

*Supplementary Materials*

***De novo* mutations in histone modifying genes in congenital heart disease**

<b>Patient cohorts</b>	4
<b>Exome sequencing</b>	4
<b>Identification and validation of <i>de novo</i> mutations</b>	4-5
<b>RNA sequencing dataset and analysis</b>	5
<b>Principal component analysis</b>	6
<b>Statistical analyses</b>	6-7
<b>Estimating the number of genes in which <i>de novo</i> mutations contribute to CHD</b>	7
<b>Table S1: Description of the cohorts subjected to sequencing</b>	8
<b>Table S2: Sequencing QA/QC for CHD and control cohorts</b>	9
<b>Table S3: Relationship of <i>de novo</i> Bayesian quality score vs. confirmation by Sanger sequencing</b>	10
<b>Table S4: All <i>de novo</i> mutations with Bayesian quality score <math>\geq 50</math></b>	11-16
<b>Table S5: <i>De novo</i> mutations in CHD are enriched in genes that are more highly expressed in the developing heart at e14.5</b>	17
<b>Table S6. Increased frequency of <i>de novo</i> mutations (damaging and missense at conserved positions) in CHD cases and controls stratified for gene expression in developing heart at e9.5</b>	18
<b>Table S7: Comparison of <i>de novo</i> mutation frequency using RNA expression data at e14.5</b>	19
<b>Table S8. <i>De novo</i> mutations in CHD probands and controls stratified for gene expression in developing heart; categorical analysis of the presence or absence of any <i>de novo</i> mutation in probands</b>	20

---

<b>Table S9: Odds ratios in different CHD classes</b>	21
<b>Table S10: Chromatin modifying and other genes of interest with <i>de novo</i> mutations in CHD probands (expanded diagnosis)</b>	22
<b>Table S11: Genes with &gt; 1 <i>de novo</i> mutation in CHD probands</b>	23
<b>Table S12: Candidate genes for LVO, CTD, HTX</b>	24
<b>Table S13: CHD <i>De novo</i> variants that overlap with candidate genes</b>	25
<b>Figure S1: Distribution of 8X read coverage between cases and controls</b>	26
<b>Figure S2: Correlation of Bayesian quality score and probability of Sanger validation</b>	27
<b>Figure S3: Distribution of <i>de novo</i> mutation quality score</b>	28
<b>Figure S4: Correlation of paternal and maternal age with <i>de novo</i> mutation rate</b>	29
<b>Figure S5: Effect of ancestry of <i>de novo</i> mutation rate</b>	30
<b>Figure S6: <i>De novo</i> mutation rate closely approximates Poisson distribution in CHD cases and controls</b>	31
<b>Figure S7: <i>De novo</i> mutations in CHD probands and controls stratified for gene expression in developing heart; categorical analysis of the presence or absence of any <i>de novo</i> mutation in probands</b>	32
<b>Figure S8: Expression of genes mutated in cases and controls</b>	33
<b>Figure S9: <i>De novo</i> mutations in CHD cases and controls stratified for gene expression in adult tissues</b>	34

---

**Patient cohorts.** Probands with or without parents were recruited from 9 centers in the United States and the United Kingdom into the Congenital Heart Disease Genetic Network Study of the Pediatric Cardiac Genomics Consortium (CHD Genes: NCT01196182)<sup>7</sup>. 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 and Mt. 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 each participating subject or their parent/guardian. Probands were selected for severe congenital heart disease (excluding isolated VSDs, ASDs, PDAs or PSs), availability of both parents, and absence of any CHD in first-degree relatives. Cardiac diagnoses were obtained from review of echocardiogram, catheterization and operative reports; extracardiac findings were extracted from medical records. Controls were from 264 previously studied quartets that included one offspring with autism, an unaffected sibling and unaffected parents, all recruited with written informed consent by the Simons Foundation Autism Research Initiative<sup>28</sup>. Parents and their unaffected sibling from this cohort were analyzed in the current study.

**Exome sequencing.** Trios were sequenced at the Yale Center for Genome Analysis following the same protocol. Genomic DNA from venous blood was captured with the NimbleGen v2.0 exome capture reagent (Roche) and sequenced (Illumina HiSeq 2000, 75 base paired-end reads). Reads were mapped to the reference genome using Eland. SNV and indel calls were assigned quality scores (QS) using SAMtools<sup>8</sup> and annotated for novelty using dbSNP, build 135, 1000 genomes, May 2011 release and the Yale Exome Database, for impact on encoded proteins, and conservation of variant position.

**Identification and confirmation of *de novo* mutations.** Heterozygous SNVs and indels in the proband that showed QS  $\geq$  60 and 600, respectively, and rare

non-reference calls in both parents were selected. Read plots of all putative indels were visually inspected in trio members to eliminate false calls. A Bayesian algorithm was used to assist *de novo* mutation calls. Elements included probability of the proband being heterozygous at the test position; probability that parents are homozygous for the reference allele, given frequency of reference and non-reference reads and probability of heterozygosity in offspring; probability that a variant is *de novo* given its population frequency. Resulting QSs scaled from 0 to 100. Their correlation with bona fide *de novo* mutations was determined by Sanger sequencing of PCR amplicons harboring 181 putative mutations distributed across the QS spectrum. Additionally, all six *de novo* indels with QS > 50 in the HHE gene set were tested and confirmed by Sanger sequencing.

**RNA sequencing and analysis.** Hearts from e14.5 mouse embryos (strain 129SvEv) were isolated, rinsed, and immersed in RNALater. Left and right atria, left ventricle (with interventricular septum, aortic and mitral valves), and right ventricle (with pulmonary and tricuspid valves) were dissected. Chamber-specific RNAs were extracted and pooled from 5 embryos, selected with oligo-dT, copied into double stranded DNA, and ligated to adaptors. 150-250 bp fragments were isolated after acrylamide gel electrophoresis, amplified and sequenced (Illumina HiSeq2000), with > 40 million paired-end 50 base reads per library as previously described<sup>29</sup>. Reads were aligned to the mouse genome (mm9)<sup>30</sup>, and reads per gene per million mapped reads (rpm) was determined. The average of rpm of each gene from each chamber was used as the measure of heart expression. RNA from atria, ventricle and truncus/outflow tract at e9.5 was prepared, sequenced and analyzed by an analogous approach. RNA sequencing of control human adult tissues- lung, liver, heart and brain- from the Illumina Human Body Map (<http://www.ebi.ac.uk/arrayexpress/browse.html?keywords=E-MTAB-513>) was similarly performed and analyzed as reads per gene per million reads per kb of transcript.

**Principal component analysis.** The EIGENSTRAT program was used to compare SNP genotypes of probands and individuals of known ancestry in HapMap3 (<http://hapmap.ncbi.nlm.nih.gov/>). SNPs with MAF >5% without significant linkage disequilibrium with other SNPs were analyzed. The results of analysis correctly distinguished ancestry groups in HapMap3 samples; ancestries of CHD subjects were assigned accordingly.

**Statistical analyses.** The significance of mutation frequency differences between groups was tested with two-tailed binomial exact tests; two-tailed Fisher exact tests assessed differences in numbers of patients with one or more *de novo* mutations; tests among 3 groups was by Chi-square analysis. Gene expression at e14.5 of genes mutated in cases and controls was compared by Wilcoxon signed-rank test. Correlation of mutation rate and parental age was tested by Pearson's correlation. The expected number of genes with more than one *de novo* mutation was determined by Monte Carlo simulation ( $10^8$  iterations) specifying the total number of protein-altering mutations and 21,000 genes of observed coding length. Analogous approaches were used to determine probabilities of any gene having  $\geq 2$  damaging mutations,  $\geq 1$  damaging and  $\geq 1$  mutation at a conserved position, and  $\geq 13$  genes mutated in both CHD and autism. The fit to the Poisson distribution of the observed numbers of *de novo* mutations per subject was assessed by Chi-square test.

Overrepresentation of *de novo* mutations in the H3K4me pathway and the presence of significant enrichment of other gene pathways was tested via Gene Ontology (GO) analysis, using a modified Fisher's exact test with Bonferroni correction as implemented in DAVID (<http://david.abcc.ncifcrf.gov/>). Input was all genes with protein-altering *de novo* mutations in CHD or control subjects, and all genes sequenced. The H3K4me gene set was: *CHD8, MLL3, SETD7, WHSC1L1, CDC73, WHSC1, SETD1A, MLL2, KDM5A, MLL4, MLL5, UBE2B, ASH1L, SETD1B, MLL, LEO1, PAF1, KDM5C, CTR9, PRDM9, MEN1, CHD7, RNF20, KDM1A, RNF40, SMYD3, KDM6A, KDM5B, USP44, WDR5*. The

expected number of mutations in the H3K4me set was calculated from the fraction of the exome coding region attributable to this gene set and the total number of *de novo* mutations.

