

ORIGINAL ARTICLE



# Integration of Large-Scale Genomic Data Sources With Evolutionary History Reveals Novel Genetic Loci for Congenital Heart Disease

**BACKGROUND:** Most cases of congenital heart disease (CHD) are sporadic and nonsyndromic, with poorly understood etiology. Rare genetic variants have been found to affect the risk of sporadic, nonsyndromic CHD, but individual studies to date are of only moderate sizes, and none to date has incorporated the ohnolog status of candidate genes in the analysis. Ohnologs are genes retained from ancestral whole-genome duplications during evolution; multiple lines of evidence suggest ohnologs are overrepresented among dosage-sensitive genes. We integrated large-scale data on rare variants with evolutionary information on ohnolog status to identify novel genetic loci predisposing to CHD.

**METHODS:** We compared copy number variants present in 4634 nonsyndromic CHD cases derived from publicly available data resources and the literature, and >27 000 healthy individuals. We analyzed deletions and duplications independently and identified copy number variant regions exclusive to cases. These data were integrated with whole-exome sequencing data from 829 sporadic, nonsyndromic patients with Tetralogy of Fallot. We placed our findings in an evolutionary context by comparing the proportion of vertebrate ohnologs in CHD cases and controls.

**RESULTS:** Novel genetic loci in CHD cases were significantly enriched for ohnologs compared with the genome ( $\chi^2$  test,  $P < 0.0001$ , OR = 1.253 [95% CI, 1.199–1.309]). We identified 54 novel candidate protein-coding genes supported by both: (1) copy number variant and whole-exome sequencing data; and (2) ohnolog status.

**CONCLUSIONS:** We have identified new CHD candidate loci, and show for the first time that ohnologs are overrepresented among CHD genes. Incorporation of evolutionary metrics may be useful in refining candidate genes emerging from large-scale genetic evaluations of CHD.

Elisavet Fotiou, MRes  
Simon Williams, PhD  
Alexandra Martin-Geary,  
MSc  
David L. Robertson, PhD  
Gennadiy Tenin, PhD  
Kathryn E. Hentges, PhD  
Bernard Keavney, MD

**Key Words:** copy number variants, DNA ■ evolution ■ heart defects, congenital ■ whole exome sequencing

© 2019 The Authors. *Circulation: Genomic and Precision Medicine* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited.

<https://www.ahajournals.org/journal/circgen>

Congenital heart disease (CHD) is the most prevalent birth defect in humans, occurring in  $\approx 8$  per 1000 live births, and consisting of malformation of the heart and the great vessels.<sup>1</sup> Around 20% of all CHDs can be attributed to chromosomal imbalances such as Down and Turner, and 22q11 deletion syndromes; around 80% occur as sporadic, nonsyndromic CHD. In such cases, CHD behaves overall as a genetically complex trait with moderate heritability. Previous genome-wide investigations into CHD have found evidence for rare causative copy number variants (CNVs) and single nucleotide variants (SNVs), and associations with common SNVs in GWAS.<sup>2-6</sup> It has been estimated in previous studies that several hundred genes may be involved in polygenic CHD susceptibility; therefore, many remain to be discovered.<sup>7</sup>

CNVs are 1 kilobase (kb) to several megabase (Mb) sized regions of duplication and deletion in the genome. A 2014 meta-analysis of CNVs in 1694 nonsyndromic CHD cases identified 79 chromosomal regions in which 5 or more CHD cases had overlapping imbalances.<sup>5</sup> The estimated prevalence of pathogenic CNVs in nonsyndromic CHD patients is 4% to 14%, whereas in syndromic CHD patients it is 15% to 20% (the most common being 22q11 deletion syndrome).<sup>3,8,9</sup> There are multiple mechanisms by which a CNV may lead to disease including the disruption of chromosome structure, alteration of gene expression due to disruption of regulatory elements, and changes of the relative amounts of dosage-sensitive genes.<sup>10</sup>

The dosage-balance model postulates that, for genes that are in stoichiometric relationships (eg, forming protein complexes with other genes), any perturbation in their relative ratios will tend to be deleterious.<sup>10</sup> In the early course of vertebrate evolution, around 500 million years ago, 2 whole-genome duplications, during which gene stoichiometry throughout the genome was preserved, as all genes were duplicated, took place. Periods of gene loss followed each of these events, resulting in the retention of some whole-genome duplication paralogs in the genome (termed ohnologs) and the loss of others. The dosage-balance model would predict that ohnologs should be enriched for dosage-sensitive genes.<sup>11</sup> Ohnologs, of which there are around 7000 in the human genome, have indeed been shown to exhibit characteristics consistent with dosage-sensitivity: for example, ohnologs are enriched for haploinsufficient genes<sup>11,12</sup>; and Makino et al<sup>13</sup> reported, based on CNV data in healthy individuals from the Database of Genomic Variants, that genomic regions ( $\approx 2$  Mb in size) near ohnologs are CNV deserts, indicating that those regions are dosage-balanced.

The formation and fixation of gene duplications within the genome are subject to different evolutionary mechanisms—small scale duplications (SSD) involving relatively few genes, and whole-genome duplication. A strong relationship between the evolutionary mechanism of duplication and phenotypic consequences, including

heritable diseases, has been previously shown.<sup>14-16</sup> Ohnologs have a significant association with certain human genetic diseases; for example 12 out of 16 reported candidate genes within the Down syndrome critical region (21q22.12, 21q22.13, and 21q22.2) are dosage-balanced ohnologs.<sup>11</sup> By contrast, genes arising from SSDs lack enrichment for disease association.<sup>17</sup> In addition, ohnologs are enriched for genes involved in signaling and gene regulation, key cardiovascular developmental processes.<sup>11</sup> These considerations led us to hypothesize that ohnologs may be enriched among CHD causative genes.

We tested this hypothesis in a meta-analysis of CNV data including 4634 nonsyndromic CHD cases, and integrated these data with a WES (whole-exome sequencing) study of 829 cases of Tetralogy of Fallot (TOF), the commonest cyanotic CHD phenotype, which has been previously shown to have a significant etiological contribution from CNVs.<sup>6</sup> These were compared with control data, which were derived from large-scale genomic resources.<sup>18-21</sup>

## METHODS

The appropriate institutional review bodies approved all recruitment of human participants in this study. The study corresponded with the stipulations of the Declaration of Helsinki, and all participants (or their parents, if affected probands were children too young to themselves consent) provided informed consent. Data from consortia were accessed subject to the applicable data-sharing agreements. Summary data, analytic methods, and summary study materials will be made available to other researchers for purposes of reproducing the results or replicating the analyses reported here, on request to the corresponding authors. Full Materials and Methods are available in the [Data Supplement](#) of the article.

