Increased Frequency of De Novo Copy Number Variants in Congenital Heart Disease by Integrative Analysis of Single Nucleotide Polymorphism Array and Exome Sequence Data

Joseph T. Glessner,* Alexander G. Bick,* Kaoru Ito,* Jason G. Homsy,* Laura Rodriguez-Murillo, Menachem Fromer, Erica Mazaika, Badri Vardarajan, Michael Italia, Jeremy Leipzig, Steven R. DePalma, Ryan Golhar, Stephan J. Sanders, Boris Yamrom, Michael Ronemus, Ivan Iossifov, A. Jeremy Willsey, Matthew W. State, Jonathan R. Kaltman, Peter S. White, Yufeng Shen, Dorothy Warburton, Martina Brueckner, Christine Seidman,† Elizabeth Goldmuntz,† Bruce D. Gelb,† Richard Lifton,† Jonathan Seidman,† Hakon Hakonarson,† Wendy K. Chung†

Rationale: **Congenital heart disease (CHD) is among the most common birth defects. Most cases are of unknown pathogenesis.**

Objective: **To determine the contribution of de novo copy number variants (CNVs) in the pathogenesis of sporadic CHD.**

Methods and Results: **We studied 538 CHD trios using genome-wide dense single nucleotide polymorphism arrays and whole exome sequencing. Results were experimentally validated using digital droplet polymerase chain reaction. We compared validated CNVs in CHD cases with CNVs in 1301 healthy control trios. The 2 complementary high-resolution technologies identified 63 validated de novo CNVs in 51 CHD cases. A significant increase in CNV burden was observed when comparing CHD trios with healthy trios, using either single nucleotide polymorphism array ($P=7\times10^{-5}$; odds ratio, 4.6) or whole exome sequencing data ($P=6\times10^{-4}$; odds ratio, 3.5) and remained after removing 16% of de novo CNV loci previously reported as pathogenic ($P=0.02$; odds ratio, 2.7). We observed recurrent de novo CNVs on 15q11.2 encompassing *CYFIP1, NIPA1*, and *NIPA2* and single de novo CNVs encompassing *DUSP1, JUN, JUP, MED15, MED9, PTPRE SREBF1, TOP2A,* and *ZEB2,* genes that interact with established CHD proteins *NKX2-5* and *GATA4*. Integrating de novo variants in whole exome sequencing and CNV data suggests that *ETS1* is the pathogenic gene altered by 11q24.2-q25 deletions in Jacobsen syndrome and that *CTBP2* is the pathogenic gene in 10q subtelomeric deletions.**

Conclusions: **We demonstrate a significantly increased frequency of rare de novo CNVs in CHD patients compared with healthy controls and suggest several novel genetic loci for CHD. (*Circ Res*. 2014;115:884-896.)**

Key Words: DNA copy number variations ■ genomics ■ microarray analysis ■ polymorphism, single nucleotide

Original received May 26, 2014; revision received September 5, 2014; accepted September 9, 2014. In August, 2014, the average time from submission to first decision for all original research papers submitted to *Circulation Research* was 13.55 days.

From the Center for Applied Genomics (J.T.G., R.G., H.H.), Center for Biomedical Informatics (M.I., J.L., P.S.W.), and Division of Cardiology (E.G.), Children's Hospital of Philadelphia, PA; Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia (J.T.G., H.H.); Department of Genetics, Harvard Medical School, Boston, MA (A.G.B., K.I., J.G.H., E.M., S.R.D., C.S., J.S.); Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA (J.G.H.); Department of Pediatrics, Mindich Child Health and Development Institute (L.R.-M., B.D.G.), Department of Genetics and Genomic Sciences (L.R.-M., M.F., B.D.G.), Icahn Institute for Genomics and Multiscale Biology (M.F.), and Psychiatric Genomics in the Department of Psychiatry (M.F.), Icahn School of Medicine at Mount Sinai, New York, NY; Department of Systems Biology (B.V., Y.S.), Department of Genetics and Development (in Medicine) (D.W.), and Department of Pediatrics and Medicine (W.K.C.), Columbia University Medical Center, New York, NY; Department of Genetics (S.J.S., A.J.W., M.W.S., R.L.), Department of Pediatrics (M.B.), and Department of Medicine (R.L.), Yale University, New Haven, CT; Department of Psychiatry, University of California, San Francisco (S.J.S., A.J.W., M.W.S.); Cold Spring Harbor Laboratory, NY (B.Y., M.R., I.I.); and Department of Cardiovascular Sciences, National Heart, Lung, and Blood Institute (NHLBI), NIH, Bethesda, MD (J.R.K.).

This article was sent to Gerald Dorn, Consulting Editor, for review by expert referees, editorial decision, and final disposition.

*These authors contributed equally to this article.

†These authors are co-senior authors.

The online-only Data Supplement is available with this article at http://circres.ahajournals.org/lookup/suppl/doi:10.1161/CIRCRESAHA.115.304458/-/DC1.

Correspondence to Hakon Hakonarson, Department of Pediatrics, Center for Applied Genomics, University of Pennsylvania, Philadelphia, PA 19104. E-mail hakonarson@email.chop.edu; or Wendy K. Chung, Department of Pediatrics and Medicine, Columbia University Medical Center, 630 W 168th St, New York, NY10032. E-mail Wkc15@Cumc.Columbia.Edu

Nonstandard Abbreviations and Acronyms	
CHD	congenital heart disease
CNV	copy number variant
ddPCR	digital droplet polymerase chain reaction
LOF	loss of function
LRR	log R ratio
OR	odds ratio
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
TOF	tetralogy of Fallot
WES	whole exome sequencing

Congenital heart disease (CHD) is the most frequent birth defect, affecting ≈7 in 1000 live births[1] and is a significant cause of childhood morbidity and mortality.[2] Rare Mendelian disorders, specific chromosomal abnormalities, and copy number variants (CNVs) are known to explain a subset of CHD cases,[2–4] but the cause of >80% of CHD remains unexplained.[5–12]

Editorial, see p 821
In This Issue, see p 811

The application of evolving technologies that detect structural variation throughout the genome has demonstrated a considerable contribution of CNVs to CHD. Early cytogenetic studies recognized an increased prevalence of de novo chromosomal abnormalities in syndromic patients with CHD, observations that were replicated and extended to nonsyndromic CHD with successive generations of CNV detection technologies including array comparative genomic hybridization and low-density single nucleotide polymorphism (SNP) arrays.[3–7,13–19] Using these techniques, researchers have demonstrated significant burden of large de novo CNV in some specific CHD lesions. Such CNVs are reported to occur in 13.9% of infants with single ventricles compared with 4.4% in controls,[13] in 10% of nonsyndromic tetralogy of Fallot (TOF) compared with 4% of controls,[5] and in 12.7% children with hypoplastic left heart syndrome compared with 2% of controls.[19] Among different CHD lesions, the frequency of large de novo CNVs is similar.[19] Although many large CNVs are unique to a single patient with CHD, several are recurrent in CHD cohorts. A 3-Mb 22q11.2 deletion is the most common recurrent de novo CNV associated with syndromic conotruncal defects and is found overall in ≥10% of TOF,[20,21] 35% of truncus, and 50% of interrupted aortic arch type B cases.[22] Recurrent de novo CNVs in patients with CHD reported in multiple studies also occur at chromosomes 1q21.1, 3p25.1, 7q11.13, 8p23.1, 11q24-25, and 16p13.11.[5,19]

The identification of CHD loci that are altered by CNVs provides opportunities to elucidate disease pathogenesis. However, discerning the causal gene(s) and inferring critical networks and pathways that cause or contribute to CHD has been difficult because low-resolution technologies used in many studies (array comparative genomic hybridization and low-density SNP arrays) typically define large CNVs (>100 kb) involving many genes. To address these issues, we capitalized on 2 independent strategies, high-density SNP genotyping arrays (Illumina Omni-1.0 and 2.5M) and whole exome sequencing (WES), to detect smaller de novo CNVs in a family-based trio study of sporadic CHD cases with conotruncal, heterotaxy, and left ventricular outflow tract defects.[23] We compared CNVs found in CHD trios with those identified in healthy control trios. Through these analyses we sought to compare the robustness of genome-wide CNV detection using array-based and sequence-based technologies to determine whether there was an increased burden of smaller de novo CNVs in patients with CHD as was demonstrated with larger CNVs and to determine whether fewer genes altered by these CNVs enabled more precise detection of gene networks and pathways contributing to the pathogenesis of CHD.

Methods

Ethics Statement

The protocol was approved by the Institutional Review Boards of Boston Children's Hospital, Brigham and Women's Hospital, Great Ormond St. Hospital, Children's Hospital of Los Angeles, Children's Hospital of Philadelphia, Columbia University Medical Center, Icahn School of Medicine at Mt. Sinai, Rochester School of Medicine and Dentistry, Steven and Alexandra Cohen Children's Medical Center of New York, and Yale School of Medicine. Written informed consent was obtained from each participating subject or their parent/guardian.

Patient Cohorts

CHD probands and parents were recruited into the CHD Genes Study of the Pediatric Cardiac Genomics Consortium (CHD genes: ClinicalTrials.gov identifier NCT01196182) as previously described,[23] using protocols approved by institutional review boards of each institution. Trios selected for this study had no history of CHD in first-degree relatives. CHD diagnoses were obtained from echocardiograms, catheterization, and operative reports; extracardiac findings were extracted from medical records and included dysmorphic features, major anomalies, noncardiac medical problems, and deficiencies in growth or developmental delay. The pathogeneses for CHD were unknown; patients with previously identified cytogenetic anomalies or pathogenic CNVs identified through routine clinical evaluation were excluded. Whole blood samples were collected and genomic DNA extracted.

CHD trios were studied by SNP arrays (n=414) or by WES (n=356), including a subset (n=233) that was analyzed by both methods. The distribution by CHD lesions in patients genotyped by arrays was 403 (61%) left ventricular obstruction, 197 (30%) conotruncal defect, 49 (7%) heterotaxy, and 12 (2%) other cardiac diagnoses (Online Table I). The distribution by CHD lesions in patients studied by WES was 284 (46.1%) left ventricular obstruction, 235 (38.1%) conotruncal defects, 78 (12.7%) heterotaxy, and 19 (3.1%) with other cardiac diagnoses (Online Table II).

Control trios were the unaffected sibling and parents of a child with autism who were consented and recruited through the Simons Simplex Collection. CNVs were identified in the same way in the control trios as in the cases using SNP arrays (n=814) or WES (n=872), including a subset (n=385) analyzed by both methods.[24–26]

Additional data on the distribution and prevalence of previously reported CNVs in the general population were derived from the Database of Genomic Variants (http://dgv.tcag.ca) and from 649 deidentified control subjects who had participated in an unrelated psychiatric case–control study, genotyped on the same high-density SNP array platforms at the same genotyping center as the CHD probands (438 on the Illumina Omni-1M and 211 on the Illumina Omni-2.5M). These controls were used only to prioritize the de novo CNVs identified by SNP array methods that were selected for confirmation analyses.

Array Genotyping and CNV Identification

A total of 360 CHD parental samples genotyped on the Omni-1M and 654 on Omni-2.5M arrays were applied for cluster definition using Illumina Genome Studio clustering algorithm. Raw data are publicly available through the database of genotypes and phenotypes National Heart, Lung, and Blood Institute Bench to Bassinet Program: The Pediatric Cardiac Genetics Consortium under database of genotypes and phenotypes Study Accession: phs000571.v1.p1.We removed clusters with outlier values of SNP call rate, Hardy–Weinberg equilibrium, AA/AB/BB cluster means, and minor allele frequency to improve the intensity noise (Log R ratio [LRR] SD) from a mean of 0.2 (using the default cluster file from Illumina) to 0.1 for CHD samples. Briefly, individual samples were filtered through a standard quality control pipeline.[24] B-allele frequency and LRR values were exported from Illumina Genome Studio. Only samples with SNP call rate >98%, SD of normalized intensity (LRR) <0.3, absolute value of guanine cytosine-corrected LRR <0.005, as well as CNV call count <800 for Omni1-Quadv1 or <300 for Omni2.5-8v1 were included.[27] Samples with high inbreeding coefficients that were duplicated or had sex mismatches and trios with Mendelian errors >1% were removed from analyses. We started with 1536 genotyped samples (512 trios), including 561 on the Illumina Omni-1M and 969 on the Illumina Omni-2.5M. Four hundred sixty-one trios had the same array version for all family members. On completion of these quality control procedures, 1245 samples, including 447 genotyped on the Illumina Omni-1M and 798 on the Illumina Omni-2.5M high-density SNP array platforms, were taken forward for analysis, constituting 415 complete trios (Online Table III).

Three groups (Children's Hospital of Philadelphia, Harvard, and Yale) independently analyzed genotyping data using slightly different algorithms to detect putative de novo CNVs. For each of the 3 independent analyses, CNVs were called for each subject using PennCNV[28] with the hidden Markov model algorithm and custom-made population frequency of B-allele and guanine cytosine model files. CNVs were called when ≥10 consecutive probes demonstrated consistent copy number change. The PennCNV detect_cnv --trio option was used to boost transmission probability of CNV calling for initially de novo scored CNVs. Fragmented CNV calls were merged using clean_cnv. All candidate CNVs were visually inspected to ensure the appropriate pattern of LRR and B-allele frequency was consistent with the CNV call. In addition, Gnosis,[24] QuantiSNP,[29] and Nexus (biodiscovery.com) were used to increase specificity. De novo CNVs were prioritized for quality by genomic length, number of probes, confidence score based on signal strength, 50% overlap of ≥2 algorithms, low parental origin *P* value using infer_snp_allele, and visual B-allele frequency/LRR review. CNVs with a minor allele frequency >1% were removed, leaving rare CNVs. All putative de novo CNVs were experimentally evaluated by digital droplet polymerase chain reaction (ddPCR; Online Figure I), and only validated CNVs are reported.

De novo CNV loci that were previously reported as pathogenic were defined by reported recurrence in ≥2 publications using independent data. Although some of the CNVs reported here overlap with previously reported CNVs in CHD patients based on review of the literature,[30] they do not meet our frequency constraint for previously reported pathogenic de novo CNV loci.

CNV Identification and Variant Calling From WES Data

WES data from 356 CHD trios were analyzed for de novo CNVs (Online Table II). WES samples were captured with the Nimblegen SeqCap Exome V2 chemistry and sequenced on the Illumina HiSeq 2000 platform as previously described.[31] Sequence reads were aligned to the human reference genome hg19 using Novoalign (http://novocraft.com), BWA (Burrows-Wheeler Aligner),[32] and ELAND (Efficient Large-Scale Alignment of Nucleotide Databases).[33] Duplicates were marked with Picard (http://picard.sourceforge.net). Indel realignment and Base Quality Score Recalibration were done with Genome Analysis Toolkit (GATK). eXome Hidden Markov Model (XHMM) is an algorithm to detect exon-level copy number variation and assign CNV quality metrics[34] and was used at 4 of the Pediatric Cardiac Genetics Consortium analysis sites (Children's Hospital of Philadelphia, Harvard, Columbia, and Mount Sinai) to identify de novo CNVs (Online Figure II). Candidate de novo CNVs were inspected visually. Putative de novo CNVs were prioritized for confirmation based on genomic length, low sequence depth variability, and low prevalence in the XHMM call set data (allele frequency<1%). All putative de novo CNVs were independently confirmed by ddPCR.

Single nucleotide variants (SNVs) and short insertions/deletions (indels) were called from the Novoalign alignment of WES trios using a pipeline derived from GATK version 2.7 best practices.[35] Briefly, aligned reads were first compressed using the GATK ReducedReads module, and variants were called on all CHD WES trios using the UnifiedGenotyper joint variant calling module. Identified variants were filtered using GATK variant quality score recalibration. Variants were annotated using SnpEff.[36] De novo SNVs and indels were independently confirmed using Sanger sequencing.

CNV Confirmation With ddPCR

Putative CNVs were experimentally confirmed with ddPCR, as previously reported,[37] using an 18- to 27-base pair FAM probe designed within each candidate CNV region, avoiding homopolymer runs or probes that began with G. A VIC probe targeting the RPP30 gene was used as reference. Reaction mixtures of 20-μL volume comprising ddPCR Master Mix (Bio-Rad), relevant forward and reverse primers and probe(s) and 100 ng of digested DNA were prepared, ensuring that ≈25% to 75% of the 10 000 droplets ultimately produced were positive for FAM or VIC signal. For de novo CNV confirmations, DNA from the patient with CHD and parents was used. After thermal cycling, plates were transferred to a droplet reader (Bio-Rad) that flows droplets single-file past a 2-color fluorescence detector. Differentiation between droplets that contain target and those that did not was achieved by applying a global fluorescence amplitude threshold in QuantaSoft (Bio-Rad). The threshold was set manually based on visual inspection at approximately the midpoint between the average fluorescence amplitude of positives and negative droplet clusters on each of the FAM and VIC channels. Confirmed CNV duplications had ≈50% increase in the ratio of positive to negative droplets as did the reference channel. Conversely, confirmed CNV deletions had approximately half the ratio of positive to negative droplets as did the reference channel.

Network Analysis

Three bioinformatic algorithms were used: Database for Annotation, Visualization and Integrated Discovery (DAVID),[38] Disease Association Protein-Protein Link Evaluator (DAPPLE),[39] and WebGestalt.[40] Four different gene lists derived from the de novo CNV loci were used (Online Table IV). The lists were constructed as follows: (1) all genes contained within de novo CNV intervals, (2) published causative genes from previously reported CHD CNVs intervals in addition to all genes in novel CHD CNV intervals. Causal genes in previously reported CNV intervals included *ELN* (Williams syndrome), *RAI1* (Smith–Magenis syndrome), *TBX1* (22q11 deletion), *GATA4* (8p23.1 deletion), *GJA5* (1q21.1 duplication), and *NKX2.5* (5q35.1 deletion); (3) genes contained solely within novel CHD CNV intervals (eg, exclude genes from previously published CNVs); (4) genes contained within de novo CNV intervals that are highly expressed in the developing mouse heart (top quartile of all genes expressed E14.5 mouse heart).[31] We anticipated that genes in list 2 and list 4 would have increased specificity for CHD in comparison with genes in list 1 and that genes in list 3 would be biased toward new disease networks.

We expanded network analysis input gene lists by including both de novo CNV genes and de novo SNVs that were previously identified in CHD probands by WES.[31] Only de novo SNVs predicted to be deleterious (eg, loss of function [LOF]: nonsense, frame-shift, and splice site mutations and missense variants that alter highly conserved amino acid residues or predicted to be deleterious by SIFT [Sorting Intolerant from Tolerant] or PolyPhen2) were included in the expanded gene list. The additional gene lists included (5) all genes within a de novo CNV interval (eg, list 1) and protein-altering SNVs

and (6) published causative genes from previously reported CHD CNVs intervals in addition to all genes in novel CHD CNV intervals (eg, list 2) and protein-altering SNVs.

Statistical Analysis

Burden calculations were done with a Fisher exact test computed in the R statistical computing environment. For analyses using DAVID, networks with an enrichment of genes affected by CNVs were assigned a *P* value with Benjamini and Hochberg correction for multiple testing with a false discovery rate of 0.05. In DAPPLE, type I error was controlled through permutation. *P* values of <0.05 were considered significant.

Results

Identification of De Novo CNVs

We studied 415 CHD trios genotyped by SNP arrays and 356 trios by WES analysis, including 233 trios studied by both methods. No trios had an affected first-degree relative, and the genetic cause of CHD in all studied children was unknown (Online Tables I and II).

Sixty-five de novo CNVs identified in CHD cases were independently confirmed by ddPCR (Table 1). De novo CNVs were identified in 51 unique probands (9.8%). These CNVs ranged in size from 0.1 kb to 12.8 Mb. Forty eight of these (76%) were <500 kb and half were smaller than 110 kb. The number of genes in the CNV intervals ranged from 1 to 175 with 42 (67%) having ≤5 genes. Four de novo intervals contained no genes. Six probands had 2 de novo CNVs, 2 had 3 CNVs and 1 had 4 CNVs.

The parental origin of deletion CNVs was determined when the haplotype of the remaining copy could be uniquely assigned to 1 parent. Seven de novo CNVs arose on maternal chromosomes and 10 on paternal chromosomes. The remainder could not be assigned because of uninformative or insufficient numbers of informative parent-of-origin SNPs.

Comparison of SNP Array and WES CNV Calling

To consider the accuracy of identifying de novo CNVs from SNP array data, we first considered a set of 40 high-confidence PennCNV de novo CNV calls that contained ≥10 adjacent SNPs, were >10 kb in length, and passed visual inspection. Among these 40 high-confidence putative CNVs, 40 were experimentally validated by ddPCR in the proband and 32 (80%) were experimentally confirmed to be de novo, representing a false-positive rate of 20%. For smaller de novo CNVs identified using the high-density array data, we considered a set of 97 high-confidence PennCNV putative de novo CNV calls based on 7 to 9 SNPs. Although 88% were experimentally validated by ddPCR in the proband, only 4 of the 97 (5%) were confirmed to be de novo.

From the WES data, we selected an initial set of 29 putative CNVs with a size range spanning 6 orders of magnitude from 530 bases in length (2 exons) to >8 Mb in length covering hundreds of exons. Twenty-six of the 29 CNVs (90%) confirmed experimentally to be de novo, representing a false-positive rate of 10%. The 3 false-positive CNVs included one 530-bp region that contained only 2 exon targets and 2 different inherited CNVs that were miscalled as de novo because both parents harbored CNVs at the locus. On the basis of these considerations, we restricted subsequent WES de novo CNV calls to those containing ≥3 exons and for which each parental data set contained no CNVs within the locus.

To evaluate false-negative rates of the 2 platforms and analyses, we tested our ability to detect 4 CNVs (2 22q11 deletions, 1 17p11 duplication, and 1 10q terminal deletion; Online Table V in clinical cases previously diagnosed with these CNVs). These 4 CNVs served as positive controls and were distinct from the Pediatric Cardiac Genetics Consortium cohort. Both the SNP array and the WES platforms detected each of these 4 large, clinically significant CNVs.

We also compared the results of de novo CNVs analysis from the 233 trios studied by both SNP array and WES. Among 42 confirmed de novo CNVs in these trios, 24% (10/42) were identified by both platforms, whereas 40% (17/42) were identified only with the SNP arrays and 35% (15/42) only by WES (Figure 1). The recognized technical limitations of each platform prevented detection of some CNVs. For example, CNVs that occur exclusively in noncoding sequences are not captured by WES, whereas CNVs in coding or noncoding genomic regions where the SNP density is sparse can escape detection by SNP arrays.

From our studies, we deduced that de novo CNVs were accurately detected by arrays when ≥10 adjacent SNPs were affected or by WES when >3 adjacent exons were affected. In our data set, 29 of 42 CNVs fulfilled both of these criteria and should have been identified by both technologies (Figure 1). However, only 34% (10/29) of these ddPCR confirmed CNVs were identified by both platforms. SNP arrays uniquely identified 34% (10/29) and WES analyses uniquely identified 31% (9/29). Taken together, the false-negative rate of each methodology is ≈30% to 35%. Overall, the genome-wide analyses of de novo CNVs identified by SNP arrays were reasonably concordant with WES data, but each also identified complementary CNVs. The minimum CNV size that we reliably detected by SNP arrays was 10 kb and by WES was 1 kb although some smaller CNVs identified by these techniques were validated.

CNV Burden Analysis

The burden of de novo CNVs in CHD cases and control trios was initially compared using analyses from SNP arrays. De novo CNVs were assessed in 841 control trios, studied using the Illumina Omni-1M array to match the case trio array resolution and called using the PennCNV algorithm using computational parameters described previously.[24] Nine de novo ddPCR-validated CNVs were identified among 841 control trios. Twenty-two de novo ddPCR-validated CNVs were identified among 462 CHD trios with SNP arrays. These data define a significant burden of CNVs in CHD cases when compared with controls (odds ratio [OR], 4.6; Fisher $P=7\times10^{-5}$; Table 2). After excluding 9 previously identified CHD-associated CNVs, the calculated burden of novel CNVs identified in CHD cases remained modestly significant (OR, 2.7; Fisher $P=0.02$).