**Estimating number of genes in which *de novo* mutations contribute to CHD.** We addressed this question using the ‘unseen species problem’<sup>9</sup>. We infer that the number of probands with non-synonymous mutations in the HHE set (81) minus the expected number (44; calculated from the number observed in controls), represents the number of subjects in whom *de novo* mutations confer CHD risk (37; 10.0% of probands). The number of genes with > 1 protein-altering *de novo* mutation (six) minus the most likely number expected by chance (three) represents risk-associated genes with more than 1 mutation (three). The number of risk-associated genes (C) is estimated as follows:

$$C = c/u + g^2 \times d \times (1-u)/u$$

c = number of observed risk-associated genes (34)

c<sub>1</sub> = number of genes mutated once (31)

d = total number of risk-associated mutations (37)

g = variation in effect size of individual *de novo* mutations (assumed to be 1, which minimizes underestimation of set size)

u = 1 – c<sub>1</sub>/d (probability that newly added mutation hits a previously mutated gene)

$$C = 401$$

From 95% confidence intervals of the number of risk-associated events, the 95% confidence interval for number of risk genes is calculated as 197-837.

**Table S1. Description of cohorts subjected to sequencing**

Demographics	Cases (362) <sup>†</sup>	Controls (264)
Male: Female	220:142	163:161
Age at enrollment ( <i>mean ± SD</i> )	7.8 ± 9.6	13.7 ± 4.6
Maternal age at birth of proband ( <i>mean ± SD</i> )	31.2 ± 5.5	29.9±4.8
Paternal age at birth of proband ( <i>mean ± SD</i> )	31.9 ± 6.1	32.5±6
African-American	32 (8.8%)	9 (3.8%)
European	305 (84.3%)	187 (79.2%)
Asian	31 (8.6%)	13 (5.5%)
Native American	4 (1.1%)	0 (0%)
Hispanic	50 (13.8%)	27 (11.5%)
<b>Cardiac Lesions in Probands<sup>§</sup></b>		
Conotruncal Defects (CTD): 153 probands		
TOF	63 (17%)	
TOF/PA	9 (2.5%)	
TOF/PA-MAPCA	12 (3%)	
D-TGA	47 (13%)	
DORV	18 (5%)	
TA	7 (2%)	
AoAA	3 (1%)	
Left Ventricular Obstruction (LVO): 132 probands		
HLHS	60 (17%)	
CoA	36 (10%)	
BAV	40 (11%)	
AS (not BAV)	13 (4%)	
MS	1 (0.3%)	
Heterotaxy (HTX): 70 probands		
RAI	8 (2%)	
LAI	16 (4%)	
DEX	19 (5%)	
CAVC	26 (7%)	
DORV	17 (5%)	
L-Loop	12 (3%)	
L-TGA	18 (5%)	
Other Cardiac Diagnosis: 6 probands		
TAPVR	1 (0.3%)	
CAVC	1 (0.3%)	
DILV	3 (1%)	
other	1 (0.3%)	
Extracardiac abnormalities	81 probands (22%)	

<sup>†</sup>Self-reported ethnicity: 15 probands self-identified as multi-ethnic, hence ethnicities total >100%

<sup>§</sup>Numerous probands have more than one cardiac feature; for this reason, sum of sub-phenotypic classification does not equal 100%.

Abbreviations: TOF-tetralogy of Fallot; TOF/PA-tetralogy of Fallot with pulmonary atresia; TOF/PA-MAPCA-tetralogy of Fallot with pulmonary atresia and multiple aortico-pulmonary collaterals; D-TGA-D-transposition of the great arteries; DORV-double outlet right ventricle; AoAA-aortic arch anomaly; TA-truncus arteriosus; HLHS-hypoplastic left heart syndrome; CoA-coarctation of the aorta; BAV-bicuspid aortic valve; AS-aortic stenosis; MS-mitral stenosis; RAI-right atrial isomerism; LAI-left atrial isomerism; Dex-dextrocardia; CAVC-complete atrioventricular canal; L-Loop-L-looped ventricles; L-TGA-L-Transposition of the great arteries; TAPVR-total anomalous pulmonary venous return; DILV-double-inlet left ventricle.



**Table S2: Sequencing QA/QC for CHD and control cohorts**

Category	Cases (1086 samples)	Controls (792 samples)
Read length (bp)	74	74
# of reads per sample (M)	102.6 ± 28	120 ± 45.5
Median coverage at each targeted base (X)	91.2 ± 23.9	99.3 ± 36.6
Mean coverage at each targeted base (X)	107 ± 27.7	117.1 ± 42
% of all bases that map to human genome	91.3% ± 0.9	91.3% ± 2
% of all bases that map to target	69.1% ± 4.8	67.4% ± 9.3
% of targeted bases read at least 8x	96.0% ± 1	95.6% ± 1.6
% of targeted bases read at least 20x	91.4% ± 3.4	90.5% ± 6.8
Mean error rate	0.5% ± 0.1	0.6% ± 0.2
% PCR duplicates	5.53% ± 2.39	5.39% ± 2.79

**Table S3. Relationship of *de novo* Bayesian quality score vs. confirmation by Sanger sequencing**

Bayesian Quality Score	# of putative <i>de novo</i> mutations	# of attempted confirmations	# validated by Sanger sequencing	% of all <i>de novo</i> mutations
≥50	324	88	88 (100%)	88.5%
40-50	9	9	7 (77%)	1.9%
30-40	11	11	8 (72%)	2.2%
20-30	34	31	13 (42 )	3.9%
10-20	64	15	3 (20%)	3.5%
5-10	77	5	0 (0%)	0%















**Table S5. *De novo* mutations (damaging and missense at conserved positions) in CHD are enriched in genes that are more highly expressed in the developing heart at e14.5**

% of all genes	$\geq n$ reads per million @ e14.5	# genes	Total # <i>de novo</i> mutations		<i>De novo</i> mutations/subject		Odds Ratios Cases: Cont (95% CI) <sup>†</sup>	P-value <sup>††</sup>
			CHD	Controls	CHD	Controls		
			362 trios	264 trios	362 trios	264 trios		
100%	0	16,676	101	63	0.279	0.239	1.48 (0.9-2.4)	0.34
50%	13	8,338	73	36	0.202	0.136	1.97 (1.1-3.6)	0.06
45%	17	7,504	71	30	0.196	0.114	2.16 (1.1-4.1)	0.01
40%	21	6,670	67	28	0.185	0.106	2.03 (1.0-4.0)	0.01
35%	26	5,837	63	24	0.174	0.091	2.35 (1.2-4.8)	0.006
30%	32	5,003	61	18	0.169	0.068	3.54 (1.6-7.7)	0.0004
25%	40	4,169	54	15	0.149	0.057	3.60 (1.6-8.3)	0.0005
20%	50	3,335	48	11	0.133	0.042	5.13 (2.0-12.9)	0.0002
15%	66	2,501	37	10	0.102	0.038	4.84 (1.8-13.2)	0.004
10%	93	1,668	24	6	0.066	0.023	4.73 (1.4-15.7)	0.02
5%	161	834	12	2	0.033	0.008	5.00 (0.7-33.8)	0.05

<sup>†</sup>The odds ratio is the ratio of protein-altering to silent variants in cases divided by the corresponding ratio in controls

<sup>††</sup>P-values compare the number of variants in each category between cases and controls using a two-tailed binomial exact test

Genes were ranked for level of expression in heart at e14.5, and partitioned at successive percentiles (e.g. '25%' denotes the genes in the top quartile of expression). Genes in resulting groups were analyzed for burden of damaging or missense mutations at conserved positions in CHD cases and controls.