## RESULTS

### Update of CHD CNV Data Set and Generation of Control CNV Data Set

We updated the previous meta-analysis of CNVs in nonsyndromic CHD cases,<sup>5</sup> in a further 2882 nonsyndromic CHD cases from DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources), ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Abberations), ISCA (International Standards for Cytogenomic Arrays) databases and further published studies investigating the role of CNVs in CHD.<sup>4,20,22-32</sup> The updated CHD case CNV data set consists of 4634 unrelated individuals of different ancestries (Table 1). The outline workflow to identify candidate genes is shown in Figure 1. Filtering of the CHD case population against DECIPHER known microdeletion/microduplication syndromes resulted in 224 cases being removed; this left

**Table 1.** Number of Cases in Previous and Current Meta-Analysis Studies as Well as Controls Used in the Current Study

Databases for CHD Cases	Thorsson et al <sup>5</sup> Study	Current Study	Databases for Controls	Current Study
DECIPHER	279	1252	1000 Genome phase 3	2504
ISCA	331	1107	DGV	>6430
Published literature	814	1900	gnomAD	10738
CHDwiki	328	328	Published literature	356
ECARUCA	0	47	WTCC2	≈6000
WES TOF (Page et al <sup>6</sup> )	N/A	829	DDD	845

CHD indicates congenital heart disease; DDD, Deciphering Developmental Disorders study;

DECIPHER, Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources; DGV, Database of Genomic Variants; ECARUCA, European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations; ISCA, International Standards for Cytogenomic Arrays; WES TOF, whole-exome sequencing of Tetralogy of Fallot; and WTCC2, Wellcome Trust Case Control Consortium 2.

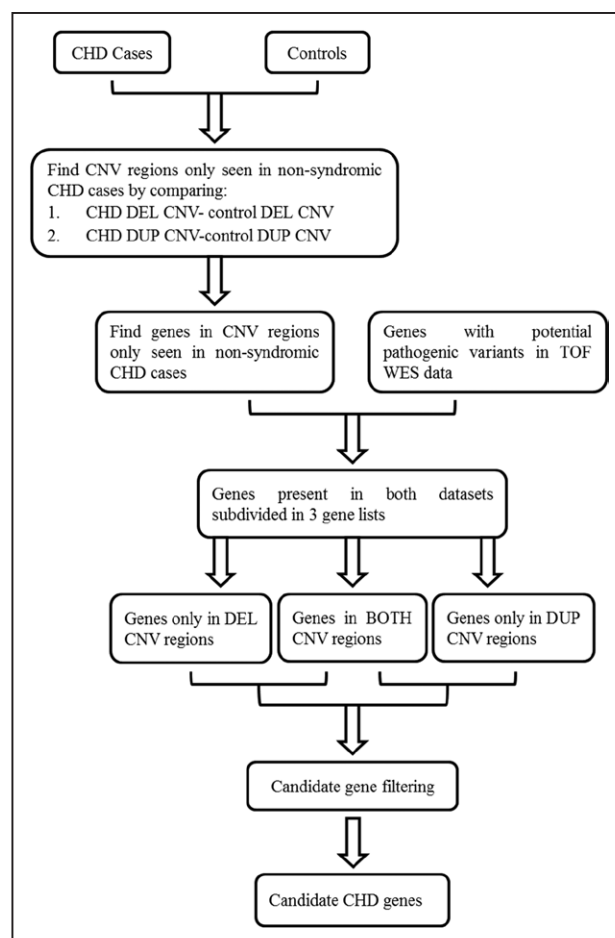
4410 CHD cases with 3362 deletion CNVs and 2540 duplication CNVs which were used for further analysis (Figure I in the Data Supplement).

A control CNV data set was generated by acquiring CNV data from individuals not explicitly identified as having a developmental disorder, who were enrolled in the 1000 Genome Project Phase 3, Database of Genomic Variants, DECIPHER, and published studies.<sup>21,27,28,33,34</sup> The control CNV data set resulted in 256511 deletion CNVs, 84343 duplication CNVs, and 6403 BOTH CNVs, that is, either deletion or duplication. gnomAD CNVs<sup>35</sup> were incorporated into the analysis as they became available and resulted in an additional 51420 duplication CNVs and 198611 deletion CNVs.

## Comparison of CHD CNV Regions With Control CNV Regions

All CHD deletion and duplication CNV regions (coordinates hg19) were compared against control deletion and duplication CNV regions, respectively. Any CHD CNV regions overlapping control CNV regions were excluded. As a result, we identified deletion and duplication CNV regions only seen in nonsyndromic CHD cases. The genes located in those regions were annotated using the Ensembl database. There were a number of genes that already had an assigned phenotype (Online Mendelian Inheritance in Man)<sup>36</sup>; among these, 59 had been previously associated with CHD pathogenesis such as *ZIC3*, *NKX2-6*, *GATA4*, *JAG1*, *GJA1*, and *TBX5*. All genes with Online Mendelian Inheritance in Man assigned phenotypes were excluded from further analysis.

Novel genes found in the CNV regions only seen in CHD cases were then compared with an in-house list of 12771 genes with novel or rare SNVs (either absent

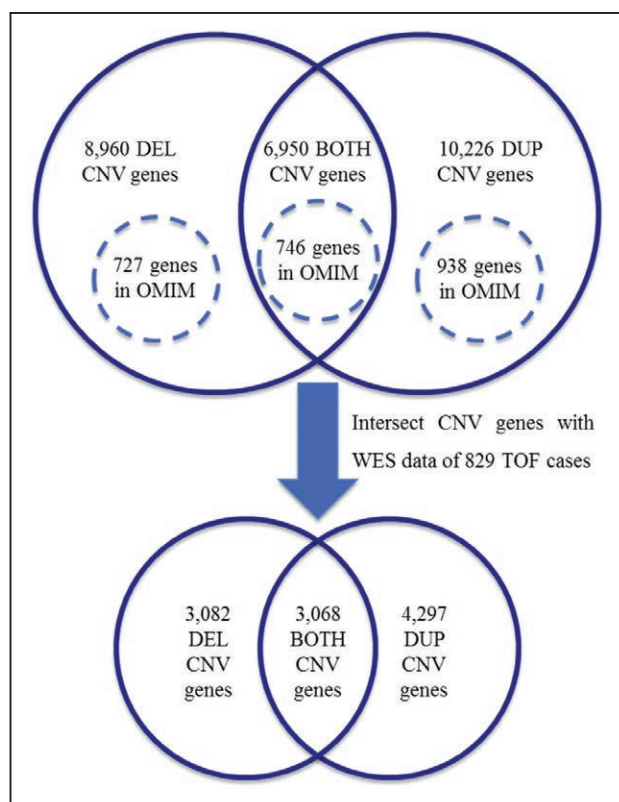
**Figure 1.** Overall methodology.

