Rare Genomic Copy Number Variants Implicate New Candidate Genes for Bicuspid Aortic Valve

Authors: Steven G. Carlisle[1], Hasan Albasha[2], Hector Michelena[3], Anna Sabate-Rotes[4], Lisa Bianco[4], Julie De Backer[5], Laura Muiño Mosquera[6], Anji T. Yetman[7], Malenka M Bissell[8], Maria Grazia Andreassi[9], Ilenia Foffa[9], Dawn S. Hui[10], Anthony Caffarelli[11], Yuli Y. Kim[12], Dong-Chuan Guo[1], Rodolfo Citro[13], Margot De Marco[14], Justin T. Tretter[15], Kim L. McBride[16], &EBAV Investigators, &BAVCon Investigators, Dianna M. Milewicz[1], Simon C. Body[17], Siddharth K. Prakash[1]

[1]Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, Texas.

[2]UCD School of Medicine, University College Dublin, Belfield, Dublin 4, Ireland.

[3]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota.

[4]Department of Pediatric Cardiology, Hospital Vall d'Hebron, Facultad de Medicina, Universidad Autònoma Barcelona, Barcelona, Spain.

[5]Centre for Medical Genetics, Ghent University Hospital, Ghent, Belgium; VASCERN HTAD European Reference Centre, Belgium; Department of Pediatrics, Division of Pediatric Cardiology, Ghent University Hospital, Ghent, Belgium; Department of Cardiology, Ghent University Hospital, Ghent, Belgium.

[6]Centre for Medical Genetics, Ghent University Hospital, Ghent, Belgium

[7]Children's Hospital and Medical Center, University of Nebraska, Omaha, Nebraska.

[8]Deparmentt of Biomedical Imaging Science, Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, United Kingdom.

[9]Consiglio Nazionale delle Richerche (CNR), Instituto di Fisiologia Clinica, Pisa, Italy.

[10]Department of Cardiothoracic Surgery, University of Texas Health Science Center San Antonio, Texas.

[11]Department of Cardiothoracic Surgery, Stanford University School of Medicine, Stanford, California.

[12]Division of Cardiovascular Medicine, The Hospital of the University of Pennsylvania, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania; Philadelphia Adult Congenital Heart Center, The Children's Hospital of Philadelphia, Perelman Center for Advanced Medicine, Penn Medicine, Philadelphia, Pennsylvania.

[13]Cardio-Thoracic and Vascular Department, University Hospital "San Giovanni di Dio e Ruggi d'Aragona," Salerno, Italy.

[14]Department of Medicine, Surgery and Dentistry Schola Medica Salernitana, University of Salerno, Baronissi, Italy.

[15]Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio.

[16]Division of Human Genetics, Ohio State University Wexner Medical Center, Columbus, Ohio.

[17]Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, Massachusetts.

Corresponding author:

Siddharth Prakash, MD, PhD, Department of Internal Medicine, John P. and Katherine G. McGovern Medical School, University of Texas Health Science Center at Houston, 6431 Fannin Street, MSB 6.116, Houston, Texas, USA

E-mail: Siddharth.K.Prakash@uth.tmc.edu; Office: 713-500-7003

[&]Complete lists of EBAV and BAVCon Investigators are provided in the Acknowledgments.

Abstract

Bicuspid aortic valve (BAV), the most common congenital heart defect, is a major cause of aortic valve disease requiring valve interventions and thoracic aortic aneurysms predisposing to acute aortic dissections. The spectrum of BAV ranges from early onset valve and aortic complications (EBAV) to sporadic late onset disease. Rare genomic copy number variants (CNVs) have previously been implicated in the development of BAV and thoracic aortic aneurysms. We determined the frequency and gene content of rare CNVs in EBAV probands (n = 272) using genome-wide SNP microarray analysis and three complementary CNV detection algorithms (cnvPartition, PennCNV, and QuantiSNP). Unselected control genotypes from the Database of Genotypes and Phenotypes were analyzed using identical methods. We filtered the data to select large genic CNVs that were detected by multiple algorithms. Findings were replicated in cohorts with late onset sporadic disease (n = 5040). We identified 34 large and rare (< 1:1000 in controls) CNVs in EBAV probands. The burden of CNVs intersecting with genes known to cause BAV when mutated was increased in case-control analysis. CNVs intersecting with *GATA4* and *DSCAM* were enriched in cases, recurrent in other datasets, and segregated with disease in families. In total, we identified potentially pathogenic CNVs in 8% of EBAV cases, implicating alterations of candidate genes at these loci in the pathogenesis of BAV.

Author Summary

Bicuspid aortic valve (BAV) is the most common form of congenital heart disease and can lead to long-term complications such as aortic stenosis, aortic regurgitation, or thoracic aortic aneurysms. Most BAV-related complications arise in late adulthood, but 10-15% of individuals with BAV develop early onset complications before age 30. Copy number variants (CNVs) are genomic structural variations that have been previously implicated in some types of congenital heart disease, including BAV. Here we demonstrate that individuals with early onset complications of BAV are enriched for specific rare CNVs compared to individuals with late-onset BAV disease. We also describe novel CNVs involving *DSCAM*, a gene on chromosome 21 that has not previously been associated with the development of BAV. These results may lead to improved risk stratification and targeted therapies for BAV patients.

Introduction

Copy number variants (CNVs) have been implicated as causes or modifiers of many human diseases [1]. Specifically, large genomic CNVs are significantly enriched in cohorts with developmental delay or congenital abnormalities, and the severity of phenotypes has been correlated with the burden of rare CNVs [2]. These observations show that large, rare, *de novo* CNVs are likely to be pathogenic and can exert clinically relevant effects on disease pathogenesis [3-4].

