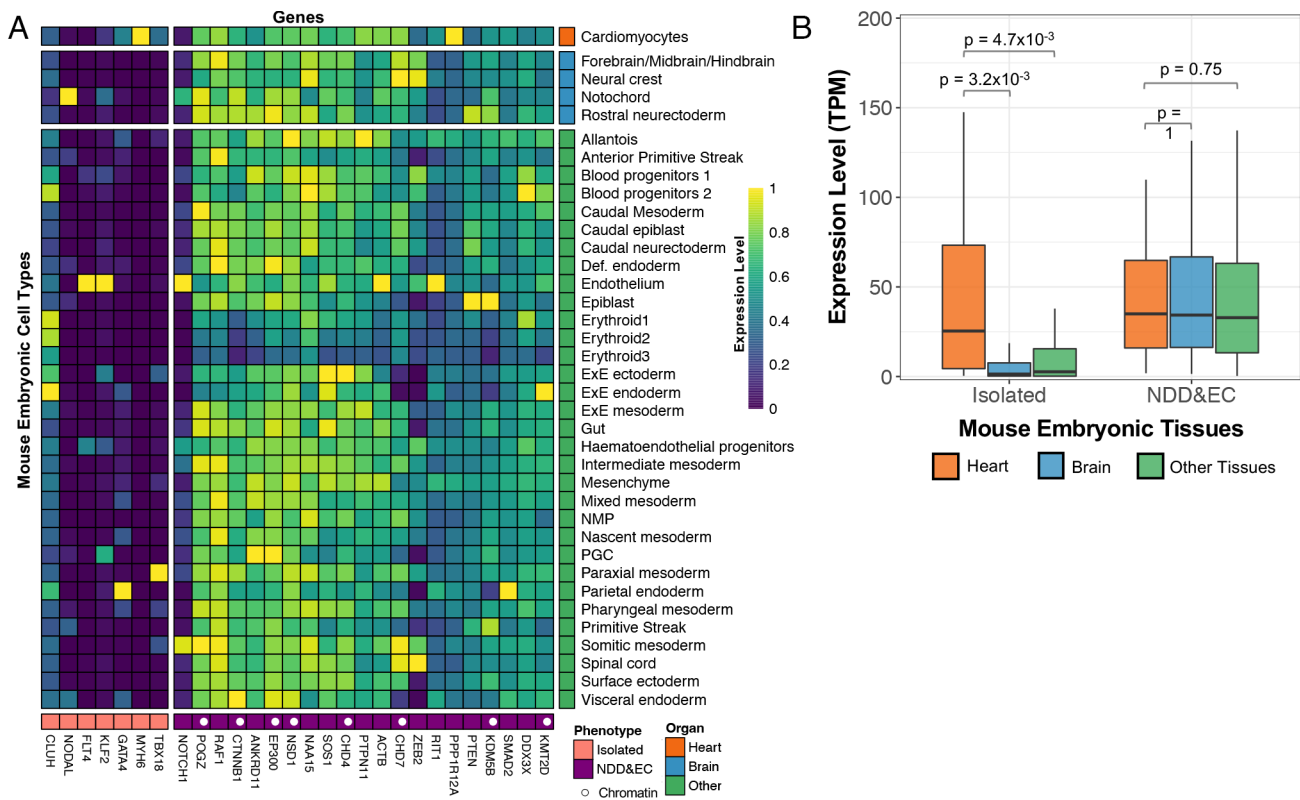
Table shows JL-FDRs for gene-based meta-analysis of damaging variants in 60 genes stratified by proband CHD phenotype (left panel) or, presence/absence of EC features and/or neurodevelopmental disease (NDD) (together: NDD & EC) in the right panel. FDR includes multiple testing correction with FDR < 0.05 highlighted in blue (*Materials and Methods*). FDR in "All" column is damaging meta-analysis of all 11,555 CHD probands (as shown in *SI Appendix, Table S7*). Chromatin gene symbols are underlined. \*, Novel CHD gene; △, Novel in humans; ▲, 1st statistical support; □, PGC previously established gene.

EC were most likely (29.2%) (*SI Appendix, Table S12*). Conversely, transmitted damaging variants were most frequent in probands with isolated CHD (6.6%), and negligible in probands with either NDD alone (0.2%) or NDD plus EC (0.002%) (*SI Appendix, Table S11*), providing evidence of impaired reproductive fitness for variants causing NDD.

**Role of Developmental Expression Patterns in Manifestation of EC Phenotypes.** We anticipated that genes associated with isolated CHD would show expression profiles restricted to cardiovascular

cell types whereas genes associated with NDD plus EC would have broader expression, including the brain. Single-cell gene expression at mouse gastrulation (6.5 to 8.5 gestational days) (23) confirmed this expectation (Fig. 2). This finding was corroborated in an expression dataset of human fetal development from 72 to 129 d postconception (24) (*SI Appendix, Fig. S9*).

**Chromatin Modifier Genes.** Damaging de novo variants were sixfold enriched ( $P < 10^{-53}$ ) among chromatin-modifying genes studied (*SI Appendix, Table S13A*). Ten showed significant mutation burden



**Fig. 2.** Single-cell expression during mouse gastrulation of genes enriched in probands with isolated CHD or NDD & EC. (A) Single-cell RNA-seq from mouse gastrulation (24) for 28 genes associated with isolated CHD probands (in pink) or NDD & EC (in purple) are shown. On the vertical axis, cell types are sorted for heart (red), brain (blue), and other tissues (green). Expression levels are scaled to the greatest TPM (Transcripts per Million Reads) for each gene across all tissues. Chromatin genes are annotated with a white dot in the purple bar. (B) Quantitation of expression of genes expressed in heart, brain, and other tissues comparing genes associated with isolated CHD and NDD & EC. Median values are shown on each boxplot. A Wilcoxon rank-sum  $P$ -value comparing expression levels in heart/brain/other tissues is shown for genes associated with isolated vs. NDD & EC probands.

(SI Appendix, Table S13B), predominantly due to de novo LOFs; nonetheless, all 13 LOFs seen in parents were transmitted to probands. Transmitted LOFs predominated for *SMARCC1* (six transmitted, zero DNMs), with six more in singletons. DNMs in these 10 genes comprise 69% of the total chromatin gene signal, 20% of the entire DNM signal and explain 1.8% of probands in trios (SI Appendix, Table S13C) and 8.3% of probands with EC and NDD (SI Appendix, Table S12). After exclusion of these 10 genes, significant signal from the remainder of this gene set remains (SI Appendix, Table S13C).

Several of these genes are associated with a single CHD subgroup (e.g., *KDM5B* with CTD, *SMARCC1*), while others (e.g., *CHD7*, *KMT2D*) are associated with multiple CHD subgroups (Table 1). All show association with EC phenotypes (Table 1) and are broadly expressed (Fig. 2).

**Transmitted Damaging Missense Variants in *MYH6* in Multiple CHD Subtypes.** *MYH6* encodes the embryonic cardiac  $\alpha$ -myosin heavy chain, and has previously been associated with a small number of patients with dilated and hypertrophic cardiomyopathy (25), ASD (25), and HLHS (26). We previously linked biallelic *MYH6* variants to multilevel left ventricular outflow obstruction (10).

*MYH6* had a twofold excess of monoallelic D-mis variants (FDR  $< 10^{-10}$  SI Appendix, Table S8A). TDT showed that 68/97 D-mis variants were transmitted from parents to probands (70%,  $P = 7.5 \times 10^{-5}$ ), while LOFs were much less frequent and not overtransmitted (SI Appendix, Table S9E). D-mis *MYH6* variants were significantly associated with and over transmitted in LVO, HLHS, and ASD (SI Appendix, Table S14 A and B). Transmitted variants contributed to 2 to 3% of probands in each of these subgroups and nearly 1% of all trios studied. The genotypic risk ratio of transmitted variants in these groups was 6.0, a large effect.

Family histories in families of probands with HLHS, LVO, or ASD and transmitted D-mis *MYH6* variants identified structural heart disease in 3 transmitting (two with ASD) and 0 nontransmitting parents, two siblings, and 7 aunts, uncles, or first cousins among 32 kindreds. 8 of 10 affected family members were on the transmitting lineage. The low frequency of clinical CHD among transmitting parents provides evidence of incomplete penetrance of these variants, though subclinical CHD cannot be excluded without echocardiography.

The distribution of transmitted, nontransmitted, and de novo very rare D-mis variants in *MYH6* is shown in SI Appendix, Fig. S10. Interestingly, two de novo D-mis variants in *MYH6* were recurrent; a both are absent in BRAVO, EVS, and gnomAD databases. Seven transmitted variants were recurrent and were transmitted to probands by 16 of 19 carrier parents (84% transmission, chi-square  $P = 2.8 \times 10^{-3}$ ) (Dataset S8).