Flowchart showing the methodology used to identify novel genetic loci for nonsyndromic congenital heart disease (CHD) cases. Potential pathogenic variants were novel or rare single nucleotide variants (either absent from ExAC or with frequency of <0.01). Candidate genes identified at the end of the workflow were subsequently analyzed for ohnolog status. BOTH indicates deleted/duplicated; CNV, copy number variant; DEL, deletion; DUP, duplication; TOF, Tetralogy of Fallot; and WES, whole-exome sequencing.

from ExAC or with frequency of <0.01) from WES data in 829 TOF cases.<sup>6</sup> Genes supported by both CNV and WES data were included for further analysis. In total, 3082 genes in deletion CNVs, 4297 genes in duplication CNVs, and 3068 in BOTH CNVs (ie, genes found in deletion and duplication CNVs) were also found in the TOF WES data with either high (nonsense variants, frameshift, splice variants) or medium (missense, splice variants) impact SNVs. This intersection of CNV and WES data led to an overall reduction of ≈60% in the number of candidate genes for CHD (Figure 2).

## Ohnologs are Highly Enriched in CHD Cases Whereas SSD and Singleton Genes are Not

Ohnologs (N=7023) were identified using data from Singh et al,<sup>37</sup> available at <http://ohnologs.curie.fr/>. SSDs (N=7014) were extracted from Ensembl gene trees.<sup>12</sup> Any



**Figure 2. Intersection of copy number variant (CNV) and whole-exome sequencing (WES) data.**

Numbers of genes involved in the final stages of the workflow depicted in Figure 1 are shown. Genes with assigned phenotypes (circles with dashed line) were excluded from further analysis. BOTH indicates deleted/duplicated; DEL, deletion; DUP, duplication; OMIM, Online Mendelian Inheritance in Man; and TOF, Tetralogy of Fallot.

remaining genes that were neither found in the ohnolog data set nor identified as having a direct paralog were considered for the purpose of this study to be singletons. The frequencies of ohnologs, SSDs, and singletons among the candidate CHD genes were compared with their frequency in the human genome. Novel genes supported by the CNV data in CHD cases were found to be enriched for ohnologs (14.65% versus 12.05%,  $\chi^2$  test,  $P < 0.0001$ ; OR=1.253 [95% CI, 1.199–1.309]; Figure 3A). There were no differences in SSDs (Figure 3B) and an under-representation for singletons (Figure 3C) compared with the human genome. There was a 2.3-fold increased enrichment of ohnologs in the genes supported by both CNV and WES data in CHD cases ( $\chi^2$  test,  $P < 0.0001$ , OR=3.751 [95% CI, 3.574–3.937]). In this instance, SSDs were also enriched in CHD cases compared with the human genome ( $\chi^2$  test,  $P < 0.0001$ ; OR=1.437 [95% CI, 1.356–1.905]). However, ohnologs were 2-times elevated compared with SSD genes (33.94% versus 16.43%). Additionally, we assessed our methodology by applying it to a group of genes with strong a priori evidence for pathogenicity. The crowd-sourced Genomics England PanelApp gene list for CHD (available at <https://panelapp.genomicsengland.co.uk/panels/212/>), which represents a consensus view of causative genes, was highly enriched for ohnologs (76.6% versus 12.05%,  $\chi^2$  test,  $P < 0.0001$ ; OR=23.89 [95% CI, 12.33–46.18]). We, therefore, used ohnolog status as an additional candidate gene filter.

## Candidate Genes Supported by Both CNV and WES Data of CHD Cases

To further refine our candidate genes, we integrated additional genomic resources including the top 5% ExAC CNV intolerance scores, probability of haploinsufficiency (pHI),<sup>38</sup> probability of loss-of-function intolerance (pLI),<sup>19</sup> and RNAseq expression data from mouse embryonic hearts.<sup>39</sup> Lastly, we incorporated ohnolog status. Genes from BOTH CNVs were analyzed twice; once with the metrics used for genes from deletion CNVs and once with the metrics used for genes from duplication CNVs (Figure 4).