To provide further support for this finding, we analyzed the burden of de novo CNVs that were identified by WES. WES in CHD cases and control trios were technically

Table 1. Confirmed Rare De Novo CNVs in Discovery Cohort

ID	Chr	Start	End	Band	CNV*	Syndrome/Gene	Analysis Observed†	Cardiac Lesion: (Diagnosis)	Parent Origin	Extracardiac	Genes, n	Size, kb
1-01401	1	59 247 993	59 251 097	p32.1	1	*JUN*	A	LVOT (HLHS)	...	...	1	3.1
1-03171	1	145 586 403	145 799 634	q21.1	3	1q21.1 dup/*GJA5*‡	A E	CTD (TOF/ APVS)	...	...	7	213.2
1-01036	1	146 631 133	147 416 212	q21.1	3	1q21.1 dup/*GJA5*‡	E	CTD (TOF)	M	...	15	785.1
1-01486	1	194 201 171	194 304 070	q24.2–q25	3	*CDC73*	A	LVOT (HLHS)	...	Yes	0	102.9
1-01518	1	248 750 565	248 795 110	q44	3	*OR2T10, OR2T11*	A	LVOT (HLHS)	...	...	2	44.5
1-01536	2	70 168 995	70 359 345	p13.3	1	*PCBP1*	A	CTD (TOF/ PA)	...	...	5	190.4
1-01401	2	102 493 466	103 001 458	q11.2–q12.1	1	*MAP4K4*	E	LVOT (HLHS)	...	...	6	508.0
1-01401	2	145 155 868	145 274 931	q22.3	1	Mowat-Wilson/*ZEB2*‡	E	LVOT (HLHS)	...	...	1	119.1
1-00762	3	60 661	11 712 230	p26.1	3	*ARL8B, ARPC4, CAMK1, CAV3, CRBN, EMC3, ITPR1, SEC13, SETD5, VGLL4*	A	ASD/PS (ASD)	...	Yes	103	11651.6
1-01049	3	15 637 812	15 643 461	p25.1	3	*BTD, HACL1*	E	CTD (TOF)	...	...	2	5.6
1-01045	3	477 80 965	48 309 270	p21.31	3	*CDC25A, DHX30, MAP4, SMARCC1*	A	LVOT (HLHS)	...	...	14	528.3
1-02093	3	197 143 652	197 186 111	q29	3	*BDH1*	A	CTD (TOF/ PA)	...	Yes	0	42.5
1-00771	4	185 603 346	185 638 397	q34.1	1	*CENPU, PRIMPOL*	E	CTD (DTGA/ VSD)	P	Yes	2	35.1
1-00789	5	136 464	232 969	p15.33	3	*CCDC127, LRRC14B, PLEKHG4B, SDHA*	A	CTD (TOF)	...	...	4	96.5
1-00113	5	133 706 994	133 730 455	q31.1	1	*UBE2B*	A	CTD (TOF/ PA)	...	Yes	1	23.5
1-00296	5	166 386 727	173 073 664	q34–q35.2	1	*NKX2.5*‡	A	CTD (TOF)	M	Yes	53	6686.9
1-01916	6	36 646 788	36 651 971	p21.2	1	*CDKN1A*	A	HTX (HTX)	...	...	1	5.2
1-01049	6	43 484 783	43 485 159	p21.1	3	*POLR1C*	E	CTD (TOF)	...	...	1	0.4
1-00096	7	50 179 707	50 191 153	p12.2	1	*C7orf72*	E	CTD (TOF/ PA)	...	Yes	1	11.4
1-00800	7	72 719 386	74 138 603	q11.23	1	Williams syndrome‡	A	CTD (VSD/ PS)	P	Yes	34	1419.2
1-00540	7	72 721 123	74 140 708	q11.23	1	Williams syndrome‡	A	LVOT (ASD)	M	Yes	34	1419.6
1-00977	7	138 258 252	143 807 632	q24–q25	1	*C7orf55, FAM115A, LUC7L2, MKRN1, NDUFB2, UBN2, ZC3HAV1L, ZYX*	E	CTD (TOF)	...	...	175	5549.4
1-01995	7	142 334 207	142 460 871	q34	1	*MTRNR2L6, PRSS1*	E	CTD (TOF)	M	...	15	126.7
1-01562	8	8 067 768	12 530 976	p22.1–p23.1	3	*GATA4*‡	A	CTD (TOF)	...	...	75	4463.2
1-02625	8	8 102 183	12 190 106	p23.1	3	*GATA4*‡	A	LVOT (CoA)	M	Yes	62	4087.9
1-00566	8	11 606 428	11 710 963	p23.1	1	*GATA4*‡	A E	CTD (TOF)	...	...	6	104.5
1-00948	8	119 053 343	119 064 098	q24.1	1	*EXT1*	A	LVOT (CoA)	P	Yes	1	10.8
1-02360	9	5 302 500	5 337 760	p24.1	3	*RLN1, RLN2*	A	CTD (ASD)	...	Yes	3	35.3
1-01852	11	34 458 230	34 460 862	p13	1	*CAT*	A	CTD (VSD)	...	...	1	2.6

(*Continued*)

Table 1. Continued

ID	Chr	Start	End	Band	CNV*	Syndrome/Gene	Analysis Observed†	Cardiac Lesion: (Diagnosis)	Parent Origin	Extracardiac	Genes, n	Size, kb
1-00565	11	42 968 283	42 970 488	p12	3	*HNRNPKP3*	A	LVOTO (ASD)	...	...	0	2.2
1-01536	11	65 157 239	65 408 708	q13.1	1	*EHBP1L1, LTBP3, MAP3K11, PCNXL3, SCYL1, SSSCA1*	A	CTD (TOF/PA)	...	...	14	251.5
1-00230	11	86 939 592	87 025 456	q14.2	1	*TMEM135*	A E	LVOT (ASD)	P	Yes	1	85.9
1-01486	11	125 641 368	134 943 190	q24.2–q25	1	Jacobsen/*ETS1*‡	A E	LVOT (HLHS)	P	Yes	73	9301.8
1-00795	11	134 598 043	134 617 838	q25	3	*LOC283177*	A	CTD (VSD)	M	...	0	19.8
1-00124	12	8 003 758	8 123 306	p13.31	3	*SLC2A14, SLC2A3*	A	LVOT (ASD/HLHS)	...	...	3	119.5
1-00050	12	52 845 952	52 862 783	q13.13	1	*KRT6C*	A	LVOT (HLHS)	...	...	1	16.8
1-02411	14	58 860 893	58 881 694	q23.1	1	*TIMM9, TOMM20L*	A	CTD (TOF)	...	...	2	20.8
1-01049	14	74 551 632	74 551 731	q24.3	3	*LIN52*	E	CTD (TOF)	...	...	1	0.1
1-00192	15	22 296 985	23 161 330	q11.2	3	1 MB from *PW/CYFIP1*‡	A	LVOT (CoA)	...	...	20	864.3
1-00315	15	22 750 305	23 140 114	q11.2	3	1 MB from *PW/CYFIP1*‡	A	LVOT (CoA)	M	...	5	389.8
1-01396	15	22 750 305	23 228 712	q11.2	1	1 MB from *PW/CYFIP1*‡	A E	CTD (TOF/PA)	P	...	6	478.4
1-00243	15	22 835 893	23 062 345	q11.2	1	1 MB from *PW/CYFIP1*‡	E	LVOT (CoA)	P	Yes	4	226.5
1-01994	15	28 389 771	28 446 734	q13.2	1	*HERC2*	E	LVOT (ASD)	P	...	1	57.0
1-01696	15	44 833 588	44 856 873	q21.1	1	*EIF3J, SPG11*	A E	CTD (TriAtresia/DTGA)	...	...	2	23.3
1-01941	15	88 761 539	88 779 300	q25.3	3	*NTRK3*	A	CTD (TOF/DTGA)	P	-	1	17.8
1-01427	17	21 562 473	22 252 439	p11.2	1	*FAM27L, FLJ36000, MTRNR2L1*	A	HTX (HTX)	...	Yes	7	690.0
1-00561	17	27 962 393	28 099 002	q11.2	1	*SSH2*	A	LVOT (ASD)	...	Yes	3	136.6
1-01995	17	38 544 624	38 548 586	q21.1	1	*TOP2A*	A E	CTD (TOF)	...	...	1	4.0
1-01049	17	39 845 210	39 846 477	q21.2	3	*EIF1*	E	CTD (TOF)	...	...	2	1.3
1-01588	18	65 138 642	78 015 180	q22.1–q23	1	*NFATC1*‡	A	LVOT (CoA)	...	Yes	58	12 876.5
1-02170	19	20 601 006	20 717 536	p12	1	*ZNF826P*	A	CTD (TOF)	...	Yes	1	116.5
1-00174	19	40 515 744	40 681 387	q13.2	1	*ZNF546, ZNF780A, ZNF780B*	A	CTD (TOF/PA)	...	Yes	4	165.6
1-01536	19	47 792 293	47 905 132	q13.33	1	*C5AR1, C5AR2, DHX34*	A	CTD (TOF/PA)	...	...	3	112.8
1-00730	20	14 529 657	14 583 899	p12.2	1	*MACROD2, MACROD2-IT1*	A	CTD (DTGA)	...	...	2	54.2
1-01194	22	18 844 632	21 500 000	q11.2	1	DiGeorge/*TBX1*‡	A	CTD (VSD)	P	Yes	80	2655.4
1-00113	22	18 886 915	22 000 000	q11.2	1	DiGeorge/*TBX1*‡	A E	CTD (TOF/PA)	P	Yes	96	3113.1
1-01836	22	19 020 529	21 380 382	q11.2	1	DiGeorge/*TBX1*‡	A E	CTD (TOF)	M	...	70	2359.9
1-00988	22	20 733 495	21 464 479	q11.2	1	DiGeorge/*TBX1*‡	A	CTD (HLHS/HTX)	M	Yes	31	731.0
1-02133	22	25 661 725	25 919 492	q11.23	3	22q11 distal duplication‡	A	CTD (TOF)	...	...	4	257.8

(*Continued*)

Table 1. Continued

ID	Chr	Start	End	Band	CNV*	Syndrome/Gene	Analysis Observed†	Cardiac Lesion: (Diagnosis)	Parent Origin	Extracardiac	Genes, n	Size, kb
1-00425	22	36 038 076	36 149 338	q12.3	1	*APOL5, APOL6, RBFOX2*	A E	LVOT (HLHS)	...	...	4	111.3
1-01427	22	42 522 638	42 531 210	q13.2	3	*CYP2D6*	A	HTX (HTX)	...	Yes	2	8.6
1-01941	X	23 003 525	23 086 619	p22.11	3	*DDX53, RP11-40F8.2*	A	CTD (TOF/DTGA)	...	...	1	83.1
1-00197	X	148 685 645	148 693 146	q28	3	*TMEM185A*	E	LVOT (HLHS)	...	Yes	1	7.5

Genomic coordinates refer to hg19. APVS indicates absent pulmonary valve syndrome; ASD, atrial septal defect; CNV, copy number variant; CoA, coarctation of the aorta; CTD, conotruncal defect; DTGA, dextro-transposition of the great arteries; HLHS, hypoplastic left heart syndrome; HTX, heterotaxy; LVOT, left ventricular outflow tract obstruction; M, maternal chromosome; P, paternal chromosome; PA, pulmonary atresia; PS, pulmonary stenosis; TOF, tetralogy of Fallot; TriAtresia, tricuspid atresia; and VSD, ventricular septal defect.

*Copy number: 1, deletion and 3, duplication.

†Analysis: A, identified with single nucleotide polymorphism array and E, identified with whole exome sequencing.

‡De novo CNV loci that were previously reported as pathogenic.

comparable, including the same Nimblegen V2 exome capture chemistry and similar sequence read depths obtained on identical Illumina platforms. Sixty percent of control trios were sequenced at the same site (Yale Center for Genome Analysis) that sequenced the cases. Raw sequence reads were processed through the identical short read aligner (Novoalign) for CNV burden analysis. SNP genotyping of CHD and control data sets and principal component analysis did not identify any systematic biases (Online Figure V). Cases and controls were matched for sex as best as possible with slight excess of male cases. Using an identical XHMM pipeline (CNVs involving ≥3 exons and no parental CNVs within 1 MB), we identified 19 de novo ddPCR-validated CNVs in 356 CHD trios, and 14 de novo ddPCR-validated CNVs in 872 control trios (OR, 3.5, Fisher $P=6\times10^{-4}$; Table 2). Excluding the 6 de novo CNVs previously identified as CHD associated, we identified a similar OR and *P* value as in the SNP array data (OR, 2.3; Fisher *P*=0.03).

Our data identify an increased burden of CNVs, detected by SNP arrays or WES, in patients with CHD when compared with controls. We observed a larger mean size of de novo CNVs with increased burden in patients with CHD (3.6 Mb) than controls (495 kb; *t* test *P*=0.035) with the distribution of CHD CNVs skewed toward the largest CNVs identified in CHD cases. The median size of de novo CNVs from CHD cases (522 kb) was also significantly larger than controls (118 kb; Mann–Whitney *P*=0.028). Of the CNVs identified by SNP array that were capable of detecting CNVs outside of coding regions, there was a trend toward an increased number of de novo CNVs in controls that contained no coding exon (4/9) when compared with Pediatric Cardiac Genetics Consortium cases (3/22; Fisher *P*=0.15).

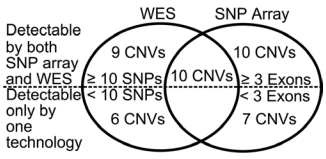


Figure 1. Comparison of single nucleotide polymorphism (SNP) array and whole exome sequencing (WES) platforms in detection of the 42 validated de novo copy number variants (CNVs) in the subset of 233 probands studied by both technologies. Ten of the 42 were detected by both methods, 32 were called by 1 method. The figures below the dotted line show the number of CNVs that were below the detection limits of the second platform (CNVs that span <10 SNPs on SNP arrays or <3 exons on WES) and hence could not have been called. The figures above the dotted line show the number of CNVs with sufficient SNPs and exons to enable high confirmation rates, but that were not called.

Putative CHD Loci at 15q11.2 and 2p13.3

Overlapping de novo CNVs found in multiple cases and not in controls likely contain disease genes. Sixteen of 65 (25%) de novo CNVs in CHD probands have been previously implicated in CHD,[5] including 4 22q11.2 deletions, 3 8p23 deletions (involving *GATA4*), 2 1q21.1 duplications (involving *PRKAB2*, *PDIA3P*, *FMO5*, *CHD1L*, *BCL9*, *ACP6*, and *GJA5)*, 1 22q11.2 distal duplication, 1 2q22.3 deletion (that causes Mowat–Wilson syndrome), 1 11q24.2-q25 deletion (that causes Jacobsen syndrome), and 4 with CNVs in 15q11.2.

CNVs in 4 CHD probands (2 deletions and 2 duplications) at the BP1-BP2 15q11.2 locus that spans ≈225 kb (chr15: 22 836 000–23 062 000) were observed as recurrent de novo events (Online Figures III and IV). Both patients with duplications (1-00192, 1-00315) and 1 with a deletion (1-00243) had left ventricular obstruction because of aortic coarctation. The remaining proband (1-01396) had TOF with pulmonary atresia. As there was no de novo CNV identified in this region among 814 and 872 control trios studied, respectively, by SNP arrays or WES, this locus has a significant burden of de novo CNVs in CHD cases (4/538 CHD versus 0/1301 controls; Fisher *P*=0.007). CNVs at the 15q11.2 locus were observed at low frequency (allele frequency<1%) in the Database for Genomic Variants. Among the 3 genes altered by this CNV (*CYFIP1*, *NIPA1*, and *NIPA2*), only *CYFIP1* is highly expressed in the developing mouse heart.[31] *CYFIP1* encodes the cytoplasmic FMR1-interacting protein 1, which has dual roles in inhibiting local protein synthesis and in promoting actin remodeling.[41] An earlier study observed an increased burden of inherited deletions in CHD cases at 15q11.2[1] and a recent article identified a single

Table 2. Case Control De Novo CNV Burden

	Probands (n)	CNVs, n (%)	OR	*P* Value
SNP array				
SSC*	841	9 (1%)	...	...
PCGC: all CNVs	462	22 (4.7%)	4.6	7×10^{-5}
PCGC: novel loci	...	13 (2.8%)	2.7	0.02
WES				
SSC†	872	14 (1.6%)	...	...
PCGC: all CNVs	356	19 (5.6%)	3.5	6×10^{-4}
PCGC: novel loci	...	13 (3.9%)	2.3	0.03

CNV indicates copy number variant; OR, odds ratio; PCGC, Pediatric Cardiac Genetics Consortium; SNP, single nucleotide polymorphism; SSC, Simons Simplex Collection; and WES, whole exome sequencing.

*Controls derived from Sanders et al.[24]

†Controls derived from 3 studies: Iossifov et al[71]; Sanders et al[26]; and an additional set of unpublished controls provided by State selected by the same criteria and sequenced as described in ref. 26.

proband with a 6-Mb de novo duplication at 15q11.2-q13.1[20] and 2 additional cases with inherited 300- to 400-kb duplications at 15q11.2. Our data provide additional evidence that de novo CNVs at 15q11.2 may contribute to disease risk in CHD.

In addition, a recurrent CNV was observed to alter a novel locus at chromosome 2p13.3. A de novo 190-kb deletion was identified in a TOF proband (1-01536) and was maternally inherited in a proband with truncus arteriosus (1-01805). No 2p13.3 CNV was found in control samples or in Database for Genomic Variants. Among 3 genes included in the CNV interval (*ASPRV1*, *PCBP1*, and *PCBP1-AS1*), only *PCBP1* is highly expressed in the developing mouse heart.[31] *PCBP1* encodes a major cellular poly(rC)-binding protein, which controls translation from mRNAs containing the differentiation control element.[42] In DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources), patient 257771 with an atrioventricular canal defect had a 7-Mb overlapping deletion of 2p13.3, suggesting that this locus may also contribute to disease risk in CHD.

Integration of CNV and Sequence Data to Identify CHD Genes

To improve the identification of specific genes altered by CNVs that might cause or contribute to CHD, we searched the WES data for de novo, rare LOF variants in genes encoded in CNV intervals. We identified a terminal deletion of chromosome 11q24.2-q25, which causes Jacobsen syndrome in 1 patient with CHD (1-01486) with clinical manifestations typical of this dominant disorder (hypoplastic left heart, coarctation of the aorta, mitral and aortic valve atresia, strabismus, and short stature). *ETS1* has been proposed as the critical CHD gene in the Jacobsen syndrome locus based on impaired ventricular development in an *Ets1*-null mouse.[43] WES analyses identified a de novo *ETS1* frameshift mutation (chr11: 128 350 159GTCCT>G, c.1046_1049delAGGA [p.K349fs]) in another patient with CHD without the chromosome 11q24.2-q25 deletion with cardiac abnormalities observed in Jacobsen syndrome (hypoplastic left heart and mitral valve atresia). Our data provide the first human genetic evidence to suggest that *ETS1* mutations contribute to the cause of cardiac malformations in Jacobsen syndrome.

We also assessed whether de novo CNVs in combination with a rare or novel deleterious variant on the other allele might produce recessive forms of CHD. One patient with CHD (1-01179) with a de novo 10q25-26 deletion also had a novel *CTBP2* variant (p.R134W) on the remaining allele. The hemizygous variant that was absent from public genome databases[44,45] is predicted to be damaging (Polyphen2 score of 0.998) and altered a phylogenetically conserved residue (PhyloP score=2.54). Cardiac abnormalities are present in approximately one third of patients with subterminal chromosome 10q deletions and recently *CTBP2* was proposed as

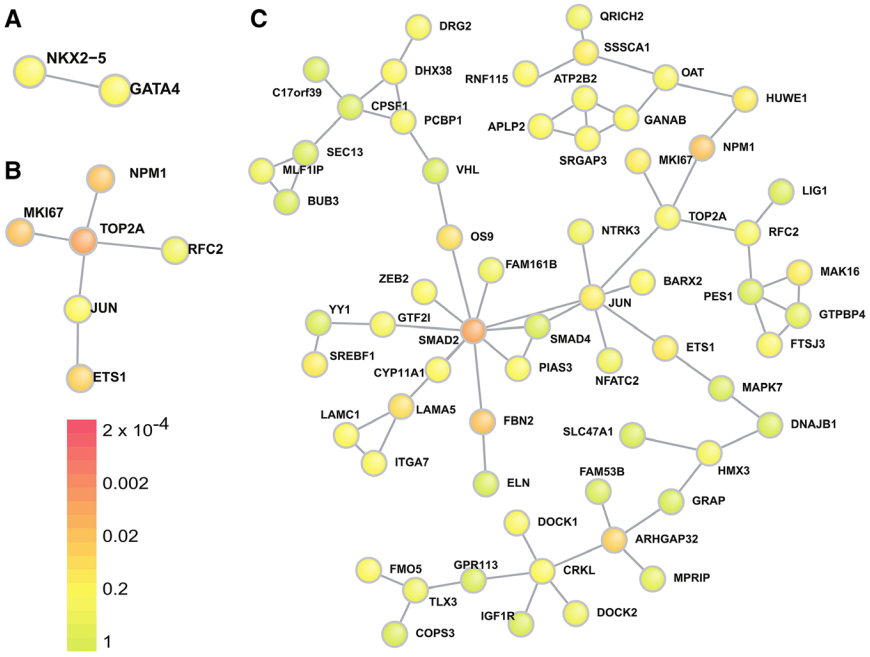


Figure 2. Network analysis of copy number variant (CNV) loci genes. Two networks of direct protein–protein interactions, (**A**) NKX2.5/Gata4 and (**B**) ETS1/JUN/TOP2A, were consistently identified in the DAPPLE (Disease Association Protein-Protein Link Evaluator) de novo CNV loci analysis. *P* values from the genes highly expressed in the developing heart, the most restrictive gene set list, are presented here. **C**, The ETS1/JUN/TOP2A network was significantly elaborated on by incorporating genes with deleterious de novo point mutations and indels in the whole exome sequencing analysis in addition to the CNV loci. Of note, 2 probands had de novo ETS1 variants (1 CNV and 1 frameshift), 2 probands had de novo SMAD2 variants (a splice site mutation and a highly conserved missense variant), and 2 probands had de novo ELN variants (both Williams syndrome CNVs).

a candidate CHD gene.[46] The clinical manifestations of our patient, truncus arteriosus and right aortic arch, resemble the phenotypes identified in a *Ctbp2*-null mouse (failure of vascular remodeling and cardiac looping).[47] We suggest that *CTBP2* sequence analyses in individuals with chromosome 10q deletions may identify additional variants in a subset of patients that modify phenotype.

Correlation of CHD Phenotypes and CNVs

The frequency of de novo CNVs was 10% among conotruncal anomalies, 6% among left-sided obstructive lesions, and 21% in heterotaxy. We observed a modest trend toward increased extracardiac manifestations, such as developmental delay in patients with de novo CNVs (Online Table VI). Approximately 31% of all patients with CHD studied with SNP arrays or WES had extracardiac manifestations, whereas 40% (21/52; OR, 1.5; Fisher P=0.2) of patients with de novo CNVs had extracardiac features. This association has been found in some,[48] but not all,[19] previous studies, perhaps because of differences in the ages of the patients with CHD studied, methods of clinical data collection, and the definition of an extracardiac anomaly.

Gene Networks Affected by CNVs in CHD

We used pathway and network analysis with DAVID,[38] DAPPLE,[39] and WebGestalt,[40] using as input 4 different lists of genes encoded within all de novo CNV loci (Methods in the Online Data Supplement; Online Table IV). Initial gene lists contained (1) all genes encoded in a de novo CNV interval; (2) genes previously defined as causative with CNVs intervals plus all genes in novel de novo CNV intervals; (3) only genes contained within novel de novo CNV intervals; (4) all genes contained within de novo CNV intervals that are highly expressed (top 25%) in the developing heart.[31]

DAVID identified enrichment of a gene pathway implicated in acetylation $P<2.3\times10^{-4}$, phosphoprotein $P<3.9\times10^{-4}$, and G-protein–activated inward rectifier potassium channel $P<2.5\times10^{-2}$ (Benjamini–Hochberg corrected). WebGestalt implicated an enrichment of previously identified CHD genes, including *ELN*, *NKX2.5, GATA4*, and *ZEB2*, contributing to Gene Ontology processes: anatomic structure formation involved in morphogenesis $P<0.03$, cardioblast differentiation $P<0.03$, and septum secundum development $P<0.02$ (Benjamini–Hochberg corrected).

Using DAPPLE, we identified 2 additional subnetworks of direct protein/protein interactions that were consistently observed across 4 gene lists. Among genes encoded within CNVs that are highly expressed in the developing heart, a subnetwork consisting of *NKX2.5* and *GATA4* ($P<0.1$; Figure 2A) and a subnetwork consisting of *ETS1*, *JUN*, *TOP2A*, and *MKI67* ($P<0.01$; Figure 2B) were identified. By further expanding the CNV gene lists to include genes with de novo LOF mutations, the *ETS1/JUN/TOP2A* subnetwork was significantly elaborated on and enriched ($P<0.005$). Each of these 3 genes was directly linked through protein–protein interactions to subnetworks containing ≥10 additional genes identified in either CNV or WES data sets.[31] This entire network incorporated >60 genes implicated in CHD (Figure 2C). Because the *ETS1/JUN/TOP2A* subnetwork was robust to the specific de novo CNV gene list (criteria 2 above) and expanded with the addition of genes containing rare de novo LOF mutations, the data suggest that this subnetwork contains genes and pathways involved in CHD.

Discussion

We report whole-genome CNV analyses using complementary detection technologies in a large cohort of patients with CHD. CNV detection in WES has been investigated in schizophrenia[34] and autism,[49] but array-based and sequence-based strategies have not previously been directly compared, and our data highlight the differences between array-based and sequence-based strategies to detect de novo CNVs. By defining small CNVs with high resolution and integrating these findings with WES data that identified rare deleterious mutations, we identified novel de novo CNVs and genes involved in the pathogenesis of CHD. We show that 9.8% (53/538) of patients with CHD without a previously identified genetic pathogenesis have rare de novo CNVs. We previously demonstrated that 10% of patients with CHD in our cohort have de novo single nucleotide or small insertion/deletion mutations in genes highly expressed in the developing heart that are likely to be damaging.[31] None of the patients with CHD with rare de novo CNVs reported here carry these variants. Even if all the de novo CNVs and de novo predicted pathogenic sequence variants we have identified were causative, we do not yet know the pathogenesis for the majority of CHD subjects in our study.

Our detection rate of ≈10% de novo CNVs in patients with CHD is equivalent to previous studies,[5,19,48] despite identifying small CNVs. Had we not excluded patients with known pathogenic CNVs identified through clinical care, we expect that de novo CNVs would have been identified in ≈15% of patients with CHD, based on the prevalence of common de novo CNVs in CHD (eg, 7% of TOF with chromosome 22q11 deletions and 1% of TOF to 1q21 CNVs). In our study, these CNV loci accounted for <1% of CHD probands.

Despite these exclusion criteria, we identified a 4-fold increased frequency of rare de novo CNVs relative to the background frequencies of 1.2% (detected by SNP arrays) and 1.8% (detected by WES) of de novo CNVs in controls ($P=7\times10^{-5}$, $P=4\times10^{-4}$, respectively). Even after excluding previously defined CNVs, we still observed an ≈2-fold increase in novel rare de novo CNVs (P=0.02).

Because the OR of de novo CNVs in cases versus controls was 3.5 to 4.6, we estimate that between 50% and 70% of de novo CNVs observed in cases may be disease causing. The possibility exists that a higher percentage of de novo CNVs increase the risk of CHD but may not be sufficient to cause CHD without other contributing genetic or environmental factors. In addition, subtle anatomic defects in the heart may not have been diagnosed in the control group because controls were not systematically examined by echocardiogram. Overall, our evidence suggests a model in which de novo CNVs contribute to CHD.

The comparison of dense array-based platforms and WES analyses to detect independently validated CNVs indicates that each strategy identifies only ≈70% of the CNVs that should be within the detection limitations of each technology. As such, these 2 CNV methodologies provide substantial complementary information. An important corollary to this conclusion is that previously published CNV analyses in human disease may have significantly underestimated the burden conveyed by these structural variants.

Among all confirmed de novo CNVs, 61% (41) were deletions and 39% (26) were duplications. The proportion of these classes of CNVs are not significantly different; whether the trend toward more CNV deletions in CHD is biologically meaningful or reflects greater sensitivity to detect deletions by these methods will require further analyses. De novo CNVs ranged in size from <1 kb to 12.8 Mb, with a median size of 110 kb. Thus, half of the independently confirmed CNVs were smaller than the reported detection limit of most previous studies. For example, 4 patients with CHD had 200-kb de novo CNVs on chromosome 15. Although the pathogenicity of the identified CNVs remains to be determined, we propose that the smaller CNVs involving fewer genes are particularly valuable in defining specific candidate CHD genes in comparison with larger CNVs that typically include many more candidates. The ability to detect small CNVs is helpful reliably, particularly if they fall within large CNVs previously identified and define a critical interval of overlap. For example, we identified 1 de novo CNV that only affected *JUN* and another that only altered *TOP2A*, 2 genes that were implicated by network analyses as interacting with transcription factors *SMAD2*, *SMAD4*, and *ETS1*, molecules that play important roles in cardiovascular development.

Although there is considerable complexity in CHD phenotypes, we observed no significant difference in the frequencies of rare de novo CNVs among distinct CHD subclassifications. Although patients with CHD and CNVs in our cohort were more likely to have extracardiac phenotypes (OR, 1.5), this trend fell short of significance. Whether this finding reflects shared developmental biological pathways among different organ systems or the possibility that CNVs perturb multiple genes that individually contribute to organ system development remains unknown.

We identified several de novo CNVs that affected established CHD genes, including *GATA4* and *GJA5*. We also identified a patient with CHD and a deletion of chromosome 5q34-q35.2, encompassing *NKX2-5*. LOF *NKX2-5* mutations are an established cause of CHD,[3,5,6,11,12,50] and CNVs encompassing *NKX2-5* have been previously recognized in CHD.[18,51,52]

We identified recurrent de novo CNVs involving deletions or duplications at chromosome 15q11.2. As the proximal region of chromosome 15 is meiotically unstable because of the segmental duplications that serve as breakpoint hotspots, recurrent de novo events at this locus might reflect locus genomic instability. However, the excess burden of de novo CNVs at this locus in patients with CHD when compared with controls (Fisher P=0.007) suggests otherwise. The report of an excess burden of inherited deletions in patients with CHD at this locus[3] lends further evidence for pathogenicity although this study lacked information on inheritance. Because CNVs at chromosome15q11.2 CNV exhibit incomplete penetrance for both neuropsychiatric and CHD phenotypes, genes affected by this could participate in inherited and sporadic CHD.

Chromosome 15q11.2 deletions and duplications are implicated in neurodevelopmental disorders, including schizophrenia, intellectual disability, and autism.[53–55] That chromosome 15q11.2 CNVs are also associated with CHD adds to a growing list of loci (22q11,[56] 1q21,[5,57] 7q11.23,[58] 16p11.2,[59,60] and 16p13.11[19,60]) that link cardiac malformations and neurocognitive disorders. These (and other) genetic loci may explain, in part, the significant coexpression of heart and brain developmental phenotypes in many children.

By integrating CNV and sequencing data from WES, we also identified candidate genes within CNV regions that may cause dominant or recessive forms of CHD. We present the first human *ETS1* LOF mutation that likely contributes to Jacobsen syndrome. We also identified a rare inherited and predicted deleterious *CTBP2* missense variant that is hemizygous because of a de novo CNV deletion, associated with a CHD phenotype comparable with that observed in *Ctbp2*-null mice. Continued integration of CNV and sequence data should enable more comprehensive assessments of genetic causes of disease. The current study provides suggestive data, and sequencing large cohorts of patients with CHD for mutations in these 2 genes will be necessary to prove the role of these genes in CHD unambiguously.