Congenital heart disease (CHD) has a worldwide prevalence of 8.2 per 1000 live births [5]. CNVs have been implicated in both syndromic and non-syndromic forms of CHD [6-10]. The pathogenicity and penetrance of CNVs was initially established for clinical syndromes such as velocardiofacial syndrome, Turner syndrome, or Williams–Beuren syndrome, which involve chromosomal or megabase scale duplications or deletions, but has since been expanded to include additional CHD subtypes [10]. CNVs contribute to 10% of all CHD cases and up to 25% of cases with extracardiac anomalies or other syndromic features [11]. The role of pathogenic CNVs affecting genes that are known to cause CHD when mutated, such as *GATA4* and *TBX1*, has been established [12]. Furthermore, population-level analysis has consistently demonstrated an increased burden of CHD in carriers of CNVs at specific genomic hotspots compared to controls, displaying the pathogenic potential of rare or *de novo* CNVs [12-14].

Bicuspid Aortic Valve (BAV) is the most common congenital heart malformation with a population prevalence of 0.5 – 2% [15]. BAV predisposes to aortic valve stenosis and thoracic aortic aneurysms and is associated with other left ventricular outflow tract lesions such as mitral valve disease and coarctation [16]. The high heritability of BAV was demonstrated in first- and

second-degree relatives, who are more than ten times more likely to be diagnosed with BAV compared to matched controls [17]. BAV can occur as an isolated congenital lesion or as part of a clinical syndrome. For example, the prevalence of BAV is increased in Velocardiofacial, Loeys-Dietz, Kabuki, and Turner syndromes. Pathogenic variants of several genes are implicated in familial non-syndromic BAV, which is typically inherited as an autosomal dominant trait with reduced penetrance and variable expressivity. There is strong cumulative evidence that *GATA4, GATA6, NOTCH1, ROBO4, SMAD4*, *MUC4*, and *SMAD6* each contribute to a small percentage of non-syndromic BAV cases. Phenotypic expression of BAV disease ranges from incidental discovery in late adulthood to neonatal or childhood onset of complications. In comparison to patients with later disease onset, younger BAV cohorts tend to present with syndromic features or complex congenital malformations that are more likely to have a genetic cause, thereby increasing the power of association studies to discover clinically relevant CNVs [18]. Recently, we identified recurrent rare CNVs that were enriched for cardiac developmental genes in a young cohort with early-onset thoracic aortic aneurysms or acute aortic dissections [19].

We hypothesize that large rare genomic CNVs contribute to early onset complications of BAV. Consistent with previous observations, we predict that the burden and penetrance of rare CNVs will be increased in individuals with early onset disease when compared to elderly sporadic BAV cases and population controls. Identification of novel pathogenic CNVs can provide new insights into the genetic complexity of BAV and may be useful for personalized risk stratification or clinical guidance based on the specific recurrent CNV [20]. Therefore, we set out to describe the burden and penetrance of rare CNVs in a young cohort with early onset complications of BAV disease (EBAV).

Materials and Methods

The study protocol was approved by the Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston (HSC-MS-11-0185). After written informed consent, we enrolled 272 probands of European ancestry with early onset BAV disease (EBAV), which we defined as individuals with BAV who were under the age of 30 at the time of first clinical event. Clinical events were defined as aortic replacement, aortic valve surgery, aortic dissection, moderate or severe aortic stenosis or aortic regurgitation, large aneurysm ($Z > 4.5$), or intervention for BAV-related conditions. Those with hypoplastic left heart, known genetic mutations, genetic syndromes, or complex congenital heart disease were excluded. Affected and unaffected family members of probands were included in this cohort for a total of 544 individuals in 293 families (26 trios and 16 multiplex families). Samples were collected and genotyped similar to our previous study [21]. For comparison, we analyzed a cohort of older individuals of European ancestry with sporadic BAV disease selected from the International BAV Consortium (Table 1) [22].

Table 1. Summary of Case Cohorts.

Cohort	Source	Sample Size	Array
EBAV	UTHealth Houston	544	Illumina GSA-24v1.0/2.0
BAVGWAS	International BAV Consortium	5040	Illumina GSA-24v3.0

Cohort: name of case cohort; EBAV: family-based cohort selected for early onset complications of bicuspid aortic valve (BAV); BAVGWAS: unrelated probands with sporadic BAV disease. Source: origin of genotypes; Array: microarray used for genotyping.

Phenotypes were derived from record review with confirmation of image data whenever possible [23-24]. The computational pipeline for CNV analysis of Illumina single nucleotide polymorphism (SNP) array data included three independent CNV detection algorithms (Fig 1).