Consistent with selective expression in the cardiomyocyte lineage (Fig. 2 and SI Appendix, Fig. S9), probands with *MYH6* D-mis variants predominantly had isolated CHD (Table 1); only 1/24 probands with transmitted variants and LVO, HLHS, or ASD, had NDD.

**LOF and Cysteine Variants in *NOTCH1*.** Notch signaling plays a broad role in developmental cell fate decisions (27). *NOTCH1* is a transmembrane cell surface receptor; its 36 extracellular EGF repeats, when bound by cell surface ligands such as JAGGED 1 on adjacent cells, trigger a signaling cascade. There were 37 LOF variants distributed throughout the *NOTCH1* coding region, highly enriched in DNM and case-control analysis (SI Appendix, Fig. S11 and Table S15A), and associated with CTD, TOF, LVO, and HLHS (Table 1 and SI Appendix,

**Table 2. Enrichment of cysteine-altering NOTCH1 D-mis mutations in CHD probands with TOF or CTD, but not other cardiac phenotypes**

Domain	TOF/CTD Probands (n = 4,240)			gnomAD Controls (n = 135,743)			FET	
	# D-mis	WT	AF	# D-mis	WT	AF	Enrich	P-Val
<i>NOTCH1</i>	68	8,030	8.40E-03	885	252,239	3.50E-03	2.4	<b>4.12E-10</b>
EGFs	58	8,040	7.20E-03	661	252,463	2.60E-03	2.7	<b>8.08E-11</b>
Non-EGFs	10	8,088	1.20E-03	224	252,900	8.80E-04	1.4	1.93E-01
EGFs ΔCys	24	8,074	3.00E-03	84	253,040	3.30E-04	8.9	<b>3.10E-14</b>
EGFs w/o ΔCys	34	8,064	4.20E-03	577	252,547	2.30E-03	1.8	<b>9.33E-04</b>
EGF #5 ΔCys	8	8,090	9.90E-04	1	253,123	4.00E-06	250.1	<b>7.44E-12</b>
EGF ΔCys not in EGF #5	16	8,082	2.00E-03	83	253,041	3.30E-04	6	<b>7.05E-08</b>

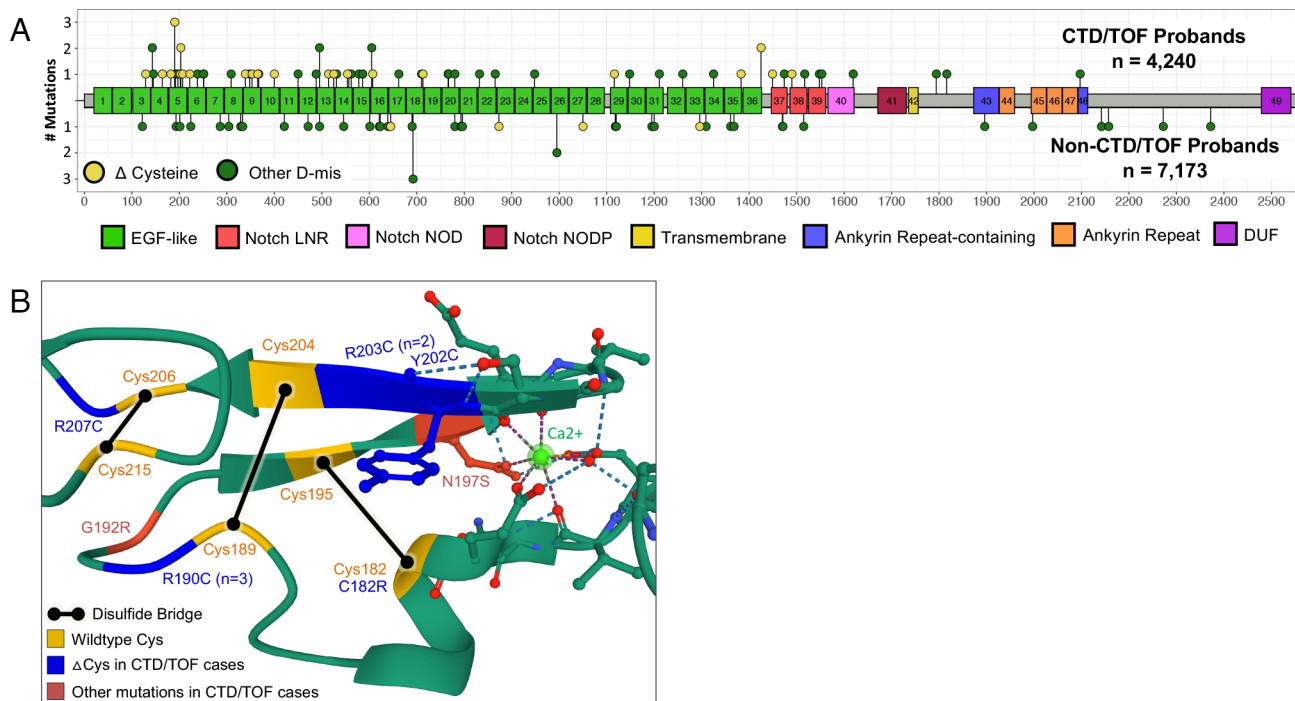
  

Domain	Non-TOF/CTD Probands (n = 7,173)			gnomAD Controls (n = 135,743)			FET	
	# D-mis	WT	AF	# D-mis	WT	AF	Enrich	P-Val
<i>NOTCH1</i>	45	13,654	3.30E-03	885	252,239	3.50E-03	0.9	6.79E-01
EGFs	36	13,663	2.60E-03	661	252,463	2.60E-03	1.0	5.09E-01
Non-EGFs	9	13,690	6.60E-04	224	252,900	8.80E-04	0.7	8.49E-01
EGFs ΔCys	4	13,695	2.90E-04	84	253,040	3.30E-04	0.9	6.67E-01
EGFs w/o ΔCys	32	13,667	2.30E-03	577	252,547	2.30E-03	1.0	4.72E-01
EGF #5 ΔCys	0	13,699	0.00E+00	1	253,123	4.00E-06	0.0	1.00E+00
EGFs ΔCys not in EGF #5	4	13,695	2.90E-04	83	253,041	3.30E-04	0.9	6.59E-01

Case-control results are shown. Probands with CTD or TOF in the upper panel and probands with all other characterized cardiac phenotypes in the lower panel. ΔCys denotes cysteine-altering mutations. Domains from the PROSITE and PFAM databases. P-values significant in NOTCH1 subsets after Bonferroni multiple testing correction ( $P \leq 0.05/6 = 8.3 \times 10^{-3}$ ) are shown in red.

Table S15A). *NOTCH1* LOFs were associated with isolated CHD, EC, NDD, and NDD + EC phenotypes (Table 1 and SI Appendix, Table S15A).

Eleven parents had LOF *NOTCH1* variants; all were transmitted to CHD probands. Three parents, all mutation carriers, had CHDs including aortic coarctation, bicuspid aortic valve (BAV),



**Fig. 3.** Frequent cysteine-altering mutations in *NOTCH1* EGF domain #5 in TOF/CTD probands. (A) NOTCH1 protein domain structure derived from Prosite and PFAM databases is shown, with EGF-like domains in green; amino acid positions in NOTCH1 are numbered on the x-axis. Lollipops show location of very rare D-mis variants in CHD probands; those in CTD/TOF probands are shown above the protein diagram and non-CTD-TOF probands are shown below. Variants that introduce or remove a cysteine are shown in yellow; all others are in green. Height of lollipops corresponds to number of independent occurrences of specific variants. "DUF" denotes "Domain of unknown function DUF3454." (B) Crystal structure of *NOTCH1* EGF domain. Amino acids altered to or from cysteine in CTD/TOF probands (blue text), and other mutations in EGF #5 in CTD/TOF probands are in red; disulfide bridges in canonical protein are shown as black lines. Structure from ref. 30.

