European Journal of Human Genetics (2013) 21, 173–181
© 2013 Macmillan Publishers Limited All rights reserved 1018-4813/13
www.nature.com/ejhg

ARTICLE

Rare DNA copy number variants in cardiovascular malformations with extracardiac abnormalities

Seema R Lalani*[,1,10,11], Chad Shaw[1,11], Xueqing Wang[1,10], Ankita Patel[1], Lance W Patterson[2], Katarzyna Kolodziejska[1], Przemyslaw Szafranski[1], Zhishuo Ou[1], Qi Tian[1], Sung-Hae L Kang[1], Amina Jinnah[1], Sophia Ali[3], Aamir Malik[4], Patricia Hixson[1], Lorraine Potocki[1], James R Lupski[1], Pawel Stankiewicz[1], Carlos A Bacino[1], Brian Dawson[1], Arthur L Beaudet[1], Fatima M Boricha[5], Runako Whittaker[6], Chumei Li[7], Stephanie M Ware[8], Sau Wai Cheung[1], Daniel J Penny[2], John Lynn Jefferies[9] and John W Belmont[1,10]

Clinically significant cardiovascular malformations (CVMs) occur in 5–8 per 1000 live births. Recurrent copy number variations (CNVs) are among the known causes of syndromic CVMs, accounting for an important fraction of cases. We hypothesized that many additional rare CNVs also cause CVMs and can be detected in patients with CVMs plus extracardiac anomalies (ECAs). Through a genome-wide survey of 203 subjects with CVMs and ECAs, we identified 55 CNVs >50 kb in length that were not present in children without known cardiovascular defects (*n*=872). Sixteen unique CNVs overlapping these variants were found in an independent CVM plus ECA cohort (*n*=511), which were not observed in 2011 controls. The study identified 12/16 (75%) novel loci including non-recurrent *de novo* 16q24.3 loss (4/714) and *de novo* 2q31.3q32.1 loss encompassing *PPP1R1C* and *PDE1A* (2/714). The study also narrowed critical intervals in three well-recognized genomic disorders of CVM, such as the cat-eye syndrome region on 22q11.1, 8p23.1 loss encompassing *GATA4* and *SOX7*, and 17p13.3-p13.2 loss. An analysis of protein-interaction databases shows that the rare inherited and *de novo* CNVs detected in the combined cohort are enriched for genes encoding proteins that are direct or indirect partners of proteins known to be required for normal cardiac development. Our findings implicate rare variants such as 16q24.3 loss and 2q31.3-q32.1 loss, and delineate regions within previously reported structural variants known to cause CVMs.

European Journal of Human Genetics (2013) **21,** 173–181; doi:10.1038/ejhg.2012.155; published online 29 August 2012

Keywords: rare copy number variations; extracardiac anomalies (ECAs); cardiovascular malformations (CVMs); 16q24.3 microdeletion; protein-interaction network

INTRODUCTION

Congenital cardiovascular malformations (CVMs) are among the most common of all medically significant birth defects and are a leading contributor to infant mortality in the United States.[1] Yet, the cause of these malformations is unknown in ~85–90% of cases.[2,3] Several large population-based studies estimate the prevalence of CVMs to range from 3 to 6 cases per 1000 live births.[4–6] The birth prevalence for severe CVMs in other studies is reported to be ~1.5 cases per 1000 live births.[7,8] Despite a wealth of information on the developmental pathways that operate during normal cardiogenesis, only a limited number of genes have been identified in which disease-causing mutations are associated with CVMs. Chromosomal abnormalities including trisomy 18, trisomy 13, Turner syndrome, Down syndrome and DiGeorge syndrome account for ~10% of all cases of CVMs,[9] whereas another 3–5% of cases are caused by single gene disorders including CHARGE syndrome, Noonan syndrome and Holt–Oram syndrome.[10,11] Estimating the fraction with a syndrome or genetic condition is challenging, although it can be approximated from the population-based Baltimore-Washington Infant Study in which nearly 17% of infants with a CVM had an identifiable syndrome.[3]

The copy number variations (CNVs) resulting from instability of regional genomic architecture[12] are an important cause of CVMs[13] such as DiGeorge syndrome (22q11.2 deletion)[14] and Williams syndrome (7q11.23 deletion).[15] Other genomic disorders associated with CVMs include Smith–Magenis syndrome,[16] 17q21.3 microdeletion syndrome[17] and 17q23.1q23.2 recurrent microdeletion syndrome.[18] These genomic disorders affect several contiguous genes, but often as exemplified by the *TBX1* gene in 22q11del syndrome[19] and *ELN* in Williams syndrome,[20] a single gene is thought to be the major factor in causing the cardiovascular developmental defects because of haploinsufficiency.

For genes known to cause CVMs, there is typically variable expressivity, and extensive allelic heterogeneity. Currently, a working model for the genetic architecture of CVMs involves both rare and

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; [2]Section of Pediatric Cardiology, Texas Children's Hospital, Houston, TX, USA; [3]The University of Texas at Austin, Austin, TX, USA; [4]The Aga Khan University Hospital, Karachi, Pakistan; [5]Department of Pediatrics, University of Texas, Houston, TX, USA; [6]St John Medical Center, Tulsa, OK, USA; [7]Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada; [8]Division of Molecular Cardiovascular Biology, Cincinnati Children's Hospital and University of Cincinnati College of Medicine, Cincinnati, OH, USA; [9]Division of Pediatric Cardiology, Cincinnati Children's Hospital and University of Cincinnati College of Medicine, Cincinnati, OH, USA; [10]Children's Nutrition Research Center, Houston, TX, USA
[11]These authors contributed equally to this work.
*Correspondence: Dr SR Lalani, Department of Molecular and Human Genetics, Baylor College of Medicine, 1100 Bates Street, Room 8005, Houston, TX 77030, USA. Tel: +1 713 798 8921; Fax: +1 832 825 4294; E-mail: seemal@bcm.tmc.edu
Received 11 January 2012; revised 12 June 2012; accepted 13 June 2012; published online 29 August 2012

common variants, possibly with interaction of multiple genes and gene–environment interactions.[21] Rare variants are known to cause CVMs[22] and if these account for the major fraction of cases, the high frequency of CVMs could be explained by the large number of genes that play a role in normal cardiac development, called the 'large mutational target' hypothesis.[21,23] We designed this study, to define critical regions within both novel and established genome structural variations likely to be important for human cardiac patterning. To establish a genome-wide schema of minimal critical regions encompassing haploinsufficient/dosage-sensitive genes in CVMs, a large pediatric cohort with extracardiac anomalies (ECAs) and normal cardiac studies was selected as controls for the study. We predicted that a genome-wide comparative analysis of cases and affected controls for segmental aneusomies would delineate a global signature of critical regions that are specific to cases.