**Table S6. Increased frequency of *de novo* mutations (damaging and missense at conserved positions) in CHD cases and controls stratified for gene expression in developing heart at e9.5**

<b>Cases vs. controls, High heart expressed genes at e9.5 (top 25%)</b>	Total # of <i>de novo</i> mutations		<i>De novo</i> mutations/subject		Odds Ratio Cases:Cont (95% CI) <sup>†</sup>	P-value <sup>††</sup>
	CHD 362 trios	Controls 264 trios	CHD 362 trios	Controls 264 trios		
Silent	18	23	0.050	0.087	1.00 (0.42 - 2.39)	0.08
Nonconserved Missense	29	21	0.080	0.080	1.76 (0.77 - 4.06)	1.00
Silent and Protein Changing	93	58	0.257	0.220	2.05 (1.02 - 4.12)	0.37
All Protein Changing	75	35	0.207	0.133	2.74 (1.31 - 5.71)	0.03
Conserved Missense	32	12	0.088	0.045	3.41 (1.38 - 8.43)	0.05
Conserved and Damaging Protein Altering	46	14	0.127	0.053	4.20 (1.78 - 9.91)	0.004
Damaging	14	2	0.039	0.008	8.94 (1.80 - 44.52)	0.02

<sup>†</sup>The odds ratio is the ratio of protein-altering to silent variants in cases divided by the corresponding ratio in controls

<sup>††</sup>P-values compare the number of variants in each category between cases and controls using a two-tailed binomial exact test

**Table S7. Comparisons of *de novo* mutation frequencies using RNA expression data at e14.5**

<b>a. Cases vs. controls, Low heart expressed genes (bottom 75%)</b>	Total # of <i>de novo</i> mutations		<i>De novo</i> mutations/subject		Odds Ratio Cases:Cont (95% CI) <sup>†</sup>	P-value <sup>††</sup>
	CHD 362 trios	Controls 264 trios	CHD 362 trios	Controls 264 trios		
Silent	44	39	0.12	0.15	1.00 (0.54-1.84)	0.38
Nonconserved Missense	103	70	0.28	0.27	1.30 (0.77-2.21)	0.70
Silent and Protein Changing	194	157	0.54	0.59	1.10 (0.68-1.77)	0.33
All Protein Changing	150	118	0.41	0.45	1.13 (0.69-1.85)	0.54
Conserved Missense	37	33	0.10	0.13	0.99 (0.53-1.88)	0.40
Conserved and Damaging Protein Altering	47	48	0.13	0.18	0.87 (0.48-1.56)	0.12
Damaging	10	15	0.03	0.06	0.59 (0.24-1.47)	0.10
<b>b. High vs. Low heart- expressed genes in CHD cases</b>	Total # of <i>de novo</i> mutations		<i>De novo</i> mutations/subject		Odds Ratio High:Low (95% CI) <sup>†</sup>	P-value <sup>††</sup>
	High heart 4,169 genes	Low heart 12,507 genes	High heart 4,169 genes	Low heart 12,507 genes		
Silent	21	44	0.06	0.12	1.00 (0.48-2.09)	0.59
Nonconserved Missense	27	103	0.07	0.28	0.55 (0.28-1.07)	0.04
Silent and Protein Changing	102	194	0.28	0.54	1.10 (0.62-1.95)	0.05
All Protein Changing	81	150	0.22	0.41	1.13 (0.63-2.03)	0.05
Conserved Missense	39	37	0.11	0.1	2.21 (1.11-4.39)	6.40E-05
Conserved and Damaging Protein Altering	55	47	0.15	0.13	2.45 (1.28-4.69)	1.63E-07
Damaging	15	10	0.04	0.03	3.14 (1.21-8.16)	0.001
<b>c. High vs. Low heart- expressed genes in Controls</b>	Total # of <i>de novo</i> mutations		<i>De novo</i> mutations/subject		Odds Ratio High:Low (95% CI) <sup>†</sup>	P-value <sup>††</sup>
	High heart 4,169 genes	Low heart 12,507 genes	High heart 4,169 genes	Low heart 12,507 genes		
Silent	21	39	0.08	0.15	1.00 (0.47-2.12)	0.32
Nonconserved Missense	17	70	0.06	0.27	0.45 (0.21-0.95)	0.06
Silent and Protein Changing	53	157	0.20	0.59	0.63 (0.34-1.16)	0.25
All Protein Changing	32	118	0.12	0.45	0.50 (0.26-0.97)	0.04
Conserved Missense	13	33	0.05	0.13	0.73 (0.32-1.68)	1.00
Conserved and Damaging Protein Altering	15	48	0.06	0.18	0.58 (0.26-1.27)	0.41
Damaging	2	15	0.01	0.06	0.25 (0.05-1.19)	0.18

<sup>†</sup> The odds ratio is the ratio of protein-altering to silent variants in high-heart genes divided by the corresponding ratio in low-heart genes for cases (a) or controls (b)

<sup>††</sup> P-values compare the number of bases in each category between high-heart expressed genes and low-heart expressed genes using a two-tailed binomial exact test

**Table S8. *De novo* mutations in CHD probands and controls stratified for gene expression in developing heart at e14.5; categorical analysis of the presence or absence of any *de novo* mutation in probands**

<b>a. Cases vs. controls, High heart expressed genes (top 25%)</b>	Total # of subjects with $\geq 1$ <i>de novo</i> mutations		Fraction of subjects with $\geq 1$ <i>de novo</i> mutation		Odds Ratio Cases:Cont (95% CI)	P-value
	CHD	Controls	CHD	Contr Is		
	362 trios	264 trios	362 trios	264 trios		
Silent	20	21	0.052	0.080	0.68 (0.34-1.35)	0.25
Nonconserved Missense	27	17	0.075	0.064	1.17 (0.6-2.34)	0.75
Silent and Protein Changing	86	49	0.238	0.186	1.37 (0.91-2.07)	0.14
All Protein Changing	77	31	0.213	0.117	2.03 (1.27-3.3)	0.002
Conserved Missense	38	13	0.105	0.049	2.26 (1.15-4.73)	0.01
Conserved and Damaging Protein Altering	51	15	0.141	0.057	2.72 (1.46-5.33)	0.0006
Damaging	14	2	0.039	0.008	5.26 (1.19-48.08)	0.02
<b>b. Cases vs. controls, Low heart expressed genes (bottom 75%)</b>	Total # of subjects with $\geq 1$ <i>de novo</i> mutations		Fraction of subjects with $\geq 1$ <i>de novo</i> mutation		Odds Ratio Cases:Cont (95% CI)	P-value
	CHD	Controls	CHD	Controls		
	362 trios	264 trios	362 trios	264 trios		
Silent	43	34	0.119	0.129	0.91 (0.55-1.52)	0.71
Nonconserved Missense	85	61	0.235	0.231	1.02 (0.69-1.52)	0.92
Silent and Protein Changing	150	115	0.414	0.436	0.92 (0.66-1.28)	0.62
All Protein Changing	122	94	0.337	0.356	0.92 (0.65-1.3)	0.67
Conserved Missense	35	32	0.097	0.121	0.78 (0.45-1.34)	0.36
Conserved and Damaging Protein Altering	44	46	0.122	0.174	0.66 (0.41-1.05)	0.07
Damaging	10	15	0.028	0.057	0.47 (0.19-1.14)	0.10
<b>c. High vs. Low heart-expressed genes, CHD cases</b>	Total # of subjects with $\geq 1$ <i>de novo</i> mutations		Fraction of subjects with $\geq 1$ <i>de novo</i> mutation		Odds Ratio High:Low (95% CI)	P-value
	High heart	Low heart	High heart	Low heart		
	4,169 genes	12,507 genes	4,169 genes	12,507 genes		
Silent	20	43	0.052	0.163	1.13 (0.63-1.97)	0.68
Nonconserved Missense	27	85	0.075	0.322	0.77 (0.48-1.21)	0.30
Silent and Protein Changing	86	150	0.238	0.568	1.40 (1.06-1.83)	0.02
All Protein Changing	77	122	0.213	0.462	1.54 (1.14-2.06)	0.004
Conserved Missense	38	35	0.105	0.133	2.65 (1.63-4.31)	4.51 $\times 10^{-5}$
Conserved and Damaging Protein Altering	51	44	0.141	0.167	2.83 (1.85-4.33)	6.33 $\times 10^{-7}$
Damaging	14	10	0.039	0.038	3.41 (1.41-8.59)	0.003
<b>d. High vs. Low heart-expressed genes, Controls</b>	Total # of subjects with $\geq 1$ <i>de novo</i> mutations		Fraction of subjects with $\geq 1$ <i>de novo</i> mutation		Odds Ratio High:Low (95% CI)	P-value
	High heart	Low heart	High heart	Low heart		
	4,169 genes	12,507 genes	4,169 genes	12,507 genes		
Silent	21	34	0.058	0.129	1.51 (0.83-2.67)	0.14
Nonconserved Missense	17	61	0.047	0.231	0.68 (0.37-1.18)	0.17
Silent and Protein Changing	49	115	0.135	0.436	1.04 (0.73-1.46)	0.86
All Protein Changing	31	94	0.086	0.356	0.80 (0.52-1.22)	0.33
Conserved Missense	13	32	0.036	0.121	0.99 (0.48-1.94)	1.00
Conserved and Damaging Protein Altering	15	46	0.041	0.174	0.79 (0.41-1.45)	0.48
Damaging	2	15	0.006	0.057	0.33 (0.04-1.4)	0.18

The odds ratio is calculated from the ratio of cases with and without *de novo* mutations in each category in cases divided by the corresponding ratio in controls. The *P*-values compares the number of subjects with and without variants in a specific category between cases and controls using a two-tailed Fisher exact test

