HHS Public Access

Author manuscript

Birth Defects Res. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Birth Defects Res. 2017 March 01; 109(4): 271–295. doi:10.1002/bdra.23609.

Rare Copy Number Variants in Patients with Congenital Conotruncal Heart Defects

Hongbo M. Xie[*],
The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

Petra Werner[*],
Division of Cardiology, Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104 USA

Dwight Stambolian,
Department of Ophthalmology and Human Genetics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Joan E. Bailey-Wilson,
Statistical Genetics Section, National Human Genome Research Institute, National Institutes of Health, 333 Cassell Drive, Baltimore, MD 21224

Hakon Hakonarson,
The Center for Applied Genomics, Department of Pediatrics, The Children's Hospital of Philadelphia, Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Peter S. White,
Division of Biomedical Informatics, Cincinnati Children's Hospital, Department of Biomedical Informatics, University of Cincinnati, Cincinnati, Ohio 45229

Deanne M. Taylor, and
The Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

Elizabeth Goldmuntz
Division of Cardiology, Children's Hospital of Philadelphia, Department of Pediatrics, University of Pennsylvania, Perelman School of Medicine, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104 USA

Abstract

Background—Previous studies using different cardiac phenotypes, technologies and designs suggest a burden of large, rare or de novo copy number variants (CNVs) in subjects with congenital heart defects (CHD). We sought to identify disease-related CNVs, candidate genes and

Corresponding Author: Elizabeth Goldmuntz, MD, Abramson Research Center, Rm 702A, 3615 Center Civic Boulevard, Philadelphia, PA USA 19104-4318, 215-590-5820 (phone), 215-590-5454 (fax), goldmuntz@email.chop.edu.
[*]Co-first authors

functional pathways in a large number of cases with conotruncal and related defects that carried no known genetic syndrome.

Methods—Cases and control samples were divided into two cohorts and genotyped in order to assess each subject's CNV content. Analyses were performed to ascertain differences in overall CNV prevalence and to identify enrichment of specific genes and functional pathways in conotruncal cases relative to healthy controls.

Results—Only findings present in both cohorts are presented. From 973 total conotruncal cases, a burden of rare CNVs was detected in both cohorts. Candidate genes from rare CNVs found in both cohorts were identified based on their association with cardiac development or disease, and/or their reported disruption in published studies. Functional and pathway analyses revealed significant enrichment of terms involved in either heart or early embryonic development.

Conclusions—Our study tested one of the largest cohorts specifically with cardiac conotruncal and related defects. These results confirm and extend previous findings that CNVs contribute to disease risk for CHDs in general and conotruncal defects in particular. As disease heterogeneity renders identification of single recurrent genes or loci difficult, functional pathway and gene regulation network analyses appear to be more informative.

Keywords

Congenital heart defects; conotruncal defects; copy number variants; CNVs; functional analysis; pathway analysis

Introduction

Congenital heart defects (CHDs), which comprise the most common, severe birth defect, occur in 4–9 per 1,000 liveborn and are thought to be caused by both genetic and environmental factors (Pierpont et al., 2007). Conventional karyotyping detects chromosomal anomalies in approximately 13% of all CHD cases, most of which fall into aneuploidy syndromes (e.g. trisomy 18 or 21) (reviewed in Hartman et al., 2011). Array-based technologies have revealed submicroscopic chromosomal deletions or duplications (copy number variants (CNVs)) in an additional 3–20% of CHD cases, with a higher frequency observed in those with syndromic or additional non-cardiac features (reviewed in Andersen et al., 2014; Lalani and Belmont, 2014). Despite differences in study cohort phenotypes and genomic surveillance approach, most studies report a significant burden of large, rare, and/or de novo CNVs in CHD cases (Glessner et al., 2014; Greenway et al., 2009; Lalani et al., 2013; Silversides et al., 2012; Soemedi et al., 2012b; Tomita-Mitchell et al., 2012). Some of these CNVs encompass genes usually disrupted by single nucleotide mutations for which CHD is part of the clinical spectrum, such as *TBX1* (22q11.2 deletion, OMIM#188400, MIM:602054), *EHMT1* (9q34.3 deletion or the Kleefstra syndrome OMIM#610253, MIM:607001), *GATA4* (MIM:600576, mapping in to the 8p23.1 deletion), and other genes deemed critical for heart development (reviewed by Andersen et al., 2014; Lalani and Belmont, 2014). However, many of the newly discovered CNVs do not contain a yet well-established cardiac-related gene, and few are recurrent. We and others (Glessner et al., 2014; White et al., 2014) have therefore applied functional and pathway analyses to identify additional candidate genes, in order to establish mechanistic and/or developmental

relationships between these rare events. To date, most studies have employed a limited repertoire of functional approaches and few have replicated findings from other studies (Glessner et al., 2014; Lalani et al., 2013; Silversides et al., 2012).

In an attempt to reduce disease heterogeneity, we sought to identify recurrent CNVs, candidate gene sets and developmental mechanisms associated with a specific subset of CHD, namely conotruncal and related defects. These defects are thought to share a common genetic etiology based on family and animal studies (Digilio et al., 2000; Gobel et al., 1993; Miller and Smith, 1979). To that end we studied one of the largest cohorts to date with conotruncal defects whose cases did not carry a known genetic diagnosis, used denser SNP-based arrays to increase resolution in a subset of cases, applied a range of pathway and functional analyses, and compared our results to those previously published.

Methods

Study Cohorts

This study was approved by The Children's Hospital of Philadelphia (CHOP) Institutional Review Board. Study subjects and their parents were recruited, consented, and diagnosed in a uniform manner at the CHOP Cardiac Center. Study subjects were approached to participate if they had a conotruncal or related cardiac defect and had not been diagnosed with a recognized genetic syndrome upon review of their medical record (e.g. 22q11.2 deletion syndrome, Trisomy 21, Alagille syndrome). Reports from echocardiograms, cardiac catheterizations, cardiac magnetic resonance imaging or cardiac operative notes were reviewed to detail the cardiac anatomy. Medical records, including available consults performed by clinical geneticists, were reviewed to detail non-cardiac congenital anomalies. Family medical history was obtained by an interview conducted by a genetic counselor. DNA was extracted from whole blood collected from parents; proband DNA was either extracted from whole blood or in certain cases, from an established lymphoblastoid cell line, using the Puregene DNA isolation kit (Gentra Systems Inc., Minneapolis, MN).

Three independent groups of healthy controls were used in this study. Healthy control samples (N=4255, Healthy_CHOP) were recruited from well-child visits (ages 3–18 years) within CHOP's healthcare network as previously described (Glessner et al., 2009). All healthy control samples for this study were carefully examined by genotype and health record to exclude samples with any indications of CHD, evidence of chronic health issues, documented genetic abnormalities, or syndromic genomic diseases. Genomic DNA was obtained from whole blood using standard protocols.

A second group of healthy adult controls (N=2156), which were part of a previously published study of candidate genes for ocular refraction in the Age Related Eye Diseases Study (AREDS), were downloaded from dbGaP (dbGaP Study Accession: phs000001.v3.p1) (Wojciechowski et al., 2013).

A third control cohort, 179 HapMap CEU samples genotyped using Illumina HumanOmni 2.5M Beadchip Array, was downloaded from the Illumina data depository (ftp.illumina.com).

Array Genotyping

All CHOP samples, including all conotruncal patients and controls in the healthy CHOP cohort (N=4255), were genotyped following a consistent protocol at CHOP's Center for Applied Genomics. The majority of conotruncal cases (n= 627) and all of the healthy controls were array genotyped on the Illumina Infinium™ II HumanHap550 v1 or v3, or BeadChip 610 array (Illumina, San Diego, CA) as previously described (Elia et al., 2012). The remaining cases (n= 346) were array genotyped using the HumanOmni2.5-8 BeadChip array. The standard Illumina cluster file downloaded from the Illumina website was used for the analysis and running the GenomeStudio clustering algorithm. Control samples from the AREDS study were genotyped using the Illumina HumanOmni2.5 Quad BeadChip array with the standard Illumina cluster file as previously described (dbGaP Study Accession: phs000429.v1.p1 (Simpson et al., 2013)).

Sample Quality Control

Subject gender was verified by the CNV Workshop software package (Gai et al., 2010; Gai et al., 2012). Exclusion criteria for genotypes included SNP call rate <98%, probe intensity LRR ≥3 standard deviations from the cohort mean (0.36), excess of inheritance errors within trios, non-European ancestry as determined by Plink sample stratification (Patterson et al., 2006; Price et al., 2006; Purcell et al., 2007), or gender inconsistencies between self-reported and genotype-derived values.

CNV detection and analysis

We grouped cases and controls into two mutually exclusive cohorts. Cohort 1 included all cases and controls genotyped using the Illumina Infinium™ II HumanHap550 v1 or v3, or BeadChip 610 array. Cohort 2 included cases and AREDS control samples genotyped using the Illumina 2.5M BeadChip.

In order to correct for differences in SNP probe content among all three SNP array versions used in Cohort 1, analysis was limited to the subset of SNPs shared by all three genotyping arrays (535,591 SNPs). CNV Workshop (Gai et al., 2010; Gai et al., 2012) and PennCNV (Wang et al., 2007) were used to define CNV regions as previously described (White et al., 2014).

We applied the same approach for samples in Cohort 2 to adjust for the different versions of Illumina 2.5M BeadChip arrays between cases (Illumina HumanOmni2.5-8v1) and controls (Illumina HumanOmni2.5-4). For the 2.5M arrays, the subset of 2,332,843 SNPs in common between the two platforms was used to predict CNV regions in genotyped samples. In addition, we used 179 Hapmap Caucasian samples that were genotyped using HumanOmni2.5-8v1 BeadChip array (Illumina) to further reduce any systemic bias potentially introduced by different genotyping technologies used in Cohort 2. Hapmap samples were processed in a manner consistent with the Cohort 2 cases. Quality filtered CNV calls from HapMap samples were used as a validation set. Any genes, functional terms, or gene network clusters deemed as significant by comparing HapMap samples to the AREDS cohort control samples (nominal p-value< 0.05) were removed from further consideration, as these findings could be due to systemic bias.

All of the analyses described below were performed in each cohort independently and repeated in the Combined Cohort, generated by merging Cohort 1 and Cohort 2.

CNV Quality Control

CNV calls were considered for further review only if predicted by both algorithms for ≥60% of the predicted CNV span, with the exception of certain large CNVs as specified below. Subject genotypes with total CNV burden ≥3 standard derivations from the cohort mean were removed from further analysis (Pankratz et al., 2011). To reduce the possibility of type I error, deletions spanning less than 5 consecutive SNPs and duplications spanning less than 10 consecutive SNPs in Cohort 1 were excluded. Given that Cohort 2 was genotyped on a higher density array, we adopted a higher threshold for Cohort 2 such that deletions spanning less than 10 consecutive SNPs and duplications spanning less than 20 consecutive SNPs were excluded. In both cohorts, deletions spanning less than 10 kilobases and duplications spanning less than 20 kilobases were removed. CNV SNP and length thresholds were selected based upon previous studies from our group (Elia et al., 2012; Gai et al., 2012; Shaikh et al., 2009; White et al., 2014), examination of size-based concordance rates between the two algorithms (White et al., 2014), and extensive experience with samples undergoing array-based clinical diagnostics at our institution (Conlin et al., 2010).

Additional CNV exclusion criteria included: CNVs with ≥50% overlap with centromere, telomere, and immunoglobulin variable regions; CNVs within olfactory receptor genes; and CNVs with SNP densities ≤ 1 SNP/30 kilobases, as described in (Hasin et al., 2008; Hellemans et al., 2007; Young et al., 2008). CNVs were considered equivalent if their genomic regions reciprocally overlapped for ≥60% of their length. ***Large CNVs*** were defined as those falling within the top 5% of CNVs observed in the corresponding control cohorts, ***inherited CNVs*** as equivalent CNVs identified in a subject and either parent, ***rare CNVs*** as being observed in one or fewer controls (<0.05% frequency in controls), and ***very rare CNVs*** as those not observed in the control cohort (White et al., 2014). B-allele frequencies (BAF) and signal intensity Log R ratios (LRR) of large CNVs were also visually inspected in GenomeStudio (Illumina). Large CNVs within 10 kilobases of each other were also visually inspected in GenomeStudio, and if the BAF and LRR traces indicated likelihood of a single contiguous event, the CNV regions were merged. Predicted CNVs were annotated using the RefSeq gene list (Pruitt et al., 2005), as represented in the UCSC Genome Browser (Kent et al., 2002) (genome.ucsc.edu).