MATERIALS AND METHODS

Cases

The study was approved by the Institutional Review Board of the Baylor College of Medicine in Houston, Texas. The discovery set of 203 unrelated patients (BCM1 cohort) comprised 104 Hispanic/Latino Americans and 99 non-Hispanic patients of European descent, all evaluated at Texas Children's Hospital (TCH) in Houston, Texas. The cases included 104 males and 99 females, with a mean age of 3.0 years (0–18 years). All cases in this cohort were assessed by cardiologists, neurologists and/or dysmorphologists at TCH between January 2008 and July 2010. The echocardiographic diagnoses were used to classify cases according to the scheme of Botto *et al*[24] and encompassed eight categories: left ventricular outflow tract obstructive defects (LVOTOs, $n=30$), right ventricular outflow tract obstructive defects (RVOTOs, $n=26$), septal defects ($n=54$), atrioventricular septal defects (AVSDs, $n=3$), anomalous pulmonary venous return (APVR, $n=4$), conotruncal defects ($n=29$), heterotaxy ($n=14$) and complex ($n=17$) anomalies (Supplementary Figure 1). The complex category included patients with multiple cardiac anomalies, L-transposition of great arteries (L-TGA) with RVOTOs, L-TGA plus LVOTOs, single ventricle or double-inlet left ventricle type with either d- or l-malposed great arteries. Individuals with dilated aortic root ($n=11$) and cardiomyopathy ($n=15$) were additionally included if there were no known molecular diagnoses. Those with patent ductus arteriosus (PDA), patent foramen ovale (PFO) and arrhythmias were excluded from the study. All cases were affected with ECAs, unexplained developmental delay and/or facial dysmorphisms (Supplementary Table 1). Of the 203 cases in the BCM1 cohort, 159 (78%) had renal ultrasound study, performed for ECA evaluation. Normal results were confirmed in 107 (67%) cases. Renal findings observed in the other 52 cases (33%) included hydronephrosis, unilateral renal agenesis, crossed renal ectopia and nephrocalcinosis. Brain imaging including cranial ultrasound, magnetic resonance imaging or CT scan were completed in 158/203 cases ($\sim$78%). Of these, 83 (52%) were normal studies. The abnormal findings observed in 75 cases (48%) are detailed in Supplementary Table 1.

The second cohort consisted of 511 unrelated pediatric patients, also evaluated at TCH (BCM2 cohort). This group consisted of approximately equal numbers of Hispanic/Latino Americans and non-Hispanic patients of European descent. Cardiac diagnoses were confirmed by echocardiography in all cases. The subjects were similarly classified according to Botto *et al*[24] (Supplementary Figure 2). This cohort included 169 cases with septal defects, 94 with LVOTOs, 67 with conotruncal malformations, 57 with RVOTOs, 33 with complex defects, 25 with heterotaxy, 4 with APVR, 24 with AVSDs, 23 with vasculopathy and 15 with cardiomyopathy diagnoses. All cases were assessed by cardiologists, developmental pediatricians, neurologists and/or geneticists at TCH.

Controls

The primary controls ($n=872$) had normal echocardiogram studies and/or normal clinical cardiovascular exam, performed by cardiologists at TCH, between January 2006 and July 2010. The controls used for second cohort included 2011 subjects from TCH with normal cardiovascular studies, confirmed similarly by echocardiogram and/or clinical exam at TCH (Supplementary Figure 3). Both control cohorts had similar rates of ECAs, with or without unexplained developmental delay, and/or facial dysmorphisms (Supplementary Table 2). They were composed of approximately equal number of non-Hispanic Caucasians and Hispanic/Latino Americans, between the ages of 0 and 18 years. Both the first and second control sets were matched for age and ethnicity with the cases, as illustrated by the principal component analysis utilizing 10 000 oligonucleotides involving polymorphic loci (Supplementary Figure 4). Using controls who had ECAs but were not affected with CVMs has several advantages that improve the specificity of the analysis, when a genomic region is a candidate to cause cardiac malformations: (1) all cases and controls came from the same clinic-based cohort; (2) were well matched for age and (3) ethnic background; (4) all studies for cases and controls were performed with similar Agilent technology; and were (5) analyzed with an identical segmentation statistical method, confirmed by a (6) large number of confirmatory observations using orthogonal technology; (7) all assays were performed on freshly isolated peripheral blood lymphocytes; and finally (8) the specificity of the CNV association with cardiac malformations makes it more likely that genes would be ascertained with specific cardiac development functions within this group of conditions.

To exclude the possibility of ascertaining polymorphic CNVs, Illumina genotypes of 2024 pediatric subjects from Children's Hospital of Philadelphia (CHOP) Control CNV Study, dbGaP Study Accession: phs000199.v1.p1 were also obtained for comparison from the Database of Genotypes and Phenotypes (dbGAP) found at http://www.ncbi.nih.gov.ezproxyhost.library.tmc.edu/gap.

All 203 cases (BCM1 cohort) and 872 controls were studied with customized 105k genome-wide arrays, using array comparative genomic hybridization (CGH) with $\sim$105 000 oligonucleotides covering the whole genome at an average resolution of 30 kb, with denser coverage at disease loci. The array was designed by Baylor Medical Genetics Laboratories and manufactured by Agilent Technology (Santa Clara, CA, USA; http://www.bcm.edu/geneticlabs/cma/tables.html).[25–27] In the second cohort, the cases and controls were studied by either 44k, 105k or 180k oligonucleotide arrays on similar Agilent platform and including the 55 CNVs identified in the discovery set. Peripheral blood samples from the subjects were submitted to the Baylor Medical Genetics Laboratories. DNA was extracted from whole blood using the Puregene DNA Blood Kit (Gentra, Minneapolis, MN, USA) according to the manufacturer's instructions. Data analysis methods are included in the Supplementary Text.

RESULTS

CNVs in cases

Analysis of the 105k custom Agilent array data from BCM1 cohort including 203 subjects with CVMs plus ECAs detected a total of 547 CNVs (excluding Y chromosome, 276 copy number gains and 271 copy number losses), ranging from 54 kb to 36 Mb in size. Of these CNVs, 334 (61%) had $\geq$93% overlap with copy number polymorphisms in the human genome, based on DGV. Large cytogenetic aberrations including unbalanced rearrangements were observed in 11/203 (3.9%) cases (Table 1).

Multiple CNVs known to be associated with syndromic CVMs were observed, including 22q11.2 loss involving the DiGeorge critical region, seen in 7/203 cases, 7q11.23 loss corresponding to the Williams–Beuren syndrome (WBS) critical region, observed in 5/203 cases, and 16p13.11 gain including *MYH11*,[28] observed in 2/203 cases. Other rare genomic disorders in this cohort observed as singleton events included 1q21.1 recurrent microduplication, 22q11.1-q11.2 gain associated with Cat-Eye syndrome (CES), 17p13.3 loss corresponding to Miller–Dieker lissencephaly syndrome, 17p11.2 loss seen in Smith–Magenis syndrome and 17p11.2 reciprocal gain, observed in Potocki–Lupski syndrome (Table 1).