**Table S9. Odds ratios in different disease classes**

<b>a. Cases vs. controls, High heart expressed genes (top 25%)</b>	Total # of <i>de novo</i> mutations				<i>De novo</i> mutations/subject				Odds Ratio (95% CI) <sup>†</sup>		
	CTD	LVO	HTX	Controls	CTD	LVO	HTX	Controls	CTD	LVO	HTX
	154 trios	132 trios	70 trios	264 trios	154 trios	132 trios	70 trios	264 trios			
Silent	9	7	3	21	0.06	0.05	0.04	0.08	N/A	N/A	N/A
Nonconserved Missense	14	10	2	17	0.08	0.08	0.03	0.06	1.92 (0.67 - 5.51)	1.76 (0.55 - 5.6)	0.82 (0.12 - 5.5)
Silent and All Protein Changing	42	46	11	53	0.27	0.35	0.16	0.20	N/A	N/A	N/A
All Protein Changing	33	39	8	32	0.21	0.30	0.11	0.12	2.41 (0.96 - 6.04)	3.66 (1.38 - 9.7)	1.75 (0.4 - 7.36)
Conserved Missense	13	22	4	13	0.08	0.17	0.06	0.05	2.33 (0.78 - 6.98)	5.08 (1.7 - 15.2)	2.15 (0.4 - 11.2)
Conserved and Damaging Protein Altering	20	29	6	15	0.13	0.22	0.09	0.06	3.11 (1.11 - 8.7)	5.80 (2 - 16.7)	2.80 (0.6 - 13)
Damaging	6	7	2	2	0.05	0.05	0.03	0.01	7.00 (1.2 - 41.54)	10.50 (1.8 - 62.8)	7.00 (0.7 - 70.1)

<b>b. Cases vs. controls, Low heart expressed genes (bottom 75%)</b>	Total # of <i>de novo</i> mutations				<i>De novo</i> mutations/subject				Odds Ratio (95% CI) <sup>†</sup>		
	CTD	LVO	HTX	Controls	CTD	LVO	HTX	Controls	CTD	LVO	HTX
	154 trios	132 trios	70 trios	264 trios	154 trios	132 trios	70 trios	264 trios			
Silent	15	20	9	39	0.10	0.15	0.13	0.15	N/A	N/A	N/A
Nonconserved Missense	42	33	25	70	0.27	0.25	0.36	0.27	1.56 (0.8 - 3.2)	0.92 (0.47 - 1.8)	1.55 (0.66 - 3.65)
Silent and All Protein Changing	80	67	44	157	0.52	0.51	0.63	0.59	N/A	N/A	N/A
All Protein Changing	65	47	35	118	0.42	0.36	0.50	0.45	1.43 (0.7 - 2.79)	0.78 (0.4 - 1.47)	1.29 (0.57 - 2.91)
Conserved Missense	20	11	6	33	0.13	0.08	0.09	0.13	1.58 (0.7 - 3.56)	0.65 (0.27 - 1.6)	0.79 (0.25 - 2.44)
Conserved and Damaging Protein Altering	23	14	10	48	0.15	0.11	0.14	0.18	1.25 (0.57 - 2.7)	0.57 (0.3 - 1.27)	0.90 (0.33 - 2.44)
Damaging	3	3	4	15	0.02	0.02	0.06	0.06	0.52 (0.1 - 2.06)	0.39 (0.1 - 1.51)	1.16 (0.31 - 4.32)

<sup>†</sup>The odds ratio is the ratio of protein-altering to silent variants in each disease class divided by the corresponding ratio in controls  
Abbreviations: CTD: Conotruncal defects, LVO: Left ventricular obstruction, HTX: Heterotaxy

**Table S10. Chromatin modifying and other genes of interest with *de novo* mutations in CHD probands**

ID	Gene	Heart Exp <sup>†</sup>	Mutation	Primary Classification: Specific Cardiovascular Diagnoses <sup>§</sup>	Extracardiac Structural Anomalies	Neuro-Developmental	Somatic Growth	
							Ht(%)	Wt(%)
1-00596	<i>MLL2</i>	216	p.Ser1722 Argfs*9	LVO: Mitral atresia, HLHS, aortic atresia, dbl AA	Epicanthal folds, telecanthus, large low-set ears, excess nuchal skin, high arched palate, wide-spaced nipples, undescended testes, club foot, hyperpigmented lesions,	motor delay, hypotonia	50	<5
1-00853	<i>WDR5</i>	39	p.Lys7Gln	CTD: TOF, right aortic arch, aberrant LSA, coronary abnormality	No	abnormal	90	90
1-00534	<i>CHD7</i>	125	p.Gln1599*	CTD: TOF-PA	Cleft lip, cleft palate, inguinal hernia, micropenis, sensorineural hearing loss	abnormal	<5	<5
1-00230	<i>KDM5A</i>	70	p.Arg1508Trp	LVO: LSVC, Primum ASD, cleft MV, sub-AS, BAV	No	normal	<5	<5
1-01965	<i>KDM5B</i>	68	p.IVS12+1 G>A	LVO: Coarctation	No	n/a	<5	<5
1-01907	<i>UBE2B</i>	146	p.Arg8Thr	CTD: TOF	No	normal	50	10
1-00075	<i>RNF20</i>	58	p.Gln83*	HTX: Dextrocardia, RAI, TAPVR, L-ventricular loop, CAVC unbalanced-right dominant, PA	Low-set ears, excess nuchal skin, hydronephrosis, wide-spaced 2nd toe, Asplenia, primary cilia dyskinesia	abnormal	<5	<5
1-01260	<i>USP44</i>	0	p.Glu71Asp	LVO: ASD, mitral atresia, aortic atresia, HLHS	No	normal	25	25
1-02020	<i>SMAD2</i>	38	p.IVS6+1 G>A	HTX: Dextrocardia, ASD, CAVC-unbalanced, DORV, D-TGA, PS	Asplenia	normal	95	10
1-02621	<i>SMAD2</i>	38	p.Trp244Cys	HTX: Dextrocardia, LSVC to LA, PAPVR, CAVC unbalanced-right dominant, DORV, PS	Abnormal nose, foot syndactyly, malrotation	n/a	50	50
1-01451	<i>MED20</i>	25	p.IVS2+2 T>C	HTX: Dextrocardia, PAPVR, mitral atresia, HLHS, aortic atresia, hypoplastic AA	No	abnormal	<5	10
1-01151	<i>SUV420H1</i>	44	p.Arg143Cys	CTD: dbl AA	No	abnormal	50	10
1-00750	<i>HUWE1</i>	260	p.Arg3219Cys	LVO: Mitral stenosis, aortic stenosis, HLHS	No	mild abnormal	25	25
1-00577	<i>CUL3</i>	57	p.Iso144Phe fs*23	LVO: Hypoplastic mitral valve, hypoplastic aortic annulus, aortic stenosis, coarctation	Congenital hip dysplasia, congenital scoliosis	n/a	12	60
1-00116	<i>NUB1</i>	45	p.Asp310His	CTD: LSVC, sinus venosus ASD, truncus arteriosus, VSD-muscular	Spine lipoma	abnormal	<5	5
1-01828	<i>DAPK3</i>	55	p.Pro193Leu	CTD: TOF	No	n/a	n/a	n/a
1-03151	<i>SUPT5H</i>	133	p.Glu451Asp	LVO: BAV, aortic stenosis	No	n/a	75	90
1-00455	<i>NAA15</i>	214	p.Lys335Lys fs*6	HTX: Dextrocardia, TAPVR, LSVC, hypoplastic TV, DORV, hypoplastic RV, D-TGA, PS	Hydronephrosis, asplenia, malrotation	normal	50	50
1-00141	<i>NAA15</i>	214	p.Ser761*	CTD: TOF, single LCA	No	n/a	<5	20
1-01138	<i>USP34</i>	65	p.Leu432Pro	LVO: supra MS, BAV, CoA	No	n/a	25	>95
1-00448	<i>NF1</i>	55	p.IVS6+4 del A	CTD: PA VSD-MAPCAs	No	n/a	5	5
1-00802	<i>PTCH1</i>	32	p.Arg831Gln	LVO: HLHS	No	n/a	50	50
1-02458	<i>SOS1</i>	28	p.Thr266Lys	Other: ASD (multiple), dysplastic mitral, tricuspid and pulmonic valves	Macrocephaly, dolichocephaly, low-set ears, hyperextensible fingers, foot syndactyly, café-au-lait spots	abnormal	<5	5
1-02952	<i>PITX2</i>	18	p.Ala47Val	LVO: CoA	No	n/a	75	>95
1-01913	<i>RAB10</i>	119	p.Asn112Ser	Other: DILV, D-TGA, BAV, CoA	No	n/a	30	95

<sup>†</sup>Heart expression refers to # reads per million at murine e14.5. Mutation denotes the impact on encoded protein in three letter code; \* denotes termination mutation. *Frameshift* mutation in *MLL2*, *CUL3* and *NAA15*. 'IVS' stands for intervening sequence. 'fs' stands for frameshift. *Splice site* mutation in *KDM5B*, *MED20*, and *SMAD2* occur at 1st base of canonical splice donor of intron 12, at 2nd base of canonical splice donor of intron 2 and 1<sup>st</sup> base of canonical splice donor of intron 6 respectively.