**Table 3. Phenotypes of probands with characteristic mutations causing syndromic CHD**

Gene (syndrome)	Phenotypic feature	Phenotype % in reported patients	Phenotype freq in Mut+ probands	Gene (syndrome)	Phenotypic feature	CHD patients with given phenotype		
						Syndromic Dx Yes	Syndromic Dx No	P-Val
CHD7 (CHARGE syndrome)	Clinical diagnosis	100%	23/32 (71.9%)	CHD7 (CHARGE syndrome)	CHD	23/23 (100.0%)	9/9 (100.0%)	1.00
	CHD	75 to 92%	32/32 (100.0%)		NDD	20/21 (95.2%)	6/8 (75.0%)	0.18
	NDD	76 to 100%	26/29 (90.0%)		TEF	4/23 (17.4%)	2/9 (22.2%)	1.00
	TEF	18 to 29%	6/32 (18.8%)		Coloboma	16/22 (71.4%)	1/8 (12.5%)	<b>9.36E-03</b>
	Coloboma	75 to 89%	17/30 (56.7%)		Choanal atresia	6/22 (27.3%)	2/9 (22.2%)	1.00
	Choanal atresia	44 to 60%	8/31 (25.8%)		Hearing loss	21/22 (95.5%)	4/9 (44.4%)	<b>4.60E-03</b>
	Hearing loss	89 to 94%	25/31 (80.6%)		Feeding difficulty	22/23 (95.6%)	5/9 (55.5%)	<b>1.50E-02</b>
	Feeding difficulty	80 to 82%	27/32 (84.4%)		G-tube	18/22 (81.8%)	4/9 (44.4%)	0.08
	G-tube	N.R.	22/31 (71.0%)		Tracheostomy	3/22 (13.6%)	2/9 (22.2%)	0.61
	Tracheostomy	N.R.	5/31 (16.1%)		Deceased	1/23 (4.5%)	2/9 (22.2%)	0.18
KMT2D/KDM6A (Kabuki Syndrome)	Clinical diagnosis	100%	18/25 (72.0%)	KMT2D/KDM6A (Kabuki syndrome)	̄ Age ± SD [Range]	10.1 ± 6.7 [2-27]	12.1 ± 16.5 [0.1-55]	0.84
	CHD	37 to 70%	25/25 (100.0%)		CHD	18/18 (100.0%)	7/7 (100.0%)	1.00
	NDD	87 to 89%	21/25 (84.0%)		NDD	17/18 (94.4%)	4/7 (57.1%)	5.3E-02
	Dysmorphic facies	80%	18/21 (85.7%)		Dysmorphic facies	17/17 (100.0%)	1/4 (25.0%)	<b>3.01E-03</b>
	Hypotonia	30%	13/25 (52.0%)		Hypotonia	11/18 (61.1%)	2/7 (28.5%)	0.20
	Feeding difficulty	70%	17/25 (68.0%)		Feeding difficulty	15/18 (83.3%)	2/7 (28.5%)	<b>1.69E-02</b>
	G-Tube	N.R.	12/25 (48.0%)		G-Tube	10/18 (55.5%)	2/7 (28.5%)	0.38
	Ab. Renal Anatomy	25%	8/23 (34.7%)		Ab. Renal Anatomy	7/18 (38.8%)	1/5 (20.0%)	0.14
	Clinical diagnosis	100%	47/76 (61.8%)		Deceased	0/18 (0.0%)	0/7 (0.0%)	1.00
	CHD	66%	78/78 (100.0%)		̄ Age ± SD [Range]	9.1 ± 3.7 [3-16]	12.2 ± 9.2 [5-30]	0.93
RASopathy-associated (Noonan/RASopathy)	NDD	35 to 44%	35/71 (49.2%)	RASopathy-associated (Noonan/RASopathy)	CHD	47/47(100.0%)	29/29 (100.0%)	1.00
	Dysmorphic Facies	Most	30/47 (63.8%)		NDD	28/44 (63.6%)	7/25 (28.0%)	<b>5.24E-03</b>
	Lymphatic Ab.	<20%	5/75 (6.6%)		Dysmorphic facies	27/31 (87.0%)	3/14 (21.4%)	<b>4.50E-05</b>
	Feeding Difficulty	25%	20/73 (27.3%)		Lymphatic Ab.	4/45 (8.8%)	1/28 (3.5%)	0.64
	G-Tube	N.R.	10/74 (13.5%)		Feeding difficulty	17/44 (38.6%)	3/27 (11.1%)	<b>1.48E-02</b>
	HCM/Sub-AS	20%	12/78 (15.3%)		G-Tube	8/45 (17.7%)	2/27 (7.4%)	0.30
	Arrhythmia	<1%	16/71 (22.5%)		HCM/Sub-AS	10/47 (21.2%)	2/29 (6.8%)	0.12
	Clinical diagnosis	100%	11/18 (61.1%)		Arrhythmia	13/42 (31.0%)	2/28 (7.1%)	<b>3.30E-02</b>
	CHD	66%	18/18 (100.0%)		Deceased	0/47 (0.0%)	2/29 (6.8%)	0.15
	NDD	35 to 44%	6/16 (37.5%)		̄ Age ± SD [Range]	15.1 ± 9.8 [0.4-38]	14.9 ± 12.0 [4-64]	0.64
PTPN11 N308 (Noonan/RASopathy)	Dysmorphic facies	Most	7/10 (70.0%)	PTPN11 N308 (Noonan/RASopathy)	CHD	11/11 (100.0%)	7/7 (100.0%)	1.00
	Lymphatic Ab.	<20%	2/17 (11.7%)		NDD	5/11 (45.4%)	1/5 (20.0%)	0.59
	Feeding difficulty	25%	2/16 (12.5%)		Dysmorphic facies	6/6 (100.0%)	1/4 (25.0%)	<b>3.33E-02</b>
	G-Tube	N.R.	2/16 (12.5%)		Lymphatic Ab.	2/10 (20.0%)	0/7 (0.0%)	0.49
	HCM/Sub-AS	20%	3/18 (16.6%)		Feeding difficulty	2/11 (18.1%)	0/5 (0.0%)	1.00
	Arrhythmia	<1%	2/15 (13.3%)		G-Tube	2/11 (18.1%)	0/5 (0.0%)	1.00
	Clinical diagnosis	100%	11/18 (61.1%)		HCM/Sub-AS	2/11 (18.1%)	1/7 (14.2%)	1.00
	CHD	66%	18/18 (100.0%)		Arrhythmia	2/8 (25.0%)	0/7 (0.0%)	0.47
	NDD	35 to 44%	6/16 (37.5%)		Deceased	0/11 (0.0%)	0/7 (0.0%)	1.00
	̄ Age ± SD [Range]				̄ Age ± SD [Range]	19.4 ± 10.4 [0.4-38]	13.8 ± 6.2 [8-25]	0.16

The left panel compares reported and observed frequencies of clinical phenotypes in probands with LOF mutations in *CHD7* and *KMT2D/KDM6A* and ClinVar+ mutations in *PTPN11*, *BRAF*, *RAF1*, *RIT1*, *SHOC2*, *SOS1* and *LZTR1* genes together and N308 mutations in *PTPN11*. Citations for reported phenotype frequencies for CHARGE syndrome are (31–33); for, Kabuki syndrome (34–37); and Noonan/Rasopathy (10, 38, 39). HCM, hypertrophic cardiomyopathy. The right panel compares clinical phenotypes in CHD probands with pathogenic mutations that were or were not clinically diagnosed. *P*-values from two-tailed FET for binary traits, and Wilcoxon rank-sum test for continuous variables.

and pulmonary atresia/ventricular septal defect. Other relatives on the transmitting lineages had TOF, HLHS, and VSD, consistent with the spectrum of phenotypes found in probands. The absence of clinical CHD among the other 8 transmitting parents implies incomplete penetrance or subclinical CHD; echocardiograms have not been performed.

D-mis variants in *NOTCH1* were selectively enriched for two cardiac phenotypes, TOF (threefold enrichment;  $P = 2.3 \times 10^{-8}$ ) and CTD (twofold enrichment;  $P = 4.3 \times 10^{-4}$ ) (*SI Appendix, Table S15B*), with enrichment limited to *NOTCH1*'s 36 EGF domains ( $P < 10^{-10}$ ). Each EGF domain has six cysteines which form three intradomain disulfide bonds that are critical domain structure (28). D-mis variants introducing or removing one cysteine, producing an odd number of cysteines, were enriched 8.9-fold ( $P < 10^{-13}$ ) and accounted for 41% (24/58) of the D-mis mutations in EGF domains (Table 2). These variants all leave an unpaired cysteine available for electrophilic attack. Cysteine-altering variants were highly clustered, with eight occurring

in EGF domain #5 including two recurrent variants (250-fold enrichment,  $P < 10^{-11}$ ; Table 2 and Fig. 3 *A* and *B*), and 20 (83%) among the first 18 EGF domains. In contrast, there is only one cysteine mutation in the first 18 EGF domains in non-CTD/TOF probands. All these cysteine-altering variants had allele frequencies of zero in the BRAVO, ESP, and gnomAD databases (*SI Appendix, Table S16*). An additional missense variant altered asparagine 197 in EGF #5; this residue helps coordinate  $Ca^{2+}$  in this EGF domain;  $Ca^{2+}$  binding is known to be critical to *NOTCH1* signaling (29).

