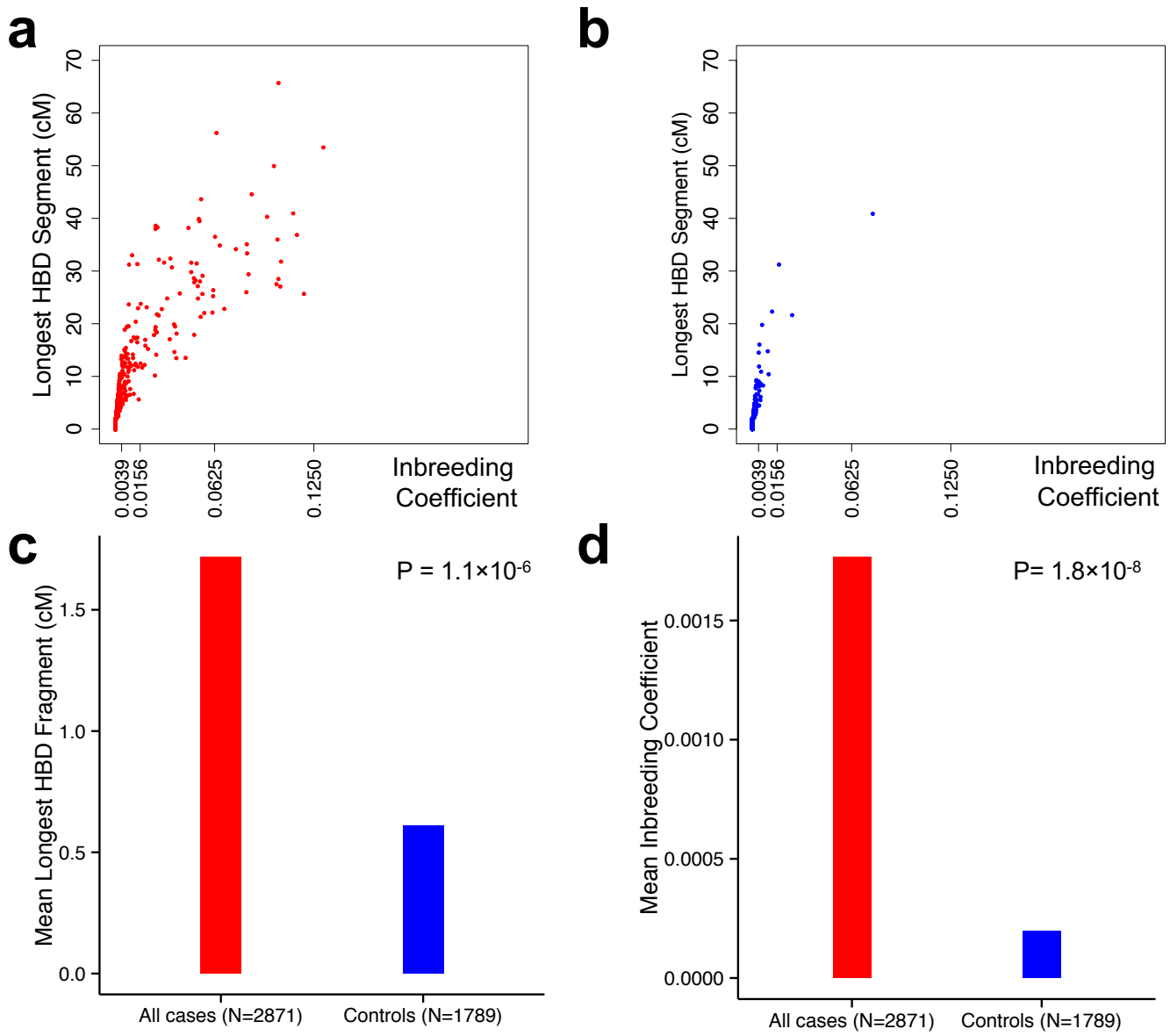
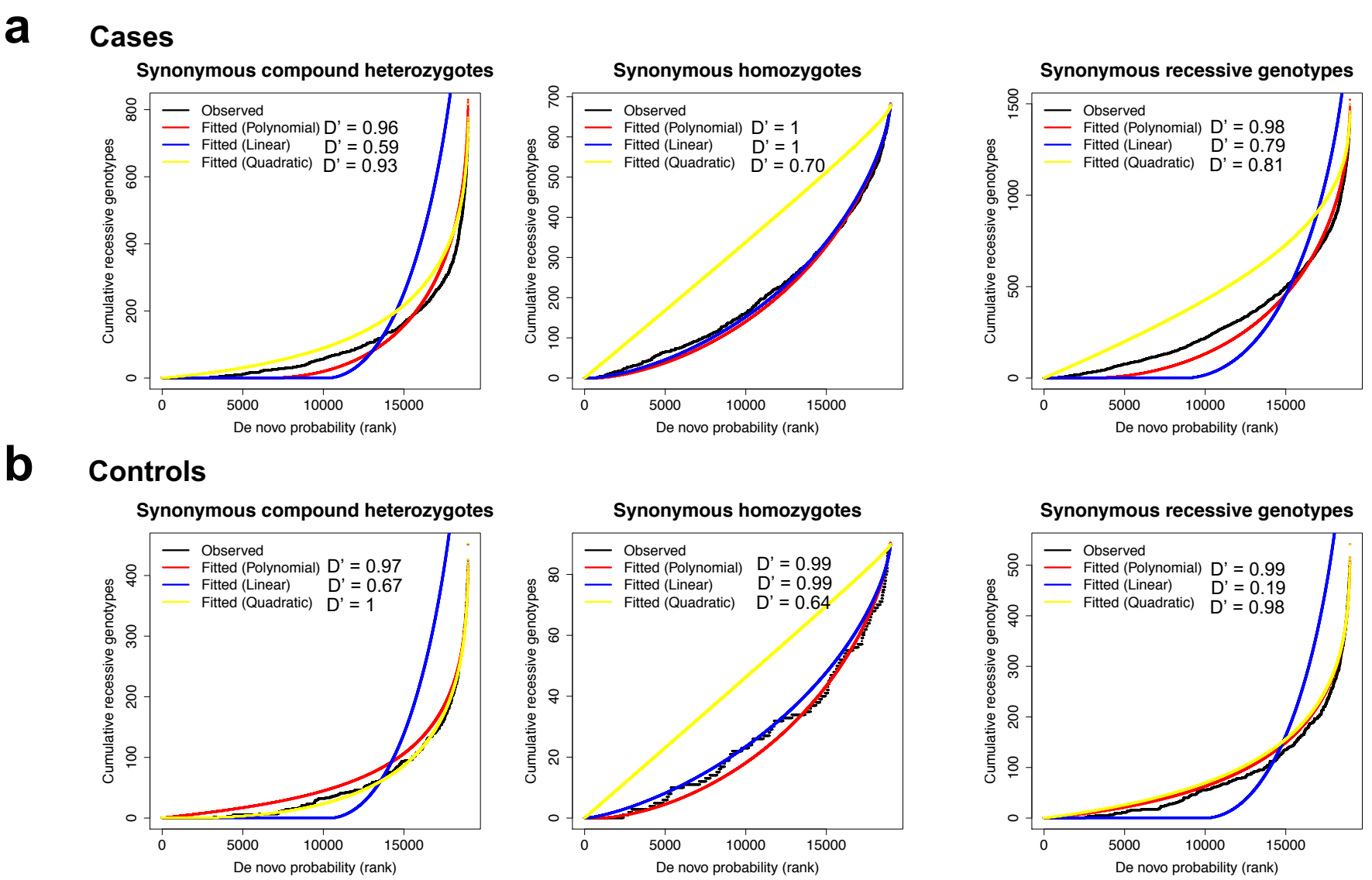
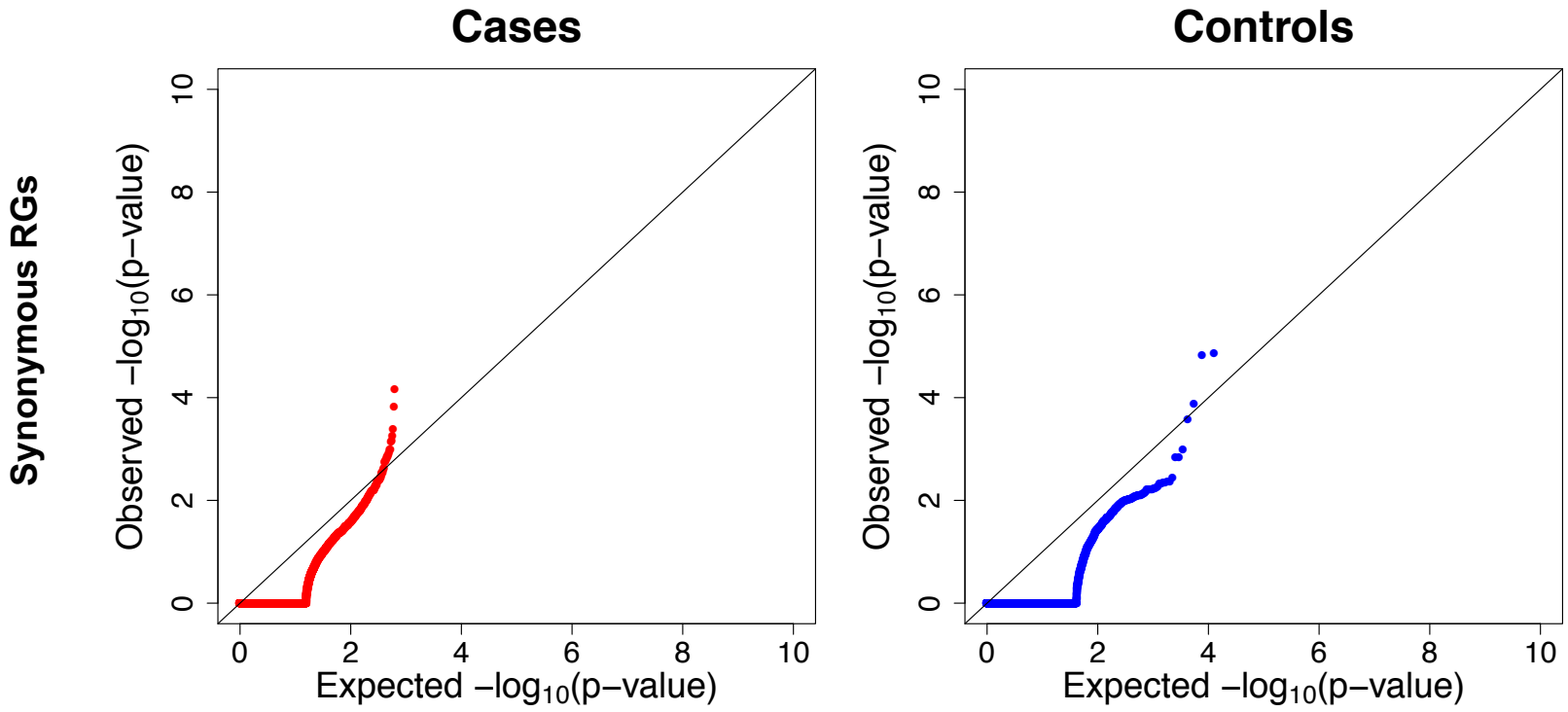


Supplementary Figure 1. Significant difference in the degree of consanguinity between cases and controls. Red and blue dots represent the longest homozygosity-by-descent (HBD) segments and inbreeding coefficients of case **(a)** or control **(b)** subjects, respectively. For reference, the ticks on the x-axis correspond to the inbreeding coefficient for third, second, first cousin, and half-siblings (from smallest to largest). The distribution of the mean longest HBD segment **(c)** and inbreeding coefficient **(d)** respectively between cases and controls was compared using a Kolomogorov-Smirnov non-parametric test.

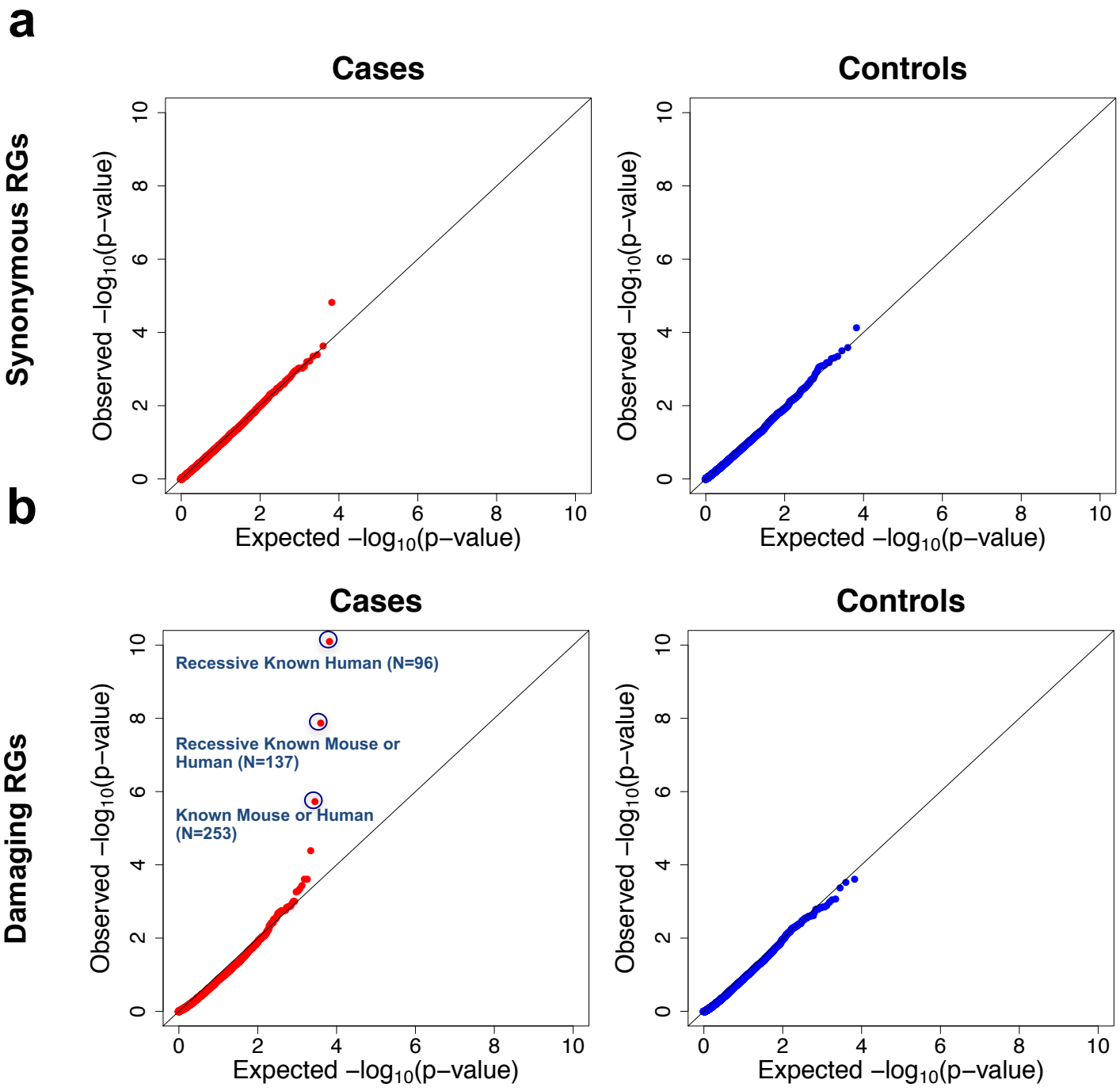
Supplementary Figure 2. Correlation between the observed and expected allele frequencies for rare (allele frequency < 10^{-5}) damaging heterozygous variants in cases (a) and controls (b). The expected and observed allele frequencies (expected values based on each gene's probability of *de novo* mutation) for each gene is shown. Two genes, *TTN* and *RYR1*, were extreme outliers for expected and observed allele frequency and were excluded from the plot for clarity, but included in the correlation calculation. The expected number of rare damaging heterozygotes in each gene was calculated using a linear model including the *de novo* probability stratified by the pLI score (see **Online Methods**). The Pearson correlation coefficients without taking into account *TTN* and *RYR1* are 0.80 in cases and 0.82 in controls, respectively.



Supplementary Figure 3. The polynomial model using *de novo* probabilities as predictors fits the observed number of synonymous recessive genotypes (RG) appropriately in cases and controls at the gene level. A polynomial regression model (RG Count = $\beta_0 + \beta_1 \times$ (mutability) + $\beta_2 \times$ (mutability^2) + ε) was fitted to determine the relationship between the *de novo* probability and the number of RGs for synonymous compound heterozygotes, homozygotes, and all RGs, respectively, in cases (a) and controls (b). The x-axis represents the rank of the synonymous *de novo* probability. The y-axis represents the cumulative count of synonymous RGs. The model fit was evaluated by the distance goodness-of-fit metric D' = $|\ 1 - (\Sigma all(observed_i - fitted_i)^2 / \Sigma all(observed_i)^2)|$; 1 is best fit and 0 is worst).



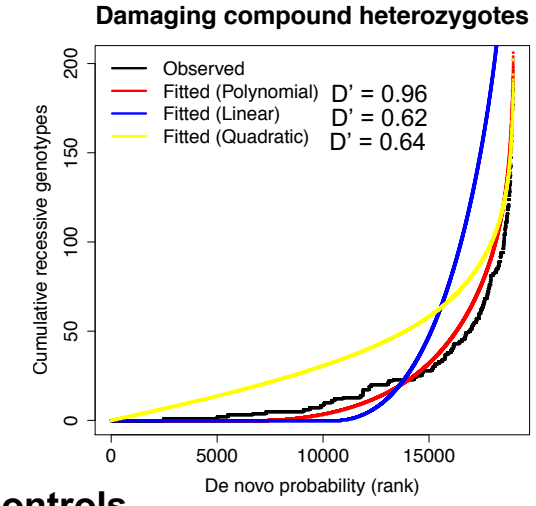
Supplementary Figure 4. Quantile-quantile plots of the observed versus expected p-values comparing the burden of synonymous recessive genotypes (RGs) adjusted for fitted values from the polynomial model for each gene. A polynomial regression model (RG Count = $\beta_0 + \beta_1 \times$ (mutability) + $\beta_2 \times$ (mutability^2) + ε) was fitted to determine the relationship between the *de novo* probability and the number of synonymous RGs within each gene in cases and controls, respectively. A one-sided binomial test was conducted to compare the observed number of RGs within each gene versus expected adjusting for the fitted values from the polynomial model. The quantile-quantile plots show that there is no inflation of p-values in both cases and controls for synonymous RGs, suggesting that the observed numbers of synonymous RGs in each gene match the expected number estimated using the polynomial model.

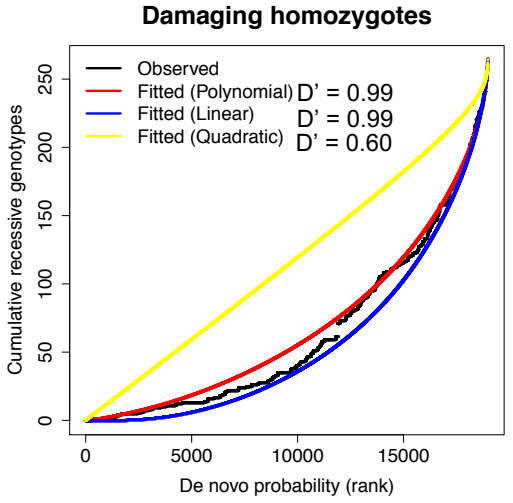


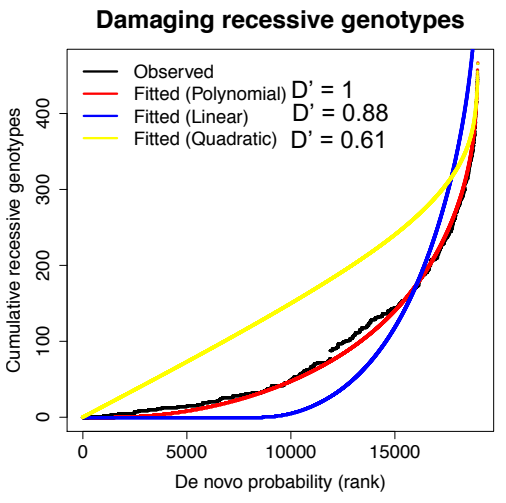
Supplementary Figure 5. Quantile-quantile plots of the observed versus expected p-values comparing the burden of synonymous (a) and damaging (b) recessive genotypes (RGs) adjusted for fitted values from the polynomial model *for a* gene-set. A polynomial regression model (RG Count = $\beta_0 + \beta_1 \times$ (mutability) + $\beta_2 \times$ (mutability^2) + ε) was fitted to determine the relationship between the *de novo* probability and the number of damaging RGs and synonymous RGs, respectively. We performed simulations by selecting gene sets of 100-1,000 genes at random from all genes. A one-sided binomial test was then conducted to compare the observed number of RGs in the randomly chosen gene set versus expected adjusting for the fitted values from the polynomial model. A total of 10,000 iterations were conducted. *GDF1*, *MYH6*, and *TTN* were removed in the simulation of damaging RGs in cases because they were known outliers. The numbers of damaging RGs in 'Known Mouse and Human' and 'Recessive Known Mouse and Human' gene sets far exceed the expected in cases; however, there was no enrichment in controls and for synonymous RGs.