<sup>§</sup>HLHS-hypoplastic left heart syndrome; Dbl AA-double aortic arch; TOF-tetralogy of Fallot; PAPVR-partial anomalous pulmonary venous return; LSVC-left superior vena cava; LA-left atrium; CAVC-complete atrioventricular canal defect; TAPVR- total anomalous pulmonary venous return; MV-mitral valve; BAV-bicuspid aortic valve; ASD-atrial septal defect; VSD- ventricular septal defect; PA-pulmonary atresia; RAI- right atrial isomerization.

**Table S11. Genes with > 1 de novo mutation in CHD probands**

ID	Gene	Heart Exp <sup>†</sup>	Mutation	Primary Classification: Specific Cardiovascular Diagnoses <sup>§</sup>	Extracardiac Anomalies	Neuro-Developmental Abnormal	Somatic Growth	
							HT(%)	Wt(%)
1-00455	<i>NAA15</i>	214	p.Lys335Lys fs*6	HTX: Dextrocardia, TAPVR, LSVC, hypoplastic TV, DORV, hypoplastic RV, D-TGA, PS	Hydronephrosis, asplenia, malrotation	Yes	50	50
1-00141	<i>NAA15</i>	214	p.Ser761*	CTD: TOF, single LCA	No	n/a	<5	20
1-02020	<i>SMAD2</i>	38	p.IVS6+1 G>A	HTX: Dextrocardia, ASD, CAVC-unbalanced, DORV, D-TGA, PS	Asplenia	No	95	10
1-02621	<i>SMAD2</i>	38	p.Trp244Cys	HTX: Dextrocardia, LSVC to LA, PAPVR, CAVC unbalanced-right dominant, DORV, PS	Abnormal nose, foot syndactyly, malrotation	n/a	50	50
1-02121	<i>DST</i>	122	p.Gly2936Asp	CTD: Right aortic arch-abn branching, vascular ring	Absent right kidney, dysplastic left kidney	Yes	<5	<5
1-02394	<i>DST</i>	122	p.Lys2653Ile	LVO: Coarctation	No	Yes	75	50
1-01538	<i>MKRN2</i>	19	p.Ala251Val	HTX: Mesocardia, atrial situs inversus, VSD-malalignment, D-TGA, PA	Abdominal situs inversus	No	50	50
1-00230	<i>MKRN2</i>	19	p.Arg50Trp	LVO: LSVC, primum ASD, cleft MV, sub-AS, BAV	No	No	<5	<5
1-01664	<i>OBSCN</i>	298	p.Phe5295Ser	HTX: Dextrocardia, LSVC, hypoplastic MV, sub AS, BAV, congenital coronary abnormality, coarctation	No	n/a	25	10
1-03190	<i>OBSCN</i>	298	p.Thr4421Met	CTD: TOF-PA with MAPCAs	No	No	75	75
1-01341	<i>UMODL1</i>	0	p.Val1090Met	CTD: TOF, ASD, LPA stenosis	No	No	n/a	n/a
1-02408	<i>UMODL1</i>	0	p.Lys1251Glu	HTX: Dextrocardia, mitral atresia, PA, L-TGA	Epicanthic folds, flat nasal bridge, short neck	No	10	10

<sup>†</sup>Heart expression refers to # reads per million in murine hearts at e14.5. Mutation denotes the impact on encoded protein in single letter code; \* denotes termination mutation. 'IVS' stands for intervening sequence. 'fs' stands for frameshift. *Frameshift* mutation in *NAA15*. *Splice site mutation* in *SMAD2* occurs at 1st base of canonical splice donor of intron 6.

<sup>§</sup>LSVC-left superior vena cava; LA-left atrium; PAPVR-partial anomalous pulmonary venous return; CAVC-complete atrioventricular canal defect; TAPVR- total anomalous pulmonary venous return; ASD-atrial septal defect; DORV-double outlet right ventricle; D-TGA- dextro transposition of great arteries; PS-pulmonic stenosis; MV-mitral valve; BAV-bicuspid aortic valve; TOF-tetralogy of Fallot; LCA-left coronary artery, VSD- ventricular septal defect; LPA stenosis- late left pulmonary artery stenosis .

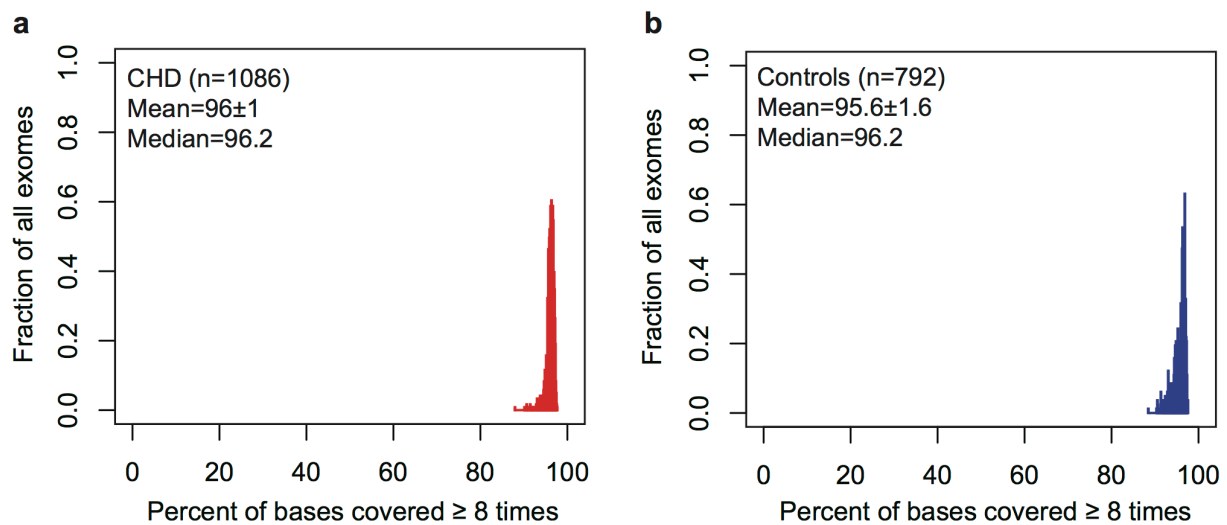
**Table S12. Candidate genes for LVO, CTD and HTX**

ACP6	CCDC40	FBN2	IGFBP4	MYH11	PPP3CA	SUPT3H
ACTA2	CCT4	FGF8	IGFBP5	MYH6	PQPB1	TBX1
ACTC1	CDH2	FGFR1	IHH	NEK2	PRKAB2	TBX20
ACVR1	CER1	FIBP1	INVERSIN	NF1	PROX1	TBX3
ACVR2B	CFC1	FLNA	IPPK	NFATC1	PTCH1	TBX5
AHSA2	CHD1L	FMO5	ISL1	NFATC3	PTCH2	TCF21
AKAP9	CHRA1	FOXA2	JAG1	NFATC4	PTPN11	TDGF1
ANKRD1	CHRD	FOXC1	JAZF1	NIPBL	RAB10	TFAP2B
APOBEC2	CITED2	FOXH1	KCNE1	NKD1	RAB23	TGFBR2
ARL13B	CLDN7	FOXJ1	KCNJ2	NKX2-5	RAI1	TLL1
ASXL2	CLUL1	FOXL2	KCNQ1	NKX2-6	RAI2	TMBIM4
ATE1	CREBBP	FTO	KIAA1841	NKX2.5	RAPGEF5	TMEM195
ATP1A2	CRELD1	GADL1	KIF3A	NKX3-2	REL	TNFRSF21
ATP4A	CRHBP	GALNT11	KIF3B	NODAL	RFX2	TSC1
ATP4B	CRX	GATA4	KIF3C	NOTCH1	RFX3	TSEN15
BBS1	CSRP1	GATA5	KIFAP3	NOTCH2	ROCK2	TTC21B
BBS10	CTNNA3	GATA6	KLF13	NOTCH2NL	ROR2	TTC30A
BBS11	DAND5	GDF1	LBR	NOTCH3	ROTATIN	TWIST1
BBS12	DHCR7	GJA1	LEFTY1	NOTCH4	RPGRIP1L	TXNDC3
BBS2	DLL1	GJA5	LEFTY2	NOTO	RUNX2	UBR1
BBS3	DMRT2	GJA8	LEMD3	NPHP3	S100Z	USP34
BBS4	DNAH2	GJA9	LLPH	NPPA	SALL1	VANGL2
BBS5	DNAH5	GPC3	LPIN1	NSD1	SALL4	VEGFA
BBS6	DNAI1	GPR161	LRP2	NUMBL	SDC2	VEGFC
BBS7	DNAI2	GPRC6A	LRRC50	NUP188	SEMA3E	VIT
BBS8	DOLK	GSK3B	LRRC6	OFD1	SESN1	WNT3A
BBS9	DPPA4	HAND1	MARK2	OSR1	SHH	XPO1
BCL11A	DQ983818	HAND2	MAX	PAPOLG	SHOC2	ZAC1
BCL6	DVL1	HES1	MED13L	PCMTD2	SIL	ZEB2
BCL9	DVL2	HES4	MEF2A	PCNT	SLC2A10	ZFPM1
BCOR	DZIP1	HEY2	MEF2C	PCSK5	SMAD2	ZIC3
BICC1	EED	HOXA1	METT10D	PEX1	SMAD5	ZNF480
BMP4	EHMT1	ID2	MGAT1	PEX13	SMARCD3	ZNF528
BMP7	ELN	IER2	MGP	PHYHD1	SMO	ZNF534
BMPR1A	EP300	IFT122	MID1	PIFO	SNAI1	ZNF610
BMPR1B	ESCO2	IFT172	MKKS	PITX2	SOS1	ZNF638
BMPR2	EVC	IFT20	MKS1	PKD1L1	SOX17	ZNHIT3
BUB1B	EVC2	IFT57	MNDA	PKD2	SRF	
C1orf106	EZH1	IFT88	MSX1	PLAGL1	STIL	
CCDC39	FBN1	MYH10	MSX2	PPM1K	SUFU	

