Supplementary Materials for

De novo mutations in Congenital Heart Disease with Neurodevelopmental and Other Birth Defects

Jason Homsy[§], Samir Zaidi[§], Yufeng Shen[§], James S. Ware[§], Kaitlin E. Samocha, Konrad J. Karczewski, Steven R. DePalma, David McKean, Hiroko Wakimoto, Josh Gorham, Sheng Chih Jin, John Deanfield, Alessandro Giardini, George A. Porter Jr., Richard Kim, Kaya Bilguvar, Francesc Lopez, Irina Tikhonova, Shrikant Mane, Angela Romano-Adesman, Hongjian Qi, Badri Vardarajan, Lijiang Ma, Mark Daly, Amy E. Roberts, Mark W. Russell, Seema Mital, Jane W. Newburger, J. William Gaynor, Roger E. Breitbart, Ivan Iossifov, Michael Ronemus, Stephan J. Sanders, Jonathan R. Kaltman, Jonathan G. Seidman, Martina Brueckner[†], Bruce D. Gelb[†], Elizabeth Goldmuntz[†], Richard P. Lifton[†,*], Christine E. Seidman[†,*], Wendy K. Chung[†,*]

[§]These authors contributed equally to this work.
[*]Co-senior authors
[†]Correspondence to: bruce.gelb@mssm.edu, goldmuntz@email.chop.edu, martina.brueckner@yale.edu, richard.lifton@yale.edu, csediman@genetics.med.harvard.edu, or wkc15@cumc.columbia.edu

This PDF file includes:

Other Suplementary Materials for this manuscript includes the following:

Databases S1 to S10 as zipped archives:

Database S1: Phenotypes for each case proband, including cardiac, neurodevelopmental disorders and extra-cardiac congenital anomalies.
Database S2: List of *de novo* Mutations in CHD case cohort.
Database S3: List of *de novo* Mutations in Control cohort.
Database S4: List of *de novo* probabilities for each variant class in each protein-coding gene on the Nimblegen V2 exome, adjusted for depth in Cases.
Database S5: List of *de novo* probabilities for each variant class in each protein-coding gene on the Nimblegen V2 exome, adjusted for depth in Controls.
Database S6: Functional term enrichment analysis of all Genes with Damaging (loss of function + deleterious missense) *de novo* mutations in all cases.
Database S7: Functional term enrichment analysis of all Genes with Loss of Function *de novo* mutations in 860 new cases.
Database S8: List of 1,563 variants (1,161 unique genes) with damaging *de novo* mutations from 7 independent NDD cohorts.
Database S9: Functional term enrichment analysis among 69 genes with Damaging *de novo* mutations overlapping between CHD cases and the published NDD (P-NDD) cohort.
Database S10: Percentile ranks of genes by expression in the developing mouse heart and brain.

Abbreviations and Conventions used in this Supplement

HHE: High Heart Expressed genes (genes in the top quartile of expression)
LHE: Lower Heart Expressed genes (genes in the bottom three quartiles of expression)
HBE: High Brain Expressed genes (genes in the top quartile of expression)
LBE: Lower Brain Expressed genes (genes in the bottom three quartiles of expression)
LoF: Loss of function
D-Mis: Deleterious missense variants as predicted by Meta-SVM
NDD: Neurodevelopmental disorders
CA: Congenital Anomaly (extra-cardiac)
Bold face in tables: OR or Enrichments > 2; $p < 0.005$

Materials and Methods

Patient cohorts

Pediatric Cardiac Genomics Consortium (PCGC):

Probands were recruited from 10 centers in the United States and United Kingdom (Congenital Heart Disease Genetic Network Study of the PCGC) (*4*). The protocol was approved by the Institutional Review Boards of Boston Children's Hospital, Brigham and Women's Hospital, Great Ormond St. Hospital, Children's Hospital of Los Angeles, Children's Hospital of Philadelphia, Columbia University Medical Center, Icahn School of Medicine at Mount Sinai, Rochester School of Medicine and Dentistry, Steven and Alexandra Cohen Children's Medical Center of New York, and Yale School of Medicine. Written informed consent was obtained from participating subjects or their parents. Probands without any first-degree relative with CHD were selected. Cardiac diagnoses were obtained from review of echocardiogram, catheterization and operative reports. Extra-cardiac congenital anomalies were determined from review of medical records. Determination of neurodevelopmental status was based on parental report when the proband was at least 12 months of age (answering "Yes" to the presence of at least one of the following conditions: developmental delay, learning disability, mental retardation, or autism).

Pediatric Heart Network:

Samples were selected from the DNA biorepository of the Single Ventricle Reconstruction (SVR) trial (*5*)**.** In this trial, subjects underwent in-person neurodevelopmental testing at age 14 months with the Psychomotor Development Index (PDI) and Mental Development Index (MDI) of the Bayley Scales of Infant Development-II (*24*). In addition, subjects were evaluated with the Ages and Stages Questionnaire (ASQ), from which scores at age 3 years were analyzed (*24*). Probands were defined as having NDD if they had a PDI or MDI score of <70, or an "at risk" score in at least one of the five domains of the ASQ at 3 years. Blood or sputum samples for DNA isolation were collected from parent-child trios at or after the 3-year follow-up visit.

Controls:
Control trios were kindly provided by the Simons Foundation Autism Research Initiative Simplex Collection. Simplex families (two unaffected parents, one child with autism spectrum disorder, and one unaffected sibling) underwent whole exome sequencing (*7-9*). Trios of unaffected family members served as controls for our study.

Cardiac Phenotypes
Cardiac phenotypes were divided into 5 major categories (Table S1) on basis of the major cardiac lesion: conotruncal defects (CTD, n=425), left ventricular outflow tract obstruction (LVO, n=426), heterotaxy (HTX, n=195), Other (n=131), Complex (n=36). CTD phenotypes include tetralogy of Fallot, D-loop transposition of the great arteries, double-outlet right ventricle, truncus arteriosus, ventricular septal defects, and aortic arch abnormalities. LVO phenotypes include hypoplastic left heart syndrome, coarctation of the aorta, and aortic stenosis/bicuspid aortic valve. HTX syndromes include left-right isomerism as the major malformation, and may include other defects such as transposition of the great arteries, atrioventricular canal defects, anomalous pulmonary venous drainage, and double outlet right ventricle. Isomerism of other organs was not considered a separate extra-cardiac CA for this study. Lesions in the "other" category include pulmonary valve abnormalities, anomalous pulmonary venous drainage, atrial septal defects, atrioventricular canal defects, double inlet left ventricle and tricuspid valve atresia.