a

Cases

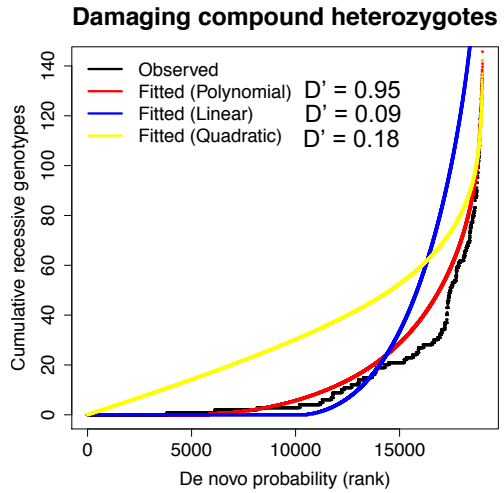


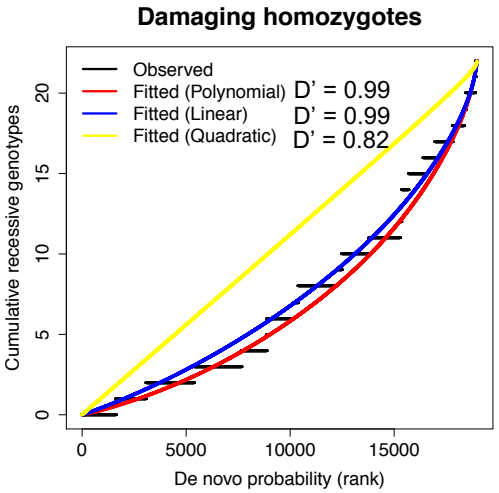


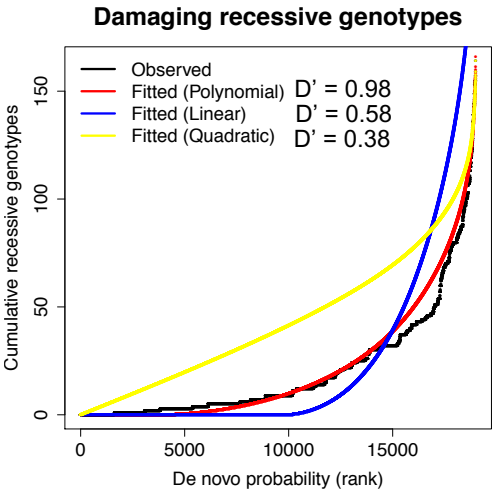


b

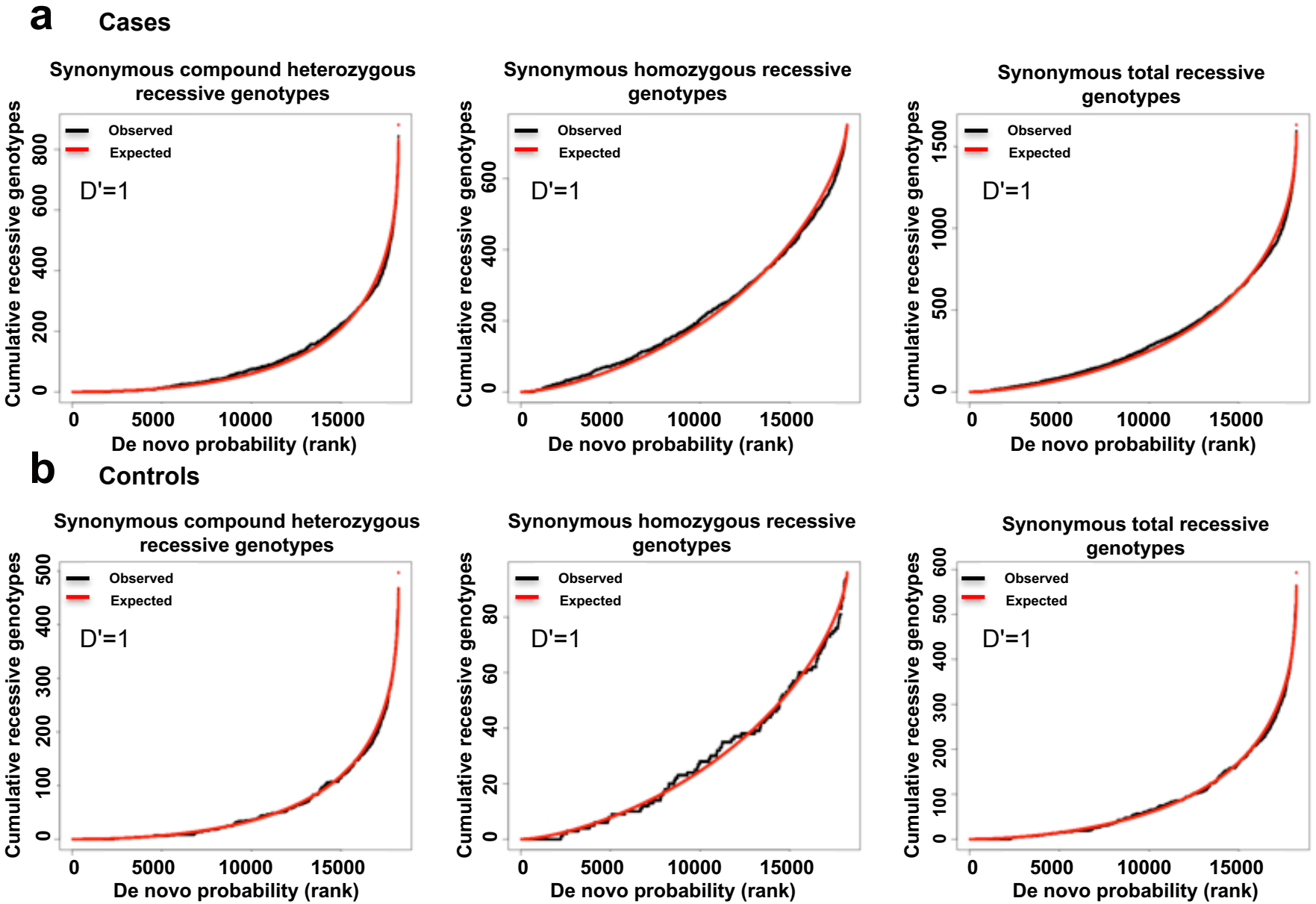
Controls



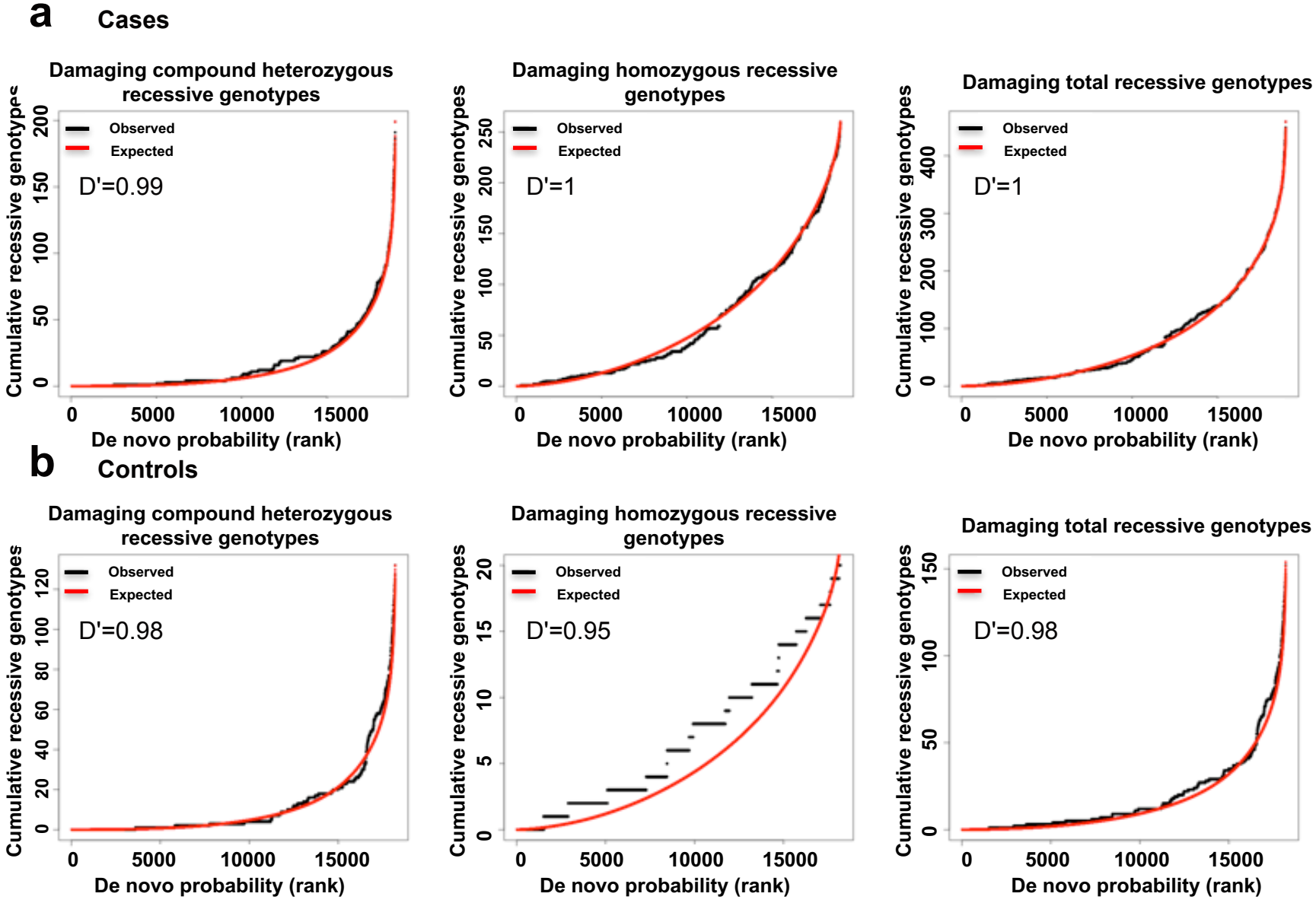




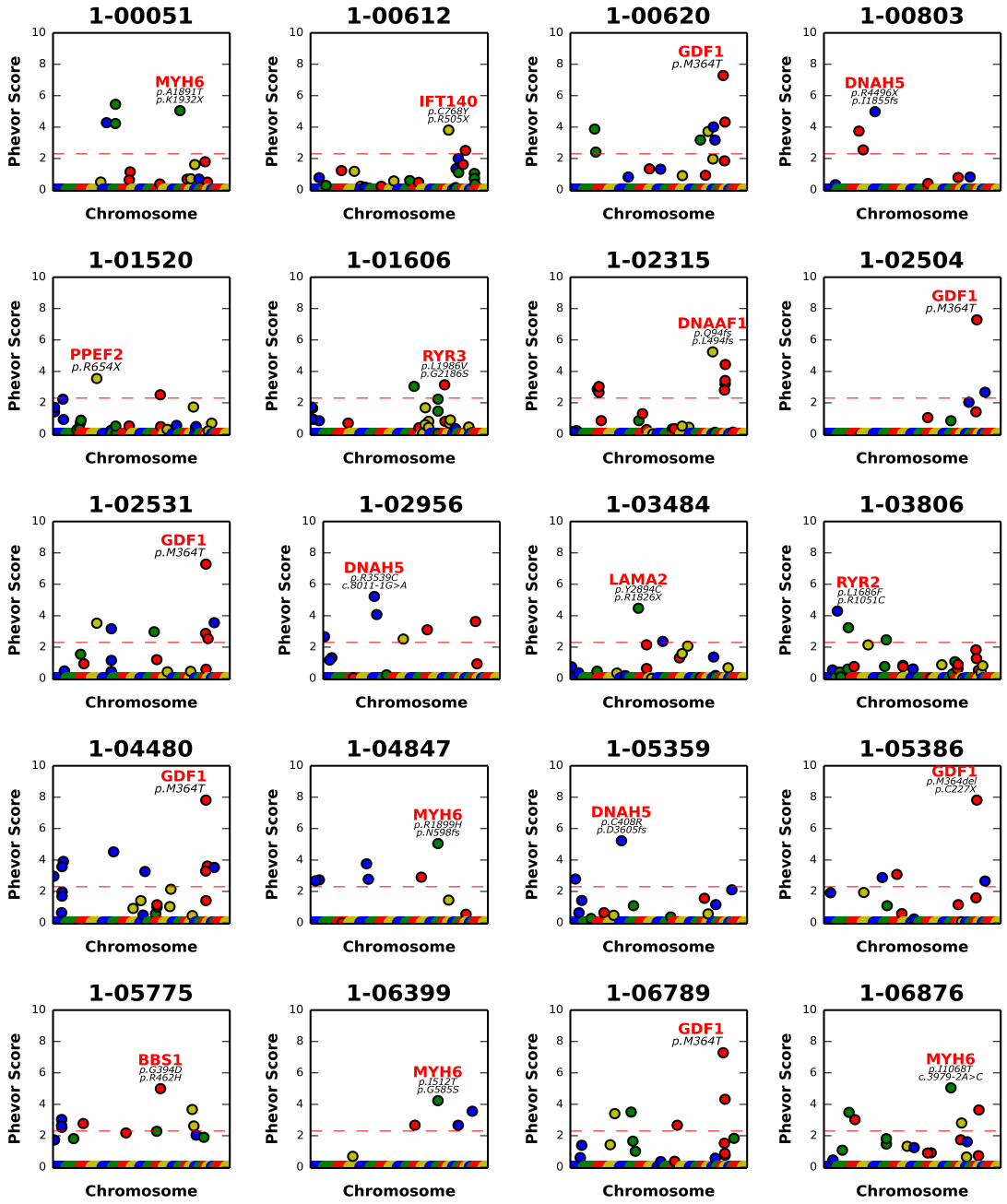
Supplementary Figure 6. The polynomial model using *de novo* probabilities as predictors fits the observed number of damaging recessive genotypes appropriately in cases and controls at the gene level. A polynomial regression (Recessive genotype (RG) Count = $\beta_0 + \beta_1 \times$ (mutability) + $\beta_2 \times$ (mutability^2) + ε) model was fitted to determine the relationship between the *de novo* probability and the number of RGs for damaging compound heterozygotes, homozygotes, and all RGs, respectively, in cases (a) and controls (b). The x-axis represents the rank of the damaging *de novo* probability. The y-axis represents the cumulative count of damaging RGs. The model fit was evaluated by the distance goodness-of-fit metric D' = $|\ 1 - (\Sigma all(observed_i - fitted_i)^2 / \Sigma all(observed_i)^2)|$; 1 is best fit and 0 is worst).



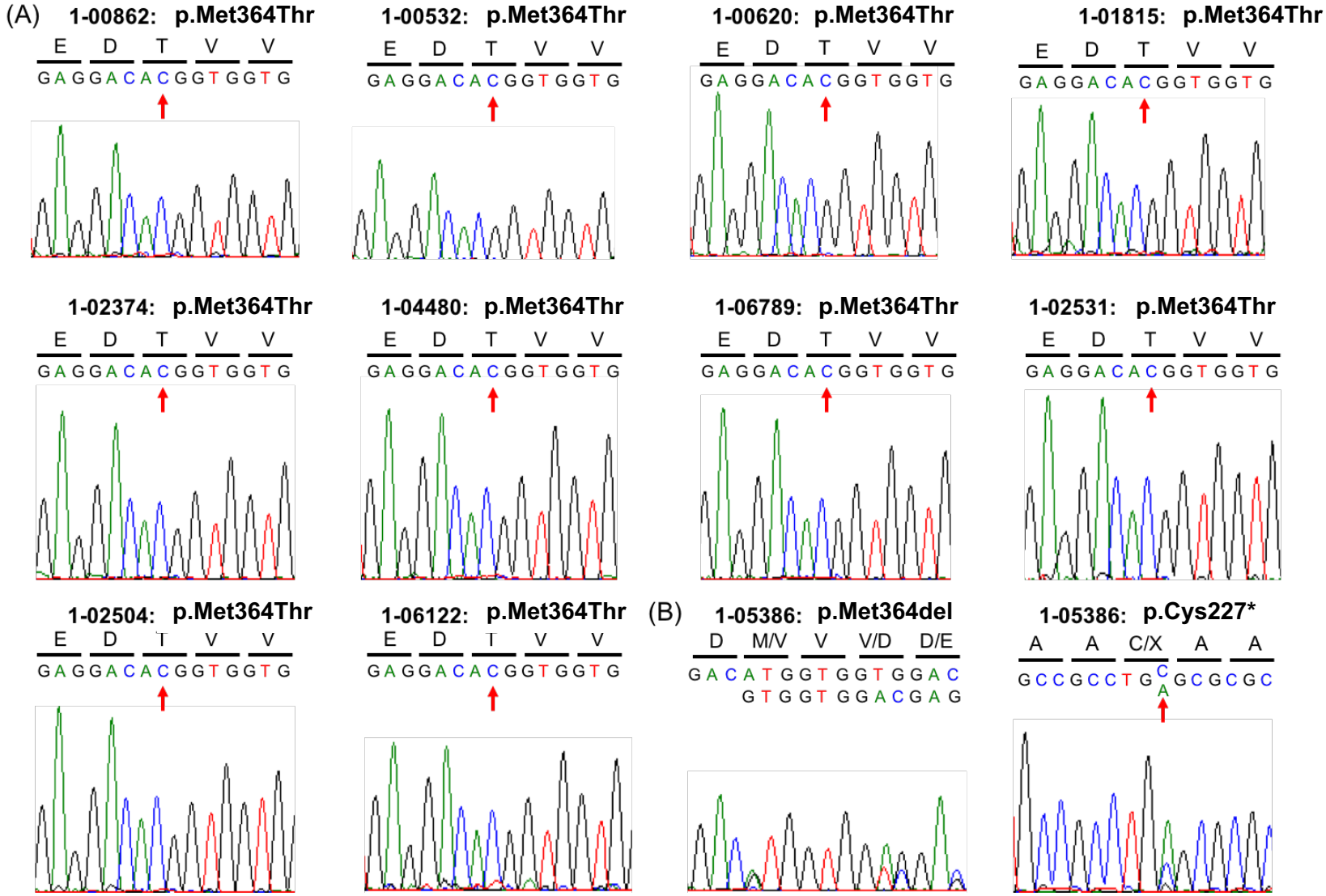
Supplementary Figure 7. Expected distribution of synonymous recessive genotypes when independently modeling compound heterozygous variants as a quadratic fit and homozygous variants as a linear fit using *de novo* probabilities. The cumulative number of recessive genotypes is plotted by the rank ordered probability of synonymous *de novo* mutation for each gene (smallest to largest). For each cohort, (a) cases and (b) controls, compound heterozygous and homozygous genotypes were considered independently to calculate the expected distribution of genotypes. The total number of recessive genotypes in each cohort is derived from the sum of the compound heterozygous and homozygous distributions. The model fit was evaluated by the distance goodness-of-fit metric D' (scaled from 0, worst fit, to 1, best fit). See **Online Methods** for details.



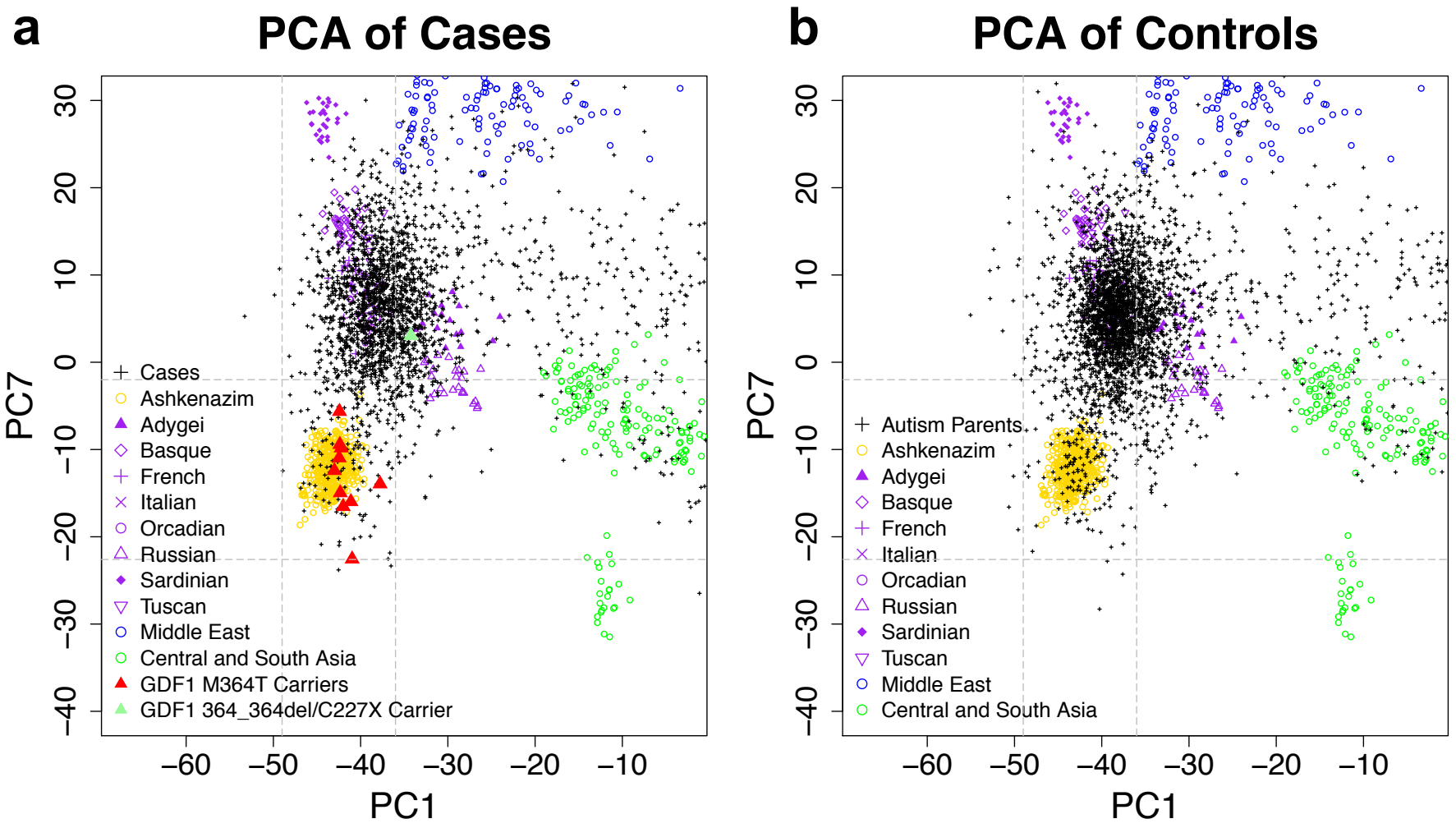
Supplementary Figure 8. Expected distribution of damaging recessive genotypes when independently modeling compound heterozygous variants as a quadratic fit and homozygous variants as a linear fit using *de novo* probabilities. The cumulative number of recessive genotypes is plotted by the rank ordered probability of damaging *de novo* mutation for each gene (smallest to largest). For each cohort, (a) cases and (b) controls, compound heterozygous and homozygous genotypes were considered independently to calculate the expected distribution of genotypes. The total number of recessive genotypes in each cohort is derived from the sum of the compound heterozygous and homozygous distributions. The model fit was evaluated by the distance goodness-of-fit metric D' (scaled from 0, worst fit, to 1, best fit). See **Online Methods** for details. The stepwise appearance of homozygous genotypes in controls (b, middle) is due to the low number of homozygous observations in this cohort (22).



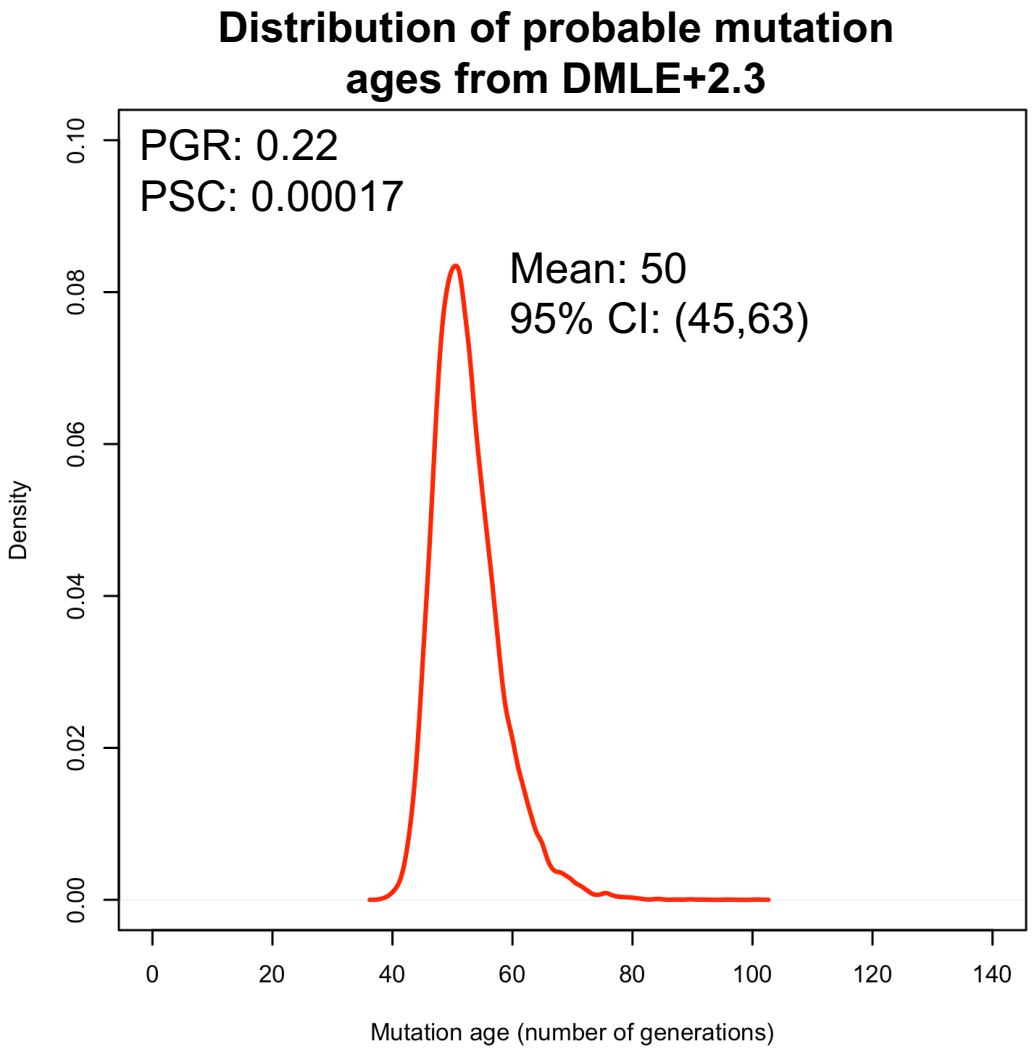
Supplementary Figure 9. Burden-test approach for recessive variants implicated in congenital heart disease. Trios were scored using the VAAST burden test, and candidate genes were prioritized using the PHEVOR ontological re-ranking tool (see **Supplementary Note**). Each Manhattan-style panel displays genes along chromosomes 1-23, X, and Y on the x-axis and PHEVOR scores on the y-axis (log_{10}). Genes with a PHEVOR score above the red dashed line have an HPO-based prior probability of association ≥ 0.5 with the proband's phenotype (e.g. conotruncal, left outflow obstruction, heterotaxy, etc.) and a VAAST p-value of ≤ 0.005. .