Network analyses by DAPPLE was more successful in elucidating novel network biology than DAVID and WebGestalt, which rely heavily on previously annotated gene sets and are challenged by the addition of unrelated genes encoded with CNV intervals along with pathogenic genes. If pathogenic CNVs on average contain 1 main causal gene and ≈5 unrelated genes, then we might expect DAVID and WebGestalt to be less informative for CNV network analyses.[61] Conversely, DAPPLE, based on proteome-wide protein–protein interaction data rather than previously curated gene lists, calculates P values through within-degree node-label permutation, which is more permissive to background noise.[39]

DAPPLE network analysis reinforced the central role of transcriptional regulation in CHD. The identification of 1 network, including NKX2.5/GATA4, provided a robust positive control as protein–protein interactions and substantial contributions by these molecules to CHD are previously described.[62,63] Direct protein–protein interactions between ETS1/JUN/TOP2A have also been reported,[64–66] but this network has not been previously implicated in CHD. In an expanded network analysis of these molecules that included rare LOF mutations identified from exome sequencing, JUN was linked to SMAD2 and SMAD4, molecules that participate in cardiac development through the transforming growth factor-β signaling pathway.[67–70]

We focused our current analysis solely on rare de novo CNVs. Because the pathogenesis of CHDs is known to be polygenic, and incomplete penetrance of genes for CHD has been previously described, future analyses of rare inherited CNVs may expand these findings.

The novel de novo CNVs we report should be considered provisional pending replication in independent studies. Replication of the overall effect and the magnitude of the risk of these identified variants is needed. Although it is not yet possible to draw a conclusion about whether any particular de novo CNV is causal, the identification of additional CNVs and mutations in specific genes within the CNV intervals will be required to validate the new loci identified here.

In summary, integration of high-resolution complementary platforms for CNV and sequence data on large numbers of patients with CHD has proven valuable to define the underlying genomic architecture of CHD and expand the genes and networks involved in cardiac development and is likely applicable to the study of other diseases.

Acknowledgments

We are grateful to the patients and families who participated in this research. We thank the following team members for contributions to patient recruitment: D. Awad, K. Celis, D. Etwaru, J. Kline, R. Korsin, A. Lanz, E. Marquez, J.K. Sond, A. Wilpers, R. Yee (Columbia Medical School); A. Roberts, K. Boardman, J. Geva, J. Gorham, B. McDonough, A. Monafo, J. Stryker (Harvard Medical School); N. Cross (Yale School of Medicine); S.M. Edman, J.L. Garbarini, J.E. Tusi, S.H. Woyciechowski (Children's Hospital of Philadelphia); R. Kim, J. Ellashek, and N. Tran (Children's Hospital of Los Angeles); K. Flack (University College London); A. Romano, D.Gruber, N. Stellato (Steve and Alexandra Cohen Children's Medical Center of New York); D. Guevara, A. Julian, M. Mac Neal, C. Mintz (Icahn School of Medicine at Mount Sinai); and G. Porter and E. Taillie (University of Rochester School of Medicine and Dentistry). We also thank V. Spotlow, P. Candrea, K. Pavlik, and M. Sotiropoulos for their expert production of exome sequences, and we thank M. Lemma, C. Kim, F.G. Otieno, M. Khan and K. Thomas for their expert production of genome-wide genotypes. We are grateful to all of the families at the participating SFARI (Simons Foundation Autism Research Initiative) Simplex Collection sites (SSC), as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, and E. Wijsman) and the coordinators and staff at the SSC clinical sites. We thank New England Research Institutes, S. Tennstedt, B. Williams, D. Nash, J. Barenholtz, K. Cucchi, K. Dandreo, S. Yates, T. Hamza, and C. Taglienti.

Sources of Funding

This work was supported by the National Institutes of Health (NIH) National Heart, Lung, and Blood Institute (NHLBI) Pediatric Cardiac Genomics Consortium (U01-HL098188, U01-HL098147, U01-HL098153, U01-HL098163, U01-HL098123, and U01-HL098162) and in part by the Simons Foundation for Autism Research and the NIH Centers for Mendelian Genomics (5U54HG006504). J.G. Homsy is supported by the John S. LaDue Fellowship of Harvard Medical School.

Disclosures

None.

References

1. Hoffman JI, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol*. 2002;39:1890–1900.
2. Tennant PW, Pearce MS, Bythell M, Rankin J. 20-year survival of children born with congenital anomalies: a population-based study. *Lancet*. 2010;375:649–656.
3. Soemedi R, Wilson IJ, Bentham J, et al. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am J Hum Genet*. 2012;91:489–501.
4. Fahed AC, Gelb BD, Seidman JG, Seidman CE. Genetics of congenital heart disease: the glass half empty. *Circ Res*. 2013;112:707–720.
5. Greenway SC, Pereira AC, Lin JC, et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat Genet*. 2009;41:931–935.
6. Silversides CK, Lionel AC, Costain G, Merico D, Migita O, Liu B, Yuen T, Rickaby J, Thiruvahindrapuram B, Marshall CR, Scherer SW, Bassett AS. Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS Genet*. 2012;8:e1002843.
7. Erdogan F, Larsen LA, Zhang L, Tümer Z, Tommerup N, Chen W, Jacobsen JR, Schubert M, Jurkatis J, Tzschach A, Ropers HH, Ullmann R. High frequency of submicroscopic genomic aberrations detected by tiling path array comparative genome hybridisation in patients with isolated congenital heart disease. *J Med Genet*. 2008;45:704–709.
8. Goldmuntz E, Paluru P, Glessner J, Hakonarson H, Biegel JA, White PS, Gai X, Shaikh TH. Microdeletions and microduplications in patients with congenital heart disease and multiple congenital anomalies. *Congenit Heart Dis*. 2011;6:592–602.
9. Luo C, Yang YF, Yin BL, Chen JL, Huang C, Zhang WZ, Wang J, Zhang H, Yang JF, Tan ZP. Microduplication of 3p25.2 encompassing RAF1 associated with congenital heart disease suggestive of Noonan syndrome. *Am J Med. Genet A*. 2012;158A:1918–1923.
10. Priest JR, Girirajan S, Vu TH, Olson A, Eichler EE, Portman MA. Rare copy number variants in isolated sporadic and syndromic atrioventricular septal defects. *Am J Med. Genet A*. 2012;158A:1279–1284.
11. Tomita-Mitchell A, Maslen CL, Morris CD, Garg V, Goldmuntz E. GATA4 sequence variants in patients with congenital heart disease. *J Med Genet*. 2007;44:779–783.
12. Christiansen J, Dyck JD, Elyas BG, Lilley M, Bamforth JS, Hicks M, Sprysak KA, Tomaszewski R, Haase SM, Vicen-Wyhony LM, Somerville MJ. Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circ Res*. 2004;94:1429–1435.
13. Carey AS, Liang L, Edwards J, et al. Effect of copy number variants on outcomes for infants with single ventricle heart defects. *Circ Cardiovasc Genet*. 2013;6:444–451.
14. Breckpot J, Thienpont B, Peeters H, de Ravel T, Singer A, Rayyan M, Allegaert K, Vanhole C, Eyskens B, Vermeesch JR, Gewillig M, Devriendt K. Array comparative genomic hybridization as a diagnostic tool for syndromic heart defects. *J Pediatr*. 2010;156:810–7, 817.e1.
15. Hitz MP, Lemieux-Perreault LP, Marshall C, et al. Rare copy number variants contribute to congenital left-sided heart disease. *PLoS Genet*. 2012;8:e1002903.
16. Payne AR, Chang SW, Koenig SN, Zinn AR, Garg V. Submicroscopic chromosomal copy number variations identified in children with hypoplastic left heart syndrome. *Pediatr Cardiol*. 2012;33:757–763.
17. Richards AA, Santos LJ, Nichols HA, Crider BP, Elder FF, Hauser NS, Zinn AR, Garg V. Cryptic chromosomal abnormalities identified in children with congenital heart disease. *Pediatr Res*. 2008;64:358–363.
18. Thienpont B, Mertens L, de Ravel T, Eyskens B, Boshoff D, Maas N, Fryns JP, Gewillig M, Vermeesch JR, Devriendt K. Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients. *Eur Heart J*. 2007;28:2778–2784.
19. Warburton D, Ronemus M, Kline J, et al. The contribution of de novo and rare inherited copy number changes to congenital heart disease in an unselected sample of children with conotruncal defects or hypoplastic left heart disease. *Hum Genet*. 2014;133:11–27.
20. Goldmuntz E, Clark BJ, Mitchell LE, Jawad AF, Cuneo BF, Reed L, McDonald-McGinn D, Chien P, Feuer J, Zackai EH, Emanuel BS, Driscoll DA. Frequency of 22q11 deletions in patients with conotruncal defects. *J Am Coll Cardiol*. 1998;32:492–498.
21. Rauch R, Hofbeck M, Zweier C, Koch A, Zink S, Trautmann U, Hoyer J, Kaulitz R, Singer H, Rauch A. Comprehensive genotype-phenotype analysis in 230 patients with tetralogy of Fallot. *J Med Genet*. 2010;47:321–331.
22. Chieffo C, Garvey N, Gong W, Roe B, Zhang G, Silver L, Emanuel BS, Budarf ML. Isolation and characterization of a gene from the DiGeorge chromosomal region homologous to the mouse Tbx1 gene. *Genomics*. 1997;43:267–277.

23. Gelb B, Brueckner M, Chung W, et al; Pediatric Cardiac Genomics Consortium. The Congenital Heart Disease Genetic Network Study: rationale, design, and early results. *Circ Res*. 2013;112:698–706.
24. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011;70:863–885.
25. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron*. 2010;68:192–195.
26. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012;485:237–241.
27. Glessner JT, Li J, Hakonarson H. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res*. 2013;41:e64.
28. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17:1665–1674.
29. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*. 2007;35:2013–2025.
30. van Karnebeek CD, Hennekam RC. Associations between chromosomal anomalies and congenital heart defects: a database search. *Am J Med. Genet*. 1999;84:158–166.
31. Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013;498:220–223.
32. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589–595.
33. Cox AJ. Eland: Efficient large-scale alignment of nucleotide databases. *Illumina, San Diego*. 2007
34. Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, Handsaker RE, McCarroll SA, O'Donovan MC, Owen MJ, Kirov G, Sullivan PF, Hultman CM, Sklar P, Purcell SM. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet*. 2012;91:597–607.
35. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–498.
36. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92.
37. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, Emslie KR. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem*. 2012;84:1003–1011.
38. Huang da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*. 2007;35:W169–W175.
39. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, Cotsapas C, Daly MJ; International Inflammatory Bowel Disease Genetics Constortium. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet*. 2011;7:e1001273.
40. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res*. 2013;41:W77–W83.
41. De Rubeis S, Pasciuto E, Li KW, et al. CYFIP1 coordinates mRNA translation and cytoskeleton remodeling to ensure proper dendritic spine formation. *Neuron*. 2013;79:1169–1182.
42. Meng Q, Rayala SK, Gururaj AE, Talukder AH, O'Malley BW, Kumar R. Signaling-dependent and coordinated regulation of transcription, splicing, and translation resides in a single coregulator, PCBP1. *Proc Natl Acad Sci USA*. 2007;104:5866–5871.
43. Ye M, Coldren C, Liang X, Mattina T, Goldmuntz E, Benson DW, Ivy D, Perryman MB, Garrett-Sinha LA, Grossfeld P. Deletion of ETS-1, a gene in the Jacobsen syndrome critical region, causes ventricular septal defects and abnormal ventricular morphology in mice. *Hum Mol Genet*. 2010;19:648–656.
44. Exome Variant Server. *NHLBI Go Exome Sequencing Project (ESP). Seattle, WA*. http://evs.Gs.Washington.Edu/evs/. Accessed January 2014.
45. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA; 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
46. Courtens W, Wuyts W, Rooms L, Pera SB, Wauters J. A subterminal deletion of the long arm of chromosome 10: a clinical report and review. *Am J Med. Genet A*. 2006;140:402–409.
47. Hildebrand JD, Soriano P. Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development. *Mol Cell Biol*. 2002;22:5296–5307.
48. Breckpot J, Tranchevent LC, Thienpont B, Bauters M, Troost E, Gewillig M, Vermeesch JR, Moreau Y, Devriendt K, Van Esch H. BMPR1A is a candidate gene for congenital heart defects associated with the recurrent 10q22q23 deletion syndrome. *Eur J Med Genet*. 2012;55:12–16.
49. Poultney CS, Goldberg AP, Drapeau E, Kou Y, Harony-Nicolas H, Kajiwara Y, De Rubeis S, Durand S, Stevens C, Rehnström K, Palotie A, Daly MJ, Ma'ayan A, Fromer M, Buxbaum JD. Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *Am J Hum Genet*. 2013;93:607–619.
50. Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, Maron BJ, Seidman CE, Seidman JG. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science*. 1998;281:108–111.
51. Baekvad-Hansen M, Tümer Z, Delicado A, Erdogan F, Tommerup N, Larsen LA. Delineation of a 2.2 Mb microdeletion at 5q35 associated with microcephaly and congenital heart disease. *Am J Med. Genet A*. 2006;140:427–433.
52. Buysse K, Crepel A, Menten B, Pattyn F, Antonacci F, Veltman JA, Larsen LA, Tümer Z, de Klein A, van de Laar I, Devriendt K, Mortier G, Speleman F. Mapping of 5q35 chromosomal rearrangements within a genomically unstable region. *J Med Genet*. 2008;45:672–678.
53. Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, Craddock N, Owen MJ, O'Donovan MC; International Schizophrenia Consortium; Wellcome Trust Case Control Consortium. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet*. 2009;18:1497–1503.
54. Stefansson H, Meyer-Lindenberg A, Steinberg S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 2014;505:361–366.
55. Stefansson H, Rujescu D, Cichon S, et al; GROUP. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232–236.
56. Kobrynski LJ, Sullivan KE. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet*. 2007;370:1443–1452.
57. Mefford HC, Sharp AJ, Baker C, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med*. 2008;359:1685–1699.
58. Golzio C, Katsanis N. Genetic architecture of reciprocal CNVs. *Curr Opin Genet Dev*. 2013;23:240–248.
59. Ghebranious N, Giampietro PF, Wesbrook FP, Rezkalla SH. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am J Med. Genet A*. 2007;143A:1462–1471.
60. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, Reymond A, Sun M, Sawa A, Gusella JF, Kamiya A, Beckmann JS, Katsanis N. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature*. 2012;485:363–367.
61. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1–13
62. Schlesinger J, Schueler M, Grunert M, Fischer JJ, Zhang Q, Krueger T, Lange M, Tönjes M, Dunkel I, Sperling SR. The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. *PLoS Genet*. 2011;7:e1001313.
63. Stennard FA, Costa MW, Elliott DA, Rankin S, Haast SJ, Lai D, McDonald LP, Niederreither K, Dolle P, Bruneau BG, Zorn AM, Harvey RP. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. *Dev Biol*. 2003;262:206–224.
64. Kroll DJ, Sullivan DM, Gutierrez-Hartmann A, Hoeffler JP. Modification of DNA topoisomerase II activity via direct interactions with the cyclic adenosine-3',5'-monophosphate response element-binding protein and related transcription factors. *Mol Endocrinol*. 1993;7:305–318.
65. Logan SK, Garabedian MJ, Campbell CE, Werb Z. Synergistic transcriptional activation of the tissue inhibitor of metalloproteinases-1 promoter via functional interaction of AP-1 and Ets-1 transcription factors. *J Biol Chem*. 1996;271:774–782.

66. Miyamoto-Sato E, Fujimori S, Ishizaka M, et al. A comprehensive resource of interacting protein regions for refining human transcription factor networks. *PLoS One*. 2010;5:e9289.
67. Caputo V, Cianetti L, Niceta M, et al. A restricted spectrum of mutations in the SMAD4 tumor-suppressor gene underlies Myhre syndrome. *Am J Hum Genet*. 2012;90:161–169.
68. Chen CR, Kang Y, Siegel PM, Massagué J. E2F4/5 and p107 as Smad cofactors linking the TGFbeta receptor to c-myc repression. *Cell*. 2002;110:19–32.
69. Macías-Silva M, Abdollah S, Hoodless PA, Pirone R, Attisano L, Wrana JL. MADR2 is a substrate of the TGFbeta receptor and its phosphorylation is required for nuclear accumulation and signaling. *Cell*. 1996;87:1215–1224.
70. Waldrip WR, Bikoff EK, Hoodless PA, Wrana JL, Robertson EJ. Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. *Cell*. 1998;92:797–808.
71. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012;74:285–299.

Novelty and Significance

What Is Known?

- Congenital heart disease (CHD) is among the most common birth defects.
- Many genomic loci are implicated in CHD, but most cases are of unknown pathogenesis.

What New Information Does This Article Contribute?

- To determine the impact of de novo copy number variants (CNVs) in CHD, we found an increased de novo CNV frequency in CHD families compared with healthy control families (termed burden).
- We found true recurrent de novos to find and define significant CHD CNV genes.
- We performed network analysis to assess biological gene function of single de novo CNVs.

CHD is among the most common birth defects. Many genomic loci are implicated in CHD, but most cases are of unknown pathogenesis. A significant increase in CNV burden was observed when comparing CHD trios with healthy trios, using either single nucleotide polymorphism array ($P=7\times10^{-5}$; odds ratio, 4.6) or whole exome sequencing data ($P=6\times10^{-4}$; odds ratio, 3.5) and remained after removing 16% of de novo CNV loci previously reported as pathogenic ($P=0.02$; odds ratio, 2.7). We observed recurrent de novo CNVs on 15q11.2 encompassing *CYFIP1, NIPA1*, and *NIPA2* and single de novo CNVs encompassing *DUSP1, JUN, JUP, MED15, MED9, PTPRE SREBF1, TOP2A,* and *ZEB2,* genes that interact with established CHD proteins *NKX2-5* and *GATA4*. Integrating de novo variants in whole exome sequencing and CNV data suggests that *ETS1* is the pathogenic gene altered by 11q24.2-q25 deletions in Jacobsen syndrome and that *CTBP2* is the pathogenic gene in 10q subtelomeric deletions. This is the first large cohort study of CHD families using whole exome sequencing and dense state-of-the-art single nucleotide polymorphism arrays for integrative de novo CNV discovery. The new loci implicated here provide novel diagnostic markers for early detection of CHD and novel therapeutic drug targets.

Supplemental Material

Increased Frequency of *De novo* Copy Number Variations in Congenital Heart Disease by Integrative Analysis of SNP Array and Exome Sequence Data

Joseph T. Glessner[1,2*], Alexander G. Bick[3*], Kaoru Ito[3*], Jason Homsy[3*], Laura Rodriguez-Murillo[4,5], Menachem Fromer[5,6,7], Erica Mazaika[3], Badri Vardarajan[8], Michael Italia[9], Jeremy Leipzig[9], Steven R. DePalma[3], Ryan Golhar[1], Stephan J. Sanders[10,11], Boris Yamrom[12], Michael Ronemus[12], Ivan Iossifov[12], A. Jeremy Willsey[10,11], Matthew W. State[10,11], Jonathan R. Kaltman[13], Peter S. White[9], Yufeng Shen[8], Dorothy Warburton[14], Martina Brueckner[15], Christine Seidman[3+], Elizabeth Goldmuntz[16+], Bruce D. Gelb[4,5+], Richard Lifton[10,17+], Jonathan Seidman[3+], Hakon Hakonarson[1,2+‡], Wendy K. Chung[18+‡]

Supplementary Methods

Ethics Statement

The protocol was approved by the Institutional Review Boards of Boston Children's Hospital, Brigham and Women's Hospital, Great Ormond St. Hospital, Children's Hospital of Los Angeles, Children's Hospital of Philadelphia, Columbia University Medical Center, Icahn School of Medicine and Mt. Sinai, Rochester School of Medicine and Dentistry, Steven and Alexandra Cohen Children's Medical Center of New York, and Yale School of Medicine. Written informed consent was obtained from each participating subject or their parent/guardian.

Patient cohorts

CHD probands and parents were recruited into the CHD Genes Study of the Pediatric Cardiac Genomics Consortium (CHD genes: ClinicalTrials.gov identifier NCT01196182) as previously described,[1] using protocols approved by Institutional Review Boards of each institution. Trios selected for this study had no history of CHD in first-degree relatives. CHD diagnoses were obtained from echocardiograms, catheterization and operative reports; extra-cardiac findings were extracted from medical records and included dysmorphic features, major anomalies, non-cardiac medical problems, and deficiencies in growth or developmental delay. The etiologies for CHD were unknown; patients with previously identified cytogenetic anomalies or pathogenic CNVs identified through routine clinical evaluation were excluded. Whole blood samples were collected and genomic DNA extracted.

CHD trios were studied by SNP arrays (n=414) or by WES (n=356), including a subset (n=233) that were analyzed by both methods. The distribution by CHD lesions in patients genotyped by arrays was: 403 (61%) left ventricular obstruction (LVO); 197 (30%) conotruncal defects (CTD); 49 (7%) heterotaxy (HTX); and 12 (2%) other cardiac diagnoses (Supplementary Table I). The distribution by CHD lesions in patients studied by WES was 284 (46.1%) left ventricular obstruction (LVO); 235 (38.1%) conotruncal defects (CTD); 78 (12.7%) heterotaxy (HTX); and 19 (3.1%) with other cardiac diagnoses (Supplemental Table II).

Control trios were the unaffected sibling and parents of a child with autism who were consented and recruited through the Simons Simplex Collection (SSC). CNVs were identified in the same way in the control trios as in the cases using SNP arrays (n=814) or WES (n=872), including a subset (n=385) analyzed by both methods.[2-4]

Additional data on the distribution and prevalence of previously reported CNVs in the general population was derived from the Database of Genomic Variants (http://dgv.tcag.ca) and from 649 de-identified control subjects who had participated in an unrelated psychiatric case-control study, genotyped on the same high density SNP array platforms at the same genotyping center as the CHD probands (438 on the Illumina Omni-1M and 211 on the Illumina Omni-2.5M). These controls were used only to prioritize the *de novo* CNVs identified by SNP array methods that were selected for confirmation analyses.

Array Genotyping and CNV identification

A total of 360 CHD parental samples genotyped on the Omni1M and 654 on Omni2.5M arrays were applied for cluster definition using Illumina Genome Studio clustering algorithm. Raw data is publicly available through the database of genotypes and phenotypes (dbGaP) National Heart, Lung, and Blood Institute (NHLBI) Bench to Bassinet Program: The Pediatric Cardiac Genetics Consortium (PCGC) under dbGaP Study Accession: phs000571.v1.p1.We removed clusters with outlier values of SNP call rate, Hardy-Weinberg equilibrium, AA/AB/BB cluster means, and minor allele frequency to improve the intensity noise (Log R ratio standard deviation) from a mean of 0.2 (using the default cluster file from Illumina) to 0.1 for CHD samples. Briefly, individual samples were filtered through a standard quality control pipeline.[3] B-allele frequency (BAF) and Log R ratio (LRR) values were exported from Illumina Genome Studio. Only samples with SNP call rate > 98%, standard deviation (SD) of normalized intensity (LRR) < 0.3, absolute value of GC-corrected LRR <0.005, as well as CNV call count <800 for Omni1-Quadv1 or <300 for Omni2.5-8v1 were included.[5] Samples with high inbreeding coefficients, that were duplicated, or had gender mismatches, and trios with Mendelian errors > 1% were removed from analyses. We started with 1,536 genotyped samples (512 trios), including 561 on the Illumina Omni-1M and 969 on the Illumina Omni-2.5M. Four hundred and sixty-one trios had the same array version for all family members. Upon completion of these QC procedures 1,245 samples, including 447 genotyped on the Illumina Omni-1M and 798 on the Illumina Omni-2.5M high-density SNP array platforms, were taken forward for analysis, constituting 415 complete trios (Supplemental Table III).

Three groups (CHOP, Harvard, Yale) independently analyzed genotyping data using slightly different algorithms to detect putative *de novo* CNVs. For each of the three independent analyses, CNVs were called for each subject using PennCNV [6] with the hidden Markov model algorithm and custom-made population frequency of B-allele (PFB) and GC model files. CNVs were called when 10 or more consecutive probes demonstrated consistent copy number

change. The PennCNV detect_cnv --trio option was used to boost transmission probability of CNV calling for initially *de novo* scored CNVs. Fragmented CNV calls were merged using clean_cnv. All candidate CNVs were visually inspected to ensure the appropriate pattern of LRR and B-allele frequency was consistent with the CNV call. Additionally, Gnosis,[3] QuantiSNP,[7] and Nexus (biodiscovery.com) were used to increase specificity. *De novo* CNVs were prioritized for quality by genomic length, number of probes, confidence score based on signal strength, 50% overlap of two or more algorithms, low parental origin p-value using infer_snp_allele, and visual BAF/LRR review. CNVs with a minor allele frequency > 1% were removed, leaving rare CNVs. All putative *de novo* CNVs were experimentally evaluated by digital droplet PCR (ddPCR, Supplemental Figure I), and only validated CNVs are reported.

De novo CNV loci that were previously reported as pathogenic were defined by reported recurrence in at least two publications using independent data. Although some of the CNVs reported here overlap with previously reported CNVs in CHD patients based on review of the literature,[8], they do not meet our frequency constraint for previously reported pathogenic *de novo* CNV loci.

CNV identification and variant calling from WES Data

WES data from 356 CHD trios were analyzed for *de novo* CNVs (Supplemental Table II). WES samples were captured with the Nimblegen SeqCap Exome V2 chemistry and sequenced on the Illumina HiSeq 2000 platform as previously described.[9] Sequence reads were aligned to the human reference genome hg19 using Novoalign (http://novocraft.com), BWA,[10] and ELAND.[11] Duplicates were marked with Picard (http://picard.sourceforge.net). Indel realignment and Base Quality Score Recalibration was done with GATK. XHMM is an algorithm to detect exon-level copy number variation and assign CNV quality metrics[12] and was used at four of the PCGC analysis sites (CHOP, Harvard, Columbia and Mount Sinai) to identify *de novo* CNVs (Supplemental Figure II). Candidate *de novo* CNVs were inspected visually. Putative *de novo* CNVs were prioritized for confirmation based on genomic length, low sequence depth variability and low prevalence in the XHMM call set data (AF<1%). All putative *de novo* CNVs were independently confirmed by ddPCR.

SNVs and short insertions/deletions (indels) were called from the Novoalign alignment of WES trios using a pipeline derived from GATK version 2.7 best practices.[13] Briefly, aligned reads were first compressed using the GATK ReducedReads module and variants were called on all CHD WES trios using the UnifiedGenotyper joint variant calling module. Identified variants were filtered using GATK variant quality score recalibration. Variants were annotated using SnpEff.[14] *De novo* SNVs and indels were independently confirmed using Sanger sequencing.

CNV confirmation with digital droplet PCR

Putative CNVs were experimentally confirmed with ddPCR as previously reported[15] using an 18-27 base pair FAM probe designed within each candidate CNV region, avoiding homopolymer runs or probes that began with G. A VIC probe targeting the RPP30 gene was used as reference. Reaction mixtures of 20 µL volume comprising ddPCR Master Mix (Bio-Rad), relevant forward and reverse primers and probe(s) and 100 ng of digested DNA were prepared, ensuring that approximately 25-75% of the 10,000 droplets ultimately produced were positive for FAM or VIC signal. For *de novo* CNV confirmations, DNA from the CHD patient and parents was used. After thermal cycling, plates were transferred to a droplet reader (Bio-Rad) that flows droplets single-file past a two-color fluorescence detector. Differentiation between droplets that contain target and those that did not was achieved by applying a global fluorescence amplitude threshold in QuantaSoft (Bio-Rad). The threshold was set manually based on visual inspection at approximately the midpoint between the average fluorescence amplitude of positives and negative droplet clusters on each of the FAM and VIC channels. Confirmed CNV duplications had approximately 50% increase in the ratio of positive to negative droplets as did the reference channel. Conversely confirmed CNV deletions had approximately half the ratio of positive to negative droplets as did the reference channel.