We further compared cases with age- and ethnicity-matched controls to identify novel rare variants, observed in 203 cases but

Table 1 Summary of the CNVs associated with known DECIPHER syndromes and large cytogenetic aberrations observed in 203 cases

Chromosome region	*Event*	*Length (bp)*	*Cytoband*	*Cases (*n*)*	*Decipher syndromes*
chr1:0–6 675 076	Loss	6675077	p36.33–p36.31	4	1p36 microdeletion syndrome
chr1:144 768 813–145 860 886	Gain	1092074	q21.1	1	1q21.1 recurrent microduplication syndrome
chr7:72,031,958–73 625 107	Loss	1593150	q11.23	5	7q11.23 deletion syndrome, Williams–Beuren Syndrome
chr8:8,448,638–12 203 210	Loss	3754573	p23.1	1	8p23.1 deletion syndrome
chr16:14 977 157–16 515 701	Gain	1538545	p13.11	2	16p13.11 recurrent microduplication syndrome
chr17:1 009 930–1 068 719	Loss	58789	p13.3	1	Miller–Dieker syndrome (MDS)
chr17:18 988 039–20 380 184	Gain	1392146	p11.2	1	Potocki–Lupski syndrome (17p11.2 duplication syndrome)
chr17:18 472 298–18 988 039	Loss	515742	p11.2	1	Smith–Magenis Syndrome (17p11.2 deletion syndrome)
chr22:17 218 693–20 282 517	Loss	3063825	q11.21	7	22q11 deletion (Velocardiofacial/ DiGeorge syndrome)
chr22:15 659 961–19 066 406	Gain	3406445	q11.1–q11.21	1	Cat-Eye Syndrome (Type I), 22q11 duplication syndrome
Large cytogenetic aberrations					
chr1:212 404 889–247 249 719	Gain	34844831	q41–q44	1	
chr4:6 059 477–22 915 724	Loss	16856248	p16.1–p15.31	1	
chr8:121 053 939–145 707 549	Gain	24653611	q24.12–q24.3	1	
chr10:0–30 769 852	Gain	30769852	p15.3–p11.23	1	
chr11:116 158 107–134 452 384	Gain	18294278	q23.3–q25	1	
chr13:99 572 022–114 142 980	Loss	14570959	q32.3–q34	1	
chr14:69 737 214–106 368 585	Gain	36631372	q24.2–q32.33	1	
chrX:140 239 251–152 088 951	Gain	11849700	q27.2–q28	1	

none in controls. Considering both known and novel CNVs, we identified 55 variants that were >50 kb, had DGV overlap of $\leq 75\%$, contained at least one known gene and were not present in the 872 controls (Table 2). Of these, 44 (80%) were singleton events. The CNVs included 1p36.33-p36.32 loss, corresponding to 1p36 microdeletion syndrome, observed in 4/203 cases and 0/872 controls. The 8p23.1 loss encompassing *GATA4* and corresponding to 8p23.1 deletion syndrome was observed as a singleton event, not seen in 872 controls. The 22q11.2 loss corresponding to DGS region and the 7q11.23 loss associated with Williams–Beuren syndrome critical region were also seen in the 872 affected controls without CVMs, and were thus excluded after comparative analysis.

We then evaluated an independent cohort of 511 subjects with CVMs plus ECAs (BCM2). Of the 55 CNVs detected in the BCM1 cohort, 16 were observed in the BCM2 cohort with significant overlap (Table 2). These 16 variants, absent in the primary controls ($n=511$), were also absent in the second control set ($n=2011$) and were validated by FISH studies using BAC clones or long-range PCR (Supplementary Figures 5 and 6). None of these 16 variants, including the inherited events, were found to be present in the healthy 2024 pediatric subjects from CHOP Control (phs000199.v1.p1) assayed using an independent SNP array platform.

Of the 16 loci, 12 (75%) were novel, including 16q24.3 loss of *ANKRD11*, 5p13.2 gain encompassing *C50orf42, NUP155* and *WDR70*, and 2q31.3-q32.1 loss including *SSFA2*, *PPP1R1C*, *PDE1A* and 13q32.3-q33.1 loss. The known genomic disorders including 8p23.1 loss encompassing *GATA4* and 22q11.1-q11.21 gain corresponding to CES region were also observed in the replication set, with none of the 2011 controls harboring these CNVs. Interestingly, the rate at which CNVs from cases in the BCM2 cohort overlapped the 55 CNVs from the BCM1 cohort was significantly greater among cases than for controls (odds ratio$=$1.683; 95% CI, 1.23–2.27; $P=0.0008634$ for CNVs; as well as at the level of individual subjects, odds ratio$=$1.972; 95% CI, 1.367–2.8159; $P=0.0002321$, scoring subjects as positive when they have at least one CNV overlapping the 55 variants). This result remains significant even when excluding the two segmental calls from chromosome 8, including the *GATA4* gene (odds ratio$=$1.588; 95% CI, 1.153– 2.163; $P=0.0049$ for CNVs; odds ratio$=$1.8159, 95% CI, 1.24–2.62; $P=0.001728$ for individual subjects).

Combining the results from the two groups involving 714 CVM cases (BCM1 and BCM2 groups), we identified 16 CNV regions, present in two or more cases and absent in 2883 controls. The most significant variant observed in the affected was at the 22q11.21 locus, involving 705-kb gain (*CECR1)* within the CES region (MIM #115470), seen in 4/714 cases and 0/2883 controls, and 16q24.3 *de novo* loss, observed in 4/714 cases and 0/2883 controls. The 8p23.1 loss (including *GATA4* and 25 other genes) was observed in 3/714 cases and 0/2883 controls. Although the CES region gain and *GATA4* loss on 8p23.1 are known to be associated with CVMs, loss of 16q24.3 region has not previously been described in association with structural cardiac defects.

To precisely map the deletion breakpoints for 16q24.3 CNV in these cases, genome-wide analysis for DNA copy number alterations was performed using NimbleGen array HG18 WG_CGH_v1 with 385 000 oligonucleotide probes. Breakpoint mapping was performed using long-range PCR (Supplementary Methods). In subject 7658 with a 245-kb deletion with AVSD, the proximal breakpoint of the deletion mapped at 87 853 995 within *Alu*Jb and the distal breakpoint at 88 138 377 with an 'A' insertion, likely because of non-homologous end joining (NHEJ) (Figure 1). In subject 0340 with septal defect, the proximal breakpoint of the $\sim$1.5-Mb deletion mapped within poly T-rich tract between 86 406 011 and 86 406 037 and the distal breakpoint mapped within poly A tract between 87 962 518 and 87 962 533. In subject 0585 with a 2-Mb deletion and AVSD defect, the proximal and the distal breakpoints mapped within the SINE/*Alu* repeats at 86 051 611 and 88 133 224, respectively. In subject 4535 with septal defect, the proximal breakpoint of the smallest 139-kb *de novo* deletion mapped between 87 822 867 and 87 862 929 at the 3′ end of *ANKRD11* and the distal breakpoint mapped between 88 001 859 and 88 011 936 within intron 2 of this gene. Subject 2779 with conoventricular VSD and supravalvular pulmonic stenosis was found