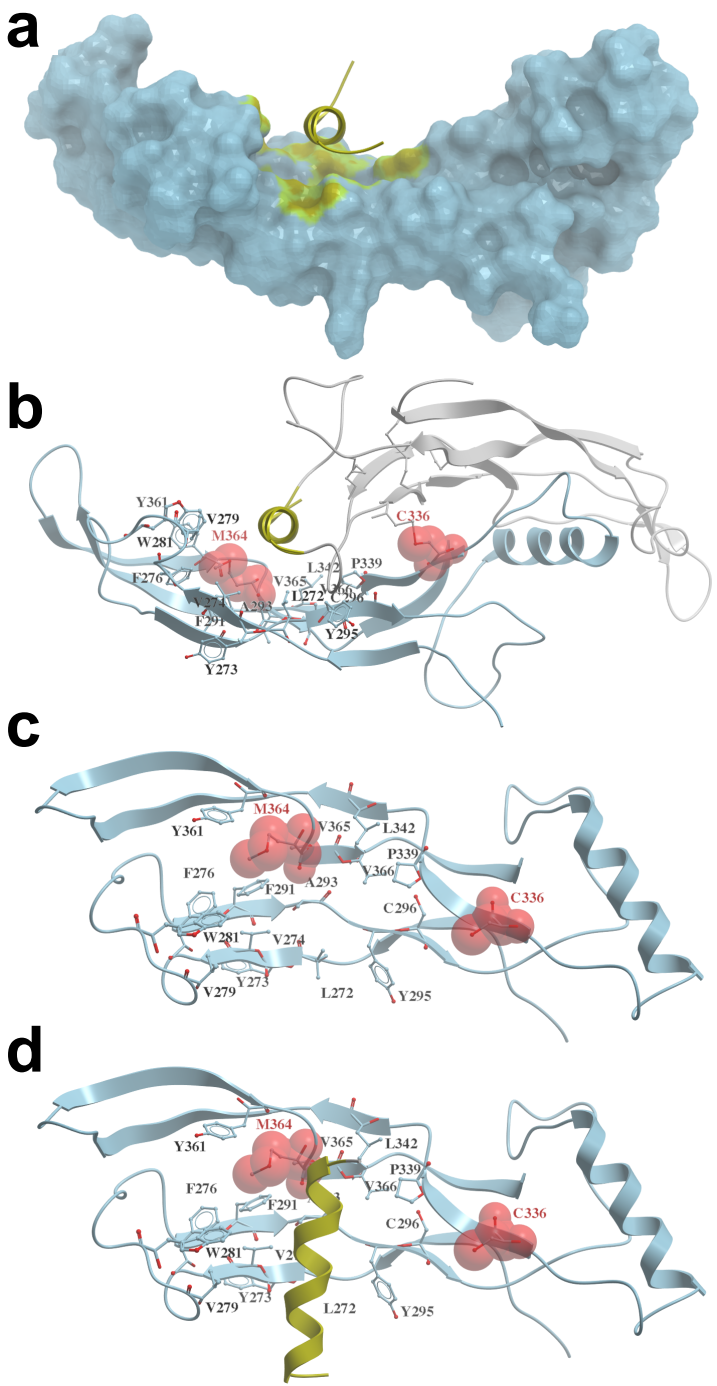
Supplementary Figure 10. *GDF1* variants in 11 CHD cases. (A) Sanger sequence chromatograms for 10 CHD cases carrying homozygous p.Met364Thr (c.1091T>C) mutations. (B) Sanger sequence chromatograms for a CHD case carrying a non-frameshift (p.Met364del; c.1090_1092delATG) and a stop-gain variant (p.Cys227*;c.681C>A).



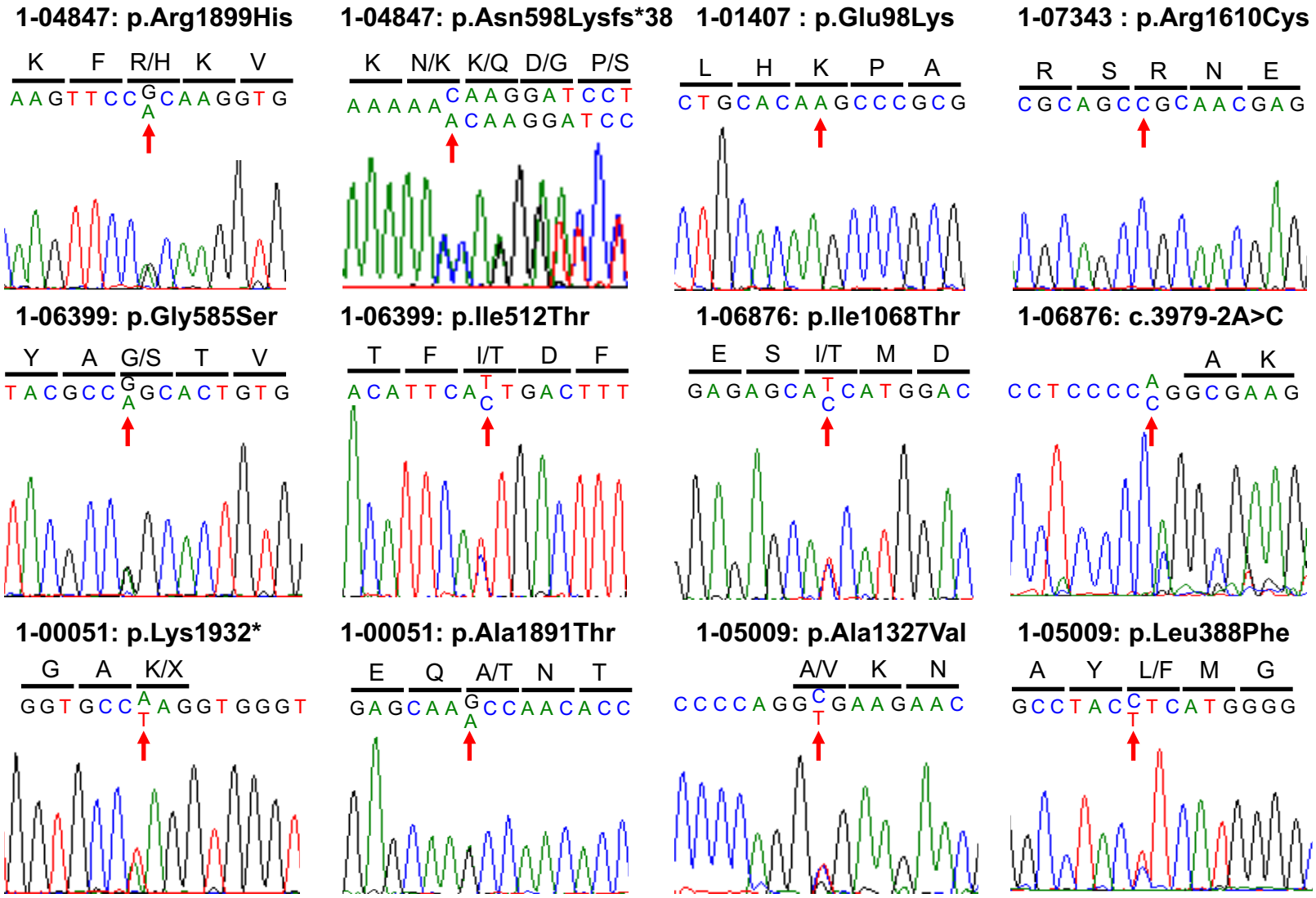
Supplementary Figure 11. Principal component analysis of 2,871 CHD cases and 3,578 parental controls. (a) 7.11% (204 out of 2,871) of all CHD cases best cluster to Ashkenazim (defined by the gray dash lines in the figures). Eleven patients with *GDF1* recessive genotypes are denoted by filled triangles (red for p.Met364Thr mutation carriers and pale green for p.Met364del/p.Cys227* variant carrier, respectively) **(b)** 8.44% (302 out of 3,578) of all autism parental controls best cluster to Ashkenazim (defined by the gray dash lines in the figures).



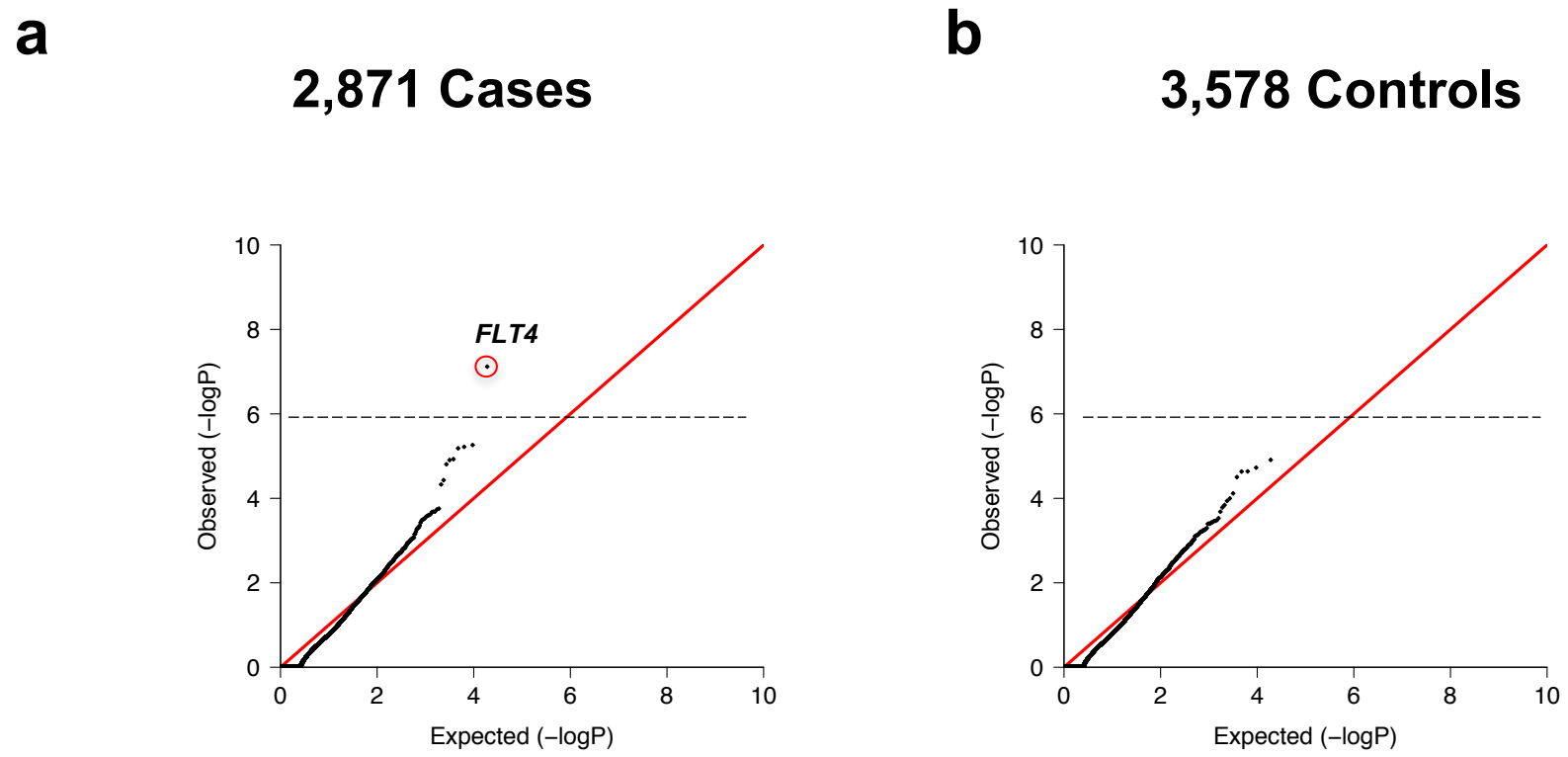
Supplementary Figure 12. Estimation of the mutation age for *GDF1*-p.Met364Thr. Mutation age (the number of generations, x-axis) was estimated using DMLE+2.3 software based on a Bayesian inference approach. A total of 20,000 iterations were performed. The y-axis shows the relative frequency of occurrences for each mutation age estimate. The optimum population growth rate (PGR) and the proportion of sampled chromosome (PSC) for *GDF1*-p.Met364Thr mutation in Ashkenazim were estimated to be 0.22 and 0.00017, respectively (see **Supplementary Note**). The average mutation age is 50 and the 95% confidence interval is between 45 and 63.



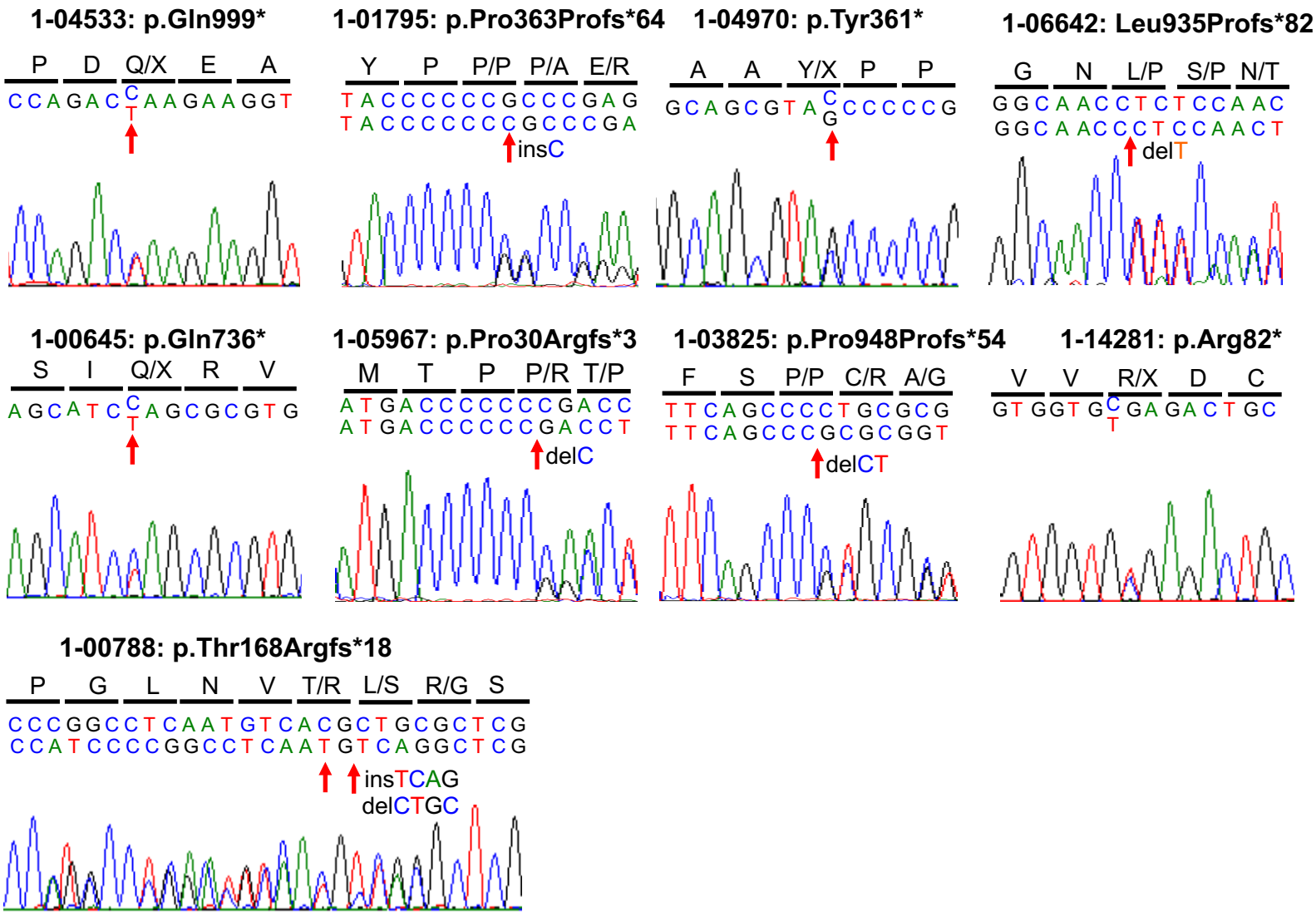
Supplementary Figure 13. Structural representation of the a-helix and flanking residues (residues 314-329; shown in yellow) binding in *GDF1*. (**a**) The predominantly hydrophobic a-helix binds in a shallow cavity on the adjacent monomer (shown as blue surface). The surface of the residues that interact with the a-helix on the adjacent monomer have been colored yellow. (**b**) The two monomers (white and blue) bind with a two-fold inverted symmetry, where the predominantly hydrophobic α-helix (shown in yellow) from one monomer binds in a shallow hydrophobic cavity from the the adjacent monomer. The interacting residues from the cavity have been labelled and shown as sticks. The position of M364 and C336 have been illustrated as red balls. (**c**) Top view of the residues that interact with the α-helix, illustrated as sticks. The floor of the cavity is hydrophobic. (**d**) Top view of the α-helix (yellow) binding to the adjacent monomer. The interacting residues have been illustrated as sticks. Only the α-helix has been drawn for clarity purpose.



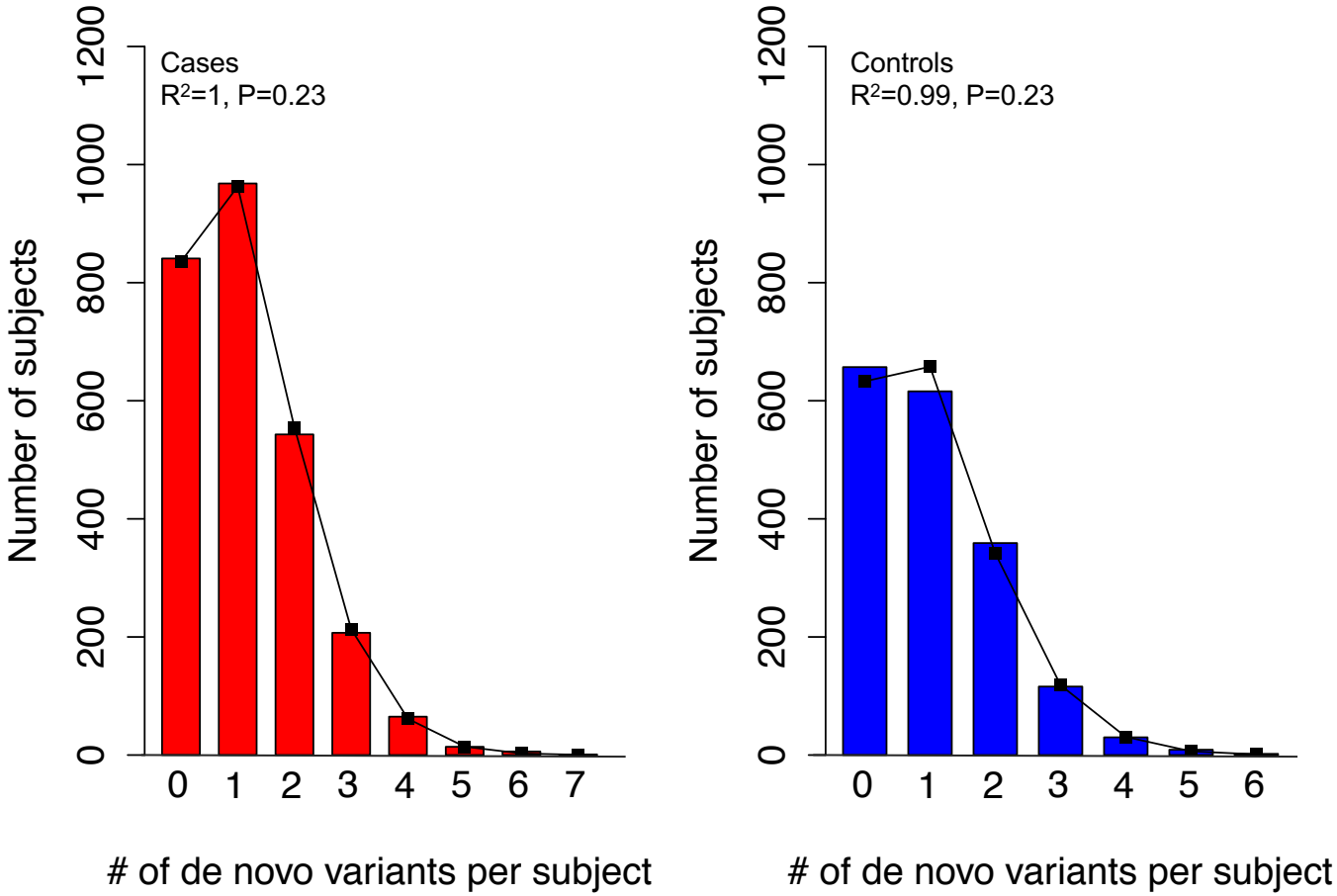
Supplementary Figure 14. Sanger sequence chromatograms for 7 CHD cases carrying recessive genotypes in *MYH6*.



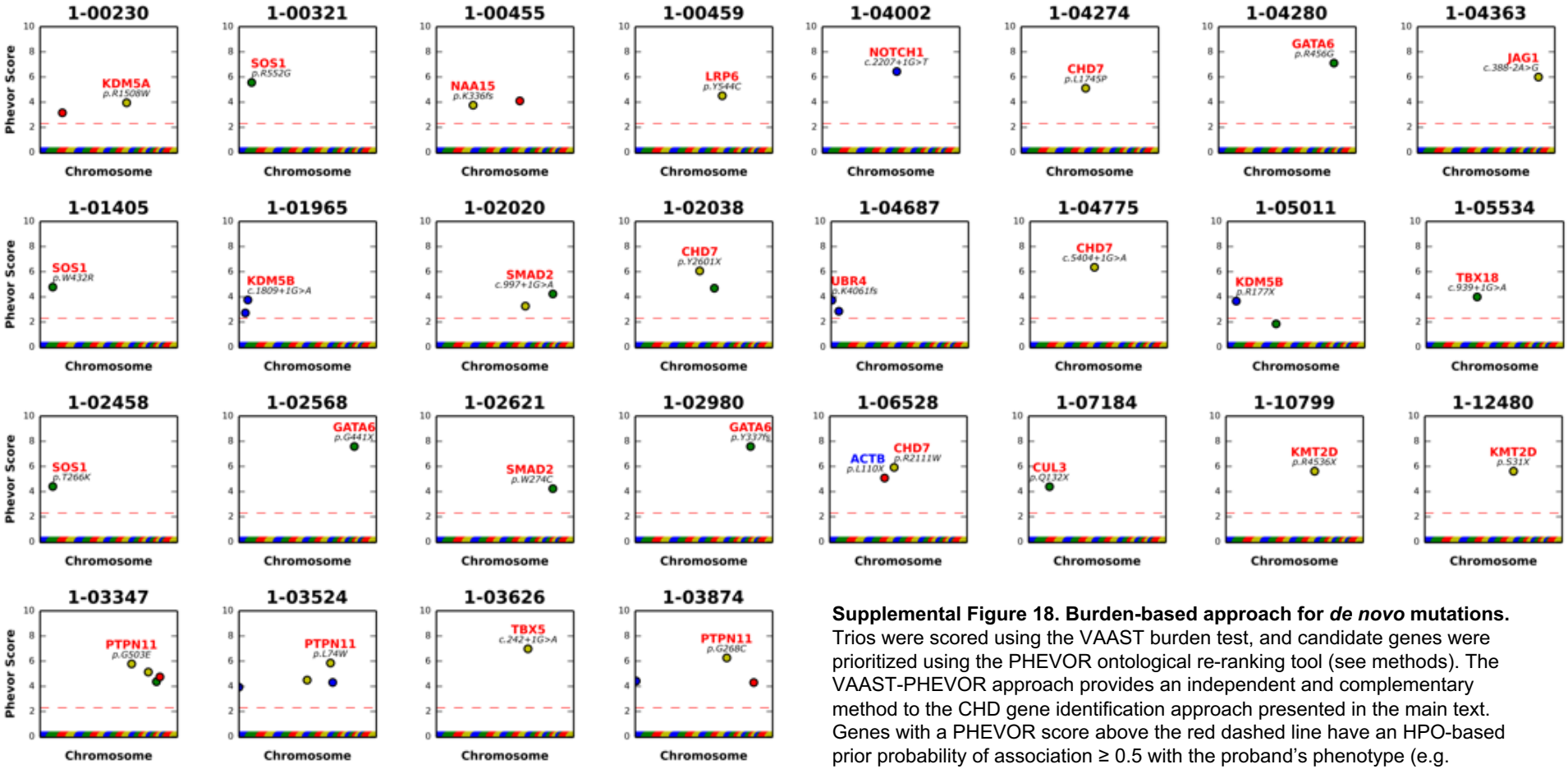
Supplementary Figure 15. *FLT4* surpasses genome-wide significance in gene burden analysis of rare loss-of-function heterozygous variants in all genes (a) Quantile-quantile plot of observed versus expected p-values in cases is shown. The probability of the observed number of variants in each gene was calculated from the total number of loss-of-function (LoF) heterozygous variants and proportion of LoF *de novo* probability using the binomial test. **(b)** Quantile-quantile plot of observed versus expected p-values in controls is shown. *FLT4* is the only gene surpassing genome-wide significance (2.6×10^{-6}, 0.05/(18,989)) in cases. No genes reach genome-wide significance in controls.