Functional analysis

Gene Ontology (GO) (Ashburner et al., 2000) annotations were retrieved from Ensembl.org (huseast.ensembl.org/index.html) using the BioMart data-mining tool (Smedley et al., 2015). Mammalian Phenotype Ontology (MPO) term annotations were obtained from the Mammalian Genome Informatics resource (MGI) (www.informatics.jax.org)(Eppig et al., 2015). Functional annotation of Reactome (www.reatome.org) (Croft et al., 2014; Milacic et al., 2012) and KEGG (www.kegg.jp) (Kanehisa and Goto, 2000; Kanehisa et al., 2016) gene set collections were downloaded from the GSEA database (www.broadinstitute.org/gsea/msigdb/index.jsp) (Mootha et al., 2003). All annotations were studied to assess gene set enrichments in cases as compared to controls. Gene Ontology and Mammalian Phenotype

Ontology analyses included child and antecedent parental terms associated with a given gene. The extent of statistical enrichment for each functional term was determined by applying Fisher's Exact Test (two-sided), which directly compared the frequency of occurrence in case and control cohorts for each gene or CNV being considered. We applied the Benjamini-Hochberg False Discovery Rate procedure (Benjamini and Hochberg, 1995) to further eliminate any potential family-wise type I error. For global CNV and gene analyses, amplification and deletion events were considered both in aggregate and separately at each locus considered. We only reported a finding when the functions' nominal p-value was less than 0.05 in each cohort and the False Discovery Rate measured in the merged cohort was less than 0.05 (Figure 1).

Knowledge-based Analysis

A subset of genes of particular interest for cardiac development and congenital cardiac defects was compiled in an unsupervised manner by considering prior knowledge of the biomedical literature or expression status in heart tissue. We used 47 terms descriptive of conotruncal defects or general cardiac development through an analysis of MEDLINE articles using natural language processing methods. Gene-Cardiac terms were required to be associated with at least three articles in order to eliminate type I error.

Gene network construction

To construct a network among our genes of interest, especially rare genes among patient cohorts, we used the Cytoscape ReactomeFIViz Gene Set/Mutation Analysis application with default parameters. (Cytoscape version 3.2, f1000research.com/articles/3–146/v2) (Shannon et al., 2003; Wu et al., 2014) Gene interaction networks obtained were clustered into modules using ReactomeFIViz's Cluster FI Network function. A pathway enrichment analysis was employed on each individual network module using the Analyze Module Functions tool. Only pathways with a FDR <0.05 were reported in order to reduce family wise type I error.

Cardiac Gene sets

Two mouse gene expression profiles were compiled and tested for enrichment among our collection of case CNVs using Fisher's Exact test. Known cardiac relevance was assayed by using previously reported gene lists that compiled mouse genes ranked by level of expression in the developing mouse heart at days E9.5 and E14.5 (Zaidi et al., 2013). All mouse transcripts were converted to human gene homologs and subsequently ranked by their relative expression levels. The "high heart expressed 9.5" (HHE_9.5) list contains genes within the top quartile of expression levels (n = 4402) at E9.5, while the "high heart expressed _14.5" (HHE_14.5) list contains genes within the top quartile of expression levels at E14.5. Gene lists with expression levels ranked in the lowest quartile were also compiled ("low heart expressed 9.5" (LHE_9.5), and "low heart expressed_14.5" (LHE_14.5). For each gene list, differing thresholds of inclusion were also explored to measure the trend of enrichments among conotruncal patient cohorts.

We repeated our gene function and network studies restricting the gene list to those present in very rare CNVs and a third high-heart expressed gene list that combined HHE_9.5 and

HHE 14.5 (HHE: combined HHE_14.5 and 9.5) given that HHE_9.5 and HHE 14.5 shared approximately 80% gene identity. Selected genes were imported into DAVID Bioinformatics website (Huang da et al., 2009a; b) and Reactome FI application to evaluate gene functional and regulation network properties as previously described. We also repeated our analysis restricting the gene list to those present in very rare CNVs and the low-heart expressed gene list (LHE: combined LHE_14.5 and 9.5) to eliminate any false positive findings resulted from systemic gene set annotation bias by either DAVID Bioinformatics or Reactome FI.

Statistics Test Utility

The Wilkoxon rank sum test, two way ANOVA test (Type III Sums of Squares), or two tailed Fisher's Exact Test, as appropriate, were used to test significance in case-control CNV and gene enrichment analyses. The Benjamini Hochberg False Discovery Rate (BH-FDR) procedure was applied to adjust for family-wise multiple hypotheses testing.

CNV validation

Selected CNVs, based on likely candidacy, statistical likelihood, or putative function, were validated using TaqMan® copy number assays (Life Technologies, Grand Island, NY). Selection was based on CNV size (<100 kb) and on available human disease information (OMIM: omim.org). An RNAse P TaqMan assay was used as the internal control. Assays were performed on an ABI 7500 Fast Realtime PCR System (Life Technologies) using standard conditions and analyzed with the 7500 Fast System SDS v.1.4.0 software (Life Technologies). All samples were assayed in triplicate and negative results were verified at least twice in independent experiments.

Results

Study cohort

A total of 973 cases (Cohort 1 + Cohort 2) with a definitive diagnosis of a conotruncal or related heart malformation who upon review of medical records did not carry the diagnosis of a known genetic syndrome were used for these analyses (Table 1). All cases were recruited at the CHOP Cardiac Center and passed our rigid quality control process as detailed in Methods. Most cases were ascertained at less than one year of age (63% of Cohort 1, 52% Cohort 2, 59% overall), and 71% of cases were ascertained at less than five years of age. As such, while we divided the cohort into those with and without additional congenital anomalies for subgroup analyses, we could not consider the presence of neurodevelopmental disorders given the young age of the study population. A first-degree relative was reported to have CHD in 6% (n=59) of cases. Array genotyped parental samples were only available for Cohort 1 for which there were 367 complete case-parent trios (both parents and case) and 199 incomplete case-parent trios (one parent and case). The type, number, and frequency of specific cardiac abnormalities from both cohorts are listed in Table 1. All Cohort 1 (n=627) and Cohort 2 (n=346) cases were of European descent. There was no gender difference between the two cohorts with a proband gender ratio of 1.5:1 (376 males) and 1.34:1 (198 males) in Cohort 1 and 2, respectively (p-value=0.44, Fisher's Exact Test). A total of 4833 healthy subjects (2980 in Cohort 1 and 1853 in Cohort 2) passed our quality control steps outlined above and were used as controls as detailed in Methods.

CNV burden in conotruncal patient cohorts

Structural variation content of the 627 cases in Cohort 1 totaled 2735 CNVs, consisting of 553 duplications, 2083 heterozygous deletions, 90 homozygous deletions, and 9 hemizygous deletions (deletions in male X-chromosome) (Figure 2; Supplemental Table S1a). Of these, 1407 (51.4%) could be definitively identified as inherited (710 maternal, 636 paternal, and 61 present in both parents), while 487 were present in neither parent and were thus suggestive of *de novo* events. Of these *de novo* CNVs, 145 were very rare (5.3% of total CNVs) and identified in 105 subjects (16.7% of subjects). Previous work had established bias towards Type II error using the protocol proposed by (Itsara et al., 2010). Therefore, certain of these *de novo* events were likely due to Type II error and present in a parent; those of interest were validated by quantitative PCR, as described in Methods. We detected no significant differences in the overall CNV frequency ($P>0.05$, case/control ratio=1.00) or CNV size ($P>0.05$, case/control ratio=1.05) between cases and controls. This lack of correlation was upheld when considering only the subset of CNVs overlapping transcribed regions between cases and controls ($P>0.05$, mean case/control ratio=1.00 for CNV frequency, mean case/control ratio=1.08 for CNV size). The same conclusion was observed when we restricted the CNV-derived gene list to those overlapping with the HHE genes (CNV frequency: p-value>0.05, mean case/control ratio =1; CNV size: p-value >0.05, mean case/control ratio=1.04). When restricting CNV burden analysis to the 367 conotruncal trios, parental transmission of inherited CNVs to probands was found to be independent of parent gender ($P>0.05$; 654 maternal vs. 655 paternal).

We detected 3192 total CNVs from 346 singletons of Cohort 2, including 2270 heterozygous deletions, 283 homozygous deletions, and 639 duplications (Supplemental Table S1b). We again detected no significant differences in the overall CNV frequency ($P>0.05$, case/control ratio=1.00) or CNV size ($P>0.05$, case/control ratio=1.05) between cases and controls in Cohort 2. As Cohort 2 had no trio data, we were unable to determine inheritance status.

We defined rare CNVs as those present in less than 0.05% of healthy controls whether inherited or *de novo*. By this definition, Cohort 1 contained 836 rare CNVs (263 duplications, 568 heterozygous deletions, and 5 hemizygous X chromosome deletions) and Cohort 2 contained 888 rare CNVs (276 duplications, 611 heterozygous deletions, and one homozygous deletion). The overall distribution of CNVs in both cohorts is depicted in Figure 2.

The burden of rare CNVs was assessed in each cohort (Table 2). Rare CNVs were significantly overrepresented in cases, both when comparing the proportion of subjects with rare CNVs or the frequency of rare CNVs in cases and controls. Rare CNV burden remained significant for overall large CNVs (CNVs with size larger than 3 times of standard derivation of mean CNV size in controls), suggesting similar overall CNV burden characteristics for each cohort. A subgroup analysis comparing the burden of rare CNVs in cases with and without additional non-cardiac anomalies showed significant enrichment as compared to controls (Table 2) while there was no difference comparing one to the other (Supplemental Table S2).

Gene analysis

In Cohort 1, a total of 1217 CNVs included one or more genes, collectively representing 1816 individual genes (Supplemental Table S3). We determined that 314 of these genes were included in CNVs in two or more individuals; of these, only 42 genes were not included in CNVs in controls. In Cohort 2, 1412 CNVs included 1458 individual genes (Supplemental Table S3). We determined that 364 of these genes were included in CNVs in two or more individuals; of these, only 54 genes were not included in CNVs in controls. When combined, 55 genes were included in CNVs in both cohorts at least once but not in any controls (23 genes were in deletions in both cohorts, 22 genes were in duplications in both cohorts, and 10 genes were in different types of CNVs in the two case cohorts; Supplemental Table S4).

We performed a gene-based case-control enrichment analysis of conotruncal CNV-associated genes to determine if any genes were overrepresented in cases. No genes remained significantly enriched in our cases when all CNVs or only deletions or duplications were considered after correcting for multiple tests in the Combined Cohort (see Figure 1). We observed the same conclusion when the analysis was restricted to the subset of HHE genes.

We next restricted our analysis to include only a subset of genes (1534 genes in total) previously implicated in cardiovascular development from the biomedical literature, as described in Methods. Using this process, we identified 37 such genes within 39 CNVs (10 duplications and 29 heterozygous deletions) in Cohort 1 and 40 genes within 89 CNVs (21 duplications and 68 heterozygous deletions) in Cohort 2. Among those CNVs, 29 of 39 were rare CNVs in Cohort 1 (7 duplications and 22 deletions) and 27 of 89 were rare in Cohort 2 (10 duplications and 17 deletions). Three of these rare CNVs were present in both Cohort 1 and 2, all of which have been identified in other CHD studies. These included 2 very rare chromosome 1q21 deletions that overlapped with previously reported CNVs deleting the gene *GJA5* (Digilio et al., 2013; Glessner et al., 2014; Greenway et al., 2009; Silversides et al., 2012; Soemedi et al., 2012a; Tomita-Mitchell et al., 2012; Warburton et al., 2014). A smaller very rare CNV in the same region deleting only *CHD1L* was found in a single case from Cohort 2. The other two recurrent CNVs in our cohort disrupted genes *ANGPT2* (Silversides et al., 2012) and *FLT4*, respectively (Serra-Juhe et al., 2012; Soemedi et al., 2012b) (Table 3). Several other rare CNVs found only in one of our cohorts were also reported in other CHD studies. These CNVs are listed in Table 3 and overlapped genes of interest at 5q14.1 (*SSBP2*) (Silversides et al., 2012; Soemedi et al., 2012b), and 3q22.1 (*NPHP3*) (Tomita-Mitchell et al., 2012).

We compared genes included in rare CNVs in the conotruncal cases to healthy controls in order to determine whether those genes were preferentially enriched among heart specific mouse-human homolog gene expression sets. We did not observe conclusive association of the HHE as compared to LHE genes in neither case cohorts as compared to controls (Supplemental Table S5).