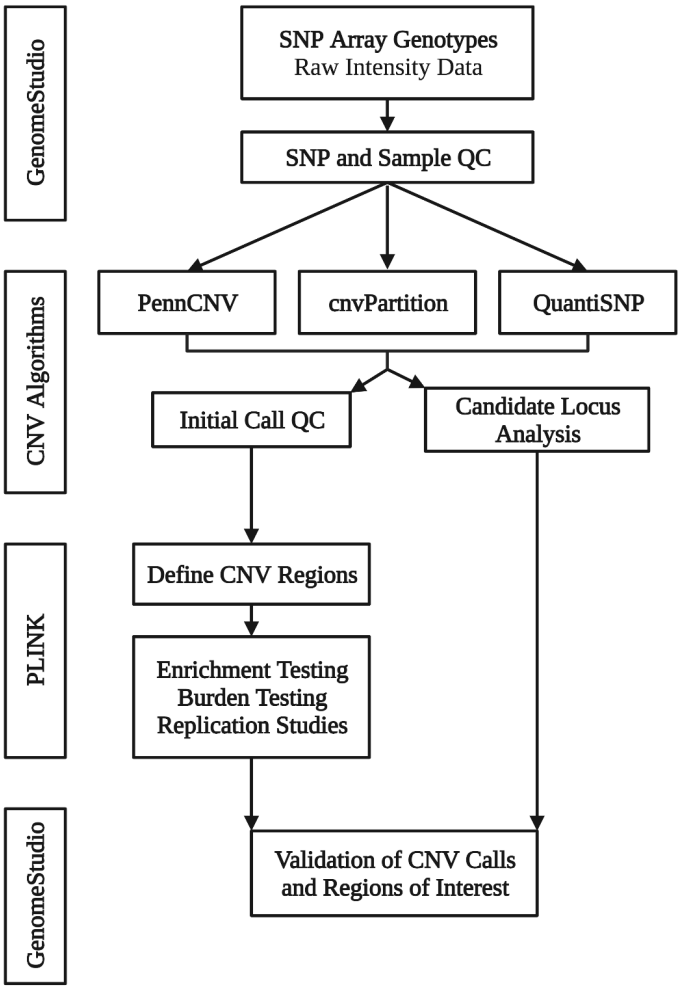


Fig 1. Overview of Pipeline for CNV Identification and Validation.
SNP, single nucleotide polymorphism. QC, Quality control. CNV, copy number variant. The software and algorithms used for the analysis are provided in boxes to the left of the corresponding steps. Illumina raw signal intensity data was trimmed and exported using GenomeStudio. The intensity data was then analyzed with three different CNV calling algorithms (PennCNV [25], cnvPartition, and QuantiSNP [26]) to generate initial CNV calls and sample-level statistics. Sample-level quality control analysis was performed using PennCNV. PLINK [27] toolset was used to define CNV regions from initial CNV calls for subsequent burden testing, enrichment studies, and replication studies. The initial CNV calls were individually screened for CNVs intersecting with candidate loci, which we defined as genes implicated in bicuspid aortic valve disease and those discovered in our enrichment studies. CNVs of interest were then validated in GenomeStudio.

GenomeStudio was used to exclude samples with indeterminate sex or more than 5% missing genotypes, and single nucleotide polymorphisms (SNPs) with GenTrain = 0. Principal component analysis was used to remove outliers that did not cluster with European ancestry. Only SNPs common to all microarray platforms were included.

Three independent algorithms (PennCNV, cnvPartition, and QuantiSNP) were used to generate CNV calls and sample-level quality statistics from SNP intensity data. PennCNV and QuantiSNP were run on Unix clusters and cnvPartition data were exported from GenomeStudio. The analysis was run using default configurations.

PennCNV was used to generate QC data and remove CNV calls that intersect with polymorphic genomic regions. Samples that met any of the following criteria were excluded: standard deviation of the LogR ratio (obtained from PennCNV) > 0.35 or number of CNVs > 2 standard deviations above the mean for each data set. CNV calls less than 20 kilobase pairs and/or spanned by less than 6 SNP probes were excluded. The overlap function for rare CNVs in PLINK was used to construct CNV regions (CNVRs) and adjacent regions were merged using PennCNV.

LogR ratio (LRR) and B allele frequency (BAF) data at CNVRs and calls of interest were visualized in GenomeStudio for validation. For segregation analysis, GenomeStudio was used to determine the presence of CNVs in relatives.

A total of 22,014 unselected control Illumina Genotypes obtained from the Database of Genotypes and Phenotypes were analyzed using identical methods (Table in S1Table). Cohorts were paired as follows for case-control analysis based on the concordance of sample-level quality control statistics (mean number of CNV calls and mean standard deviation of the LogR Ratio): EBAV and WLS, BAVGWAS and HRS.

PLINK was used to catalog CNV calls and perform burden and enrichment studies. Case - control burden tests were restricted to large (250 - 5000 kilobase pairs), rare (occurring in less than 1 in 1000 samples; total of cases and controls), and validated CNV calls in EBAV probands. Genome Reference Consortium Human Build 37 [28] was used for CNV annotation.

Results

Compared to BAVGWAS probands, EBAV probands were significantly younger at diagnosis, had more frequent co-existing congenital heart and vascular lesions, and underwent more frequent valve or aortic operations. A phenotype summary of the EBAV and BAVGWAS Cohorts is provided in Table 2.

Table 2. Characteristics of EBAV and BAVGWAS Probands.

	EBAV (n = 279)	BAVGWAS (n = 3141)
Female (%)	33	29
Age at diagnosis (years)	17 □ 13	52 □ 16
TAA (%)	20	37
Predominant AR (%)	12	40
Predominant AS (%)	20	37
Other Lesions (%)	53	1
Aortic Replacement (%)	27	16
Aortic Valve Surgery (%)	40	16

N: number of cases; □, standard deviation; TAA, thoracic aortic aneurysm; AR: aortic regurgitation; AS, aortic stenosis; Other Lesions, other congenital heart malformations (primarily coarctation or ventricular septal defect). We had phenotype information for 279 EBAV probands but did not have access to genotype information for all samples.

CNV analysis is summarized in Table 3. The percentages of individuals with large and rare CNV regions were relatively consistent throughout datasets. The prevalence of large and rare CNVs, specifically large genomic deletions, was increased in EBAV cases compared to controls (Table S2).

Table 3. Summary of CNV Calls for EBAV Cohort.

	RATE	P^E	P^B	PROP	P^E	P^B	TOT	P^E	P^B	AVG	P^E	P^B
Large	0.51	$1x10^{-7}$	1	0.17	1	1	2648	$1x10^{-7}$	$2.2x10^{-2}$	690	$1x10^{-7}$	$6x10^{-5}$
Rare	0.36	0.79	1	0.21	1	1	426	$6.1x10^{-4}$	0.87	288	$4.1x10^{-2}$	0.6
Duplications	$7.1x10^{-2}$	0.96	1	$6.8x10^{-2}$	0.96	1	648	0.25	0.98	615	0.18	0.98
Deletions	0.11	$1x10^{-7}$	1	$4.8x10^{-2}$	$1.9x10^{-2}$	1	1477	$1.1x10^{-3}$	$4.1x10^{-2}$	608	0.23	$1.7x10^{-2}$

Large: CNV regions between 250 Kb and 5 Mb in length. Rare: occur in fewer than 1 in 1000 individuals; Rate: number of CNVs per individual; Prop: proportion of samples with one or more CNVs; TOT: total length of all CNVs in kilobases; AVG: mean CNV length. p^E, p-value for EBAV cohort in respective category. p^B, p-value for

BAVGWAS in respective category. Tests are 1-sided with 100,000 permutations. A subset of CNV calls from the EBAV and BAVGWAS datasets were validated by examining GenomeStudio plots. In total, 125/347 (36%) of EBAV and 289/600 (48%) of BAVGWAS CNVs were validated.