These clustered cysteine mutations are highly reminiscent of cysteine-altering variants in *NOTCH3* that cause cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). Notably, in trios, all 6 parental cysteine mutations in EGFs 1-18 were transmitted to probands with TOF or CTD. Among four kindreds with family history data, one transmitting parent had TOF while others had no CHD history.

**Clinical Characteristics of CHD Probands with Pathogenic Variants in Syndromic Genes.** Because clinically significant CHD was the sole requirement for study enrollment, and clinical phenotypes were recorded independently of mutation status, this cohort allowed unbiased estimates of the association of genotypes and specific phenotypes. We performed follow-up chart review of clinical features of probands with LOF variants in the most frequently mutated syndromic genes in our cohort—*CHD7* and *KMT2D*, and known GOF mutations in *BRAF*, *PTPN11*, *RAF1*, *RIT1*, *SHOC2*, and *SOS1* (and LOFs in *LZTR1*) causing Noonan syndrome and related RASopathies. Follow-up occurred 0 to 10 y after initial recruitment. Overall, among 133 probands with disease-causing variants in these genes, only 88 (66%) were clinically diagnosed with their respective syndromes, indicating incomplete sensitivity for the clinical diagnosis.

***CHD7* and CHARGE Syndrome.** LOF variants in *CHD7* cause CHARGE syndrome, featuring coloboma, atresia choanae, slow growth and development, genital malformations, ear malformations, and congenital heart disease (31–33). Thirty-two probands with *CHD7* LOFs had available medical records (Table 3). Twenty-three of these probands (72%) were clinically diagnosed with CHARGE.

The frequency of characteristic CHARGE phenotypes in these probands and those reported in the literature showed that frequencies of coloboma (57% vs. 75 to 89% reported), choanal atresia (26% vs. 44 to 60% reported), and hearing loss (81% vs. 89 to 94% reported) were lower than previously estimated (Table 3) (31). Other frequencies were like those previously reported.

Overestimates of the frequency of syndromic features likely contributed to underdiagnosis of CHARGE. Thus, while 16 of 22 (71%) probands with coloboma were diagnosed with CHARGE, only 1 of 8 (13%) without coloboma was diagnosed. Similarly, 21 of 22 (96%) with hearing loss but only 4 of 9 (46%) without were diagnosed. The age at evaluation of those who were and were not diagnosed with CHARGE were similar (means of 10.1 and 12.1 y, respectively;  $P = 0.84$ ), as was the distribution of LOF variants (*SI Appendix, Fig. S12A*).

***KMT2D* and Kabuki Syndrome.** Kabuki syndrome features infantile hypotonia, NDD, palpebral fissures, and characteristic facial dysmorphism (34). Sixty to seventy-five percent of Kabuki probands have LOF DNMs in *KMT2D*, and 3 to 8% are due to LOF DNMs in *KDM6A* (40). Our cohort included 21 with LOF DNMs and 12 phase-unknown LOFs in *KMT2D* (*SI Appendix, Table S19*) and one LOF in *KDM6A*. CHD phenotypes were HLHS, CTD, and LVO, and all but two probands had EC, NDD, or both (*SI Appendix, Table S19*). Clinical details were available for 25 probands (*SI Appendix, Table S20*); 18 (72%) had a clinical diagnosis of Kabuki syndrome. The frequency of characteristic Kabuki phenotypes in our cohort was very similar to reported values, except increased hypotonia (52% vs. 30% reported) and abnormal renal anatomy (35% vs. 25% reported) (Table 3) (34–36). Only one of seven probands not diagnosed (Table 3), had dysmorphic facies reported, vs. 100% of those diagnosed with Kabuki ( $P = 3.0 \times 10^{-3}$ ). Similarly, feeding difficulty was reported less frequently in those without a Kabuki diagnosis (29% vs. 83%;  $P = 1.7 \times 10^{-2}$ ). There was no significant difference in age between diagnosed and undiagnosed probands (9.1 and 12.2 y, respectively;  $P = 0.93$ ) and no difference in the distribution of LOF variants between these groups (*SI Appendix, Fig. S12B*).

There were also six D-mis DNMs in *KMT2D*, all clustered in its C-terminus, which includes domains found in lysine methyltransferases (*SI Appendix, Fig. S13 and Table S21*) (41); these DNMs were absent in BRAVO, ESP, and gnomAD databases.

Two DNMs were identical R5351Q mutations. Four probands had HLHS (*SI Appendix, Fig. S13*). Of five probands with clinical data, only one was clinically diagnosed. One proband had NDD and thrombocytopenia, two were deceased, and three had feeding difficulty. One additional proband had a transmitted C-terminal variant, p.R5179H, identical to a DNM in a proband; this proband had HLHS and EC, consistent with a pathogenic variant. C-terminal missense variants in *KMT2D* have been reported in CHD and Kabuki syndrome, with variable expressivity of cardiac and EC features (37, 42, 43).

**RAS/MAPK Pathway and Noonan Syndrome.** Noonan syndrome and related RASopathies share distinctive facial features, NDD, short stature, lymphatic malformations, hematological abnormalities, and CHD (including congenital hypertrophic cardiomyopathy) (38, 39). They are caused by variants in the RAS/MAPK pathway, including GOF missense variants in activators of the pathway (*PTPN11*, *SOS1*, *RAF1*, *BRAF*, *SHOC2*, *RIT1*) and recessive LOF variants or recurrent dominant-negative missense variants in the 4th Kelch domain of *LZTR1*, an inhibitor of RAS signaling (44).

Ninety variants in these genes were identified as pathogenic in ClinVar (ClinVar+, *SI Appendix, Table S22*). Twenty-two were de novo, four were transmitted, and 64 were unphased (*SI Appendix, Table S23*). Seventy-eight probands had available medical records. Noonan syndrome/RASopathy-related phenotypes occurred near their reported frequencies (38, 45) (Table 3). Overall, 38% were not clinically diagnosed with Noonan/RASopathy (Table 3) (38). Only 21% of undiagnosed probands had dysmorphic facies reported, vs. 87% in those diagnosed ( $P = 4.5 \times 10^{-5}$ ) and NDD was also less frequent in undiagnosed (28%) vs. diagnosed probands (64%) ( $P = 5.2 \times 10^{-3}$ ) (Table 3). Notably, both deceased probands were from the undiagnosed group. Age at evaluation was similar (mean of 15.1 y in diagnosed vs. 14.9 y in undiagnosed probands;  $P = 0.64$ ).

In contrast to the expected RASopathy-associated cardiac phenotypes (75 to 90% accounted for by PS, HCM, and ASD), probands had more ASD, less LVH, and more polyvalvular disease, BAV/aortic coarctation, and AVC (*SI Appendix, Table S24A*). Unexpectedly, arrhythmia requiring medical management or pacemaker/ICD was reported in 16 (23%) ClinVar+ probands (*SI Appendix, Table S24B*). Patients with arrhythmia were diagnosed with RASopathy more frequently than those without (87% vs. 52%,  $P = 1.9 \times 10^{-2}$ ), were older (22.7 vs. 13.5 y,  $P = 0.046$ ) and were more likely to have LVH (31% vs. 13%,  $P = 0.13$ ).

A clinical diagnosis was made more frequently in probands with variants in *RAF1*, *RIT1*, and *SOS1* (collectively 15/18; 83% diagnosed) vs. those with variants in *LZTR1* (0/5 diagnosed) or *PTPN11* (29/49; 59% diagnosed) (*SI Appendix, Table S25*).

Variants in *PTPN11* included many at known mutational hotspots (*SI Appendix, Fig. S12C*). Among 18 probands with an N308 variant (46), 11 were diagnosed with Noonan/RASopathy and seven were not. Dysmorphic facies were present in all diagnosed probands but only 25% of undiagnosed ( $P = 0.03$ ) (Table 3) and ages were similar (mean 19.4 and 13.8 y, respectively;  $P = 0.16$ ).

The absence of a clinical diagnosis of these syndromes was correlated with the absence of characteristic clinical features that were variable among probands harboring disease-causing variants.

## Discussion

This study represents, to our knowledge, the largest investigation to date of the impact of rare variants on structural CHD. Use of MIPseq and WES has increased the number of probands with

genomic data fourfold compared with our previous study (10) and enabled substantial progress in gene discovery. The robust MIPseq gene panel captured over 1.4 Mb of genomic DNA sequence from 248 genes with high completeness and precision. MIPseq can provide a low-cost platform for discovery or clinical genetic diagnosis for diseases featuring high locus heterogeneity, though continued reductions in sequencing costs will narrow the potential advantages of this method compared with whole exome or whole genome sequencing.