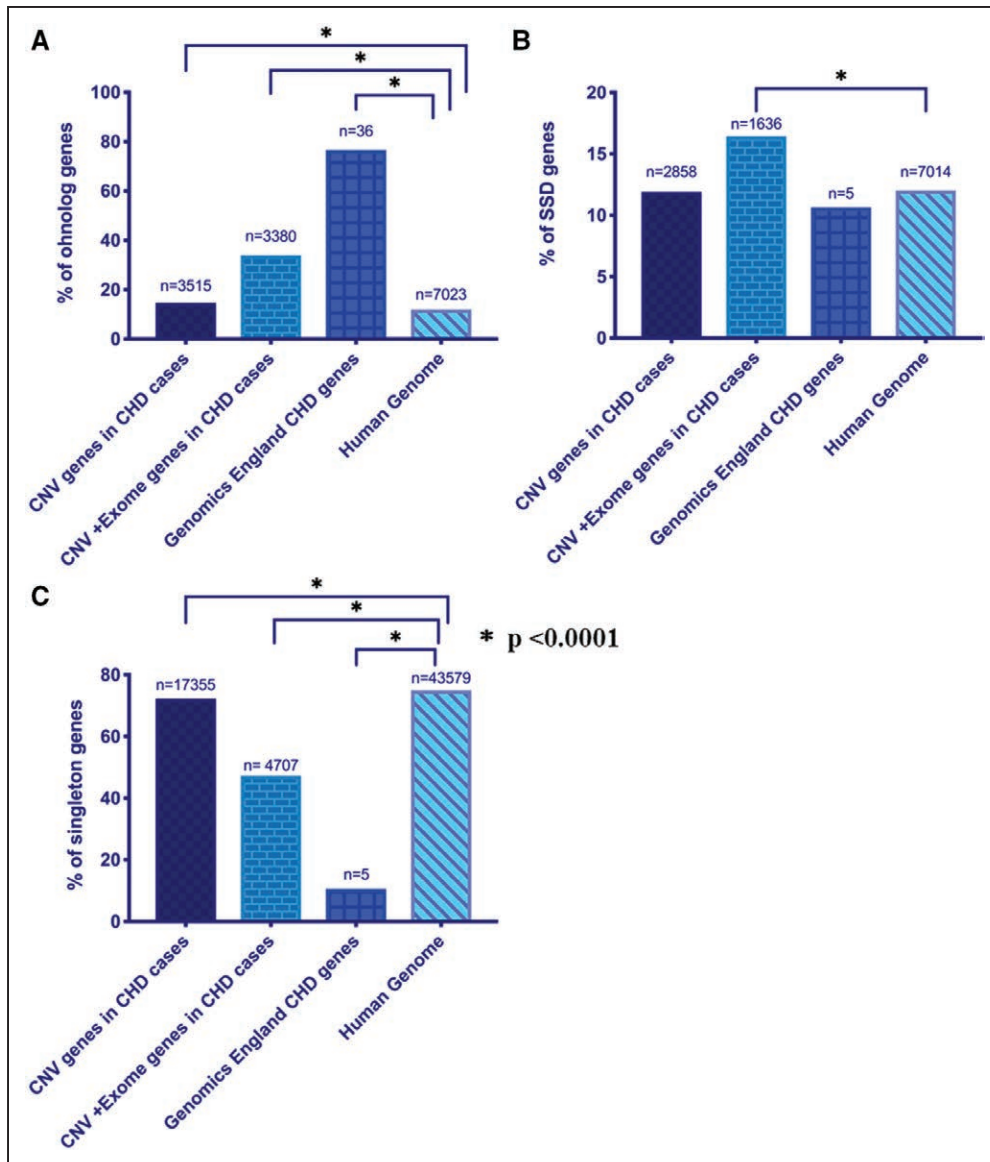
This led to the identification of 9 candidate genes from deletion and BOTH CNVs: *BRWD1*, *DIP2C*, *EYA3*, *GRB10*, *HNRNPC*, *RC3H2*, *SLIT3*, *TLN1*, and *UBASH3B*. All 9 have the following properties: (1) loss-of-function (LoF) variants in the WES data, (2) found in deletion or BOTH CNV regions only seen in nonsyndromic CHD cases, (3) top 5% of ExAC deletion CNV intolerance scores, (4) haploinsufficient (pHI  $\geq 0.65$ ) and unable to tolerate loss-of-function variants (pLI  $\geq 0.9$ ), (5) in the top 25% of highly expressed genes in mouse heart at E9.5 and E14.5, (6) ohnolog, (7) not present in the list of genes curated from the DDD study (Deciphering Developmental Disorders), and (8) not classified as human nonessential genes from the Sudmant study<sup>21</sup> (Table 2).

In addition, we found 45 candidate genes from duplication and BOTH CNVs, which had the following properties: (1) high or medium impact SNVs in the WES data, (2) found in duplication and BOTH CNV regions only seen in nonsyndromic CHD cases, (3) top 5% of ExAC duplication CNV intolerance scores, (4) in the top 25% of highly expressed genes in mouse heart at E9.5 and E14.5, (5) ohnolog, (6) not present in the list of genes curated from the DDD study, and (7) not in the list of nonessential human genes from the Sudmant study<sup>21</sup> (Table 2).

## Pathway Enrichment and Gene Ontology Analysis

We performed pathway enrichment analysis, using the Reactome Pathways Analysis tool,<sup>40</sup> on the final 54 candidate genes supported by both CNV and WES data in nonsyndromic CHD cases. This resulted in 11 pathways, where  $>5$  of our candidate genes were involved in those pathways (Table 3). The top 3 pathways based on entities ratio (entities found/total entities) from Reactome were axon guidance, signaling by receptor tyro-





**Figure 3. Ohnologs are enriched in congenital heart disease (CHD) cases.**

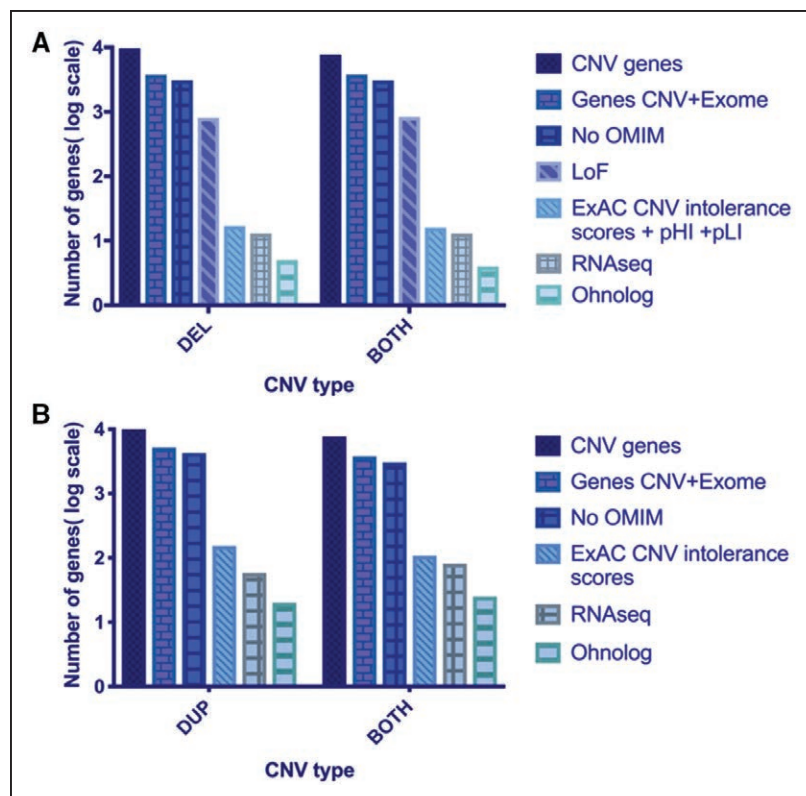
Graphs show the percentage of genes that are (A) ohnologs, (B) small scale duplications (SSD), and (C) singletons. Statistical significance was tested using 2-tailed  $\chi^2$  test with Yates's correction,  $P < 0.05$  was considered statistically significant. CNV indicates copy number variant.

sine kinases, and cellular responses to external stimuli. In addition, ingenuity pathway analysis was also used with the only pathway including >5 genes being axon guidance signaling. Gene ontology (GO) analysis<sup>41</sup> of our candidate genes revealed 22 GO terms with particular enrichment on 4 GO terms; apoptotic process involved in luteolysis (GO0061364; FDR corrected  $P$  value=0.0462), ventricular septum morphogenesis (GO0060412; FDR corrected  $P$  value=0.00921), ventricular septum development (GO0003281; FDR corrected  $P$  value=0.0343), and cardiac septum morphogenesis (GO0060411; FDR corrected  $P$  value=0.036). Both pathway and GO analysis identified processes in which the genes *ABLIM1*, *ARHGEF12*, *SLIT2*, and *SLIT3* are involved (Figure 5).