There were 34 large (>250 Kb), rare (<1:1000 in dbGAP controls) CNV regions that involved protein-coding genes in EBAV cases (Table S3). Seven of these genic CNVs were enriched in EBAV cases compared to WLS controls with a genome-wide adjusted empiric P < 0.05. These CNVs included the genes *PCP4*, *DSCAM*, *MIR4760*, and *DSCAM-AS1* in 21q22 and *GATA4*, *C8orf49*, *NEIL2*, *FDFT1*, and *CTSB* in 8p23. Large duplications involving the Velocardiofacial (VCFS) region in 22q11.2 and 1q21.1 microduplications were also enriched in EBAV cases (Table S4). The overall burden of large, rare, genic CNVs was not different between EBAV cases and WLS controls. However, the burden of large, rare genic CNVs intersecting with genes known to cause BAV when mutated or implicated in syndromic BAV was significantly increased in EBAV cases (Table 4).

Table 4. Burden Testing of Rare EBAV CNVs.

	EBAV		WLS			
	Calls	**Rate**	**Calls**	**Rate**	**RR**	***P***
Genic	28	0.8	1151	0.65	1.2	0.23
Deletions	11	3.8×10^{-2}	439	4.6×10^{-2}	0.81	0.78
BAV	3	1.0×10^{-2}	1	1.1×10^{-4}	97	1.1×10^{-3}
Total	34	-	1443	-	-	-

Calls: total number of CNVs that met the specified criteria. Rate: number of CNVs per individual; RR: relative risk; *P:* p-value; Genic; CNVs that intersect with genes; BAV: CNVs that intersect with genes that are known to cause bicuspid aortic valve (BAV) when mutated or implicated in syndromic BAV. Total: total number of large, rare CNVs or CNVRs. Tests are 2-sided using 100,000 permutations.

We also scrutinized genomic regions that are implicated in CHD by careful analysis of data from individual CNV algorithms to detect subtle copy number alterations. We identified additional rare EBAV CNVs that intersect with CHD candidate genes *CELSR1*, *GJA5*, *RAF1*, *LTBP1, KIF1A*, *MYH11*, *MAPK3*, *TTN*, and the VCFS region in 22q11.2. We detected additional *GATA4* and *DSCAM* CNVs in multiplex families. These CNVs were enriched in EBAV cases compared to WLS controls (Table 5).

Table 5. CNVs Affecting Congenital Heart Disease Genes in EBAV Cohort.

Region	Genes	Case	Control	OR	95% CI
Chr22:46261909-51187440	*CELSR1*	1	1	33	2.1 to 530
Chr1:146326373-147340734	*GJA5*	1	2	17	1.5 to 183
Chr3:12599717-12803792	*RAF1*	1	2	17	1.5 to 183
Chr22:41278694-41813285	*DSCAM*	4	2	67	12 to 367
Chr8:11495032-11856903	*GATA4*	4	0	301	16 to 5599
Chr22:19000000-22000000	*TBX1, CRKL*	4	10	13	4.2 to 43
Chr16:15484868-16295863	*MYH11*	2	22	3.0	0.70 to 13
Chr2:241652252-241678528	*KIF1A*	3	22	4.5	1.3 to 15
Chr2:32775984-33331219	*LTBP1*	2	26	2.5	0.60 to 11

Region: coordinates corresponding to the minimum overlap region of CNVs; Genes: cardio-developmental candidate genes in the region. Case: number of large and rare CNVs in EBAV cases that intersect with region of interest. Control: number of CNVs in WLS cohort that intersect with region of interest. OR: odds ratio; 95% CI, 95% confidence interval for respective odds ratio.

Next, we attempted to replicate our observations by identifying CNVs in the BAVGWAS dataset that overlapped with rare EBAV CNVs. We found that large duplications involving *SOX7* and *GATA4* in 8p23 and the VCFS region in 22q11.2 were also significantly enriched in BAVGWAS cases compared to HRS controls (Table 6, Table S6 and S7).

Table 6. CNVs Affecting Congenital Heart Disease Genes in BAVGWAS Cohort.