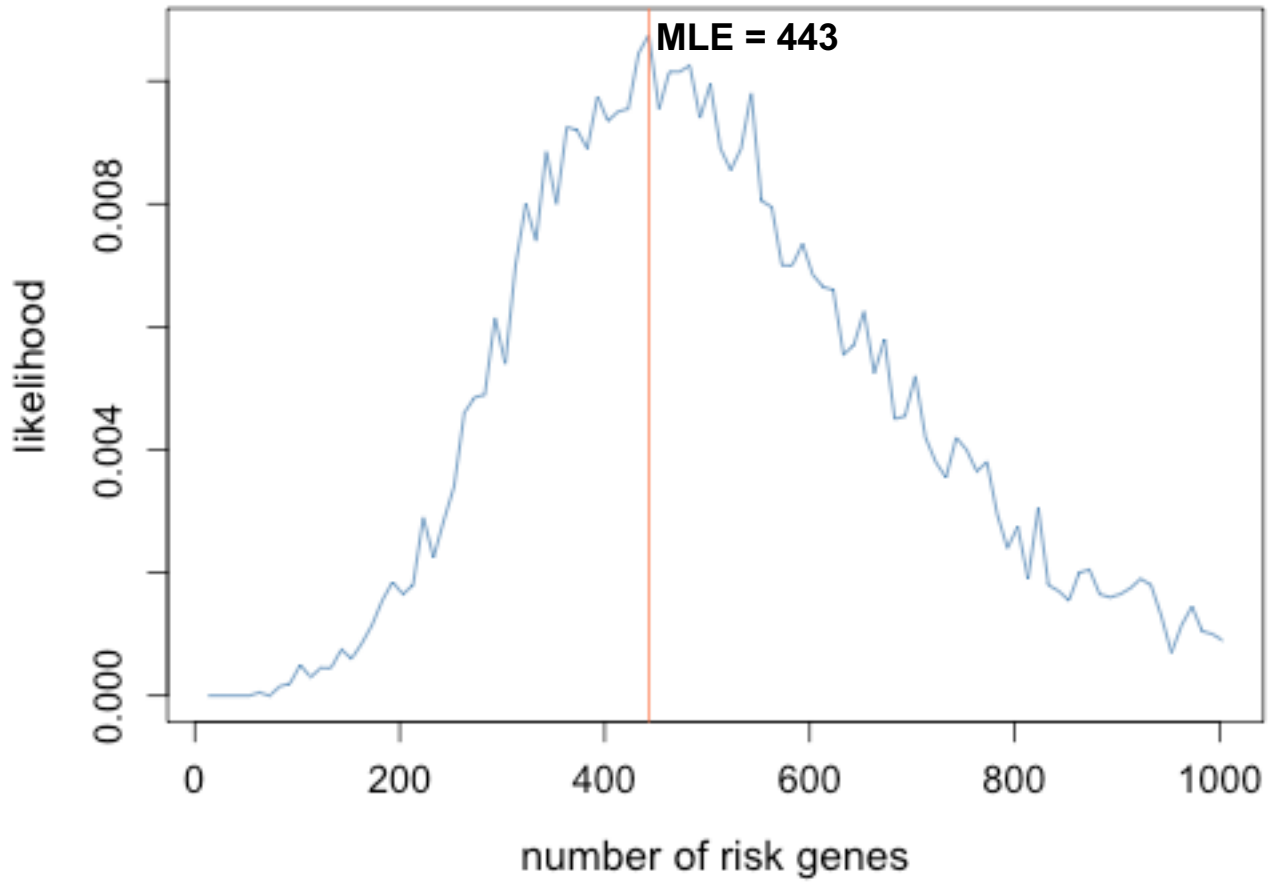
Supplementary Figure 16. Sanger chromatograms for 10 CHD cases carrying *FLT4* loss-of-function heterozygous mutations



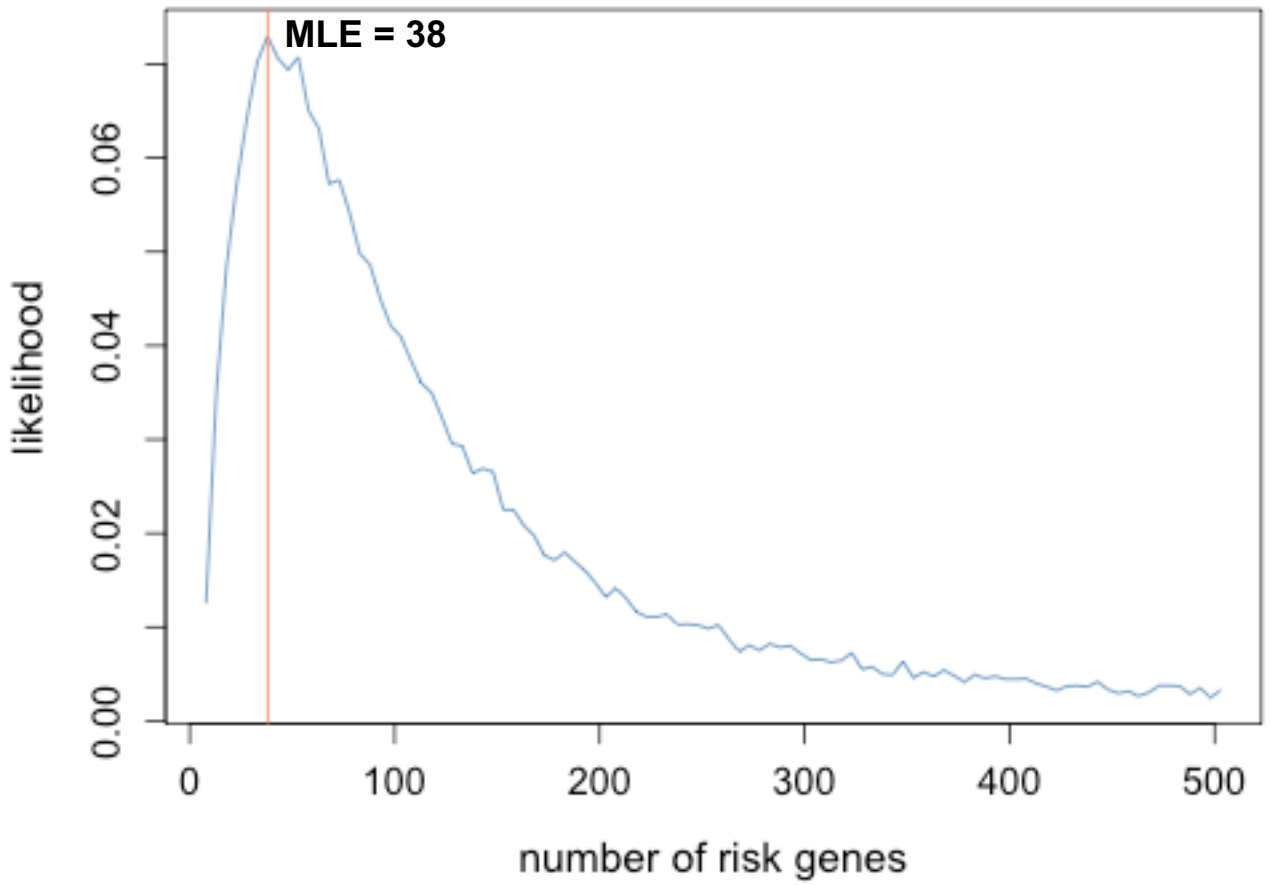
Supplementary Figure 17. *De novo* mutation rate closely approximates Poisson distribution in cases and controls. Observed number of *de novo* mutations per subject (bars) compared to the numbers expected (line) from the Poisson distribution in the case (red) and control (blue) cohorts. ‘p' denotes chi-squared p-value.



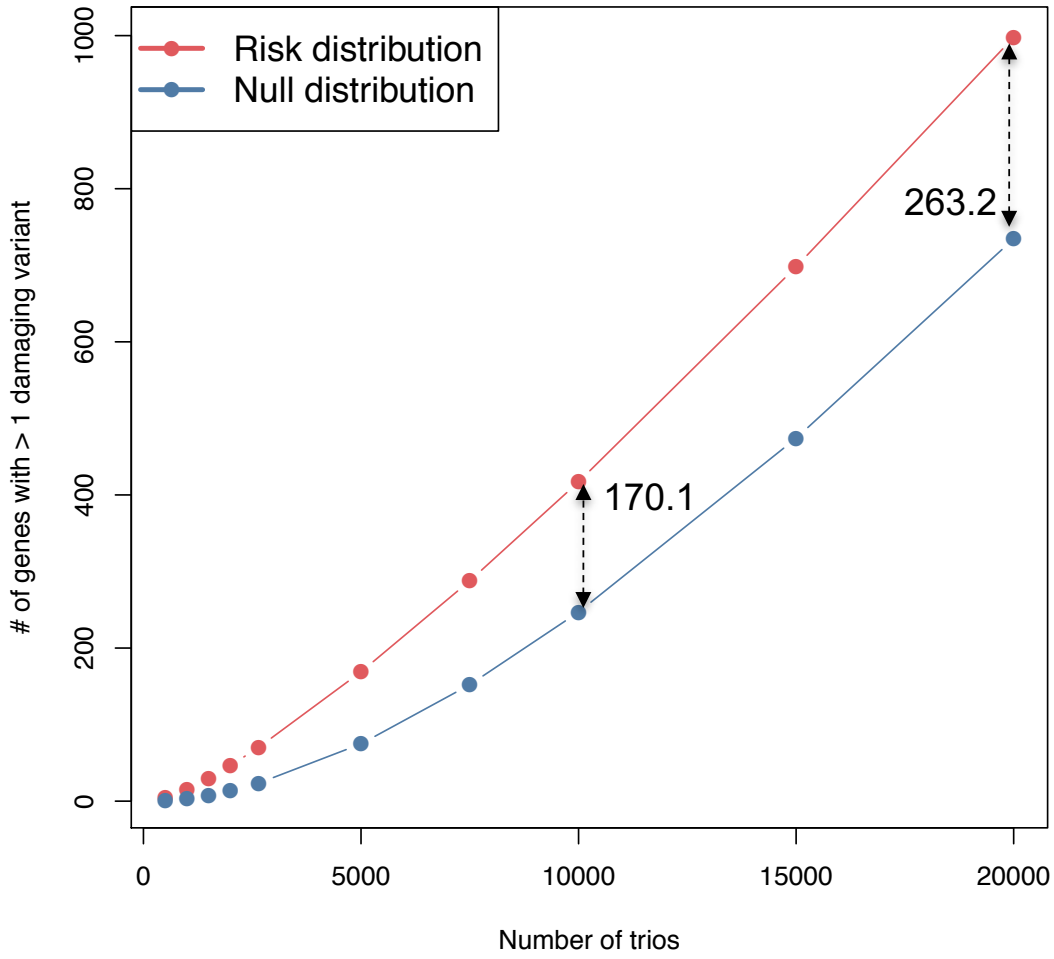
Supplemental Figure 18. Burden-based approach for *de novo* mutations. Trios were scored using the VAAST burden test, and candidate genes were prioritized using the PHEVOR ontological re-ranking tool (see methods). The VAAST-PHEVOR approach provides an independent and complementary method to the CHD gene identification approach presented in the main text. Genes with a PHEVOR score above the red dashed line have an HPO-based prior probability of association ≥ 0.5 with the proband's phenotype (e.g. conotruncal, left outflow obstruction, heterotaxy, etc.) and a VAAST p-value of ≤ 0.005.



Supplementary Figure 19. Estimation of the number of congenital heart disease risk genes. Monte Carlo simulation was performed based on observed damaging *de novo* mutations in the top 25% of high heart expressed genes using 20,000 iterations. Because this likelihood based approach assumes the penetrance of all damaging *de novo* mutations in all risk genes is identical, *CHD7*, *KMT2D*, and *PTPN11* were removed from the simulation due to each having ≥ 9 damaging *de novo* mutations. These three genes were added to the optimum estimate for the number of the risk genes later. We estimate that the number of risk genes via *de novo* events is ~443 (95% confidence interval: [154.1, 731.9]).



Supplementary Figure 20. Estimation of the number of chromatin modification genes contributing to congenital heart disease. Monte Carlo simulation was performed based on observed damaging *de novo* mutations in 546 chromatin modification genes (GO:0016569) using 20,000 iterations. Because this likelihood based approach assumes the penetrance of all damaging *de novo* mutations in all risk genes is identical, *CHD7*, *KMT2D*, and *NSD1* were removed from the simulation due to each having ≥ 5 damaging *de novo* mutations. These three genes were added to the optimum estimate for the number of the risk genes later. We estimate that the number of chromatin modification genes contributing to CHD via *de novo* events is ~38 (95% confidence interval: [7, 69]).



Supplementary Figure 21. Estimation of the number of recurrent genes. The number of trios and the number of genes with more than one damaging *de novo* mutation are specified on the x and y-axis, respectively. The expected number of risk genes was determined by Monte Carlo simulation to be ~443. We modeled the expected rate of damaging *de novo* mutations in cases (red, risk distribution) and controls (blue, null distribution) given an increasing sample size. A total of 10,000 iterations were performed to estimate the number of genes with more than1 damaging *de novo* mutations taking into account of the damaging *de novo* mutation probability. WES of 10,000 and 20,000 trios will yield a 38.4% and 59.4 % saturation for all CHD risk genes.

Supplementary Table 1. Number of studied cases and controls

Category	N
# of case trios	2,645
# of case singletons	226
# of control trios	1,789

N: Number of case trios and singletons, and control trios.

Supplementary Table 2. Demographic characteristics of CHD cases and controls

	CHD Cases	Autism Sibling Controls
Sample size	2,871	1,789
Gender		
Male	1,691 (58.9%)	840 (47.0%)
Female	1,180 (41.1%)	949 (53.0%)
Ethnicity		
European	2,063 (71.9%)	1,408 (78.7%)
African American	189 (6.6%)	79 (4.4%)
East Asian	36 (1.3%)	41 (2.3%)
South Asian	136 (4.7%)	87 (4.9%)
Mexican	280 (9.8%)	125 (7.0%)
Other	167 (5.8%)	49 (2.7%)

The number of samples is shown in each category with the corresponding percentage in parentheses. Ethnicity is determined by principal component analysis compared to HapMap samples using EIGENSTRAT (see **Online Methods**).

Supplementary Table 3a. Cardiac lesions in 2,871 probands

Conotruncal Defects (CTD): 872 probands	
Tetralogy of Fallot	416 (14.5%)
Double-Outlet Right Ventricle	68 (2.4%)
Truncus Arteriosus	68 (2.4%)
Ventricular Septal Defect (conoventicular)	119 (4.1%)
Aortic Arch Abnormalities	71 (2.5%)
Pulmonary Stenosis/Atresia	128 (4.5%)
Other	11 (0.4%)
D-Transposition of the Great Arteries (D-TGA): 251 probands	
Isolated D-Transposition of the Great Arteries	74 (2.6%)
Ventricular Septal Defect	175 (6.1%)
Coarctation	26 (0.9%)
Double-Outlet Right Ventricle	26 (0.9%)
Heterotaxy (HTX): 272 probands	
L-Transposition of the Great Arteries	119 (4.1%)
Double-Outlet Right Ventricle	88 (3.1%)
Atrio-Ventricular Canal Defect	82 (2.9%)
Anomalous Pulmonary Veins	61 (2.1%)
Dextrocardia	78 (2.7%)
Right or Left Atrial Isomerism	77 (2.7%)
Left Ventricular Obstruction (LVO): 797 probands	
Bicuspid Aortic Valve-Aortic Stenosis	259 (9.0%)
Coarctation	285 (9.9%)
Hypoplastic Left Heart Syndrome	293 (10.2%)
Other	99 (3.4%)
Other phenotypes: 679 probands	
Atrial Septal Defect	337 (11.7%)
Ventricular Septal Defect	122 (4.2%)
Atrio-Ventricular Canal Defect	107 (3.7%)
Tricuspid Atresia	47 (1.6%)
Double Inlet Left Ventricle	22 (0.8%)
Anomalous Pulmonary Veins	92 (3.2%)
Coronary Artery Abnormalities	19 (0.7%)
Ebstein's Anomaly	39 (1.4%)

Supplemental Table 3b. Distribution of EA and NDD phenotype in CHD cases

Group	2,871 Probands	2,645 Trios
NDD status obtained	1,844	1,708
NDD positive	577	561
EA status obtained	2,840	2,614
EA positive	998	935
EA and NDD	231	222
EA or NDD	1,344	1,274

NDD: neurodevelopmental deficit; EA: extracardiac congenital anomalies pertaining to structural defect outside of heart; Those probands who did not have obtainable NDD or EA status were excluded from the above tabulation

Supplemental Table 3c. Clinical syndromes diagnosed among probands prior to sequencing

Syndrome	N
Alagille	2
CHARGE	5
Suspect DiGeorge (no del22q11)	8
Suspect Down (no tri 21)	3
Ehlers-Danlos	2
Ellis van Creveld	1
Goldenhar	6
Holt-Oram	6
Kabuki	3
Marfan	1
Noonan	17
VATER/VACTERL	28

Supplementary Table 4. Summary sequencing statistics for the CHD and control cohorts

Category	Cases (Roche V2; N=5,433)	Cases (MedExome; N=2,728)[&]	Controls (Roche V2; N=5,367)
Read length (bp)	74	74/99	50-94
# of reads per sample (M)	75.3	52.6*	99.7
Median coverage at each targeted base (X)	62.7	35.0*	67.0
Mean coverage at each targeted base (X)	74.9	40.7*	79.1
% of all reads that map to target	69.2%	48.7%*	57.6%
% of all bases that map to target	59.0%	38.8%*	49.0%
% of targeted bases read at least 8x	94.9%	96.3%*	94.6%
% of targeted bases read at least 10x	93.7%	94.9%*	93.4%
% of targeted bases read at least 15x	90.0%	89.2%*	89.9%
% Mean error rate	0.3%	0.4%*	0.4%

[&]2,728 case samples were sequenced using the MedExome capture reagent. All other samples were sequenced using the Roche V2 capture reagent. 8X, 10X and 15X were comparable across the platforms. A Wilcox rank sum test was performed to determine whether there is statistically significant difference of sequencing metrics between V2-captured samples and MedExome-captured samples in cases. * denotes Wilcox p-value < 0.05.

Supplemental Table 5. 51 damaging recessive genotypes in known CHD genes in CHD cases

ID	Gene	AA Change	Type	Ethnicity-specific ExAC Freq	OMIM Inheritance Pattern	OMIM Cardiac Phenotype	Patient Diagnosis	Cardiac Phenotype Concordant	Extracardiac	Neuro
Genes with known recessive transmission pattern										
1-13318	*COL1A2*	p.Val416Ile	Hom	0	Monoallelic/Biallelic	Ehlers-Danlos	CTD	No	Yes	n/a
1-00532	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	HTX	Yes	No	Yes
1-00620	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	CTD (TGA)	Yes	Yes	No
1-00862	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	HTX	Yes	Yes	Yes
1-01815	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	CTD (TGA)	Yes	No	n/a
1-02374	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	CTD (TGA)	Yes	No	n/a
1-02504	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	CTD	Yes	No	No
1-02531	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	CTD	Yes	No	No
1-04480	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	HTX	Yes	No	Yes
1-06122	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	CTD	Yes	No	No
1-06789	*GDF1*	p.Met364Thr	Hom	5.6×10^{-3}	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	CTD	Yes	No	No
1-05386	*GDF1*	p.Met364del/p.Cys227*	CmpHet	0/0	Monoallelic/Biallelic	DORV, RAI, TOF, D-TGA	HTX	Yes	No	No
1-00240	*ARMC4*	p.Arg286fs	Hom	6.0×10^{-4}	Biallelic	HTX	HTX	Yes	No	No
1-00129	*ATIC*	c.1227+2T>A/p.Lys426Arg	CmpHet	$0/3.0\times10^{-5}$	Biallelic	ASD	CTD (TGA)	No	Yes	Yes
1-01055	*ATIC*	p.Thr329Ile	Hom	0	Biallelic	ASD	HTX	No	No	No
1-05775	*BBS1*	p.Gly394Asp/p.Arg462His	CmpHet	$2.8\times10^{-3}/0$	Biallelic	Rare, broad spectrum	AVC	N/A	No	No
1-05059	*BBS10*	p.Tyr589*	Hom	0	Biallelic	Rare, HTX	HTX	Yes	Yes	n/a
1-05476	*C5orf42*	p.Asp563Val	Hom	0	Biallelic	Rare, broad spectrum	CTD	N/A	No	n/a
1-02315	*DNAAF1*	p.Gln94fs/p.Leu494fs	CmpHet	0/0	Biallelic	HTX	HTX	Yes	No	No
1-13221	*DNAH11*	p.Gln3166Arg	Hom	0	Biallelic	HTX	CTD	No	No	n/a
1-00803	*DNAH5*	p.Arg4496*/p.Ile1855fs	CmpHet	0/0	Biallelic	HTX	ASD, VSD, cleft MV	No	No	Yes
1-02956	*DNAH5*	p.Arg3539Cys/c.8011-1G>A	CmpHet	0/0	Biallelic	HTX	HTX	Yes	Yes	Yes
1-05359	*DNAH5*	p.Asp3605fs/p.Cys408Arg	CmpHet	$1.0\times10^{-4}/0$	Biallelic	HTX	HTX	Yes	Yes	n/a
1-07573	*DNAI2*	p.Ala54fs	Hom	0	Biallelic	Situs inversus	LVO	No	No	No
1-06387	*DYNC2H1*	p.Leu1931Met/p.Val2069Ala	CmpHet	$2.0\times10^{-4}/0$	Biallelic	TGA	HTX	No	No	Yes
1-00612	*IFT140*	p.Cys768Tyr/p.Arg505*	CmpHet	$1.5\times10^{-5}/0$	Biallelic	Rare, broad range	LVO	N/A	No	No
1-05014	*NPHP3*	p.Gly1299Asp	Hom	8.0×10^{-4}	Biallelic	HTX	LVO	No	No	No
1-06780	*STAMBP*	p.Thr351Ala	Hom	0	Biallelic	ASD/VSD	ASD	Yes	Yes	Yes
1-01687	*DAW1*	p.Ser364Thr	Hom	6.6×10^{-5}	Biallelic	HTX	HTX	Yes	No	No
1-06817	*DAW1*	p.Leu66*/p.Trp372Cys	CmpHet	0/0	Biallelic	HTX	LVO	No	Yes	n/a
1-05375	*MYH10*	p.Glu1738Lys	Hom	9.0×10^{-5}	Biallelic	CTD	CTD	Yes	Yes	No
1-06101	*PKD1L1*	p.Gln2527*/p.Arg284*	CmpHet	$3.9\times10^{-5}/0$	Biallelic	HTX	HTX	Yes	Yes	n/a
1-06808	*LRP1*	p.Val315Ala/p.Arg1734His	CmpHet	$0/1.5\times10^{-5}$	Biallelic	CTD	CTD	Yes	No	Yes
1-01111	*LRP1*	p.Ser2907Leu/p.Ala3634Thr	CmpHet	$1.5\times10^{-5}/1.1\times10^{-3}$	Biallelic	CTD	CTD	Yes	Yes	n/a