### **SLIT2 and SLIT3 Variants in CHD**

*SLIT2* and *SLIT3* were the most strongly supported genes found both by pathway analysis and GO (Figure 5). Therefore, we further explored the phenotypic associations of these genes within our population.

In the present study, individuals with CNVs including *SLIT3* were reported with malformation of the heart and great vessels (n=1), ventricular septal defect (n=1), atrial septal defect (n=3), and TOF (n=1) whereas individuals with *SLIT2* CNVs were reported with malformation of the heart and great vessels (n=1), ventricular septal defect (n=2), and double outlet right ventricle (n=1). In addition, 20 missense SNVs and 3 splice-site SNVs in *SLIT3* were found in 24 out of 829 TOF cases (2.9% [95% CI, 1.91%–4.35%]) and *SLIT2* had 12 missense



**Figure 4. Filtering process using large-scale genomic data resources.**

Both graphs are in logarithmic scale and represent the consecutive filtering of the genes using the different metrics for (A) deleted (DEL) and both copy number variant (CNV) genes (B) duplicated (DUP) and both CNV genes. There is  $\approx 70\%$  reduction in the number of candidate genes when we apply the evolutionary duplication metric—ohnolog. Also, none of our candidates were present in the list of homozygously deleted genes (nonessential) from the Sudmant study as well as not present in the list of genes curated from the DDD study. LoF indicates loss-of-function; OMIM, Online Mendelian Inheritance in Man; pHI, probability of haploinsufficiency; and pLI, probability of loss-of-function intolerance.

SNVs and 2 splice-site SNVs in 14 out of 829 TOF cases (1.7% [95% CI, 0.9%–2.9%]). Probands were available for 12 *SLIT3* variants, and 5 *SLIT2* variants which were confirmed by Sanger sequencing. Remaining variants were confirmed to have good coverage using Integrative Genomics Viewer. Samples from both parents were available for 9 probands with *SLIT3* variants and were analyzed for variant inheritance; 2 of the 9 *SLIT3* variants tested were identified as de novo. Samples from both parents were available for 5 probands with *SLIT2* variants and were all either maternally or paternally inherited.

## DISCUSSION

Here, we performed a large-scale genome-wide meta-analysis study of nonsyndromic CHD cases and identified 54 novel candidate genes for CHD. In addition to the large size of our data set, we incorporated a novel analysis strategy incorporating the evolutionary origin of gene duplications. Ohnologs tend not to be observed in CNVs in vertebrate genomes.<sup>13</sup> Moreover, McLysaght et al<sup>14</sup> have also shown that ohnologs are significantly overrepresented in pathogenic CNVs associated with schizophrenia and neurodevelopmental disorders and that they are the possible cause of the deleterious effects of these rare pathogenic CNVs. Here, we have shown for the first time that genes included in CNVs from CHD cases are significantly enriched for ohnologs compared with the human genome. Due to this signifi-

cant association between ohnologs and CHD, we incorporated ohnolog status in our methodology to identify novel genetic loci associated with CHD.

Pathway analysis and GO analysis concordantly identified the *SLIT2* and *SLIT3* genes, which have recently received increasing attention in heart development.<sup>42</sup> Invertebrates, the Slit family comprises of 3 known members (SLIT1-3), which are highly conserved secreted proteins that bind to ROBO (Roundabout receptors). *SLIT2* and *SLIT3* are expressed during mouse embryonic development and interact with ROBO1 and ROBO2.<sup>43,44</sup> They both encode proteins that consist of 4 LRR domains (leucine-rich repeats) also called D1-D4, 8 EGF repeats (epidermal growth factor) and 1 Laminin-G-like domain.<sup>43</sup> *Slit3* is expressed early in murine cardiogenesis in the cardiac crescent at E7.5 and linear heart tube at E8.5, later expression being restricted to the myocardium of the atria and OFT but not in the cardiac cushions or valves.<sup>44,45</sup> *Slit2* is strongly expressed in the pharyngeal region at E8.5–E9.5 and later in the ventricular trabecular myocardium, epicardium, aortic semilunar valves and the mesenchyme surrounding the caval veins.<sup>44,46</sup> *Slit3*<sup>-/-</sup> mutant mice exhibit ventricular septal defect, thick atrioventricular valves, and hypoplastic posterior aortic semilunar leaflet with *Slit2*<sup>-/-</sup> mutant mice exhibiting bicuspid aortic valves and immature semilunar valves.<sup>44</sup> *Robo1*<sup>-/-</sup> mutant mice also exhibit ventricular septal defect with downregulation of Notch signaling, suggesting a potential mechanism for the underlying defects.<sup>44</sup> In another study, *Slit3*<sup>-/-</sup> mice also