The results increase the number of genes statistically implicated in monoallelic CHD from this cohort to 60. Of the 13 genes not previously implicated in CHD, 10 had 2 or more damaging DNMs in the training set and 2 were nominated based on suggestion from prior biology. Of 3 genes not previously implicated in human CHD, 1 had >2 DNMs in the training set and 2 were high pLI chromatin genes. Of note, 9 genes in the gene panel that are not currently significant have 2 damaging DNMs, including chromatin genes *KDM5A*, *KMT2A*, *KMT2C*, *NSD2*, *SETD5*, and *CREBBP*, suggesting some will likely become significant with further study.

Damaging DNMs in these 60 genes are found in 5.1% of probands, accounting for 57% of the DNM signal (*SI Appendix, Table S10A*). TDT implicated transmitted variants in 5.4% of probands, a marked increase from previous estimates (*SI Appendix, Table S10C*). Collectively, 10.1% of probands had at least one disease-related variant.

Monte Carlo simulation provided a maximum likelihood estimate that 297 genes (95% CI: 169.4 to 424.6) contribute to CHD by DNM demonstrating that much remains to be discovered. From the 248 gene panel, the current results predict that 111 will ultimately prove to be significant (*Materials and Methods*).

The DNM frequency is not distributed normally across these CHD genes, even after adjusting for gene mutability. Damaging DNMs in just four genes (*KMT2D*, *CHD7*, *PTPN11*, and *NOTCH1*) account for a third of these DNMs (69/208) (vs. expected 8%; the Shapiro–Wilk test rejected a normal distribution,  $P < 10^{-11}$ ). The high frequency of GOF DNMs in *PTPN11* and some other RASopathy genes can be explained by positive selection of these variants in the paternal germ cell lineage (47, 48); however, germline selection has not been reported for *KMT2D*, *CHD7*, and *NOTCH1* (49). Another explanation for the high prevalence of DNM in these genes could be much greater survival to birth among embryos with mutations in these genes. This could explain the very long tail on the curve relating sample size to the discovery of new disease genes.

Some of the newly implicated genes link to known biology. *CUL3* (pLI = 1) is a ubiquitin ligase component that partners with many proteins that target specific proteins for ubiquitylation and degradation (50). One of these is *LZTR1*, a known RASopathy gene that targets proteins in the RAS pathway for degradation. Similarly, *LHX2* (pLI = 0.99), encodes a transcriptional activator; mouse knockouts produce CHD (51). Both probands with damaging DNMs had aortic and mitral atresia. *AHNAK* encodes a heart scaffolding protein that regulates L-type calcium channels (52). *SVEP1* encodes an extracellular adhesion protein that, in zebrafish, was responsive to blood flow and interacted with the VEGF pathway (53). *NR6A1*, encoding a nuclear receptor, is implicated in vascular smooth muscle cell migration (54) and was upregulated in mice with CHD born of diabetic mothers (55). *CLUH* (pLI = 0.99), encodes an mRNA-binding protein involved in intracellular distribution and biogenesis of mitochondria, linked to cardiac hypertrophy (56). Further work will be required to establish the role of these and other genes newly implicated in human CHD.

Consistent with their shared role in cardiac development, the 60 core CHD genes are significantly more connected than expected by STRING analysis ( $P < 1.0 \times 10^{-16}$ ), and many of the genes newly implicated in human CHD are connected within the network (*SI Appendix, Fig. S14*) (57).

Transmitted D-mis mutations in *MYH6* proved to play a surprisingly large role, contributing to 2 to 3% of ASD, HLHS, and LVO, and nearly 1% of the entire cohort studied. The estimated effect sizes of these variants from TDT were large. Nonetheless, only 3 of 32 transmitting parents and two siblings of probands had a clinical diagnosis of CHD.

Similarly, all 11 parental LOFs in *NOTCH1* were transmitted to probands, as were all six parental cysteine-altering mutations among the first 18 EGF domains of *NOTCH1*, but only 4/15 transmitting parents from these trios with available medical histories had clinical CHD.

The significant variability in CHD phenotypes resulting from LOF mutations in 12 of these genes is striking. Ten had pLI > 0.9; nine of these showed variable EC and NDD phenotypes, consistent with dosage sensitivity for multiple developmental processes. Explanations for the variable expressivity could be explained by common variant modifiers, environmental influences, stochastic events, or subclinical CHD, as has been described for LOF variants in *NOTCH1* (58). Genome-wide association studies in this cohort would be useful in assessing whether variable expressivity can be explained by common modifier alleles.

The results increased the number of significantly mutated chromatin genes to 10. These are predominantly de novo LOFs, with few transmitted variants. These probands are enriched for EC phenotypes, reflecting the role of chromatin genes in many cell types. Maximum likelihood estimates that 28 chromatin genes contribute to CHD, with some identified in other cohorts (11).

Cysteine-altering variants clustered in *NOTCH1* n-terminal EGF domains, were highly enriched in probands with CTD and TOF. These variants are analogous to mutations in *NOTCH3* that cause CADASIL (59). Mutant *NOTCH3* extracellular domains accumulate in the basement membrane of small arteries, arterioles, and capillaries of the brain and skin, consistent with protein misfolding and/or electrophile attack of the unpaired cysteines (60).

Pairing genetic data with systematically collected cardiac and EC phenotypes enabled estimates of the frequency of phenotypes associated with each gene. For example, NDD was uncommon in probands with damaging variants in *MYH6* variants but extremely common in those with LOF variants in *CHD7* and *KMT2D*. The ability to assess future risk of NDD at birth provides the potential for early intervention.

Limitations to the study include the challenge of accurately assessing from whether rare missense variants are deleterious or benign. Also, assessment of NDD phenotypes by questionnaire in very young probands may underestimate the ultimate frequency of NDD. Nonetheless, genes that were/were not statistically associated with NDD or EC showed significant differences in EC gene brain expression, a result not expected by chance (Fig. 2B and *SI Appendix, Fig. S9B*). Also, for *CHD7*, *KMT2D*, and Noonan/RASopathy genes (Table 3), systematic chart review up to 10 y after initial ascertainment gave similar estimates of NDD frequency as obtained by questionnaire at ascertainment. The age distributions of probands at enrollment are shown in *SI Appendix, Fig. S15*. The percentage of probands with NDD was consistently 15 to 20% of probands through age 25. The percentage subsequently declined with increasing age, potentially due to selection for survival and/or ascertainment bias.

With variable expressivity, ascertainment will influence the frequency of specific phenotypes observed. For example, a prior study

ascertaining for congenital hydrocephalus identified 5 probands with damaging DNMs or transmitted LOFs in *SMARCC1* (61). In contrast, the PCGC cohort had 10 CHD probands with *SMARCC1* LOFs, none with NDD. While a few genes (*CHD7*, *KMT2D*, *PTPN11*, *NOTCH1*) have enough probands with pathogenic mutations to have reasonable estimates of the frequency of NDD, there are insufficient data to exclude NDD or accurately estimate its frequency for many genes. Much larger cohorts will be needed to both extend CHD gene discovery and define clinical variability in phenotypes arising from mutation of each CHD gene.

Collectively, rare monoallelic variants account for at least 40% of CHD: 25% from large CNVs and aneuploidy (7), 9% via damaging DNMs, at least 5% from transmitted variants as reported herein, and 2% from recessive inheritance (62). Studies have also implicated rare noncoding variants (63, 64) and common variants with small effects (65, 66), but these have been limited by small sample size and inability to link specific variants to CHD. Nongenetic factors including pregestational maternal diabetes, hypertension, alcohol use, and others account for another 10 to 14% of CHD (6, 67); the mechanisms of these effects are unknown. Given the large role of haploinsufficiency for genes with high pLI, stochastic effects that reduce expression of these genes at critical times in particular cell types could explain phenocopies in the absence of mutation.

The results herein demonstrate the significantly greater precision of molecular diagnosis of CHD to establish an early diagnosis for the syndromic genes frequently mutated in CHD. These patients are at variable risk for NDD, hearing loss, arrhythmias, and other disorders that can impact clinical management and outcome. The clinical implications for prevention and mitigation provided by early molecular diagnosis are increasingly apparent and support the routine use of genomic analysis in evaluation of CHD.