Exome Sequencing
Case trios were sequenced at the Yale Center for Genome Analysis as described previously (*6*). Genomic DNA from whole blood was captured with the NimbleGen v2.0 exome capture reagent (Roche) and sequenced (Illumina HiSeq 2000, 75 base paired-end reads). Sequence data from control subjects were analyzed. These samples were captured and sequenced using the same methods (59% at Yale School of Medicine, 38% at Cold Spring Harbor, 3% at Washington University in St. Louis). Sequencing metrics are provided in Table S12. Reads were processed via three independent analysis pipelines at Harvard Medical School and Yale School of Medicine, and Columbia University Medical Center. At Harvard, reads were mapped to the human reference genome hg19 with Novoalign (www.novocraft.com) and processed according to the best practices of the Genome Analysis Toolkit 3.0 (*25*). Variant calls were made with GATK HaplotypeCaller and annotated with Meta-SVM (*11*) and other annotations using dbNSFP (*26*). At Columbia, reads were mapped to hg19 using BWA-mem (*27*). At Yale, reads were mapped to the human reference hg19 with BWA-mem and variants were called using SAMtools. Variants were annotated with an in-house pipeline and *de novo* mutations were called using a Bayesian algorithm as previously described (*6*). Exome sequence data for case trios are deposited in dbGaP (accession phs000571.v2.p1).

Identification of *de novo* mutations
A high stringency method for selecting *de novo* mutations was used because DNA from controls subjects (recruited by the Simons Foundation Autism Research Initiative) was not available for confirmation by Sanger sequencing. Rare (AF < 0.04%) on-target heterozygous SNVs or indels were initially identified based on presence in the child and

absence in the parents. Candidate *de novo* mutations were pooled from all three pipelines and then further filtered based on depth (minimum 10 reads total and 5 alternate allele reads), alternate allele balance (minimum 20% if alternate read depth greater than or equal to 10 or minimum 28% if alternate read depth less than 10), and parental read characteristics (minimum depth of 10 reference reads; alternate allele balance less than 3.5%). False positives were removed by *in silico* visualization. Of 409 case variants that were submitted for validation by Sanger sequencing, 394 were confirmed (specificity 96.3%). The frequency of *de novo* mutations detected per proband followed an expected Poisson distribution in both case and control cohorts (Fig. S3).

Mutational Model

We used sequence context to derive the probability of observing a *de novo* variant in each gene as described previously (*10*). All protein-coding Gencode transcripts intersecting with target regions present on the NimbleGen v2.0 capture array and with annotations in the dbNSFP database were considered (*26*). Briefly, for each base in the exome, the probability of observing each of the three possible single nucleotide substitutions was determined. The coding consequence of each potential substitution was determined, and the probability of mutation was summed for each variant class (synonymous, missense, nonsense, essential splice site, frameshift, start lost, stop lost) and for each gene. The probability of a frameshift mutation was determined by multiplying the probability of a nonsense mutation by 1.25 as described previously (*10*). In-frame insertions or deletions are currently not accounted for by the model and were not considered in the analysis. Deleterious missense prediction by Meta-SVM (*11*) was provided by dbNSFP (version 2.8, corresponding to a dbNSFP Meta-SVM rank score of greater than 0.83357). Where conflicts between two or more transcripts of a given gene occurred, the annotation corresponding to the longest transcript was taken. Each probability was adjusted to control for variable sequencing coverage as previously described: the raw probability was multiplied by a factor in the range 0.9 - 1, according to the percentage of trios covering that base with at least 10x depth. Positions with a coverage of zero resulted in a probability of zero for that base. The sequencing coverage adjustment was calculated separately for cases and controls, which were sequenced in separate batches. Each probability was further adjusted by a divergence score (derived from the number of divergent sites between humans-macaques for the gene region as well as 1MB upstream and downstream), as previously described (*10*).

Statistical Analyses

Statistical analyses were performed using R. The “denovolyzeR” package (http://denovolyzer.org, downloadable from the Comprehensive R Archive Network) implemented analyses based on the mutational model above. A detailed protocol is available (*28*).

Global or gene set burden:

Briefly, the expected number of *de novo* mutations in case and control cohorts across each variant class was determined by taking the sum of each class-specific probability multiplied by the number of probands in the study, multiplied by 2 (for diploid genomes). We tested for an excess of observed *de novo* mutations over expectation using Poisson

statistics. For gene set enrichment, the expected rates were obtained from the probabilities corresponding to that gene set only.

Number of genes with multiple de novo mutations:
The number of genes containing multiple damaging *de novo* mutations was compared with an empirical distribution derived by permutation. The number of *de novo* mutations observed in each class was randomly distributed across all genes, weighted according to gene mutability. For each permutation, the number of genes with multiple mutations was tallied. For both cases and controls, 1 million permutations were performed.

Single genes with multiple de novo mutations:
For each gene, the expected number of *de novo* mutations for each class (LoF, D-Mis and damaging) was calculated from the corresponding *de novo* probability adjusting for cohort size. The number of *de novo* mutations for each gene was compared to that expected using a Poisson test. For each gene, we compared the number of LoF variants, D-Mis variants, and damaging variants (=LoF + D-Mis), using a Bonferroni corrected significance threshold ($9x10^{-7}$, i.e.: (0.05 / (18,515 genes * 3 tests)).

Percent of CHD attributable to de novo variation:
For each phenotype under consideration (such as CHD plus both NDD and CA) the fraction of individuals carrying at least one damaging *de novo* mutation in an HHE gene was determined. From this rate, the expected rate of damaging *de novo* mutations in HHE genes per individual was subtracted. The expected rate was estimated by 10,000 simulations of a cohort of 1,213 subjects harboring the expected number of mutations (~76; Table 1) distributed with replacement among the simulated subjects designed to carry an approximate Poisson distribution of mutations per person (~427 individuals carry 0 mutations, 448 carry 1, 235 carry 2, etc). Using this method, the expected rate was determined to be 0.06.

Estimation of the number of risk genes:
We followed a previously described method (*7*) to estimate the number of risk genes that are *de novo* mutation targets with two modifications: (a) we used our gene-specific background mutation rates to replace gene-size based background rates (Database S4); (b) we counted the number of genes mutated in two cases and three (or more) cases separately, and utilized such information to stabilize likelihood estimation at the lower range. An implicit assumption is that the penetrance of all damaging mutations in all risk genes is identical.

We started with the number of observed damaging mutations (K) in HHE among all cases, the observed number of HHE genes mutated twice R1 and mutated at least three times R2. We set the fraction (E) of damaging mutations in risk genes based on point estimate of enrichment in cases compared to expectation (E = (M1 – M2)/M1, where M1 and M2 are average number of damaging mutations per subject in cases and expectation, respectively).

We then estimated the likelihood of risk gene number L(G). Specifically, in the simulation, the number of damaging mutations (K) was fixed by the observed value in cases. We generated simulated mutations from two parts, i.e., risk HHE genes and random HHE genes. We first randomly selected G risk genes from all HHE genes with equal probability. We set the number of contributing damaging mutations in risk genes (C1) by sampling from a binomial distribution Binom(K,E), and the number of non-contributing damaging mutations C2 = K-C1. We then simulated C1 contributing damaging mutations by sampling with replacement from G genes and C2 non-contributed mutations from all HHE genes using their background mutation rate as probability weights.

We performed 20,000 simulations for every G from 1 to 1000, and set L(G) to be the proportion of simulations in which the number of genes with two damaging mutations is exactly R1 and the number of genes with more than two damaging mutations is exactly R2.