Network analysis

Three bioinformatic algorithms were utilized: DAVID,[16] DAPPLE,[17] and WebGestalt.[18] Four different gene lists derived from the *de novo* CNV loci were used (Supplemental Table IV). The lists were constructed as follows: (1) All genes contained within *de novo* CNV intervals; (2) Published "causative" genes from previously reported CHD CNVs intervals in addition to all genes in novel CHD CNV intervals. "Causal" genes in previously reported CNV intervals included *ELN* (Williams syndrome), *RAI1* (Smith-Magenis syndrome),*TBX1* (22q11 deletion), *GATA4* (8p23.1 deletion), *GJA5* (1q21.1 duplication), and *NKX2.5* (5q35.1 deletion); (3) Genes contained solely within novel CHD CNV intervals (e.g., exclude genes from previously published CNVs); (4) Genes contained within *de novo* CNV intervals that are highly expressed in the developing mouse heart (top quartile of all genes expressed E14.5 mouse heart).[9] We anticipated that genes in list 2 and list 4 would have increased specificity for CHD in comparison to genes in list 1 and that genes in list 3 would be biased towards new disease networks.

We expanded network analysis input gene lists by including both *de novo* CNV genes and *de novo* single nucleotide variants (SNV) that were previously identified in CHD probands by WES.[9] Only *de novo* SNVs predicted to be deleterious (e.g., loss of function (LOF): nonsense, frame-shift, and splice site mutations and missense variants that alter highly conserved amino acid residues or predicted to be deleterious by SIFT or PolyPhen2) were included in the expanded gene list. The additional gene lists included: (5) All genes within a *de novo* CNV interval (e.g., list 1) and protein-altering SNVs and (6) Published

"causative" genes from previously reported CHD CNVs intervals in addition to all genes in novel CHD CNV intervals (e.g., list 2) and protein altering SNVs.

Statistical analysis

Burden calculations were done with a Fisher exact test computed in the R statistical computing environment. For analyses using DAVID, networks with an enrichment of genes impacted by CNVs were assigned a *p*-value with Benjamini and Hochberg correction for multiple testing with a false discovery rate of 0.05. In DAPPLE, type I error was controlled through permutation. *p*-values of less than 0.05 were considered significant.

Supplementary Tables

Online Table I. Patient characteristics of PCGC cohort studied by SNP array (n=414)

Proband Demographics (probands)[†]	
Male	261 (63.0%)
Female	153 (37.0%)
Age at enrollment *(mean ± SD in years)*	6.6±7.9
Maternal age at birth of proband *(mean ± SD in years)*	30.5±5.6
Paternal age at birth of proband *(mean ± SD in years)*	32.8±6.6
Race	N=400
African-American	34 (8.2%)
White	364 (97.9%)
Asian	19 (4.6%)
Native American	7 (1.7%)
Hispanic	77 (18.6%)
Cardiac Lesions[§]	
Conotruncal Defects (CTD): probands	
TOF	37 (8.9%)
TOF/PA	20 (4.8%)
TOF/PA-MAPCA	13 (3.1%)
D-TGA	38 (9.2%)
DORV	31 (7.5%)
TA	10 (2.4%)
Interrupted AoAA	13 (3.1%)
Isolated AoAA	35 (8.4%)
Left Ventricular Obstruction (LVO): probands	
HLHS	60 (14.5%)
CoA	97 (23.4%)
BAV	87 (21.0%)
AS (not BAV)	117 (28.3%)
MS	41 (9.9%)
LVOT	1 (0.2%)
Heterotaxy (HTX): probands	
RAI	4 (1.0%)
LAI	5 (1.2%)
DEX	13 (3.1%)
CAVC	11 (2.7%)
DORV	0
L-Loop	10 (2.4%)
L-TGA	5 (1.2%)
D-TGA	1 (0.2%)

Other Cardiac Diagnosis: probands	
TAPVR	5 (1.2%)
DILV	7 (1.7%)
Extracardiac abnormalities	130 (31.4%)

[†]Self-reported ethnicity: 17 probands self-identfied as multi-ethnic, hence ethnicities total >100%
[§]Numerous probands have more than one cardiac feature; hence, sum of sub-phenotypic classification does not equal 100%.
Abbreviations: TOF-tetralogy of Fallot; TOF/PA-tetralogy of Fallot with pulmonary atresia; TOF/PA-MAPCA-tetralogy of Fallot with pulmonary atresia and multiple aortico-pulmonary collaterals; D-TGA-D-transposition of the great arteries; DORV-double outlet right ventricle; AoAA-aortic arch anomaly; TA-truncus arteriosus; HLHS-hypoplastic left heart syndrome; CoA-coarctation of the aorta; BAV-bicuspid aortic valve; AS-aortic stenosis; MS-mitral stenosis; RAI-right atrial isomerism; LAI-left atrial isomerism; Dex-dextrocardia; CAVC-complete atrioventricular canal; L-Loop-L-looped ventricles; L-TGA- L-transposition of the great arteries; TAPVR-total anomalous pulmonary venous return; DILV-double-inlet left ventricle.

Online Table II. Patient characteristics of PGCG cohort with WES data (n=358)

Proband Demographics (probands)[†]	
Male	228 (63.7%)
Female	130 (36.3%)
Age at enrollment *(mean ± SD in years)*	8.2±9.5
Maternal age at birth of proband *(mean ± SD in years)*	30.3±5.3
Paternal age at birth of proband *(mean ± SD in years)*	32.7±6.1
Race	N=345
African-American	33 (9.2%)
White	305 (85.2%)
Asian	31 (8.7%)
Native American	5 (1.4%)
Hispanic	50 (13.9%)
Cardiac Lesions[§]	
Conotruncal Defects (CTD): probands	
TOF	63 (17.6%)
TOF/PA	16 (4.4%)
TOF/PA-MAPCA	14 (3.9%)
D-TGA	60 (16.7%)
DORV	40 (11.1%)
TA	6 (1.7%)
Interrupted AoAA	4 (1.1%)
Isolated AoAA	32 (8.9%)
Left Ventricular Obstruction (LVO): probands	
HLHS	46 (12.8%)
CoA	65 (18.1%)
BAV	57 (15.9%)
AS (not BAV)	89 (24.9%)
MS	26 (7.2%)
LVOT	1 (0.3%)
Heterotaxy (HTX): probands	
RAI	8 (2.2%)
LAI	7 (2.0%)
DEX	22 (6.1%)
CAVC	20 (5.6%)
DORV	0
L-Loop	15 (4.2%)
L-TGA	5 (1.4%)
D-TGA	1 (0.3%)
Other Cardiac Diagnosis: probands	
TAPVR	10 (2.8%)
DILV	9 (2.5%)
Extracardiac abnormalities	112 (31.2%)

[†]Self-reported ethnicity: 17 probands self-identfied as multi-ethnic, hence ethnicities total >100%
[§]Numerous probands have more than one cardiac feature; hence, sum of sub-phenotypic classification does not equal 100%.
Abbreviations: TOF-tetralogy of Fallot; TOF/PA-tetralogy of Fallot with pulmonary atresia; TOF/PA-MAPCA-tetralogy of Fallot with pulmonary atresia and multiple aortico-pulmonary collaterals; D-TGA-D-transposition of the great arteries; DORV-double outlet right ventricle; AoAA-aortic arch anomaly; TA-truncus arteriosus; HLHS-hypoplastic left heart syndrome; CoA-coarctation of the aorta; BAV-bicuspid aortic valve; AS-aortic stenosis; MS-mitral stenosis; RAI-right atrial isomerism; LAI-left atrial isomerism; Dex-dextrocardia; CAVC-complete atrioventricular canal; L-Loop-L-looped ventricles; L-TGA- L-transposition of the great arteries; TAPVR-total anomalous pulmonary venous return; DILV-double-inlet left ventricle.

Online Table III. Array version distribution

	Illumina HumanOmni1-Quadv1-0 SNP Array	Illumina HumanOmni25-8v1 SNP Array
Proband	149	266
Mother	148	267
Father	148	267

Online Table IV. Gene lists derived from *de novo* CNV loci and used for network analysis.

(1) All genes contained within *de novo* CNV intervals

IFI44L	IFI44	PIAS3	NBPF10	RNF115	NUDT17	CD160	POLR3C	PDZK1	GPR89A
AX747161	LOC100288142	PRKAB2	PDIA3PLOC100505794	FMO5	CHD1L	LINC00624	BCL9	ACP6	GJA5
GJA8	GPR89B	OR2T10	OR2T11	JUN	LOC100131060	ASPRV1	LOC100133985	MXD1	PCBP1
PCBP1-AS1	ZEB2	CHL1-AS2	CHL1	CNTN6	CNTN4	CNTN4-AS2	IL5RA	TRNT1	CRBN
LRRN1	SETMAR	SUMF1	ITPR1-AS1	ITPR1	EGOT	BHLHE40-AS1	BHLHE40	ARL8B	EDEM1
MIR4790	GRM7	LOC100996542	LMCD1-AS1	LMCD1	LINC00312	SSUH2	CAV3	OXTR	RAD18
SRGAP3	SRGAP3-AS3	THUMPD3	SETD5-AS1	LOC100996546	SETD5	LHFPL4	MTMR14	CPNE9	BRPF1
OGG1	CAMK1	TADA3	ARPC4	ARPC4-TTLL3	TTLL3	RPUSD3	CIDEC	JAGN1	IL17RE
IL17RC	CRELD1	PRRT3	PRRT3-AS1	EMC3	LOC401052	FANCD2	FANCD2OS	BRK1	VHL
IRAK2	TATDN2	LINC00852	GHRL	GHRLOS	SEC13	ATP2B2	MIR378B	MIR885	LINC00606
SLC6A11	SLC6A1	SLC6A1-AS1	HRH1	ATG7	VGLL4	LOC100133039	HACL1	BTD	n/a
CCDC111	MLF1IP	TENM2	WWC1	RARS	FBLL1	SLC2A3P1	PANK3	MIR103A1	MIR103B1
RPL10P9	SLIT3	MIR218-2	MIR585	SPDL1	DOCK2	FAM196B	MIR378E	FOXI1	KRT18P41
C5orf58	LCP2	KCNIP1	KCNMB1	GABRP	RANBP17	USP12P1	TLX3	RPL19P10	RPL10P8
MIR3912	NPM1	FGF18	C5orf50	FBXW11	STK10	EFCAB9	UBTD2	KLF3P1	SH3PXD2B
NEURL1B	MIR5003	DUSP1	ERGIC1	RPL26L1	ATP6V0E1	SNORA74B	CREBRF	CDC42P5	BNIP1
RPL7AP33	NKX2-5	STC2	BOD1	UBE2B	POLR1C	CDKN1A	C7orf72	ABHD11	ABHD11-AS1
BAZ1B	BC050599	BCL7B	CLDN3	CLDN4	CLIP2	DNAJC30	EIF4H	ELN	FKBP6
FZD9	LAT2	LIMK1	MIR4284	MIR590	MLXIPL	NSUN5	RFC2	STX1A	TBL2
TRIM50	VPS37D	WBSCR22	WBSCR27	WBSCR28	hCPE-R	BC070376	GTF2I	GTF2IRD1	TCRVB
TCRVB2S1	TRBV	TCRBV19S1P	TCRBV15S1	MTRNR2L6	TCRBV14S1	TCRBV11S1A1T	AX748042	TCRBV3S1	TCRBV4S1A1T
PRSS1	SGK223	CLDN23	MFHAS1	ERI1	MIR4660	PPP1R3B	LOC157273	TNKS	MIR597
LINC00599	MIR124-1	MSRA	PRSS55	RP1L1	MIR4286	C8orf74	SOX7	PINX1	MIR1322
XKR6	MIR598	LOC100996519	MTMR9	SLC35G5	C8orf12	FAM167A	BLK	LINC00208	GATA4
C8orf49	NEIL2	FDFT1	CTSB	DEFB136	DEFB135	DEFB134	LOC100133267	ZNF705D	FAM66D
USP17L7	USP17L2	FAM86B1	FAM85A	DEFB130	FAM66A	FAM86B2	LOC100506990	LOC729732	AK091593
AK094525	AK307207	BC017578	BC038546	BC043573	C8orf15	CR749668	DQ580909	FLJ10661	LOC157627
LOC392196	SNORD112	TDH	TRNA_Pseudo	U6	EXT1	RLN1	RLN2	ACADSB	ADAM12
ARMS2	BCCIP	BTBD16	BUB3	C10orf88	C10orf90	C10orf120	C10orf137	CLRN3	CTBP2
CPXM2	CHST15	CUZD1	DHX32	DMBT1	DOCK1	EBF3	FAM24B	FAM175B	FAM196A
FAM24A	FAM53B	FANK1	FLJ37035	FOXI2	GPR26	HMX3	HMX2	HTRA1	IKZF5
LHPP	LOC100506255	LOC387720	LOC283038	LOC100169752	LOC728065	METTL10	MIR4296	MIR3941	MMP21
MIR4484	MKI67	MGMT	MIR4297	NPSPTPRE	NKX1-2	OAT	OATP1	PLEKHA1	PSTK
TACC2	TEX36	UROS	ZRANB1	AF113016	BC010924	EHBP1L1	FAM89B	FRMD8	KCNK7
LOC254100	LTBP3	MALAT1	MAP3K11	MIR4690	MascRNA_menRNA	NEAT1	PCNXL3	SCYL1	SIPA1
SSSCA1	U7	TMEM135	PKNOX2	FEZ1	MGC39545	EI24	STT3A	CHEK1	ACRV1
PATE1	PATE2	PATE3	PATE4	HYLS1	PUS3	DDX25	LOC338667	CDON	RPUSD4
FAM118B	SRPR	FOXRED1	TIRAP	DCPS	FLJ39051	ST3GAL4	KIRREL3	KIRREL3-AS2	MIR3167
KIRREL3-AS3	ETS1	FLI1	FLI1-AS1	KCNJ1	KCNJ5	C11orf45	TP53AIP1	ARHGAP32	ZNF123P
BARX2	TMEM45B	NFRKB	PRDM10	LINC00167	APLP2	ST14	ZBTB44	LOC100996646	ADAMTS8
ADAMTS15	C11orf44	LOC100507431	SNX19	NTM	C11orf39	OPCML	SPATA19	LOC283174	MIR4697
LOC646543	IGSF9B	LOC100128239	JAM3	AK125040	MRGPRX1	CAT	KRT6C	TIMM9	TOMM20L
CYFIP1	LOC283683	NIPA1	NIPA2	TUBGCP5	abParts	WHAMMP3	HERC2	EIF3J	SPG11
NTRK3	AK296148	AK301679	AKAP10	ALDH3A1	ALDH3A2	ALKBH5	ATPAF2	AX747021	AX748015
AX748411	B9D1	BC029877	BC038464	BC040601	BC058012	BC112347	BC150162	C17orf39	CCDC144B
CCDC144C	COPS3	DL490867	DM110819	DQ570768	DQ571055	DQ572309	DQ573130	DQ576083	DQ578124
DQ578661	DQ579174	DQ579248	DQ584883	DQ585755	DQ585853	DQ586004	DQ586114	DQ592578	DQ592713
DQ593368	DQ595299	DQ595911	DQ596932	DRG2	EPN2	EVPLL	FAM106A	FAM18B1	FAM83G
FBXW10	FLCN	FLII	FLJ00050	FOXO3B	GRAP	GRAPL	LGALS9C	LLGL1	LOC339240
LRRC48	MAPK7	MED9	MFAP4	MIR1180	MIR33B	MPRIP	MYO15A	NT5M	PEMT
PLD6	PRPSAP2	RAI1	RASD1	RNF112	SHMT1	SLC47A1	SLC47A2	SLC5A10	SMCR5
SMCR7	SMCR8	SMCR9	SNORA59B	SPECC1	SREBF1	SREBP-1	TBC1D28	TNFRSF13B	TOM1L2
TOP3A	TRIM16L	TRNA_Gly	TRNA_Trp	ULK2	USP32P2	Y_RNA	ZNF286B	DHRS7B	TMEM11
C17orf103	MAP2K3	KCNJ12	C17orf51	FAM27L	FLJ36000	MTRNR2L1	SSH2	TOP2A	RARA
JUP	EIF1	ZNF826P	ZNF546	ZNF780A	ZNF780B	C5AR1	C5AR2	DHX34	GPR77
MACROD2	CRKL	TBX1	APOL5	APOL6	RBFOX2	CYP2D6	NDUFA6-AS1	DDX53	LOC100873065
TMEM185A									

(2) Previously reported “causative” genes contained in CHD CNVs intervals plus all genes in novel CHD CNV intervals. “Causal” genes in previously reported CNV intervals included *ELN* (Williams syndrome), *RAI1* (Smith-Magenis syndrome),*TBX1* (22q11 deletion), *GATA4* (8p23.1 deletion), *GJA5* (1q21.1 duplication), and *NKX2.5* (5q35.1 deletion)

IFI44L	IFI44	PIAS3	NBPF10	RNF115	NUDT17	CD160	POLR3C	PDZK1	GPR89A
AX747161	LOC100288142	PRKAB2	PDIA3P	LOC100505794	FMO5	CHD1L	LINC00624	BCL9	ACP6
GJA5	GJA8	GPR89B	OR2T10	OR2T11	JUN	LOC100131060	ASPRV1	LOC100133985	MXD1
PCBP1	ZEB2	CHL1	CNTN6	CNTN4	IL5RA	TRNT1	CRBN	LRRN1	SETMAR
SUMF1	ITPR1	EGOT	BHLHE40	ARL8B	EDEM1	MIR4790	GRM7	LOC100996542	LMCD1
LINC00312	SSUH2	CAV3	OXTR	RAD18	SRGAP3	THUMPD3	SETD5	LOC100996546	LHFPL4
MTMR14	CPNE9	BRPF1	OGG1	CAMK1	TADA3	ARPC4	TTLL3	RPUSD3	CIDEC
JAGN1	IL17RE	IL17RC	CRELD1	PRRT3	EMC3	LOC401052	FANCD2	FANCD2OS	BRK1
VHL	IRAK2	TATDN2	LINC00852	GHRL	GHRLOS	SEC13	ATP2B2	MIR378B	MIR885
LINC00606	SLC6A11	SLC6A1	HRH1	ATG7	VGLL4	LOC100133039	HACL1	BTD	CCDC111
MLF1IP	UBE2B	POLR1C	CDKN1A	C7orf72	TCRVB	TCRVB2S1	TRBV	TCRBV19S1P	TCRBV15S1
MTRNR2L6	TCRBV14S1	TCRBV11S1A1T	AX748042	TCRBV3S1	TCRBV4S1A1T	PRSS1	GATA4	EXT1	RLN1
RLN2	ACADSB	ADAM12	ARMS2	BCCIP	BTBD16	BUB3	C10orf88	C10orf90	C10orf120
C10orf137	CLRN3	CTBP2	CPXM2	CHST15	CUZD1	DHX32	DMBT1	DOCK1	EBF3
FAM24B-CUZD1	FAM175B	FAM196A	FAM24B	FAM24A	FAM53B	FANK1	FLJ37035	FOXI2	GPR26
HMX3	HMX2	HTRA1	IKZF5	LHPP	LOC100506255	LOC387720	LOC283038	LOC100169752	LOC728065
METTL10	MIR4296	MIR3941	MMP21	MIR4484	MKI67	MGMT	MIR4297	NPSPTPRE	NKX1-2
OAT	OATP1	PLEKHA1	PSTK	TACC2	TEX36	UROS	ZRANB1	AF113016	BC010924
EHBP1L1	FAM89B	FRMD8	KCNK7	LOC254100	LTBP3	MALAT1	MAP3K11	MIR4690	MascRNA_menRNA
NEAT1	PCNXL3	SCYL1	SIPA1	SSSCA1	U7	TMEM135	PKNOX2	FEZ1	MGC39545
EI24	STT3A	CHEK1	ACRV1	PATE1	PATE2	PATE3	PATE4	HYLS1	PUS3
DDX25	LOC338667	CDON	RPUSD4	FAM118B	SRPR	FOXRED1	TIRAP	DCPS	FLJ39051
ST3GAL4	KIRREL3	MIR3167	ETS1	FLI1	KCNJ1	KCNJ5	C11orf45	TP53AIP1	ARHGAP32
ZNF123P	BARX2	TMEM45B	NFRKB	PRDM10	LINC00167	APLP2	ST14	ZBTB44	LOC100996646
ADAMTS8	ADAMTS15	C11orf44	LOC100507431	SNX19	NTM	C11orf39	OPCML	SPATA19	LOC283174
MIR4697	LOC646543	IGSF9B	LOC100128239	JAM3	AK125040	MRGPRX1	CAT	KRT6C	TIMM9
TOMM20L	CYFIP1	LOC283683	NIPA1	NIPA2	TUBGCP5	WHAMMP3	HERC2	EIF3J	SPG11
NTRK3	TRNA	DHRS7B	TMEM11	C17orf103	MAP2K3	KCNJ12	C17orf51	FAM27L	FLJ36000
MTRNR2L1	SSH2	TOP2A	RARA	JUP	EIF1	ZNF826P	ZNF546	ZNF780A	ZNF780B
C5AR1	C5AR2	DHX34	GPR77	MACROD2	TBX1	APOL5	APOL6	RBFOX2	CYP2D6
NDUFA6	DDX53	LOC100873065	TMEM185A	NKX2-5	ABHD11	BAZ1B	BC050599	BCL7B	CLDN3
CLDN4	CLIP2	DNAJC30	EIF4H	ELN	FKBP6	FZD9	LAT2	LIMK1	MIR4284
MIR590	MLXIPL	NSUN5	RFC2	STX1A	TBL2	TRIM50	VPS37D	WBSCR22	WBSCR27
WBSCR28	hCPE-R	RAI1							

(3) Genes contained solely within novel CHD CNV intervals (e.g., exclude genes from previously published CNVs)

IFI44L	IFI44	PIAS3	NBPF10	RNF115	NUDT17	CD160	POLR3C	PDZK1	GPR89A
AX747161	LOC100288142	PRKAB2	PDIA3P	LOC100505794	FMO5	CHD1L	LINC00624	BCL9	ACP6
GJA5	GJA8	GPR89B	OR2T10	OR2T11	JUN	LOC100131060	ASPRV1	LOC100133985	MXD1
PCBP1	CHL1	CNTN6	CNTN4	IL5RA	TRNT1	CRBN	LRRN1	SETMAR	SUMF1
ITPR1	EGOT	BHLHE40	ARL8B	EDEM1	MIR4790	GRM7	LOC100996542	LMCD1	LINC00312
SSUH2	CAV3	OXTR	RAD18	SRGAP3	THUMPD3	SETD5	LOC100996546	LHFPL4	MTMR14
CPNE9	BRPF1	OGG1	CAMK1	TADA3	ARPC4	TTLL3	RPUSD3	CIDEC	JAGN1
IL17RE	IL17RC	CRELD1	PRRT3	EMC3	LOC401052	FANCD2	FANCD2OS	BRK1	VHL
IRAK2	TATDN2	LINC00852	GHRL	GHRLOS	SEC13	ATP2B2	MIR378B	MIR885	LINC00606
SLC6A11	SLC6A1	HRH1	ATG7	VGLL4	LOC100133039	HACL1	BTD	CCDC111	MLF1IP
UBE2B	POLR1C	CDKN1A	C7orf72	TCRVB	TCRVB2S1	TRBV	TCRBV19S1P	TCRBV15S1	MTRNR2L6
TCRBV14S1	TCRBV11S1A1T	AX748042	TCRBV3S1	TCRBV4S1A1T	PRSS1	EXT1	RLN1	RLN2	ACADSB
ADAM12	ARMS2	BCCIP	BTBD16	BUB3	C10orf88	C10orf90	C10orf120	C10orf137	CLRN3
CTBP2	CPXM2	CHST15	CUZD1	DHX32	DMBT1	DOCK1	EBF3	FAM24B-CUZD1	FAM175B
FAM196A	FAM24B	FAM24A	FAM53B	FANK1	FLJ37035	FOXI2	GPR26	HMX3	HMX2
HTRA1	IKZF5	LHPP	LOC100506255	LOC387720	LOC283038	LOC100169752	LOC728065	METTL10	MIR4296
MIR3941	MMP21	MIR4484	MKI67	MGMT	MIR4297	NPSPTPRE	NKX1-2	OAT	OATP1
PLEKHA1	PSTK	TACC2	TEX36	UROS	ZRANB1	AF113016	BC010924	EHBP1L1	FAM89B
FRMD8	KCNK7	LOC254100	LTBP3	MALAT1	MAP3K11	MIR4690	MascRNA_menRNA	NEAT1	PCNXL3
SCYL1	SIPA1	SSSCA1	U7	TMEM135	PKNOX2	FEZ1	MGC39545	EI24	STT3A
CHEK1	ACRV1	PATE1	PATE2	PATE3	PATE4	HYLS1	PUS3	DDX25	LOC338667
CDON	RPUSD4	FAM118B	SRPR	FOXRED1	TIRAP	DCPS	FLJ39051	ST3GAL4	KIRREL3
MIR3167	ETS1	FLI1	KCNJ1	KCNJ5	C11orf45	TP53AIP1	ARHGAP32	ZNF123P	BARX2
TMEM45B	NFRKB	PRDM10	LINC00167	APLP2	ST14	ZBTB44	LOC100996646	ADAMTS8	ADAMTS15
C11orf44	LOC100507431	SNX19	NTM	C11orf39	OPCML	SPATA19	LOC283174	MIR4697	LOC646543
IGSF9B	LOC100128239	JAM3	AK125040	MRGPRX1	CAT	KRT6C	TIMM9	TOMM20L	CYFIP1
LOC283683	NIPA1	NIPA2	TUBGCP5	WHAMMP3	HERC2	EIF3J	SPG11	NTRK3	TRNA
DHRS7B	TMEM11	C17orf103	MAP2K3	KCNJ12	C17orf51	FAM27L	FLJ36000	MTRNR2L1	SSH2
TOP2A	RARA	JUP	EIF1	ZNF826P	ZNF546	ZNF780A	ZNF780B	C5AR1	C5AR2
DHX34	GPR77	MACROD2	APOL5	APOL6	RBFOX2	CYP2D6	NDUFA6	DDX53	LOC100873065
TMEM185A									

(4) Genes contained within *de novo* CNV intervals that are highly expressed in the developing mouse heart (top quartile of all genes expressed E14.5 mouse heart)

ACADSB	ACP6	ALKBH5	APLP2	ARL8B	ARPC4	ATP6V0E1	BAZ1B	BCCIP	BCL9
BUB3	C17orf103	CAMK1	CAT	CAV3	CDON	COPS3	CRBN	CTBP2	CTSB
CYFIP1	DCPS	DOCK1	EHBP1L1	EI24	EIF1	EIF3J	EIF4H	ELN	EMC3
ERGIC1	ETS1	EXT1	FDFT1	FLII	FOXRED1	GATA4	GJA5	GTF2I	HERC2
ITPR1	JUN	JUP	KCNJ5	LLGL1	LTBP3	MAP3K11	MFAP4	MFHAS1	MKI67
MPRIP	NEURL1B	NKX2-5	NPM1	NT5M	OAT	PCBP1	PCNXL3	PLEKHA1	PPP1R3B
RAI1	RARA	RARS	RBFOX2	RFC2	SCYL1	SEC13	SETD5	SH3PXD2B	SLIT3
SREBF1	SRPR	SSSCA1	STT3A	TACC2	TIMM9	TMEM11	TNKS	TOM1L2	TOP2A
UBE2B	VGLL4	ZEB2							

(5) Genes within all *de novo* CNV intervals (list 1) and gene containing rare, damaging *de novo* variants (Snps and indels)