Table 2 Rare structural variants in CVM cases

Region	*Cytoband location*	*Event*	*Region length*	*BCM1* (n = *203)*[a]	*BCM2* (n = *511)*[b]
1 chr1:1,957,637–2,702,000	1p36.33–p36.32	CN Loss	744363	4	
2 chr1:2,702,000–4,357,339	1p36.32	CN Loss	1655339	3	
3 chr1:4,357,339–6,675,076	1p36.32–p36.31	CN Loss	2317737	2	
4 chr1:6,675,076–6,836,192	1p36.31	CN Loss	161116	1	
5 chr1:240,360,060–241,078,237	1q43	CN Gain	718177	2	
6 chr1:243,972,455–244,111,443	1q44	CN Loss	138988	1	1
7 chr1:245,409,645–245,486,584	1q44	CN Gain	76939	2	
8 chr2:111,168,238–111,227,870	2q13	CN Gain	59632	1	
9 chr2:128,793,105–129,745,885	2q14.3–q21.1	CN Loss	952780	2	
10 chr2:132,935,612–133,349,584	2q21.2	CN Loss	413972	1	
11 chr2:182,438,570–183,063,803	2q31.3–q32.1	CN Loss	625233	1	1
12 chr3:1,441,845–1,656,796	3p26.3	CN Loss	214951	2	
13 chr3:1,656,796–1,827,988	3p26.3	CN Loss	171192	1	
14 chr3:12,539,818–12,607,512	3p25.1	CN Gain	67694	2	
15 chr4:4,881,718–5,115,187	4p16.2	CN Gain	233469	1	
16 chr4:108,324,837–109,055,741	4q25	CN Gain	730904	1	
17 chr4:171,810,173–172,781,947	4q33–q34.1	CN Loss	971774	1	1
18 chr5:118,223,272–118,380,518	5q23.1	CN Loss	157246	1	
19 chr5:37,454,475–37,668,954	5p13.2	CN Gain	214479	2	1
20 chr5:60,126,916–60,912,562	5q12.1	CN Gain	785646	1	
21 chr6:16,659,770–17,143,875	6p22.3	CN Gain	484105	1	
22 chr6:149,377,071–151,589,034	6q25.1	CN Gain	2211963	1	
23 chr7:99,119,295–99,994,909	7q22.1	CN Loss	875614	1	
24 chr7:153,538,856–153,687,354	7q36.2	CN Loss	148498	1	
25 chr7:154,466,965–155,285,789	7q36.2–q36.3	CN Loss	818824	1	
26 chr8:11,074,772–11,560,684	8p23.1	CN Gain	485912	1	
27 chr8:9,829,449–12,203,210	8p23.1	CN Loss	2373761	1	2
28 chr8:47,069,023–47,692,803	8q11.1	CN Loss	623780	2	
29 chr8:125,309,916–125,617,442	q24.13	CN Gain	307526	1	1
30 chr8:134,955,134–140,713,904	8q24.22–q24.3	CN Gain	5758770	1	
31 chr8:140,713,904–140,800,494	8q24.3	CN Gain	86590	2	
32 chr9:96,866,314–98,473,195	9q22.32–q22.33	CN Loss	1606881	1	
33 chr10:3,338,653–4,416,690	10p15.2–p15.1	CN Loss	1078037	1	1
34 chr10:15,090,628–30,769,852	10p13–p11.23	CN Gain	15679224	1	1
35 chr11:27,042,077–27,194,788	11p14.2	CN Gain	152711	1	1
36 chr11:47,947,514–48,251,768	11p11.2	CN Gain	304254	1	
37 chr11:129,919,450–130,610,355	11q24.3–q25	CN Gain	690905	1	1
38 chr12:85,517,972–86,117,266	12q21.32	CN Loss	599294	1	
39 chr13:31,586,602–31,976,803	13q13.1	CN Gain	390201	1	
40 chr13:109,728,316–110,151,349	13q34	CN Gain	423033	1	1
41 chr13:100,197,463–100,838,781	13q32.3–q33.1	CN Loss	641319	1	2
42 chr14:68,212,155–68,737,172	14q24.1	CN Gain	525017	1	
43 chr14:72,729,142–86,943,554	14q24.2–q31.3	CN Gain	14214412	1	
44 chr16:15,864,047–16,515,701	16p13.11	CN Gain	651654	2	
45 chr16:87,848,043–88,143,163	16q24.3	CN Loss	295120	1	3
46 chr17:1,068,719–1,418,167	17p13.3	CN Loss	349448	1	1
47 chr17:3,577,945–5,010,536	17p13.3–p13.2	CN Loss	1432591	1	1
48 chr17:19,205,934–20,380,184	17p11.2	CN Gain	1174250	1	
49 chr17:57,679,028–58,467,335	17q23.2–q23.3	CN Loss	788307	1	
50 chr17:77,561,334–77,903,212	17q25.3	CN Loss	341878	1	
51 chr18:1,851,694–2,771,690	18p11.32	CN Gain	919996	1	
52 chr18:72,801,963–76,117,153	18q23	CN Gain	3315190	1	
53 chr20:2,997,598–3,815,149	20p13	CN Loss	817551	1	
54 chr20:10,160,813–11,174,283	20p12.2	CN Gain	1013470	1	
55 chr22:15,992,974–16,698,423	22q11.1–q11.21	CN Gain	705449	1	3

Chromosome coordinates and genes are based on hg18 genome assembly.
[a]In 872 controls, 55 variants were not observed.
[b]In 2011 controls, 16/55 variants were not observed.

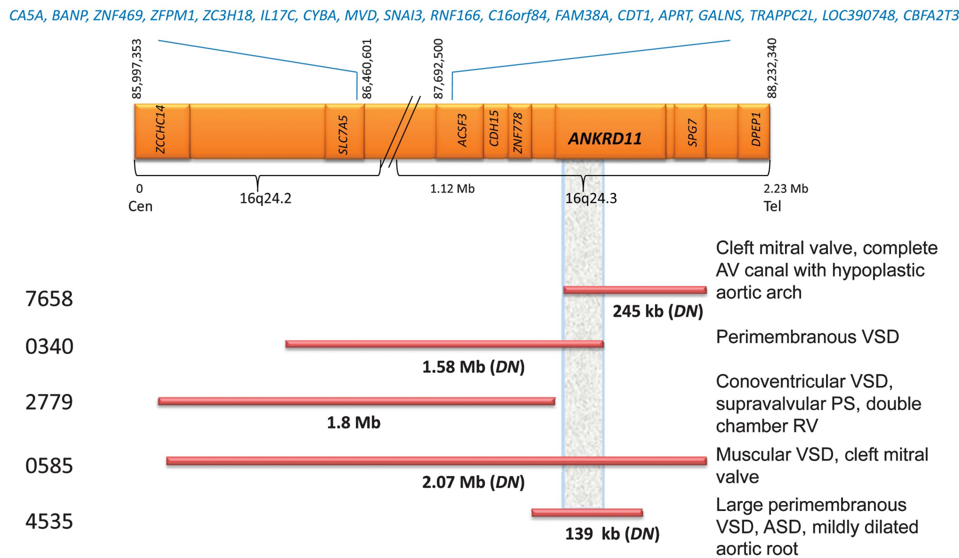

Figure 1 Variable *de novo* (*DN*) deletions of 16q24.3 observed in multiple subjects with CVMs and ECAs. Five subjects with facial dysmorphisms, developmental delay, septal and AVSD defects are shown to have variable-size 16q24.3 microdeletions. In subject 7658, the proximal breakpoint of the deletion maps within *Alu*Jb and the distal breakpoint has an 'A' insertion, likely mediated by NHEJ. In subject 0340, the proximal and the distal breakpoints map within poly T-rich tracts. In subject 0585, the proximal and the distal breakpoints map within the SINE/*Alu* repeats.

to have a 1.8-Mb deletion just outside the region of the overlap observed in these four cases. The heterozygous loss found in the four unrelated individuals with overlapping intervals was *de novo* in all cases. These results are consistent with a strong association of 16q24.3 segmental loss with a syndromic form of CVMs.