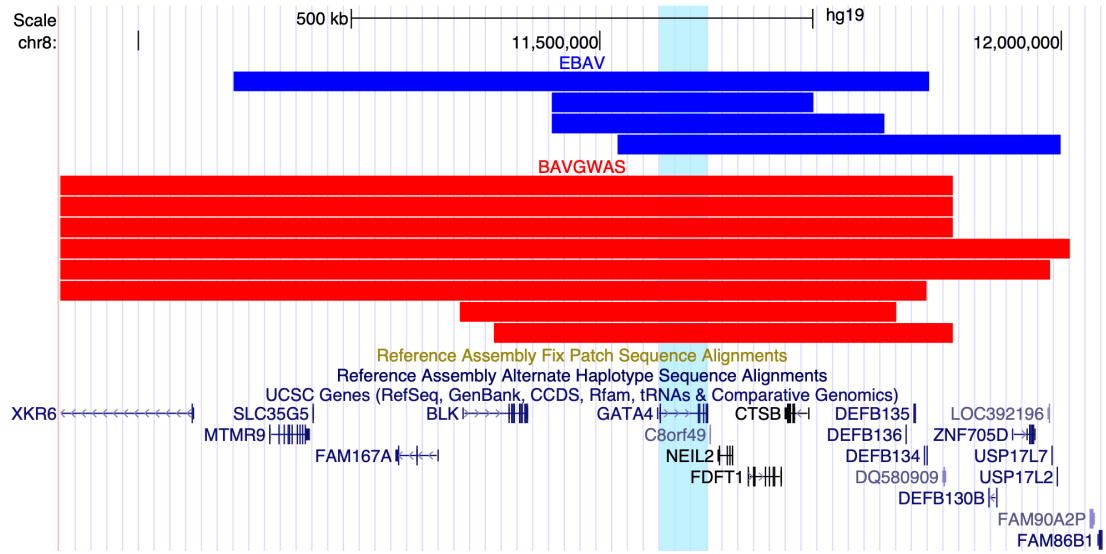
Region	Genes	Case	Control	OR	95% CI
Chr3:29993977-31273870	*TGFBR2*	1	0	5.6	0.23 to 138
Chr9:101861767-102092282	*TGFBR1*	1	0	5.6	0.23 to 138
Chr21:41577819-41842252	*DSCAM*	2	1	3.7	0.34 to 41
Chr22:46924254-46931077	*CELSR1*	3	1	5.6	0.58 to 54
Chr2:111404636-11310378	*TMEM87B, FBLN7*	3	2	2.8	0.47 to 17
Chr8:11385469-11821835	*GATA4*	8	1	15	1.9 to 120
Chr12:7918339-8130958	*NANOG*	10	2	9.4	2.1 to 43
Chr2:147166377-147308112	*GJA5*	4	10	0.75	0.23 to 2.4
Chr16:29664753-30199713	*MAPK3*	3	15	0.37	0.11 to 1.3
Chr22:19000000-22000000	*TBX1, CRKL*	18	11	3.1	1.4 to 6.5
Chr2:32689829-33299434	*LTBP1*	9	22	0.76	0.35 to 1.7
Chr16:15240816-16281154	*MYH11*	13	27	0.90	0.46 to 1.7
Chr2:241640262-241689833	*KIF1A*	13	30	0.81	0.42 to 1.6

Region: coordinates corresponding to the minimum overlap region of CNVs; Genes: cardio-developmental candidate genes in the region. Case: number of large and rare CNVs in BAVGWAS cases that intersect with region of interest. Control: number of CNVs in HRS cohort that intersect with region of interest. OR: odds ratio; 95% CI, 95% confidence interval for respective odds ratio.

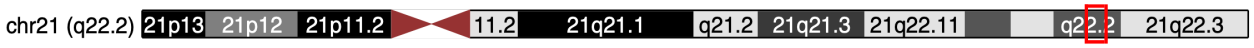
CNVs intersecting with *GATA4* and *DSCAM* significantly overlapped between EBAV and BAVGWAS datasets (Fig 2). On average, the *GATA4* CNVs were larger in the BAVGWAS dataset while the *DSCAM* CNVs were larger in the EBAV dataset.

(a)

(b)



(c)



(d)

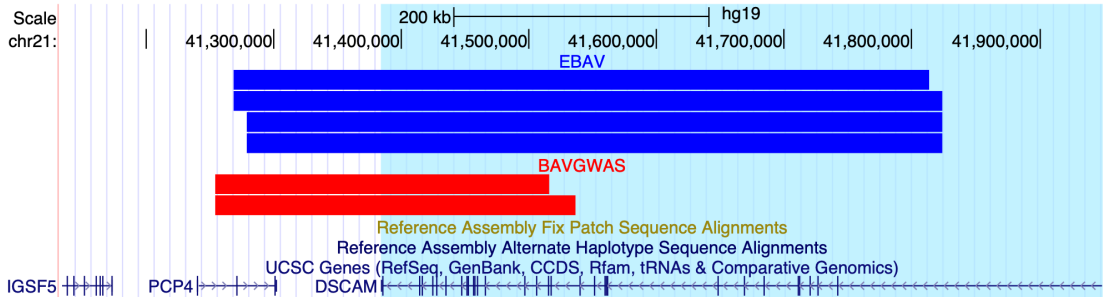


Fig 2. UCSC Genome Browser Plots of *GATA4* and *DSCAM* Variants.
(a) Ideogram of Chromosome 8 with view of image in (b) outlined in red box. (b) Plot of *GATA4* variants. Each bar represents a copy number variant (CNV). CNVs from the EBAV cohort are in blue and CNVs from the BAVGWAS cohort are in red. The region spanned by *GATA4* has been highlighted in light blue. (c) Ideogram of Chromosome 21 with view of image in (d) outlined in red box. (d) Plot of *DSCAM* variants. EBAV CNVs are in blue and BAVGWAS CNVs are in red. The region spanned by *DSCAM* is highlighted in light blue. Figures constructed using the UCSC Genome Browser, http://genome.ucsc.edu [29].

We identified 7 additional CNV regions that are enriched in BAVGWAS cases but not in EBAV and are rare or absent in controls (Table S5). *NANOG* and *NIBPL* are essential for early

heart development, and mutation of *NIBPL* causes Cornelia-de Lange syndrome with a spectrum of congenital heart malformations including BAV.

We also identified 21 very large genomic CNVs more than 5 Mb in length in the BAVGWAS dataset. Analysis of GenomeStudio data showed that most of these were mosaic loss of heterozygosity regions or duplications. Nine were large germline chromosome-scale aberrations, including two cases of trisomy 21 (Table S8). We did not identify any large X chromosome copy variants that may be consistent with Turner syndrome. There were no megabase-scale copy number variants in the EBAV dataset.

Pedigree analysis showed that several CNVs involving *CELSR1*, *LTBP1*, *KIF1A*, *GATA4*, and *DSCAM* segregate with BAV in EBAV families (Table S9). CNV carriers tended to present due to moderate or severe aortic regurgitation requiring valvular surgery. One proband had aortic coarctation. The youngest age at presentation was 13 years. There were no sex differences in presentation between CNV carriers.

Discussion

We identified large, rare, and likely pathogenic CNVs in almost 10% of EBAV probands that are enriched in genes that cause BAV when mutated. The percentage of EBAV cases with likely pathogenic CNVs is similar to our previous observations in a cohort with early onset TAD [30]. Enrichment of CNVs involving *GATA4* and *DSCAM* in EBAV cases replicated in two additional BAV datasets and thousands of unselected control genotypes. This analysis provides compelling evidence that rare CNVs collectively cause more BAV cases than any single mutated gene.