RNA sequencing

RNA sequencing on mouse hearts at embryonic day 14.5 was performed as described (*6*). For the developing brain dataset, mouse brains were harvested at embryonic day 9.5 from total of 36 embryos from 5 pregnant females (129SvEv background). Tissues containing the procencephalon, mesencephalon and metencephalon were pooled and RNA was extracted. RNA quality was confirmed with all RNA integrated number greater than 6.3. Two rounds of mRNA purification (polyA-selection) were performed on pooled total RNA using the Dynabeads mRNA DIRECT Kit (Life Technologies). cDNA was generated using the Superscript III First-Strand Synthesis System (Life Technologies) and cDNA libraries were constructed using the Nextera XT DNA Sample Preparation Kit (Illumina). Sequencing was performed to a depth of > 30 million paired-end 50-base reads. Reads were processed as described previously (*6*), and gene expression was calculated as fragments per kb of transcript per million fragments mapped (FPKM). Gene expression values are expressed as ranked centiles (Database S10; calculated for each heart and brain data sets) for comparison.

Gene group functional profiling

Functional gene group analysis was performed using g:Profiler (*29*). Genes with damaging *de novo* mutations were queried using a background gene set of Nimblegen v2.0 genes. All p-values were Bonferroni corrected and only statistically significant terms were displayed.

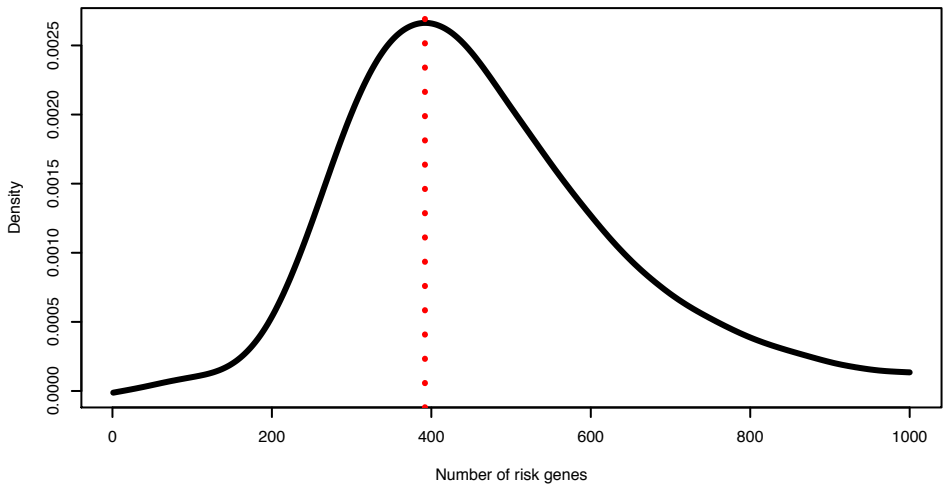

Fig. S1: Estimation of the number of Risk Genes in the top quartile of heart expression (HHE) by simulation. The maxiumum likelihood esitmate is ~392 genes. See Methods for details.

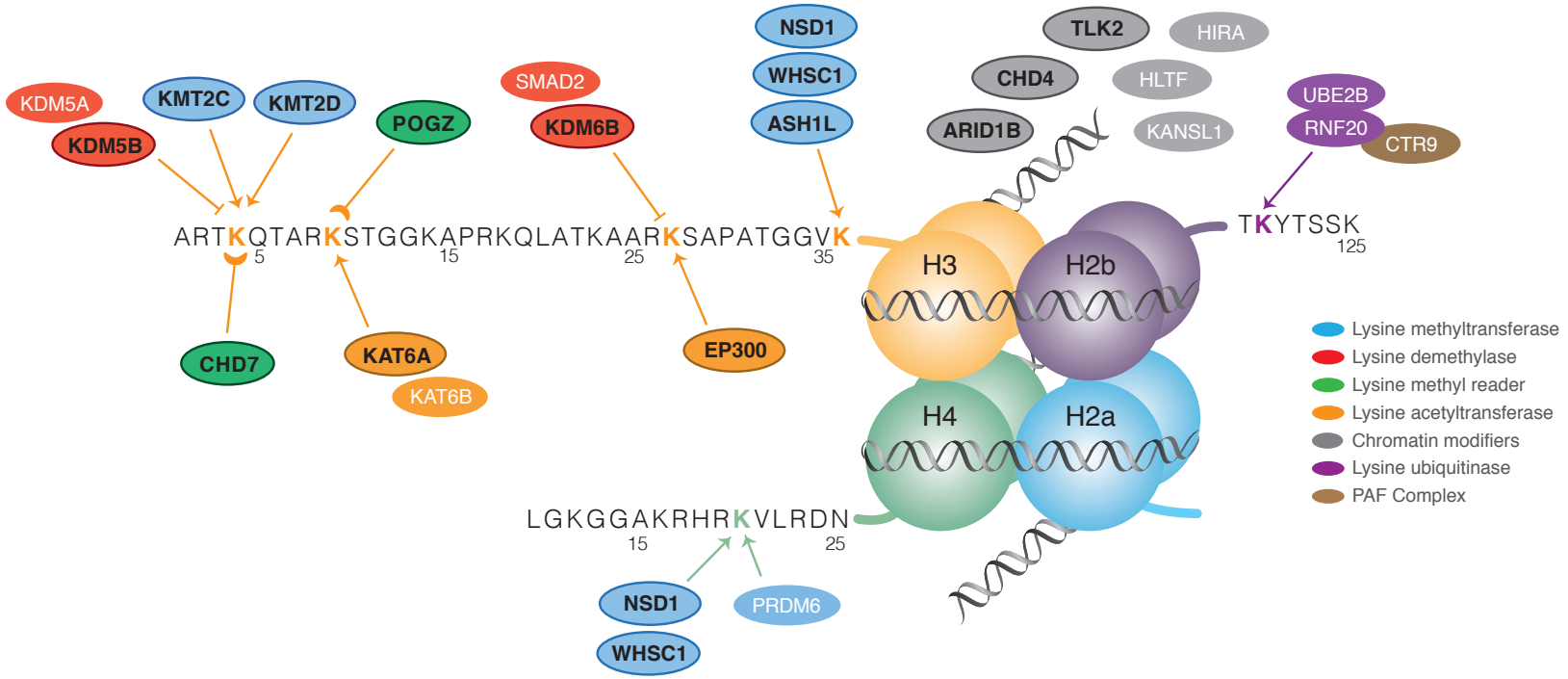

Fig. S2: *De Novo* Mutations in Chromatin Modification Genes in CHD Cases Overlap with Mutated NDD Genes. Nucleosome with numerous chromatin marks, namely H3K4, H3K9, H3K27, H3K36, and H4K20 methylation and/or acetylation, and H2BK120 ubiquitylation, are shown. Mutated proteins are marked within circles and relate to function by color. Genes in bolded black text are mutated with damaging de novo mutations in both CHD and NDD cohorts, while genes in white text are mutated uniquely in the CHD cohort. Genes with overlapping circles are found in a complex, e.g. *RNF20* and *UBE2B*.