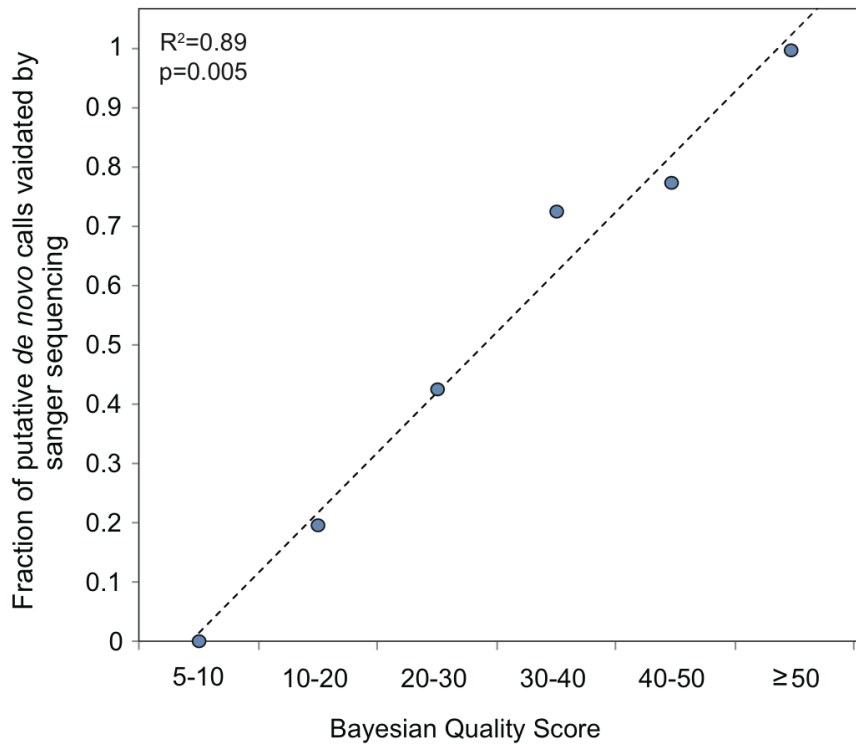


**Table S13. *De novo* mutations in CHD cohort that occur in candidate gene set**

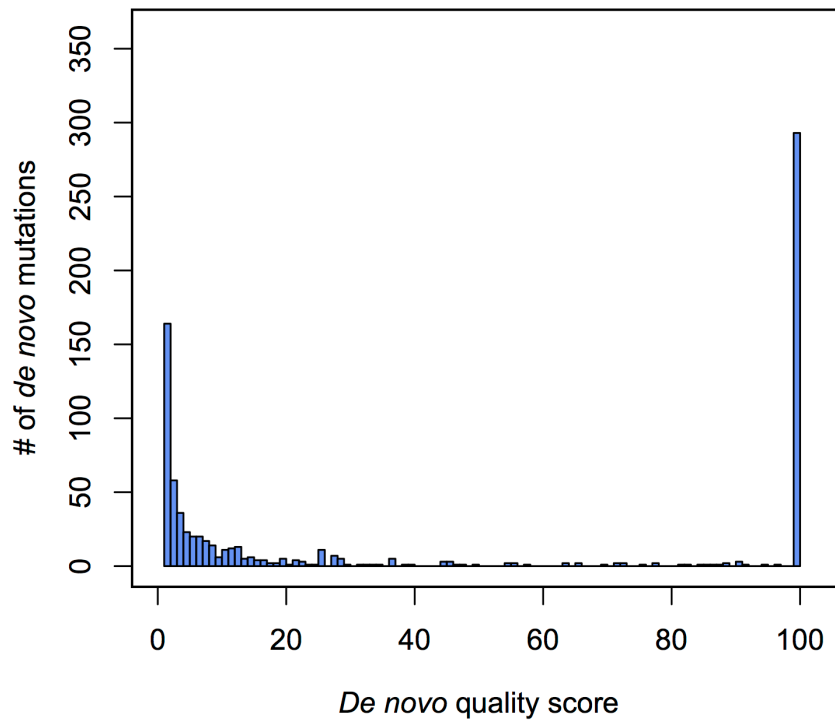
CHD Cohort	Gene	AA Change
1-00638	FBN2	p.Asp2191Asn
1-00596	MLL2	p.Ser1722Argfs*9
1-00534	CHD7	p.Gln1599*
1-01913	RAB10	p.Asn112Ser
1-00197	BCL9	p.Met1395Lys
1-01138	USP34	p.Leu432Pro
1-00448	NF1	p.IVS6+4 delA
1-02020	SMAD2	p.IVS6+1 G>A
1-02621	SMAD2	p.Trp224Cys
1-00802	PTCH1	p.Arg831Gln
1-02458	SOS1	p.Thr266Lys
1-02952	PITX2	p.Ala47Val
1-02598	LRP2	p.Glu4372Lys



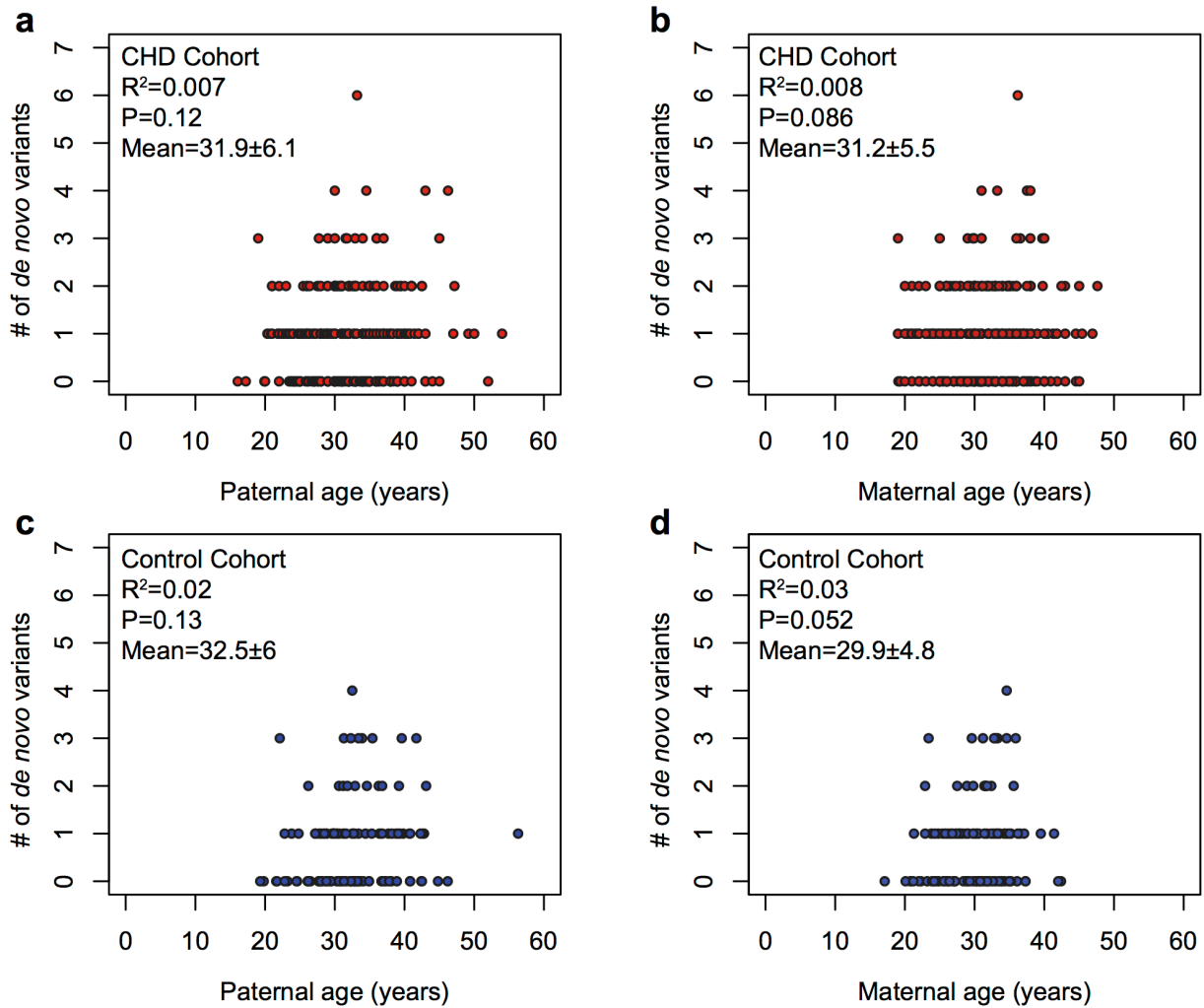
**Figure S1. Distribution of 8X read coverage between cases and controls.** Percent of bases covered  $\geq 8$  times plotted for CHD cases (a) and controls (b). Distribution, mean and median are highly concordant.