ID	Gene	AA Change	Type	Ethnicity-specific ExAC Freq	OMIM Inheritance Pattern	OMIM Cardiac Phenotype	Patient Diagnosis	Cardiac Phenotype Concordant	Extracardiac	Neuro
Genes with known dominant transmission pattern										
1-00688	*CHD7*	c.1717-1G>A	Hom	1.0×10^{-4}	Monoallelic	CTD	CTD	Yes	Yes	n/a
1-00875	*COL1A1*	p.Ala628Thr/p.Arg528His	CmpHet	2.0×10^{-4}/9.0×10^{-5}	Monoallelic	Mitral valve prolapse, Ao dilatation	HTX	No	Yes	No
1-02287	*COL5A2*	p.Arg1070His/p.Pro706Leu	CmpHet	4.0×10^{-4}/8.0×10^{-4}	Monoallelic	Mitral valve prolapse, Ao dilatation	CTD	No	No	n/a
1-04062	*COL5A2*	p.Arg356Gln	Hom	0	Monoallelic	Mitral valve prolapse, Ao dilatation	HTX	No	No	n/a
1-06907	*FBN2*	p.Asp2493Asn	Hom	1.2×10^{-3}	Monoallelic	Mitral valve prolapse, Ao dilatation	CTD	No	Yes	Yes
1-00051	*MYH6*	p.Lys1932*/p.Ala1891Thr	CmpHet	3.0×10^{-5}/0	Monoallelic	ASD, cardiomyopathy	LVO	No	No	Yes
1-01407	*MYH6*	p.Glu98Lys	Hom	3.0×10^{-4}	Monoallelic	ASD, cardiomyopathy	D-TGA	No	Yes	Yes
1-04847	*MYH6*	p.Arg1899His/p.Asn598fs	CmpHet	0/0	Monoallelic	ASD, cardiomyopathy	LVO	No	No	No
1-05009	*MYH6*	p.Ala1327Val/p.Leu388Phe	CmpHet	2.7×10^{-3}/0	Monoallelic	ASD, cardiomyopathy	AVC	No	No	n/a
1-06399	*MYH6*	p.Gly585Ser/p.Ile512Thr	CmpHet	2.0×10^{-4}/3.0×10^{-5}	Monoallelic	ASD, cardiomyopathy	LVO	No	No	n/a
1-06876	*MYH6*	p.Ile1068Thr/c.3979-2A>C	CmpHet	1.5×10^{-5}/2.0×10^{-5}	Monoallelic	ASD, cardiomyopathy	LVO	No	No	No
1-07343	*MYH6*	p.Arg1610Cys	Hom	3.0×10^{-5}	Monoallelic	ASD, cardiomyopathy	ASD/VSD	Yes	No	n/a
1-06058	*NOTCH1*	p.Pro284Leu/p.Thr123Met	CmpHet	2.0×10^{-4}/1.3×10^{-3}	Monoallelic	LVO	LVO	Yes	Yes	Yes
1-05499	*NOTCH1*	p.Phe1773fs/p.Thr123Met	CmpHet	0//1.3×10^{-3}	Monoallelic	LVO	CTD	No	No	Yes
1-01387	*NSD1*	p.Phe621Val/p.Gln990His	CmpHet	0/1.5×10^{-5}	Monoallelic	Cardiomyopathy	PA/IVS	No	No	Yes
1-07598	*TSC2*	p.Val1673Leu	Hom	0	Monoallelic	Cardiac rhabdomyosarcoma, WPW	LVO	No	Yes	No
1-00395	*DGCR2*	p.Thr152Ser	Hom	1.0×10^{-3}	Monoallelic	CTD	D-TGA	No	Yes	Yes

CTD: Conotruncal defects; HTX: Heterotaxy; LVO: Left ventricular hypertrophy: TGA: Transposition of the great arteries; D-TGA: D-transposition of the great arteries ; ASD: Atrial septal defect; TOF: Tetralogy of Fallot; DORV: Double outlet right ventricle; WPW: Wolff-Parkinson-White; PA/IVS: Pulmonary atresia with intact ventricular septum; VSD: ventricular septal defect; RAI: Right atrial isomerism. ‘Hom’ denotes homozygous variants and ‘CmpHet’ denotes compound heterozygous variants.

Supplementary Table 6a. Recessive analysis when separately modeling homozygous and compound heterozygous variants in 2,871 CHD cases

Gene set (# genes)	Observed				Expected	Enrichment	Meta P-value
	# homozygotes	# compound heterozygous	# unique genes	# recessive genotypes	# recessive genotypes		
All genes (18,989)	265	202	391	467	-	-	-
Recessive Known Human (96)	19	10	16	29	6.75	**4.29**	**1.2×10^{-10}**
Recessive Known Mouse or Human (137)	21	13	19	34	11.53	**2.95**	**3.6×10^{-8}**
Known Mouse or Human CHD (253)	28	23	28	51	26.24	**1.94**	**5.9×10^{-6}**

Supplementary Table 6b. Recessive analysis when separately modeling homozygous and compound heterozygous variants in 1,789 controls

Gene set (# genes)	Observed				Expected	Enrichment	Meta P-value
	# homozygotes	# compound heterozygous	# unique genes	# recessive genotypes	# recessive genotypes		
All genes (18,989)	22	131	146	165	-	-	-
Recessive Known Human (96)	0	0	0	0	2.98	0	1
Recessive Known Mouse or Human (137)	1	1	2	2	5.48	0.36	0.97
Known Mouse or Human CHD (253)	2	3	5	5	12.58	0.40	1

The expected number of homozygous variant was determined by multiplying the total number of homozygotes by the sum of damaging *de novo* probabilities of the gene set divided by the sum of all probabilities. The expected number of compound heterozygous variants was determined by multiplying the total number of compound heterozygous variants by the sum of the square of damaging *de novo* probabilities of the gene set divided by the sum of the square of all probabilities. The binomial test was performed to evaluate the enrichment of total RGs. Values in bold are p-values exceeding Bonferroni multiple testing cutoff = $0.05/(3\times2) = 8.3\times10^{-3}$.

Supplementary Table 7a. Damaging recessive genotypes in known CHD genes in 161 probands consanguineous by genotype

Gene set (# genes)	Observed				Expected	Enrichment	P-value
	# homozygotes	# compound heterozygous	# unique genes	# recessive genotypes	# recessive genotypes		
All genes (18,989)	211	7	210	218	-	-	-
Recessive Known Human (96)	12	1	9	13	2.42	**5.37**	**1.3×10^{-6}**
Recessive Known Mouse or Human (137)	14	1	11	15	3.76	**3.99**	**7.3×10^{-6}**
Known Mouse or Human CHD (253)	19	2	16	21	8.31	**2.53**	**1.1×10^{-4}**

Supplementary Table 7b. Damaging recessive genotypes in known CHD genes in 2,710 non-consanguineous probands by genotype

Gene set (# genes)	Observed				Expected	Enrichment	P-value
	# homozygotes	# compound heterozygous	# unique genes	# recessive genotypes	# recessive genotypes		
All genes (18,989)	54	195	202	249	-	-	-
Recessive Known Human (96)	7	9	9	16	4.12	**3.89**	**5.3×10^{-6}**
Recessive Known Mouse or Human (137)	7	12	11	19	7.12	**2.67**	**1.2×10^{-4}**
Known Mouse or Human CHD (253)	9	21	17	30	16.40	**1.83**	**1.1×10^{-3}**

The expected number of recessive genotypes was determined based on fitted values from the polynomial regression model using the damaging *de novo* probabilities. P-values were calculated using the one-tailed binomial probability. The haplotype phase was performed using Beagle v.3.3.2. Consanguinity was defined as a an estimated autozygosity of at least 0.35% of the genome in 2cM or greater homozygous segments (see **Supplementary Note)**.

Supplementary Table 8. Damaging recessive genotypes in known CHD genes after removal of nine of ten probands carrying the *GDF1* founder mutation

Gene set (# genes)	Observed				Expected	Enrichment	P-value
	# homozygotes	# compound heterozygous	# unique genes	# recessive genotypes	# recessive genotypes		
All genes (18,989)	256	202	391	458	-	-	-
Recessive Known Human (96)	10	10	16	20	6.60	**3.03**	**1.6×10^{-5}**
Recessive Known Mouse or Human (137)	12	13	19	25	10.99	**2.27**	**1.6×10^{-4}**
Known Mouse or Human CHD (253)	19	23	28	42	25.01	**1.68**	**8.4×10^{-4}**

The expected number of recessive genotypes was determined based on fitted values from the polynomial regression model using the damaging *de novo* probabilities. P-values were calculated using the one-tailed binomial probability.

Supplementary Table 9. Two genes show an excess of recessive genotypes in cases by the polynomial approach

Gene	Observed # RG	Expected # RG	Enrichment	P-value	HHE Rank
GDF1*	**11**	**0.016**	**692.6**	**3.6×10^{-28}**	**N/A**
MYH6*	**7**	**0.482**	**14.5**	**7.6×10^{-7}**	**100.0**
EPB41L4A	3	0.074	40.7	6.3×10^{-5}	61.0
HEPHL1	4	0.209	19.1	6.7×10^{-5}	10.3
*DAW1**	2	0.013	159.1	7.8×10^{-5}	16.8
CD36&*	2	0.017	119.8	1.4×10^{-4}	84.7
PPEF2	2	0.018	112.7	1.6×10^{-4}	16.3
H6PD	3	0.106	28.4	1.8×10^{-4}	70.4
OR2AG2	1	0.000	3884.3	2.6×10^{-4}	0.0
OR52N2	1	0.000	2422.5	4.1×10^{-4}	0.0
KEL	2	0.034	58.1	5.8×10^{-4}	35.2
OR5J2	1	0.001	1549.1	6.5×10^{-4}	0.0
KRT72	2	0.038	52.1	7.2×10^{-4}	0.0
CAPN14	2	0.038	52.0	7.2×10^{-4}	N/A
F11	2	0.044	45.6	9.3×10^{-4}	14.4
*ATIC**	2	0.045	44.9	9.6×10^{-4}	79.8
FUT2	2	0.045	44.2	9.9×10^{-4}	22.6
MFGE8&*	2	0.050	39.6	1.2×10^{-3}	89.2
PROC	2	0.052	38.4	1.3×10^{-3}	0.0
RIPPLY3	1	0.001	721.4	1.4×10^{-3}	33.7
FAM151B	1	0.001	719.9	1.4×10^{-3}	31.7
HEXA	2	0.054	37.1	1.4×10^{-3}	74.6
OR5AC2	1	0.002	617.5	1.6×10^{-3}	0.0
OR2F2	1	0.002	521.4	1.9×10^{-3}	0.0
TMEM134	1	0.002	467.6	2.1×10^{-3}	N/A

Enrichment and p-value were calculated using the binomial probability adjusted for fitted values from the polynomial model using damaging *de novo* probability. Genes bolded exceed multiple testing cutoff = $0.05/(18{,}989) = 2.6\times10^{-6}$. **(*)** denotes known Human and Mouse CHD genes and **(&)** denotes genes with a p-value ≤ 0.1 in controls. HHE rank represents rank of gene expression at E14.5 in the developing mouse heart. In HHE Rank, 'N/A' refers to mRNA expression in developing murine heart that does not have a human ortholog.

Supplementary Table 10. Recessive damaging genotypes of two genes are significantly enriched in CHD cases when modeling compound heterozygous and homozygous separately

Gene	Total Het RG	P-value Het RG	Total Hom RG	P-value Hom RG	Total RG	P-value total
GDF1	**1**	2.9×10^{-3}	**10**	1.2×10^{-26}	11	$\mathbf{1.5\times10^{-28}}$
MYH6	**5**	4.6×10^{-5}	**2**	8.3×10^{-3}	7	$\mathbf{1.2\times10^{-6}}$
EPB41L4A	2	4.6×10^{-4}	1	3.8×10^{-2}	3	5.2×10^{-5}
HEPHL1	3	3.4×10^{-4}	1	7.6×10^{-2}	4	6.9×10^{-5}
DAW1	1	2.0×10^{-3}	1	9.8×10^{-3}	2	7.0×10^{-5}
CD36	2	4.8×10^{-6}	NA	NA	2	1.2×10^{-4}
PPEF2	NA	NA	2	8.2×10^{-5}	2	1.3×10^{-4}
H6PD	2	1.3×10^{-3}	1	4.9×10^{-2}	3	1.6×10^{-4}
KEL	NA	NA	2	2.3×10^{-4}	2	4.8×10^{-4}
KRT72	NA	NA	2	2.7×10^{-4}	2	$6.0\times10\text{-}^{4}$
CAPN14	1	1.2×10^{-2}	1	2.3×10^{-2}	2	6.0×10^{-4}
MT1HL1	NA	NA	1	6.4×10^{-4}	1	6.5×10^{-4}
F11	NA	NA	2	3.3×10^{-4}	2	7.8×10^{-4}
ATIC	1	1.4×10^{-2}	1	2.6×10^{-2}	2	8.1×10^{-4}
FUT2	1	1.5×10^{-2}	1	2.6×10^{-2}	2	8.4×10^{-4}
MFGE8	NA	NA	2	4.1×10^{-4}	2	1.0×10^{-3}
PROC	1	1.8×10^{-2}	1	2.9×10^{-2}	2	1.1×10^{-3}
HEXA	1	1.9×10^{-2}	1	3.0×10^{-2}	2	1.2×10^{-3}
PATE1	NA	NA	1	1.5×10^{-3}	1	1.5×10^{-3}
OR2AG2	NA	NA	1	1.8×10^{-3}	1	1.9×10^{-3}

Significant enrichment of damaging recessive genotypes were identified in *GDF1* and *MYH*. The number of compound heterozygous recessive genotypes (Total Het RG), homozygous recessive genotypes (Total Hom RG), and total recessive genotypes (Total RG) are provided for each gene. P-values are calculated using the binomial test for compound heterozygous (P-value Het RG), homozygous (P-value Hom RG), or total recessive genotypes (P-value total; significant threshold is 2.6×10^{-6}).

Supplementary Table 11a. *GDF1*-p.Met364Thr variant violates the Hardy Weinberg equilibrium in 204 Ashkenazim cases

GDF1-p.Met364Thr Genotypes	T/T	T/C	C/C	HWE p-value
Observed	193	1	10	5.5×10^{-38}
Expected	183.54	19.92	0.54	

A one-degree of freedom chi-square test with Yate's correction was used to test whether the *GDF1*-p.Met364Thr recessive mutation is deviated from Hardy Weinberg equilibrium.

Supplementary Table 11b. *GDF1*-p.Met364Thr allele frequency in Ashkenazi cases and controls

Mutation	204 Ashkenazi Cases				302 Ashkenazi Autism Parental controls + 926 independent Ashkenazi adults without CHD				OR	P-value
	# T/T	# T/C	# C/C	MAF (%)	# T/T	# T/C	# C/C	MAF (%)		
GDF1-p.Met364Thr	193	1	10	5.15	1,216	12	0	0.49	Inf	2.8×10^{-9}

Ashkenazim was determined by the PCA analysis. T and C alleles are major and minor alleles of *GDF1*-p.Met364Thr respectively. Two-sided Fisher's exact tests which compare the frequencies of the *GDF1*-p.Met364Thr recessive genotype in Ashkenazim suggest a significant difference between 204 Ashkenazi cases and 1,228 Ashkenazi controls.

Supplementary Table 12. Overrepresentation of HTX, D-TGA, CTD, and LVO with damaging recessive genotypes in known Human or Mouse CHD genes

Diagnosis	# Probands	# RGs	Observed # RGs in CHD Genes	Expected # RGs in CHD Genes	Enrichment	P-value
HTX/D-TGA	**523**	**90**	**21**	**4.79**	**4.38**	**8.5×10^{-9}**
LVO	797	98	10	5.73	1.75	0.06
CTD	872	183	14	9.36	1.50	0.09
Other	572	78	4	4.05	0.99	0.58
AVC	107	18	2	0.85	2.35	0.21

Enrichment and p-values for known Human or Mouse CHD genes were calculated using the binomial test adjusted for fitted values from the polynomial model using damaging *de novo* probabilities. Bolded values are p-values which surpass multiple testing cutoff 0.01 = 0.05/5. Abbreviations: D-transposition of the great arteries (D-TGA) and heterotaxy (HTX). Conotruncal defects (CTD). Left ventricular outflow (LVO). Atrioventricular canal defects (AVC). RG: Recessive genotypes; CHD genes include known Human/Mouse CHD genes.

Supplementary Table 13a. Rare loss-of-function heterozygous variants in known CHD genes in 2,871 cases

Gene set (# genes)	Observed		Expected	Enrichment	P-value
	# heterozygous LoF	# unique genes	# heterozygous LoF		
All genes (18,989)	12,332	7,285	-	-	-
Known Dominant Human CHD (115)	66	37	61.95	1.07	0.32
Known Human CHD (212)	198	99	211.52	0.94	0.83

Supplementary Table 13b. Rare loss-of-function heterozygous variants in known CHD genes in 3,578 parental controls

Gene set (# genes)	Observed		Expected	Enrichment	P-value
	# heterozygous LoF	# unique genes	# heterozygous LoF		
All genes (18,989)	13,270	7,532	-	-	-
Known Dominant Human CHD (115)	40	31	62.36	0.65	1
Known Human CHD (212)	179	90	225.19	0.79	1

The expected number of heterozygous loss-of-function (LoF) variants was determined based on the total number of heterozygous LoF variants multiplied by the sum of LoF *de novo* probabilities for the gene set, divided by the sum of LoF *de novo* probabilities for all genes. Enrichment and p-values were calculated using the one-tailed binomial probability adjusted for the LoF *de novo* probability. Values in bold are p-values exceeding Bonferroni multiple testing cutoff = $0.05/(2\times2) = 1.3\times10^{-2}$

Supplementary Table 14. Gene-based burden analysis for all loss-of-function heterozygous variants in 2,871 cases

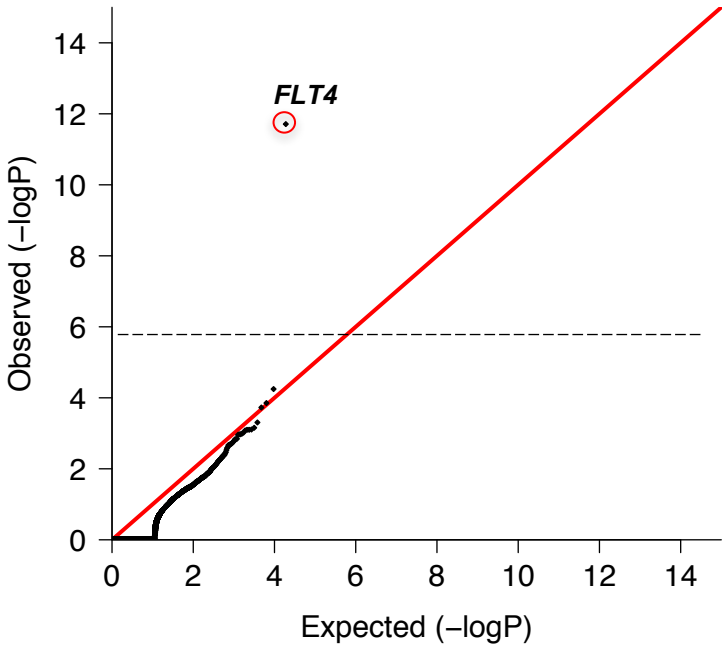
Gene	Observed # LOF heterozygotes	Expected # LOF heterozygotes	Enrichment	P-value	pLI	HHE Rank
FLT4	**8**	**0.515**	**15.5**	**7.6×10^{-8}**	**1.00**	**74.4**
CCDC154	7	0.650	10.8	5.5×10^{-6}	0.31	18.4
SMAD6	8	0.929	8.6	6.0×10^{-6}	0.00	78.3
SLCO1B3	9	1.247	7.2	6.6×10^{-6}	0.00	11.7
C21ORF2[&]	5	0.283	17.7	1.2×10^{-5}	NA	46.7
*NODAL**	4	0.135	29.6	1.2×10^{-5}	0.95	16.4
H1FOO	4	0.143	27.9	1.6×10^{-5}	0.40	10.3
RNF167	5	0.360	13.9	3.7×10^{-5}	0.13	77.7
DNAH14[&]	22	8.200	2.7	4.7×10^{-5}	0.00	N/A
*NOTCH1**	6	0.794	7.6	1.8×10^{-4}	1.00	87.9
SUMO4[&]	3	0.106	28.3	1.8×10^{-4}	0.22	N/A
CCDC168[&]	10	2.418	4.1	2.1×10^{-4}	NA	N/A
SSC5D[&]	7	1.164	6.0	2.1×10^{-4}	0.35	39.8
CYBA	4	0.283	14.1	2.1×10^{-4}	0.00	80.2
TLCD2	4	0.293	13.7	2.4×10^{-4}	NA	38.0
SYCE1L	4	0.295	13.6	2.5×10^{-4}	0.07	28.0
IGFLR1[&]	5	0.546	9.2	2.6×10^{-4}	0.00	74.5
OR51T1	4	0.301	13.3	2.7×10^{-4}	0.02	0.0
SRRD	5	0.557	9.0	2.8×10^{-4}	0.00	51.2
MUC1	3	0.126	23.7	3.1×10^{-4}	0.73	21.8
VWDE[&]	9	2.085	4.3	3.1×10^{-4}	NA	0.0
C10ORF90	6	0.892	6.7	3.2×10^{-4}	NA	28.5
TMEM232[&]	6	0.902	6.7	3.4×10^{-4}	NA	13.3
UNC5D[&]	4	0.330	12.1	3.8×10^{-4}	0.85	24.7
SIAE	6	0.945	6.4	4.4×10^{-4}	0.00	48.2

One-tailed binomial test was performed to test the enrichment of loss-of-function (LoF) heterozygous variants by comparing observed to expected. The expected number of LoF heterozygous variants within each gene was calculated by multiplying the total number of LoF heterozygous variants by the ratio between the LoF *de novo* probability for a gene and the sum of LoF *de novo* probabilities for all genes. Enrichment was calculated as the number of observed LoF heterozygous variants divided by the expected number. (*) Denotes known Human CHD genes and (&) denote genes with a p-value ≤ 0.05 in controls. The genome-wide significance cutoff is $2.6\text{x}10^{-6}$ (0.05/18,989).

a

Gene	Observed # LoF heterozygotes	Expected # LoF heterozygotes	Enrichment	P-value	pLI	HHE Rank
FLT4	**7**	**0.074**	**95.2**	**1.9×10^{-12}**	**1.00**	**74.4**
SMTN	3	0.071	42.0	5.7×10^{-5}	0.76	91.5
ZNF79	3	0.097	30.9	1.4×10^{-4}	0.02	N/A
RHBG	3	0.107	28.1	1.9×10^{-4}	0.00	36.3
OR10AG1	2	0.032	62.8	4.9×10^{-4}	0.00	10.3
ZNHIT3	2	0.038	52.8	7.0×10^{-4}	0.00	51.9
TMEM33	2	0.040	49.8	7.8×10^{-4}	0.23	76.3
ZRANB2	2	0.041	49.4	7.9×10^{-4}	0.99	89.1
SSC5D	3	0.177	17.0	8.0×10^{-4}	0.35	39.8
OR51L1	2	0.041	48.4	8.3×10^{-4}	0.02	N/A
TNPO2	2	0.044	45.0	9.6×10^{-4}	1.00	91.9
ANKLE1	3	0.192	15.6	1.0×10^{-3}	0.00	46.6
SLCO1B3	3	0.195	15.4	1.1×10^{-3}	0.00	11.7
UNC5D	2	0.047	42.5	1.1×10^{-3}	0.85	24.7
PROKR2	2	0.048	41.8	1.1×10^{-3}	0.00	25.9
IFNGR1	2	0.055	36.7	1.4×10^{-3}	0.33	57.3
CAPN11	3	0.218	13.8	1.5×10^{-3}	0.00	15.6
ATP2A2	2	0.056	35.9	1.5×10^{-3}	1.00	99.9
GSTK1	2	0.059	33.9	1.7×10^{-3}	0.00	47.5
PLEKHA6	2	0.061	32.9	1.8×10^{-3}	0.98	80.9

b



Supplementary Table 15. Rare loss-of-function heterozygous mutations in *FLT4* are significantly enriched in Tetralogy of Fallot patients. (a) Top 20 genes that have the most significant enrichment from the one-tailed binomial test was shown. The expected number of loss-of-function (LoF) heterozygous variants within each gene was calculated by multiplying the total number of LoF heterozygous variants by the ratio between LoF *de novo* probability for a gene and the sum of LoF *de novo* probability for all genes. Enrichment was calculated as the number of observed LoF heterozygous variants divided by the expected number. (b) The quantile-quantile plot of expected versus observed p-values for all LoF heterozygous variants in 426 Tetralogy of Fallot patients.