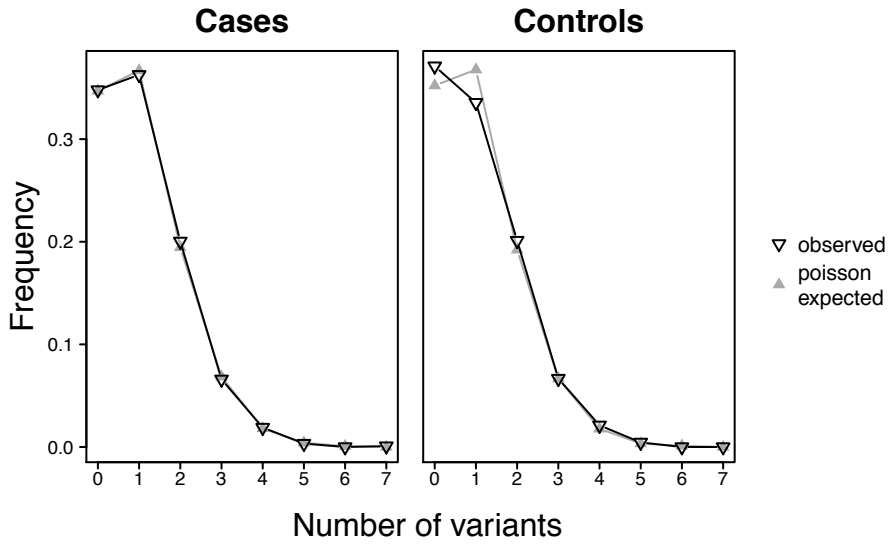

Fig. S3: Distribution of *de novo* mutations per person follows an expected Poisson distribution.

Table S1: Overview of Cardiac Phenotypes

	N
CTD (425)	
Tetralogy of Fallot	224
Transposition of the Great Arteries, D-Loop	101
Double Outlet Right Ventricle	39
Truncus Arteriosus	28
Ventricular Septal Defect (conoventricular)	14
Aortic Arch Anomalies/Interrupted Aortic Arch	16
Other Conotruncal Lesions	3
LVO (426)	
Hypoplastic Left Heart Syndrome	249
Coarctation of the Aorta	101
Aortic Stenosis/Bicuspid Aortic Valve	67
Other Left-sided Lesions	9
HTX (195)	
Transposition of the Great Arteries, L-Loop	67
Double Outlet Right Ventricle	53
Anomalous Venous Drainage (systemic and/or pulmonic)	16
Atrioventricular Canal Defect	14
Transposition of the Great Arteries, D-Loop	14
Other	31
Other (131)	
Pulmonary Valve Atresia/Pulmonary Valve Stenosis	33
Atrial Septal Defect (secundum)	32
Tricuspid Valve Atresia	7
Atrioventricular Canal Defect	27
Anomalous Pulmonary Venous Return	10
Other	22
Complex	36
Total	1213

CTD: Conotruncal Defects; LVO: Left Ventricular Outflow Tract Obstruction; HTX: Heterotaxy. Cases appear only once in table and were assigned to categories based on major cardiac malformation. Additional findings (e.g. *situs* abnormalities diagnostic of HTX) account for a cardiac malformation in two categories.

Table S2: *de novo* Burden Analysis by Case vs. Control Comparison Demonstrates Results Concordant to the Expecation Analysis

	Number of Variants		OR	p
	Cases, *N*=1213	Controls, *N*=900		
All genes				
Total	1273	925	1.1	0.64
Synonymous	277	229	0.9	0.23
Missense	846	614	1.1	0.69
D-Mis	212	119	1.3	0.014
LoF	150	82	1.4	0.028
Damaging	362	201	1.4	**0.00088**
HHE genes				
Total	448	272	1.3	0.0093
Synonymous	81	80	0.8	0.079
Missense	288	164	1.3	0.0067
D-Mis	99	37	2	**0.00025**
LoF	79	28	2.1	**0.00056**
Damaging	178	65	2.2	**$3.5x10^{-07}$**
LHE genes				
Total	825	653	1	0.22
Synonymous	196	149	1	0.83
Missense	558	450	1	0.19
D-Mis	113	82	1	0.94
LoF	71	54	0.9	0.93
Damaging	184	136	1	1

HHE: High Heart Expressed genes (top quartile of expression). LHE: Lower Heart Expressed genes (bottom three quartiles of expression). D-Mis: Deleterious Missense predicted by Meta-SVM. P-value based on binomial test. Damaging: D-Mis + LoF. Bold: OR > 2 or p < 0.005.

Table S3: *de novo* Expectation Analysis of 860 new CHD cases only

	Observed		Expected		Enrichment	p
	n	Rate	n	Rate		
All genes						
Total	916	0.76	930.7	0.77	1.0	0.69
Synonymous	193	0.16	263.3	0.22	0.7	1
Missense	605	0.50	584.8	0.48	1.0	0.21
D-Mis	154	0.13	94.4	0.08	1.6	**$1.1x10^{-08}$**
LoF	118	0.10	82.6	0.07	1.4	**0.00014**
Damaging	272	0.22	176.9	0.15	1.5	**$2.1x10^{-11}$**
HHE genes						
Total	326	0.27	264.0	0.22	1.2	**0.00013**
Synonymous	58	0.05	73.4	0.06	0.8	0.97
Missense	207	0.17	166.1	0.14	1.2	**0.0012**
D-Mis	72	0.06	28.8	0.02	**2.5**	**$1x10^{-11}$**
LoF	61	0.05	24.5	0.02	**2.5**	**$3.8x10^{-10}$**
Damaging	133	0.11	53.3	0.04	**2.5**	**$3.4x10^{-20}$**
LHE genes						
Total	590	0.49	666.7	0.55	0.9	1
Synonymous	135	0.11	189.9	0.16	0.7	1
Missense	398	0.33	418.7	0.35	1.0	0.85
D-Mis	82	0.07	65.5	0.05	1.2	0.028
LoF	57	0.05	58.1	0.05	1.0	0.58
Damaging	139	0.11	123.7	0.10	1.1	0.093

n: Number of *de novo* mutations. Rate: Number of *de novo* mutations divided by number of individuals in cohort (*N*). HHE: High Heart Expressed genes (top quartile of expression). LHE: Lower Heart Expressed genes (bottom three quartiles of expression). D-Mis: Deleterious Missense predicted by Meta-SVM. Damaging: D-Mis + LoF. Bold: Enrichment > 2 or p < 0.005

Table S4: *de novo* Expectation Analysis by Cardiac Phenotype in HHE Genes

	Observed		Expected		Enrichment	p
	n	Rate	n	Rate		
CTD (425)						
Synonymous	31	0.073	36.3	0.085	0.855	0.83
D-Mis	39	0.092	14.2	0.033	**2.740**	**$4.8x10^{-08}$**
LoF	20	0.047	12.1	0.028	1.650	0.023
Damaging	59	0.140	26.3	0.062	**2.240**	**$3x10^{-08}$**
LVO (426)						
Synonymous	28	0.066	36.4	0.085	0.770	0.93
D-Mis	33	0.077	14.3	0.034	**2.310**	**$1.6x10^{-05}$**
LoF	37	0.087	12.1	0.028	**3.050**	**$7x10^{-09}$**
Damaging	70	0.160	26.4	0.062	**2.650**	**$1.4x10^{-12}$**
Other (131)						
Synonymous	10	0.076	11.2	0.085	0.894	0.68
D-Mis	15	0.110	4.4	0.034	**3.420**	**$5.6x10^{-05}$**
LoF	12	0.092	3.7	0.028	**3.220**	**0.0005**
Damaging	27	0.210	8.1	0.062	**3.330**	**$1.4x10^{-07}$**
HTX (195)						
Synonymous	10	0.051	16.6	0.085	0.601	0.97
D-Mis	8	0.041	6.5	0.033	1.220	0.33
LoF	9	0.046	5.5	0.028	1.620	0.11
Damaging	17	0.087	12.1	0.062	1.410	0.11

CTD: Conotruncal Defects; LVO: Left Ventricular Outflow Tract Obstruction; HTX: Heterotaxy. Cardiac Phenotype Categories are defined further in Methods. Analysis performed on a minimum of 50 probands per category, thus table excludes 36 probands with a "Complex" phenotype that were unable to be categorized in either CTD, LVO or HTX. n: Number of *de novo* mutations. Rate: Number of *de novo* mutations divided by number of individuals (provided in parenthesis next to category title) in each category. Bold: OR > 2 or p < 0.005.