Functional and pathway analysis

Several approaches were used to determine whether genes sharing particular biological functions were enriched within rare CNVs in conotruncal subjects. Using the full gene set from rare CNVs, we studied each case-control cohort independently and then rare CNVs in combined cohorts for the analysis, as described in Methods (*Functional Analysis*). We intended to determine whether Gene Ontology (GO) terms assigned to genes overlapping detected CNVs were significantly enriched in conotruncal cases versus controls. Sixty-six unique Gene Ontology terms were found to be significantly enriched. Several terms relevant to heart or early embryonic development, and terms that included known cardiac-related genes were significant after multiple testing correction (Table 4). GO terms of significance and interest included: "***Regulation of sequence-specific DNA binding transcription factor activity***" for its inclusion of TGFβ1 (FDR<2.38E-04), and the potentially related GO term "***Regulation of transforming growth factor beta receptor signaling pathway***" (FDR<3.26E-02) given the relationship of TGFβ1 to heart development (Gordon and Blobe, 2008). A recent study showed that cilium-related genes were highly correlated with heart formation and defects in a mouse model (Li et al., 2015). We found that the Gene Ontology term "***Non-motile primary cilium***" was highly enriched in our case cohorts (FDR<8.48E-03). Other GO terms of interest included "***Cardiac muscle cell differentiation***" (FDR<6.17E-04), "***Positive regulation of Rho GTPase activity***" (FDR<3.46E-02) and "***Chromosome organization***" (FDR<1.9E-02). A full list of all the significant Gene Ontology terms is provided in the supplemental material (Supplemental Table S6).

As a complementary case-control approach, we evaluated whether rare conotruncal CNVs were preferentially enriched for gene orthologs responsible for specific phenotypes found in mouse models for congenital heart defects. For this analysis, we used MGI-derived MPO assignments reported for CNV-associated genes in the conotruncal cohort, as compared to such genes in the control cohort. Forty-two mouse phenotype terms were identified as significantly enriched in conotruncal subjects (Supplemental Table S7). The top significant terms of interest included "***prenatal lethality***" (FDR<3.3E-06) and "***partial embryonic lethality***" (FDR<2.9E-04) as altered function of a wide range of genes contributing to cardiac development have been shown to result in embryonic lethality (reviewed by Clowes et al., 2014; Lockhart et al., 2011; Solloway and Robertson, 1999) (Table 5). As with the Gene Ontology analysis, "***abnormal apoptosis***" (FDR<5.12E-04) was also among the top significant terms of interest.

We extended our functional study to Reactome (www.reatome.org) and KEGG (www.kegg.jp) gene sets. Using KEGG's classification, four terms were identified as significantly over-represented in conotruncal cases including the "***TGF-beta signaling pathway***" (corrected p<1.30E-02) (Table 6). There was no Reactome term significantly enriched in overall conotruncal subjects compared to controls.

We further interrogated gene interaction networks in our cohorts. We collected all genes included in CNVs that were deemed as rare in both case cohorts: 1085 genes from Cohort 1 and 770 genes from Cohort 2. The gene sets from each cohort were imported and analyzed separately and also jointly by the ReactomeFIViz component of CytoScape 3.2 (http://

f1000research.com/articles/3-146/v2). Genes were clustered based on their connectivity, followed by annotating each cluster with pathway enrichment ranks. To reduce type I error, we only studied modules with a false discovery rate less than 0.05. The gene network for each cohort was plotted and is shown in (Figure 3). As indicated in previous result section, "***TGF-beta signaling pathway***" network was the most significantly enriched function among the clusters obtained from both cohorts. Other implied functions included "***Assembly of the primary cilium***", and "***Rho GTPases signaling***," also previously identified in our GO Ontology analysis. Many of those functions were established as playing a role in cardiac development (Clement et al., 2009; Koefoed et al., 2014; Li et al., 2015; Wei et al., 2002).

We subsequently restricted our pathway and functional analyses to HHE genes included in rare conotruncal CNVs. Gene Ontology analysis identified a number of enriched functions that are known to be involved in early development with the term "***Cardiac Muscle Cell Differentiation***" ($p<4.66E-6$) being one of the most significantly enriched functions. Supporting our previous findings the GO terms "***Regulation of transforming growth factor beta receptor signaling pathway***" (GO:0017015, $p<3.523E-4$) and "***Regulation of Rho protein signal transduction***" (GO:0035023, $p<1.857E-3$) were again found to be significantly enriched (Supplemental Table S8a). In Reactome FI analysis, other pathways of interest that were found to be significantly enriched included "***Pre-NOTCH Expression and Processing***" (FDR<0.05) and ***SHP2 signaling*** (FDR<0.05), which includes the gene *ANGPT1*. Other significantly enriched pathways using these and other analytical methods are listed in the Supplemental Tables S8a–e. The list of significant terms using the HHE genes included many functions previously identified using the full gene list.

Discussion

Our CNV study represents one of the largest conducted to date with cardiac conotruncal and related anomalies in cases without a recognized genetic syndrome. In keeping with previous studies, we found an increased burden of rare (and rare large) CNVs in cases as compared to controls. An increased burden of rare CNVs was found in cases regardless of the presence or absence of non-cardiac congenital anomalies. Unfortunately, we could not further sub-divide the study cohort by neurodevelopmental status given that most cases were ascertained at less than one and even five years of age, and our study was not designed to test for or ascertain such issues longitudinally. Whether there is a subset of cases without non-cardiac congenital anomalies but with neurodevelopmental disorders that drives the CNV burden in the subset with no additional congenital anomalies cannot be discerned. However, this situation is identical to that faced by the physician/caregiver examining a fetus/newborn/infant with a conotruncal defect that has no other overt findings and for whom the neurodevelopmental status cannot yet be defined.

Though challenging to compare results from different studies given the range of case phenotypes enrolled and study designs employed, most studies to date have suggested an increased burden of some class of CNVs in CHD cases. In particular, Warburton et al. (2014), Greenway et al. (2009) and Glessner et al. (2014) found an increased prevalence of *de novo* CNVs in cases with conotruncal, and sporadic, severe CHD respectively, while Soemedi and colleagues (2012b) found an increased burden of rare genic deletions in a

cohort that included a large number of TOF patients. Silversides et al. (2012) found an association between large rare CNVs and large rare exonic duplications in a study cohort incorporating syndromic and nonsyndromic TOF cases, which disappeared in the nonsyndromic subset. Thus multiple studies suggest that CNVs contribute to disease risk for CHDs in general and conotruncal defects in particular, but defining a set of recurrent events or disease-associated genes has been difficult to replicate between studies.

As in most other reports, no statistically associated disease-related recurrent CNVs or genes were identified in our study when correction for multiple testing was applied. However, we identified several CNVs in our cases that were previously reported in other CHD studies that included identical candidate genes, thus adding validation to their disease-based impact in CHD. In particular, we found recurrent CNVs at chromosome 1q21 in both of our CHD cohorts, which is one of the most frequently reported CNVs in CHD cases in other reports (reviewed in Digilio et al., 2013; Glessner et al., 2014; Greenway et al., 2009; Silversides et al., 2012; Soemedi et al., 2012a; Soemedi et al., 2012b; Tomita-Mitchell et al., 2012; Warburton et al., 2014). The disease-associated gene within this region is thought to be *GJA5* given the finding of a single nucleotide variant in *GJA5* associated with TOF (Guida et al., 2013) and the finding that mice deleted for Gja5 develop a TOF phenotype (Gu et al., 2003).

In addition, we identified recurrent rare CNVs in both of our cohorts that overlapped with those reported in other studies that did not include genes previously listed as the likely disease-related candidate gene. For example, our CNVs at 8p23.1 and the one reported by Silversides et al. (2012) did not include *GATA4*, but instead deleted *RP1L1*. Such findings suggest that additional genes in these regions may be important for the cardiac phenotype. Alternatively our CNVs could delete regulatory domains that exert a more distant effect on gene expression of purported candidate genes, but ultimately such comparisons between studies are hampered by the use of different technologies and the difficulty defining end points.

We also found CNVs spanning purported candidate genes exclusively in our controls or in both cases and controls, decreasing the likelihood that these candidates are indeed related to CHD. In particular, a previously reported CHD-associated CNV on 15q11.2 (Glessner et al., 2014; Soemedi et al., 2012b), was present in both of our CHD cohorts (8 cases) as well as control samples (11 controls) both as deletions and duplications, and did not show any association. We also identified eight normal parents carrying the 15q11.2 CNV though their affected offspring did not inherit this CNV, an observation we confirmed by qPCR. This observation brings in to question whether this region is related to CHD. Alternatively, the conflicting results between studies may be due to systemic bias introduced by different genotyping arrays in each study or could result from a more complex model of CHD risk.

Because we did not find a significant association of either single genes or CNVs with disease risk examining the full genome, we tested whether restricting the analysis to high heart expressed (HHE) developmental genes found in rare CNVs would identify a set of heart-related genes associated with CTD. We expected this focused approach to reduce heterogeneity by preferentially eliminating unrelated genes concurrently included in CNVs,

thereby increasing power to detect meaningful associations. Unlike recent results reported for *de novo*, damaging mutations identified in whole exome sequence data (Zaidi et al., 2013), the results of this restricted analysis were inconsistent between our cohorts and expressions levels, and thus, were inconclusive. It is possible that a different or more restricted gene set (e.g. a gene set specific to conotruncal developmental such as the second heart field or cardiac neural crest cells) would be more informative.

Given the increasing evidence for marked genetic heterogeneity of CHD, we undertook an extensive pathway analysis to test whether the genetic burden of rare CNVs could be explained by the disruption of one or more distinct but functionally related genes. Our pathway analyses of genes included in rare CNVs suggest significant enrichment of pathways that have been previously associated with cardiac development, such as the TGF-beta signaling pathway, which we identified using multiple approaches. Likewise, genes associated with chromosome organization were enriched in both of our cohorts, previously identified in TOF patients by Silversides et al. (2012), and contained genes previously associated with CHD such as *CHD7, WDR5*, and *USP44* (Zaidi et al., 2013). The fact that many seemingly unrelated pathways also reached significance, such as many immunology centered pathways, might be due to the inclusion of all genes disrupted by the rare CNVs, many of which are likely irrelevant to CTD. Indeed, restricting the analysis to genes highly expressed in the developing heart resulted in more biological specification relative to heart development among significantly associated pathways. Notably, pathways enriched in other studies were not replicated by our analyses (Glessner et al., 2014; Soemedi et al., 2012a; Warburton et al., 2014). The apparent disparate results between studies could derive from different phenotypic cohorts, analytical approaches and/or genotyping platforms, or it may also reflect a lack of statistical power due to the underlying complexity of these disorders. The variability in study design clearly complicates efforts to synchronize findings on this complex genetic disorder.

In summary, our study demonstrates that rare CNVs contribute to disease risk for CTDs and once again highlights the enormous genetic heterogeneity of even this subset of CHD given the paucity of recurrent events. Comparison with other studies both confirms and questions previous associated loci and genes, but the highly variable study design employed by different investigators makes the compilation and comparison of findings between studies challenging. Given the rarity of recurrent single events, the pathway and functional based analyses based on gene content from the rare CNVs appear to be more informative as several developmentally related pathways and networks were enriched in our cohorts, particularly when the gene set examined was confined to those expressed during early heart development. These results suggest that the association of rare CNVs with disease-risk is explained by way of alteration of copy number of developmentally-related genes. Future studies will focus on relevant gene subsets as defined by expression data and defined gene networks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all of the families for their participation and the members of the Cardiac Center for their support of case ascertainment. We thank Jennifer Garbarini and Stacy Woyciechowski for case ascertainment, and Sharon Edman for data management. We are very grateful to Dr. Wendy K. Chung for her review of the manuscript. We thank the genotyping and quality control staff at the Center for Applied Genomics at CHOP for their contributions.

Contract grant sponsors: This research was supported by the US National Institutes of Health (P01-HD070454, P50-HL074731 (E.G.)), National Center for Research Resources, Grant UL1RR024134 (E.G.), which is now at the National Center for Advancing Translational Sciences, Grant UL1TR000003 (E.G.), the National Eye Institute (R01 EY020483 (D.S.)) and Pennsylvania Department of Health (SAP 4100037707 (P.S.W.)).

The content is the sole responsibility of the authors and does not necessarily represent the official views of the NIH.

References

Ackerman KG, Herron BJ, Vargas SO, et al. Fog2 is required for normal diaphragm and lung development in mice and humans. PLoS Genet. 2005; 1:58–65. [PubMed: 16103912]

Andersen TA, Troelsen Kde L, Larsen LA. Of mice and men: molecular genetics of congenital heart disease. Cell Mol Life Sci. 2014; 71:1327–1352. [PubMed: 23934094]

Arthur HM, Bamforth SD. TGFbeta signaling and congenital heart disease: Insights from mouse studies. Birt Defects Res A Clin Mol Teratol. 2011; 91:423–434.

Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]

Banks AS, Kim-Muller JY, Mastracci TL, et al. Dissociation of the glucose and lipid regulatory functions of FoxO1 by targeted knockin of acetylation-defective alleles in mice. Cell Metab. 2011; 14:587–597. [PubMed: 22055502]

Bell SE, Sanchez MJ, Spasic-Boskovic O, et al. The RNA binding protein Zfp36l1 is required for normal vascularisation and post-transcriptionally regulates VEGF expression. Dev Dyn. 2006; 235:3144–3155. [PubMed: 17013884]

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc B. 1995; 57:289–300.