Supplementary Table 16. *De novo* enrichment analysis in 2,645 cases *versus* 1,789 controls

	Cases, N=2,645						Controls, N=1,789					
	Observed		Expected		Enrichment	*P*	Observed		Expected		Enrichment	*P*
	N	Rate	N	Rate			N	Rate	N	Rate		
All Genes							**All Genes**					
Total	2990	1.130	2956.2	1.118	1.0	0.27	Total	1830	1.023	1949.9	1.09	0.9	1.0
Syn	701	0.265	839.2	0.317	0.8	1.00	Syn	484	0.271	549.6	0.307	0.9	1.0
T-Mis	1462	0.553	1508.7	0.570	1.0	0.8	T-Mis	974	0.544	993.3	0.555	1.0	0.7
D-Mis	453	0.171	348.2	0.132	1.3	$4.4x10^{-8}$	D-Mis	222	0.124	232.8	0.13	1.0	0.8
LoF	374	0.141	260.1	0.098	1.4	$2.0x10^{-11}$	LoF	150	0.084	174.3	0.097	0.9	1.0
Damaging	827	0.313	608.3	0.23	1.4	$2.4x10^{-17}$	Damaging	372	0.208	407.1	0.228	0.9	1.0
HHE Genes							**HHE Genes**					
Total	1043	0.394	835.8	0.316	1.3	$2.8x10^{-12}$	Total	533	0.298	555.0	0.310	1.0	0.8
Syn	210	0.079	233.1	0.088	0.9	0.94	Syn	135	0.075	153.8	0.086	0.9	0.9
T-Mis	443	0.167	417.8	0.158	1.1	0.11	T-Mis	272	0.152	276.9	0.155	1.0	0.6
D-Mis	199	0.075	108.1	0.041	1.8	$3.2x10^{-15}$	D-Mis	74	0.041	72.5	0.041	1.0	0.4
LoF	190	0.072	76.9	0.029	2.5	$1.4x10^{-27}$	LoF	52	0.029	51.7	0.029	1.0	0.5
Damaging	389	0.147	184.9	0.070	2.1	$3.9x10^{-39}$	Damaging	126	0.070	124.2	0.069	1.0	0.4
Intolerant Genes							**Intolerant Genes**					
Total	710	0.27	557.5	0.21	1.3	$3.2x10^{-10}$	Total	359	0.20	370.3	0.21	1	0.73
Syn	139	0.05	152.9	0.06	0.9	0.88	Syn	86	0.05	101.0	0.06	0.9	0.94
T-Mis	299	0.11	275.4	0.10	1.1	0.08	T-Mis	195	0.11	182.5	0.10	1.1	0.19
D-Mis	140	0.05	75.5	0.03	1.9	$2.1x10^{-11}$	D-Mis	49	0.03	50.8	0.03	1	0.62
LoF	132	0.05	53.7	0.02	2.5	$1.6x10^{-19}$	LoF	29	0.02	36.1	0.02	0.8	0.90
Damaging	272	0.10	129.1	0.05	2.1	$5.1x10^{-28}$	Damaging	78	0.04	86.8	0.05	0.9	0.84
Intolerant + HHE Genes							**Intolerant + HHE Genes**					
Total	449	0.17	325.2	0.12	1.4	$4.9x10^{-11}$	Total	196	0.11	216.6	0.12	0.9	0.93
Syn	73	0.03	88.3	0.03	0.8	0.96	Syn	51	0.03	58.5	0.03	0.9	0.85
T-Mis	174	0.07	161.7	0.06	1.1	0.18	T-Mis	99	0.06	107.4	0.06	0.9	0.80
D-Mis	94	0.04	43.4	0.02	2.2	$2.1x10^{-11}$	D-Mis	31	0.02	29.3	0.02	1.1	0.40
LoF	108	0.04	31.8	0.01	3.4	$2.7x10^{-26}$	LoF	15	0.01	21.4	0.01	0.7	0.94
Damaging	202	0.08	75.2	0.03	2.7	$1.1x10^{-33}$	Damaging	46	0.03	50.6	0.03	0.9	0.76
LHE Genes							**LHE Genes**					
Total	1811	0.685	1955.4	0.739	0.9	1.0	Total	1214	0.679	1286.1	0.719	0.9	1.0
Syn	463	0.175	559.3	0.211	0.8	1.0	Syn	326	0.182	365.0	0.204	0.9	1.0
T-Mis	935	0.353	997.8	0.377	0.9	1.0	T-Mis	653	0.365	654.9	0.366	1.0	0.5
D-Mis	242	0.092	229.2	0.087	1.1	0.2	D-Mis	145	0.081	153.0	0.086	0.9	0.8
LoF	170	0.064	169.1	0.064	1.0	0.5	LoF	90	0.050	113.2	0.063	0.8	1.0
Damaging	412	0.156	398.3	0.151	1.0	0.3	Damaging	235	0.131	266.2	0.149	0.9	1.0

N: the number of *de novo* mutations; Rate: the number of *de novo* mutations divided by the number of individuals in the cohort; Enrichment: ratio of observed to expected numbers of mutations; Intolerant genes: Genes with a pLI score ≥ 0.99. 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); D-Mis: Damaging missense mutations as predicted by MetaSVM; T-Mis: Tolerated missense mutations as predicted by MetaSVM; Damaging: D-Mis+LoF.

Supplementary Table 17. 46 chromatin modification genes with at least one loss-of-function or deleterious missense *de novo* mutation (GO:0016569)

Gene	# LoF	# D-Mis	HHE Rank	pLI	p
*CHD7**	11	3	93.4	1.00	$1.6x10^{-20}$
*KMT2D**	12	4	96.8	1.00	$2.1x10^{-20}$
NSD1	3	2	94.8	1.00	$1.0x10^{-5}$
KDM5B	3	0	86.0	0.00	$2.9x10^{-5}$
POGZ	2	1	83.8	1	$2.5x10^{-5}$
CTNNB1	2	0	99.0	1.00	$2.1x10^{-4}$
CHD4	0	3	99.1	1.00	$2.3x10^{-3}$
KDM5A	0	2	86.5	1.00	$4.8x10^{-3}$
UBE2B	0	1	94.6	0.32	$8.9x10^{-3}$
BAG6	1	0	97.6	1.00	$9.2x10^{-3}$
BRMS1L	1	0	80.7	0.37	0.01
TP53	1	0	88.9	0.91	0.01
RUVBL1	1	0	61.0	1.00	0.01
WHSC1	1	1	89.0	1.00	0.01
PRKAA2	1	0	91.8	0.00	0.02
TRIM37	0	1	71.0	0.17	0.02
PRDM6	1	0	40.8	NA	0.02
USP16	1	0	76.4	0.01	0.02
WAC	1	0	87.3	1.00	0.02
HDAC7	1	0	86.2	1.00	0.02
HIRA	1	0	73.7	1.00	0.03
KANSL1	1	0	84.8	1.00	0.03
TLK2	1	0	71.6	1.00	0.03
ASXL1	1	0	85.5	0.00	0.03
CTR9	1	0	81.9	1.00	0.03
RNF20	1	0	83.4	0.00	0.04
KDM6B	1	0	94.7	1.00	0.04
HR	0	1	24.3	0.00	0.04
CUL4B	0	1	68.7	1.00	0.04
KAT6B	1	0	81.7	1.00	0.04
ARID1B	1	0	83.2	1.00	0.05
KMT2A	1	1	86.1	1.00	0.05
PRKD1	0	1	51.5	0.00	0.05
KAT6A	1	0	88.9	1.00	0.05
NOS1	0	1	24.2	1.00	0.06
UHRF2	0	1	79.0	1.00	0.06
SMAD4	0	1	78.6	0.97	0.06
KMT2C	1	1	80.0	1.00	0.06
EP300	1	0	88.1	1.00	0.06
HLTF	0	1	59.3	0.00	0.07
EP400	1	0	84.9	1.00	0.08
ASH1L	1	0	87.0	1.00	0.08
GATA3	0	1	34.9	0.88	0.09
EHMT1	0	1	84.8	1.00	0.09
SMARCA1	0	1	77.4	1.00	0.11
ATRX	0	1	95.4	1.00	0.20

**GO Term: Chromatin Modification (GO:0016569), including 546 genes. The Bonferroni corrected threshold for genome-wide significance = $P < 8.8x10^{-7}$.

Supplementary Table 18. ***De novo*** **enrichment analysis in all chromatin, HHE chromatin, and LoF-intolerant chromatin genes in cases and controls (GO:0016569)**

	Cases, N=2,645						Controls, N=1,789					
	Observed		Expected		Enrichment	*p*	Observed		Expected		Enrichment	*p*
	N	Rate	N	Rate			N	Rate	N	Rate		
All Chromatin genes (N = 546)							**All Chromatin genes (N = 546)**					
Total	180	0.068	117.3	0.044	1.5	4.8×10^{-8}	Total	65	0.036	78	0.044	0.8	0.94
Syn	21	0.008	32.2	0.012	0.7	0.99	Syn	12	0.007	21.3	0.012	0.6	0.99
T-mis	70	0.026	56.7	0.021	1.2	4.8×10^{-2}	T-mis	35	0.020	37.7	0.021	0.9	0.69
D-Mis	31	0.012	17.2	0.007	1.8	1.7×10^{-3}	D-Mis	12	0.007	11.5	0.006	1.0	0.49
LoF	58	0.022	11.2	0.004	5.2	5.0×10^{-23}	LoF	6	0.003	7.5	0.004	0.8	0.76
Damaging	89	0.034	28.4	0.011	3.1	8.7×10^{-20}	Damaging	18	0.010	19.1	0.011	0.9	0.63
Intolerant Chromatin genes (N = 187)							**Intolerant Chromatin genes (N = 187)**					
Total	129	0.049	65.6	0.025	2.0	3.1×10^{-12}	Total	40	0.022	43.7	0.024	0.9	0.73
Syn	15	0.006	17.6	0.007	0.9	0.77	Syn	8	0.004	11.7	0.007	0.7	0.90
T-mis	42	0.016	30.6	0.012	1.4	2.9×10^{-2}	T-mis	25	0.014	20.4	0.011	1.2	0.18
D-Mis	24	0.009	10.8	0.004	2.2	3.8×10^{-4}	D-Mis	4	0.002	7.3	0.004	0.5	0.93
LoF	48	0.018	6.5	0.002	7.4	1.7×10^{-25}	LoF	3	0.002	4.4	0.002	0.7	0.81
Damaging	72	0.027	17.4	0.007	4.2	1.1×10^{-22}	Damaging	7	0.004	11.7	0.007	0.6	0.95
HHE Chromatin genes (N = 257)							**HHE Chromatin genes (N = 257)**					
Total	141	0.053	69.8	0.026	2.0	4.7×10^{-14}	Total	51	0.029	46.4	0.026	1.1	0.27
Syn	14	0.005	18.9	0.007	0.7	0.9	Syn	10	0.006	12.5	0.007	0.8	0.80
T-mis	49	0.019	32.9	0.012	1.5	5.1×10^{-3}	T-mis	28	0.016	21.9	0.012	1.3	0.12
D-Mis	24	0.009	11.2	0.004	2.1	6.1×10^{-4}	D-Mis	8	0.004	7.5	0.004	1.06	0.48
LoF	54	0.020	6.7	0.003	8.0	3.2×10^{-30}	LoF	5	0.003	4.5	0.003	1.1	0.47
Damaging	78	0.029	18	0.007	4.4	1.2×10^{-25}	Damaging	13	0.007	12	0.007	1.08	0.43
Intolerant + HHE Chromatin genes (N = 134)							**Intolerant + HHE Chromatin genes (N = 134)**					
Total	111	0.042	51.5	0.019	2.2	4.6×10^{-13}	Total	36	0.020	34.3	0.019	1.1	0.41
Syn	11	0.004	13.8	0.005	0.8	0.81	Syn	8	0.004	9.1	0.005	0.9	0.69
T-Mis	33	0.012	23.7	0.009	1.4	4.1×10^{-2}	T-Mis	22	0.012	15.7	0.009	1.4	0.08
D-Mis	22	0.008	8.9	0.003	2.5	1.4×10^{-4}	D-Mis	4	0.002	6	0.003	0.7	0.85
LoF	45	0.017	5.1	0.002	8.8	3.9×10^{-27}	LoF	2	0.001	3.4	0.002	0.6	0.86
Damaging	67	0.025	14	0.005	4.8	1.5×10^{-24}	Damaging	6	0.003	9.4	0.005	0.6	0.91

Supplementary Table 19a. Genes with multiple *de novo* mutations in 2,645 cases (observed vs. expected)

	Observed	Expected	Enrichment	P-value
Total	357	317.0	1.1	2.2×10^{-3}
Syn	24	23.0	1.0	0.45
Missense	172	145.2	1.2	5.2×10^{-3}
Damaging	66	41.3	1.6	3.1×10^{-5}
D-Mis	35	20.4	1.7	5.8×10^{-4}
LoF	19	7.0	2.7	4.0×10^{-5}

Observed and expected numbers of genes with > 1 *de novo* mutation in each variant category. 1 million simulations were performed based on the per-base probability of mutations in each category to determine the likelihood and the expected number of genes with > 1 *de novo* mutation.

Supplementary Table 19b. HHE genes with multiple *de novo* mutations in 2,645 cases (observed vs. expected)

	Observed	Expected	Enrichment	P-value
Total	149	106.0	1.4	5.0×10^{-6}
Syn	11	8.2	1.3	0.19
Missense	70	50.1	1.4	2.1×10^{-3}
Damaging	45	16.0	2.8	$< 10^{-6}$
D-Mis	21	8.0	2.6	2.5×10^{-5}
LoF	17	2.8	6.1	$< 10^{-6}$

Observed and expected numbers of HHE genes with > 1 *de novo* mutation in each variant category. 1 million simulations were performed based on the per-base probability of mutation in each category to determine the likelihood and the expected number of **HHE** genes with > 1 *de novo* mutation.

Supplementary Table 19c. LoF-intolerant genes with multiple *de novo* mutations in 2,645 cases (observed vs. expected)

	Observed	Expected	Enrichment	P-value
Total	133	86.8	1.5	$< 10^{-6}$
Syn	9	6.3	1.4	0.18
Missense	63	41.0	1.5	3.6×10^{-4}
Damaging	38	14.1	2.7	$< 10^{-6}$
D-Mis	20	7.2	2.8	2.3×10^{-5}
LoF	13	2.2	5.9	$< 10^{-6}$

Observed and expected numbers of LoF-intolerant genes (pLI ≥ 0.99) with > 1 *de novo* mutation in each variant category. 1 million simulations were performed based on the per-base probability of mutation in each category to determine the likelihood and the expected number of intolerant genes with > 1 *de novo* mutation.

Supplementary Table 19d. HHE + LoF-intolerant genes with multiple *de novo* mutations in 2,645 cases (observed vs. expected)

	Observed	Expected	Enrichment	P-value
Total	88	52.4	1.7	$< 10^{-6}$
Syn	6	3.8	1.6	0.18
Missense	39	25.0	1.6	3.3×10^{-3}
Damaging	31	8.6	3.6	$< 10^{-6}$
D-Mis	15	4.3	3.5	1.7×10^{-5}
LoF	12	1.4	8.6	$< 10^{-6}$

Observed and expected numbers of HHE + LoF-intolerant genes with > 1 *de novo* mutation in each variant category. 1 million simulations were performed based on the per-base probability of mutation in each category to determine the likelihood and the expected number of HHE + intolerant genes with > 1 *de novo* mutation.

Supplementary Table 20. 66 genes with more than one damaging *de novo* mutation in cases

Gene	# LoF	# D-Mis	P-value	HHE Rank	pLI	Gene	# LoF	# D-Mis	P-value	HHE Rank	pLI
*CHD7**	11	3	$1.6x10^{-20}$	93.4	1.00	*KRT13*	0	2	$1.2x10^{-3}$	15.5	0.00
*KMT2D**	12	4	$2.1x10^{-20}$	96.8	1.00	*TBX5*	1	1	$1.2x10^{-3}$	95.7	0.99
*PTPN11**	0	9	$4.6x10^{-17}$	94.2	1.00	*U2SURP*	1	1	$1.2x10^{-3}$	88.0	1.00
*GATA6**	3	1	$2.4x10^{-7}$	94.8	NA	*ZEB2*	1	1	$1.4x10^{-3}$	81.6	1.00
*RBFOX2**	3	0	$3.4x10^{-7}$	97.8	0.99	*NGFR*	0	2	$1.5x10^{-3}$	36.2	0.24
NSD1	3	2	$1.0x10^{-5}$	94.8	1.00	*CDK13*	1	1	$1.5x10^{-3}$	80.0	0.75
POGZ	2	1	$2.5x10^{-5}$	83.8	1.00	*DTNA*	1	1	$1.5x10^{-3}$	47.3	0.92
SOS1	0	3	$2.6x10^{-5}$	67.9	1.00	*MINK1*	0	2	$1.6x10^{-3}$	88.1	1.00
GPBAR1	1	1	$2.6x10^{-5}$	19.9	0.00	*PRRC2B*	2	0	$2.0x10^{-3}$	98.2	1.00
RAF1	0	3	$2.6x10^{-5}$	91.4	1.00	*NR6A1*	0	2	$2.2x10^{-3}$	36.8	0.99
NOTCH1	2	3	$2.7x10^{-5}$	87.9	1.00	*CHD4*	0	3	$2.3x10^{-3}$	99.1	1.00
KDM5B	3	0	$2.9x10^{-5}$	86.0	0.00	*DSCAML1*	0	2	$2.5x10^{-3}$	9.5	1.00
SMAD2	1	2	$5.5x10^{-5}$	74.7	0.99	*FRYL*	2	0	$2.8x10^{-3}$	84.4	1.00
PTEN	2	0	$6.0x10^{-5}$	77.9	0.98	*PPL*	1	1	$2.8x10^{-3}$	47.1	0.00
RPL5	2	0	$6.2x10^{-5}$	97.9	0.99	*ANK3*	2	0	$3.1x10^{-3}$	95.0	1.00
CYP21A2	1	1	$7.4x10^{-5}$	14.0	0.67	*KIAA0196*	0	2	$3.9x10^{-3}$	79.7	0.00
MYRF	0	2	$9.7x10^{-5}$	85.6	1.00	*CAD*	0	3	$4.3x10^{-3}$	85.9	1.00
ACTB	1	2	$1.2x10^{-4}$	99.9	0.94	*FBN1*	0	3	$4.5x10^{-3}$	93.1	1.00
RIT1	0	2	$1.2x10^{-4}$	61.4	0.67	*KDM5A*	0	2	$4.8x10^{-3}$	86.5	1.00
ELN	2	0	$1.3x10^{-4}$	79.8	0.00	*TSC1*	1	1	$7.5x10^{-3}$	73.8	1.00
SAMD11	2	0	$1.8x10^{-4}$	NA	NA	*GANAB*	1	1	$9.3x10^{-3}$	93.7	1.00
CTNNB1	2	0	$2.1x10^{-4}$	99.0	1.00	*CLUH*	0	2	0.01	95.9	1.00
NAA15	2	0	$2.2x10^{-4}$	96.7	1.00	*JAG1*	1	1	0.01	76.8	1.00
PYGL	0	3	$2.7x10^{-4}$	82.6	0.00	*CACNA1A*	1	2	0.01	48.7	1.00
AKAP12	2	0	$3.8x10^{-4}$	94.2	0.37	*WHSC1*	1	1	0.01	89.0	1.00
MYH6	1	3	$4.2x10^{-4}$	100.0	0.00	*MYO7B*	0	2	0.02	20.7	0.00
LZTR1	0	2	$4.8x10^{-4}$	84.1	0.00	*SCN10A*	0	2	0.02	43.9	0.00
FLT4	2	0	$5.2x10^{-4}$	74.4	1.00	*NALCN*	0	2	0.03	21.1	0.00
RABGAP1L	1	1	$5.5x10^{-4}$	78.4	0.09	*RYR3*	1	2	0.03	76.7	1.00
BRAF	0	2	$6.1x10^{-4}$	65.8	1.00	*FBN3*	0	2	0.04	N/A	N/A
DGUOK	0	2	$7.2x10^{-4}$	55.2	0.37	*KMT2A*	1	1	0.05	86.1	1.00
ITSN2	0	2	$8.2x10^{-4}$	77.3	0.02	*KMT2C*	1	1	0.06	80.0	1.00
AHNAK	1	1	$1.1x10^{-3}$	95.7	0.34	*LRP1*	0	2	0.16	92.9	1.00

Loss-of-function (LoF: non-sense, frameshift, canonical splice disruptions, and start loss mutaitons), and MetaSVM deleterious missense mutations were included. Analysis was performed by compared the observed versus expected number of *de novo* mutations using the denovolyzeR package. The Bonferroni correction for genome-wide significance is = 0.05/(18,989x3) = $8.8x10^{-7}$.

Supplementary Table 21a. *De novo* enrichment analysis in known monoallelic human CHD genes in cases and controls

	Cases, N=2,645						Controls, N=1,789					
	Observed		Expected		Enrichment	*p*	Observed		Expected		Enrichment	*p*
	N	Rate	N	Rate			N	Rate	N	Rate		
All monoallelic human CHD genes (N = 104)							**All monoallelic genes (N = 104)**					
Total	128	0.048	32.6	0.012	3.9	1.2×10^{-36}	Total	19	0.011	21.1	0.012	0.9	0.70
Syn	7	0.003	9.5	0.004	0.7	0.84	Syn	5	0.003	6.1	0.003	0.8	0.72
T-Mis	13	0.005	11.4	0.004	1.1	0.32	T-Mis	7	0.004	7.4	0.004	0.9	0.61
D-Mis	60	0.023	8.9	0.003	6.8	1.3×10^{-29}	D-Mis	6	0.003	5.8	0.003	1.0	0.53
LoF	48	0.018	2.7	0.001	17.7	3.8×10^{-42}	LoF	1	0.001	1.8	0.001	0.6	0.84
Damaging	108	0.041	11.6	0.004	9.3	5.5×10^{-65}	Damaging	7	0.004	7.6	0.004	0.9	0.64
Intolerant monoallelic human CHD genes (N = 54)							**Intolerant monoallelic human CHD genes (N = 54)**					
Total	103	0.039	24.5	0.009	4.2	3.1×10^{-32}	Total	12	0.007	16.1	0.009	0.7	0.88
Syn	5	0.002	7	0.003	0.7	0.82	Syn	3	0.002	4.6	0.003	0.7	0.83
T-Mis	12	0.005	9.1	0.003	1.3	0.21	T-Mis	5	0.003	6.0	0.003	0.8	0.71
D-Mis	47	0.018	6.3	0.002	7.5	2.5×10^{-25}	D-Mis	3	0.002	4.2	0.002	0.7	0.79
LoF	39	0.015	2.1	0.001	18.2	4.8×10^{-35}	LoF	1	0.001	1.4	0.001	0.7	0.76
Damaging	86	0.033	8.4	0.003	10.2	3.5×10^{-55}	Damaging	4	0.002	5.7	0.003	0.7	0.82
HHE monoallelic human CHD genes (N = 75)							**HHE monoallelic human CHD genes (N = 75)**					
Total	114	0.043	26.6	0.010	4.3	3.6×10^{-36}	Total	16	0.009	17.3	0.010	0.9	0.66
Syn	6	0.002	7.7	0.003	0.8	0.78	Syn	4	0.002	5	0.003	0.8	0.73
T-Mis	10	0.004	9.4	0.004	1.1	0.47	T-Mis	7	0.004	6.0	0.003	1.2	0.39
D-Mis	51	0.019	7.3	0.003	7.0	6.8×10^{-26}	D-Mis	4	0.002	4.9	0.003	0.8	0.72
LoF	47	0.018	2.2	0.001	21.3	6.4×10^{-45}	LoF	1	0.001	1.5	0.001	0.7	0.77
Damaging	98	0.037	9.5	0.004	10.3	8.8×10^{-63}	Damaging	5	0.003	6.3	0.004	0.8	0.76
Intolerant + HHE monoallelic human CHD genes (N = 44)							**Intolerant + HHE monoallelic human CHD genes (N = 44)**					
Total	94	0.036	21.4	0.008	4.4	7.1×10^{-31}	Total	12	0.007	14.2	0.008	0.8	0.75
Syn	5	0.002	6.1	0.002	0.8	0.73	Syn	3	0.002	4	0.002	0.8	0.76
T-Mis	10	0.004	8.0	0.003	1.3	0.28	T-Mis	5	0.003	5.2	0.003	1.0	0.59
D-Mis	41	0.016	5.5	0.002	7.4	4.4×10^{-22}	D-Mis	3	0.002	3.7	0.002	0.8	0.72
LoF	38	0.014	1.9	0.001	20.5	4.5×10^{-36}	LoF	1	0.001	1.2	0.001	0.8	0.71
Damaging	79	0.030	7.4	0.003	10.7	3.5×10^{-52}	Damaging	4	0.002	5	0.003	0.8	0.73

Intolerant genes: Genes with a pLI score ≥ 0.99. HHE: High heart-expressed genes (genes in the top quartile of expression).