Table S5: Genes with Multiple *de novo* Mutations in Controls (Observed vs. Expected, 1M Permutations)

	Observed	Median Expected	Max Expected	p
Synonymous	3	3	13	0.5
D-Mis	4	1	10	0.02
LoF	2	0	5	0.05
Damaging	6	2	12	0.03

Shown are the Observed number of genes with multiple *de novo* mutations for each indicated variant class, including the Median and Maximum expected number of genes with multiple hits seen in 1 million permutations. P-values were calculated by permutation. Significance threshold correcting for 4 tests is 0.01.

Table S6: Characteristics of Genes with Multiple Damaging *de novo* Mutations in Cases

Gene	LoF	D-Mis	p	Cardiac Pheno	NDD Pheno	CA Pheno	Rank Heart Expr	Rank Brain Expr
PTPN11	0	4	2.9×10^{-11}	O/O/C/O	+/+/+/+	-/+/-/+	94	90
KMT2D	4	2	4.1×10^{-09}	L/L/L/C/L/L	+/U/U/U/+/+	-/+/-/+/+/+	97	85
RBFOX2	3	0	3.4×10^{-08}	L/L/L	U/U/+	-/-/-	98	100
KDM5B	3	0	2.9×10^{-06}	L/O/H	U/+/+	-/+/-	86	91
KRT13	0	2	1.0×10^{-05}	C/L	U/+	-/+	16	0
MYH6	0	3	2.4×10^{-05}	L/L/C	U/U/U	+/-/U	100	42
CAD	0	3	3.7×10^{-05}	C/H/O	+/+/+	+/-/+	86	98
NAA15	2	0	4.7×10^{-05}	C/H	-/-	-/+	97	95
SMAD2	1	1	1.1×10^{-04}	H/H	-/U	-/+	75	79
RABGAP1L	1	1	4.0×10^{-04}	L/L	U/-	-/-	78	39
POGZ	1	1	4.3×10^{-04}	O/L	+/-	+/-	84	81
JAG1	1	1	4.5×10^{-04}	L/C	-/U	-/-	77	81
GANAB	1	1	4.5×10^{-04}	L/L	+/+	-/-	94	98
DTNA	1	1	4.7×10^{-04}	C/C	-/+	-/-	47	48
PPL	1	1	6.0×10^{-04}	C/O	U/+	+/+	47	37
CHD7	2	0	6.2×10^{-04}	C/O	+/U	+/+	93	89
ZEB2	1	1	6.2×10^{-04}	C/O	U/+	-/+	82	63
FBN1	0	2	6.8×10^{-04}	C/L	+/-	-/+	93	61
CHD4	0	2	1.2×10^{-03}	O/L	+/+	+/+	99	99
AHNAK	1	1	2.9×10^{-03}	O/L/C	+/U/+	-/-/-	96	56
NOTCH1	1	1	4.4×10^{-03}	L/C/L	U/U/+	+/-/-	88	92

The phenotypes (Cardiac, NDD, CA) for each proband with a mutation in the given gene are shown in respective order. Cardiac phenotypes are abbreviated as L (Left Ventricular Outflow Tract Obstruction); C (Conotruncal defects); H (Heterotaxy); O (Other). U: Unknown. Expression data are given as ranked centiles (i.e. ordered from 0 to 100, low to high). The cutoff for high heart or high brain expression is >= 75.

Table S7: *de novo* Enrichment Analysis in 2234 *RBFOX2* Target Genes

	Cases, *N*=1213				**Controls, *N*=900**			
	Observed	Expected	Enrichment	p	Observed	Expected	Enrichment	p
Total	232	202.3	1.1	0.022	141	150.8	0.9	0.8
Syn	44	56.2	0.8	0.96	41	41.9	1.0	0.58
Missense	148	127.1	1.2	0.038	78	94.7	0.8	0.96
D-Mis	41	23.1	1.8	**0.00047**	15	17.2	0.9	0.73
LoF	40	19.0	**2.1**	**$1.8x10^{-05}$**	22	14.2	1.6	0.032
Damaging	81	42.1	1.9	**$6.6x10^{-08}$**	37	31.4	1.2	0.18

Shown are the observed and expected numbers of *de novo* variants restricted to a set of 2,234 *RBFOX2* target genes (gene set enrichment analysis) *(16)*. Enrichment = Observed / Expected. P-values are from Poisson test comparing Observed vs. Expected.

Table S8: *de novo* Enrichment Analysis in 524 Genes involved in Chromatin Modification (*GO:0016568*)

	Cases, *N*=1213				**Controls, *N*=900**			
	Observed	Expected	Enrichment	p	Observed	Expected	Enrichment	p
Total	74	49.1	1.5	**0.00056**	29	36.6	0.8	0.91
Syn	2	13.4	0.1	1	6	10.0	0.6	0.93
Missense	47	31.0	1.5	**0.0044**	20	23.0	0.9	0.77
D-Mis	11	5.3	**2.1**	0.02	3	3.9	0.8	0.75
LoF	25	4.8	**5.3**	$\mathbf{5.7x10^{-11}}$	3	3.5	0.8	0.69
Damaging	36	10.0	**3.6**	$\mathbf{1.8x10^{-10}}$	6	7.5	0.8	0.75

Shown are the observed and expected numbers of *de novo* variants restricted to a set of 524 genes in the *GO:0016568* category (gene set enrichment analysis). Enrichment = Observed / Expected. P-values are from Poisson test comparing Observed vs. Expected.