Bergmann C, Fliegauf M, Bruchle NO, et al. Loss of nephrocystin-3 function can cause embryonic lethality, Meckel-Gruber-like syndrome, situs inversus, and renal-hepatic-pancreatic dysplasia. Am J Hum Genet. 2008; 82:959–970. [PubMed: 18371931]

Bleyl SB, Saijoh Y, Bax NA, et al. Dysregulation of the PDGFRA gene causes inflow tract anomalies including TAPVR: integrating evidence from human genetics and model organisms. Hum Mol Genet. 2010; 19:1286–1301. [PubMed: 20071345]

Bonnafe E, Touka M, AitLounis A, et al. The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. Mol Cell Biol. 2004; 24:4417–4427. [PubMed: 15121860]

Breckpot J, Thienpont B, Arens Y, et al. Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects. Cytogenet Genome Res. 2011; 135:251–259. [PubMed: 21921585]

Chaki M, Hoefele J, Allen SJ, et al. Genotype-phenotype correlation in 440 patients with NPHP-related ciliopathies. Kidney Int. 2011; 80:1239–1245. [PubMed: 21866095]

Chang AC, Fu Y, Garside VC, et al. Notch initiates the endothelial-to-mesenchymal transition in the atrioventricular canal through autocrine activation of soluble guanylyl cyclase. Dev Cell. 2011; 21:288–300. [PubMed: 21839921]

Chang Z, Zhang Q, Feng Q, et al. Deletion of Akt1 causes heart defects and abnormal cardiomyocyte proliferation. Dev Biol. 2010; 347:384–391. [PubMed: 20816796]

Choudhary B, Ito Y, Makita T, et al. Cardiovascular malformations with normal smooth muscle differentiation in neural crest-specific type II TGFbeta receptor (Tgfbr2) mutant mice. Dev Biol. 2006; 289:420–429. [PubMed: 16332365]

Christiansen J, Dyck JD, Elyas BG, et al. Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. Circ Res. 2004; 94:1429–1435. [PubMed: 15117819]

Clement CA, Kristensen SG, Mollgard K, et al. The primary cilium coordinates early cardiogenesis and hedgehog signaling in cardiomyocyte differentiation. J Cell Sci. 2009; 122:3070–3082. [PubMed: 19654211]

Clowes C, Boylan MG, Ridge LA, et al. The functional diversity of essential genes required for mammalian cardiac development. Genesis. 2014; 52:713–737. [PubMed: 24866031]

Conlin LK, Thiel BD, Bonnemann CG, et al. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. Hum Mol Genet. 2010; 19:1263–1275. [PubMed: 20053666]

Conlon FL, Lyons KM, Takaesu N, et al. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development. 1994; 120:1919–1928. [PubMed: 7924997]

Croft D, Mundo AF, Haw R, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014; 42:D472–D477. [PubMed: 24243840]

Digilio MC, Bernardini L, Consoli F, et al. Congenital heart defects in recurrent reciprocal 1q21.1 deletion and duplication syndromes: rare association with pulmonary valve stenosis. Eur J Med Genet. 2013; 56:144–149. [PubMed: 23270675]

Digilio MC, Marino B, Musolino AM, et al. Familial recurrence of nonsyndromic interrupted aortic arch and truncus arteriosus with atrioventricular canal. Teratology. 2000; 61:329–331. [PubMed: 10777827]

Dodou E, Verzi MP, Anderson JP, et al. Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development. Development. 2004; 131:3931–3942. [PubMed: 15253934]

Donovan J, Kordylewska A, Jan YN, Utset MF. Tetralogy of fallot and other congenital heart defects in Hey2 mutant mice. Curr Biol. 2002; 12:1605–1610. [PubMed: 12372254]

Dor Y, Camenisch TD, Itin A, et al. A novel role for VEGF in endocardial cushion formation and its potential contribution to congenital heart defects. Development. 2001; 128:1531–1538. [PubMed: 11290292]

Dumont DJ, Jussila L, Taipale J, et al. Cardiovascular failure in mouse embryos deficient in VEGF receptor-3. Science. 1998; 282:946–949. [PubMed: 9794766]

Elia J, Glessner JT, Wang K, et al. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. Nat Genet. 2012; 44:78–84.

Engelmann GL, Boehm KD, Birchenall-Roberts MC, Ruscetti FW. Transforming growth factor-beta 1 in heart development. Mech Dev. 1992; 38:85–97. [PubMed: 1419851]

Eppig JT, Blake JA, Bult CJ, et al. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res. 2015; 43:D726–D736. [PubMed: 25348401]

Erdogan F, Larsen LA, Zhang L, et al. 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. [PubMed: 18713793]

Fisher SA, Langille BL, Srivastava D. Apoptosis during cardiovascular development. Circ Res. 2000; 87:856–864. [PubMed: 11073880]

Fouladkou F, Lu C, Jiang C, et al. The ubiquitin ligase Nedd4-1 is required for heart development and is a suppressor of thrombospondin-1. J Biol Chem. 2010; 285:6770–6780. [PubMed: 20026598]

Fryer CJ, White JB, Jones KA. Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. Mol Cell. 2004; 16:509–520. [PubMed: 15546612]

Gai X, Perin JC, Murphy K, et al. CNV Workshop: an integrated platform for high-throughput copy number variation discovery and clinical diagnostics. BMC Bioinformatics. 2010; 11:74. [PubMed: 20132550]

Gai X, Xie HM, Perin JC, et al. Rare structural variation of synapse and neurotransmission genes in autism. Mol Psychiatry. 2012; 17:402–411. [PubMed: 21358714]

Gao Z, Kim GH, Mackinnon AC, et al. Ets1 is required for proper migration and differentiation of the cardiac neural crest. Development. 2010; 137:1543–1551. [PubMed: 20356956]

Gaspar P, Cases O, Maroteaux L. The developmental role of serotonin: news from mouse molecular genetics. Nat Rev Neurosci. 2003; 4:1002–1012. [PubMed: 14618156]

Glessner JT, Bick AG, Ito K, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. Circ Res. 2014; 115:884–896. [PubMed: 25205790]

Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. 2009; 459:569–573. [PubMed: 19404257]

Gobel JW, Pierpont ME, Moller JH, et al. Familial interruption of the aortic arch. Pediatr Cardiol. 1993; 14:110–115. [PubMed: 8469627]

Gordon KJ, Blobe GC. Role of transforming growth factor-beta superfamily signaling pathways in human disease. Biochim Biophys Acta. 2008; 1782:197–228. [PubMed: 18313409]

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. [PubMed: 19597493]

Gu H, Smith FC, Taffet SM, Delmar M. High incidence of cardiac malformations in connexin40-deficient mice. Circ Res. 2003; 93:201–206. [PubMed: 12842919]

Guida V, Ferese R, Rocchetti M, et al. A variant in the carboxyl-terminus of connexin 40 alters GAP junctions and increases risk for tetralogy of Fallot. Eur J Hum Genet. 2013; 21:69–75. [PubMed: 22713807]

Hamilton TG, Klinghoffer RA, Corrin PD, Soriano P. Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms. Mol Cell Biol. 2003; 23:4013–4025. [PubMed: 12748302]

Hartman RJ, Rasmussen SA, Botto LD, et al. The contribution of chromosomal abnormalities to congenital heart defects: a population-based study. Pediatr Cardiol. 2011; 32:1147–1157. [PubMed: 21728077]

Hasin Y, Olender T, Khen M, et al. High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. PLoS Genet. 2008; 4:e1000249. [PubMed: 18989455]

Hellemans J, Mortier G, De Paepe A, et al. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 2007; 8:R19. [PubMed: 17291332]

Hoff S, Halbritter J, Epting D, et al. ANKS6 is a central component of a nephronophthisis module linking NEK8 to INVS and NPHP3. Nat Genet. 2013; 45:951–956. [PubMed: 23793029]

Huang, da W., Sherman, BT., Lempicki, RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009a; 37:1–13. [PubMed: 19033363]

Huang, da W., Sherman, BT., Lempicki, RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009b; 4:44–57. [PubMed: 19131956]

Huang J, Cheng L, Li J, et al. Myocardin regulates expression of contractile genes in smooth muscle cells and is required for closure of the ductus arteriosus in mice. J Clin Invest. 2008; 118:515–525. [PubMed: 18188448]

Huang J, Elicker J, Bowens N, et al. Myocardin regulates BMP10 expression and is required for heart development. J Clin Invest. 2012; 122:3678–3691. [PubMed: 22996691]

Huang J, Wang T, Wright AC, et al. Myocardin is required for maintenance of vascular and visceral smooth muscle homeostasis during postnatal development. Proc Natl Acad Sci U S A. 2015; 112:4447–4452. [PubMed: 25805819]

Itsara A, Wu H, Smith JD, et al. De novo rates and selection of large copy number variation. Genome Res. 2010; 20:1469–1481. [PubMed: 20841430]

Jorge AA, Malaquias AC, Arnhold IJ, Mendonca BB. Noonan syndrome and related disorders: a review of clinical features and mutations in genes of the RAS/MAPK pathway. Horm Res. 2009; 71:185–193. [PubMed: 19258709]

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30. [PubMed: 10592173]

Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016; 44:D457–D462. [PubMed: 26476454]

Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]

Koefoed K, Veland IR, Pedersen LB, et al. Cilia and coordination of signaling networks during heart development. Organogenesis. 2014; 10:108–125. [PubMed: 24345806]

Kume T, Jiang H, Topczewska JM, Hogan BL. The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. Genes Dev. 2001; 15:2470–2482. [PubMed: 11562355]

Lalani SR, Belmont JW. Genetic basis of congenital cardiovascular malformations. Eur J Med Genet. 2014; 57:402–413. [PubMed: 24793338]

Lalani SR, Shaw C, Wang X, et al. Rare DNA copy number variants in cardiovascular malformations with extracardiac abnormalities. Eur J Hum Genet. 2013; 21:173–181. [PubMed: 22929023]

Lee YC, Chang CJ, Bali D, et al. Glycogen-branching enzyme deficiency leads to abnormal cardiac development: novel insights into glycogen storage disease IV. Hum Mol Genet. 2011; 20:455–465. [PubMed: 21075835]

Li Y, Klena NT, Gabriel GC, et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. Nature. 2015; 521:520–524. [PubMed: 25807483]

Lincoln J, Kist R, Scherer G, Yutzey KE. Sox9 is required for precursor cell expansion and extracellular matrix organization during mouse heart valve development. Dev Biol. 2007; 305:120–132. [PubMed: 17350610]

Lockhart M, Wirrig E, Phelps A, Wessels A. Extracellular matrix and heart development. Birt Defects Res A Clin Mol Teratol. 2011; 91:535–550.

Lowe LA, Yamada S, Kuehn MR. Genetic dissection of nodal function in patterning the mouse embryo. Development. 2001; 128:1831–1843. [PubMed: 11311163]

Lu, Jenher, Tsai, Tzuchun, Choo, Sielin, et al. Induction of apoptosis and inhibition of cell growth by tbx5 knockdown contribute to dysmorphogenesis in Zebrafish embryos. J Biomed Sci. 2011; 18:73–73. [PubMed: 21982178]

Ma MC, Li P, Shen H, et al. Dysregulated endocardial TGFbeta signaling and mesenchymal transformation result in heart outflow tract septation failure. Dev Biol. 2016; 409:272–276. [PubMed: 26522286]

McElhinney DB, Geiger E, Blinder J, et al. NKX2.5 mutations in patients with congenital heart disease. J Am Coll Cardiol. 2003; 42:1650–1655. [PubMed: 14607454]

Milacic M, Haw R, Rothfels K, et al. Annotating cancer variants and anti-cancer therapeutics in reactome. Cancers (Basel). 2012; 4:1180–1211. [PubMed: 24213504]

Miller ME, Smith DW. Conotruncal malformation complex: examples of possible monogenic inheritance. Pediatrics. 1979; 63:890–893. [PubMed: 450526]

Mohapatra B, Casey B, Li H, et al. Identification and functional characterization of NODAL rare variants in heterotaxy and isolated cardiovascular malformations. Hum Mol Genet. 2009; 18:861–871. [PubMed: 19064609]

Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34:267–273. [PubMed: 12808457]

Nebigil CG, Choi DS, Dierich A, et al. Serotonin 2B receptor is required for heart development. Proc Natl Acad Sci U S A. 2000; 97:9508–9513. [PubMed: 10944220]

Nebigil CG, Maroteaux L. Functional consequence of serotonin/5-HT2B receptor signaling in heart: role of mitochondria in transition between hypertrophy and heart failure? Circulation. 2003; 108:902–908. [PubMed: 12925446]