IFI44L	IFI44	PIAS3	NBPF10	RNF115	NUDT17	CD160	POLR3C	PDZK1	GPR89A
AX747161	LOC100288142	PRKAB2	PDIA3P	LOC100505794	FMO5	CHD1L	LINC00624	BCL9	ACP6
GJA5	GJA8	GPR89B	OR2T10	OR2T11	JUN	LOC100131060	ASPRV1	LOC100133985	MXD1
PCBP1	ZEB2	CHL1	CNTN6	CNTN4	IL5RA	TRNT1	CRBN	LRRN1	SETMAR
SUMF1	ITPR1	EGOT	BHLHE40	ARL8B	EDEM1	MIR4790	GRM7	LOC100996542	LMCD1
LINC00312	SSUH2	CAV3	OXTR	RAD18	SRGAP3	THUMPD3	SETD5	LOC100996546	LHFPL4
MTMR14	CPNE9	BRPF1	OGG1	CAMK1	TADA3	ARPC4	TTLL3	RPUSD3	CIDEC
JAGN1	IL17RE	IL17RC	CRELD1	PRRT3	EMC3	LOC401052	FANCD2	FANCD2OS	BRK1
VHL	IRAK2	TATDN2	LINC00852	GHRL	GHRLOS	SEC13	ATP2B2	MIR378B	MIR885
LINC00606	SLC6A11	SLC6A1	HRH1	ATG7	VGLL4	LOC100133039	HACL1	BTD	CCDC111
MLF1IP	TENM2	WWC1	RARS	FBLL1	SLC2A3P1	PANK3	MIR103A1	MIR103B1	RPL10P9
SLIT3	MIR218-2	MIR585	SPDL1	DOCK2	FAM196B	MIR378E	FOXI1	KRT18P41	C5orf58
LCP2	KCNIP1	KCNMB1	GABRP	RANBP17	USP12P1	TLX3	RPL19P10	RPL10P8	MIR3912
NPM1	FGF18	C5orf50	FBXW11	STK10	EFCAB9	UBTD2	KLF3P1	SH3PXD2B	NEURL1B
MIR5003	DUSP1	ERGIC1	RPL26L1	ATP6V0E1	SNORA74B	CREBRF	CDC42P5	BNIP1	RPL7AP33
NKX2-5	STC2	BOD1	UBE2B	POLR1C	CDKN1A	C7orf72	ABHD11	BAZ1B	BC050599
BCL7B	CLDN3	CLDN4	CLIP2	DNAJC30	EIF4H	ELN	FKBP6	FZD9	LAT2
LIMK1	MIR4284	MIR590	MLXIPL	NSUN5	RFC2	STX1A	TBL2	TRIM50	VPS37D
WBSCR22	WBSCR27	WBSCR28	hCPE-R	BC070376	GTF2I	GTF2IRD1	TCRVB	TCRVB2S1	TRBV
TCRBV19S1P	TCRBV15S1	MTRNR2L6	TCRBV14S1	TCRBV11S1A1T	AX748042	TCRBV3S1	TCRBV4S1A1T	PRSS1	LINC00208
GATA4	C8orf49	NEIL2	FDFT1	CTSB	DEFB136	DEFB135	DEFB134	LOC100133267	ZNF705D
FAM66D	USP17L7	USP17L2	FAM86B1	FAM85A	DEFB130	FAM66A	FAM86B2	LOC100506990	LOC729732
AK091593	AK094525	AK307207	BC017578	BC038546	BC043573	C8orf15	CR749668	DQ580909	FLJ10661
LOC157627	LOC392196	SNORD112	TDH	TRNA_Pseudo	U6	EXT1	RLN1	RLN2	ACADSB
ADAM12	ARMS2	BCCIP	BTBD16	BUB3	C10orf88	C10orf90	C10orf120	C10orf137	CLRN3
CTBP2	CPXM2	CHST15	CUZD1	DHX32	DMBT1	DOCK1	EBF3	FAM24B-CUZD1	FAM175B
FAM196A	FAM24B	FAM24A	FAM53B	FANK1	FLJ37035	FOXI2	GPR26	HMX3	HMX2
HTRA1	IKZF5	LHPP	LOC100506255	LOC387720	LOC283038	LOC100169752	LOC728065	METTL10	MIR4296
MIR3941	MMP21	MIR4484	MKI67	MGMT	MIR4297	NPSPTPRE	NKX1-2	OAT	OATP1
PLEKHA1	PSTK	TACC2	TEX36	UROS	ZRANB1	AF113016	BC010924	EHBP1L1	FAM89B
FRMD8	KCNK7	LOC254100	LTBP3	MALAT1	MAP3K11	MIR4690	MascRNA_menRNA	NEAT1	PCNXL3
SCYL1	SIPA1	SSSCA1	U7	TMEM135	PKNOX2	FEZ1	MGC39545	EI24	STT3A
CHEK1	ACRV1	PATE1	PATE2	PATE3	PATE4	HYLS1	PUS3	DDX25	LOC338667
CDON	RPUSD4	FAM118B	SRPR	FOXRED1	TIRAP	DCPS	FLJ39051	ST3GAL4	KIRREL3
MIR3167	ETS1	FLI1	KCNJ1	KCNJ5	C11orf45	TP53AIP1	ARHGAP32	ZNF123P	BARX2
TMEM45B	NFRKB	PRDM10	LINC00167	APLP2	ST14	ZBTB44	LOC100996646	ADAMTS8	ADAMTS15
C11orf44	LOC100507431	SNX19	NTM	C11orf39	OPCML	SPATA19	LOC283174	MIR4697	LOC646543
IGSF9B	LOC100128239	JAM3	AK125040	MRGPRX1	CAT	KRT6C	TIMM9	TOMM20L	CYFIP1
LOC283683	NIPA1	NIPA2	TUBGCP5	WHAMMP3	HERC2	EIF3J	SPG11	NTRK3	AK296148
AK301679	AKAP10	ALDH3A1	ALDH3A2	ALKBH5	ATPAF2	AX747021	AX748015	AX748411	B9D1
BC029877	BC038464	BC040601	BC058012	BC112347	BC150162	C17orf39	CCDC144B	CCDC144C	COPS3
DL490867	DM110819	DQ570768	DQ571055	DQ572309	DQ573130	DQ576083	DQ578124	DQ578661	DQ579174
DQ579248	DQ584883	DQ585755	DQ585853	DQ586004	DQ586114	DQ592578	DQ592713	DQ593368	DQ595299
DQ595911	DQ596932	DRG2	EPN2	EVPLL	FAM106A	FAM18B1	FAM83G	FBXW10	FLCN
FLII	FLJ00050	FOXO3B	GRAP	GRAPL	LGALS9C	LLGL1	LOC339240	LRRC48	MAPK7
MED9	MFAP4	MIR1180	MIR33B	MPRIP	MYO15A	NT5M	PEMT	PLD6	PRPSAP2
RAI1	RASD1	RNF112	SHMT1	SLC47A1	SLC47A2	SLC5A10	SMCR5	SMCR7	SMCR8
SMCR9	SNORA59B	SPECC1	SREBF1	SREBP-1	TBC1D28	TNFRSF13B	TOM1L2	TOP3A	TRIM16L
TRNA	ULK2	USP32P2	Y_RNA	ZNF286B	DHRS7B	TMEM11	C17orf103	MAP2K3	KCNJ12
C17orf51	FAM27L	FLJ36000	MTRNR2L1	SSH2	TOP2A	RARA	JUP	EIF1	ZNF826P
ZNF546	ZNF780A	ZNF780B	C5AR1	C5AR2	DHX34	GPR77	MACROD2	CRKL	TBX1
APOL5	APOL6	RBFOX2	CYP2D6	NDUFA6	DDX53	LOC100873065	TMEM185A	ABHD16B	AHNAK
AKAP12	AKNA	AL591025.1	CATSPERG	CCDC88B	CHD7	COL11A2	COL4A2-AS2	CUL3	DNAH10
DNAJB1	ELAVL4	ETS1	ETS1	FAM161B	FTSJ3	GPR113	GPSM2	GREB1L	GRINA
GTPBP4	GZMB	IQCB1	KRTAP4-9	LIG1	LIPJ	MED20	MLL2	MSLNL	MUC17
NAA15	NAA15	NAGS	NAT8L	NCKAP1	NTM	OR52H1	OS9	PCDHB11	PLCZ1
PPL	QRICH2	RNF20	SCLY	SERHL2	SERPINA9	SERPINH1	SHANK1	SLC5A2	SLC7A10
SMAD2	TMEM190	TMEM237	TRAPPC2L	UCHL3	ZFR2	ZNF521	CALML6	EFHD2	TMEM54
ZNF326	BCL9	TARS2	ECM1	TOMM40L	LAMC1	USP34	BCL2L11	MKRN2	FYCO1
ALS2CL	KIAA2018	GHSR	TNK2	FABP2	NKD2	COL4A3BP	FBN2	UBE2B	GABRA6
RNF44	FGFR4	FAM135A	DSE	EPDR1	GRM8	RP1L1	LOXL2	MAK16	UNC5D
EIF3H	MTSS1	KIAA0196	CPSF1	PDCD1LG2	CCDC171	NR6A1	ZNF79	MASTL	JMJD1C
RUFY2	UNC5B	WDR96	KNDC1	PIDD	USH1C	C11orf9	GANAB	ATG2A	TIGD3
CLPB	HEPACAM	KDM5A	ITGA7	RDH5	SBNO1	FREM2	NAA16	PCDH17	TBC1D4
MYO16	KIAA1737	YY1	SERINC4	CYP11A1	IGF1R	MAPK8IP3	PKD1	SPN	PDP2
DHX38	JPH3	NEURL4	ALOX15B	CPD	RAB11FIP4	OSBPL7	KCNH6	KCNJ16	DSG2
SMAD2	SMAD4	RNF152	TSHZ1	ZNF653	KLF2	GLT25D1	ZNF536	SIPA1L3	B3GNT8
PVRL2	LIG1	NUCB1	NUP62	SIGLEC5	DEFB128	SPATA2	NFATC2	LAMA5	NKAIN4
KCNJ15	UMODL1	LZTR1	MTMR3	PES1	IL2RB	SHROOM2	ZNF630	HUWE1	

(6) Published “causative” genes from previously reported CHD CNVs intervals in addition to all genes in novel CHD CNV intervals (e.g., list 2) and protein altering SNVs.

IFI44L	IFI44	PIAS3	NBPF10	RNF115	NUDT17	CD160	POLR3C	PDZK1	GPR89A
AX747161	LOC100288142	PRKAB2	PDIA3P	LOC100505794	FMO5	CHD1L	LINC00624	BCL9	ACP6
GJA5	GJA8	GPR89B	OR2T10	OR2T11	JUN	LOC100131060	ASPRV1	LOC100133985	MXD1
PCBP1	ZEB2	CHL1	CNTN6	CNTN4	IL5RA	TRNT1	CRBN	LRRN1	SETMAR
SUMF1	ITPR1	EGOT	BHLHE40	ARL8B	EDEM1	MIR4790	GRM7	LOC100996542	LMCD1
LINC00312	SSUH2	CAV3	OXTR	RAD18	SRGAP3	THUMPD3	SETD5	LOC100996546	LHFPL4
MTMR14	CPNE9	BRPF1	OGG1	CAMK1	TADA3	ARPC4	TTLL3	RPUSD3	CIDEC
JAGN1	IL17RE	IL17RC	CRELD1	PRRT3	EMC3	LOC401052	FANCD2	FANCD2OS	BRK1
VHL	IRAK2	TATDN2	LINC00852	GHRL	GHRLOS	SEC13	ATP2B2	MIR378B	MIR885
LINC00606	SLC6A11	SLC6A1	HRH1	ATG7	VGLL4	LOC100133039	HACL1	BTD	CCDC111
MLF1IP	UBE2B	POLR1C	CDKN1A	C7orf72	TCRVB	TCRVB2S1	TRBV	TCRBV19S1P	TCRBV15S1
MTRNR2L6	TCRBV14S1	TCRBV11S1A1T	AX748042	TCRBV3S1	TCRBV4S1A1T	PRSS1	GATA4	EXT1	RLN1
RLN2	ACADSB	ADAM12	ARMS2	BCCIP	BTBD16	BUB3	C10orf88	C10orf90	C10orf120
C10orf137	CLRN3	CTBP2	CPXM2	CHST15	CUZD1	DHX32	DMBT1	DOCK1	EBF3
FAM24B-CUZD1	FAM175B	FAM196A	FAM24B	FAM24A	FAM53B	FANK1	FLJ37035	FOXI2	GPR26
HMX3	HMX2	HTRA1	IKZF5	LHPP	LOC100506255	LOC387720	LOC283038	LOC100169752	LOC728065
METTL10	MIR4296	MIR3941	MMP21	MIR4484	MKI67	MGMT	MIR4297	NPSPTPRE	NKX1-2
OAT	OATP1	PLEKHA1	PSTK	TACC2	TEX36	UROS	ZRANB1	AF113016	BC010924
EHBP1L1	FAM89B	FRMD8	KCNK7	LOC254100	LTBP3	MALAT1	MAP3K11	MIR4690	MascRNA_menRNA
NEAT1	PCNXL3	SCYL1	SIPA1	SSSCA1	U7	TMEM135	PKNOX2	FEZ1	MGC39545
EI24	STT3A	CHEK1	ACRV1	PATE1	PATE2	PATE3	PATE4	HYLS1	PUS3
DDX25	LOC338667	CDON	RPUSD4	FAM118B	SRPR	FOXRED1	TIRAP	DCPS	FLJ39051
ST3GAL4	KIRREL3	MIR3167	ETS1	FLI1	KCNJ1	KCNJ5	C11orf45	TP53AIP1	ARHGAP32
ZNF123P	BARX2	TMEM45B	NFRKB	PRDM10	LINC00167	APLP2	ST14	ZBTB44	LOC100996646
ADAMTS8	ADAMTS15	C11orf44	LOC100507431	SNX19	NTM	C11orf39	OPCML	SPATA19	LOC283174
MIR4697	LOC646543	IGSF9B	LOC100128239	JAM3	AK125040	MRGPRX1	CAT	KRT6C	TIMM9
TOMM20L	CYFIP1	LOC283683	NIPA1	NIPA2	TUBGCP5	WHAMMP3	HERC2	EIF3J	SPG11
NTRK3	TRNA	DHRS7B	TMEM11	C17orf103	MAP2K3	KCNJ12	C17orf51	FAM27L	FLJ36000
MTRNR2L1	SSH2	TOP2A	RARA	JUP	EIF1	ZNF826P	ZNF546	ZNF780A	ZNF780B
C5AR1	C5AR2	DHX34	GPR77	MACROD2	TBX1	APOL5	APOL6	RBFOX2	CYP2D6
NDUFA6	DDX53	LOC100873065	TMEM185A	NKX2-5	ABHD11	BAZ1B	BC050599	BCL7B	CLDN3
CLDN4	CLIP2	DNAJC30	EIF4H	ELN	FKBP6	FZD9	LAT2	LIMK1	MIR4284
MIR590	MLXIPL	NSUN5	RFC2	STX1A	TBL2	TRIM50	VPS37D	WBSCR22	WBSCR27
WBSCR28	hCPE-R	RAI1	ABHD16B	AHNAK	AKAP12	AKNA	AL591025.1	CATSPERG	CCDC88B
CHD7	COL11A2	COL4A2-AS2	CUL3	DNAH10	DNAJB1	ELAVL4	ETS1	FAM161B	FTSJ3
GPR113	GPSM2	GREB1L	GRINA	GTPBP4	GZMB	IQCB1	KRTAP4-9	LIG1	LIPJ
MED20	MLL2	MSLNL	MUC17	NAA15	NAGS	NAT8L	NCKAP1	NTM	OR52H1
OS9	PCDHB11	PLCZ1	PPL	QRICH2	RNF20	SCLY	SERHL2	SERPINA9	SERPINH1
SHANK1	SLC5A2	SLC7A10	SMAD2	TMEM190	TMEM237	TRAPPC2L	UCHL3	ZFR2	ZNF521
CALML6	EFHD2	TMEM54	ZNF326	BCL9	TARS2	ECM1	TOMM40L	LAMC1	USP34
BCL2L11	MKRN2	FYCO1	ALS2CL	KIAA2018	GHSR	TNK2	FABP2	NKD2	COL4A3BP
FBN2	UBE2B	GABRA6	RNF44	FGFR4	FAM135A	DSE	EPDR1	GRM8	RP1L1
LOXL2	MAK16	UNC5D	EIF3H	MTSS1	KIAA0196	CPSF1	PDCD1LG2	CCDC171	NR6A1
ZNF79	MASTL	JMJD1C	RUFY2	UNC5B	WDR96	KNDC1	PIDD	USH1C	C11orf9
GANAB	ATG2A	TIGD3	CLPB	HEPACAM	KDM5A	ITGA7	RDH5	SBNO1	FREM2
NAA16	PCDH17	TBC1D4	MYO16	KIAA1737	YY1	SERINC4	CYP11A1	IGF1R	MAPK8IP3
PKD1	SPN	PDP2	DHX38	JPH3	NEURL4	ALOX15B	CPD	RAB11FIP4	OSBPL7
KCNH6	KCNJ16	DSG2	SMAD4	RNF152	TSHZ1	ZNF653	KLF2	GLT25D1	ZNF536
SIPA1L3	B3GNT8	PVRL2	NUCB1	NUP62	SIGLEC5	DEFB128	SPATA2	NFATC2	LAMA5
NKAIN4	KCNJ15	UMODL1	LZTR1	MTMR3	PES1	IL2RB	SHROOM2	ZNF630	HUWE1

Online Table V. Probands with CNVs previously identified by clinical cytogenetics or microarray used as positive controls for CNV detection by SNP Array and WES

ID	*Chrom*	*Band*	*Size (kb)*	*Copy number*[1]	*Clinical syndrome/ significant gene*	*Analysis*[2]	*Cardiac Lesion (diagnosis)*	*Parental origin*[3]
1-01179	10	q26.13-q26.2	7161.2	1	10q terminal del	A E	CTD (TA)	M
1-02527	22	q11.2	2834.1	1	DiGeorge *TBX1*	A E	CTD (TOF)	P
1-01617	17	p11.2	3533.7	3	Smith-Magenis *RAI1*	A E	LVOT (HLHS)	
1-02313	22	q11.2	2574.7	1	DiGeorge *TBX1*	A E	CTD (TOF)	M

[1]Copy number: 1- deletion; 3- duplication
[2]Analysis: A- identified with SNP Array; E- identified with WES
[3]Parental Origin: M- maternal chromosome; P- paternal chromosome
Abbreviations: CTD-conotruncal defect; LVOT-Left Ventricular Outflow Tract Obstruction; TA-truncus arteriosus; TOF-tetralogy of Fallot; HLHS-hypoplastic left heart syndrome

Online Table VI. Extracardiac manifestations in confirmed de novo CNVs in discovery cohort. Genomic coordinates refer to hg19.

Blinded ID	Chrom	Start	End	Band	Copy number	Syndrome/ significant gene	Extracardiac
1-01486	1	194201171	194304070	q24.2-q25	3		Strabismus, short stature
1-00762	3	60661	11712230	p26.1	3		Cleft palate, Pierre Robin sequence, significant hearing loss, developmental delay
1-02093	3	197143652	197186111	q29	3		Persistent feeding difficulties, less than the first percentile for both height and weight.
1-00771	4	185603346	185638397	q34.1	1		Urethral stricture
1-00113	5	133706994	133730455	q31.1	1		Hypocalcemia
1-00296	5	166386727	173073664	q34-q35.2	1	*NKX2.5*	Dysmorphic face, strabismus
1-00096	7	50179707	50191153	p12.2	1		Microcephalic, esotropia, choanal atresia, imperforate anus, vaginal atresia, abnormal thumb, overfolded ears, small mouth, highly arched palate, sacral dimple, left proximal thumb placement.
1-00800	7	72719386	74138603	q11.23	1	Williams syndrome	Periorbital fullness, wide mouth, full cheeks, full lips
1-00540	7	72721123	74140708	q11.23	1	Williams syndrome	Asymetrical eye size, hypercalcemia
1-02625	8	8102183	12190106	p23.1	3	*GATA4*	Bilateral slipped capital femoral epiphysis
1-00948	8	119053343	119064098	q24.1	1		Bilateral pelvicaliectasis
1-02360	9	5302500	5337760	p24.1	3		Developmental delay, G-tube dependent, lactic acidosis
1-00561	11	18949220	18956690	p15.1	1		Prominent forehead, micrognathia, congenital hip dsyplasia, bilaternal congenital knee dislocations, abnormal fingers
1-00230	11	86939592	87025456	q14.2	1		Short stature
1-01486	11	125641368	134943190	q24.2-q25	1	Jacobsen / *ETS1*	Strabismus, mildly dysmporphic, hypertelorism
1-00243	15	22835893	23062345	q11.2	1		Pectus excavatum
1-01427	17	21562473	22252439	p11.2	1		Polysplenia, malrotation, transverse liver, G-tube

							dependent, behavioral problems
1-00561	17	27962393	28099002	q11.2	1		Prominent forehead, micrognathia, congenital hip dsyplasia, bilaternal congenital knee dislocations, abnormal fingers
1-01588	18	65138642	78015180	q22.1-q23	1	*NFATC1*	Cleft palate, downward slanting palpebral fissures, low set ears, prominant occiput and wide anterior/posterior fontanelle, excess nucal skin, clenched fist, single palmar crease, tapering fingers, clubbed feet
1-02170	19	20601006	20717536	p12	1		Seizures, learning disabilities
1-00174	19	40515744	40681387	q13.2	1		High myopia, nystagmus, right-sided uterocele, non-functioning right kidney, bladder outlet obstruction
1-01194	22	18844632	21500000	q11.2	1	DiGeorge / *TBX1*	Hirsuit, short downslanting palpebral fissures, large posteriorly rotated ears with overlapping pinna, short neck, wide spaced nipples, imperforate anus, sacral agensis
1-00113	22	18886915	22000000	q11.2	1	DiGeorge / *TBX1*	Hypocalcemia
1-00988	22	20733495	21464479	q11.2	1	DiGeorge / *TBX1*	Hooded eyelids, hypospadias
1-01427	22	42522638	42531210	q13.2	3		Polysplenia, malrotation, transverse liver, G-tube dependent, behavioral problems
1-00197	X	148685645	148693146	q28	3		Osteopenia, sacral dimple and hypothyroid

Online Table VII. High Heart Expressed Genes

CSTF2	EI24	TFRC	FAM160B2	ING4	SHROOM3	ARF3	PRMT1	EIF3A
YIF1A	MRPS28	DUS3L	COMT	LRRC41	HSD17B4	USMG5	DDX6	PALLD
HEATR5B	FBXL5	MAP2K7	ATP6V0E1	UBR3	HIST1H2AB	DCTN2	PDAP1	DDB1
PSAT1	TK1	RQCD1	EFNB2	AHCYL1	UBR5	NUCB1	PCDHGA10	SERBP1
NEURL1B	KIAA0317	RFK	BMS1	DVL2	LRPPRC	KLHDC8B	MAP1B	DDX17
THOC1	SMAD4	VAMP3	DNAJC11	RBMS1	FLOT2	NSA2	LAMP1	HSPB1
LIG3	BCL7A	SPATA13	PPP1R12B	AP1S1	RPL27	SQSTM1	WDR38	MDH2
CLTB	PACS1	TSPAN2	THBS2	GPBP1	HIST1H2AE	TNFAIP1	MSI2	ENO1
SCAF4	XIAP	SFSWAP	NCOA3	CD34	ZMYM3	ITM2A	PEA15	ANP32A
IMPACT	PRPF40B	PLOD2	VWF	TBRG1	FIS1	PPP2R4	PKP2	CANX
YTHDF2	PCMTD1	GNG11	LASP1	WIZ	EPB41L3	PSMD8	PPIB	SORBS2
FAM114A1	KDM3A	MED21	FITM1	MCM7	Mobkl3	NDRG4	OXCT1	RPL32
NSMAF	PTBP2	NTAN1	QSOX1	SNRPG	EPN1	PUM1	Tubb2c	NDUFV1
ASXL2	SPG7	KIAA0101	XPR1	LRRC39	UBE2M	SERINC1	PCDHGA12	CHD4
AHCTF1	KAT7	ARL3	ATP2C1	TET3	CHD2	B2M	SLC16A1	CCND2
PARP2	TMEM11	C20orf112	RPS12	SHOC2	NFAT5	TNRC6A	MICAL3	RPL22
ORAOV1	NSUN2	OTUD4	FAM127B	MEPCE	DDX42	YAP1	C22orf32	PRRC2A
XRCC6	MGA	KIAA1429	FBXO21	PTCD3	C19orf42	DLAT	GPX1	MICAL2
ADSL	CHPF2	CDC34	RNF2	SLC12A7	STX12	PPP1R12A	SEC31A	TPT1
UBE4A	CCDC90B	SYNGR2	LIMK2	ITGB5	MFAP2	GLYR1	PAIP2	GPI
PMP22	AQR	PPM1A	PLEKHA1	CTC1	NEO1	RSRC2	RALY	HNRNPK
ILKAP	FOXJ3	MFAP1	ARHGAP23	TGFBR3	SMC4	ARPC5	ARPP19	SON
PML	CSK	MBD2	GEMIN4	PMPCB	CLU	CPE	TUBB2B	HBZ
OLFML2B	TEF	ZEB2	HIGD1A	DCTN4	EIF3F	EIF4E2	TBCA	COX8A
CEP170	LHFP	ATL2	TLK1	CCDC23	CREBBP	IPO9	C1orf63	RPLP0
NFS1	HYAL2	NLRP10	ZNHIT1	AHDC1	TAX1BP1	UTRN	RPS17	RBM24
SREBF1	SMG1	PPDPF	SPG20	DARS	ELAVL1	WBP11	PPP2R1A	AKAP2
KAT5	OSBPL5	ULK1	VAT1	LAMA2	PECAM1	COPB2	SF1	EIF4G1
ARRB1	PRPF6	FAM73B	SART3	PRKACB	ATP11A	CHD7	PDHB	NISCH
PRRG3	PIGS	KAT6B	SNRK	TMEM147	PSMC4	ANKRD17	NDRG2	ACTG1
RBX1	DBT	DAP	TOMM40L	PCDH7	PTPRK	RBM39	SOD2	COX6B1
PFAS	GMPS	KIAA0947	LGMN	GNG2	SLC25A51	FKBP9	GLO1	RPL13A
IMPA1	PPP1R8	SPATS2	EPHA7	SMG5	FKBP2	GDI1	TGFB1I1	PLN
NOTCH3	MBNL1	KCMF1	ZNF740	ARL2	C1QBP	ATP5G2	ATXN7L3B	LGALS1
CHKB	CDH13	CGGBP1	KIAA1211	ARMCX3	SMYD2	SSB	TRIM28	NDUFV2
C1orf27	DLG4	FAM69A	Scand3	DNLZ	BTF3	SLC16A3	SMC4	RPS5
ZNF652	ALCAM	PRKD3	CPSF6	ARF6	RBBP6	RAB8A	HSBP1	PPP1CA
CEP120	TMBIM1	COX10	C10orf2	NDFIP1	OSBP	AHSA1	CCDC141	RPS28
MTF2	WDR77	PAN2	TRIB2	PPFIBP1	VPS26A	RRM1	PHB2	COL3A1
FMR1	TTC14	PLIN3	ASNSD1	DTX3	HPRT1	BTG1	NDUFA10	CALR
ANKRD10	ARL1	SIK3	CAPN12	KLHL24	RAD23A	FHL2	RAN	LAMC1
FBXO30	MRPL24	CD99L2	NOS3	ATP6V1B2	RBMX	SPAG7	NDUFA8	CS
CCNL1	TMEM108	ZNF512	MAP2K2	CXADR	IRF2BP2	CNOT1	PSMB7	PGAM2
KIAA2013	ALDH1L2	CCNK	SEC24D	CTCF	SF3B3	STT3B	HNRPDL	NDUFV3
RHOBTB1	HDAC9	GAPVD1	NUP62	PLSCR3	PPIF	DAAM1	PXDN	NFIB
SPCS2	HOMER3	SNX1	SEH1L	PRDX4	PPP5C	DBI	CAND2	COX7A2
HIC1	SSBP2	SMS	GOLGA7	FAM78A	AGL	RAB18	TEAD2	RRBP1
C2orf43	CPNE5	SMPD4	MBD3	HMG20A	AK2	HSPB2	NDUFA11	COL4A2
LYPLA2	SLC35F5	HNRPLL	U2AF1	FNBP1L	HSD17B10	SORT1	RTN3	CD81
MOSPD3	MRPS35	ATAD3A	HS6ST1	DCTN6	NDUFAF2	HAND2	PCBP2	LDHA
ZKSCAN3	MRPL30	MRPS18A	CD36	SS18L2	ETFDH	MEF2D	CHCHD3	DES
POLE4	RNPEPL1	NUP85	RABGGTB	TNKS	CTBP1	SIN3B	EIF3L	ATP5D
YARS	ATG9A	RPL9	GTF2H5	GPRASP1	MCM3	ARHGAP35	ALPK3	PTRF
SENP2	ZNF521	PPT1	PNRC1	USP7	GAB1	FLII	FAM120A	COX5A
ZDHHC3	ZDHHC9	EIF4ENIF1	MPHOSPH8	Brp44	ACTR2	ASPH	HRC	COX7B
IMPAD1	AGPS	MMADHC	Erh	CLPTM1	TRA2A	NDUFS8	C5orf15	EEF1B2
COX15	PPP4R1	ATP2B4	FIP1L1	CCNB2	DROSHA	ADAM10	LRRFIP1	UQCR10
PAIP2B	CRAMP1L	MRPL2	CDC27	GSTP1	EIF3D	MICALL1	NNT	CDKN1C
ZNF462	PPFIA4	GNG12	SUV420H2	MAP1A	PRR12	SLC25A23	RPS24	HNRNPU
LYRM5	ZNF384	USP14	SUZ12	CLIP4	SRPK2	IQSEC1	FBXL22	RPS16
POGK	PATL1	EZH2	SUGT1	LONP1	QRICH1	TIMM23	ARHGAP31	GYG1
VPRBP	NUSAP1	ATAT1	GRB2	RBM14	NF2	PNISR	RPL37A	COX4I1
EPS15	RALGPS2	CISD3	GTF2A2	NCAPH2	AKAP1	SSR1	GABARAP	NACA
SYPL1	DHX36	VPS4A	LUZP1	COPS3	TEK	NDUFC1	PDLIM5	SMPX
PGLS	LRRC47	RPA3	ARID2	TXLNB	CRAT	COPS5	BYSL	HSPA8
SYVN1	TPD52L2	PPM1B	NOTCH2	ARHGAP11A	KRTCAP2	TCF4	GDI2	GNB2L1
ZNF768	MAN2B1	TRIM41	CPSF1	C15orf23	MEF2C	HMGN1	NRP1	FLNC
KIFAP3	CLIC5	SCAF8	FDFT1	PMF1	MAP1LC3B	ERBB2	KDELR1	SPTBN1
RRAGA	NAA20	EMC4	SRP72	PLXNB1	UNC5B	PEX19	INPPL1	ATP5J
MAPK9	ZNF710	CTR9	EHMT1	AKT1S1	COPB1	KIF13A	SMARCC1	FXYD1
GAK	MFN1	PHTF2	DBNL	MGST3	ADCY5	PRPF4B	CHTF8	UQCRQ
GTF2A1	PDCL	IPO4	ZCCHC8	ACOT2	IRF2BPL	AAMP	PSMC2	FSTL1
DHX38	UHRF2	RALBP1	Vps24	ID3	GATAD2A	PUM2	SVIL	SRRM2
KIAA0247	CLCN4	PQBP1	FAM115A	DDX1	FYCO1	FHL1	HNRNPUL1	SLC25A3
ZNF436	ITGA9	WHSC1L1	SETD2	PCNXL3	HJURP	GANAB	ATXN2L	MYOM1
STK16	FBLN2	COG4	ERC1	OTUD7B	JAK1	ADPRHL1	MTSS1	NEXN
NPEPL1	NR3C1	SLC4A2	EP400	PTGR2	SUMO3	DAZAP2	NDUFB5	FTH1
RBM18	SV2A	KIAA0391	KBTBD10	TIMM8B	JMJD8	NASP	SRRM1	Rps9
ELOF1	NUP107	AKT2	ESRRA	ACADS	GNAQ	VCAN	HIGD2A	COX6A2