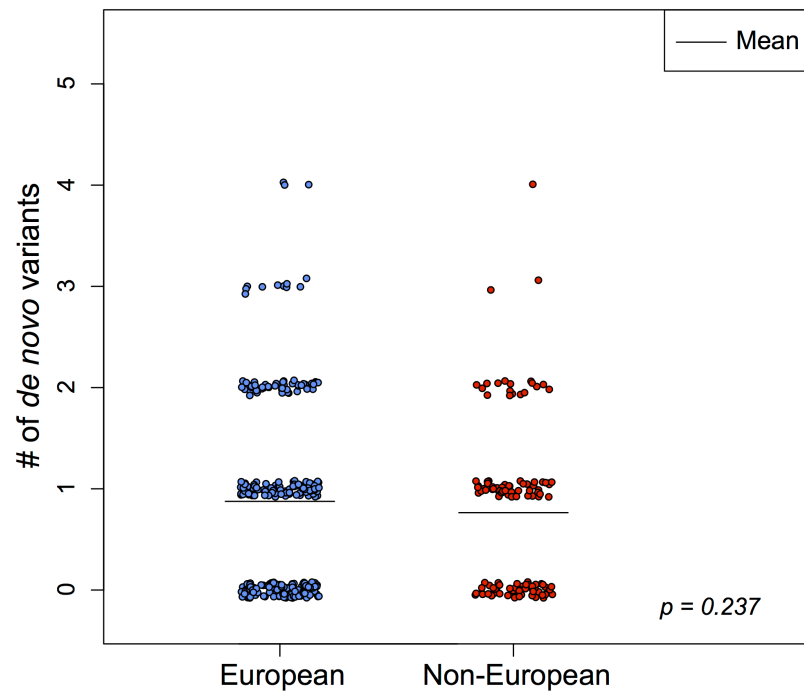
**Figure S2. Correlation of Bayesian quality score and probability of Sanger validation.** A strong correlation between the fraction of putative *de novo* variants and specific ranges of Bayesian quality scores,  $R^2=0.89$ . Notably, Sanger sequencing validated a subset (88) of all *de novo* calls with a Bayesian quality score  $\geq 50$ , with a specificity of 100%.



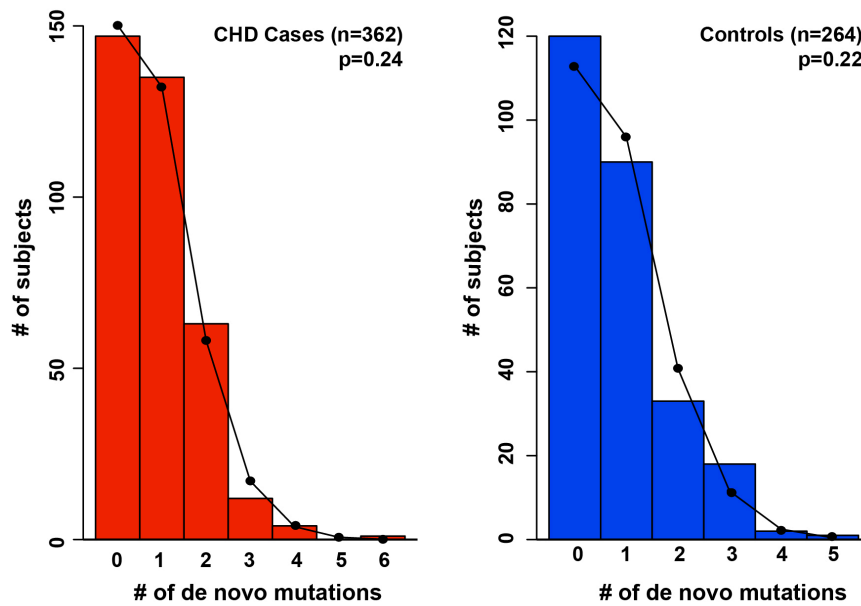
**Figure S3. Distribution of *de novo* mutation quality scores.** The frequency of potential *de novo* mutations in different bins of *de novo* quality scores are shown. 100% of variants with scores  $\geq 50$  confirmed as *de novo* mutations by Sanger sequencing. Of these, ~90% had the maximum QS of 100. (see Supplementary Table 3 and Supplementary Figure 2).



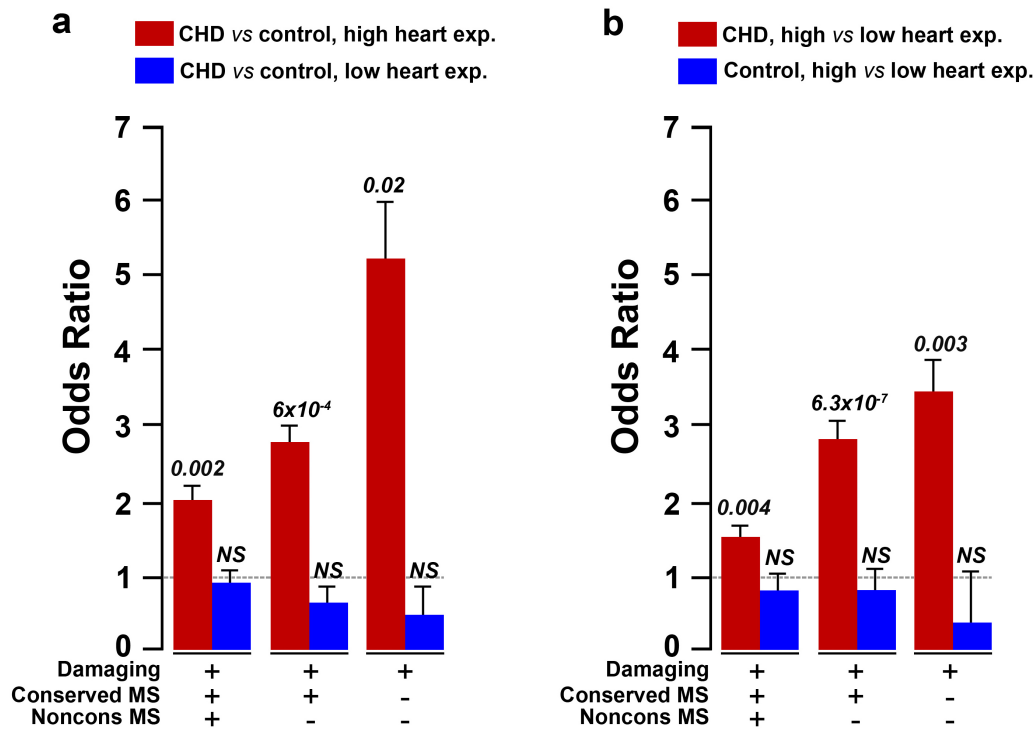
**Figure S4. Correlation of paternal and maternal age with *de novo* mutation rate.** Weak correlation between paternal (a,c) or maternal (b,d) age and number of *de novo* mutations per subject in the CHD (a,b) or control (c,d) cohorts.



**Figure S5. Effect of ancestry on *de novo* mutation rate.** No significant difference in *de novo* mutation rate between European and Non-European (Indian, Mexican, African-American, and East Asian) ancestries,  $p=0.24$ .