## Materials and Methods

**Ascertainment and Clinical Characteristics of Study Population.** 11,555 probands and 7,774 parents, comprising 3,887 trios and 7,668 singleton probands recruited to the Congenital Heart Disease Network Study of the Pediatric Cardiac Genomics Consortium were investigated (CHD GENES: ClinicalTrials.gov identifier NCT01196182) (68). All participants or their parents provided informed consent using protocols that were reviewed and approved by institutional review boards of Boston's Children's Hospital, Brigham and Women's Hospital, Children's Hospital of Los Angeles, Children's Hospital of Philadelphia, Columbia University Medical Center, Great Ormond Street Hospital, Icahn School of Medicine at Mount Sinai, Rochester School of Medicine and Dentistry, Steven and Alexandra Cohen Children's Medical Center of New York, University of California-San Francisco School of Medicine, University of Utah School of Medicine, Stanford University School of Medicine and Yale School of Medicine. All data are stripped of identifiers and labeled with a study number. Classification of CHD and EC features were as previously described (68) (*SI Appendix, Table S3*). Further evaluation of probands found to have mutations associated with CHARGE syndrome, Kabuki syndrome, or Noonan syndrome/RASopathies was done through focused chart review.

**Molecular Inversion Probe Design, Sequencing, and Processing.** A MIPseq panel of 248 genes were selected using data from our earlier study of 1,213 CHD trios (9) (*SI Appendix*). Probes were designed using the MIPGen algorithm applying default parameters (69) with modifications, including pooling of sets of probes of similar G-C content, iterative rounds of development to replace ineffective probes, and balancing of pool inputs into sequencing reaction to produce similar levels of sequence coverage. Gap filling reactions spanned ~250 bp of genomic sequence and adjacent probes overlapped by at least 30 bp. Design included a random 6 bp unique molecular identifier barcode added to the extension arm of each probe, permitting recognition and exclusion of PCR duplicates. Probe targets covered all coding regions plus at least 12 intronic bases flanking each exon (70). Probes were synthesized by Integrated DNA Technologies at a

scale sufficient to genotype 50,000 samples. The probe design file in [Dataset S2](#). Library preparation and sequencing of MIPseq samples was modified from previously published methods (71) and fully described in *SI Appendix*. MIPseq and WES was performed on the Illumina Platform at the Yale Center for Genome Analysis (YCGA). The total production cost of research MIPseq for each sample was ~\$25.50 per sample.

**Variant Calling.** Variants were called using GATK v3.7 (72) using default parameters, disabling of variant quality score recalibration due to lack of training SNPs in the MIPseq panel; variants were also called with FreeBayes v1.3.2 (<https://github.com/ekg/freebayes>) using default parameters. The union of these variant calls was annotated using ANNOVAR (73), multiallelic sites were split with BCFtools (74), and insertion-deletion variants were left-aligned with BCFtools (*SI Appendix*). Sensitivity and specificity for calling rare variants from the MIPseq pipeline was established by comparing MIPseq calls from Genome in a Bottle sample NA12878 compared to gold-standard variants for this sample (*SI Appendix, Fig. S6A*) (17). We also compared MIPseq and WES of 170 PCGC CHD samples (*SI Appendix, Fig. S6B*). Variants from WES of PCGC samples were called as described previously (10).

To validate candidate variants called by the MIPseq pipeline in new CHD probands, we performed PCR amplification using custom primers followed by Sanger sequencing of 59 very rare variants comprising a wide range of mutation types (*SI Appendix, Table S1*).

**Cohort Sequencing.** All MIPseq and WES was performed on the Illumina Platform at YCGA. The study cohort contains 11,555 CHD probands, 5,929 with MIPseq, and 5,626 with WES, including 1,739 singletons and 3,887 probands with both parents (7774).

The control cohort comprises 133,743 samples with WES or WGS from the gnomAD database after exclusion of samples that are also present in the TopMed database, which is included in the BRAVO variant database (20). Parental and sample-level data for gnomAD samples were not available.

**Annotation of Variants.** Variants were annotated for the most severe functional consequence among all potential isoforms using ANNOVAR and Meta-SVM from gnomAD v2.1.1 (20) in both cases and controls as previously described (10).

**Analysis of DNMs in Trios.** DNMs in trios were identified using TrioDeNovo (75) and filtered as previously described (10). All candidate variants were visually inspected in silico in the proband and both parents.

Enrichment of DNMs in probands was calculated from the observed frequency compared to the mutability-based expectation as previously described (10). The genomic inflation per thousand samples ( $\lambda_{1000}$ ) was calculated (76). To determine the enrichment of DNMs in sets of multiple genes, the observed and expected counts were summed across all genes included in a given set. We estimated the number of genes expected to have >1 DNM in a cohort of 3,887 trios as described previously (10).

**Identification and Analysis of Very Rare TUVs in Probands and Controls.** We filtered for variants with a MAF  $\leq 10^{-5}$  in both BRAVO (18) and EVS (77) databases. We also removed variants that had a within-cohort MAF  $\leq 1.3 \times 10^{-4}$  independently for CHD cases and gnomAD controls. BRAVO variants were lifted over from hg38 to hg19 using LiftOver (78). In gnomAD controls, the 100 most frequent very rare variants were visualized in silico using at least three representative variants from the online gnomAD browser (20); those found to be false positives were removed. We further harmonized cases and controls by removing exons or flanking splice regions that were not targeted for capture by all platforms.

To calculate enrichment of TUVs in cases, we compared the frequency of alleles of each class (i.e., synonymous, T-mis, D-mis, LOF) in cases and controls and calculated *P*-values using a one-tailed Fisher's exact test (FET).

**Meta-Analysis of DNMs and Very Rare TUVs.** Utilizing all CHD probands, we performed a gene-level meta-analysis on all 248 panel genes using the Poisson *P*-values from DNMs in CHD trios and the FET *P*-values from the case-control. We used the Joint-Local False Discovery Rate ("JL-FDR") method (21) to generate meta-analysis *P*-values and FDR statistics that account for the number of genes tested. To perform the meta-analysis of DNMs and very rare TUVs at a phenotype-level on the 60 genes identified as significantly enriched in all probands, we repeated the meta-analysis procedure using the JL-FDR method to generate

*P*-values and FDR statistics that accounted for the eight cardiac and four NDD/EC phenotypes tested.

When performing meta-analysis of DNMs and TUVs on individual genes, we used Fisher's Method to calculate the meta-analysis *P*-value and applied a Bonferroni correction to determine the threshold for significance. When performing multiple testing correction on the *P*-values from the Poisson test or FET independently, the Benjamini–Hochberg procedure was used to generate FDR statistics.

**Transmission Disequilibrium Test in Rare Variants from WES Trios.** For the 60 significant genes in the MIPseq panel, transmission disequilibrium test (TDT) was performed (79), for rare damaging variants in the 3,887 WES trios. The significance of transmission disequilibrium was calculated by chi-square tests and the genotypic risk ratio was calculated (22).

**Estimating the Total Number of Risk Genes in MIPseq Panel.** We estimated the total number of CHD risk genes in the 248 gene MIPseq panel using the JL-FDR framework (21), fitting a two-component bivariate normal mixture distribution with an EM-algorithm (80) (*SI Appendix*).

**Calculation of Proportion of Cases Attributable to DNMs and Very Rare TUVs.** We calculated the fraction of probands having a likely damaging de novo variant of a tested class from the fraction of probands with such a variant minus the expected fraction from mutability (10), and similarly calculated this from the difference in the observed frequency of TUVs in probands and controls.

**Data, Materials, and Software Availability.** Phenotype and DNA sequence data have been deposited in the NIH National Heart Lung and Blood BioData Catalyst <https://biodatacatalyst.nih.gov> under "congenital heart disease" [phs001194](https://biodatacatalyst.nih.gov/phenotype/001194) (81) and the main PGC substudy [phs000571](https://biodatacatalyst.nih.gov/phenotype/000571) (82).

**ACKNOWLEDGMENTS.** We are grateful to the patients and families who participated in this research and the many individuals who made outstanding contributions to the recruitment of these individuals. We thank S. Zaidi for advice and helpful discussions. This work was supported by grants from the NIH: U01 HL098153 and U11TR000003 from the National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH (E.G.); the Pediatric Cardiac Genomics Consortium (U01 HL131003, U01-HL153009, U01-HL098147, U01-HL128711, U01-HL098162); the Pediatric Heart Network (1UG1HL135680-01); Centers for Mendelian Genomics, 5U54HG006504 (S. Mane and R.P.L.); R03HD100883-A1 (H.Z.); T32HD007149 (M.C.S.); RM1HG011014 (J. Cao); R00HL143036-02 and the Children's Discovery Institute Faculty Scholar award CDI-FR-2021-926 (S.C.J.); American Heart Association Predoctoral Fellowship 19PRE34380842 (W.D.); American Heart Association Career Development Award, and NIH/National Heart Lung and Blood Institute K08 HL157653 (S.U.M.). Funding organizations had no role in the: design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; decision to submit the manuscript for publication.