Table S9: Gene set *de novo* Enrichment Analysis in Cases and Controls in all Genes, Genes in the Top Quartile or Lower 75th Percentile of the Developing Brain

	Cases, *N*=1213						**Controls, *N*=900**					
	Observed		Expected		Enrichment	p	Observed		Expected		Enrichment	p
	n	Rate	n	Rate			n	Rate	n	Rate		
All genes												
Total	1273	1.05	1312.7	1.08	1.0	0.87	925	1.03	979.7	1.09	0.9	0.96
Synonymous	277	0.23	371.4	0.31	0.7	1	229	0.25	277.4	0.31	0.8	1
Missense	846	0.70	824.9	0.68	1.0	0.24	614	0.68	615.6	0.68	1.0	0.53
D-Mis	212	0.17	133.1	0.11	1.6	**$1.8x10^{-10}$**	119	0.13	99.3	0.11	1.2	0.03
LoF	150	0.12	116.5	0.10	1.3	**0.0016**	82	0.09	86.7	0.10	0.9	0.71
Damaging	362	0.30	249.5	0.21	1.4	**$1.5x10^{-11}$**	201	0.22	186.0	0.21	1.1	0.14
HBE genes												
Total	404	0.33	346.4	0.29	1.2	**0.0014**	253	0.28	258.3	0.29	1.0	0.64
Synonymous	77	0.06	96.8	0.08	0.8	0.98	66	0.07	72.2	0.08	0.9	0.78
Missense	255	0.21	217.8	0.18	1.2	0.0075	166	0.18	162.4	0.18	1.0	0.4
D-Mis	77	0.06	36.1	0.03	**2.1**	**$2.3x10^{-09}$**	30	0.03	26.9	0.03	1.1	0.3
LoF	72	0.06	31.9	0.03	**2.3**	**$7.2x10^{-10}$**	21	0.02	23.7	0.03	0.9	0.74
Damaging	149	0.12	68.0	0.06	**2.2**	**$1.5x10^{-17}$**	51	0.06	50.6	0.06	1.0	0.5
LBE genes												
Total	815	0.67	892.1	0.74	0.9	1	634	0.70	666.2	0.74	1.0	0.9
Synonymous	191	0.16	253.6	0.21	0.8	1	155	0.17	189.5	0.21	0.8	1
Missense	552	0.46	560.3	0.46	1.0	0.64	423	0.47	418.4	0.46	1.0	0.42
D-Mis	131	0.11	91.7	0.08	1.4	**$6.5x10^{-05}$**	89	0.10	68.4	0.08	1.3	0.0097
LoF	72	0.06	78.2	0.06	0.9	0.77	56	0.06	58.3	0.06	1.0	0.63
Damaging	203	0.17	169.9	0.14	1.2	0.0074	145	0.16	126.7	0.14	1.1	0.059

n: Number of *de novo* mutations. Rate: Number of *de novo* mutations divided by number of individuals in cohort (*N*). HBE: High Brain Expressed genes (top quartile of expression). LBE: Lower Brain Expressed genes (bottom three quartiles of expression). D-Mis: Damaging Missense predicted by Meta-SVM. Damaging: D-Mis + LoF. Bold: Enrichment > 2 or $p < 0.005$

Table S10: Characteristics of 69 Genes with Damaging *de novo* Mutations Overlapping between CHD and P-NDD Cohorts

Gene	LoF	D-Mis	Cardiac Pheno	NDD Pheno	CA Pheno	Rank Heart Expression	Rank Brain Expression	Overlapping Study	Chromatin Modification	Transcription Regulation
ABCG1	0	1	O	U	-	51	39	S		✓
ABL2	0	1	L	+	+	67	62	S		
ACTB	0	1	X	+	+	100	100	A	✓	
ANK3	1	0	C	+	-	95	47	S/E		
ARID1B	1	0	L	U	+	83	73	A/S/D3	✓	✓
ASB17	0	1	C	U	+	24	20	A		
ASH1L	1	0	C	+	-	87	64	S/A	✓	✓
ATIC	0	1	L	+	+	80	92	S		
ATRX	0	1	L	U	-	95	75	D3	✓	✓
CACNA1C	0	1	X	+	-	93	35	S		
CACNA1H	0	1	O	+	U	93	54	S		
CDK13	1	0	L	U	-	80	82	S		✓
CHD4	0	2	O/L	+/+	+/+	99	99	S	✓	✓
CHD7	2	0	C/O	+/U	+/+	93	89	S	✓	✓
CNOT1	1	0	C	U	-	93	92	S		✓
CTNNB1	1	0	O	+	+	99	100	S/D/D3	✓	✓
CUL3	1	0	L	+	+	83	97	S		
DGCR2	0	1	L	+	+	88	96	Z		
DISP1	0	1	H	-	-	61	0	E		
DPYD	0	1	L	+	+	39	28	Z		
DVL3	0	1	L	U	-	87	90	S		✓
EP300	1	0	L	+	+	88	87	A/D3	✓	✓
ETS1	1	0	H	+	+	87	75	E		✓
FBN1	0	2	C/L	+/-	-/+	93	61	A/S		
FOXM1	1	0	L	-	+	81	96	A		✓
FREM2	0	1	L	U	-	81	76	A		
FRMD3	0	1	L	-	-	22	28	S		
FTSJ3	1	0	L	-	+	84	94	S		
INTS6	0	1	C	-	-	61	54	S		
KANSL1	1	0	L	+	-	85	100	D3	✓	

Table S10 (con't)

Gene	LoF	D-Mis	Cardiac Pheno	NDD Pheno	CA Pheno	Rank Heart Expression	Rank Brain Expression	Overlapping Study	Chromatin Modification	Transcription Regulation
KAT6A	1	0	O	+	+	89	89	S	✓	✓
KAT6B	1	0	C	+	+	82	58	D3	✓	✓
KCTD20	0	1	H	+	-	77	92	S		
KDM5B	3	0	L/O/H	U/+/+	-/+/-	86	91	S/A	✓	✓
KDM6B	1	0	O	+	+	95	90	S/A	✓	✓
KIAA0100	1	0	O	+	-	95	97	S		
KMT2C	1	0	C	+	-	80	49	S/A	✓	✓
KMT2D	4	2	L/L/L/C/L/L	+/U/U/U/+/+	-/+/-/+/+/+	97	85	S/D3	✓	✓
LRP1	0	1	H	U	-	93	94	S		
LRP2	0	1	H	-	-	39	58	S/D		
LRP5	0	1	C	-	-	76	85	A		✓
LRRFIP1	1	0	L	-	-	96	56	S		✓
LZTR1	0	1	L/H	+/U	-/+	84	88	S		✓
MED13L	0	1	C	U	-	83	68	S/D3		✓
MINK1	0	1	C	-	+	88	86	S		
MTOR	1	0	L	-	-	80	82	S		✓
MYO5A	0	1	L	+	-	70	97	S/A/E		
NAA15	2	0	C/H	-/-	-/+	97	95	A		✓
NCKAP1	1	0	L	+	-	92	97	S		
NF1	0	1	L/O	U/+	-/+	82	63	S		
NOTCH1	1	1	L/C/L	U/U/+	+/-/-	88	92	S		✓
NSD1	1	0	H	U	+	95	84	A/D3	✓	✓
POGZ	1	1	O/L	+/-	+/-	84	81	S/A/D3		
PTPN11	0	4	O/O/C/O	+/+/+/+	-/+/-/+	94	90	S		
RYR1	0	1	C	-	-	27	27	S		
RYR3	0	1	C	+	-	77	22	S/E		
SF3B1	1	0	C	U	-	95	96	S/A		
SMAD4	0	1	L	U	-	79	94	S/D3		✓
SPRED2	1	0	H	-	+	62	77	A		
SRRM2	1	0	L	+	-	99	99	A		

Table S10 (con't)