Norris RA, Moreno-Rodriguez RA, Sugi Y, et al. Periostin regulates atrioventricular valve maturation. Dev Biol. 2008; 316:200–213. [PubMed: 18313657]

Oswald F, Kostezka U, Astrahantseff K, et al. SHARP is a novel component of the Notch/RBP-Jkappa signalling pathway. EMBO J. 2002; 21:5417–5426. [PubMed: 12374742]

Pankratz N, Dumitriu A, Hetrick KN, et al. Copy number variation in familial Parkinson disease. PLoS ONE. 2011; 6:e20988. [PubMed: 21829596]

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]

Pierpont ME, Basson CT, Benson DW Jr, et al. Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. Circulation. 2007; 115:3015–3038. [PubMed: 17519398]

Pizzuti A, Sarkozy A, Newton AL, et al. Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. Hum Mutat. 2003; 22:372–377. [PubMed: 14517948]

Poelmann RE, Gittenberger-de Groot AC. Apoptosis as an instrument in cardiovascular development. Birth Defects Res C Embryo Today. 2005; 75:305–313. [PubMed: 16425248]

Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]

Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2005; 33:D501–D504. [PubMed: 15608248]

Purandare SM, Ware SM, Kwan KM, et al. A complex syndrome of left-right axis, central nervous system and axial skeleton defects in Zic3 mutant mice. Development. 2002; 129:2293–2302. [PubMed: 11959836]

Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

Rezvani M, Barrans JD, Dai KS, Liew CC. Apoptosis-related genes expressed in cardiovascular development and disease: an EST approach. Cardiovasc Res. 2000; 45:621–629. [PubMed: 10728383]

Rigor DL, Bodyak N, Bae S, et al. Phosphoinositide 3-kinase Akt signaling pathway interacts with protein kinase Cbeta2 in the regulation of physiologic developmental hypertrophy and heart function. Am J Physiol Heart Circ Physiol. 2009; 296:H566–H572. [PubMed: 19122165]

Rios H, Koushik SV, Wang H, et al. periostin null mice exhibit dwarfism, incisor enamel defects, and an early-onset periodontal disease-like phenotype. Mol Cell Biol. 2005; 25:11131–11144. [PubMed: 16314533]

Rochais F, Dandonneau M, Mesbah K, et al. Hes1 is expressed in the second heart field and is required for outflow tract development. PLoS ONE. 2009; 4:e6267. [PubMed: 19609448]

Ruiz-Perez VL, Ide SE, Strom TM, et al. Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodental dysostosis. Nat Genet. 2000; 24:283–286. [PubMed: 10700184]

Sakata Y, Koibuchi N, Xiang F, et al. The spectrum of cardiovascular anomalies in CHF1/Hey2 deficient mice reveals roles in endocardial cushion, myocardial and vascular maturation. J Mol Cell Cardiol. 2006; 40:267–273. [PubMed: 16242143]

Schachterle W, Rojas A, Xu SM, Black BL. ETS-dependent regulation of a distal Gata4 cardiac enhancer. Dev Biol. 2012; 361:439–449. [PubMed: 22056786]

Sengupta A, Chakraborty S, Paik J, et al. FoxO1 is required in endothelial but not myocardial cell lineages during cardiovascular development. Dev Dyn. 2012; 241:803–813. [PubMed: 22411556]

Serra-Juhe C, Rodriguez-Santiago B, Cusco I, et al. Contribution of rare copy number variants to isolated human malformations. PLoS ONE. 2012; 7:e45530. [PubMed: 23056206]

Shaikh TH, Gai X, Perin JC, et al. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res. 2009; 19:1682–1690. [PubMed: 19592680]

Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504. [PubMed: 14597658]

Silversides CK, Lionel AC, Costain G, et al. Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. PLoS Genet. 2012; 8:e1002843. [PubMed: 22912587]

Simpson CL, Wojciechowski R, Yee SS, et al. Regional replication of association with refractive error on 15q14 and 15q25 in the Age-Related Eye Disease Study cohort. Mol Vis. 2013; 19:2173–2186. [PubMed: 24227913]

Smedley D, Haider S, Durinck S, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015; 43:W589–W598. [PubMed: 25897122]

Soemedi R, Topf A, Wilson IJ, et al. Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. Hum Mol Genet. 2012a; 21:1513–1520. [PubMed: 22199024]

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. 2012b; 91:489–501. [PubMed: 22939634]

Solloway MJ, Robertson EJ. Early embryonic lethality in Bmp5;Bmp7 double mutant mice suggests functional redundancy within the 60A subgroup. Development. 1999; 126:1753–1768. [PubMed: 10079236]

Soriano P. The PDGF alpha receptor is required for neural crest cell development and for normal patterning of the somites. Development. 1997; 124:2691–2700. [PubMed: 9226440]

Stumpo DJ, Byrd NA, Phillips RS, et al. Chorioallantoic fusion defects and embryonic lethality resulting from disruption of Zfp36L1, a gene encoding a CCCH tandem zinc finger protein of the Tristetraprolin family. Mol Cell Biol. 2004; 24:6445–6455. [PubMed: 15226444]

Tartaglia M, Gelb BD. Noonan syndrome and related disorders: genetics and pathogenesis. Annu Rev Genomics Hum Genet. 2005; 6:45–68. [PubMed: 16124853]

Tobin JL, Beales PL. The nonmotile ciliopathies. Genet Med. 2009; 11:386–402. [PubMed: 19421068]

Tomita-Mitchell A, Mahnke DK, Struble CA, et al. Human gene copy number spectra analysis in congenital heart malformations. Physiol Genomics. 2012; 44:518–541. [PubMed: 22318994]

Tomita-Mitchell A, Maslen CL, Morris CD, et al. GATA4 sequence variants in patients with congenital heart disease. J Med Genet. 2007; 44:779–783. [PubMed: 18055909]

Tory K, Rousset-Rouviere C, Gubler MC, et al. Mutations of NPHP2 and NPHP3 in infantile nephronophthisis. Kidney Int. 2009; 75:839–847. [PubMed: 19177160]

van Gijn ME, Blankesteijn WM, Smits JF, et al. Frizzled 2 is transiently expressed in neural crest-containing areas during development of the heart and great arteries in the mouse. Anat Embryol (Berl). 2001; 203:185–192. [PubMed: 11303904]

Ver Heyen M, Heymans S, Antoons G, et al. Replacement of the muscle-specific sarcoplasmic reticulum Ca(2+)-ATPase isoform SERCA2a by the nonmuscle SERCA2b homologue causes mild concentric hypertrophy and impairs contraction-relaxation of the heart. Circ Res. 2001; 89:838–846. [PubMed: 11679415]

Vincentz JW, Barnes RM, Firulli AB. Hand factors as regulators of cardiac morphogenesis and implications for congenital heart defects. Birt Defects Res A Clin Mol Teratol. 2011; 91:485–494.

Wang K, Li M, Hadley D, et al. 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. [PubMed: 17921354]

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. [PubMed: 23979609]

Ware SM, Harutyunyan KG, Belmont JW. Heart defects in X-linked heterotaxy: evidence for a genetic interaction of Zic3 with the nodal signaling pathway. Dev Dyn. 2006; 235:1631–1637. [PubMed: 16496285]

Washington Smoak I, Byrd NA, Abu-Issa R, et al. Sonic hedgehog is required for cardiac outflow tract and neural crest cell development. Dev Biol. 2005; 283:357–372. [PubMed: 15936751]

Wei L, Imanaka-Yoshida K, Wang L, et al. Inhibition of Rho family GTPases by Rho GDP dissociation inhibitor disrupts cardiac morphogenesis and inhibits cardiomyocyte proliferation. Development. 2002; 129:1705–1714. [PubMed: 11923206]

Weninger WJ, Lopes Floro K, Bennett MB, et al. Cited2 is required both for heart morphogenesis and establishment of the left-right axis in mouse development. Development. 2005; 132:1337–1348. [PubMed: 15750185]

White PS, Xie HM, Werner P, et al. Analysis of chromosomal structural variation in patients with congenital left-sided cardiac lesions. Birt Defects Res A Clin Mol Teratol. 2014; 100:951–964.

Wojciechowski R, Yee SS, Simpson CL, et al. Matrix metalloproteinases and educational attainment in refractive error: evidence of gene-environment interactions in the Age-Related Eye Disease Study. Ophthalmology. 2013; 120:298–305. [PubMed: 23098370]

Wu G, Dawson E, Duong A, et al. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis [version 2; referees: 2 approved]. F1000Res. 2014; 3:146. [PubMed: 25309732]

Xu D, Qu CK. Protein tyrosine phosphatases in the JAK/STAT pathway. Front Biosci. 2008; 13:4925–4932. [PubMed: 18508557]

Ye M, Coldren C, Liang X, et al. 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. [PubMed: 19942620]

Young JM, Endicott RM, Parghi SS, et al. Extensive copy-number variation of the human olfactory receptor gene family. Am J Hum Genet. 2008; 83:228–242. [PubMed: 18674749]

Yu H, Smallwood PM, Wang Y, et al. Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube closure: general implications for tissue fusion processes. Development. 2010; 137:3707–3717. [PubMed: 20940229]

Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature. 2013; 498:220–223. [PubMed: 23665959]

Zlotogora J, Schimmel MS, Glaser Y. Familial situs inversus and congenital heart defects. Am J Med Genet. 1987; 26:181–184. [PubMed: 3812559]

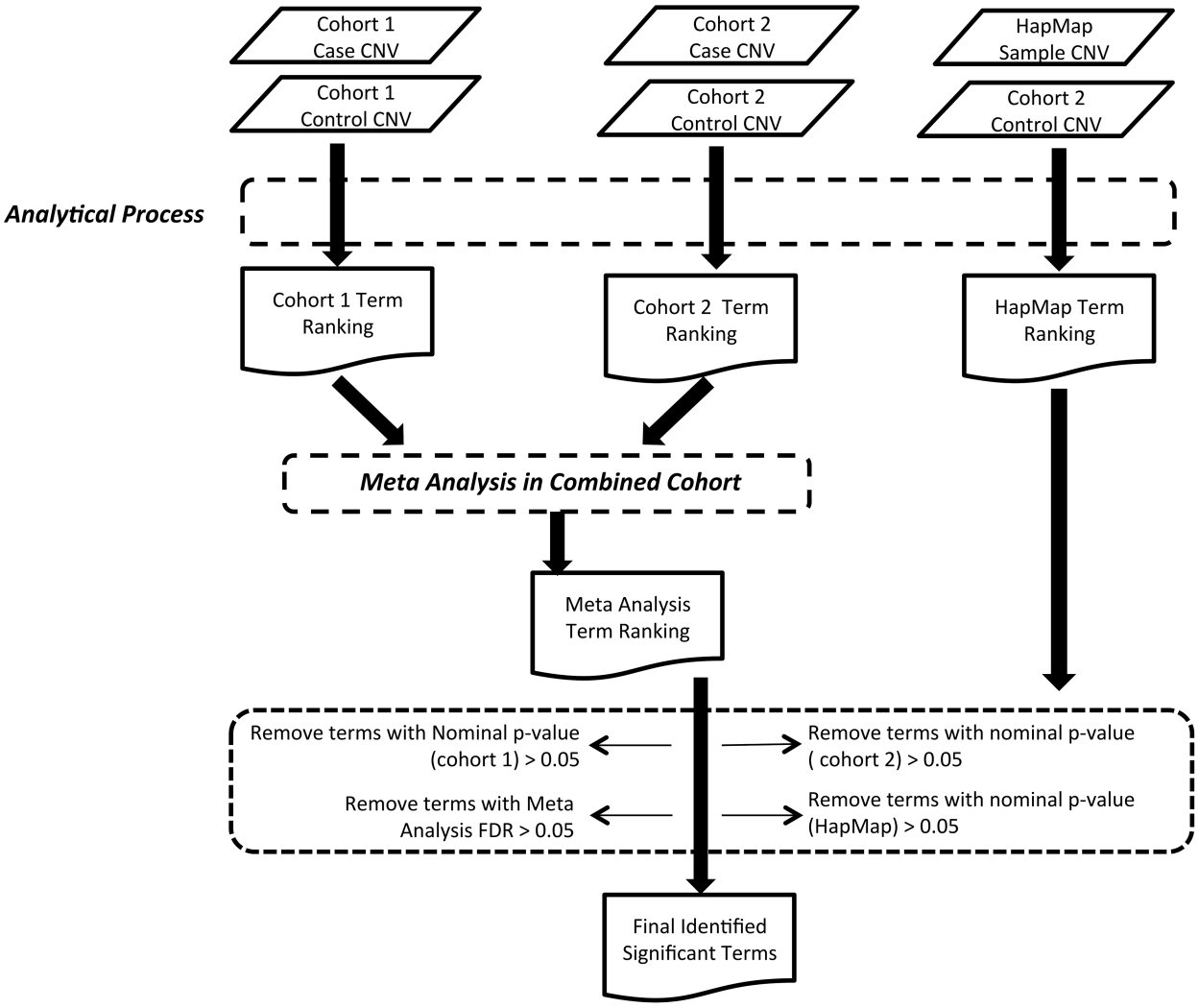

FIGURE 1. Flow chart outlining process of data analysis

For CNV detection workflow refer to White et al. (2014).