**Table 2.** Candidate Genes Supported by both CNV and WES Data of CHD Cases

ENS Gene ID	Gene Name	Chr	Start (hg19)	End (hg19)	DEL/DUP/BOTH	TOF LOF Var Count	TOF HIGH Impact Var Count	TOF MED impact Var count	Case CNVs Overlap: FULL	Case CNVs Overlap: PARTIAL
ENSG00000058668	<i>ATP2B4</i>	1	203595689	203713209	DUP	0	0	13	3	1
ENSG00000064042	<i>LIMCH1</i>	4	41361624	41702061	DUP	6	6	21	4	0
ENSG00000083223	<i>ZCCHC6</i>	9	88902648	88969369	DUP	1	1	8	4	0
ENSG00000092847	<i>AGO1</i>	1	36335409	36395211	DUP	0	0	2	3	0
ENSG00000094916	<i>CBX5</i>	12	54624724	54673886	DUP	0	0	1	2	0
ENSG00000101367	<i>MAPRE1</i>	20	31407699	31438211	DUP	0	0	1	5	0
ENSG00000108387	<i>SEPT4</i>	17	56597611	56618179	DUP	0	0	6	5	1
ENSG00000108389	<i>MTMR4</i>	17	56566898	56595266	DUP	0	0	9	10	0
ENSG00000109332	<i>UBE2D3</i>	4	103715540	103790053	DUP	2	2	0	3	0
ENSG00000109685	<i>WHSC1</i>	4	1873151	1983934	DUP	0	0	12	6	0
ENSG00000112079	<i>STK38</i>	6	36461669	36515247	DUP	0	0	3	5	1
ENSG00000113108	<i>APBB3</i>	5	139937853	139973337	DUP	1	2	8	8	0
ENSG00000122515	<i>ZMIZ2</i>	7	44788180	44809477	DUP	1	1	10	10	0
ENSG00000138641	<i>HERC3</i>	4	89442199	89629693	DUP	0	0	5	14	0
ENSG00000138835	<i>RGS3</i>	9	116207011	116360018	DUP	1	1	19	5	0
ENSG00000140403	<i>DNAJA4</i>	15	78556428	78574538	DUP	0	0	8	9	0
ENSG00000140497	<i>SCAMP2</i>	15	75136071	75165706	DUP	0	0	1	6	0
ENSG00000145147	<i>SLIT2</i>	4	20254883	20622184	DUP	0	0	14	3	0
ENSG00000146463	<i>ZMYM4</i>	1	35734568	35887659	DUP	0	0	12	26	1
ENSG00000179361	<i>ARID3B</i>	15	74833518	74890472	DUP	0	0	6	5	0
ENSG00000185658	<i>BRWD1</i>	21	40556102	40693485	DEL	5	5	16	10	1
ENSG00000151240	<i>DIP2C</i>	10	320130	735683	DEL	1	1	15	4	0
ENSG00000158161	<i>EYA3</i>	1	28296855	28415207	DEL	1	1	2	4	0
ENSG00000106070	<i>GRB10</i>	7	50657760	50861159	DEL	1	2	8	3	0
ENSG00000154127	<i>UBASH3B</i>	11	122526383	122685181	DEL	1	1	5	9	0
ENSG00000092199	<i>HNRNPC</i>	14	21677295	21737653	BOTH	1	1	2	6	0
ENSG00000056586	<i>RC3H2</i>	9	125606835	125667620	BOTH	1	1	8	3	1
ENSG00000184347	<i>SLIT3</i>	5	168088745	168728133	BOTH	2	2	21	3	0
ENSG00000137076	<i>TLN1</i>	9	35696945	35732392	BOTH	4	4	22	9	0
ENSG00000100117	<i>RANBP9</i>	6	13621730	13711796	BOTH	0	0	14	3	0
ENSG00000020577	<i>SAMD4A</i>	14	55033815	55260033	BOTH	0	0	11	8	0
ENSG00000033800	<i>PIAS1</i>	15	68346517	68483096	BOTH	0	0	4	7	1
ENSG00000064726	<i>BTBD1</i>	15	83685174	83736106	BOTH	1	1	7	5	0
ENSG00000083312	<i>TNPO1</i>	5	72112139	72212560	BOTH	0	0	4	5	0
ENSG00000091009	<i>RBM27</i>	5	145583113	145718814	BOTH	0	0	7	4	0
ENSG00000099204	<i>ABLIM1</i>	10	116190872	116444762	BOTH	3	3	14	12	0
ENSG00000100320	<i>RBFOX2</i>	22	36134783	36424473	BOTH	0	0	5	2	0
ENSG00000100330	<i>MTMR3</i>	22	30279144	30426855	BOTH	0	0	11	4	0
ENSG00000100592	<i>DAAM1</i>	14	59655364	59838123	BOTH	0	0	9	14	0
ENSG00000113649	<i>TCERG1</i>	5	145826874	145891524	BOTH	0	0	6	9	0
ENSG00000116191	<i>RALGPS2</i>	1	178694300	178889238	BOTH	0	0	1	2	0
ENSG00000120899	<i>PTK2B</i>	8	27168999	27316903	BOTH	0	0	8	4	0
ENSG00000127022	<i>CANX</i>	5	179105629	179157926	BOTH	0	0	10	2	0
ENSG00000135074	<i>ADAM19</i>	5	156822542	157002783	BOTH	2	2	10	7	1

(Continued)

**Table 2. Continued**

ENS Gene ID	Gene Name	Chr	Start (hg19)	End (hg19)	DEL/DUP/BOTH	TOF LOF Var Count	TOF HIGH Impact Var Count	TOF MED impact Var count	Case CNVs Overlap: FULL	Case CNVs Overlap: PARTIAL
ENSG00000137573	<i>SULF1</i>	8	70378859	70573150	BOTH	1	1	10	3	0
ENSG00000137962	<i>ARHGAP29</i>	1	94614544	94740624	BOTH	0	0	12	3	0
ENSG00000138107	<i>ACTR1A</i>	10	104238986	104262482	BOTH	0	0	2	2	1
ENSG00000155506	<i>LARP1</i>	5	154092462	154197167	BOTH	2	2	13	6	0
ENSG00000166747	<i>AP1G1</i>	16	71762913	71843104	BOTH	0	1	4	4	1
ENSG00000166888	<i>STAT6</i>	12	57489191	57525922	BOTH	0	0	7	5	0
ENSG00000180340	<i>FZD2</i>	17	42634925	42636907	BOTH	0	0	8	8	0
ENSG00000180776	<i>ZDHC20</i>	13	21950263	22033509	BOTH	1	1	2	6	0
ENSG00000196914	<i>ARHGEF12</i>	11	120207787	120360645	BOTH	0	0	14	12	0
ENSG00000213079	<i>SCAF8</i>	6	155054459	155155192	BOTH	0	0	13	14	0

Fifty-four protein-coding candidate genes supported by CNV and WES data in nonsyndromic CHD cases. All genes in the list are strict ohnologs. Data presented includes the Ensembl ID and the nature of the chromosomal imbalance for which the gene is either deleted (DEL), duplicated (DUP), or deleted/duplicated (BOTH). Chr indicates chromosome; ENS, ensemble; LoF, loss-of-function; MED, medium impact; TOF, tetralogy of Fallot; and Var, variants.

exhibit congenital diaphragmatic hernia.<sup>47</sup> Congenital heart defects ranging from bicuspid aortic valves to septal, and outflow tract defects are, therefore, observed in variety of animal models in which genes involved in the SLIT/ROBO pathway have been inactivated.