Gene	LoF	D-Mis	Cardiac Pheno	NDD Pheno	CA Pheno	Rank Heart Expression	Rank Brain Expression	Overlapping Study	Chromatin Modification	Transcription Regulation
TANC2	1	0	C	+	-	84	44	S		
TLK2	1	0	O	+	-	72	78	S	✓	
TRIM37	0	1	O	+	-	71	80	A	✓	✓
TTN	0	1	C/H/L/C/L	U/-/-/-/U	-/-/-/-/+	100	13	S/A/E		
UBR3	1	0	C/H	+/+	-/-	87	74	S		
UMODL1	0	1	C/H	U/-	-/-	0	13	A		
USP46	1	0	H	+	-	75	80	S		
WHSC1	1	0	O	+	+	89	90	A	✓	✓
ZNF292	1	0	L	U	-	77	65	A/D		✓

The phenotypes (Cardiac, NDD, CA) for each proband with a mutation in the given gene are shown in respective order. Cardiac phenotypes are abbreviated as L (Left Ventricular Outflow Tract Obstruction); C (Conotruncal defects); H (Heterotaxy); O (Other); X (Complex). U: Unknown. Expression data are given as ranked centiles (i.e. ordered from 0 to 100, low to high). The cutoff for high heart or high brain expression is >= 75. Gene membership in Chromatin Modification (GO:0016568) or Transcription Regulation (GO:0006355) GO terms is indicated. Overlapping Studies: A (Autism Sequencing Consortium *(21)*), S (Simons Simplex Collection *(7)*), D (deLigt Intellectual Disability *(19)*), E (Epileptic encephalopathies *(18)*), R (Rauch Intellectual Diability *(20)*), Z (Schizophrenia *(22)*), D3 (Deciperhing Developmental Disorders *(23)*).

Table S11: Gene set *de novo* Enrichment Analysis using 1,161 Genes with Damaging *de novo* Mutations from 7 Published NDD (P-NDD) Studies (Companion Table to Fig. 3) *(7, 18-23)*

	P-NDD Gene Set (1,143 genes)						**P-NDD and HHE Gene Set (564 genes)**					
	Observed		**Expected**		Enrichment	p	**Observed**		**Expected**		Enrichment	p
	n	Rate	n	Rate			n	Rate	n	Rate		
Controls (900)												
Missense	80	0.09	73.5	0.08	1.1	0.24	32	0.04	36	0.04	0.9	0.77
D-Mis	23	0.03	13.3	0.01	1.7	0.01	9	0.01	7	0.01	1.3	0.23
LoF	13	0.01	10.6	0.01	1.2	0.27	8	0.01	5	0.01	1.5	0.18
Damaging	36	0.04	24.0	0.03	1.5	0.013	17	0.02	12	0.01	1.4	0.11
All CHD (1213)												
Missense	109	0.09	98.7	0.08	1.1	0.16	64	0.05	48	0.04	1.3	0.018
D-Mis	42	0.03	17.9	0.01	**2.4**	**$8.5x10^{-07}$**	30	0.02	9	0.01	**3.3**	**$2.8x10^{-08}$**
LoF	43	0.04	14.3	0.01	**3.0**	**$7.4x10^{-10}$**	41	0.03	7	0.01	**5.7**	**$4.7x10^{-18}$**
Damaging	85	0.07	32.2	0.03	**2.6**	**$8.9x10^{-15}$**	71	0.06	16	0.01	**4.4**	**$1.2x10^{-23}$**
CHD without NDD (438)												
Missense	33	0.08	35.6	0.08	0.9	0.69	17	0.04	18	0.04	1.0	0.57
D-Mis	9	0.02	6.5	0.01	1.4	0.2	4	0.01	3	0.01	1.2	0.41
LoF	7	0.02	5.2	0.01	1.4	0.26	6	0.01	3	0.01	**2.3**	0.05
Damaging	16	0.04	11.6	0.03	1.4	0.13	10	0.02	6	0.01	1.7	0.075
Unknown NDD (362)												
Missense	32	0.09	29.5	0.08	1.1	0.34	21	0.06	14	0.04	1.5	0.061
D-Mis	14	0.04	5.3	0.01	**2.6**	**0.0013**	11	0.03	3	0.01	**4.1**	**0.00012**
LoF	9	0.02	4.3	0.01	**2.1**	0.031	9	0.02	2	0.01	**4.2**	**0.00042**
Damaging	23	0.06	9.6	0.03	**2.4**	**0.00017**	20	0.06	5	0.01	**4.1**	**$2.2x10^{-07}$**
CHD with NDD (413)												
Missense	44	0.11	33.6	0.08	1.3	0.049	26	0.06	16	0.04	1.6	0.018
D-Mis	19	0.05	6.1	0.01	**3.1**	**$2.2x10^{-05}$**	15	0.04	3	0.01	**4.9**	**$8.8x10^{-07}$**
LoF	27	0.07	4.9	0.01	**5.5**	**$3.2x10^{-12}$**	26	0.06	2	0.01	**10.5**	**$3.6x10^{-18}$**
Damaging	46	0.11	11.0	0.03	**4.2**	**$2.9x10^{-15}$**	41	0.10	6	0.01	**7.4**	**$3.9x10^{-22}$**

Table S12: Sequencing Metrics for CHD Case and Control cohorts

	Cases, n=3639	Controls, n=2700
Read Length (bp)	74	74
Reads per sample (M)	79.2 ± 25.2	127.5 ± 55.8
Median Coverage at each Targeted base (X)	75.6 ± 20.9	100.2 ± 34.4
Mean Coverage at each Targeted base (X)	84.6 ± 28.1	106 ± 41.1
% of all bases that map to the human genome	98.1 ± 0.7	94.5 ± 6.8
% of all bases that map to targets	57.3 ± 3.6	48.1 ± 14.7
% of targeted bases read at least 20x	85.5 ± 5.1	88.7 ± 5
% PCR duplicates	5.2 ± 1.9	8.5 ± 5.7

Captions for databases S1 to S10

Database S1: Phenotypes for each case proband, including cardiac, neurodevelopmental disorders (NDD) and extra-cardiac congenital anomalies (CA). Cohort labels: PCGC (Pediatric Cardiac Genomics Consortium) or PHN (Pediatric Heart Network). Cardiac Phenotypes: LVO (Left Ventricular Outflow Tract Obstruction), HTX (Heterotaxy), CTD (Conotruncal Defects), Other and Complex, (see Methods for details on cardiac phenotypes). NDD determination is shown (separate for PCGC probands vs. PHN probands; see Materials and Methods for details).

Database S2: List of *de novo* Mutations in CHD case cohort. Chromosome and positions are in hg19 coordinates. Deleterious missense variants are determined by the dbNSFP Meta-SVM Rankscore greater than 0.83357.

Database S3: List of *de novo* Mutations in Control cohort. Chromosome and positions are in hg19 coordinates. Deleterious missense variants are determined by the dbNSFP Meta-SVM Rankscore greater than 0.83357.

Database S4: List of *de novo* probabilities for each variant class in each protein-coding gene on the Nimblegen V2 exome, adjusted for depth in Cases. Probabilities are transformed by $\log_{10} prob$.

Database S5: List of *de novo* probabilities for each variant class in each protein-coding gene on the Nimblegen V2 exome, adjusted for depth in Controls. Probabilities are transformed by $\log_{10} prob$.