1. GBD 2017 Congenital Heart Disease Collaborators, Global, regional, and national burden of congenital heart disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet Child. Adolesc. Health* **4**, 185–200 (2020).
2. D. van der Linde *et al.*, Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. *J. Am. Coll. Cardiol.* **58**, 2241–2247 (2011).
3. P. D. Morton, N. Ishibashi, R. A. Jonas, Neurodevelopmental abnormalities and congenital heart disease: Insights into altered brain maturation. *Circ. Res.* **120**, 960–977 (2017).
4. M. El-Chouli *et al.*, Lifetime risk of comorbidity in patients with simple congenital heart disease: A Danish nationwide study. *Eur. Heart J.* **44**, 741–748 (2023).
5. R. Whittemore, J. A. Wells, X. Castellsague, A second-generation study of 427 probands with congenital heart defects and their 837 children. *J. Am. Coll. Cardiol.* **23**, 1459–1467 (1994).
6. N. Oyen *et al.*, Risk of congenital heart defects in offspring of affected mothers and fathers. *Circ. Genom. Precis. Med.* **15**, e003533 (2022).
7. S. Zaidi, M. Brueckner, Genetics and genomics of congenital heart disease. *Circ. Res.* **120**, 923–940 (2017).
8. S. Zaidi *et al.*, De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
9. J. Homsy *et al.*, De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262–1266 (2015).
10. S. C. Jin *et al.*, Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet.* **49**, 1593–1601 (2017).
11. A. Sifrim *et al.*, Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat. Genet.* **48**, 1060–1065 (2016).
12. N. S. Diab *et al.*, Molecular genetics and complex inheritance of congenital heart disease. *Genes (Basel)* **12**, 1020 (2021).

Author affiliations: <sup>a</sup>Department of Genetics, Yale School of Medicine, New Haven, CT 06510; <sup>b</sup>Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, NY 10065; <sup>c</sup>Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110; <sup>d</sup>Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110; <sup>e</sup>Yale Center for Genome Analysis, Yale University, New Haven, CT 06516; <sup>f</sup>Department of Neurosurgery, Yale School of Medicine, New Haven, CT 06510; <sup>g</sup>Yale Program on Neurogenetics, Yale School of Medicine, New Haven, CT 06510; <sup>h</sup>Department of Medical Genetics, School of Medicine, Acibadem University, Istanbul 34752, Türkiye; <sup>i</sup>Department of Translational Medicine, Health Sciences Institute, Acibadem University, Istanbul 34752, Türkiye; <sup>j</sup>Division of Newborn Medicine, Department of Pediatrics, Boston Children's Hospital, Boston, MA 02115; <sup>k</sup>Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA 02115; <sup>l</sup>Broad Institute of Massachusetts Institute of Technology and Harvard, Boston, MA 02142; <sup>m</sup>Department of Biostatistics, Yale School of Public Health, New Haven, CT 06510; <sup>n</sup>Laboratory of Single-Cell Genomics and Population Dynamics, The Rockefeller University, New York, NY 10065; <sup>o</sup>Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI 53706; <sup>p</sup>Department of Biomedical Sciences, Korea University College of Medicine, Seoul, South Korea; <sup>q</sup>Department of Pediatrics, Section of Cardiology, Yale School of Medicine, New Haven, CT 06510; <sup>r</sup>Department of Pediatrics, Michigan State University College of Human Medicine, Grand Rapids, MI 48824; <sup>s</sup>Department of Surgery, Yale University School of Medicine, New Haven, CT 06510; <sup>t</sup>Division of Cardiology, Children's Hospital of Philadelphia, Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; <sup>u</sup>Institute of Cardiovascular Science, University College London, London WC1E 6BT, United Kingdom; <sup>v</sup>Pediatric Cardiology, Great Ormond Street Hospital, London WC1N 3JH, United Kingdom; <sup>w</sup>Division of Cardiology, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON M5G1X8, Canada; <sup>x</sup>Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI 48109; <sup>y</sup>Division of Cardiothoracic Surgery, Children's Hospital of Philadelphia, Philadelphia, PA 19104; <sup>z</sup>Department of Pediatrics, University of Cincinnati, Cincinnati, OH 45229; <sup>aa</sup>Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; <sup>ab</sup>Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; <sup>ac</sup>Gladstone Institute of Cardiovascular Disease and University of California San Francisco, San Francisco, CA 94158; <sup>ad</sup>Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032; <sup>ae</sup>Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10032; <sup>af</sup>Department of Pediatrics, Cardiology, Stanford University, Stanford, CA 94304; <sup>ag</sup>Department of Pediatrics, The School of Medicine and Dentistry, University of Rochester Medical Center, Rochester, NY 14642; <sup>ah</sup>Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115; <sup>ai</sup>Department of Pediatrics, Harvard Medical School, Boston, MA 02115; <sup>aj</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115; <sup>ak</sup>Department of Human Genetics, University of Utah and School of Medicine, Salt Lake City, UT 84112; <sup>al</sup>The Catholic University of America, Washington, DC 20064; <sup>am</sup>Division of Pediatric Cardiology, University of Utah, Salt Lake City, UT 84112; <sup>an</sup>Pediatric Cardiac Surgery, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048; <sup>ao</sup>Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115; <sup>ap</sup>Department of Pediatrics and Medicine, Columbia University Medical Center, New York, NY 10032; <sup>aq</sup>Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029; <sup>ar</sup>Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY 10029; <sup>as</sup>Cardiovascular Division, Brigham and Women's Hospital, Boston, MA 02115; and <sup>at</sup>HHMI, Chevy Chase, MD 20815

Author contributions: M.C.S., K.B., F.L.-G., J.R.K., B.D.G., C.E.S., M.B., and R.P.L. designed research; M.C.S., S.C.J., K.B., F.L.-G., I.T., J.Z., C.N.-W., J.R.K., J. Cao, S. Mane, S.C.S., P.J.G., E.G., J.D., A.G., S. Mital, M.R., J.W.G., E.K., M.W., D.S., Y.S., D.B., G.A.P., J.W.N., M.T.-F., R.K., W.K.C., M.B., and R.P.L. performed research; M.C.S., S.C.J., S.U.M., W.D., W.J., Z.L., B.L., X.Z., Q.L., J. Choi, H.Z., J. Cao, M.L., J.G.S., A.E.R., M.Y., H.J.Y., and R.P.L. analyzed data; B.D.G., C.E.S., and M.B. cohort ascertainment, phenotypic characterization, and recruitment; and M.C.S., M.B., and R.P.L. wrote the paper.

Reviewers: J.S., University of Washington; and M.S., Langley Porter Psychiatric Institute, University of California, San Francisco.

13. K. Schwarze, J. Buchanan, J. C. Taylor, S. Wordsworth, Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet. Med.* **20**, 1122–1130 (2018).
14. B. J. O'Roak *et al.*, Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nat. Commun.* **5**, 5595 (2014).
15. H. A. Stessman *et al.*, Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).
16. T. Wang *et al.*, Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nat. Commun.* **11**, 4932 (2020).
17. J. M. Zook *et al.*, Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci. Data* **3**, 160025 (2016).
18. D. Taliun *et al.*, Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
19. W. Fu *et al.*, Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
20. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
21. W. Jiang, W. Yu, Controlling the joint local false discovery rate is more powerful than meta-analysis methods in joint analysis of summary statistics from multiple genome-wide association studies. *Bioinformatics* **33**, 500–507 (2017).
22. N. Risch, K. Merikangas, The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
23. B. Pijuan-Sala *et al.*, A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).