GATA-Binding Protein 4 is a transcription factor that is required for cardiac and neuronal differentiation during embryogenesis [31]. Mutations of *GATA4* and its homologs *GATA5* and *GATA6* cause congenital heart lesions [32]. Mutations in the *GATA4* gene have been linked to a range of congenital heart diseases in humans, such as cardiac septal defects, tetralogy of Fallot, amd patent ductus arteriosus [33]. Patients with BAV who have rare functional variants in the *GATA* family exhibit varying degrees of aortopathy expression, including aortic aneurysm, dissection, and/or aortic stenosis. Alonso-Montes et al. described 4 predicted deleterious *GATA4* mutations in 122 non-syndromic BAV probands who did not have affected relatives [34]. Rare *GATA4* deletions and putative loss of function mutations are also implicated in CHD with distinctive features, underlining the importance of *GATA4* dosage to cardiac development [35-36]. Glessner et al. discovered large *de novo* (~4Mb) duplications involving *GATA4* in CHD trios with conotruncal defects or left ventricular outflow tract obstructive lesions [37]. Some duplications were inherited from apparently unaffected parents. Zogopoulos and Yu described similar genomic duplications in unaffected individuals and in unselected control genotypes [38-39].

These observations are consistent with low-penetrance CHD in *GATA4* duplication carriers. Similar to other complex and multifactorial disorders, CHD pathogenesis is likely caused by the cumulative impact of multiple CNVs or mutations, each exerting small to moderate effects to collectively disrupt cardiac development. For example, the frequency of congenital heart lesions is increased in individuals with velocardiofacial syndrome who have 22q1.2 deletions and a common 12p13.31 duplication involving the *SLC2A3* gene. The *SLC2A3* CNV likely functions as a modifier of the cardiac phenotype associated with 22q11 deletion syndrome, exemplifying a “two-hit” model [40].

More than half of patients with Down syndrome have congenital heart malformations due to the interaction of multiple dosage-sensitive CHD genes on chromosome 21 [41-43]. Down syndrome cell adhesion molecule, previously shown to play a critical role in neurogenesis, has also been implicated in the pathophysiology of CHD [44]. Analysis of rare segmental trisomies of chromosome 21 suggested that duplication of *DSCAM* and the contiguous *COL6A1* and *COL6A2* genes may cause septal abnormalities and other Down Syndrome-related CHD lesions, including BAV. Overexpression of *DSCAM* and *COL6A2* causes cardiac malformations in mice [45]. Our findings suggest that rare CNVs involving *DSCAM* may contribute to some non-syndromic BAV cases.

Consistent with previous observations, *GATA4* and *DSCAM* CNVs segregated with disease in multiple families, but are not fully penetrant and were detected in some unaffected relatives. Intriguingly, large 22q11.2, *GATA4* and *DSCAM* CNVs were more highly enriched in EBAV than in BAVGWAS cases, suggesting that these CNVs may drive early onset BAV disease. These results are consistent with our observation that pathogenic CNVs involving candidate BAV genes are also enriched in EBAV compared to BAVGWAS cases. Our data suggests that pathogenic CNVs at these loci may predict accelerated disease onset or more severe complications.

We also identified recurrent rare CNVs of specific dosage-sensitive regions that affect cardiac developmental genes and are implicated in non-syndromic CHD. Recurrent 1q21.1 distal deletions encompassing *GJA5*, the gene encoding Connexin-40, are associated with CHD lesions including BAV. A study of 807 TOF cases showed significant enrichment of small duplications spanning the *GJA5* gene, providing compelling evidence that it acted as the primary candidate gene, supporting the association of *GJA5* and CHD [31]. Additionally, cardiac abnormalities

344 have been documented in mice with a targeted *GJA5* deletion, implying that haploinsufficiency
345 of *GJA5* might contribute to cardiac defects in individuals affected by 1q21.1 deletions [46].
346 *CELSR1*, a cadherin superfamily member, is mutated in families with BAV and hypoplastic left
347 heart syndrome [47]. *LTBP1* encodes an extracellular matrix protein that regulates TGF-beta and
348 fibrillin and has been implicated in congenital heart lesions [48]. *KIF1A*, encoding a kinesin
349 microtubule transporter, was implicated in a dominant multisystem syndromic disorder with
350 valvular and cardiac defects [49]. Mutation of *MYH11* causes familial thoracic aortic aneurysms
351 and dissections with an increased prevalence of BAV [50]. *TTN* mutations cause dilated
352 cardiomyopathy and are associated with other left-sided congenital lesions [51]. Mutations or
353 copy number changes involving these genes all cause a wide spectrum of penetrance and
354 phenotypic severity, consistent with sensitivity to genetic or clinical modifiers.

355 Our combinatorial analysis method eliminated many CNVs that were detected by single
356 algorithms or did not meet quality control benchmarks. Therefore, our analysis likely
357 underestimated the contribution of rare pathogenic CNVs to BAV. We also recognize that
358 cardiac development involves the complex interaction of many genes. We selectively validated
359 individual CNVs at loci of interest but may have underrepresented CNVs that had no *a priori*
360 relationship with CHD. The apparent penetrance of some CNVs may be less than expected due
361 to missing phenotypic information. The available clinical data was not sufficiently detailed to
362 permit genotype-phenotype correlations with specific CHD clinical features.
363 In conclusion, we identified large rare CNVs in a significant proportion of BAV cases,
364 including a subset of CNVs that may predict early onset complications of BAV disease. These
365 observations add to the evidence that rare CNVs may eventually have clinical utility for risk
366 stratification and personalized disease management.