Database S6: Functional term enrichment analysis of all Genes with Damaging (LoF + D-Mis) *de novo* mutations in all cases. Analysis was performed using g:profiler (see Methods). P-values are Bonferroni corrected.

Database S7: Functional term enrichment analysis of all Genes with Loss of Function (LoF) *de novo* mutations in 860 new cases. Analysis was performed using g:profiler (see Methods). P-values are Bonferroni corrected.

Database S8: List of 1,563 variants (1,161 unique genes) with damaging *de novo* mutations from 7 independent NDD cohorts *(7, 18-23)*. Cohorts: SSC (Simons Simplex Collection, n=2517 probands *(7)*), epi (Epileptic encephalopathies, n=356 probands *(18)*), deLigt (Intellectual Disability, n=100 probands *(19)*), Rauch (Intellectual Diability, n=51 probands *(20)*), ASC (Autism Sequencing Consortium, n=1445 probands *(21)*), SCHIZOP (Schizophrenia, n=53 probands *(22)*), DDD (Deciperhing Developmental Disorders, n=191 of 1,133 probands with putative pathogenic variants and no CHD *(23)*). Only variants of class LoF or D-Mis (by dbNSFP Meta-SVM Rankscore greater than 0.83357) were considered. Chromosome and positions are in hg19 coordinates. AA1: Amino acid in reference sequence; AA2: Amino acid in alternate sequence.

Database S9: Functional term enrichment analysis among 69 genes with Damaging (LoF + D-Mis) *de novo* mutations overlapping between CHD cases and the published NDD (P-NDD) cohort. Genes with damaging *de novo* mutations in cases (n=331 genes) were overlapped between those from the external NDD cohort (n=1,161 genes, Table S8) to identify 69 genes. Analysis was performed using g:profiler (see Methods). P-values are Bonferroni corrected.

Database S10: Percentile ranks of genes by expression (i.e. ordered from 0 to 100, low to high) in the developing mouse heart (embryonic day 14.5) and brain (embryonic day 9.5). See Materials and Methods for details.

Additional Acknowledgments

We gratefully acknowledge the participation of *Pediatric Heart Network* investigators:

National Heart, Lung, and Blood Institute: Gail Pearson, Victoria Pemberton, Rae-Ellen Kavey*, Mario Stylianou, Marsha Mathis

Network Chair: *University of Texas Southwestern Medical Center*, Lynn Mahony

Data Coordinating Center: *New England Research Institutes*, Lynn Sleeper (PI)*, Sharon Tennstedt (PI)*, Steven Colan, Lisa Virzi*, Patty Connell*, Victoria Muratov*, Lisa Wruck*, Minmin Lu*, Dianne Gallagher*, Anne Devine*, Thomas Travison*, David F. Teitel

Core Clinical Site Investigators: *Children's Hospital Boston*, Jane W. Newburger (PI), Peter Laussen*, Pedro del Nido, Roger Breitbart, Jami Levine, Ellen McGrath, Carolyn Dunbar-Masterson; *Children's Hospital of New York,* Wyman Lai (PI), Beth Printz (currently at Rady Children's Hospital), Daphne Hsu (currently at Montefiore Medical Center), William Hellenbrand (currently at Yale New Haven Medical Center), Ismee Williams, Ashwin Prakash (currently at Children's Hospital Boston), Ralph Mosca (currently at New York University Medical Center), Darlene Servedio, Rozelle Corda, Rosalind Korsin, Mary Nash*; *Children's Hospital of Philadelphia*, Robert Shaddy (PI), Victoria L. Vetter, Sarah Tabbutt (currently at the University of California, San Francisco), J. William Gaynor (Study Co-Chair), Chitra Ravishankar, Thomas Spray, Meryl Cohen, Marisa Nolan, Stephanie Piacentino*, Sandra DiLullo*, Nicole Mirarchi; *Cincinnati Children's Medical Center*, Jame Cnota, D. Woodrow Benson (PI; currently at Children's Hospital of Wisconsin), Catherine Dent Krawczeski (currently at Stanford University), Lois Bogenschutz, Teresa Barnard, Michelle Hamstra, Rachel Griffiths*, Kathryn Hogan*, Steven Schwartz (currently at the Hospital for Sick Children, Toronto), David Nelson; *North Carolina Consortium: Duke University, East Carolina University, Wake Forest University*, Page A. W. Anderson (PI; deceased), Jennifer Li (PI), Wesley Covitz, Kari Crawford*, Michael Hines, James Jaggers, Theodore Koutlas, Charlie Sang, Jr, Lori Jo Sutton, Mingfen Xu; *Medical University of South Carolina*, J. Philip Saul (PI; currently at Nationwide Children's Hospital, Columbus, OH), Andrew Atz, Girish Shirali (PI; currently at Children's Mercy Hospital, Kansas City, MO), Scott Bradley, Eric Graham, Teresa Atz, Patricia Infinger; *Primary Children's Medical Center and the University of Utah, Salt Lake City, Utah*, L. LuAnn Minich (PI), John Hawkins (deceased), Michael Puchalski, Richard Williams, Linda Lambert, Jun Porter*, Marian Shearrow; *Hospital for Sick Children, Toronto*, Brian McCrindle (PI), Seema Mital, Steven Schwartz, Joel Kirsh, Chris Caldarone, Elizabeth Radojewski, Svetlana Khaikin, Susan McIntyre, Nancy Slater; *University of Michigan*, Caren S. Goldberg (PI), Richard G. Ohye (Study Chair), Cheryl Nowak*; *Children's Hospital of Wisconsin*, Nancy Ghanayem (PI), James Tweddell, Kathleen Mussatto, Michele Frommelt, Lisa Young-Borkowski

Auxiliary Sites: *Children's Hospital Los Angeles*, Alan Lewis (PI), Vaughn Starnes, Nancy Pike; *The Congenital Heart Institute of Florida* (CHIF), Jeffrey P. Jacobs, MD (PI), James A. Quintessenza, Paul J. Chai, David S. Cooper*, J. Blaine John, James C. Huhta, Tina Merola, Tracey Cox; *Emory University*, Kirk Kanter, William Mahle (PI), Joel Bond, Leslie French*, Jeryl Huckaby; *Nemours Cardiac Center*, Christian Pizarro, Carol Prospero; Julie Simons, Gina Baffa; *University of Texas Southwestern Medical Center*, Ilana Zeltser (PI), Tia Tortoriello, Deborah McElroy*, Deborah Town
*no longer at the institution listed

This work was supported by grants from the **National Heart, Lung, and Blood Institute** including the **Pediatric Cardiac Genomics Consortium** (U01-HL098188, U01-HL098147, U01-HL098153, U01-HL098163, U01-HL098123, U01-HL098162), the **Pediatric Heart Network** (U01-HL068269, U01-HL068270, U01-HL068279, U01-HL068281, U01-HL068285, U01-HL068292, U01-HL068290, U01-HL068288, U01-HL085057, U01-HL109781, U01 HL109816, U01-HL109737) and **the Cardiovascular Development Consortium** (U01-HL098166), and from the **National Human Genome Research Institute** (U54HG006504), and the **National Research Science Foundation** (Award 1F30HL123238).