We identified *SLIT2* and *SLIT3* heterozygous SNVs in 2.9% and 1.7% nonsyndromic TOF patients, respectively. All SNVs were novel or rare (either absent from ExAC or with frequency of <0.01) and predicted with in silico tools to be pathogenic. The majority of the SNVs in both genes were missense, although a few splice-site SNVs were also found. Their functional relevance will be of interest in future studies. Both genes were also present in CNVs in patients with CHD with varying phenotypes, including septal defects and malformation of the great arteries. This is the first study to find an association of *SLIT2* and *SLIT3* with predisposition to human CHD, although, of note, a recent study iden-

tified *ROBO1* loss-of-function SNVs in cases with TOF and septal defects.<sup>48</sup>

## Limitations

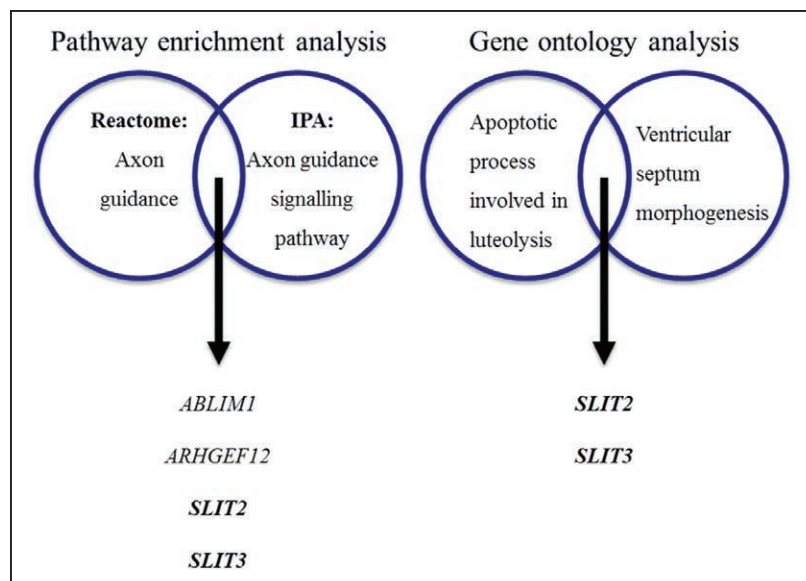
This study has certain limitations. The databases and publications included in this analysis incorporated different CNV detection platforms and analysis algorithms.<sup>49</sup> Irrespective of the method used in the studies identifying pathogenic CNVs, we only included studies that used the same genotyping method between cases and controls and confirmed their CNV detection by an additional methodology like quantitative polymerase chain reaction, which to a degree addresses this limitation. Another potential limitation is the fact that during our filtering strategy, we might have missed some important genes crucial for cardiac development. Though we accept that all important genes will not have been captured, we detected 54 strong candidate

**Table 3. Top Pathways Overrepresented in Our 54 Candidate Genes**

Pathway Tool	Pathway Name	*Entities Found	*Entities Total	Entities Ratio (%)
Reactome	Axon guidance	8	583	1.372212693
Ingenuity pathway analysis	Axon guidance signaling pathway	7	501	1.397206
Reactome	Signaling by receptor tyrosine kinases	5	521	0.959692898
Reactome	Cellular responses to external stimuli	5	621	0.805152979
Reactome	Signaling by interleukins	5	641	0.780031201
Reactome	Developmental biology	8	1177	0.679694138
Reactome	Adaptive immune system	6	998	0.601202405
Reactome	Cytokine signaling in immune system	6	1056	0.568181818
Reactome	Signal transduction	15	3202	0.468457214
Reactome	Post-translational protein modification	7	1594	0.439146801
Reactome	Immune system	11	2662	0.41322314
Reactome	Metabolism of proteins	9	2354	0.382327952

\*Number.





**Figure 5. Genes in the top significant pathways and biological processes.**

*SLIT2* and *SLIT3* genes were supported by multiple lines of evidence. IPA indicates ingenuity pathway analysis.

genes supported by multiple lines of evidence as having a causative role in nonsyndromic CHD. Further research in much larger numbers of comprehensively genetically characterized CHD cases is warranted to establish the magnitude of the contribution of these genes, and to discover novel loci.

In conclusion, we show that ohnologs are overrepresented in CHD cases and that incorporation of the evolutionary origins of genes is useful in refining candidate genes emerging from large-scale genetic evaluations of CHD. We also observe that CNVs and SNVs in *SLIT2* and *SLIT3* are associated with CHD involving TOF, septal defects, and outflow tract defects, supporting the importance of the SLIT-ROBO signaling pathway in heart development.

## ARTICLE INFORMATION

Received July 21, 2019; accepted September 20, 2019.

The Data Supplement is available at <https://www.ahajournals.org/doi/suppl/10.1161/CIRCGEN.119.002694>.

## Correspondence

Bernard D. Keavney, The University of Manchester, AV Hill Bldg, Manchester, M13 9PT, Email [bernard.keavney@manchester.ac.uk](mailto:bernard.keavney@manchester.ac.uk) or Elisavet Fotiou, The University of Manchester, AV Hill Bldg, Manchester, M13 9PT, Email [elisavet.fotiou@postgrad.manchester.ac.uk](mailto:elisavet.fotiou@postgrad.manchester.ac.uk)

## Affiliations

Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine, and Health, Manchester Academic Health Science Centre (E.F., S.W., G.T., B.K.) and Division of Evolution and Genomic science (A.M.-G., D.L.R., K.E.H.), University of Manchester. MRC-University of Glasgow Centre for Virus Research (D.L.R.), Manchester Heart Centre, Manchester University NHS Foundation Trust, Manchester (B.K.).

## Acknowledgments

This study makes use of data generated by the DECIPHER Consortium. A full list of centers that contributed to the generation of the data is available from

<https://decipher.sanger.ac.uk/> and via email from [decipher@sanger.ac.uk](mailto:decipher@sanger.ac.uk). Funding for the DECIPHER project was provided by the Wellcome Trust.<sup>20</sup>

## Sources of Funding

This study was supported by the British Heart Foundation. A. Martin-Geary is funded by a PhD studentship from the Medical Research Council (number 1622139). Prof Robertson is partially funded by the Medical Research Council (MC UU 1201412). Prof Keavney holds a British Heart Foundation Personal Chair.

## Disclosures

None.