Acknowledgments

We thank Joana Castillo and Jacqueline Jennings for sample preparation, and Gladys Zapata, Nitesh Mehta, and the Laboratory for Translational Genomics at Baylor College of Medicine for microarray genotyping. This study was supported in part by R01HL137028 (SP). Fig 1 created with BioRender.com.

The EBAV Investigators are: Shaine A. Morris, Rita Milewski, Giuseppe Limongelli, Allesandro Della Corte, Laura Perrone, Yuli Y. Kim, Hector Michelena, Maria G. Andreassi, Arturo Evangelista, Denver Sallee, Anji Yetman, Kim McBride, Eduardo Bossone, Rodolfo Citro, Dawn S. Hui, Malenka M. Bissell, Andrea Ballotti, Ilenia Foffa, Margot De Marco, Anthony Caffarelli, Rita Weise, Julie DeBacker, Laura Muino Mosquera, Robbin Cohen, Laura Dos Subira, Justin T. Tretter, Anna Sabe Rotes, Martina Caiazza, Lamia Ait Ali, Francesca Pluchinotta, and Simon C. Body.

The BAVCon Investigators are: Simon Body, Alessandro Della Corte, Alessandro Frigiola, Andrea Ballotta, Arturo Evangelista, Betti Giusti, Bo Yang, Carlo de Vincentiis, Dan Gilon, David Messika Zeitoun, Dianna M. Milewicz, Eduardo Bossone, Eric Eisselbacher, Francesca R. Pluchinotta, Giuseppe Limongelli, Gordon S. Huggins, Hector I. Michelena, J. Daniel Muehlschelgel, Kim Eagle, Lasse Folkerson, Malenka M. Bissell, María Brion, Patrick Mathieu, Per Eriksson, Peter Lichtner, Rodolfo Citro, Ronen Durst, Sébastien Thériault, Siddharth K. Prakash, Thoralf M. Sundt, Vicenza Stefano Nistri, Yahn Bossé.

References

1. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annual review of genomics and human genetics. 2009;10:451.

2. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H. A copy number variation morbidity map of developmental delay. Nature genetics. 2011 Sep;43(9):838-46.

3. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, Filipink RA, McConnell JS, Angle B, Meschino WS, Nezarati MM. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. New England Journal of Medicine. 2012 Oct 4;367(14):1321-31.

4. Kaufman L, Ayub M, Vincent JB. The genetic basis of non-syndromic intellectual disability: a review. Journal of neurodevelopmental disorders. 2010 Dec;2(4):182-209.

5. Liu, Y., Chen, S., Zühlke, L., Black, G., Choy, M. K., Li, N., & Keavney, B. (2019). Global birth prevalence of congenital heart defects 1970–2017: Updated systematic review and meta-analysis of 260 studies. International Journal of Epidemiology, 48(42), 455–463.

6. Hitz MP, Lemieux-Perreault LP, Marshall C, Feroz-Zada Y, Davies R, Yang SW, Lionel AC, D'Amours G, Lemyre E, Cullum R, Bigras JL. Rare copy number variants contribute to congenital left-sided heart disease.

7. Warburton D, Ronemus M, Kline J, Jobanputra V, Williams I, Anyane-Yeboa K, Chung W, Yu L, Wong N, Awad D, Yu CY. The contribution of de novo and rare inherited copy number changes to congenital heart disease in an unselected sample of children with conotruncal defects or hypoplastic left heart disease. Human genetics. 2014 Jan;133:11-27.

8. Silversides CK, Lionel AC, Costain G, Merico D, Migita O, Liu B, Yuen T, Rickaby J, Thiruvahindrapuram B, Marshall CR, Scherer SW. Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways.

9. Ware SM, Jefferies JL. New genetic insights into congenital heart disease. Journal of clinical & experimental cardiology. 2012 Jun 6.

10. Sørensen KM, El-Segaier M, Fernlund E, Errami A, Bouvagnet P, Nehme N, Steensberg J, Hjortdal V, Soller M, Behjati M, Werge T. Screening of congenital heart disease patients using multiplex ligation-dependent probe amplification: Early diagnosis of syndromic patients. American journal of medical genetics Part A. 2012 Apr;158(4):720-5.

11. Lander J, Ware SM. Copy number variation in congenital heart defects. Current Genetic Medicine Reports. 2014 Sep;2:168-78.

12. Tomita-Mitchell A, Mahnke DK, Struble CA, Tuffnell ME, Stamm KD, Hidestrand M, Harris SE, Goetsch MA, Simpson PM, Bick DP, Broeckel U. Human gene copy number spectra analysis in congenital heart malformations. Physiological genomics. 2012 May 1;44(9):518-41.

13. Kim DS, Kim JH, Burt AA, Crosslin DR, Burnham N, Kim CE, McDonald-McGinn DM, Zackai EH, Nicolson SC, Spray TL, Stanaway IB. Burden of potentially pathologic copy number variants is higher in children with isolated congenital heart disease and significantly impairs covariate-adjusted transplant-free survival. The Journal of thoracic and cardiovascular surgery. 2016 Apr 1;151(4):1147-51.

14. Soemedi R, Wilson IJ, Bentham J, Darlay R, Töpf A, Zelenika D, Cosgrove C, Setchfield K, Thornborough C, Granados-Riveron J, Blue GM. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. The American Journal of Human Genetics. 2012 Sep 7;91(3):489-501.

15. Chandra S, Lang RM, Nicolarsen J, Gayat E, Spencer KT, Mor-Avi V, Hofmann Bowman MA. Bicuspid aortic valve: inter-racial difference in frequency and aortic dimensions. JACC: Cardiovascular Imaging. 2012 Oct;5(10):981-9.

16. Michelena HI, Prakash SK, Della Corte A, Bissell MM, Anavekar N, Mathieu P, Bossé Y, Limongelli G, Bossone E, Benson DW, Lancellotti P. Bicuspid aortic valve: identifying knowledge gaps and rising to the challenge from the International Bicuspid Aortic Valve Consortium (BAVCon). Circulation. 2014 Jun 24;129(25):2691-704.