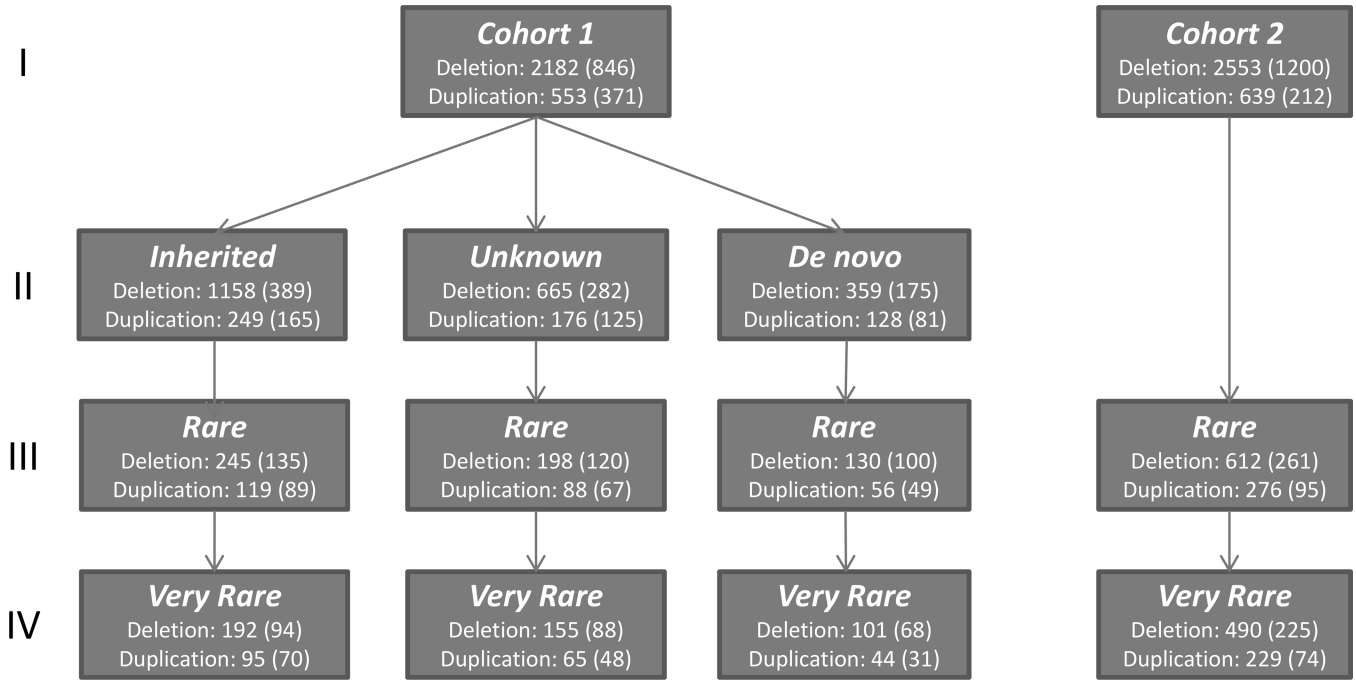

FIGURE 2. Flow chart depicting the distribution of CNVs in each cohort

The total count of CNVs and in parenthesis, the subset of CNVs containing genes, are presented. Row I reports all CNVs; Row II describes inheritance status for Cohort 1; Rows III and IV report the number of rare and very rare CNVs as defined in Methods, respectively.

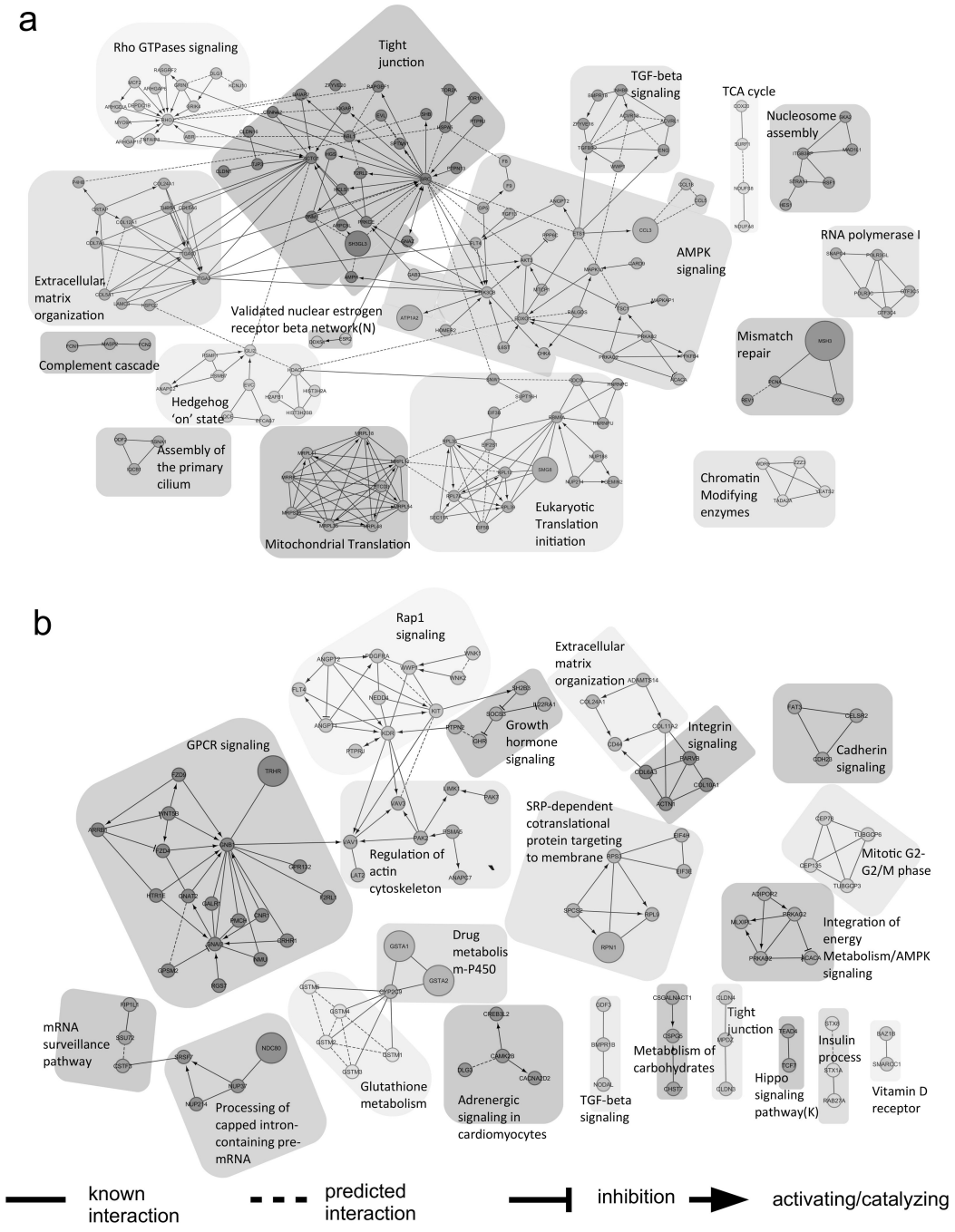

FIGURE 3. Gene interaction network clustering using ReactomeFIViz

Top function within each cluster is highlighted on top of each cluster. **a** The figure is constructed using genes from rare CNVs from Cohort 1. **b** The figure is constructed using genes from rare CNVs from Cohort 2. Each circle represents one unique Refseq gene with different shades representing different interaction network clusters identified from those genes. To simplify figure presentation, we annotated each module using its top enriched function or more abundant functional categories to illustrate each module's functional

characterization. Different connecting lines represent different biological events as illustrated in the legend in the figure.

TABLE 1

Phenotype distribution for both cohorts

	Count (%)	
Cardiac Lesion*	**Cohort 1**	**Cohort 2**
Tetralogy of Fallot	249 (39.7)	118 (34.1)
Pulmonary Stenosis	195 (78.3)	79 (66.9)
Pulmonary Atresia	41 (16.5)	27 (22.9)
Absent Pulmonary Valve	6 (2.4)	1 (0.8)
Unspecified Pulmonary Anatomy	7 (2.8)	11 (9.3)
Ventricular Septal Defect†	120 (19.1)	93 (26.9)
Conoventricular	101 (84.2)	72 (77.4)
Conal Septal Hypoplasia	5 (4.2)	4 (4.3)
Malalignment	14 (11.7)	15 (16.1)
Unspecified Type	0	2 (2.2)
D-Transposition of the Great Arteries	124 (19.8)	68 (19.6)
With Ventricular Septal Defect	61 (49.2)	30 (44.1)
Without Ventricular Septal Defect	60 (48.4)	33 (48.5)
Unspecified if Ventricular Septal Defect Present	3 (2.4)	5 (7.4)
Transposition of the Great Arteries - other/unknown ~	6 (1)	4 (1.2)
Double Outlet Right Ventricle^	68 (10.8)	19 (5.5)
Pulmonary Stenosis/Atresia	28 (41.2)	8 (42.1)
Aortic Stenosis/Atresia	9 (13.2)	1 (5.3)
Tricuspid Stenosis/Atresia	8 (11.8)	2 (10.5)
Mitral Stenosis/Atresia	26 (38.2)	5 (26.3)
Common Atrioventricular Valve	6 (8.8)	5 (26.3)
Single Ventricle (Double Inlet Right or Left Ventricle)	1 (1.5)	1 (5.3)
Isolated Aortic Arch Anomaly	29 (4.7)	18 (5.2)
Left Aortic Arch with Aberrant Right Subclavian Artery	1 (3.4)	4 (22.2)
Right Aortic Arch with Mirror Image Branching	3 (10.3)	2 (11.1)
Right Aortic Arch with Aberrant Left Subclavian Artery	9 (31.0)	7 (38.9)
Double Aortic Arch	16 (55.2)	5 (27.8)
Truncus Arteriosus	18 (2.9)	16 (4.6)
Type 1	8 (44.4)	11 (68.8)
Type 2	6 (33.3)	4 (25.0)
Type 3	1 (5.6)	0
Type 4	1 (5.6)	0
Type Unspecified	2 (11.1)	1 (6.3)
Interrupted Aortic Arch	12 (1.9)	8 (2.3)
Type A	3 (25.0)	1 (12.5)

	Count (%)	
Cardiac Lesion*	**Cohort 1**	**Cohort 2**
Type B	8 (66.7)	7 (87.5)
Type Unspecified	1 (8.3)	0
Other #	1 (0.1)	2 (0.6)
Total	627 (100)	346 (100)

*2.7% and 3.2% of the subjects were also diagnosed with heterotaxy in Cohort1 and Cohort 2, respectively.

†17.5% and 14% of the subjects were also diagnosed with coarctation of the aorta in Cohort 1 and Cohort 2, respectively; and 9.2% and 6.5% had concurrent muscular VSDs in Cohort 1 and Cohort 2, respectively.

~Cardiac segments SDL or unknown

^Subsets are not mutually exclusive.

#Single subjects with atrial septal defect and muscular VSD, muscular VSD, right ventricle aorta and pulmonary atresia.