TNIP1	RNF114	MAPK14	TMEM214	CDC20	SLK	FASN	Rpl23a	NDUFA3
MRPS27	WSB2	PTBP3	TOPBP1	PLA2G16	ATOX1	SOX12	HSP90AA1	SCD
PCIF1	WBP1	PCNT	EFHA1	EFHD2	SUPT16H	LAMB2	SMARCA4	LDB3
ERCC1	RBM22	ACP6	NKIRAS2	ACOX1	NARS	OAT	TOMM7	TUBA1C
Brp44l	MRPS6	PRKRIR	KLHL21	UBN2	TMA7	RCN1	TNRC6B	FUS
SLC25A20	SLC29A1	SPECC1L	DIDO1	VAMP2	ANGPT1	PLVAP	COX6C	VDAC1
SLC9A1	TARS	PTK2	VPS35	RUFY2	NPTN	EZR	ADD1	UQCR11
RANBP2	CDK12	ATP1B3	SPPL3	RIOK3	HDAC5	RPN2	RPL30	PTMA
ADORA1	ADAM11	EPS15L1	GALK1	ENTPD4	ILF2	BRD2	MKI67	RPL4
CCNB1	ARHGAP18	MRPS17	INTS1	TSEN34	SEC62	GABPA	EIF5B	ATP5E
TNFRSF1A	TIMM10	PPP1R11	AP3B1	PRPF40A	KDM4A	PIEZO1	KIF1C	NDUFA4
SP2	PWWP2A	ASPSCR1	EMP2	NOTCH1	CDKN1B	LPHN1	CAPZA2	HCFC1R1
GRAMD1A	ADRBK1	TOR1AIP1	FABP4	RRAGD	CSNK1G2	SRRT	TCEAL8	ITGB1
TMEM33	LSM6	MRPS15	CTBP2	ELOVL5	SAR1A	DOCK6	UBAP2L	EIF4G2
TCEA1	DCPS	ACTR10	APC	IVD	FKBP10	MAPRE1	ACADL	TUBA1A
GPSM1	PTPN9	HIST1H2AK	WWC2	INPP5A	BMP7	SLC25A12	FXR1	CRIP2
PANK4	WDR43	TNFRSF12A	NHSL1	ZDHHC5	ANLN	RAVER1	VCL	HSPG2
USP48	MRPL9	YTHDF1	CHSY1	SNRNP27	PDCD6IP	FURIN	SSR3	DSTN
ITGAV	SKIV2L2	NFX1	PCGF2	CXCL12	CORO1C	CSNK1D	GATA4	GNAS
C11orf48	DLG1	USP20	SNRPD1	CARM1	C7orf59	ALPK2	LAMA5	COL4A1
MAD2L1	UBE2D1	C11orf49	CCDC50	PHLDB2	SMTNL2	LUC7L3	MAP4K4	NDUFA12
ANAPC4	PERP	ASAH1	MRPL33	MESDC2	F2R	BPTF	PEG3	PFN1
JOSD1	TCF19	KEAP1	DCAKD	NUP88	STK25	NDUFB4	NDUFS2	ATP1A1
FGFR1	IMP3	ISCA2	PCYOX1	DGCR2	TXNL4A	UBE4B	THRAP3	RPS29
RHOG	FZD1	EIF3J	MFHAS1	VPS29	NOP58	VAPA	ECE1	CKM
BCL2L11	COL6A6	AGAP1	SNX4	FAM40B	MEF2A	SMC1A	SNRPD3	UQCRH
FBXO38	CITED2	ALAS1	ERI3	KLF13	MARK2	APRT	NID1	GPX3
NOP14	PPM1K	C6orf48	MLLT10	PODXL	TRIM44	RAB1B	SET	VIM
NUP50	GPR107	ELF2	CD63	PRMT5	ITGA6	ALKBH5	NDUFB6	RPS3
PPP1R12C	CAPN7	C4orf3	MRPL51	PTTG1IP	POLR3H	TAGLN2	C17orf61	SERPINH1
HIAT1	MUM1	SYDE1	EXOSC10	U2SURP	SMC6	SETD7	SFPQ	PABPC1
CDC25A	Iqcj-schip1	SLC48A1	MED16	RNASEK	TAF15	FOXP1	ILK	EIF1
MCRS1	BOLA3	PCMT1	MRPL49	TMEM63B	KCNH2	ZMYND11	MATR3	TPI1
ZNF638	UBL4A	TBC1D9B	HMG20B	ZNF704	PSMC5	TMBIM6	PRPF8	Pkm2
HGS	ARID4A	SPCS3	FOXP4	KDM1A	SNRPA	TRIM35	JUND	RPL41
UGCG	PPIL2	TMEM55A	SUN2	AXL	CUL4A	LRRC4B	CDC42	TUBA1B
SALL1	WASL	ABCE1	IGF2BP2	SDC3	NUTF2	SUPT5H	TSTD2	SLC8A1
ABCA3	PON2	KLHL30	TMEM131	PLOD3	ANXA5	CNIH	PLXNB2	CSRP3
TOR1B	TOX4	RBM17	COLEC12	POM121	USP9X	FAT1	DYNC1I2	TMSB4X
BPGM	PPP4R2	SLC4A3	SELM	RREB1	GSPT1	SLMAP	NFIA	ACTA2
RXRB	ADAMTS10	DAPK3	SF3A2	MINK1	APH1A	SRP54	DNMT3A	IDH2
HERPUD1	PARM1	CCDC34	BMPR1A	ZCCHC24	FAM32A	MKNK2	NAA15	HBE1
SMARCD1	FAM193A	MRPS5	RMND5A	RPN1	CPD	TRA2B	FKBP1A	RPS20
IFRD2	SNURF	PPP1R14B	SETD8	MAN2A2	PHB	CKAP4	CASQ1	HSPB7
FBXW2	TP53I11	YTHDC1	FAM108C1	CKAP2L	RCC2	EIF5	HN1	LBH
GTF3C5	HEG1	RXRA	DHX30	POLDIP2	SMG7	ADIPOR1	MAT2A	YWHAE
RNF126	RBCK1	CLSTN1	AGPAT6	TPX2	PRELID1	EPRS	YWHAQ	RPL8
ISOC1	PPP3CA	FAM111A	USP34	ARAP3	COL6A1	NOP10	PGK1	ANKRD1
PPP2CB	GTF3C6	PSENEN	MRPL38	FAM65A	LPP	PRNP	PGM5	ACTN2
MAN2C1	ST5	HDGFRP2	UBE2G2	SRSF10	CHTOP	LSP1	RPL36AL	FN1
NDUFAF4	QSER1	STAG2	GAS1	AFF4	PAFAH1B2	FASTK	ITM2B	NEDD4
KHDRBS3	CDC42SE1	Hddc3	CYB5A	RNF4	NACC1	NDUFA7	MAPRE2	TNNI3
SH3BGRL3	UBR7	ETNK1	CAT	TSC22D3	ARHGEF2	DCTN1	TCF25	EEF2
MRPL53	NUP153	GIPC1	LRRC8A	RAMP2	DNAJC5	MAPK8IP3	SH3PXD2A	HNRNPA2B1
Camsap1l1	POMGNT1	GPC4	BCCIP	ZNF862	PRC1	PTOV1	RPL12	SPARC
ETV6	MRPL11	C1orf21	H2AFY	CNIH4	RPS27L	LMAN1	MLL2	ATP5G3
ZMYM4	EYA3	USO1	CIZ1	PTP4A1	FAM168A	EIF3B	COX17	IGFBP5
UBE2O	RBM33	S100A10	ERLIN2	EP300	PFDN5	NFIC	IPO5	GRB10
SELENBP1	MLXIP	ANKFY1	ODF2	HIP1	SUPT6H	SLC39A7	TP53INP2	MYBPC3
PGP	PTCD2	USP24	DPP8	ANKHD1	GLTSCR2	KDM5C	CTDSP2	ATP5A1
LONP2	NCBP1	NF1	CNOT2	AP2A1	PITPNA	PRDX1	PFKM	PAM
DGCR8	SERTAD2	DYNC1LI1	LZTS2	LRRFIP2	KCTD10	GCN1L1	CALM2	CFL1
PEAK1	HTATSF1	BAP1	PREPL	YTHDF3	IL6ST	STX4	IDH3G	MYL3
TFDP1	MTX2	SEC63	PRDM2	PPARGC1A	UBE2S	DNAJA2	CNBP	TUBB
KCTD20	YY1	SMG6	PHF12	CTSZ	ARHGAP29	NDUFS3	FSD2	MEST
DGKK	KPNA4	ZNF326	HAUS1	MARK3	UBE2K	NUDC	SULF2	HSP90AB1
DNAJB6	SLC27A4	ANAPC2	PTPRA	CAV3	BIRC5	BGN	DDR1	IGF2R
MPND	DCAF6	MRPL37	POLR2L	SRI	SYNCRIP	IGF2BP1	TCP1	ATP5B
DHDDS	ELP2	EIF2B2	NUP205	EEF1D	IPO7	ARMCX2	FAM162A	SLC25A4
GNL3L	SLC22A23	TCEB3	FNBP1	SIK1	SMARCD3	PFKP	RBBP4	TNNI1
TOMM5	CARHSP1	RPS6KB1	DDX19A	UBN1	DUSP27	GLG1	ALDH1B1	ATP2A2
LIX1L	UBA5	LLGL1	C16orf80	DCN	CHCHD1	UBE2L3	PCBP1	WDR89
DOHH	BTBD2	MRPS25	PPP1R1A	ARL6IP5	FBXO18	PSMA6	RHOA	FTL
RBMS3	MAGI1	CPSF2	SULF1	TACC1	GNAO1	CASZ1	RANBP1	TTN
OTUD6B	CELSR1	HERC1	C19orf48	TM9SF4	SSR4	UBE2I	NAV2	TNNC1
GPC2	GNL3	IFLTD1	DNAJA3	Kiaa1310	AGXT2L2	ARPC2	ADD3	ACTB
HBS1L	ATP6V1C1	GUCY1B3	NHP2L1	GLUD1	XPO1	KDELR2	MFN2	TPM1
TOB2	GFM2	MYOM2	HBG1	QARS	ZNF609	PTGES3	METAP2	RPLP1
WIPF2	CYTH3	ZC3HAV1L	TKT	SH3PXD2B	S100A1	RPL37	HADH	MYL9
TIMM17B	ZNF445	PPP2R5E	COL4A6	RNF14	ARHGEF12	TIMP2	RNF10	NPPA
NAGK	UBXN7	PRND	ARHGAP21	RARS	PALM2	C14orf166	LAMA4	EEF1A1
RYR3	PSMG1	DTYMK	ARHGEF40	MED12	C6orf106	CERS2	FKBP3	MYL4

ID1	ASH2L	SH3GL1	SYNJ2BP	ARL8B	SORBS1	COPZ1	CLIC1	MYL7
POLR3GL	MSH6	TCN2	SLC30A9	RNF7	C17orf49	SSPN	ABCF1	MYL2
COQ6	BCAT2	GLRX	MARS	CDK2	C21orf33	ZFP106	ANP32E	MYH6
IRF2BP1	SLC35A4	FAM20B	RHEB	SGTA	TXNDC12	PTPN11	UNC45B	TNNT2
CNPPD1	MYPN	NEK7	CYFIP1	TXNDC17	IFNGR2	AKAP12	MLF2	ACTC1
ZC3H18	PSEN1	ADAMTS1	DYNLT3	HNRNPH2	C5orf24	TJP1	COL4A5	MYH7
TMEM50A	PRR13	MRPL39	CPPED1	PGRMC2	MAP3K12	SARS	TARDBP	IGF2
ASCC2	ALDH7A1	TRAM1	BTF3L4	PREP	NES	SLC38A1	SF3B2	HBA1
SNCA	PHF20	HAT1	TPP2	SBF2	SNRNP200	NT5DC2	H2AFX	
CNOT10	ATAD2	NAA35	MTA1	RNF40	DEGS1	CAPNS1	TTC3	
ACO1	Thoc4	ARVCF	FLRT2	TMEM184B	SEC14L1	RPL35A	SDHD	
NADK	FAM40A	BCL2L13	RND3	MRPS9	PMPCA	UBE3B	CAMK2A	
NOP2	PIK3C2B	ZBTB7A	SHANK3	DOCK1	SH3GLB1	ARHGAP1	COX7C	
TBPL1	TMEM64	EIF2A	LARP4B	LTBP4	CAMTA1	RPL7	QKI	
CTDSPL	PSMB3	ABI1	COPZ2	THOC2	NUDCD1	SREBF2	U2AF2	
HERC2	PTTG1	TNRC6C	FADS3	GOLGA2	BLOC1S1	ACTN1	MCAM	
CCDC12	Ankrd50	PYGL	PTPN14	CDC123	PSMA2	CGNL1	SPEG	
VPS41	SFT2D2	TMCO1	PIH1D1	GATAD1	CAND1	EPHB4	SND1	
CLK3	SMO	CMTM3	MECP2	COPE	TRAK1	TRIP12	EIF4B	
FAR1	PINK1	CACNA2D2	MAT2B	IK	CYB5B	UBE2H	UQCRFS1	
JAG1	RBM6	FGFR1OP2	CTPS1	LTBP3	RPL22L1	CYB5R3	H3F3A	
STAT6	NCOA5	SIN3A	MAGI3	SRSF7	CRK	CSNK1A1	LUC7L2	
HOMER2	GRIK5	CCDC80	INCENP	RCAN1	PXN	EMILIN2	SEPW1	
EPC1	TBCB	RGP1	ASXL1	HMGA1	TXN	USP28	RHOC	
ARL6IP6	KDM4B	PHF5A	CNOT7	EFNA5	TECR	AZIN1	NDUFS1	
WFS1	JMJD6	AGPAT3	MMRN2	NIPBL	UBE2C	DAD1	PAPOLA	
C19orf60	PDZD2	CUTA	GIGYF2	TRPC4AP	ATF4	CCNI	EWSR1	
CASD1	EDC4	BMP1	SBF1	KIAA0355	ECHS1	STMN1	IVNS1ABP	
LIFR	IRAK1	BCAR1	FAM129B	ZC3H13	NDN	GAS6	FERMT2	
GLIPR2	DCAF12	NCOA1	PPP2R3C	PACSIN2	RAF1	SMC3	NFE2L1	
SNRPF	KIAA0196	AEBP2	HBP1	NDST1	EID1	THRA	CD93	
DHRS7	PIK3CA	KREMEN1	COQ7	ECH1	TMED2	PDK1	BEX4	
HAUS8	TBC1D1	ZCCHC14	MRPS16	DDAH2	FZD2	RERE	LMNB1	
SEC61A2	PPAT	FAM100B	STRAP	UBE2N	FAM168B	PJA2	ATN1	
TRAPPC6A	ATIC	RSU1	C6orf162	ANKRD40	HIST1H2AG	CTSL1	FKBP8	
VPS52	MRPS34	SLC12A6	UACA	ARIH1	SUMO2	TMED10	ANKRD11	
PTPRM	ITFG3	SNX22	MYO9B	KIDINS220	MED25	IGF1R	CAPRIN1	
TOM1L1	SMURF2	PLK1	CABIN1	NRBP1	GLUL	ETFA	ANP32B	
IGFBP2	LSM2	STK11	IDE	C1orf131	NKTR	BANF1	FAM127C	
TARBP2	EXOC3	UBE2G1	DYSF	APRT	YPEL2	TXNDC5	PPP1CB	
SNRNP48	OSBPL3	COQ9	SYMPK	CNOT3	SRP14	AKAP13	NDUFB7	
FAM60A	ID2	STXBP1	ADPRH	MAGOH	ANAPC11	PRDX3	NCOR1	
ICAM2	TAF10	PDGFRB	C11orf9	MGEA5	ATP6V0A1	BRD3	SOD1	
C3orf70	MID1	LMO4	MTMR6	ATP6V1F	NAPA	GUCY1A3	DYNLL2	
GLRX3	MED24	CCDC88A	RABGAP1	SUCLG2	CSNK2B	TAB2	PYGM	
LIAS	EFR3A	HMGN3	CCNY	POLR1D	C1orf198	SLC25A11	DPYSL3	
HINT2	ELN	CKAP2	MRPL45	SNN	SMTN	KLC1	ROMO1	
DCTPP1	KBTBD2	GPBP1L1	MIA3	ALAD	HMGCS1	STK39	COL5A1	
MITF	TRABD	UBE3C	RBM38	AARS	SF3A1	DLK1	NCAM1	
NDE1	AURKAIP1	AIMP1	PHF23	AFG3L2	MRPL54	ZC3H7B	PEBP1	
TRAFD1	CLCA1	MED13L	NCBP2	VDAC3	MRPL15	PAFAH1B1	JPH2	
CRTC1	PDZD8	WDR48	SCYL1	NCS1	MMP14	TRIM63	TRDN	
SNTA1	BEND4	DCHS1	TH1L	PNRC2	HNRNPR	ETF1	NEDD8	
GPR125	ATXN1L	SLC4A1	STT3A	NOMO1	RSL1D1	MTUS2	SRSF3	
LIN7C	AP3S2	SETD1A	CHERP	TOMM40	PPP6R3	STUB1	HNRNPUL2	
ATP13A2	C14orf37	ZCCHC11	TMEM14C	XPO7	MBNL2	DPYSL2	SLC25A5	
NCAPH	PPRC1	TAF6	CENPE	EGLN1	NCKAP1	CCNA2	CREB3L2	
PPME1	ERP44	DNAJC10	ITSN1	RHOT2	RANGAP1	AMOTL1	RBBP7	
NUS1	HSPA14	DOK4	VCAM1	ATP6V0B	SPIN1	MYO1C	RPL35	
RAB12	MAP2K1	ADIPOR2	NUFIP2	RACGAP1	LMOD2	RPS21	CLIC4	
BAK1	HMBS	GTPBP1	CTDNEP1	GALNT1	EIF4EBP2	MTUS1	IDH3A	
THBS1	MDP1	NT5C2	DDX50	SIPA1L1	ATP6AP1	RRP1	PLXND1	
EIF2D	CLPTM1L	GTPBP2	LARP7	OPA1	USP13	MTCH1	POLR2A	
RFXANK	BAG1	MON2	CAPN2	UPF1	NBR1	PSMB2	PSMB1	
MTAP	ZCCHC6	TRAPPC3	KCNJ5	GLS	OTUB1	POSTN	ACIN1	
ATP6V0D1	TRMT112	SERPINB6	FILIP1	FAM54B	MRPS21	WBP5	DNAJC7	
TSR1	TIMM17A	OTUD5	ARID4B	ATP2B1	PPP1CC	DLD	MARCKSL1	
EIF4E	LPCAT3	COPS4	SLC12A4	SCAF11	ORMDL3	CDC37	ANXA2	
RAB35	TMOD3	CDC42BPA	VPS72	AKAP9	KLHL9	SSRP1	CAPZB	
COQ5	MSTO1	SGCE	STAT5B	SLC44A2	VEGFB	RPS27A	PTPRS	
CCDC9	ECT2	USP15	ZCCHC3	ZMYND8	MAP1LC3A	TSPAN3	CDH5	
ZNF217	ARMCX1	COL23A1	EIF2C1	EML2	CFDP1	S100A11	ST13	
TPD52	APBB1	UBP1	Gstm5	GIGYF1	SPCS1	UBE2B	COX7A1	
NBEA	OSBPL1A	VPS26B	TMEM66	SYNPO2	HNRNPH3	FBLN1	EEF1G	
MUL1	AUP1	IQGAP3	PITPNM2	ELK3	GNPAT	PSMB5	LAPTM4A	
PLP2	HM13	PPP1R1B	MARVELD1	SYNE2	NTN1	TM9SF3	YWHAZ	
STRBP	C6orf70	XBP1	PKDCC	UBE2Z	ABHD4	DCAF8	PRRC2C	
PHOSPHO1	CCDC85B	MAP3K7	CD151	LYPLA1	AP2S1	CHD3	CDK4	
PPP1R15B	SAP130	ARRDC3	PPP4C	TSHZ1	HIST1H2AI	GYS1	POPDC2	
PDE4A	PPP1R10	KIAA1967	HIRIP3	NARF	SHC1	SLC38A4	CNN2	
KIF2A	CAPZA1	APOA1BP	KIF23	TRIOBP	MCM2	ACAA2	RPL13	
CSNK2A2	ZNF22	B4GALT6	DAP3	TNNI2	RCN3	RBM20	KIF5B	