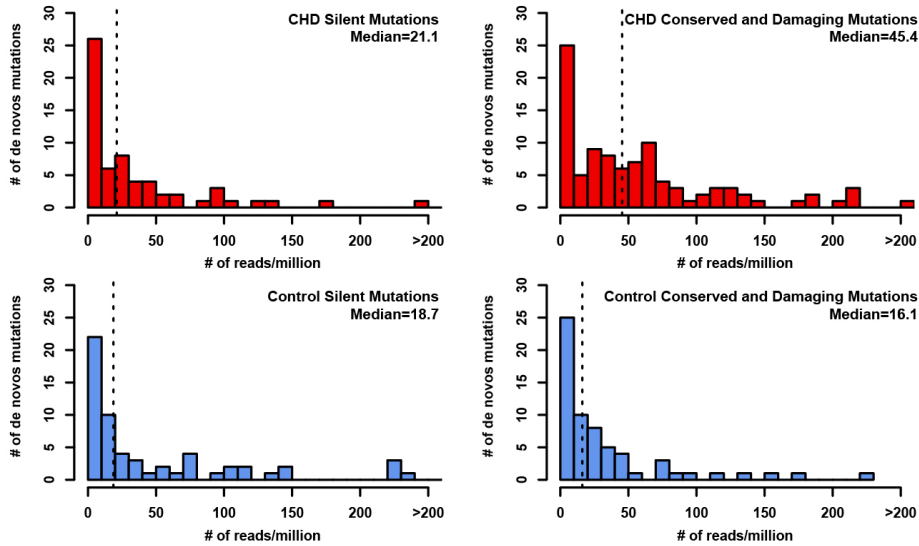
**Figure S6. *De novo* mutation rate closely approximates Poisson distribution in CHD cases and controls.** Observed number of *de novo* mutations per subject (bars) compared to the numbers expected (line) from the Poisson distribution in CHD (red) and control (blue) cohorts. ‘p’ denotes Chi-squared value.



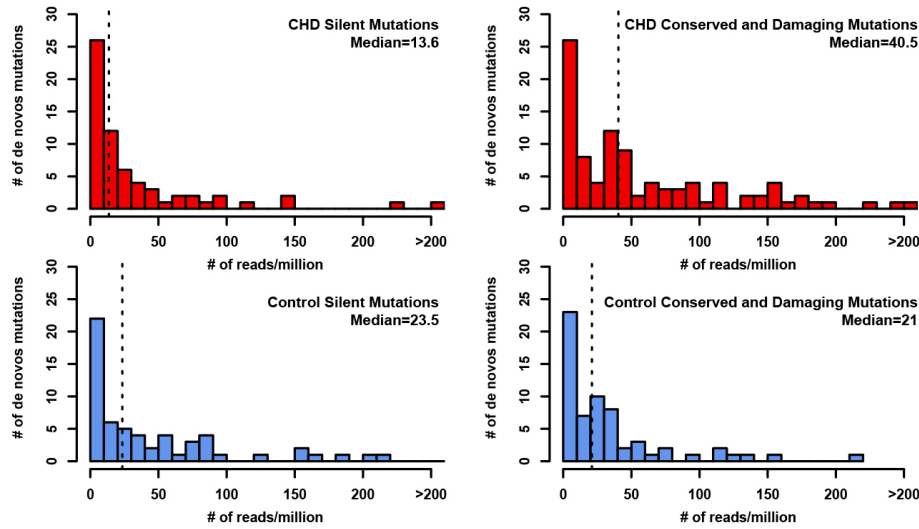
**Figure S7. *De novo* mutations in CHD probands and controls stratified for gene expression in developing heart; categorical analysis of the presence or absence of any *de novo* mutation in probands a**, Odds ratios (ORs) comparing fraction of patients with and without *de novo* mutations of indicated classes in CHD cases and controls for genes in top quartile (red bars) and bottom 75% (blue bars) of expression in developing heart. **b**, ORs comparing the presence of *de novo* mutations in genes in top quartile versus bottom 75% of expression in developing heart in CHD cases (red bars) and controls (blue bars). *De novo* mutations are classified as missense mutations at poorly conserved positions (noncons MS), missense mutations at highly conserved positions (cons MS), and damaging (nonsense, splice site, or frameshift mutations). Odds ratio + SEM is shown and significance of the difference between groups is indicated (*P* values from two-tailed Fisher exact test). Odds ratio compares the proportion of # of individuals with *de novo* mutations to the # of individuals without a *de novo* mutation in each specific category. NS, not significant.



**Mouse heart e14.5**

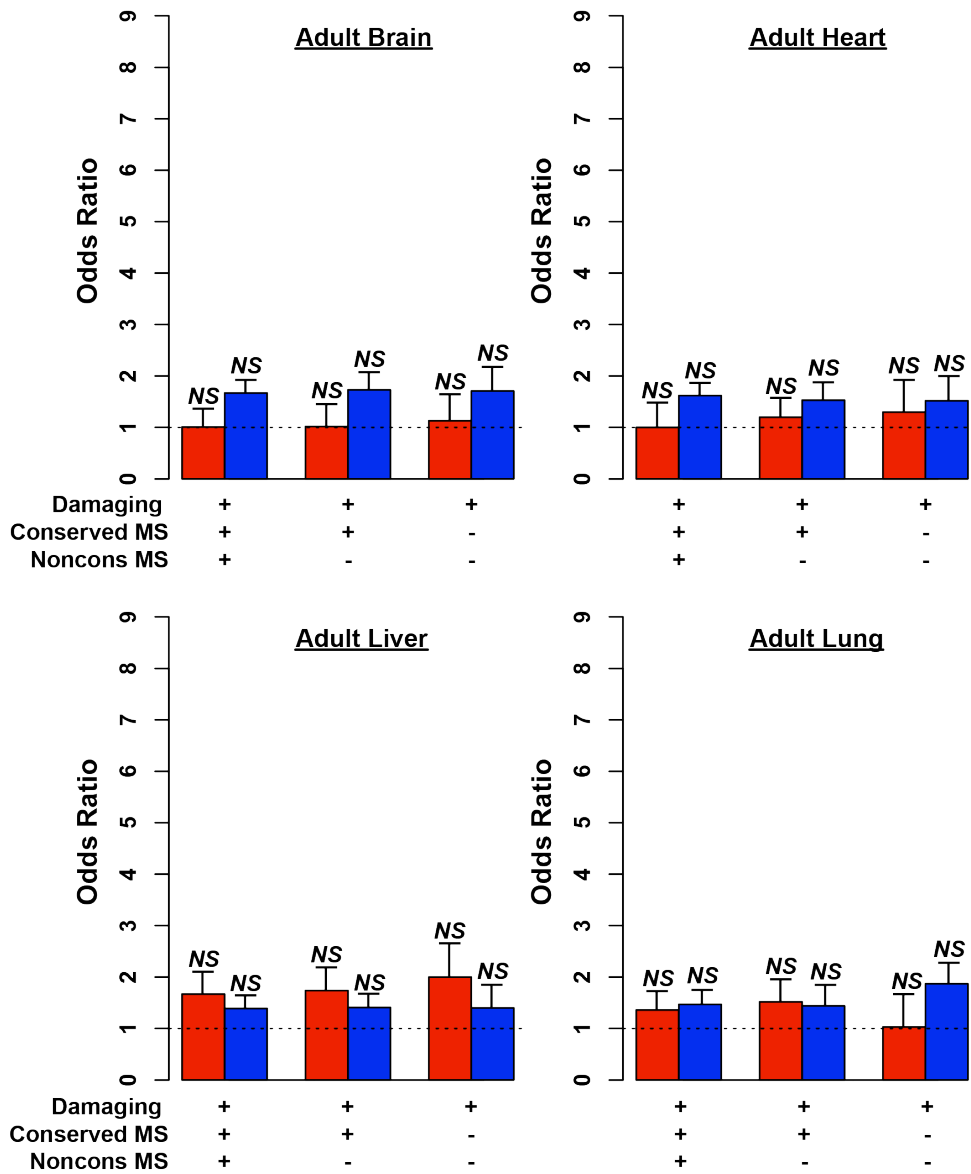


**Mouse heart e9.5**



**Figure S8. Expression of genes mutated in cases and controls.** Expression levels (# of reads per million at e14.5 and e9.5) are shown for silent (a,c) and damaging and conserved (b,d) *de novo* mutations for cases (a,b) and controls (c,d). Vertical dashed lines indicate median values in each category. Conserved and damaging *de novo* mutations show higher expression in CHD cases than controls ( $P=5 \times 10^{-4}$  at e14.5 and  $P=1.6 \times 10^{-3}$  at e9.5), while silent mutations show no significant difference ( $P=0.7$  at e14.5 and  $P=0.5$  at e9.5) (comparison via Wilcoxon signed-ranked test).

■ CHD vs. control, high tissue exp.  
■ CHD vs. control, low tissue exp.



**Figure S9. *De novo* mutations in CHD cases and controls stratified for gene expression in adult tissues.** Odds ratios (ORs) comparing incidence of indicated classes of *de novo* mutations in CHD cases and controls for genes in top quartile (red bars) and bottom 75% (blue bars) of expression in different adult tissues. Odds ratio + SEM is shown and significance of the difference between groups is indicated (*P* values calculated from two-tailed binomial exact test). The odds ratio is the ratio of protein altering to silent variants in cases divided by the corresponding ratio in controls. NS, not significant.