TABLE 2

Rare CNV Burden Analysis

		All patients					CTD[#] Patients with no other anomalies					CTD[#] patients with additional anomalies				
	CNV type	**Count**	**CNV burden**	**Case/control CNV burden odds ratio**	**Significance (sample count based)***	**Significance (CNV count based)***	**Count**	**CNV burden**	**Case/control CNV burden odds ratio**	**Significance (sample count based)***	**Significance (CNV count based)***	**Count**	**CNV burden**	**Case/control CNV burden odds ratio**	**Significance (sample count based)***	**Significance (CNV count based)***
Cohort 1	*Duplications*	263	0.420	1.462	***1.13E-06***	***8.58E-10***	213	0.4235	1.4759	***7.98E-06***	***2.52E-08***	49	0.3984	1.3885	***3.05E-02***	***3.78E-03***
	Deletions	573	0.914	1.460	***1.53E-07***	***9.45E-22***	460	0.9145	1.4605	***4.23E-06***	***2.58E-19***	113	0.9187	1.4672	***3.01E-03***	***2.61E-05***
	All CNVs	836	1.333	1.460	***1.72E-09***	***1.89E-30***	673	1.338	1.4653	***1.06E-07***	***8.65E-27***	162	1.3171	1.4424	***1.99E-03***	***6.70E-07***
	Large CNVs															
	Duplications	75	0.120	1.498	***3.72E-03***	***2.01E-04***	60	0.1193	1.4936	**5.86E-03**	***7.27E-04***	15	0.122	1.527	2.23E-01	8.23E-02
	Deletions	32	0.051	1.690	***1.91E-02***	***9.55E-03***	22	0.0437	1.4482	1.01E-01	***3.02E-02***	10	0.0813	2.692	***1.44E-02***	9.12E-02
	All CNVs	107	0.171	1.551	***1.90E-04***	***3.94E-06***	82	0.163	1.4811	**1.40E-03**	***5.58E-05***	25	0.2033	1.8466	***2.21E-02***	***1.42E-02***
Cohort 2	*Duplications*	276	0.798	1.668	***3.03E-12***	***4.03E-47***	235	0.7993	1.6717	***2.67E-10***	***8.69E-42***	40	0.8	1.6731	***5.36E-04***	***5.00E-09***
	Deletions	612	1.769	1.847	***6.28E-11***	***7.10E-34***	526	1.7891	1.8677	***8.53E-10***	***9.60E-32***	82	1.64	1.7121	***1.82E-02***	***4.61E-05***
	All CNVs	888	2.567	1.787	***5.05E-15***	***1.10E-65***	761	2.5884	1.8025	***3.72E-12***	***3.14E-60***	122	2.44	1.6991	***1.71E-04***	***5.29E-10***
	Large CNVs															
	Duplications	61	0.176	1.675	***1.93E-04***	***2.47E-08***	50	0.1701	1.6161	***1.38E-03***	***1.03E-07***	11	0.22	2.0906	***1.29E-02***	***1.97E-02***
	Deletions	30	0.087	1.530	2.05E-01	6.36E-02	23	0.0782	1.3806	5.83E-01	2.40E-01	6	0.12	2.1177	5.59E-02	1.45E-01
	All CNVs	91	0.263	1.625	***1.47E-04***	***5.49E-09***	73	0.2483	1.5337	***1.71E-03***	***1.70E-07***	17	0.34	2.1001	***1.36E-02***	***3.13E-03***

[#]CTD: Conotruncal defect,

*Fisher Exact Test, two-side, bold type indicates significance

TABLE 3

Candidate Genes in Rare and Very Rare CNVs

Cyto band	ID	DX[#]	Hg19 Coordinate start	Hg19 Coordinate end	CNV Size (Kb)	CNV Type	CNV Inherited[‡]	Genes	Gene of Interest and supporting Information: Gene(s) of interest	Animal Model	Human Phenotype	Function and/or Expression	Frequency (type if different) of Gene in controls	Cohort	References citing genes disrupted by CNVs	Non-cardiac clinical features
Recurrent CNVs in our cases and reported literature																
1q21	2200	D-TGA	146501348	149160437~	2659	Del	N	*LOC728989, PDZK1P1, GJA5, BCL9 GPR89C, NBPF16, PPIAL4E, GJA8 MIR5087, NBPF24, ACP6, CHD1L NBPF15, PPIAL4D, PPIAL4A, LINC00624 LOC100130000, NBPF11, FMO5 LOC645166, PRKAB2, NBPF14, PDIA3P GPR89B, PPIAL4F, FLJ39739*	***GJA5****	(Gu et al., 2003)	MIM614049, (Christiansen et al., 2004)		1 (Dup)	Cohort 1	(Digilio et al., 2013; Erdogan et al., 2008; Glessner et al., 2014; Serra-Juhe et al., 2012; Silversides et al., 2012; Soemedi et al., 2012a,b; Tomita-Mitchell et al., 2012; Warburton et al., 2014)	Seizures
	3246	TA	146089254	147829352	1740	Del	(Y)	*CHD1L, PRKAB2, GJA8, PDIA3P, HYDIN2 LOC728989, GPR89B, ACP6, FMO5 GPR89C, MIR5087, LINC00624, NBPF11 PDZK1P1, BCL9, NBPF24, GJA5*	***GJA5****					Cohort 2	(Digilio et al., 2013; Erdogan et al., 2008; Glessner et al., 2014; Serra-Juhe et al., 2012; Silversides et al., 2012; Soemedi et al., 2012a,b; Tomita-Mitchell et al., 2012; Warburton et al., 2014)	Simple auricular helix; Extra nuchal skin; Anteriorly placed anus; Autism
	841	TOF	146762064	146791729	30	Del	Y	*CHD1L*					1 (Dup)	Cohort 1	(Greenway et al., 2009; Silversides et al., 2012;	None

Cyto band	ID	DX[#]	Hg19 Coordinate start	Hg19 Coordinate end	CNV Size (Kb)	CNV Type	CNV Inherited[‡]	Genes	Gene of Interest and supporting Information: Gene(s) of interest	Animal Model	Human Phenotype	Function and/or Expression	Frequency (type if different) of Gene in controls	Cohort	References citing genes disrupted by CNVs	Non-cardiac clinical features
															Soemedi et al., 2012b; Tomita-Mitchell et al., 2012)	
2q24.1	996	VSD	159,259,303	159,308,789	49	Del	NA	*CCDC148*					1	Cohort 2	Recurrent in (Silversides et al., 2012)	Clinodactyly
3q22.1	356	D-TGA	130,854,903	130,888,597	34	Del	NA	*NEK11*					0	Cohort 2	(Silversides et al., 2012; Tomita-Mitchell et al., 2012)	Pyloric stenosis
3q22.1	3193	TA	132324055	132712925	389	Del	(Y)	*NPHP3-ACAD11, ACAD11, NPHP3-AS1 NPHP3, UBA5*	***NPHP3****	(Bergmann et al., 2008; Hoff et al., 2013)	(Bergmann et al., 2008; Chaki et al., 2011; Tory et al., 2009)		0	Cohort 2	(Tomita-Mitchell et al., 2012)	None
3q22.3	2360	TOF	138380656	138517384	137	Del	N	*PIK3CB*	***PIK3CB****			(Chang et al., 2011; Chang et al., 2010; Rigor et al., 2009)	0	Cohort 1	(Erdogan et al., 2008; Tomita-Mitchell et al., 2012)	None
3q29	157	TA-2	196,368,501	196,482,211	114	Dup	NA	*LRRC33, PIGX, PAK2, CEP19*				PAK expressed in fetal heart based on (Soemedi et al., 2012b)	*PAK2*: 2 Others 0	Cohort 2	(Soemedi et al., 2012b)	Coloboma; Simple helix; Hypocalcemia with stress
3q29	2495	TOF	198,278,255	198,502,522	224	Del	N	*DLG1*				Expressed in heart based on (Soemedi et al., 2012b)	1 (Dup)	Cohort 1	(Soemedi et al., 2012b)	None
4q12	645	AAA	54133328	57134389	3001	Dup	NA	*FIP1L1, GSX2, KIT, LNX1-AS2, CHIC2 PDGFRA, KIAA1211, SRD5A3*	**PDGFRA***	(Bleyl et al., 2010;			*SCFD2*: 1 (Del) Others: 0	Cohort 2	(Breckpot et al., 2011)	None

Cyto band	ID	DX#	Hg19 Coordinate start	Hg19 Coordinate end	CNV Size (Kb)	CNV Type	CNV Inherited‡	Genes	Gene of Interest and supporting Information: Gene(s) of interest	Animal Model	Human Phenotype	Function and/or Expression	Frequency (type if different) of Gene in controls	Cohort	References citing genes disrupted by CNVs	Non-cardiac clinical features
								LOC644145, CLOCK, LNX1-AS1, NMU CEP135, RPL21P44, SCFD2, EXOC1 LNX1, SRD5A3-AS1, TMEM165, KDR PDCL2		Hamilton et al., 2003; Soriano, 1997)					
5p15.2	939	IAA	9,307,909	9,340,073	32	Del	Y	*SEMA5A*				(Tomita-Mitchell et al., 2012)	1	Cohort 1	(Tomita-Mitchell et al., 2012)	None
5q14.1	2914	TOF	80,665,128	80,772,968	108	Dup	NA	*SSBP2, ACOT12*				(Soemedi et al., 2012b)	1 (Del)	Cohort 1	(Silversides et al., 2012; Soemedi et al., 2012b)	None
	2437	VSD	80,863,776	80,918,866	55	Dup	Y	*SSBP2*				(Soemedi et al., 2012b)	1 (Del)	Cohort 1	(Silversides et al., 2012; Soemedi et al., 2012b)	Hypospadias; Left frontal subependymal hemorrhage
5q35.3	2503	TOF	179985934	180067675	82	Del	Y	*SCGB3A1, CNOT6, FLT4*	***FLT4****	(Dumont et al., 1998)			0	Cohort 1	(Serra-Juhe et al., 2012; Soemedi et al., 2012b) (CNOT6 recurrent in (Soemedi et al., 2012b))	Speech delay
	3227	AAA	180,043,388	180,074,248	31	Dup	NA	*SCGB3A1, CNOT6, FLT4*	***FLT4****	(Dumont et al., 1998)			0	Cohort 2	(Serra-Juhe et al., 2012; Soemedi et al., 2012b) (CNOT6 recurrent in (Soemedi et al., 2012b))	Long philtrum; Tented upper lip
8p23.1	955	TOF	6,341,567	6,401,499	60	Del	NA	*ANGPT2, MCPH1*	***ANGPT2***				*ANGPT2*: 0 *MCPH1*: 3	Cohort 1	(Silversides et al., 2012)	Asymmetric crying facies: Pyloric stenosis;

Cyto band	ID	DX[#]	Hg19 Coordinate start	Hg19 Coordinate end	CNV Size (Kb)	CNV Type	CNV Inherited[‡]	Genes	Gene of Interest and supporting Information: Gene(s) of interest	Animal Model	Human Phenotype	Function and/or Expression	Frequency (type if different) of Gene in controls	Cohort	References citing genes disrupted by CNVs	Non-cardiac clinical features
	222	TOF/PS	6,354,159	6,414,091	60	Del	NA	*ANGPT2, MCPH1*					*ANGPT2*: 0 *MCPH1*: 3	Cohort 1	(Silversides et al., 2012)	None
8p23.1	686	D-TGA	10,263,765	10,510,048	246	Del	Y	*RP1L1, PRSS55, MSRA*					*RP1L1*: 0	Cohort 1	(Silversides et al., 2012; Tomita-Mitchell et al., 2012; Glessner et al., 2014)	Incomplete vertebral fusion at the L5 and S1 levels
	710	TOF	10,497,783	10,534,893	37	Del	Y	*RP1L1*						Cohort 1	(Silversides et al., 2012; Tomita-Mitchell et al., 2012; Glessner et al., 2014)	Clinodactyly; Cafeau lait spots
10q22.1	2955	D-TGA	72129976	73296259	1166	Dup	(Y)	*PRF1, PALD1, SLC29A3, CDH23 LOC728978, ADAMTS14, SGPL1, LRRC20 UNC5B, PCBD1, EIF4EBP2, TBATA NODAL*	***NODAL****	(Conlon et al., 1994; Lowe et al., 2001)	(MIM270100) (Mohapatra et al., 2009; Zlotogora et al., 1987)		0	Cohort 2	{Warburton et al., 2014)	Bipolar disorder
11q23.3	1046	VSD	128383077	128395628	13	Del	NA	*ETS1*	***ETS1****	(Gao et al., 2010; Ye et al., 2010)	OMIM (147791)	(Gao et al., 2010; Schachterle et al., 2012)	0	Cohort 1	(Glessner et al., 2014)	None
Rare or very rare CNVs containing candidate genes not previously reported																
3q22	614	VSD	30567666	31089530	522	Dup	NA**	*TGFBR2, GADL1*	***TGFBR2****	(Choudhary et al., 2006)			0	Cohort 1		None
3p12.3	2314	D-TGA	81698119	81723365	25	Del	NA	*GBE1*	***GBE1****	(Lee et al., 2011)			1 (Dup)	Cohort 2		None
3q28-q29	2338	D-TGA	193423253	194586088	1163	Dup	Y	*TMEM44, DPPA2P3, GP5, LRRC15 TMEM44-AS1, CPN2, LOC100507391 LSG1, FLJ34208, FAM43A*	***HES1****	(Rochais et al., 2009)			0	Cohort 1		None