Some of the other novel loci found in cases included 5p13.2 gain and 13q32.3-q33.1 loss that were observed in 3/714 cases and 0/2883 controls. Other rare CNVs that were observed in 2/714 cases and 0/2883 controls included *de novo* loss of 1q44 and 2q31.3-q32.1 (Table 3). The range of cardiac defects was highly diverse and encompassed all major classes of malformations, including LVOTOs, AVSDs, RVOTOs, conotruncal, APVR, heterotaxy, septal and complex defects. We also identified CNVs in regions previously implicated in CVMs (DECIPHER syndromes), such as 17p13.3 loss (Miller–Dieker lissencephaly syndrome), 8p23.1 deletion syndrome (including *GATA4*) and 22q11.1–q11.21 gain (CES region). These regions are represented in Table 4.

There were an additional 2860 segmental events found in controls that overlapped events in the cases, involving 232 non-overlapping genomic segments; there were 575 segmental events involving 226 non-overlapping genomic segments found in the controls but absent from the cases. Overall, there were 38 cases with the 16 CNVs shared by the BCM1 and BCM2 cohorts. Both parental samples were available for further studies in 20/38 of these cases. Of these, 18 were found to be *de novo* (Supplementary Figures 5 and 6) and 2 were inherited. The 152-kb gain involving 11p14.2 in one case and the 971-kb loss of 4q33–q34.1 in another individual were both found to be inherited. Echocardiographic studies were not performed in the carrier parents and we cannot exclude subclinical morphologic abnormalities or anatomic variants in these individuals.

Protein-interaction network

To test the hypothesis that the rare inherited and *de novo* recurrent CNVs within the CVM plus ECA cohort are enriched for genes encoding proteins that directly interact with proteins known to be required for normal cardiac development, we selected 276 proteins from Gene Ontology cardiac development categories (Supplementary Tables 3 and 4) and identified protein–protein interactions in the Human Protein Reference Database (HPRD) that involved any of the proteins encoded within the candidate CNVs from our study (Supplementary Methods). We found that within this network, which contained a total of 234 connections, 11 of the candidate proteins have at least one connection with a human cardiac-specific protein (Supplementary Figure 7). Interactions were noted for calmodulin-binding transcription activator, CAMTA2 with NKX2-5; CRK with CRKL and ERBB4; SOX7 with SMAD7 and SMAD5; NEIL2 and PELP1 with EP300; ARRB2 with DVL2, OXYR, and TGFBR3 and GATA4 with NKX2-5, SRF, TBX5, EP300, HAND2 and FOG2 (Supplementary Table 5).

Haploinsufficiency score analysis

As an additional study, to assess the developmental relevance of the variants identified in this study, we tabulated the haploinsufficiency scores compiled by Huang *et al.*[29] for the genes within the 16 identified regions. We found that the genes within these variants scored higher than would be expected by chance (Wilcoxon test *P*-value $P<5\times10^{-5}$). A graphic representation of these results is presented in Supplementary Figure 8. The high haploinsufficiency score for these regions further support our findings that the variants identified in our study are more likely to be pathogenic and are functionally relevant.

DISCUSSION

This study describes 16 rare non-recurrent structural variants, >50 kb in length, present in individuals with CVMs plus ECAs, and not observed in those without CVMs. Several published reports have examined smaller cohorts of patients with complex clinical presentations and described non-recurrent contiguous gene deletion syndromes contributing to CVMs, including 1p36 monosomy,[30] 15q26 deletion,[31–35] Wolf–Hirschhorn syndrome (4p16.3 deletion),[36] Cri-du-chat syndrome (5p15.2 monosomy),[37] Miller–Dieker lissencephaly syndrome (17p13.3 deletion)[38] and 17q23.1q23.2 microdeletion syndrome.[18] Examples where such an approach was successful in identifying specific dosage-sensitive genes include *JAG1* in Allagile

Table 3 New rare CNVs observed in CVM cases with extracardiac anomalies

Genomic regions	*Band*	*Event*	*Region length (bp)*	*CA (n)*[a]	*CO (n)*[a]	*ID*	*Cardiac phenotype*	*Extracardiac phenotype*	*Inh*	*Candidate genes and microRNAs within intervals*
chr1:243,972,455–244,111,443	q44	Loss	138 988	2	0	CVM0103	Complex	Microcephaly, epilepsy	NA	
						CVM1260	Septal	Polysyndactyly, dysgenesis of corpus callosum, microcephaly and developmental delay	DN	
chr2:182,438,570–183,063,803	q31.3–q32.1	Loss	625 233	2	0	CVM3491	LVOTO	Cleft palate, micropenis, hearing loss, coloboma	DN	*SSFA2, PPP1R1C,* ***PDE1A***
						CVM0114	APVR	Microcephaly	DN	
chr4:171,810,173–172,781,947	q33–q34.1	Loss	971 774	2	0	CVM0858	Complex	Hearing loss, epilepsy, developmental delay	NA	
						CVM1144	Septal	Microtia, developmental delay	Pat	
chr5:37,454,475–37,668,954	p13.2	Gain	214 479	3	0	CVM0127	RVOTO	Large for gestational age	NA	*C5Orf42, NUP155, WDR70*
						CVM3443	Septal	Seizures	NA	
						CVM3149	Septal	PHACE syndrome, vascular anomalies, developmental delay	NA	
chr11:27,042,077–27,194,788	p14.2	Gain	152 711	2	0	CVM1581	Septal	Renal cystic disease	Mat	*BBOX1*
						CVM3871	Aortic vasculopathy	Speech delay, facial dysmorphisms	NA	
chr13:100,197,463–100,838,781	q32.3–q33.1	Loss	641 319	3	0	CVM3495	Septal	Agenesis of corpus callosum, growth retardation, facial dysmorphisms, imperforate anus	DN	***NALCN***
						CVM0116	Complex	Agenesis of corpus callosum, Peter's anomaly *versus* microphthalmia	NA	
						CVM2233	Septal	Seizures, developmental delay, ileal atresia	NA	
chr16:87,848,043–88,143,163	q24.3	Loss	295 120	4	0	CVM7658	AVSD	Microcephaly, facial dysmorphisms	DN	*ZNF778,* ***ANKRD11***
						CVM0340	Septal	Facial dysmorphisms, seizures	DN	
						CVM0585	AVSD	Facial dysmorphisms, hearing loss, seizures, corpus callosum hypoplasia, optic nerve hypoplasia	DN	
						CVM4535	Septal	Facial dysmorphisms	DN	

Chromosome coordinates and genes are based on hg18 genome assembly.
Abbreviations: Inh, inheritance; DN, *de novo*; Pat, paternal; Mat, maternal; NA, not available.
Genes in bold font are candidate genes for CVM.
[a]CA represents CVM cases; CO indicates non-CVM controls.