Supplementary Table 21b. *De novo* enrichment analysis in known biallelic human CHD genes in cases and controls

	Cases, N=2,645						Controls, N=1,789					
	Observed		Expected		Enrichment	*p*		Observed		Expected		Enrichment	*p*
	N	Rate	N	Rate				N	Rate	N	Rate		
All biallelic human CHD genes (N = 85)							**All biallelic human CHD genes (N = 85)**						
Total	22	0.008	23.1	0.009	1.0	0.62	Total	16	0.009	15.4	0.009	1.0	0.47
Syn	8	0.003	6.3	0.002	1.3	0.30	Syn	3	0.002	4.2	0.002	0.7	0.79
T-Mis	6	0.002	11.1	0.004	0.5	0.96	T-Mis	9	0.005	7.4	0.004	1.2	0.32
D-Mis	6	0.002	3.4	0.001	1.8	0.13	D-Mis	1	0.001	2.3	0.001	0.4	0.90
LoF	2	0.001	2.3	0.001	0.9	0.67	LoF	3	0.002	1.5	0.001	2.0	0.20
Damaging	8	0.003	5.7	0.002	1.4	0.22	Damaging	4	0.002	3.8	0.002	1.0	0.53
Intolerant biallelic human CHD genes (N = 2)							**Intolerant biallelic human CHD genes (N = 2)**						
Total	1	0.000	1.9	0.001	0.5	0.85	Total	2	0.001	1.2	0.001	1.6	0.36
Syn	0	0.000	-	-	-	-	Syn	0	0.000	-	-	-	-
T-Mis	0	0.000	-	-	-	-	T-Mis	1	0.001	0.3	0.000	3.3	0.26
D-Mis	1	0.000	0.7	0.000	1.4	0.50	D-Mis	1	0.001	0.5	0.000	2.1	0.37
LoF	0	0.000	-	-	-	-	LoF	0	0.000	-	-	-	-
Damaging	1	0.000	0.9	0.000	1.2	0.58	Damaging	1	0.001	0.6	0.000	1.7	0.44
HHE biallelic human CHD genes (N = 11)							**HHE biallelic human CHD genes (N = 11)**						
Total	3	0.001	4.1	0.002	0.7	0.77	Total	3	0.002	2.7	0.002	1.1	0.50
Syn	1	0.001	1.2	0.000	0.8	0.70	Syn	1	0.001	0.8	0.000	1.3	0.55
T-Mis	0	0	-	-	-	-	T-Mis	1	0.001	1.7	0.001	0.6	0.81
D-Mis	1	0.001	2.5	0.001	2.1	0.39	D-Mis	0	0.000	-	-	-	-
LoF	1	0.001	0.3	0.000	3.4	0.25	LoF	1	0.001	0.2	0.000	5.1	0.18
Damaging	2	0.001	0.8	0.000	2.6	0.18	Damaging	1	0.001	0.5	0.000	1.9	0.41
Intolerant + HHE biallelic human CHD genes (N = 1)							**Intolerant + HHE biallelic human CHD genes (N = 1)**						
Total	0	0.000	-	-	-	-	Total	1	0.001	0.5	0.000	2.1	0.38
Syn	0	0.000	-	-	-	-	Syn	0	0.000	-	-	-	-
T-Mis	0	0.000	-	-	-	-	T-Mis	1	0.001	0.3	0.000	3.4	0.25
D-Mis	0	0.000	-	-	-	-	D-Mis	0	0.000	-	-	-	-
LoF	0	0.000	-	-	-	-	LoF	0	0.000	-	-	-	-
Damaging	0	0.000	-	-	-	-	Damaging	0	0.000	-	-	-	-

Intolerant genes: Genes with a pLI score ≥ 0.99. HHE: High heart-expressed genes (genes in the top quartile of expression).

Supplementary Table 22a *De novo* enrichment analysis in high heart expressed genes in cases stratified by extracardiac manifestations and neurodevelopmental deficit status

	Observed		Expected		Enrichment	*P*
	N	Rate	N	Rate		
Isolated CHD (820)						
Syn	64	0.078	72.3	0.088	0.9	0.85
T-Mis	152	0.185	129.5	0.158	1.2	2.9×10^{-2}
D-Mis	43	0.052	33.5	0.041	1.3	0.06
LoF	40	0.049	23.8	0.029	1.7	1.5×10^{-3}
Damaging	83	0.101	57.3	0.07	1.5	8.5×10^{-4}
CHD+EA only (318)						
Syn	27	0.085	28	0.088	1	0.6
T-Mis	56	0.176	50.2	0.158	1.1	0.22
D-Mis	29	0.091	13	0.041	2.2	8.8×10^{-5}
LoF	13	0.041	9.2	0.029	1.4	0.14
Damaging	42	0.132	22.2	0.07	1.9	1.2×10^{-4}
CHD+NDD only (327)						
Syn	27	0.083	28.8	0.088	0.9	0.66
T-Mis	40	0.122	51.6	0.158	0.8	0.96
D-Mis	21	0.064	13.4	0.041	1.6	3.2×10^{-2}
LoF	26	0.08	9.5	0.029	2.7	7.6×10^{-6}
Damaging	47	0.144	22.9	0.07	2.1	6.5×10^{-6}
CHD+Both (222)						
Syn	9	0.041	19.6	0.088	0.5	1
T-Mis	39	0.176	35	0.158	1.1	0.27
D-Mis	36	0.162	9.1	0.041	4	1.2×10^{-11}
LoF	42	0.189	6.5	0.029	6.5	1.4×10^{-20}
Damaging	78	0.351	15.5	0.07	5	1.6×10^{-29}

	Observed		Expected		Enrichment	*P*
	N	Rate	N	Rate		
CHD+NDD±EA (561)						
Syn	36	0.064	49.4	0.088	0.7	0.98
T-Mis	80	0.143	88.6	0.158	0.9	0.83
D-Mis	58	0.103	22.9	0.041	2.5	6.0×10^{-10}
LoF	68	0.121	16.3	0.029	4.2	1.2×10^{-21}
Damaging	126	0.225	39.2	0.070	3.2	3.5×10^{-28}
CHD+EA±NDD (935)						
Syn	65	0.070	82.4	0.088	0.8	0.98
T-Mis	259	0.277	185.9	0.199	1.4	2.4×10^{-7}
D-Mis	94	0.101	38.2	0.041	2.5	2.0×10^{-14}
LoF	95	0.102	27.2	0.029	3.5	3.8×10^{-24}
Damaging	189	0.202	65.4	0.070	2.9	1.6×10^{-35}
CHD+Either (1274)						
Syn	92	0.072	112.3	0.088	0.8	0.98
T-Mis	206	0.162	201.3	0.158	1.0	0.38
D-Mis	116	0.091	51.9	0.041	2.2	1.7×10^{-14}
LoF	121	0.095	37	0.029	3.3	9.3×10^{-28}
Damaging	237	0.186	89.1	0.07	2.7	1.4×10^{-38}

Enrichment (ratio of observed to expected) of damaging *de novo* mutations in high heart expressed (HHE) genes is shown for each phenotype. P-values are shown and calculated using the *denovolyzer* R package. CHD cases were characterized as isolated, associated with extracardiac congenital anomalies (EA), neurodevelopmental deficit (NDD) or NDD and/or EA.

Supplementary Table 22b *De novo* enrichment analysis in LoF-intolerant genes in cases stratified by extracardiac manifestations and neurodevelopmental deficit status

	Observed		Expected		Enrichment	*P*		Observed		Expected		Enrichment	*P*
	N	Rate	N	Rate				N	Rate	N	Rate		
Isolated CHD (820)							**CHD+NDD±EA (561)**						
Syn	44	0.054	47.2	0.058	0.9	0.7	Syn	21	0.037	32.3	0.058	0.7	0.99
T-Mis	88	0.107	85.0	0.104	1.0	0.39	T-Mis	68	0.121	58.1	0.104	1.2	0.11
D-Mis	34	0.041	23.3	0.028	1.5	2.2×10^{-2}	D-Mis	43	0.077	16	0.029	2.7	1.6×10^{-8}
LoF	22	0.027	16.6	0.020	1.3	0.12	LoF	51	0.091	11.3	0.020	4.5	5.8×10^{-18}
Damaging	56	0.068	39.9	0.049	1.4	9.2×10^{-3}	Damaging	94	0.168	27.3	0.049	3.5	1.7×10^{-23}
CHD+EA only (318)							**CHD+EA±NDD (935)**						
Syn	19	0.060	18.3	0.058	1.0	0.47	Syn	45	0.048	53.8	0.058	0.8	0.9
T-Mis	41	0.129	33.0	0.104	1.2	0.01	T-Mis	117	0.125	96.9	0.104	1.2	2.6×10^{-2}
D-Mis	21	0.066	9.0	0.028	2.3	4.7×10^{-4}	D-Mis	65	0.070	26.6	0.028	2.4	2.3×10^{-10}
LoF	12	0.038	6.4	0.020	1.9	3.1×10^{-2}	LoF	76	0.081	18.9	0.020	4.0	4.2×10^{-23}
Damaging	33	0.104	15.5	0.049	2.1	7.1×10^{-5}	Damaging	141	0.151	45.5	0.049	3.1	7.7×10^{-30}
CHD+NDD only (327)							**CHD+Either (1,274)**						
Syn	12	0.037	18.8	0.057	0.6	0.96	Syn	57	0.045	73.3	0.058	0.8	0.98
T-Mis	34	0.104	33.9	0.104	1.0	0.52	T-Mis	152	0.119	132.0	0.104	1.2	4.7×10^{-2}
D-Mis	17	0.052	9.3	0.028	1.8	1.5×10^{-2}	D-Mis	82	0.064	36.2	0.028	2.3	4.8×10^{-11}
LoF	18	0.055	6.6	0.020	2.7	1.8×10^{-4}	LoF	94	0.074	25.7	0.020	3.7	3.2×10^{-25}
Damaging	35	0.107	15.9	0.049	2.2	2.4×10^{-5}	Damaging	176	0.138	62	0.049	2.8	2.5×10^{-32}
CHD+Both (222)													
Syn	9	0.041	12.8	0.058	0.7	0.89							
T-Mis	33	0.149	23.0	0.104	1.4	2.9×10^{-2}							
D-Mis	26	0.117	6.3	0.028	4.1	3.8×10^{-9}							
LoF	33	0.149	4.5	0.020	7.4	4.8×10^{-18}							
Damaging	59	0.266	10.8	0.049	5.5	1.7×10^{-24}							

Enrichment (ratio of observed to expected) of damaging *de novo* mutations in LoF-intolerant genes (pLI ≥ 0.99) is shown for each phenotype. P-values are shown and calculated using the *denovolyzer* R package. CHD cases were characterized as isolated, associated with extracardiac congenital anomalies (EA), neurodevelopmental deficit (NDD) or NDD and/or EA.

Supplementary Table 22c *De novo* expectation analysis in LoF-intolerant and high heart expressed genes in cases stratified by extracardiac manifestations and neurodevelopmental deficit status

	Observed		Expected		Enrichment	*P*		Observed		Expected		Enrichment	*P*
	N	Rate	N	Rate				N	Rate	N	Rate		
Isolated CHD (820)							**CHD+NDD±EA (561)**						
Syn	26	0.032	27.4	0.033	0.9	0.63	Syn	9	0.016	18.7	0.033	0.5	1
T-Mis	56	0.068	50.1	0.061	1.1	0.22	T-Mis	33	0.059	34.3	0.061	1.0	0.29
D-Mis	20	0.024	13.5	0.016	1.5	5.7×10^{-2}	D-Mis	28	0.050	9.2	0.016	3.0	4.8×10^{-7}
LoF	13	0.016	9.8	0.012	1.3	0.19	LoF	46	0.082	6.7	0.012	6.8	3.2×10^{-23}
Damaging	33	0.040	23.3	0.028	1.4	3.4×10^{-2}	Damaging	74	0.132	15.9	0.028	4.6	4.6×10^{-26}
CHD+EA only (318)							**CHD+EA±NDD (935)**						
Syn	13	0.041	10.6	0.033	1.2	0.27	Syn	24	0.026	31.2	0.033	0.8	0.92
T-Mis	23	0.072	19.5	0.061	1.2	0.24	T-Mis	70	0.075	57.1	0.061	1.2	5.4×10^{-2}
D-Mis	17	0.053	5.2	0.016	3.3	3.4×10^{-5}	D-Mis	47	0.050	15.4	0.016	3.1	6.9×10^{-11}
LoF	9	0.028	3.8	0.012	2.4	1.6×10^{-2}	LoF	67	0.072	11.2	0.012	6.0	1.0×10^{-29}
Damaging	26	0.082	9.0	0.028	2.9	3.2×10^{-6}	Damaging	114	0.122	26.6	0.028	4.3	3.7×10^{-36}
CHD+NDD only (327)							**CHD+Either (1274)**						
Syn	5	0.015	10.9	0.033	0.5	0.98	Syn	29	0.023	42.5	0.033	0.7	0.99
T-Mis	16	0.049	20.0	0.061	0.8	0.84	T-Mis	87	0.068	77.9	0.061	1.1	0.16
D-Mis	9	0.028	5.4	0.017	1.7	9.5×10^{-2}	D-Mis	56	0.044	20.9	0.016	2.7	1.6×10^{-10}
LoF	15	0.046	3.9	0.012	3.8	1.6×10^{-5}	LoF	82	0.064	15.3	0.012	5.4	8.1×10^{-33}
Damaging	24	0.073	9.3	0.028	2.6	4.0×10^{-5}	Damaging	138	0.108	36.2	0.028	3.8	5.0×10^{-38}
CHD+Both (222)													
Syn	4	0.018	7.4	0.033	0.5	0.94							
T-Mis	16	0.072	13.6	0.061	1.2	0.29							
D-Mis	19	0.086	3.6	0.016	5.2	1.2×10^{-8}							
LoF	31	0.140	2.7	0.012	11.6	1.5×10^{-22}							
Damaging	50	0.225	6.3	0.028	7.9	6.9×10^{-28}							

Enrichment (ratio of observed to expected) of damaging *de novo* mutations in LoF-intolerant (pLI ≥ 0.99) and high heart expressed (HHE) genes is shown for each phenotype. P-values are shown and calculated using the *denovolyzer* R package. CHD cases were characterized as isolated, associated with extracardiac congenital anomalies (EA), neurodevelopmental deficit (NDD) or NDD and/or EA.

Supplementary Table 22d. *De novo* enrichment analysis in all genes in cases stratified by extracardiac manifestations and neurodevelopmental deficit status

	Observed		Expected		Enrichment	*P*
	N	Rate	N	Rate		
Isolated CHD (820)						
Syn	219	0.267	260.2	0.317	0.8	1
T-Mis	481	0.587	467.8	0.570	1.0	0.28
D-Mis	118	0.144	107.9	0.132	1.1	0.18
LoF	98	0.120	80.6	0.098	1.2	3.3×10^{-2}
Damaging	216	0.263	188.6	0.230	1.2	2.7×10^{-2}
CHD+EA only (318)						
Syn	102	0.321	100.9	0.317	1.0	0.47
T-Mis	179	0.563	181.3	0.570	1.0	0.28
D-Mis	60	0.189	41.9	0.132	1.4	4.9×10^{-3}
LoF	44	0.138	31.3	0.098	1.4	1.8×10^{-2}
Damaging	104	0.327	73.1	0.230	1.4	3.9×10^{-4}
CHD+NDD only (327)						
Syn	88	0.269	103.7	0.317	0.8	0.95
T-Mis	161	0.492	186.6	0.571	0.9	0.97
D-Mis	56	0.171	43	0.131	1.3	3.3×10^{-2}
LoF	50	0.153	32.2	0.098	1.6	2.1×10^{-3}
Damaging	106	0.324	75.2	0.230	1.4	4.7×10^{-4}
CHD+Both (222)						
Syn	39	0.176	70.4	0.317	0.6	1
T-Mis	131	0.590	126.7	0.571	1.0	0.36
D-Mis	62	0.279	29.2	0.132	2.1	9.0×10^{-8}
LoF	56	0.252	21.8	0.098	2.6	7.3×10^{-10}
Damaging	118	0.532	51.1	0.230	2.3	8.9×10^{-16}

	Observed		Expected		Enrichment	*P*
	N	Rate	N	Rate		
CHD+NDD±EA (561)						
Syn	128	0.228	178	0.317	0.7	1
T-Mis	297	0.529	319.9	0.570	0.9	0.9
D-Mis	119	0.212	73.9	0.132	1.6	8.4×10^{-7}
LoF	107	0.191	55.2	0.098	1.9	4.2×10^{-10}
Damaging	226	0.403	129	0.230	1.8	7.7×10^{-15}
CHD+EA±NDD (935)						
Syn	242	0.259	296.6	0.317	0.8	1
T-Mis	520	0.556	533.3	0.570	1.0	0.72
D-Mis	189	0.202	123.1	0.132	1.5	2.2×10^{-8}
LoF	166	0.178	92	0.098	1.8	2.6×10^{-12}
Damaging	355	0.380	215	0.230	1.7	1.7×10^{-18}
CHD+Either (1274)						
Syn	331	0.260	404.2	0.317	0.8	1
T-Mis	686	0.538	726.7	0.570	0.9	0.94
D-Mis	246	0.193	167.7	0.132	1.5	9.2×10^{-9}
LoF	217	0.170	125.3	0.098	1.7	7.4×10^{-14}
Damaging	463	0.363	293	0.230	1.6	3.4×10^{-20}

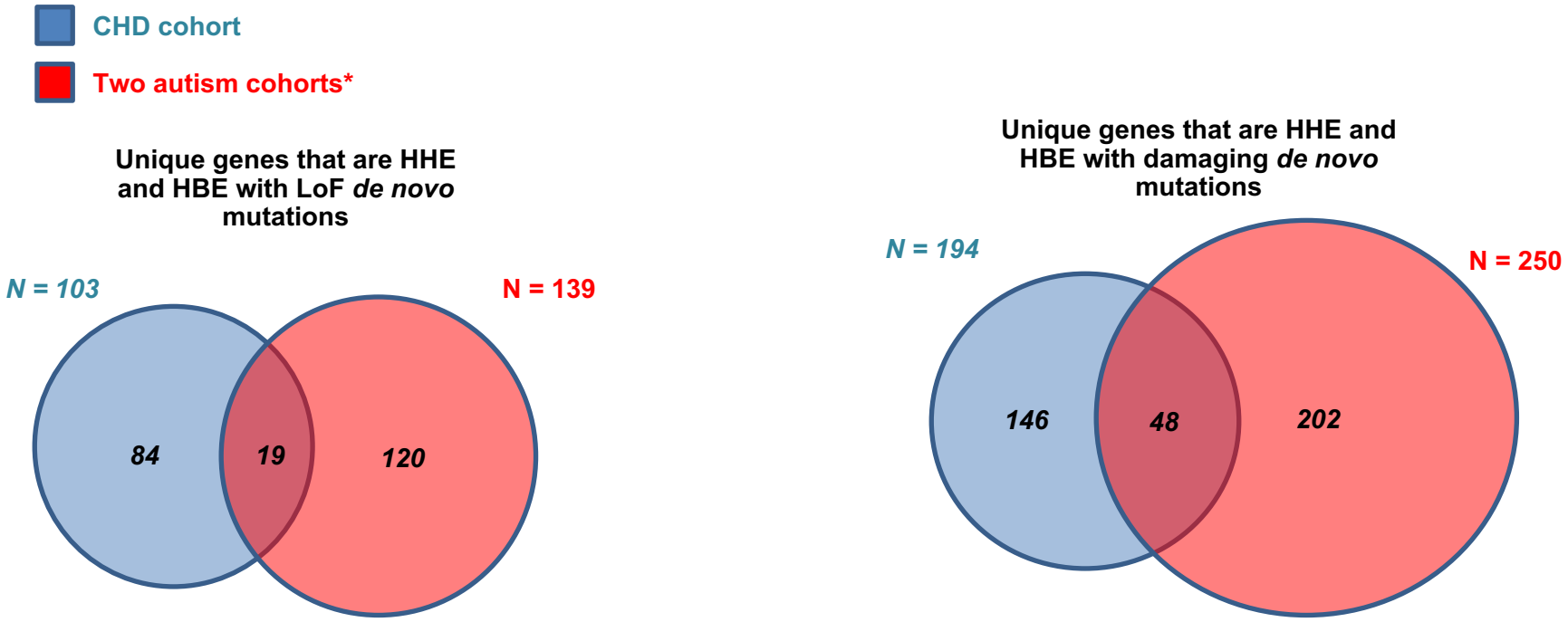
Enrichment (ratio of observed to expected) of damaging *de novo* mutations in all genes is shown for each phenotype. P-values are shown and calculated using the *denovolyzer* R package. CHD cases were characterized as isolated, associated with extracardiac congenital anomalies (EA), neurodevelopmental deficit (NDD) or NDD and/or EA.

Supplementary Table 23. Ten genes with > 1 damaging *de novo* mutation in patients with CHD + NDD + EA

Gene	# LoF	# D-Mis	HHE Rank	pLI	*P*	Known CHD Genes?	Chromatin Modifiers?
*KMT2D**	6	1	96.8	1.00	$1.6x10^{-15}$	Yes	Yes
*SOS1**	0	3	67.9	1.00	$1.6x10^{-8}$	Yes	No
*PTPN11**	0	3	94.2	1.00	$2.6x10^{-8}$	Yes	No
*CHD7**	3	0	93.4	1.00	$4.7x10^{-8}$	Yes	Yes
*RIT1**	0	2	61.4	0.67	$8.4x10^{-7}$	Yes	No
RAF1	0	2	91.4	1.00	$1.1x10^{-5}$	Yes	No
CHD4	0	2	99.1	1.00	$2.3x10^{-4}$	Yes	Yes
NSD1	1	1	94.8	1.00	$2.6x10^{-4}$	Yes	Yes
CAD	0	2	85.9	1.00	$3.6x10^{-4}$	No	No
NOTCH1	1	1	87.9	1.00	$3.9x10^{-4}$	Yes	No

Loss-of-function and MetaSVM deleterious missense mutations in patients with CHD + NDD + EA are shown. Analysis was performed by compared observed versus expected number of *de novo* mutations in the CHD + NDD + EA patient set using the denovolyzeR package. The Bonferroni correction for genome-wide significance is = 0.05/(18,989x 3) = $8.8x10^{-7}$. (*) denotes genes that surpass genome-wide significance.