## REFERENCES

- Liu Y, et al. Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. *Int J Epidemiol.* 2019; 48:455–463.
- Cordell HJ, et al. Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat Genet.* 2013;45:822–824. doi: 10.1038/ng.2637
- Soemedi R, 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. doi: 10.1016/j.ajhg.2012.08.003
- Kaminsky EB, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med.* 2011;13:777–784. doi: 10.1097/GIM.0b013e31822c79f9
- Thorsson T, et al. Chromosomal imbalances in patients with congenital cardiac defects: a meta-analysis reveals novel potential critical regions involved in heart development. *Congenit Heart Dis.* 2015;10:193–208. doi: 10.1111/chd.12179
- Page DJ, et al. Whole exome sequencing reveals the major genetic contributors to nonsyndromic tetralogy of fallot. *Circ Res.* 2019;124:553–563. doi: 10.1161/CIRCGEN.118.313250
- Jin SC, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet.* 2017;49:1593–1601. doi: 10.1038/ng.3970
- Greenway SC, et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of fallot. *Nat Genet.* 2009;41:931–935. doi: 10.1038/ng.415
- Costain G, et al. Rare copy number variations in an adult with transposition of the great arteries emphasize the importance of updated genetic assessments in syndromic congenital cardiac disease. *Int J Cardiol.* 2016;203:516–518. doi: 10.1016/j.ijcard.2015.10.216

10. Rice AM, et al. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat Commun*. 2017;8:14366. doi: 10.1038/ncomms14366
11. Makino T, et al. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A*. 2010;107:9270–9274. doi: 10.1073/pnas.0914697107
12. Martin-Geary A, et al. Human genetic disease is greatly influenced by the underlying fragility of evolutionarily ancient genes. *bioRxiv*. 2019;558916. doi: 10.1101/558916
13. Makino T, et al. Genome-wide deserts for copy number variation in vertebrates. *Nat Commun*. 2013;4:2283. doi: 10.1038/ncomms3283
14. McLysaght A, et al. Ohnologs are overrepresented in pathogenic copy number mutations. *Proc Natl Acad Sci U S A*. 2014;111:361–366. doi: 10.1073/pnas.1309324111
15. Smith BH, et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. 2013;42:689–700. doi: 10.1093/ije/dys084
16. Dickerson JE, et al. On the origins of mendelian disease genes in man: the impact of gene duplication. *Mol Biol Evol*. 2012;29:61–69. doi: 10.1093/molbev/msr111
17. Singh PP, et al. Human dominant disease genes are enriched in paralogues originating from whole genome duplication. *PLoS Comput Biol*. 2014;10:e1003754. doi: 10.1371/journal.pcbi.1003754
18. Ruderfer DM, et al; Exome Aggregation Consortium. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet*. 2016;48:1107–1111. doi: 10.1038/ng.3638
19. Lek M, et al; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291. doi: 10.1038/nature19057
20. Bragin E, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res*. 2014;42:D993–D1000. doi: 10.1093/nar/gkt937
21. Sudmant PH, et al; 1000 Genomes Project Consortium. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526:75–81. doi: 10.1038/nature15394
22. Vulto-van Silfhout AT, et al. An update on ECARUCA, the European Cytogeneticists Association Register of unbalanced chromosome aberrations. *Eur J Med Genet*. 2013;56:471–474. doi: 10.1016/j.ejmg.2013.06.010
23. Fakhro KA, et al. Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci U S A*. 2011;108:2915–2920. doi: 10.1073/pnas.1019645108
24. Hitz MP, et al. Rare copy number variants contribute to congenital left-sided heart disease. *PLoS Genet*. 2012;8:e1002903. doi: 10.1371/journal.pgen.1002903
25. Glessner JT, 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. doi: 10.1161/CIRCRESAHA.115.304458
26. Rigler SL, et al. Novel copy-number variants in a population-based investigation of classic heterotaxy. *Genet Med*. 2015;17:348–357. doi: 10.1038/gim.2014.112
27. Hightower HB, et al. Array comparative genomic hybridisation testing in CHD. *Cardiol Young*. 2015;25:1155–1172. doi: 10.1017/S1047951114001838
28. Sanchez-Castro M, et al. Search for rare copy-number variants in congenital heart defects identifies novel candidate genes and a potential role for FOXC1 in patients with coarctation of the aorta. *Circ Cardiovasc Genet*. 2016;9:86–94. doi: 10.1161/CIRCGENETICS.115.001213
29. Hanchard NA, et al. Assessment of large copy number variants in patients with apparently isolated congenital left-sided cardiac lesions reveals clinically relevant genomic events. *Am J Med Genet A*. 2017;173:2176–2188. doi: 10.1002/ajmg.a.38309
30. Xie L, et al. Rare de novo copy number variants in patients with congenital pulmonary atresia. *PLoS One*. 2014;9:e96471. doi: 10.1371/journal.pone.0096471
31. Xie HM, et al. Rare copy number variants in patients with congenital conotruncal heart defects. *Birth Defects Res*. 2017;109:271–295. doi: 10.1002/bdra.23609
32. Miller DT, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*. 2010;86:749–764. doi: 10.1016/j.ajhg.2010.04.006
33. MacDonald JR, et al. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42:D986–D992. doi: 10.1093/nar/gkt958
34. Auton A, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
35. Collins RL, et al. An open resource of structural variation for medical and population genetics. *bioRxiv*. 2019;578674. doi: 10.1101/578674
36. McKusick-Nathans Institute of Genetic Medicine JHUB, MD. *Online Mendelian Inheritance in Man, OMIM*®. 2019.
37. Singh PP, et al. Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. *PLoS Comput Biol*. 2015;11:e1004394. doi: 10.1371/journal.pcbi.1004394
38. Huang N, et al. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet*. 2010;6:e1001154. doi: 10.1371/journal.pgen.1001154
39. Zaidi S, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013;498:220–223. doi: 10.1038/nature12141
40. Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2018;46:D649–D655. doi: 10.1093/nar/gkx1132
41. Mi H, et al. PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017;45:D183–D189. doi: 10.1093/nar/gkw1138
42. Zhao J, et al. Slit-robo signalling in heart development. *Cardiovasc Res*. 2018;114:794–804. doi: 10.1093/cvr/cvy061
43. Blockus H, et al. Slit-robo signaling. *Development*. 2016;143:3037–3044. doi: 10.1242/dev.132829
44. Mommersteeg MT, et al. Disrupted Slit-Robo signalling results in membranous ventricular septum defects and bicuspid aortic valves. *Cardiovasc Res*. 2015;106:55–66. doi: 10.1093/cvr/cv040
45. Medioni C, et al. Expression of slit and robo genes in the developing mouse heart. *Dev Dyn*. 2010;239:3303–3311. doi: 10.1002/dvdy.22449
46. Mommersteeg MT, et al. Slit-roundabout signaling regulates the development of the cardiac systemic venous return and pericardium. *Circ Res*. 2013;112:465–475. doi: 10.1161/CIRCRESAHA.112.277426
47. Liu J, et al. Congenital diaphragmatic hernia, kidney agenesis and cardiac defects associated with Slit3-deficiency in mice. *Mech Dev*. 2003;120:1059–1070. doi: 10.1016/s0925-4773(03)00161-8
48. Kruszka P, et al. Loss of function in ROBO1 is associated with tetralogy of fallot and septal defects. *J Med Genet*. 2017;54:825–829. doi: 10.1136/jmedgenet-2017-104611
49. Pinto D, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010;466:368–372. doi: 10.1038/nature09146