Table 4 CNVs in gene regions previously implicated in CVM with ECA (DECIPHER syndromes)

Genomic regions	DECIPHER syndrome	Band	Event	Region length (bp)	CA (n)[a]	CO (n)[a]	ID	Cardiac phenotype	Extracardiac phenotype	Inh	Genes and microRNAs within intervals
chr8:9,829, 449–12,203, 210	8p23.1 deletion syndrome	p23.1	Loss	237 3761	3	0	CVM3839	Conotruncal	Microcephaly and global developmental delay	DN	*MSRA, UNQ9391, RP1L1, C8orf74,* ***SOX7****, PINX1, hsa-mir-1322, XKR6, hsa-mir-598, MTMR9, AMAC1L2, TDH, C8orf12, FAM167A, BLK,* ***GATA4****, NEIL2, FDFT1, CTSB, DEFB136, DEFB135, DEFB134, DEFB130, ZNF705D, FAM66D, LOC392196, DUB3, FAM86B1*
							CVM0796	Complex	Dysmorphisms, thinning of the corpus callosum	DN	
							CVM3598	Vasculopathy	Short stature, developmental delay	NA	
chr17:1, 068, 719–1, 418, 167	Miller–Dieker syndrome	p13.3	Loss	349 448	2	0	CVM3843	Septal	Lissencephaly, seizures, dysmorphisms	DN	*TUSC5, YWHAE,* ***CRK****, MYO1C, INPP5K, PITPNA*
							CVM9035	Septal	Short stature and failure to thrive	DN	
chr17:3, 577, 945–5, 010, 536	Miller–Dieker syndrome	p13.3 –p13.2	Loss	1 432 591	2	0	CVM3843	Septal	Lissencephaly, seizures, dysmorphisms	DN	*ITGAE, CAMKK1, P2RX1, ATP2A3, ZZEF1, CYB5D2, ANKFY1, UBE2G1, SPNS3, SPNS2, MYBBP1A, GGT6, SMTNL2, ALOX15, PELP1, ARRB2, MED11, CXCL16, ZMYND15, TM4SF5, VMO1, GLTPD2, PSMB6, PLD2, MINK1, CHRNE, GP1BA, SLC25A11, PFN1, ENO3, SPAG7,* ***CAMTA2****, INCA1, GPR172B, ZFP3, ZFP232, USP6*
							CVM0954	Septal	Lissencephaly, seizures	DN	
chr22:15, 992, 974–16, 698, 423	Cat-Eye Syndrome	q11.1–q11.21	Gain	705 449	4	0	CVM0498[b]	Septal	Dysplastic kidney, imperforate anus, ear tags, preauricular ear pits	NA	*CECR5, CECR4,* ***CECR1****, CECR2, SLC25A18, ATP6V1E1, BCL2L13, BID, MICAL3*
							CVM1421[c]	RVOTO	Imperforate anus, preauricular sinuses	DN	
							CVM1804[c]	Heterotaxy	Biliary atresia, preauricular ear pits and tags	DN	
							CVM0092	Conotruncal	Seizures, preauricular ear pits, developmental delay	NA	

Chromosome coordinates and genes are based on hg18 genome assembly.
Abbreviations: DN, *de novo*; Pat, paternal; Mat, maternal; NA, not available.
[a]CA represents CVM cases, and CO indicates non-CVM controls.
[b]Partial tetrasomy for 22q11.2.
[c]Mosaicism for 22q11.2 gain.
Genes in bold font are candidate genes for CVM (GATA4 mutations are known to cause CVM).

syndrome,[39] the *LIS1* gene in Miller–Dieker lissencephaly syndrome[40] and *TAB2* haploinsufficiency in 6q24-q25 deletion.[41] Using the affected cohort for comparison in our study provides a potentially useful tool for defining subregions affected by genomic disorders that may be causally associated with CVMs. Although it is unlikely that any significant CVM was present in the controls who were evaluated clinically but did not have imaging studies, such occurrences would only reduce power and would not lead to false-positive associations in our analysis.

One of the most significant loci enriched in the CVM cohort in our study is the *de novo* copy number loss of 16q24.3. The phenotype of subject 0585 with AVSD was previously included in the description by Willemsen *et al*[42] of four individuals with autism spectrum disorder, facial dysmorphisms and brain abnormalities. Mutations in the *ANKRD11* gene within this 16q24.3 interval have been described in patients with KBG syndrome, characterized by intellectual disability, skeletal malformations and macrodontia.[43] Congenital heart defects including VSD, partial atrioventricular canal defect and stenosis of the left pulmonary artery have also been reported in some patients with KBG syndrome.[44–46] Rare microdeletions involving *ANKRD11* and the flanking genes *ZNF778* and *CDH15* were observed most frequently in our CVM cohort. The 2-Mb deletion in patient 0585 is mediated by *Alu* repeats; however, most deletions observed in this region are non-recurrent and are not mediated by segmental

duplications. The individuals in our study also have ECAs with neurocognitive deficits and facial dysmorphisms (Table 2), as described in other studies.[42,47]

CES caused by dup 22q11 is frequently associated with heart defects, particularly total anomalous pulmonary venous return or tetralogy of Fallot.[48] Although the critical region responsible for CVM has been ill-defined in children affected with CES, Riazi *et al.*[49] have shown that overexpression of *CECR1* causes heart defects in mice, including ASD. Our analysis identified a 705-kb region within this genomic interval that showed copy number gain in 4/714 cases and 0/2883 controls. These cases define a region that includes the *CECR1* gene and other genes such as *CECR5*, *CECR4*, *CECR2*, *SLC25A18*, *ATP6V1E1*, *BCL2L13*, *BID* and *MICAL3*. The fact that CVMs occurred in all these cases suggests that genes within this narrowed interval contribute to CVMs observed in CES. Parental samples were available for 2/4 cases and both were found to be *de novo* by FISH analyses (Supplementary Figure 6). The 5p13.2 gain involving the *NUP155* and *WDR70* genes was observed in 3/714 cases and none in controls. Parents were not available for further analysis in these cases.

We observed two cases, affected with APVR and LVOTO respectively, with *de novo* deletion of 2q31.3-q32.1. This region encompasses two genes, *PPP1R1C* and *PDE1A*. *PDE1A*, phosphodiesterase 1A, is calmodulin dependent, expressed in brain and heart[50] and has been shown to regulate cardiac hypertrophy in animal model.[51] Deletion on 13q32.3-q33 was observed in three individuals, two with septal defects and one with complex CVM phenotype. *NALCN*, a neuronal voltage-independent nonselective cation channel gene within this interval, is known to be expressed in heart.[52]

Non-recurrent variable deletions of 17p13.3 are associated with Miller–Dieker lissencephaly syndrome, responsible for lissencephaly in the affected individuals. Although cardiac abnormalities including PDA and septal defects are described in a subset of these cases, no cardiac-specific critical region has been delineated within the larger 17p13.3 region in this syndrome. In our study, septal defects were observed in both individuals with the *de novo* non-recurrent loss of 17p13.3. The 1.4 Mb enriched region within this region (2/714 cases and 0/2883 controls) includes two candidate genes, *PELP1* and *CAMTA2*, calmodulin-binding transcription activator 2. *CAMTA2* is preferentially expressed in heart and brain and acts as a coactivator of *Nkx2-5*.[53] Although point mutations in *CAMTA2* have not yet been described in humans, the gene has been shown to be necessary for maximal hypertrophic response to stresses in mice.[53] The causal role of *CAMTA2* in enhancing the ability of *Nkx2-5* to activate the ANF promoter[53] and the enrichment of the 17p13.3 CNV in cases support the involvement of this gene in human cardiac development.