Cyto band	ID	DX#	Hg19 Coordinate start	Hg19 Coordinate end	CNV Size (Kb)	CNV Type	CNV Inherited‡	Genes	Gene of Interest and supporting Information: Gene(s) of interest	Animal Model	Human Phenotype	Function and/or Expression	Frequency (type if different) of Gene in controls	Cohort	References citing genes disrupted by CNVs	Non-cardiac clinical features
								LOC100131551, HES1, ATP13A3 LOC647323								
4p16	2837	DORV	5716922	5840210	123	Dup	NA	*CRMP1, EVC*	***EVC****		OMIM: 604831 (Ellis-van Creveld Syndrome), (Ruiz-Perez et al., 2000)		1 (Del)	Cohort 1		None
7p22.2	2359	TOF	1252339	3090282	1838	Dup	NA	*LFNG*	***LFNG***			(Kume et al., 2001)	0	Cohort 1		Vesicoureteral reflux; Hypertonia; Gross motor delay
8q23	455	VSD	106296518	114763029	8467	Dup	NA	*ENY2, MIR2053, TRHR, EBAG9, KCNV1 SYBU, PKHD1L1, TMEM74, CSMD3 OXR1, EIF3E, ZFPM2, RSPO2, EMC2 ABRA, NUDCD1, ANGPT1*	***ZFPM2****	(Ackerman et al., 2005)(	MIM187500, (Pizzuti et al., 2003)		0	Cohort 2		Hooded eyes, Epicanthal folds
9p24.2	310	TA	46587	13335127	13289	Del	NA	*INSL4, KIAA1432, RANBP6, C9orf123 ERMP1, SMARCA2, GLDC, C9orf66 GLIS3-AS1, LURAP1L, CD274, KANK1 KDM4C, KIAA0020, MPDZ, PTPRD SLC1A1, DMRT2, INSL6, MIR101-2 MIR4665, PLGRKT, JAK2, KIAA2026 GLIS3, CDC37L1, FOXD4, UHRF2 DMRT3, DOCK8, KCNV2, PDCD1LG2 TPD52L3, FLJ35024, IL33, RLN1, RLN2 SPATA6L, TYRP1, CBWD1, MLANA, AK3 DMRT1, VLDLR, PPAPDC2, RCL1, RFX3*	***RFX3***	(MGI:5560494, Lo, C direct data submission) (Bonnafe et al., 2004; Li et al., 2015)			0	Cohort 2		None
10q21.1	3340	TOF	60471279	60532095	61	Dup	(Y)	*FAM133CP, BICC1*	***BICC1****	MGI:5285079,			0	Cohort 2		None

Cyto band	ID	DX[#]	Hg19 Coordinate start	Hg19 Coordinate end	CNV Size (Kb)	CNV Type	CNV Inherited[‡]	Genes	Gene of Interest and supporting Information: Gene(s) of interest	Animal Model	Human Phenotype	Function and/or Expression	Frequency (type if different) of Gene in controls	Cohort	References citing genes disrupted by CNVs	Non-cardiac clinical features
										(Li et al., 2015)						
11q24.1-q24.3	652	DORV	128582557	128644407	62	Dup	Y	*FLI1*[*]	***FLI1***[*]			(Schachterle et al., 2012)	0	Cohort 1		Abdominal hernia (NOS); Large first toe
12q24.11	3214	TOF	110743687	110888923	145	Dup	NA	*ATP2A2, ARPC3, ANAPC7*	***ATP2A2***	(Ver Heyen et al., 2001)			0	Cohort 2		None
13q13.3	2678	TOF	38118531	38269676	151	Del	Y	*TRPC4, LINC00547, POSTN*	***POSTN***[*]	(Rios et al., 2005)		(Norris et al., 2008)	0	Cohort 1		Hypospadias; Bifid scrotum
13q14.1	466	TOF	41039186	41270080	231	Del	Y	*FOXO1, LINC00598*	***FOXO1***[*]	(Banks et al., 2011; Sengupta et al., 2012)			0	Cohort 1		Small palpebral fissures; Broad nasal root; Wide spaced nipples; Jaundice – unconjugated; Seizures
14q22-q24	198	D-TGA	69188421	69452088	264	Dup	NA	*ACTN1, ACTN1-AS1, ZFP36L1*	***ZFP36L1***	(Bell et al., 2006; Stumpo et al., 2004)			0	Cohort 2		Learning disabilities
14q24.3	2570	TOF	77887681	82588573	4701	Del	N	*ALKBH1, NRXN3, GTF2A1, TSHR, AHSA1 SNW1, ISM2, VIPAS39, DIO2, DIO2-AS1 SLIRP, CEP128, ADCK1, STON2 C14orf178, SNORA79, NOXRED1, SEL1L SPTLC2*	***SNW1 (AHSA1***[*]***)***			(Fryer et al., 2004; Oswald et al., 2002)	0	Cohort 1		Scoliosis; Stroke; Global delay, Speech delay
17p11.2	2427	TOF	12575850	12647490	72	Del	NA	*MYOCD*[*]	***MYOCD***[*]	(Huang et al., 2008; Huang et al., 2012; Huang et al., 2015)			0	Cohort 2		Normal
Xq26.2	2109	VSD	136496056	144261673	7766	Del	NA	*LDOC1, MAGEC1, MAGEC2, SPANXD SPANXE, ATP11C, CXorf66, MIR504*	***ZIC3***[*]	(Purandare et al., 2002; Ware et al., 2006)	OMIM: 306955		1 (Dup)	Cohort 1		Down-slanting palpebral fissures; Bulbous nasal tip; down-turned

Cyto band	ID	DX[#]	Hg19 Coordinate start	Hg19 Coordinate end	CNV Size (Kb)	CNV Type	CNV Inherited[‡]	Genes	Gene of Interest and supporting Information Gene(s) of interest	Animal Model	Human Phenotype	Function and/or Expression	Frequency (type if different) of Gene in controls	Cohort	References citing genes disrupted by CNVs	Non-cardiac clinical features
								SPANXN4, ZIC3, MAGEC3, SPANXA1 SPANXA2-OT1, SPANXN3, CDR1 SPANXB1, F9, FGF13-AS1, SLITRK4 SPANXB2, LOC158696, MCF2 SRD5A1P1, LOC389895, RP1-177G6.2 UBE2NL, SOX3, SPANXA2, SPANXC SPANXF1, SPANXN2, FGF13, MIR320D2 MIR505							corners of mouth; TE fistula; Hydronephrosis, Calectasis; Pelviectasis;

[#]D-TGA: Dextro Transposition of the Great Arteries, TA: Truncus Arteriosus, TOF: Tetralogy of Fallot; VSD: Ventricular Septal Defect; AAA: Aortic Arch Anomaly; IAA: Interrupted Aortic Arch; DORV: Double Outlet Right Ventricle; PS: Pulmonic Stenosis.

[~]Adjusted for hg19

[‡]Parenthesis indicate that inheritance has been determined by qPCR,

[*]CNVs were validated by qPCR, Genes without asterisk were not tested

TABLE 4

Enriched Gene Ontology Related to Cardiac Development*

Term	Description	GO type	CNV type	Cohort 1			Cohort 2			Total		References citing full pathway relevant to cardiac development or component members
				Case	Control	p-value	Case	Control	p-value	p-value	FDR	
GO:0051090	Regulation of sequence-specific DNA binding transcription factor activity	Biological process	dup	11	23	3.64E-02	15	16	1.80E-05	6.45E-06	2.38E-04	*HAND2* (Vincentz et al., 2011), *SHH* (Washington Smoak et al., 2005), *MYOCD* (Huang et al., 2012), *FZD2* (van Gijn et al., 2001; Yu et al., 2010) *NODAL* (Koefoed et al., 2014) *TGFB1* (Engelmann et al., 1992)
GO:0055007	Cardiac muscle cell differentiation	Biological process	dup	2	0	3.02E-02	6	2	3.08E-04	1.98E-05	6.17E-04	*MYOCD* (see above), *NKX2.5* (*McElhinney et al., 2003*)*GATA4* (Tomita-Mitchell et al., 2007), *MEF2C* (Dodou et al., 2004), *HEY2* (Donovan et al., 2002; Sakata et al., 2006)
GO:0031513	Nonmotile primary cilium	Cellular component	all	9	13	7.95E-03	7	12	0.0204	5.42E-04	8.48E-03	(Li et al., 2015; Tobin and Beales, 2009)
GO:0035924	Cellular response to vascular endothelial growth factor stimulus	Biological process	all	3	2	3.97E-02	4	2	0.00695	6.42E-04	9.84E-03	(Dor et al., 2001)
GO:2000018	Regulation of male gonad development	Biological process	all	2	0	3.02E-02	3	1	0.0137	6.77E-04	1.02E-02	*CITED2* (Weninger et al., 2005), *ZFPM2* (Pizzuti et al., 2003), *SOX9* (Lincoln et al., 2007)
GO:0051276	Chromosome organization	Biological process	dup	18	49	4.93E-02	16	36	0.0058	1.12E-03	1.90E-02	(Zaidi et al., 2013)(Silversides et al., 2012)
GO:0046425	Regulation of JAK-STAT cascade	Biological process	dup	4	3	2.03E-02	5	7	0.0286	1.87E-03	2.90E-02	(Jorge et al., 2009; Tartaglia and Gelb, 2005; Xu and Qu, 2008)
GO:0017015	Regulation of transforming growth factor beta receptor signaling pathway	Biological process	dup	4	4	3.52E-02	4	4	0.0249	2.19E-03	3.26E-02	(Arthur and Bamforth, 2011; Ma et al., 2016)
GO:0032321	Positive regulation of Rho GTPase activity	Biological process	dup	6	10	4.50E-02	4	3	0.0142	2.34E-03	3.46E-02	(Wei et al., 2002)
GO:2001236	Regulation of extrinsic apoptotic signaling pathway	Biological process	dup	8	13	1.95E-02	6	10	0.0287	2.80E-03	3.98E-02	(Fisher et al., 2000; Lu et al., 2011; Poelmann and Gittenberger-de Groot, 2005; Rezvani et al., 2000)

Term	Description	GO type	CNV type	Cohort 1			Cohort 2			Total		References citing full pathway relevant to cardiac development or component members
				Case	Control	p-value	Case	Control	p-value	p-value	FDR	
GO:0051865	Protein autoubiquitination	Biological process	dup	4	4	3.52E-02	4	5	0.0394	3.54E-03	4.71E-02	(Fouladkou et al., 2010)
GO:0090092	Regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	Biological process	dup	6	10	4.50E-02	5	7	0.0286	3.74E-03	4.94E-02	(Gaspar et al., 2003; Nebigil et al., 2000; Nebigil and Maroteaux, 2003)

* Full list in Table S6.

TABLE 5

Top Ten of Enriched Mammalian Phenotype Terms *

Mammalian Phenotype ID	Description	CNV Type	Cohort 1			Cohort 2			Total	
			Case	Control	p-value	Case	Control	p-value	p-value	FDR
MP:0005621	Abnormal Cell Physiology	Dup	49	166	4.06E-02	65	129	9.08E-11	5.48E-09	3.30E-06
MP:0002080	Prenatal Lethality	Dup	40	129	3.69E-02	57	106	3.60E-10	5.64E-09	3.30E-06
MP:0005397	Hematopoietic System Phenotype	Dup	48	149	1.16E-02	50	117	1.30E-06	4.51E-07	1.12E-04
MP:0008247	Abnormal Mononuclear Cell Morphology	Dup	38	100	2.66E-03	33	71	2.53E-05	6.20E-07	1.38E-04
MP:0002619	Abnormal Lymphocyte Morphology	Dup	37	86	3.91E-04	26	60	6.85E-04	1.04E-06	1.98E-04
MP:0011102	Partial Embryonic Lethality	All	5	2	2.41E-03	7	5	8.71E-04	7.33E-06	2.91E-04
MP:0003945	Abnormal Lymphocyte Physiology	Dup	23	63	2.94E-02	27	44	2.95E-06	2.21E-06	3.33E-04
MP:0002221	Abnormal Lymph Organ Size	Dup	25	69	2.61E-02	26	42	4.03E-06	2.96E-06	3.82E-04
MP:0001648	Abnormal Apoptosis	Dup	28	81	2.81E-02	31	59	8.85E-06	4.21E-06	5.12E-04
MP:0002722	Abnormal Immune System Organ Morphology	Dup	29	86	3.26E-02	30	55	5.33E-06	4.66E-06	5.57E-04

* Full list in Table S7

TABLE 6

Significantly Enriched KEGG terms

KEGG term	CNV Type	Cohort 1			Cohort 2			Total	
		Case	Control	p-value	Case	Control	p-value	p-value	FDR
Adherens Junction	all	13	28	2.150E-02	11	23	1.460E-02	9.491E-04	1.261E-02
TGF-beta Signaling Pathway	all	6	9	3.280E-02	7	10	1.050E-02	1.148E-03	1.305E-02
Drug Metabolism – Cytochrome P450	del	6	6	9.660E-03	9	6	1.110E-04	4.890E-06	8.998E-04
Metabolism of Xenobiotics by Cytochrome P450	del	5	6	2.880E-02	8	4	9.640E-05	1.670E-05	1.536E-03