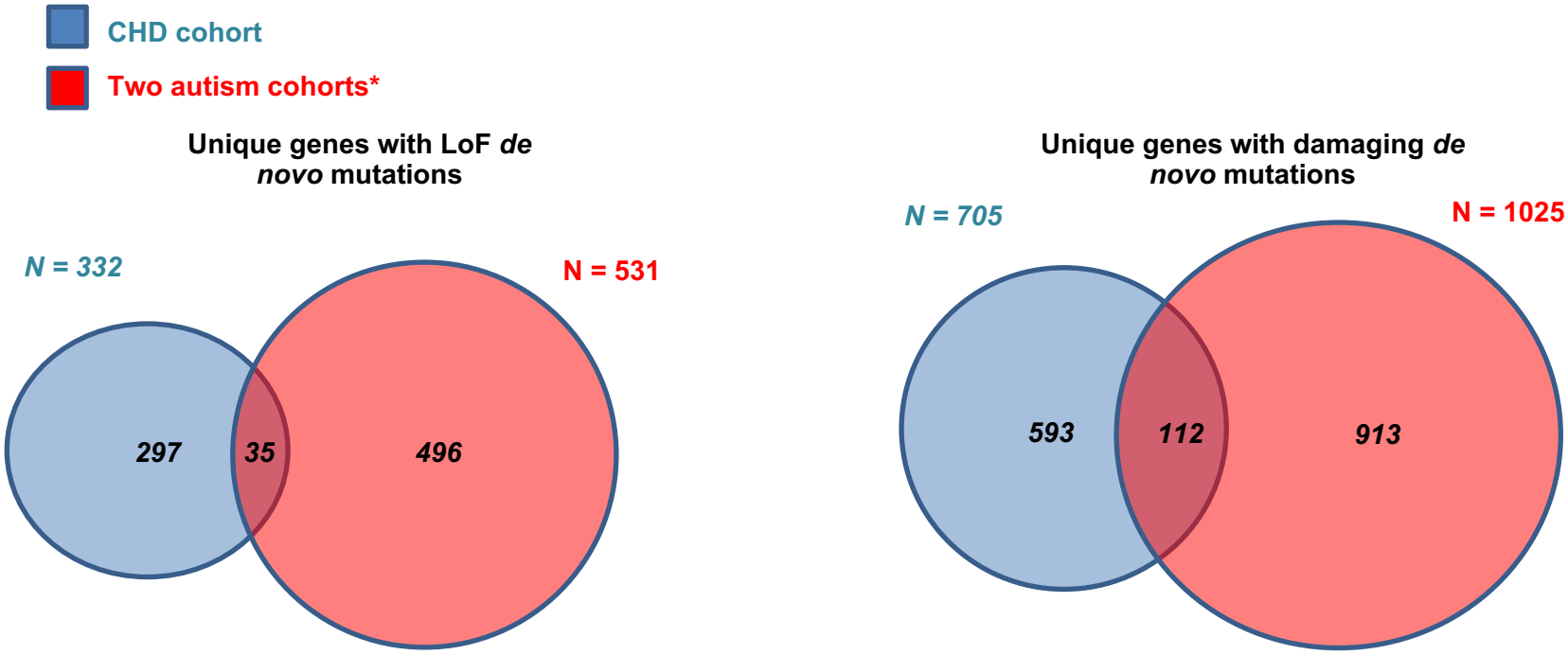
Supplementary Table 24a. Enrichment of overlapping high heart expressed and high brain expressed genes with damaging *de novo* mutations between CHD and autism cohorts



	Observed # genes	Expected # genes	Enrichment	Empirical P-value
HHE + HBE Genes with LoF *de novo* mutations overlapping between CHD and autism cohorts	19	3.68	5.16	$<10^{-6}$
HHE + HBE Genes with damaging *de novo* mutations overlapping between CHD and autism cohorts	48	17.35	2.77	$<10^{-6}$

10^6 permutations were performed to estimate the empirical distribution of the number of overlapping high heart expressed (HHE) + high brain expressed (HBE) genes between CHD and 2 autism cohorts. The empirical p-value is calculated as the proportion of the expected number of overlapping genes that exceeds the observed number of overlapping genes. For the detailed approach, please see **Online Methods.** *These two autism cohorts refer to: (1) Iossifov et al. ***Nature*** 2014 515, 216-221, and (2) De Rubeis et al. ***Nature*** 2014 515, 209-215.

Supplementary Table 24b. Enrichment of overlapping genes with LoF or damaging *de novo* mutations between CHD and autism cohorts



	Observed # genes	Expected # genes	Enrichment	Empirical P-value
Genes with LoF *de novo* mutations overlapping between CHD and autism cohorts	35	16.83	2.08	2.9×10^{-5}
Genes with damaging *de novo* mutations overlapping between CHD and autism cohorts	112	81.23	1.38	2.0×10^{-4}

10^6 permutations were performed to estimate the empirical distribution of the number of overlapping genes between CHD and 2 autism cohorts. The empirical p-value is calculated as the proportion of the expected number of overlapping genes that exceeds the observed number of overlapping genes. For the detailed approach, please see **Online Methods**.*These two autism cohorts refer to: (1) Iossifov et al. ***Nature*** 2014 515, 216-221, and (2) De Rubeis et al. ***Nature*** 2014 515, 209-215.

Supplementary Table 25. Thirty-five genes with loss-of-function *de novo* mutation shared between the CHD and autism cohorts*

Gene	# LoF	HHE Rank	pLI	Known CHD Gene?	Chromatin Modifier?	Extracardiac Status	NDD Status
KDM5B	3	86.0	0.00	No	Yes	+/-/-	+/U/+
PTEN	2	77.9	0.98	Yes	No	+/+	U/+
FRYL	2	84.4	1.00	No	No	+/+	U/-
CTNNB1	2	99.0	1.00	No	Yes	+/+	U/+
POGZ	2	83.8	1.00	No	Yes	+/+	+/U
NOTCH1	2	87.9	1.00	Yes	No	-/+	+/+
NAA15	2	96.7	1.00	No	No	-/+	-/-
ZC3H14	1	84.2	1.00	No	No	+	-
CCSER1	1	26.5	0.21	No	No	-	-
NIN	1	70.0	0.00	No	No	-	U
EP300	1	88.1	1.00	No	Yes	+	+
ASH1L	1	87.0	1.00	No	Yes	-	+
KDM6B	1	94.7	1.00	No	Yes	+	+
KAT6A	1	88.9	1.00	No	Yes	+	+
KMT2C	1	80.0	1.00	No	Yes	-	+
KIAA0100	1	95.4	0.27	No	No	-	+
TLK2	1	71.6	1.00	No	Yes	-	+
UBR3	1	87.4	1.00	No	No	-	+
ZNF292	1	77.3	1.00	No	No	-	U
SPRED2	1	62.1	0.00	No	No	+	-
LRRFIP1	1	96.4	0.37	No	No	-	-
SRRM2	1	99.4	NA	No	No	-	+
TANC2	1	84.4	1.00	No	No	-	+
ARID1B	1	83.2	1.00	No	Yes	+	U
EP400	1	84.9	1.00	No	Yes	-	U
TECTA	1	22.0	0.00	No	No	-	-
WAC	1	87.3	1.00	No	Yes	+	U
ANK2	1	72.8	1.00	No	No	+	U
KMT2A	1	86.1	1.00	Yes	Yes	+	+
SETD5	1	95.1	1.00	No	No	-	+
PSMD12	1	89.1	1.00	No	No	-	+
SPTBN1	1	99.4	1.00	No	No	+	U
CUL3	1	83.0	0.97	No	No	-	-
BAI1	1	20.0	1.00	No	No	-	-
WHSC1	1	89.0	1.00	No	Yes	+	+

(+): positive; (-): negative; (U): Unknown. *These two autism cohorts refer to: (1) Iossifov et al. ***Nature*** 2014 515, 216-221, and (2) De Rubeis et al. ***Nature*** 2014 515, 209-215.

SUPPLEMENTARY NOTE

Duplicated samples. To identify subjects which could have been recruited multiple times in case cohorts, control cohorts, or both, we calculated the overlap of high-confidence rare variants (MAF = 0% in ExAC, 1000 Genomes, and EVS) between each pair of individuals. For pairs that share ≥ 80% of rare variants, the sample with greater sequence coverage was kept in the analysis and the other discarded.

Identification of copy number variations. Samples were genotyped for CNVs as described previously[1]. Briefly, read depths for each sample at each exon were compiled using GATK DepthOfCoverage[2,3] and used for discovering copy number duplications or deletions using the XHMM (eXome-Hidden Markov Model) software[4]. *De novo* germline CNVs were called using XHMM best practices[5]. Any *de novo* CNVs involving a previously described CHD gene[6] or CHD-associated locus[1] resulted in sample exclusion from further analysis.

Known human and mouse CHD gene sets: We applied stringent criteria to define gene sets that contribute to CHD. We first manually curated a total of 212 genes that affect disease pathogenesis in humans. These genes were selected based on the following criteria: (1) identified in OMIM (On-Line Mendelian Inheritance in Man) as having an association with CHD; or, (2) being implicated in human CHD in previous publications[7]. Additionally, a set of 61 genes that has recently been shown to cause CHD in mice was downloaded from the supplementary materials[8]. Merging these two gene sets resulted in a total of 253 unique genes (**Supplementary Data Set 2**).

Analysis of inbreeding and *GDF1* p.Met364Thr Shared Haplotype Analysis. Haplotype phasing, inbreeding coefficient, and the longest homozygosity-by-descent (HBD) fragment were estimated using Beagle v3.3.2[9]. The criteria of consanguinity are defined as homozygosity in segments of 2cM or greater length that collectively comprise at least 0.35% of the genome.

The haplotype flanking *GDF1* p.Met364Thr variants was phased using the PhaseByTransmission tool of GATK[2] and the haplotype boundaries were determined for each family until the first heterozygotes were identified in both direction. This defined a shared haplotype spanning no more than 234 kb in length. Because all of *GDF1* p.Met364Thr carriers were AJs and shared a short haplotype, it suggests this is a founder mutation which originated several generations ago in the AJ. To estimate the exact age of the *GDF1* p.Met364Thr mutation, we used the DMLE+2.3[10] algorithm as previously described[11]. DMLE+2.3 performs a Bayesian inference that takes into account the proportion of disease chromosomes to estimate the mutation age based on observed linkage disequilibrium data from markers located within the shared haplotype. Genotypes for 62 polymorphic markers spanning the shared haplotype were entered. The proportion of sampled chromosomes (PSC) were estimated by $PSC = n/(2 \times P \times MAF)$ where n represents the number of disease chromosomes for CHD patients harboring the *GDF1* p.Met364Thr mutation (n=21; 10 homozygotes and 1 heterozygote in 204 AJ cases), P represents the estimated size of

the AJ population (about 1.1×10^7), and MAF represents the estimated minor allele frequency in the AJ population (0.56%). This resulted in a PSC of 1.7×10^{-4}. The population growth rates (PGR) of the AJ population were estimated by $PGR = ln(T_1/T_0)/g$, where T_1 represents the estimated size of the AJ population at the present time (about 1.1×10^7), T_0 represents the estimated size of the Eastern European population (about 7,000 in 1,170 A.D. and 10,000 in 1,500 A.D.), and g represents the number of generation between T_1 and T_0 assuming 20 years per generation. Using the T_0 values from 1,170 A.D. and 1,500 A.D. resulted in PGR estimates of 0.175 and 0.26, respectively. The average of 0.175 and 0.26 (=0.22) was used for used for estimation. Final mutation age estimates were calculated using a PSC value of 1.7×10^{-4} and a PGR value of 0.22. The point estimate and 95% confidence interval were reported.

Modeling of the distribution of recessive and dominant variant count per gene: In an outbred population, the expected frequency of damaging RGs in each subject is the square of the cumulative frequency of damaging alleles in each gene, q^2; similarly, the number of different RGs increases as the square of the number of different recessive alleles in a gene. From these expected values, one can use the binomial test to compare the observed count of damaging RGs found in a cohort to that expected by chance. We used the expected frequency of damaging DNMs in each gene as a proxy for the expected frequency of damaging rare heterozygous variants in each gene; these are highly correlated (Pearson correlation = 0.82 and 0.83 in cases and controls, respectively; **Supplementary Figure 2**), consistent with a previous study that suggested high correlation of using *de novo* probability for estimating counts of standing variations[12].

The analysis of RGs is commonly complicated by the presence of inbreeding, in which the frequency of homozygous genotypes is increased. In this setting, the number of different homozygous genotypes does not increase as the square of the number of allelic loci in a gene, but instead increases linearly with this number. Similarly, the number of different compound heterozygous genotypes increases as $q^2 – q$. To account for this effect of inbreeding, we utilized two approaches: first, we jointly analyzed the burden of all RGs by fitting a polynomial regression model and performing a binomial test to determine the enrichment of RGs; as an alternative approach, we separately modeled homozygous and compound heterozygous genotypes and then calculated the enrichment using a binomial test of the combined results. Although the joint analysis is statistically more powerful and tests the most desired hypothesis (*i.e.,* that the burden of all damaging RGs in a gene or a gene set is greater than that expected by chance), these two approaches yielded similar empiric results.

More specifically, in the absence of inbreeding, for a gene *x* with *n* allelic loci which have equal allele frequency, the expected number of RGs (sum of homozygous and compound heterozygous genotypes) is a quadratic function of the number of allelic loci within this gene (*n*), while the number of homozygous variants increases linearly with *n*. In an outbred population, the number of compound heterozygous variants, therefore, follows a polynomial function proportional to $n^2 – n$. However, in the

presence of inbreeding, the enrichment of homozygous variants distorts the expectation that the number of RGs increases with the square of the number of allelic loci; therefore, it is no longer accurate to calculate the expected number of RGs based on the quadratic function. To estimate the relationship between the number of observed RGs and mutability, we performed regression analysis and compared the following models:

(1) Polynomial model: $\text{Number of RGs} = \beta_0 + \beta_1 \times \text{mutability} + \beta_2 \times \text{mutability}^2 + \varepsilon$

(2) Quadratic model: $\text{Number of RGs} = \beta_0 + \beta_2 \times \text{mutability}^2 + \varepsilon$

(3) Linear model: $\text{Number of RGs} = \beta_0 + \beta_1 \times \text{mutability} + \varepsilon$

where β_0 is the y-axis intercept (when the mutability is 0); β_1 is the coefficient for the linear term (the difference in the predicted RG Count for each one-unit increment in mutability); β_2 is the coefficient for the quadratic term (the difference in the predicted RG Count for each one-unit increment in mutability^2). ε is a random variable accounting for the variance that cannot be explained by mutability.

To test this approach, we first fitted the polynomial regression model to the observed distribution of all rare compound heterozygous or homozygous synonymous genotypes in each gene in cases and controls (MAF ≤ 0.001 in ExAC, 1000 Genomes, and EVS). The goodness-of-fit was evaluated by the following fit metric (D'):

$$\text{D}' = \left| 1 - \frac{\sum(\#\text{ observed RG} - \#\text{ expected RG})^2}{\sum(\#\text{ observed RG})^2} \right|$$

The data closely fit the polynomial model in cases and controls (goodness of fit metric D' = 0.98 and 0.99, respectively; **Supplementary Figure 3**). Q-Q plots showed that no gene showed more synonymous RGs than expected by chance (**Supplementary Figure 4**). We found that the polynomial model fits recessive, homozygous, and compound heterozygous counts substantially better than the quadratic and linear models (**Supplementary Figures 3-6**).

This method can be extended to analysis of the burden of RGs in specified gene sets. For example, by randomly selecting 10,000 gene sets of 100 - 1000 genes, the p-values for the burden of synonymous RGs in each gene set closely matched the expected values in cases and controls (**Supplementary Figure 5a**).

Alternatively, RG can also be modeled separately as compound heterozygotes or homozygotes without the need for regression fits. In this method, the expected number of compound heterozygotes for each gene is derived from distributing the observed number of RGs, N, across all genes according to the ratio of the squared *de novo* probabilities:

$$\text{Expected Compound RG}_i = N \times \frac{probability_{de\ novo}^2}{\sum_{Genes}(probability_{de\ novo}^2)}$$

The expected number of homozygotes is derived similarly, but using the linear ratio of *de novo* probabilities:

$$\text{Expected Homozygous RG}_i = N \times \frac{probability_{de\ novo}}{\sum_{Genes}(probability_{de\ novo})}$$

The expected distributions of homozygous, compound heterozygous, and total RG among cases and controls for synonymous (**Supplementary Figure 7**) and damaging (**Supplementary Figure 8**) variants fit the observed distributions remarkably well (D' ≥ 0.95). This method reaches similar conclusions to the polynomial fit method. In comparing the enrichments of damaging RG among cases by this method (**Supplementary Table 10**) to the polynomial method (**Supplementary Table 9**), two genes, *GDF1* and *MYH6*, are found to be significantly enriched. These analyses validate the use of the *de novo* probability coupled with the polynomial model to accurately estimate the expected number of RGs under the null model.

Variant prioritization using VAAST and Phevor. As an orthogonal approach to disease gene discovery, we also prioritized disease-causing dominant and recessive/compound heterozygous genotypes using the pedigree version of VAAST (Variant Annotation, Analysis, and Search Tool). VAAST is a probabilistic disease-gene finder that employs a burden test to identity disease-genes using personal genomes data[13,14]. VAAST uses the global amino acid substitution frequencies, nucleotide-level conservation and variant population frequency data to derive a score that reflects the degree of deleteriousness of each allele. The VAAST empirical p-value is derived by comparing the observed VAAST score to the scores obtained by permutation analysis of a background population (N=2,504) that encompasses 26 ethnicities, derived from the 1000 Genomes project. The minimal attainable permutation p value is equal to 1/(# of background genomes + 1), thus 0.000399. Any variant in any proband with a p-value ≤ 0.005 was retained for further analyses. This p-value was chosen so as to be maximally concordant with previous publications[15] and thus retain > 92% of reported positives.

Variants meeting the p-value threshold ≤ 0.005 were further analyzed using Phevor[16] (Phenotype Driven Variant Ontological Re-ranking Tool). Phevor re-ranks the outputs of variant prioritization tools using a novel algorithm that propagates information across and between ontologies (e.g., the Human Phenotype Ontology[17]) in order to accurately reprioritize potentially damaging alleles in light of the gene function, disease and phenotype knowledge. We have shown that Phevor is especially useful for single exome and family trio-based diagnostic analyses, the most commonly occurring clinical scenarios[16]. The following HPO terms were used to traverse HPO or HPO+gene ontology (GO) ontologies: Conotruncal Defects (CTD) – abnormality of the pulmonary artery (HP:0004414), double outlet right ventricle (HP:0001719), Tetralogy of Fallot (HP:0001636), transposition of the great arteries (HP:0001669), truncus arteriosus (HP:0001660); Left Ventricular Outflow Tract Obstruction (LVOTO) – abnormality of the left ventricular (HP:0011103), bicuspid aortic valve (HP:0001647), subvalvular aortic stenosis (HP:0001682), supravalvular aortic stenosis (HP:0004381), endocardial fibroelastosis (HP:0001706), coarctation of aorta (HP:0001680); and Heterotaxy (HTX) - ventricular septal defect (HP:0001629), defect in the atrial septum (HP:0001631), pulmonic stenosis (HP:0001642), patent ductus arteriosus (HP:0001643), dextrocardia (HP:0001651), transposition of the great arteries (HP:0001669), complete atrioventricular canal defect (HP:0001674), situs

inversus totalis (HP:0001696), asplenia (HP:0001746), polysplenia (HP:0001748), abnormality of the respiratory system (HP:0002086), Abnormal lung lobation (HP:0002101), abdominal situs inversus (HP:0003363), pulmonary artery atresia (HP:0004935), atrioventricular canal defect (HP:0006695), ectopia of the spleen (HP:0010452), anomalous pulmonary venous return (HP:0010772), right atrial isomerism (HP:0011536), left atrial isomerism (HP:0011537), common atrium (HP:0011565), mesocardia (HP:0011599), and right aortic arch (HP:0012020).

Only genes with final Phevor scores of > 2.3 were retained. This threshold was selected as it corresponds to a neutral (flat) Bayesian prior (0.5). Thus any gene with a VAAST p-value ≤ 0.005 and a Phevor prior > 0.5 would a have final score > 2.3 ($-log_{10}$) for association with the proband's CHD phenotype based upon the contents of the Human Phenotype[17] and Gene Ontologies[18]; see Singleton et al.[16] for details. To reduce the rate of false positives, variants chosen by VAAST and Phevor were subjected to variant adjudication using a modified Smith-Waterman algorithm (Graphite, http://marthlab.org).

Damaging recessive and *de novo* genotypes identified using VAAST/Phevor. Manhattan plots for recessive genotypes in PCGC trio probands meeting criteria for VAAST p-value ≤ 0.005 and $-log_{10}$ Phevor scores > 2.3 are displayed in **Supplementary Figure 9**. PCGC trio probands with damaging recessive genotypes in *GDF1*, *MYH6*, *DNAH5* and other selected genes concordant with the analysis pipeline described in the main text are presented. Manhattan plots for damaging *de novo* alleles identified using the VAAST-Phevor pipeline are presented in **Supplementary Figure 18**. As this pipeline compares the degree of burden using a distinct background population and a distinct permutation-based approach, these data provide an orthogonal approach that complements the findings described in the main text.

Estimating the number of risk genes. We followed the Monte Carlo simulation strategy described in Homsy et al.[19] to estimate the number of risk genes that are DNM targets[19]. We defined K to be the number of observed damaging DNMs in HHE genes among cases. R_1 indicates the number of HHE genes mutated exactly twice in cases and R_2 indicates the number of HHE genes mutated three times or more. Defined as $E = (M_1-M_2)/M_1$, where M_1 and M_2 are the observed and expected count of damaging DNMs per trio, respectively. We then simulated the likelihood function as follows: First, we randomly selected G risk genes from the HHE gene set. Next, we simulated the number of contributing damaging mutations in risk genes, i.e. C, by sampling once from Binomial(K,E) distribution. Then, we simulated C contributing damaging mutations in G risk genes and K-C non-contributing damaging mutations in the complete HHE gene set using each gene's damaging mutability score as probability weights. We performed 20,000 simulations for G from 10 to 1,000, and calculated the likelihood function L(G) as the proportion of simulations in which the number of genes with two damaging mutations equals to R_1 and the number of genes with three or more damaging mutations equals to R_2. We then estimated the number of risk genes using the maximum likelihood estimate (MLE). Based on the likelihood function, we calculated

the Fisher information and constructed the confidence interval based on the MLE and estimated Fisher information using the following equation

$$MLE \pm 1.96 \times \left(\frac{1}{\sqrt{Fisher\ Information}} \right)$$

REFERENCES

1. Glessner, J.T. *et al.* Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ Res* **115**, 884-96 (2014).
2. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
3. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 1-33 (2013).
4. Fromer, M. *et al.* Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet* **91**, 597-607 (2012).
5. Fromer, M. & Purcell, S.M. Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data. *Curr Protoc Hum Genet* **81**, 7 23 1-7 23 21 (2014).
6. Fahed, A.C., Gelb, B.D., Seidman, J.G. & Seidman, C.E. Genetics of congenital heart disease: the glass half empty. *Circ Res* **112**, 707-20 (2013).
7. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet* **48**, 1060-5 (2016).
8. Li, Y. *et al.* Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature* **521**, 520-4 (2015).
9. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-97 (2007).
10. Reeve, J.P. & Rannala, B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics* **18**, 894-5 (2002).
11. Lemaire, M. *et al.* Recessive mutations in DGKE cause atypical hemolytic-uremic syndrome. *Nat Genet* **45**, 531-6 (2013).
12. Samocha, K.E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944-50 (2014).
13. Yandell, M. *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Res* **21**, 1529-42 (2011).
14. Hu, H. *et al.* VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genet Epidemiol* **37**, 622-34 (2013).
15. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220-3 (2013).
16. Singleton, M.V. *et al.* Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet* **94**, 599-610 (2014).
17. Groza, T. *et al.* The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease. *Am J Hum Genet* **97**, 111-24 (2015).
18. Harris, M.A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-61 (2004).
19. Homsy, J. *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262-6 (2015).