Our protein-interaction network analysis identified direct interactions of several of the candidate genes within the 16 structural variants with the annotated human cardiac genes. The variant on 8p23.1 includes *SOX7* in addition to *GATA4*. *SOX7* is a transcription factor, which is essential for cardiac development in *Xenopus*,[54] and has been shown to have direct interactions with both SMAD7 and SMAD5, from our HPRD network analysis. Two of the candidate genes, *NEIL2* (expressed in the heart)[55] and *PELP1*, directly interact with the protein responsible for Rubinstein–Taybi syndrome (RSTS), *EP300*.[56] Approximately one-third of the individuals with RSTS (MIM #180849) have an associated CVM. Our data suggest that the biological mechanisms underlying the cardiac phenotype in many of these individuals may have a unifying basis, likely affecting common developmental pathways. More extensive studies and further analyses of the CNVs are required to substantiate this resolutely.

A review of CNVs in relatively large studies of CVMs[13,22,57,58] showed few subjects with larger segmental aneusomies overlapping the 16 structural variants observed in our study. In a study of 60 individuals with CVMs and ECAs, Thienpont *et al.*[13] described one subject with a derivative chromosome 13 with TOF and microcephaly. The smaller 641-kb variant on 13q32.3-q33 in our study overlaps this region described in that study. Similarly, intersecting the 1q44 variant in our study encompassing 126 kb, a large 12.3-Mb deletion of 1q44 was described in another patient with AVSD and an unbalanced chromosomal translocation by Richard *et al.*[58] However, most CNVs described in these reports are unique to respective studies.

One limitation of our study is that the analysis is insensitive to copy number variants that may be associated with other developmental phenotypes in which the penetrance of the cardiac phenotype is low. A locus may be associated with CVMs, but because of incomplete penetrance will not be ascertained in this study. Nevertheless, the strength of using this approach is the specificity of the CNV association with cardiac malformations.

In summary, our results show that 16q24.3 *de novo* loss is observed in individuals with CVMs. A number of genes, known or suspected to be involved in cardiogenesis, have been brought to light with regard to their roles in cardiac malformation in ECA cases, such as *SOX7* on 8p23.1 and *PDE1A* on 2q31.3q32.1. The study provides insight into the cardiac-specific critical regions of some of the well-established genomic disorders of CVMs including duplication within the CES region and deletions associated with 8p23.1 and 17p13.3 genomic regions.

CONFLICT OF INTEREST

Multiple authors are based in the Department of Molecular and Human Genetics at Baylor College of Medicine, which derives revenue from molecular diagnostic testing (Medical Genetics Laboratories).

ACKNOWLEDGEMENTS

We thank the families for participating in the study. We thank Zhilian Xia for providing technical support. The support for this work was provided by the National Institutes of Health (RO1-HL091771) to JWB, Doris Duke Charitable Foundation and Gillson Longenbaugh Foundation to SRL and March of Dimes support (MOD: 1-FY10-401) to SRL and SMW. Pawel Stankiewicz was supported in part by Grant R13-0005-04/2008 from the Polish Ministry of Science and Higher Education.

1 Hoffman JI, Kaplan S, Liberthson RR: Prevalence of congenital heart disease. *Am Heart J* 2004; **147**: 425–439.
2 Ferencz C, Boughman JA: Congenital heart disease in adolescents and adults. Teratology, genetics, and recurrence risks. *Cardiol Clin* 1993; **11**: 557–567.
3 Ferencz C, Boughman JA, Neill CA, Brenner JI, Perry LW: Congenital cardiovascular malformations: questions on inheritance. Baltimore-Washington Infant Study Group. *J Am Coll Cardiol* 1989; **14**: 756–763.
4 Botto LD, Correa A, Erickson JD: Racial and temporal variations in the prevalence of heart defects. *Pediatrics* 2001; **107**: E32.
5 Calzolari E, Garani G, Cocchi G *et al*: Congenital heart defects: 15 years of experience of the Emilia-Romagna Registry (Italy). *Eur J Epidemiol* 2003; **18**: 773–780.
6 Pradat P, Francannet C, Harris JA, Robert E: The epidemiology of cardiovascular defects, part I: a study based on data from three large registries of congenital malformations. *Pediatr Cardiol* 2003; **24**: 195–221.
7 Bernier PL, Stefanescu A, Samoukovic G, Tchervenkov CI: The challenge of congenital heart disease worldwide: epidemiologic and demographic facts. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu* 2010; **13**: 26–34.
8 Hoffman JI, Kaplan S: The incidence of congenital heart disease. *J Am Coll Cardiol* 2002; **39**: 1890–1900.
9 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.
10 Richards AA, Garg V: Genetics of congenital heart disease. *Curr Cardiol Rev* 2010; **6**: 91–97.