DOCK7	MAP3K3	RAB7A	PCGF3	RASIP1	USP2	DENND5A	IGFBP4	
PHKB	ZNF219	MRPS24	UBQLN4	CLINT1	TSPAN18	WDR26	HK1	
ADAMTS9	MLL3	RGL2	HYOU1	BNIP3L	RAB11FIP5	PHACTR2	PNKD	
SNRPN	ZNF146	MAF1	BHLHB9	ADRM1	CHMP2A	MLL5	PSMC3	
GABARAPL1	CDK13	CUL3	CAD	IP6K1	VPS25	PICALM	ANAPC5	
WAPAL	YIF1B	POLR1A	TGFBR2	TOMM20	CDK11B	CSNK2A1	CLTC	
DCAF11	BOLA2	TOLLIP	STAU1	FNDC3B	PRKAA2	MYH7B	FAU	
TIMM9	UPF2	AK3	CDC16	KARS	TOR1AIP2	MLF1	PFKL	
AMPD2	BRD1	PKN2	HMCN1	PCP4L1	AP3D1	MAPK1	RMI1	
C6orf120	NUP93	RER1	PHIP	MT1A	SSR2	RAB2A	MYH9	
WRB	PIGP	NFYA	PHRF1	TP53	PBXIP1	UBAP2	LENG8	
TBRG4	DUSP16	MGAT1	SNAP91	MAST2	NEDD4L	RNPS1	SNRPD2	
C18orf32	UBXN6	MPDZ	SEC16A	KAT6A	ZEB1	NME2	SERF2	
MCM3AP	DAAM2	VEZF1	CHMP7	RAB5C	C12orf57	NUMA1	AKT1	
UPF3B	PTDSS1	PIP5K1C	UBA3	PCYT1A	RNF207	Grinl1a	CLASP1	
PDRG1	NFYC	SF3B5	CDCA8	DNAJC8	NCOR2	TNKS2	GJA5	
ZMYM2	SMCR7L	LUC7L	CPSF7	ANGPTL2	PFDN1	KCNK3	CAP2	
MTIF2	CXXC1	FRG1	TUFM	CDCA3	TLN2	KDM6B	P4HB	
SMURF1	SRP19	KPNA2	PCOLCE	PLEKHO1	RPS19	IMMT	MRPL12	
KIF3A	PANK1	ARAF	AP1M1	TWSG1	B4GALT1	ARGLU1	XIRP1	
ZNF292	AUTS2	GIT2	SHMT2	SNX17	ETS2	NPEPPS	NAP1L4	
CPT1A	COMMD3	FAM190B	MAEA	FTO	PSIP1	DRAP1	YWHAG	
PPP2R2A	SMPD1	PTPRD	SPAG9	SNX12	CELF1	BCLAF1	PDIA3	
ZNF687	AAAS	GRB14	EVL	LTBP1	WASF2	REEP5	ARF1	
SNRPA1	FOXRED1	FXYD6	GTPBP4	PRKCSH	TRAK2	TMEM2	HNRNPF	
TSPAN7	CAMTA2	CORO1B	SRPR	WHSC1	AP2A2	NSD1	RPL31	
ZDHHC12	E2F8	NOL7	MGAT4B	ARPC1B	CDC23	HNRNPC	CALM3	
NR2C2	ITGB1BP2	PARD3	PSMA4	TACC2	RPS13	CAPN6	ARID1A	
ITSN2	SEC11A	PYGO2	DNAJB5	MAST4	LIMS1	GNG5	RPL38	
OXR1	S100A13	MED14	ZMIZ2	ATXN2	PBX2	COPS6	CCT5	
AMBRA1	FAM134B	CUEDC2	LMF2	ZNF827	LCLAT1	HDGF	RPL10A	
ZNRF1	LRRC10	LAPTM4B	ABR	DDX46	ANTXR2	KIAA0195	TSC22D1	
ZFPM1	TMEM176A	PHF2	TXNL1	SPARCL1	PSMA1	GATA6	NDUFB11	
TRIP10	SORD	KIF11	OS9	MCL1	UBTF	KTN1	DYNC1H1	
C17orf103	MFAP4	PLEKHJ1	KDM5B	SLC20A2	TNPO2	G3BP1	NDUFA6	
CPSF3	CYBA	RALA	FBXW8	EHMT2	PPP1R3B	NDUFS7	TOP2A	
SMARCAD1	GMPPA	BCL2L1	LOXL1	EHBP1	SRSF12	EPB41	PDHA1	
GPR182	RAB21	RPL7L1	RARA	CUX1	CCT3	ADNP	HUWE1	
RNF216	PSMD14	FOXK2	USP10	BCAP31	ACADVL	CRIP1	DLST	
TNNI3K	ITFG1	MAN1A2	MEGF8	INTS3	CKAP5	ITGA5	MIDN	
LAS1L	SSU72	ZDHHC8	MTCH2	CUL1	TMEM59	PSMD4	MMP15	
THAP4	PIP4K2B	FLOT1	PLOD1	MTSS1L	ZNF516	SEC61a1	CPT1B	
LTBR	C9orf89	INO80B	DBN1	PSMD12	ABLIM1	CAMK2D	CSRP1	
FBXL3	SNX2	RBM4	RSRC1	MRPS7	BCAS2	MAPKAPK2	OBSL1	
ARFIP2	NCLN	TFPI	PPIP5K2	SAE1	DYNC1LI2	GSN	NPM1	
KIAA1524	CYTH2	ARID1B	FAM195A	AFAP1	SDCBP	SKP1	ADCY6	
ORC3	CXXC5	PPP3R1	KIRREL	CHMP4B	PSME1	FKBP4	RRAD	
MRPL33	SETDB1	ATP1A2	XRN2	SUMO1	PDCL3	PRDX6	UBE2D3	
SFI1	GOLT1B	FOXO4	MLL	CSNK1E	SBNO1	GOLGA4	NDUFA2	
APBB2	TSG101	SERPINE2	EMC2	TOMM70A	HMOX2	PSMD1	BCAM	
STBD1	APEH	PHACTR4	ATXN7L3	ATP6V1E1	UBXN1	SPON1	FBN2	
CETN2	FXC1	DNAJC13	CENPA	ARF5	MTPN	NME1	NDUFB9	
SMARCA1	TRIM8	CSTF2T	GIT1	GRPEL1	C22orf28	COL6A2	RPS8	
FAM134C	TBL1X	CCDC56	FARP1	ZRANB2	SMARCE1	KHDRBS1	PCDHGC3	
CAST	DERL1	FTSJD2	SEPP1	LPGAT1	EIF2S3	UBR4	SUCLG1	
KCTD12	SDCCAG3	ZC3H15	EMG1	BVES	ANKRD52	CIC	ASB2	
TEX2	APOOL	FAM126A	WDFY3	BAG3	LAMTOR1	NUDT4	MAP4	
SKIV2L	CPOX	KLHL7	SEPN1	YLPM1	ZFP91	RBM5	PSMB6	
C6orf136	EGFL8	AMMECR1	C20orf194	MFGE8	PPP2R5C	POLR2F	RBFOX2	
BACH1	CDK2AP1	FIBP	METTL23	WSB1	TRRAP	NDUFB2	HOPX	
UBE2Q1	MEIS2	CHMP1B	UBQLN2	CWC15	KPNA6	PUF60	PTP4A2	
VPS33A	DDX24	CSPG4	IQGAP1	LPHN2	RAB11B	PNN	MYO18B	
RNF123	CYHR1	ZFAND6	EPAS1	TIA1	RPL18	ANK3	HMGB2	
DDX47	PRKAG1	DLG5	SEL1L	EIF4A3	EPB41L2	G3BP2	EIF4EBP1	
SERINC4	BRD8	FEM1B	MRPL28	EFNB1	RPL36	PPIG	CFL2	
AKAP11	PRKCDBP	ARFGAP1	DNM2	KIAA0141	PPP6R1	PTP4A3	RPL5	
FAM172A	WDR18	DECR1	EDNRA	GORASP2	DCAF7	RPL24	RPS26	
USP25	NOL6	PDCD4	ARHGEF1	MMP2	GOT1	PARP1	PABPC4	
RNF168	SCHIP1	REXO1	PFDN6	SRP68	RPS7	TIE1	MDK	
POLD1	MGST1	LBR	HDAC7	PPP1R9B	RAB5B	CUL7	EIF4A1	
MAP4K3	RAD9A	Srsf8	EPHA4	DYNLT1	UBFD1	ARHGEF7	TRIM55	
FAM133B	IKBKAP	WIPI1	PPIL1	TRAF7	MAPK1IP1L	MYLK3	AGRN	
KAT2A	FGL2	HOMER1	ELP3	AP1B1	PCM1	SETD5	RPL26	
DDA1	EIF6	SLC25A37	CRTAP	GALNT2	ARCN1	NR2F2	C14orf2	
C14orf43	MRPL4	MRPL18	ARPC1A	BAZ2B	FHOD3	OCIAD1	MAGED2	
KIF4A	SUDS3	GSK3A	TMEM132A	UHRF1	HIST1H2AH	SAFB	PYGB	
SGPL1	BRWD1	NCOA6	TRIO	LEPROT	MYBBP1A	RAB14	MACF1	
C1QTNF6	RASA1	PPP2R1B	TAOK1	PHF3	CYCS	RBM25	PRKACA	
CDC5L	MTOR	YEATS2	FEM1A	DNAJC15	CAP1	AGPAT1	FABP3	
MRPS18C	SP1	MESDC1	KIAA0226	ROCK2	ENAH	ETFB	CBX5	
KIAA1462	SUV39H1	MEGF6	TALDO1	CCDC6	IGF2BP3	SKI	CAV1	
SNRPB2	CEP68	MRPL13	MAPK6	KIF20A	C16orf5	RPS14	SLC2A1	

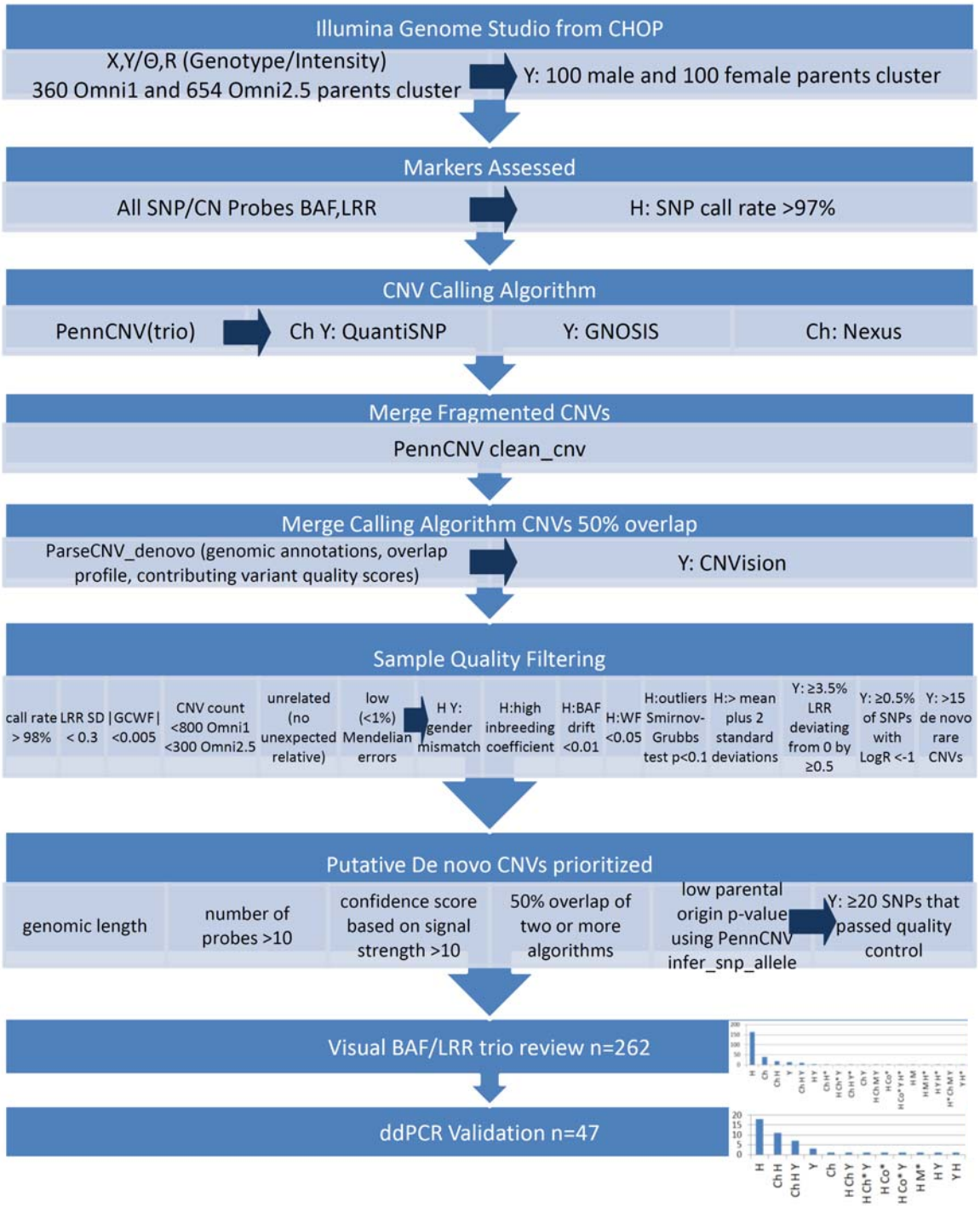
RSBN1L	RPTOR	GNL2	CASKIN2	C14orf119	CDK1	SF3B1	ACTN4	
ADRBK2	IFT46	RNF20	GFM1	RAI1	MTDH	TXNIP	CTSB	
FLT1	SERP1	CLASP2	EIF1B	CCDC124	LAMP2	ACAT1	CBX6	
IP6K2	POLR2B	SNX27	UNC13B	SLC39A1	LOXL2	MANF	YBX1	
GLCCI1	BABAM1	RNF5	TCAP	PITRM1	DCBLD2	ACTR3	HNRNPD	
MAN1B1	SEC22B	TMEM206	PHF21A	BOD1L1	MYL12B	BRD4	PABPN1	
CDC37L1	FRMD4A	PDS5B	SH3BGRL	PCCB	ZNF706	MMD	MRPL52	
CCDC142	RASSF3	PLCG1	C1orf43	TCF3	SCAF1	FRAS1	RPL27A	
RNF167	SLC7A1	SLC31A1	TCERG1	SLC6A6	GTF3C2	ATP5H	NGFRAP1	
ABCD3	ATF6B	STX7	TOM1L2	CISD1	TOP2B	ILF3	ATP5C1	
ACAA1	GBE1	EXOC4	DICER1	UBQLN1	SAFB2	NUDT3	ATP5I	
PPP1R13L	PPHLN1	CNPY2	PHF20L1	C11orf95	CDC42BPB	UCP2	ATPIF1	
CIAPIN1	SF3A3	MUT	WDR33	SUCLA2	SLN	PTBP1	PTGFRN	
SH3BP4	CACNB2	API5	SS18	CKS2	DHX15	NAV1	RPS15A	
TROVE2	TENC1	TRIM11	DYRK1A	HIBADH	CCNL2	RPL7A	COX5B	
LAGE3	PDZRN3	AIP	YME1L1	SEMA6D	HN1L	GRN	VCP	
MEN1	MTMR3	LIMD1	STRN4	SLAIN2	RPL17	DNMT1	NAP1L1	
MRPL16	SCARB1	SIDT2	CORO6	ACACB	HECTD1	HP1BP3	KPNB1	
RANBP10	CHD6	INSR	FAM114A2	ARNT	TMEM164	LARP1	CCT6A	
ABCB8	ANXA7	C6orf89	DUT	TACC3	ANK1	PARK7	UQCRC1	
POR	PIAS2	CASC3	NUP188	BAZ2A	WDR6	STIP1	ARHGDIA	
FBF1	H2AFJ	POLR2D	MYEF2	DDX23	DOT1L	ANAPC13	NDUFB10	
SIRT2	FZR1	ENSA	C11orf58	FRY	DMPK	DDX21	UQCRC2	
ECD	PREB	INO80E	MRPL34	PPP2R5D	NOSIP	TPM4	NDUFA1	
KPNA3	ZBTB47	MRPL13	GMPR	RBP1	EMP1	NID2	ZAK	
BRI3	SMARCB1	USP8	SCAND1	GPX4	BRD7	NCAPD2	RPL28	
CRBN	ACYP1	AAK1	CACYBP	TCF12	PPP1R3C	TSPAN9	COX7A2L	
FAM64A	GLOD4	MPZL1	DHCR24	HEY2	RHOQ	WLS	PSAP	
INTS7	ZFAND5	TMEM126A	KDM5A	ACOT13	RPL36A	CALD1	PLEC	
IGFBPL1	SMAD3	DNAJC14	DNAJB11	ERGIC3	PPP2CA	PSMB4	WNK1	
UBE2E3	CEBPZ	ALDH16A1	RANBP3	USP11	ZMIZ1	NDUFAB1	SUB1	
PDZD11	PIP5K1A	ELF4	PVRL2	PSME3	WBP2	BAZ1B	CST3	
YIPF3	PPP1R7	NCOA2	TMED7	RAB11A	TNPO3	CCT4	MSN	
OLA1	BRMS1L	COG6	PI4KA	PLEKHG2	MBD6	ZC3H11A	SHFM1	
DHFR	ACSL1	EGFL7	OSTC	WISP1	MLLT4	TIMP3	GNB1	
STAG1	KIF26B	CTTNBP2NL	BNIP3	TOMM6	HDAC2	FH	SRSF5	
ELMO2	MORC2	GOSR2	NCOA4	PIGT	AFAP1L1	PDIA6	MRFAP1	
BACE1	FKRP	EIF2S1	GOLM1	SAMD4A	EFTUD2	RPL14	SPTAN1	
OSBPL9	GADD45GIP1	DMWD	CLK1	BCL9L	CD24	JUP	OBSCN	
DUSP6	BUB1	RUNX1T1	TMEM30A	PLK2	CSE1L	NUCKS1	CSRP2	
STOX2	PAFAH1B3	AIFM1	SSSCA1	RPS6KA2	OGT	RPS10	SOX4	
C3orf58	MED13	FUBP3	VPS37B	RCAN2	ZFR	CITED1	ATP5L	
AMMECR1L	COMMD1	IER5	ABAT	TUBB6	SCN5A	AP2M1	PRRC2B	
CUL5	ZNF362	GPR126	SMEK1	BIRC6	POLR2J	ZNF207	GBAS	
MSL3	PSMD9	SLMO2	BEX2	UBXN4	KDM2A	HSPE1	ATP1B1	
RNASEH2C	TPCN1	ARFGAP2	KDM3B	DUSP3	ACTR1A	SFRP1	COL18A1	
FAM104A	MZT1	SPINT2	R3HDM1	RBM3	SEC24C	SDPR	C7orf55	
ANAPC16	ZFHX3	KIAA0368	PRICKLE1	SH3GLB2	LARS	KLHDC3	ATP5G1	
FAM98A	RBM42	PSMD6	USE1	ISCA1	HIST1H2AL	ATRX	RPS11	
FAM108A1	DNAJB1	WDR45L	SLIT3	RGL1	NDUFA9	MYOZ2	GTF2I	
HSDL2	OSTF1	GGA1	DPY30	GGA2	HIST1H3D	KIAA0100	PTN	
LEPROTL1	CNOT8	CDK7	ZNF512B	SPEN	ENY2	PSMC6	DAG1	
PTEN	GPR116	COPS7A	ERP29	MDC1	GPR124	TOP1	MLEC	
EXOC5	CDH3	UGGT1	MBTD1	GOLGB1	KANK2	SPTB	RPSA	
Ahsa2	UQCRB	TOPORS	RNF187	XPO6	EIF3G	LTA4H	NDUFS6	
CLK4	MCMBP	GNS	IGFBP3	MEIS1	KLHDC10	RPS3A	COPA	
UTP3	TM9SF1	C17orf81	NT5C3L	C1D	SELT	PSMD2	HINT1	
SUV420H1	KRCC1	ATF7IP	GART	TPP1	PRPF19	EHD4	MYO18A	
MKL2	DDX54	UBE3A	ARFGEF1	CENPF	RPA1	EIF2S2	RPL18A	
RAD54L2	LRPAP1	FAM48A	UCHL1	NRD1	USP22	MSI2	LPL	
PRR14L	COPS8	TRAPPC1	HDAC3	HARS	TRAP1	HSPD1	VDAC2	
ETV1	ENTPD5	FADS1	RFC1	PLXNA4A	SLC22A17	TPR	SERINC3	
NUP214	SEC24B	IFNAR1	KLHL23	BCL9	EIF3K	TMPO	PGAM1	
PPP3CB	EML1	FTSJ3	GBF1	LAP3	PDIA4	PRPF38B	DYNLL1	
MKRN1	VAPB	PIN1	SERF1A	COPS2	TBC1D16	LSM14B	SDHC	
EIF1AX	PLEKHA6	C6orf203	PSMD7	FECH	PDLIM3	NPC2	KHSRP	
NIPSNAP3B	PTPN1	RPRD2	GRINA	MTA2	ZMAT2	EEF1A2	PRDX2	
BBX	SLC7A11	RYK	TSN	DENR	MSRB3	TTYH3	CYC1	
CASP3	ITCH	POGZ	TP53BP1	RGS3	SSBP3	FSCN1	NONO	
WDR11	PCK2	SPRY2	SLC6A8	BUB1B	DEK	ARF4	ATP5O	
ANO6	MRPL14	LRP10	FLNB	RPL21	Msl1	ITM2C	WDR1	
NFE2L2	F11R	ABCB10	NET1	MCM6	DNAJA1	RPS25	CHCHD2	
LSM3	MOV10	PGD	TXNRD1	MYBPHL	LRRC59	NDUFC2	APP	
PIAS1	ANKRD13A	KCNE1	TTC28	SPOP	GSK3B	IDH3B	MYL6	
SSNA1	VBP1	USF2	SYF2	TCF7L2	DGKZ	CCT8	DHX9	
C4orf34	CCDC97	WDTC1	KLF6	HK2	FUBP1	ARPC4	GOT2	
LATS2	BRD9	DIP2B	RRN3	DLGAP4	TSC22D4	MYEOV2	TLN1	
TUSC3	TSPAN6	PKIG	MKLN1	CCND1	IGSF3	BZW2	PTMS	
SMCHD1	FBXO11	ST3GAL1	ZADH2	CALCOCO1	R3HDM2	APOE	C11orf10	
C1orf115	BCOR	FAM21C	CCNT2	LSMD1	EDNRB	SMARCC2	CNN1	
MRPS17	PRRC1	BAX	CTGF	PTK7	PLEKHA7	PDE4DIP	SDHB	
SCAMP2	ZNF187	SEPHS1	NUDT8	NSFL1C	PDK2	CNN3	RPL15	

RFC2	CHAF1A	RBM26	HADHB	MEA1	NPR3	MBTPS1	CTSD	
CLEC16A	ZNF148	NFU1	DCTN3	TGFB2	RRP8	RAC1	EHBP1L1	
GSTM2	PDPK1	IPO13	ERBB2IP	FAM198B	RBM8A	GTF3C1	GNAI2	
RAP1B	POLR2G	ALDH2	CBX1	POFUT2	COPG2	HSPA4	TMSB10	
TMEM88	ALDH6A1	IPO11	KIAA0182	ANAPC1	CACNA1H	ENG	BEX1	
EML4	LEPRE1	TEAD3	HTRA2	HIPK1	PPP2R5A	C19orf70	HDLBP	
C1orf85	MRPS36	CHURC1	S1PR1	USP4	NEBL	TEAD1	UBC	
TTC4	DNPEP	YKT6	ADH5	ATP1B2	ACADM	SRSF11	MIF	
BRI3BP	KCNJ11	ISOC2	CAB39	RNF44	SEC23A	EXTL3	TPM2	
NDUFS5	AFF1	POLR2I	BECN1	ZC3H4	UBE2R2	NPPB	RDX	
GIPR	GOLPH3	SNW1	EIF2C3	ISYNA1	UBL5	WDR82	RPL19	
PLD3	LRP6	PHLDB1	FBL	SAMD4B	EHD2	YWHAH	EIF5A	
PRCP	PDK3	NANS	TXLNA	AIF1L	TUBB2A	TBX20	APLP2	
GHR	BUB3	PKD1	NAA38	OGFR	BCKDK	PKP4	ACTA1	
RP9	FAM96A	STARD10	EIF3I	IRS1	FAM96B	CSDA	MARCKS	
MAGT1	GRIP2	NEURL4	VGLL4	PITPNB	MLLT6	RAD21	MYADM	
RNASEH2B	UBAP1	PJA1	EIF3E	PTPN23	ACLY	TXN2	RYR2	
EMC3	RAP1GDS1	VLDLR	PPIC	TMEM201	TMEM38A	HIF3A	UBA1	
FAM3C	ZNF608	UBA2	JARID2	EMD	PAICS	COL5A2	HNRNPA0	
LMTK2	IREB2	SRR	EMILIN1	UPK3B	CIRBP	TBX5	NDUFA5	
C20orf43	UFC1	HEATR7A	SLTM	OAZ2	C19orf53	PDLIM7	TGM2	
EXT2	RHOT1	FOXK1	STRN3	RTF1	ATXN10	ARL5A	GPC3	
SMARCA5	FRS2	XRCC1	MRPL42	DSG2	NCSTN	SRF	SH3BGR	
NT5M	DACT3	PER1	PLAGL2	GM2A	ARPC3	SURF4	RPL6	
SRCAP	UROD	PALM	ZFAND3	POLR2E	EIF3H	SRSF6	RPL10	
PLAA	CLIP1	PPP1R9A	BLMH	MFF	MCM4	CASQ2	HSPA9	
BAMBI	TAGLN	HIF1AN	ATP6V1D	ZNF664	NRAP	AHNAK	TNS1	
PACS2	LTV1	SCMH1	TNS3	NXF1	RCSD1	EIF3C	CDH2	
POLR2C	LYRM2	DUSP11	CENPB	CKB	LMAN2	CTDSP1	SLC38A2	
UBL3	GPR153	MAP3K15	RAB5A	VASH1	LIG1	EFNB3	PPIA	
NUB1	EDF1	MAP2K4	HDAC6	AEBP1	TSC2	EIF4A2	SMYD1	
SMAD6	TNKS1BP1	NAA40	RUVBL2	PCNP	GPX4	PCDHGA8	HNRNPL	
ZNF532	SRSF4	PGRMC1	ASH1L	ARL6IP1	TMED9	HCFC1	MB	
SMAP2	SLIT2	IARS	PPM1G	TSPYL1	TCEB2	GAA	CRYAB	
IMPDH2	PRPF31	PSME4	RNF38	BICD2	LRP1	DDOST	CALU	
C1orf123	COBRA1	MED1	DVL3	SCAMP3	PFDN2	PCDHGA9	CHCHD10	
PI4KB	PRCC	SCP2	PRPF39	PDLIM1	DYNLRB1	TCEB1	HNRNPH1	
RPRD1A	DRG1	ASB8	TIAL1	RAB1A	RAB10	PCDHGA1	ATP5J2	
AIMP2	TMEM9	SMU1	PCBP4	NAA50	OST4	PCDHGB1	GJA1	
SLC3A2	KIAA0232	TMEM161A	CORIN	APEX1	CCNG1	CDK16	ATP5F1	
SPON2	MRPS23	PHYH	SRPK1	SH3KBP1	FYTTD1	GARS	CTNND1	
NR2F1	MRPS14	LZTR1	TMED4	LSM14A	DLC1	PCDHGA6	CCT2	
DCAF5	AMZ2	FBXW5	AMOTL2	CHD8	CELF2	APOBEC2	LAMB1	
PSMA5	FREM2	TUBG1	GRSF1	CAMK1	RALB	LDB1	LRRC58	
CCDC127	GMFB	COTL1	WDR13	VASH2	DPP3	PCDHGA4	SDHA	
TM4SF1	FUNDC2	M6PR	JOSD2	SENP6	KIAA0913	SLC25A39	HADHA	
DDIT4	BAHCC1	EMCN	ESD	MCM5	TNRC18	TAOK2	PRKAR1A	
C19orf10	CLASRP	PLEKHA2	NOP56	ABCC5	CTSA	SNRPB	COL1A2	
MCC	SNAP23	PRKAR2A	C9orf123	USP5	CETN3	PCDHGB7	CTNNA1	
C20orf152	BCKDHA	SMARCA2	GGNBP2	DAB2IP	SLC38A10	RPL23	HNRNPA3	
ATP6V1A	SENP3	STAT3	MLL4	MRPL10	OXA1L	UGP2	FLNA	
NOLC1	FAF2	CLYBL	NDUFB3	UBAC2	IFITM2	GPC1	LMO7	
TMEM127	DMD	ACER2	SYNE1	CLTA	ZBED3	PPP2R3A	NDUFB8	
ITPR1	POLG	SMAD5	YPEL5	ERF	CKS1B	ERGIC1	MORF4L1	
RABGAP1L	STK38	SART1	POLDIP3	TIMM44	PCNA	CD164	SRSF1	
ENPEP	AKAP8L	KLHL31	SH2B1	C6orf62	CCT7	HIF1A	RPS15	
SLC5A3	NHSL2	RNF145	RHOB	ZNF395	C11orf31	PCDHGB2	RPS2	
LIMD2	STX5	NPLOC4	MYCBP2	SDF4	USP19	PCDHGC5	TNNT1	
NENF	STOM	KCNG2	NUDT21	SIPA1L2	PTPRF	PSMA7	MYH10	
PSPC1	PPM1F	CERS4	PCNX	HIST1H2AC	TIMM13	PCDHGA5	OGDH	
ABCF2	WDR61	BTBD1	POLD2	DKK3	TFG	PGM1	MAGED1	
CTNNBL1	TET2	ARL2BP	FXR2	PRDX5	PARVA	PCDHGC4	SRL	
KPNA1	KLHL4	RPS27	JUN	CDS2	NMT1	NKX2-5	C15orf63	
ACBD3	HABP4	TMEM134	RABAC1	CDH11	PSMD3	PCDHGB5	ANXA6	
AGBL5	PRUNE	ALDH18A1	KDR	NEK9	C19orf43	RPL3	MDH1	
PTPN13	KLHL13	ZC3H14	VAMP8	TMX4	MGRN1	KIAA0664	HNRNPM	
PHF10	KIAA1191	RRAS2	KANK1	STARD7	FBN1	NDUFS4	CALM1	
N4BP1	ACADSB	P4HA1	FAM98B	TPM3	RBPMS2	PCDHGA11	ZFP36L1	
PIK3R1	KALRN	PHC2	TANC1	SAMM50	DAZAP1	PCDHGA2	ENO3	
TPRG1L	FOXM1	URGCP	SNX5	ZWINT	ABL1	LDHB	SRSF2	
MRPL41	EXT1	CDON	NUP98	NUDCD3	LSM4	VASP	ACO2	
FBLIM1	ASNA1	AKIRIN2	GNA11	CTTN	CCAR1	MAP7D1	HNRNPAB	
ZBTB4	RBM10	SP3	VARS	CLK2	DNM1L	AK1	TMOD1	
ZNF629	ANAPC7	ADSSL1	PLXNA1	STX16	PA2G4	RAD23B	EIF4H	
HMGB3	DAB2	LMNA	PKIA	RRM2	KIF1B	PCDHGA7	DSP	
FAM171A2	UNK	METAP1	GLT25D1	TAF9	USP47	GHITM	OAZ1	
FAM49A	MYOCD	SCRIB	SYN3	SLC2A4	TMEM123	HIPK3	HSPA5	
DTNBP1	TMEM109	OSBPL8	CCDC47	PRKAG2	AKAP8	PCDHGB4	SNRNP70	
FNBP4	TMEM65	RBM12	ETS1	GOLIM4	GON4L	RPL11	HSP90B1	
EGLN2	NFKB1	PSMC1	FAM178A	RAPGEF1	SNX3	ZYX	NCL	
LANCL1	TSC22D2	BNIP2	EHD1	ABCC9	TOMM22	COPG1	COL1A1	
ADAM9	GNAI3	LGR4	SEC23IP	AMFR	DST	PBRM1	CSDE1	