24. J. Cao *et al.*, A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
25. Y. H. Ching *et al.*, Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat. Genet.* **37**, 423–428 (2005).
26. M. Anfinsen *et al.*, Significance of alpha-myosin heavy chain (MYH6) variants in hypoplastic left heart syndrome and related cardiovascular diseases. *J. Cardiovasc. Dev. Dis.* **9**, 144 (2022).
27. S. J. Bray, Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **17**, 722–735 (2016).
28. M. A. Wouters *et al.*, Evolution of distinct EGF domains with specific functions. *Protein Sci.* **14**, 1091–1103 (2005).
29. A. C. Tien, A. Rajan, H. J. Bellen, A Notch updated. *J. Cell Biol.* **184**, 621–629 (2009).
30. P. C. Weissshuhn *et al.*, Non-linear and flexible regions of the Notch1 extracellular domain revealed by high-resolution structural studies. *Structure* **24**, 555–566 (2016).
31. C. L. Hale, A. N. Niederriter, G. E. Green, D. M. Martin, Atypical phenotypes associated with pathogenic CHD7 variants and a proposal for broadening CHARGE syndrome clinical diagnostic criteria. *Am. J. Med. Genet. A* **170A**, 344–354 (2016).
32. A. Moccia *et al.*, Genetic analysis of CHARGE syndrome identifies overlapping molecular biology. *Genet. Med.* **20**, 1022–1029 (2018).
33. J. K. Meisner, D. M. Martin, Congenital heart defects in CHARGE: The molecular role of CHD7 and effects on cardiac phenotype and clinical outcomes. *Am. J. Med. Genet. C Semin. Med. Genet.* **184**, 81–89 (2020).
34. M. P. Adam *et al.*, Kabuki syndrome: International consensus diagnostic criteria. *J. Med. Genet.* **56**, 89–95 (2019).
35. M. W. Wessels, A. S. Brooks, J. Hoogbeem, M. F. Niermeijer, P. J. Willems, Kabuki syndrome: A review study of three hundred patients. *Clin. Dysmorphol.* **11**, 95–102 (2002).
36. S. Boniel, K. Szymanska, R. Smigiel, K. Szczaluba, Kabuki syndrome-clinical review with molecular aspects. *Genes (Basel)* **12**, 468 (2021).
37. D. Cacciadiferro *et al.*, Dissecting KMT2D missense mutations in Kabuki syndrome patients. *Hum. Mol. Genet.* **27**, 3651–3668 (2018).
38. A. E. Roberts, J. E. Allanson, M. Tartaglia, B. D. Gelb, Noonan syndrome. *Lancet* **381**, 333–342 (2013).
39. M. Zenker, Clinical overview on RASopathies. *Am. J. Med. Genet. C Semin. Med. Genet.* **190**, 414–424 (2022).
40. D. Baldrige *et al.*, Phenotypic expansion of KMT2D-related disorder: Beyond Kabuki syndrome. *Am. J. Med. Genet. A* **182**, 1053–1065 (2020).
41. E. Froimchuk, Y. Jang, K. Ge, Histone H3 lysine 4 methyltransferase KMT2D. *Gene* **627**, 337–342 (2017).
42. V. Faundes, G. Malone, W. G. Newman, S. Banka, A comparative analysis of KMT2D missense variants in Kabuki syndrome, cancers and the general population. *J. Hum. Genet.* **64**, 161–170 (2019).
43. S. Cuvertino *et al.*, A restricted spectrum of missense KMT2D variants cause a multiple malformations disorder distinct from Kabuki syndrome. *Genet. Med.* **22**, 867–877 (2020).
44. M. Motta *et al.*, Dominant Noonan syndrome-causing LZTR1 mutations specifically affect the Kelch domain substrate-recognition surface and enhance RAS-MAPK signaling. *Hum. Mol. Genet.* **28**, 1007–1022 (2019).
45. A. C. Shaw, K. Kalidas, A. H. Crosby, S. Jeffery, M. A. Patton, The natural history of Noonan syndrome: A long-term follow-up study. *Arch. Dis. Child.* **92**, 128–132 (2007).
46. L. Machado, D. A. Critton, R. Page, W. Petit, Redox regulation of a gain-of-function mutation (N308D) in SHP2 Noonan syndrome. *ACS Omega* **2**, 8313–8318 (2017).
47. S. R. Yoon *et al.*, Age-dependent germline mosaicism of the most common Noonan syndrome mutation shows the signature of germline selection. *Am. J. Hum. Genet.* **92**, 917–926 (2013).
48. G. J. Maher *et al.*, Selfish mutations dysregulating RAS-MAPK signaling are pervasive in aged human testes. *Genome Res.* **28**, 1779–1790 (2018).
49. M. W. Breuss *et al.*, Sperm mosaicism: Implications for genomic diversity and disease. *Trends Genet.* **37**, 890–902 (2021).
50. T. Saritas *et al.*, Disruption of CUL3-mediated ubiquitination causes proximal tubule injury and kidney fibrosis. *Sci. Rep.* **9**, 4596 (2019).
51. I. Harel *et al.*, Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18839–18844 (2012).
52. H. Haase *et al.*, Ahnak is critical for cardiac Ca(V)1.2 calcium channel function and its beta-adrenergic regulation. *FASEB J.* **19**, 1969–1977 (2005).
53. B. Coxam *et al.*, Svp1 stabilises developmental vascular anastomosis in reduced flow conditions. *Development* **149**, dev199858 (2022).
54. Y. Wang *et al.*, NR6A1 couples with cAMP response element binding protein and regulates vascular smooth muscle cell migration. *Int. J. Biochem. Cell Biol.* **69**, 225–232 (2015).
55. N. Lin, Y. Cai, L. Zhang, Y. Chen, Identification of key genes associated with congenital heart defects in embryos of diabetic mice. *Mol. Med. Rep.* **17**, 3697–3707 (2018).
56. S. Singh *et al.*, Musashi-2 causes cardiac hypertrophy and heart failure by inducing mitochondrial dysfunction through destabilizing Clu and Smyd1 mRNA. *Basic Res. Cardiol.* **118**, 46 (2023).
57. D. Szklarczyk *et al.*, STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
58. V. Garg *et al.*, Mutations in NOTCH1 cause aortic valve disease. *Nature* **437**, 270–274 (2005).
59. A. Joutel *et al.*, Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients. *Lancet* **350**, 1511–1515 (1997).
60. D. Schoemaker, J. F. Arboleda-Velasquez, Notch3 signaling and aggregation as targets for the treatment of CADASIL and other NOTCH3-associated small-vessel diseases. *Am. J. Pathol.* **191**, 1856–1870 (2021).
61. C. G. Furey *et al.*, De novo mutation in genes regulating neural stem cell fate in human congenital hydrocephalus. *Neuron* **99**, 302–314.e4 (2018).
62. W. Dong *et al.*, Recessive genetic contribution to congenital heart disease in 5,424 probands. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2419992122 (2025).
63. F. Richter *et al.*, Genomic analyses implicate noncoding de novo variants in congenital heart disease. *Nat. Genet.* **52**, 769–777 (2020).
64. M. Ameen *et al.*, Integrative single-cell analysis of cardiogenesis identifies developmental trajectories and non-coding mutations in congenital heart disease. *Cell* **185**, 4937–4953.e23 (2022).
65. H. J. Cordell *et al.*, Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat. Genet.* **45**, 822–824 (2013).
66. N. A. Hanchard *et al.*, A genome-wide association study of congenital cardiovascular left-sided lesions shows association with a locus on chromosome 20. *Hum. Mol. Genet.* **25**, 2331–2341 (2016).
67. S. Liu *et al.*, Association between maternal chronic conditions and congenital heart defects: A population-based cohort study. *Circulation* **128**, 583–589 (2013).
68. B. Gelb *et al.*, The congenital heart disease genetic network study: Rationale, design, and early results. *Circ. Res.* **112**, 698–706 (2013).
69. E. A. Boyle, B. J. O’Roak, B. K. Martin, A. Kumar, J. Shendure, MIPgen: Optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* **30**, 2670–2672 (2014).
70. D. Karolchik *et al.*, The UCSC table browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
71. S. Cantilieri *et al.*, Targeted capture and high-throughput sequencing using molecular inversion probes (MIPs). *Methods Mol. Biol.* **1492**, 95–106 (2017).
72. G. A. Van der Auwera *et al.*, From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).
73. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
74. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2992 (2011).
75. Q. Wei *et al.*, A Bayesian framework for de novo mutation calling in parents-offspring trios. *Bioinformatics* **31**, 1375–1381 (2015).
76. P. I. de Bakker *et al.*, Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
77. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. Accessed 29 October 2017.
78. M. Haessler *et al.*, The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).
79. Z. He *et al.*, Rare-variant extensions of the transmission disequilibrium test: Application to autism exome sequence data. *Am. J. Hum. Genet.* **94**, 33–46 (2014).
80. J. Wakefield, A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
81. Pediatric Cardiac Genomics Consortium, Data from “Genomic analysis of 11,555 pro bands identifies 60 dominant congenital heart disease genes.” NCBI’s Database of Phenotypes and Genotypes (dbGaP). [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001194](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001194). Deposition 31 March 2025.
82. Pediatric Cardiac Genomics Consortium, Data from “Genomic analysis of 11,555 pro bands identifies 60 dominant congenital heart disease genes.” NCBI’s Database of Phenotypes and Genotypes (dbGaP). [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000571](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000571). Deposition 31 March 2025.