11 van der Bom T, Zomer AC, Zwinderman AH, Meijboom FJ, Bouma BJ, Mulder BJ: The changing epidemiology of congenital heart disease. *Nat Rev Cardiol* 2011; **8**: 50–60.
12 Lupski JR: Genomic disorders ten years on. *Genome Med* 2009; **1**: 42.
13 Thienpont B, Mertens L, de Ravel T *et al*: Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients. *Eur Heart J* 2007; **28**: 2778–2784.
14 Moerman P, Goddeeris P, Lauwerijns J, Van der Hauwaert LG: Cardiovascular malformations in DiGeorge syndrome (congenital absence of hypoplasia of the thymus). *Br Heart J* 1980; **44**: 452–459.
15 Fontaine JL, Vernant P, Graveleau D, Lagardere B, Elchardus JF: [Elfin facies, mental retardation and cardiovascular anomalies (Williams and Beuren's syndrome). Report of two cases]. *Ann Pediatr (Paris)* 1976; **23**: 37–42.
16 Edelman EA, Girirajan S, Finucane B *et al*: Gender, genotype, and phenotype differences in Smith-Magenis syndrome: a meta-analysis of 105 cases. *Clin Genet* 2007; **71**: 540–550.
17 Shaw-Smith C, Pittman AM, Willatt L *et al*: Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 2006; **38**: 1032–1037.
18 Ballif BC, Theisen A, Rosenfeld JA *et al*: Identification of a recurrent microdeletion at 17q23.1q23.2 flanked by segmental duplications associated with heart defects and limb abnormalities. *Am J Hum Genet* 2010; **86**: 454–461.
19 Lindsay EA, Vitelli F, Su H *et al*: Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* 2001; **410**: 97–101.
20 Curran ME, Atkinson DL, Ewart AK, Morris CA, Leppert MF, Keating MT: The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. *Cell* 1993; **73**: 159–168.
21 Houle D: How should we explain variation in the genetic variance of traits? *Genetica* 1998; **102–103**: 241–253.
22 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.
23 Belmont JW: Heart patterning and congenital defects; in Moody SA (ed): *Principles of Developmental Biology*. Amsterdam: Elsevier, 2007; pp 698-720.
24 Botto LD, Lin AE, Riehle-Colarusso T, Malik S, Correa A: Seeking causes: classifying and evaluating congenital heart defects in etiologic studies. *Birth Defects Res A Clin Mol Teratol* 2007; **79**: 714–727.
25 Boone PM, Bacino CA, Shaw CA *et al*: Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 2010; **31**: 1326–1342.
26 Shao L, Shaw CA, Lu XY *et al*: Identification of chromosome abnormalities in subtelomeric regions by microarray analysis: a study of 5380 cases. *Am J Med Genet A* 2008; **146A**: 2242–2251.
27 Ou Z, Kang SH, Shaw CA *et al*: Bacterial artificial chromosome-emulation oligonucleotide arrays for targeted clinical array-comparative genomic hybridization analyses. *Genet Med* 2008; **10**: 278–289.
28 Kuang SQ, Guo DC, Prakash SK *et al*: Recurrent chromosome 16p13.1 duplications are a risk factor for aortic dissections. *PLoS Genet* 2011; **7**: e1002118.
29 Huang N, Lee I, Marcotte EM, Hurles ME: Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010; **6**: e1001154.
30 Heilstedt HA, Ballif BC, Howard LA *et al*: Physical map of 1p36, placement of breakpoints in monosomy 1p36, and clinical characterization of the syndrome. *Am J Hum Genet* 2003; **72**: 1200–1212.
31 Mosca AL, Pinson L, Andrieux J *et al*: Refining the critical region for congenital diaphragmatic hernia on chromosome 15q26 from the study of four fetuses. *Prenat Diagn* 2011; **31**: 912–914.
32 Davidsson J, Collin A, Bjorkhem G, Soller M: Array based characterization of a terminal deletion involving chromosome subband 15q26.2: an emerging syndrome associated with growth retardation, cardiac defects and developmental delay. *BMC Med Genet* 2008; **9**: 2.
33 Klaassens M, Galjaard RJ, Scott DA *et al*: Prenatal detection and outcome of congenital diaphragmatic hernia (CDH) associated with deletion of chromosome 15q26: two patients and review of the literature. *Am J Med Genet A* 2007; **143A**: 2204–2212.
34 Slavotinek AM, Moshrefi A, Davis R *et al*: Array comparative genomic hybridization in patients with congenital diaphragmatic hernia: mapping of four CDH-critical regions and sequencing of candidate genes at 15q26.1-15q26.2. *Eur J Hum Genet* 2006; **14**: 999–1008.
35 Tumer Z, Harboe TL, Blennow E, Kalscheuer VM, Tommerup N, Brondum-Nielsen K: Molecular cytogenetic characterization of ring chromosome 15 in three unrelated patients. *Am J Med Genet A* 2004; **130A**: 340–344.
36 Zollino M, Murdolo M, Marangi G *et al*: On the nosology and pathogenesis of Wolf-Hirschhorn syndrome: genotype-phenotype correlation analysis of 80 patients and literature review. *Am J Med Genet C Semin Med Genet* 2008; **148C**: 257–269.
37 Mainardi PC, Perfumo C, Cali A *et al*: Clinical and molecular characterisation of 80 patients with 5p deletion: genotype-phenotype correlation. *J Med Genet* 2001; **38**: 151–158.
38 Cardoso C, Leventer RJ, Ward HL *et al*: Refinement of a 400-kb critical region allows genotypic differentiation between isolated lissencephaly, Miller-Dieker syndrome, and other phenotypes secondary to deletions of 17p13.3. *Am J Hum Genet* 2003; **72**: 918–930.
39 Oda T, Elkahloun AG, Pike BL *et al*: Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nat Genet* 1997; **16**: 235–242.
40 Reiner O, Carrozzo R, Shen Y *et al*: Isolation of a Miller-Dieker lissencephaly gene containing G protein beta-subunit-like repeats. *Nature* 1993; **364**: 717–721.
41 Thienpont B, Zhang L, Postma AV *et al*: Haploinsufficiency of TAB2 causes congenital heart defects in humans. *Am J Hum Genet* 2010; **86**: 839–849.
42 Willemsen MH, Fernandez BA, Bacino CA *et al*: Identification of ANKRD11 and ZNF778 as candidate genes for autism and variable cognitive impairment in the novel 16q24.3 microdeletion syndrome. *Eur J Hum Genet* 2010; **18**: 429–435.
43 Sirmaci A, Spiliopoulos M, Brancati F *et al*: Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. *Am J Hum Genet* 2011; **89**: 289–294.
44 Brancati F, D'Avanzo MG, Digilio MC *et al*: KBG syndrome in a cohort of Italian patients. *Am J Med Genet A* 2004; **131**: 144–149.
45 Devriendt K, Holvoet M, Fryns JP: Further delineation of the KBG syndrome. *Genet Couns* 1998; **9**: 191–194.
46 Rivera-Vega MR, Leyva Juarez N, Cuevas-Covarrubias SA, Kofman-Alfaro SH: Congenital heart defect and conductive hypoacusia in a patient with the KBG syndrome. *Clin Genet* 1996; **50**: 278–279.
47 Isrie M, Hendriks Y, Gielissen N *et al*: Haploinsufficiency of ANKRD11 causes mild cognitive impairment, short stature and minor dysmorphisms. *Eur J Hum Genet* 2012; **20**: 131–133.
48 Rosias PR, Sijstermans JM, Theunissen PM *et al*: Phenotypic variability of the cat eye syndrome. Case report and review of the literature. *Genet Couns* 2001; **12**: 273–282.
49 Riazi AM, Van Arsdell G, Buchwald M: Transgenic expression of CECR1 adenosine deaminase in mice results in abnormal development of heart and kidney. *Transgenic Res* 2005; **14**: 333–336.
50 Michibata H, Yanaka N, Kanoh Y, Okumura K, Omori K: Human Ca2+/calmodulin-dependent phosphodiesterase PDE1A: novel splice variants, their specific expression, genomic organization, and chromosomal localization. *Biochim Biophys Acta* 2001; **1517**: 278–287.
51 Miller CL, Oikawa M, Cai Y *et al*: Role of Ca2+/calmodulin-stimulated cyclic nucleotide phosphodiesterase 1 in mediating cardiomyocyte hypertrophy. *Circ Res* 2009; **105**: 956–964.
52 Swayne LA, Mezghrani A, Varrault A *et al*: The NALCN ion channel is activated by M3 muscarinic receptors in a pancreatic beta-cell line. *EMBO Rep* 2009; **10**: 873–880.
53 Song K, Backs J, McAnally J *et al*: The transcriptional coactivator CAMTA2 stimulates cardiac growth by opposing class II histone deacetylases. *Cell* 2006; **125**: 453–466.
54 Zhang C, Basta T, Klymkowsky MW: SOX7 and SOX18 are essential for cardiogenesis in Xenopus. *Dev Dyn* 2005; **234**: 878–891.
55 Hazra TK, Kow YW, Hatahet Z *et al*: Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J Biol Chem* 2002; **277**: 30417–30420.
56 Bamforth SD, Braganca J, Farthing CR *et al*: Cited2 controls left-right patterning and heart development through a Nodal-Pitx2c pathway. *Nat Genet* 2004; **36**: 1189–1196.
57 Iascone M, Ciccone R, Galletti L *et al*: Identification of de novo mutations and rare variants in hypoplastic left heart syndrome. *Clin Genet* 2012; **81**: 542–554.
58 Richards AA, Santos LJ, Nichols HA *et al*: Cryptic chromosomal abnormalities identified in children with congenital heart disease. *Pediatr Res* 2008; **64**: 358–363.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)