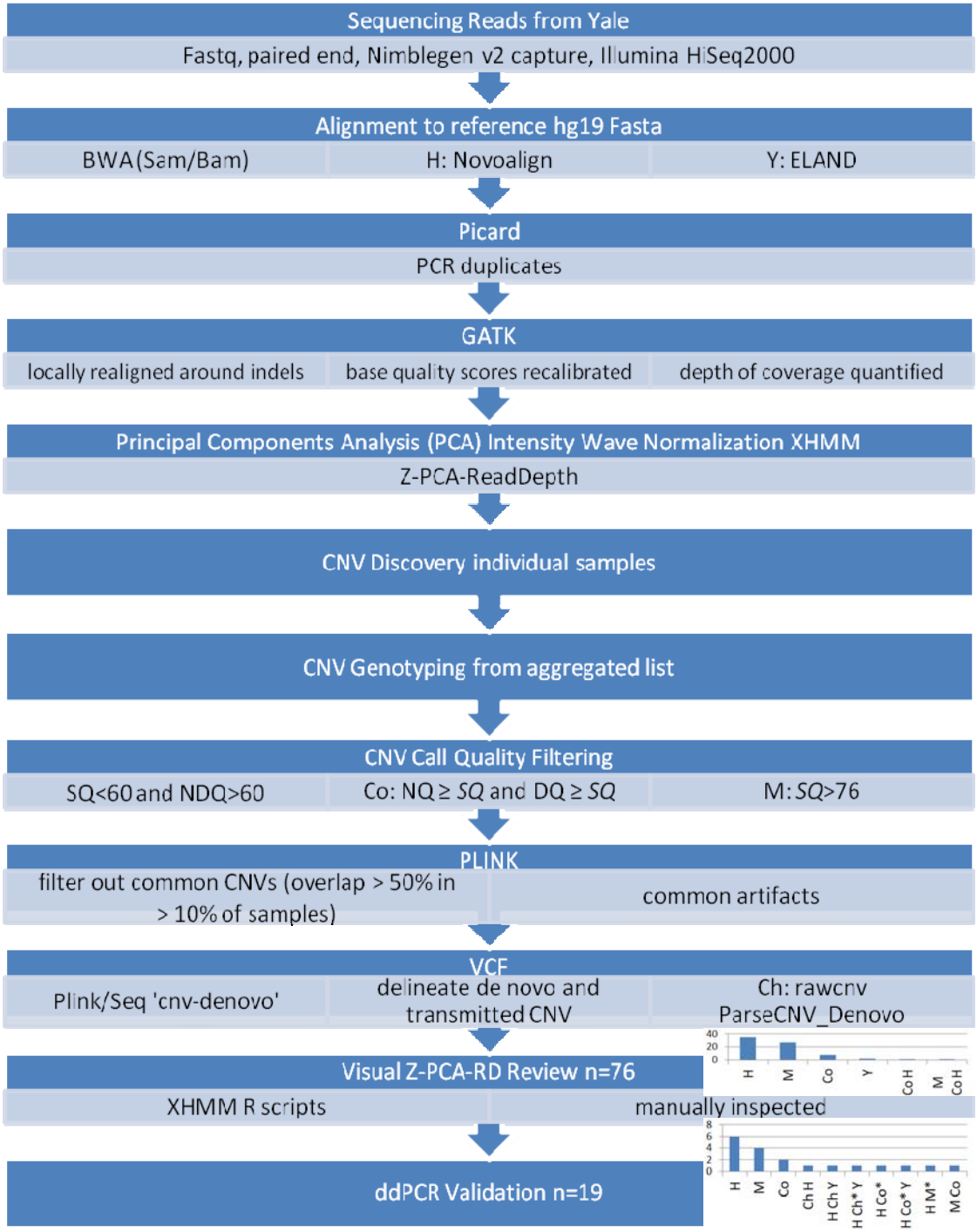
GRHPR	KIAA0907	ADK	PDXDC1	GPATCH8	RBMS2	CBX3	RPLP2	
SORBS3	ERGIC2	TANC2	MAPK3	SMC2	STAB1	VEGFA	NDUFA13	
CUL2	MLLT1	STK24	WAC	PTPLA	XPNPEP1	PCDHGB6	HMGN2	
SNX6	ATAD1	PIGQ	C16orf88	POMP	RPS18	YWHAB	CCND3	
MAP4K5	SRGAP2	MAP3K11	MAU2	MGP	BZW1	RTN4	PLAGL1	
SEMA4D	YEATS4	THOC7	TAX1BP3	PAF1	ADAM19	MPRIP	AES	
SOX11	PAIP1	TJP2	ABI2	HERC3	EIF4G3	PHPT1	DDX5	
E2F4	RGS12	TCOF1	UBR2	DPF2	ALAS2	PCDHGA3	BSG	
TCF20	AKAP6	FRYL	SEC13	AP2B1	TMEM176B	MAZ	COX6A1	
RASL11B	NDEL1	FBXO3	TM9SF2	PARVB	MEX3A	MORF4L2	SYNPO2L	
ATP8B2	CA14	WWTR1	IARS2	ARHGEF17	CACNA1C	NYNRIN	CTNNB1	

Supplementary Figures

Online Figure I. Schematic flow chart of SNP array CNV analysis performed independently by 3 groups (Ch: CHOP, Y: Yale, and H: Harvard). Modifications of the main method of analysis common to all groups indicated on the right of the flow chart by center. The distribution of CNVs identified by each method is shown in the histograms at the final two steps of analysis.

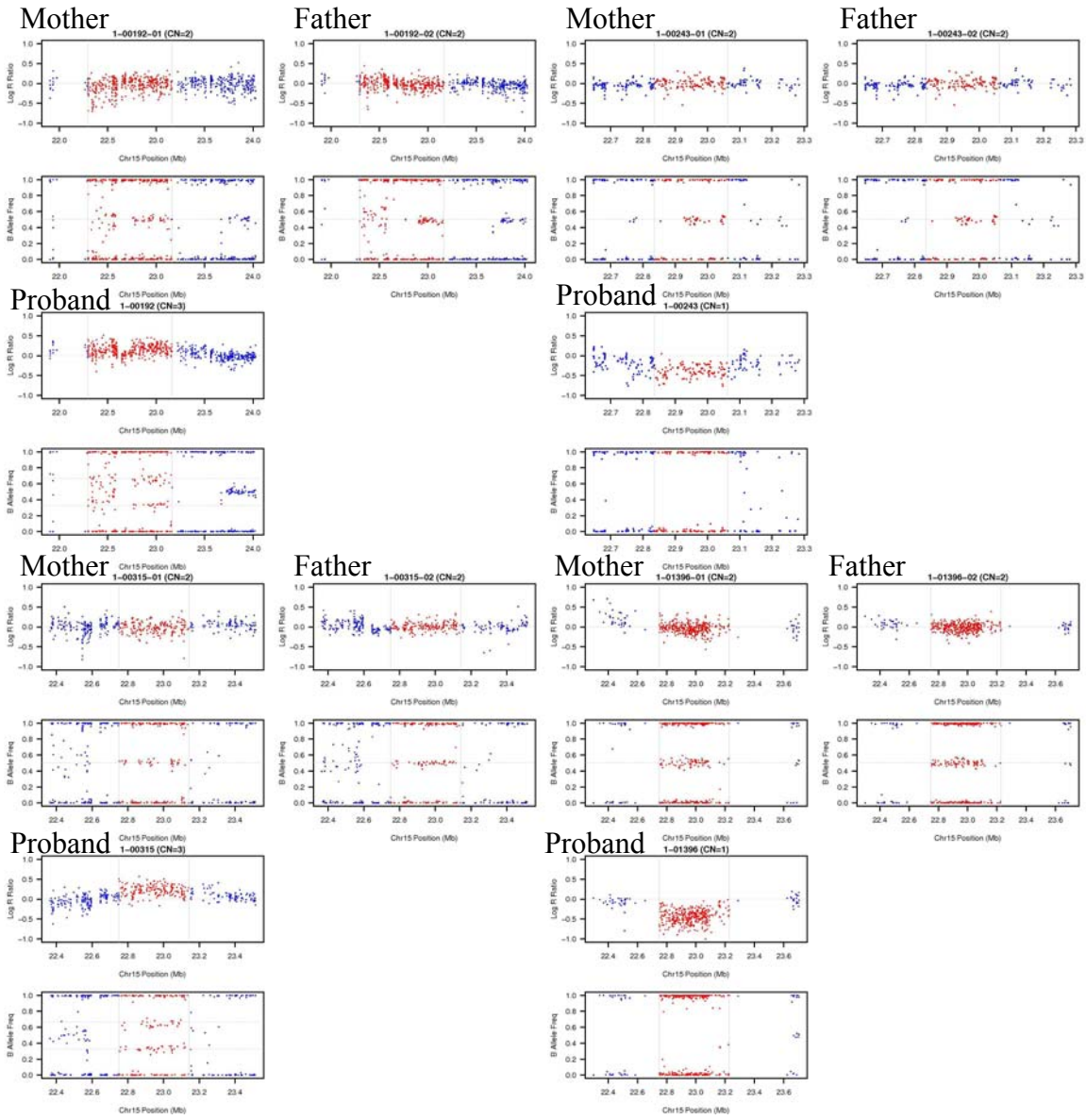


Online Figure II. Schematic flow chart of WES analysis performed independently by 5 groups (H: Harvard, Ch: CHOP, Co: Columbia, Y: Yale, M: Mount Sinai). Modifications of the main method of analysis common to all groups indicated on the right of the flow chart by group. The distribution of CNVs identified by each method is shown in the histograms at the final two steps of analysis.



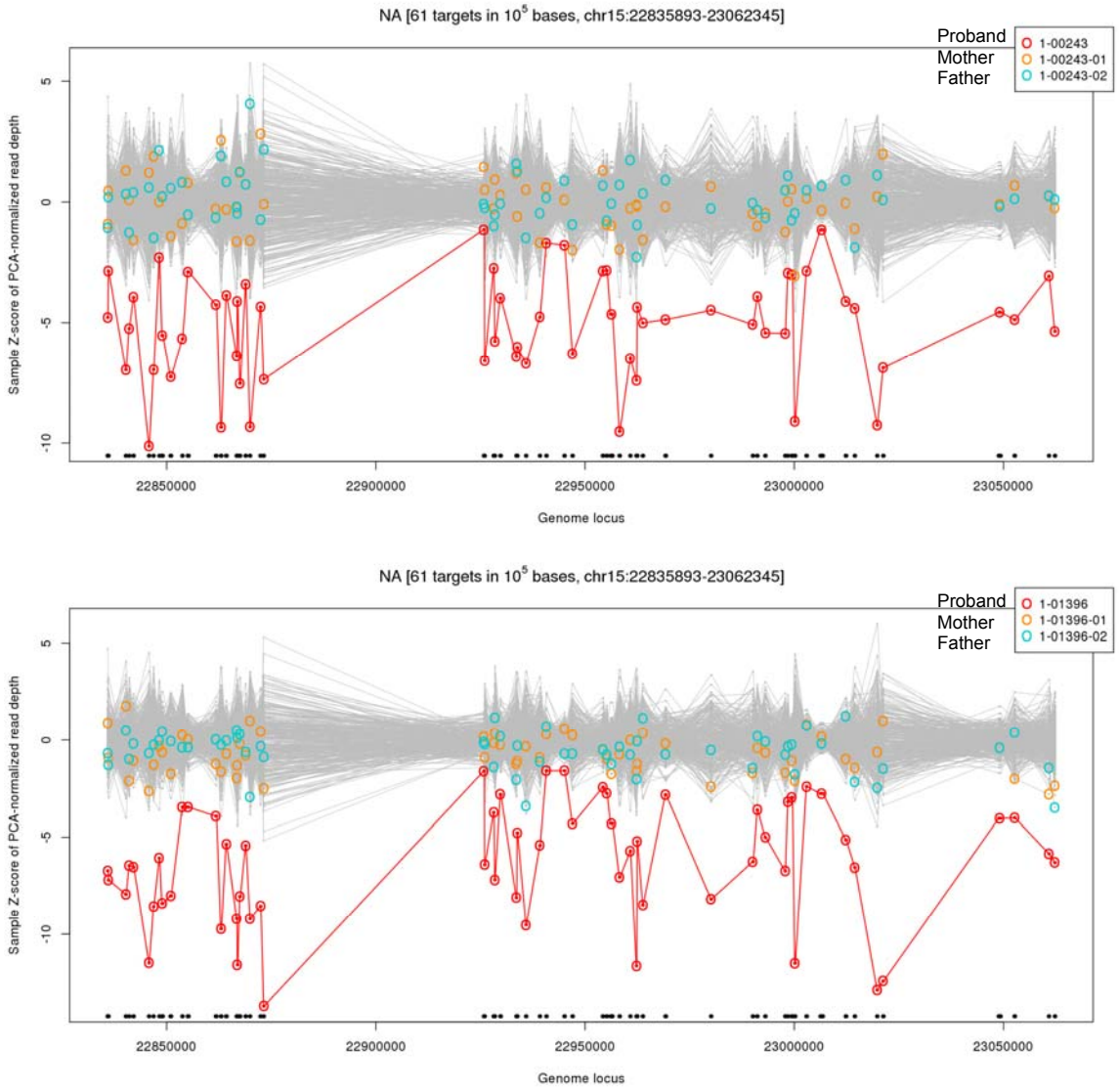
SQ: (Q_SOME) Phred-scaled quality of some CNV event in the interval, NDQ: (Q_NON_DIPLOID) Phred-scaled quality of not being diploid, i.e., DEL or DUP event in the interval, NQ: quality for deletion absence, DQ: quality for diploid event, BWA: Burrows-Wheeler Aligner.

Online Figure III. SNP Array *de novo* Plots for 15q11.2

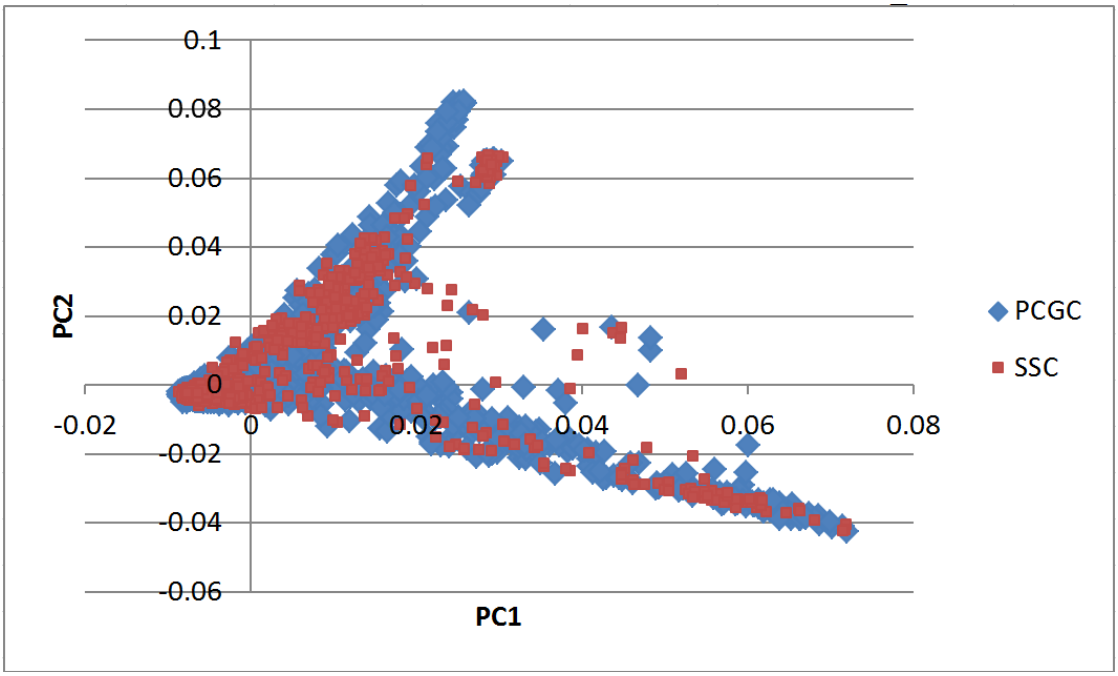


SNP array PennCNV plot for diploid mother, diploid father, and deleted (right) / duplicated (left) child with CNV region shown in red with flanking diploid shown in blue. Importantly, a decrease in LRR (0,1) and BAF are observed only in the deleted proband for this region. An increase in LRR (0,0.33,0.66,1) and BAF are observed only in the duplicated proband for this region. A mode of 0 LRR and (0,0.5,1) is observed in parents for the CNV region observed in the child, showing diploid parent.

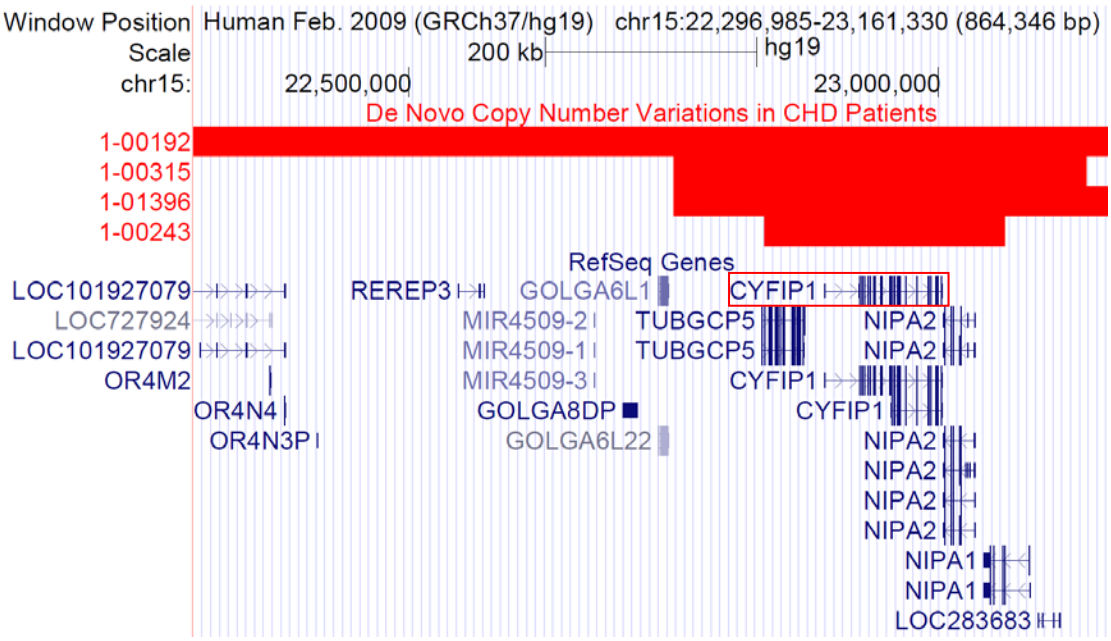
Online Figure IV. Exome Sequencing *de novo* plots for 15q11.2. XHMM Plots for corresponding regions in *de novo* samples found in exome showing the deleted proband (red circles connected by red lines) and the normal parents (orange and turquoise circles connected by gray lines). Additional gray lines represent a population view of normalized intensity (ZPCARD) values to show a reproducible diploid mode. The two *de novo* duplicated 15q11.2 samples shown in Supplemental Figure III were not assayed by exome sequencing.



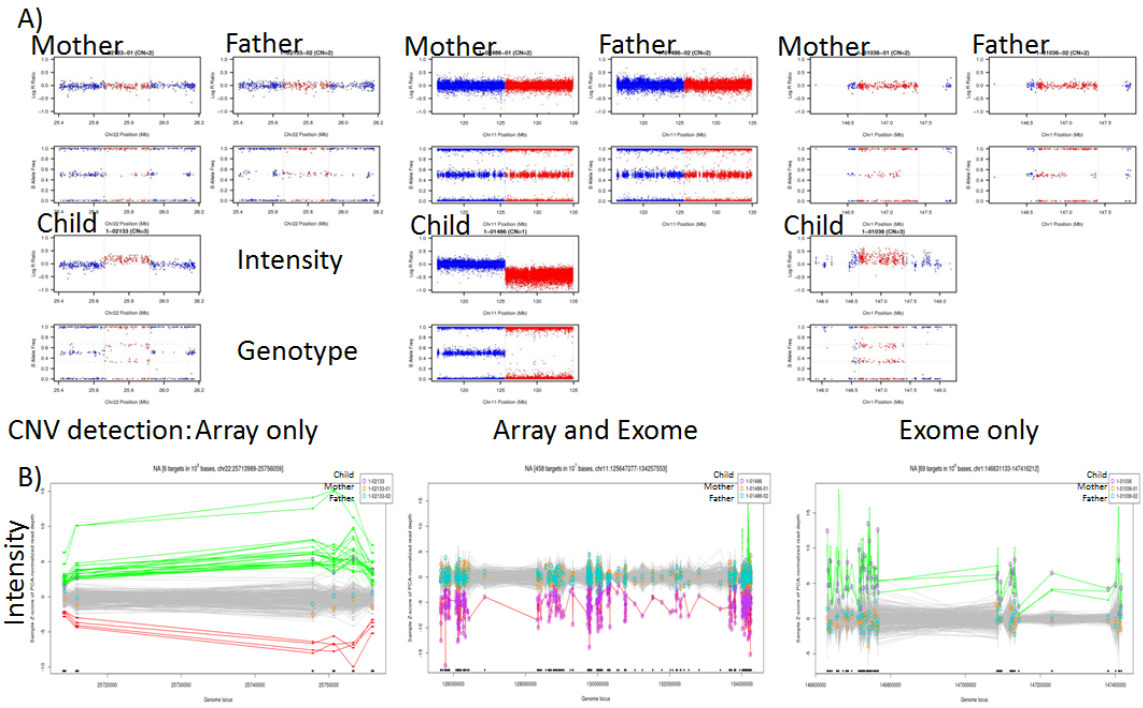
Online Figure V. Composition of PCGC and SSC controls principal components analysis showing matching between cases and controls.



Online Figure VI. Genomic Boundaries of 4 recurrent de novo CNVs



Online Figure VII. Array and WES Intensity and Genotype Comparison for Same Samples in Same Region with False Negative CNV Calls.
A) SNP Array PennCNV Plot for diploid mother, diploid father, and deleted child with CNV region in red with flanking diploid in blue. B) XHMM Plots for corresponding regions in de novo samples found in exome showing the deleted/duplicated proband (purple circles with green/red lines) and the normal parents (orange and turquoise circles with gray lines). We used exome sequencing with arrays on the same samples for integrative discovery.



References

1. Pediatric Cardiac Genomics C, Gelb B, Brueckner M, Chung W, Goldmuntz E, Kaltman J, Kaski JP, Kim R, Kline J, Mercer-Rosa L, Porter G, Roberts A, Rosenberg E, Seiden H, Seidman C, Sleeper L, Tennstedt S, Kaltman J, Schramm C, Burns K, Pearson G, Rosenberg E. The congenital heart disease genetic network study: Rationale, design, and early results. *Circ Res*. 2013;112:698-706
2. Fischbach GD, Lord C. The simons simplex collection: A resource for identification of autism genetic risk factors. *Neuron*. 2010;68:192-195
3. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ, Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW, Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Gunel M, Lifton RP, Mane SM, Martin DM, Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin CL, Brooks AI, Sutcliffe JS, Cook EH, Jr., Geschwind D, Roeder K, Devlin B, State MW. Multiple recurrent de novo cnvs, including duplications of the 7q11.23 williams syndrome region, are strongly associated with autism. *Neuron*. 2011;70:863-885
4. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Gunel M, Roeder K, Geschwind DH, Devlin B, State MW. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012;485:237-241
5. Glessner JT, Li J, Hakonarson H. Parsecnv integrative copy number variation association software with quality tracking. *Nucleic Acids Res*. 2013;41:e64
6. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. Penncnv: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome snp genotyping data. *Genome Res*. 2007;17:1665-1674
7. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. Quantisnp: An objective bayes hidden-markov model to detect and accurately map copy number variation using snp genotyping data. *Nucleic Acids Res*. 2007;35:2013-2025
8. van Karnebeek CD, Hennekam RC. Associations between chromosomal anomalies and congenital heart defects: A database search. *Am J Med Genet*. 1999;84:158-166
9. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK, Carriero NJ, Cheung YH, Deanfield J, DePalma S, Fakhro KA, Glessner J, Hakonarson H, Italia MJ, Kaltman JR, Kaski J, Kim R, Kline JK, Lee T, Leipzig J, Lopez A, Mane SM, Mitchell LE,

Newburger JW, Parfenov M, Pe'er I, Porter G, Roberts AE, Sachidanandam R, Sanders SJ, Seiden HS, State MW, Subramanian S, Tikhonova IR, Wang W, Warburton D, White PS, Williams IA, Zhao H, Seidman JG, Brueckner M, Chung WK, Gelb BD, Goldmuntz E, Seidman CE, Lifton RP. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013;498:220-223

10. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26:589-595
11. Cox AJ. Eland: Efficient large-scale alignment of nucleotide databases. *Illumina, San Diego*. 2007
12. Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, Handsaker RE, McCarroll SA, O'Donovan MC, Owen MJ, Kirov G, Sullivan PF, Hultman CM, Sklar P, Purcell SM. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet*. 2012;91:597-607
13. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491-498
14. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80-92
15. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, Emslie KR. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem*. 2012;84:1003-1011
16. Huang da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, Lempicki RA. David bioinformatics resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*. 2007;35:W169-175
17. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, International Inflammatory Bowel Disease Genetics C, Cotsapas C, Daly MJ. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet*. 2011;7:e1001273
18. Wang J, Duncan D, Shi Z, Zhang B. Web-based gene set analysis toolkit (webgestalt): Update 2013. *Nucleic Acids Res*. 2013;41:W77-83

Circulation Research
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Increased Frequency of De Novo Copy Number Variants in Congenital Heart Disease by Integrative Analysis of Single Nucleotide Polymorphism Array and Exome Sequence Data

Joseph T. Glessner, Alexander G. Bick, Kaoru Ito, Jason G. Homsy, Laura Rodriguez-Murillo, Menachem Fromer, Erica Mazaika, Badri Vardarajan, Michael Italia, Jeremy Leipzig, Steven R. DePalma, Ryan Golhar, Stephan J. Sanders, Boris Yamrom, Michael Ronemus, Ivan Iossifov, A. Jeremy Willsey, Matthew W. State, Jonathan R. Kaltman, Peter S. White, Yufeng Shen, Dorothy Warburton, Martina Brueckner, Christine Seidman, Elizabeth Goldmuntz, Bruce D. Gelb, Richard Lifton, Jonathan Seidman, Hakon Hakonarson and Wendy K. Chung

Circ Res. 2014;115:884-896; originally published online September 9, 2014;
doi: 10.1161/CIRCRESAHA.115.304458
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7330. Online ISSN: 1524-4571