17. Glotzbach JP, Hanson HA, Tonna JE, Horns JJ, McCarty Allen C, Presson AP, Griffin CL, Zak M, Sharma V, Tristani-Firouzi M, Selzman CH. Familial Associations of Prevalence and Cause-Specific Mortality for Thoracic Aortic Disease and Bicuspid Aortic Valve in a Large-Population Database. Circulation. 2023 Jun 15.

18. Prakash SK, Yetman A, Bissell MM, Kim YY, Michelena H, Hui DS, Caffarelli A, Andreassi MG, Foffa I, Jennings J, Citro R. Recurrent genomic copy number variants implicate new candidate genes for early onset bicuspid aortic valve disease. Journal of the American College of Cardiology. 2019 Mar 12;73(9S1):620-.

19. Prakash S, Kuang SQ, GenTAC Registry Investigators, Regalado E, Guo D, Milewicz D. Recurrent rare genomic copy number variants and bicuspid aortic valve are enriched in early onset thoracic aortic aneurysms and dissections. PloS one. 2016 Apr 19;11(4):e0153543.

20. Balistreri CR, Cavarretta E, Sciarretta S, Frati G. Light on the molecular and cellular mechanisms of bicuspid aortic valve to unveil phenotypic heterogeneity. Journal of Molecular and Cellular Cardiology. 2019;133: 113–114. Doi:10.1016/j.yjmcc.2019.06.004.

21. Prakash, S.K., LeMaire, S.A., Guo, D.C., Russell, L., Regalado, E.S., Golabbakhsh, H., Johnson, R.J., Safi, H.J., Estrera, A.L., Coselli, J.S. and Bray, M.S., 2010. Rare copy number - variants disrupt genes regulating vascular smooth muscle cell adhesion and contractility in sporadic thoracic aortic aneurysms and dissections. *The American Journal of Human Genetics*, *87*(6), pp.743-756.

22. Prakash SK, Bossé Y, Muehlschlegel JD, Michelena HI, Limongelli G, Della Corte A, Pluchinotta FR, Russo MG, Evangelista A, Benson DW, Body SC. A roadmap to investigate the genetic basis of bicuspid aortic valve and its complications: insights from the International BAVCon (Bicuspid Aortic Valve Consortium). Journal of the American College of Cardiology. 2014 Aug 26;64(8):832-9.

23. PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez, JG. Conde. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009 Apr;42(2):377-81.

24. PA Harris, R Taylor, BL Minor, V Elliott, M Fernandez, L O'Neal, L McLeod, G Delacqua, F Delacqua, J Kirby, SN Duda, REDCap Consortium, The REDCap consortium. Building an international community of software partners. J Biomed Inform. 2019 May 9 [doi: 10.1016/j.jbi.2019.103208].

25. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome research. 2007 Nov 1;17(11):1665-74.

26. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic acids research. 2007 Mar 1;35(6):2013-25.

27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 2007 Sep 1;81(3):559-75.

28. Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, Chen HC, Agarwala R, McLaren WM, Ritchie GR, Albracht D. Modernizing reference genome assemblies. PLoS biology. 2011 Jul 5;9(7):e1001091.

29. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome research. 2002 Jun 1;12(6):996-1006.

30. Prakash S, Kuang SQ, GenTAC Registry Investigators, Regalado E, Guo D, Milewicz D. Recurrent rare genomic copy number variants and bicuspid aortic valve are enriched in early onset thoracic aortic aneurysms and dissections. PloS one. 2016 Apr 19;11(4):e0153543.

31. Durocher, D., Charron, F., Warren, R., Schwartz, R. J., Nemer, M. The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. EMBO J. 16: 5687-5696, 1997.

32. Tremblay M, Sanchez-Ferras O, Bouchard M. GATA transcription factors in development and disease. Development. 2018 Oct 15;145(20):dev164384.

33. McCulley DJ, Black BL. Transcription factor pathways and congenital heart disease. Current topics in developmental biology. 2012 Jan 1;100:253-77.

34. Alonso-Montes C, Martín M, Martínez-Arias L, Coto E, Naves-Díaz M, Morís C, Cannata-Andía JB, Rodríguez I. Variants in cardiac GATA genes associated with bicuspid aortic valve. European journal of clinical investigation. 2018 Dec;48(12):e13027.

35. Pehlivan T, Pober BR, Brueckner M, Garrett S, Slaugh R, Van Rheeden R, Wilson DB, Watson MS, Hing AV. GATA4 haploinsufficiency in patients with interstitial deletion of chromosome region 8p23. 1 and congenital heart disease. American journal of medical genetics. 1999 Mar 19;83(3):201-6.

36. Li RG, Xu YJ, Wang J, Liu XY, Yuan F, Huang RT, Xue S, Li L, Liu H, Li YJ, Qu XK. GATA4 loss-of-function mutation and the congenitally bicuspid aortic valve. The American journal of cardiology. 2018 Feb.

37. Glessner JT, Bick AG, Ito K, Homsy JG, Rodriguez-Murillo L, Fromer M, Mazaika E, Vardarajan B, Italia M, Leipzig J, DePalma SR. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. Circulation research. 2014 Oct 24;115(10):884-96.

38. Zogopoulos G, Ha KC, Naqib F, Moore S, Kim H, Montpetit A, Robidoux F, Laflamme P, Cotterchio M, Greenwood C, Scherer SW. Germ-line DNA copy number variation frequencies in a large North American population. Human genetics. 2007 Nov;122:345-53.

39. Yu S, Zhou XG, Fiedler SD, Brawner SJ, Joyce JM, Liu HY. Cardiac defects are infrequent findings in individuals with 8p23. 1 genomic duplications containing GATA4. Circulation: Cardiovascular Genetics. 2011 Dec;4(6):620-5.

40. Mlynarski EE, Sheridan MB, Xie M, Guo T, Racedo SE, McDonald-McGinn DM, Gai X, Chow EW, Vorstman J, Swillen A, Devriendt K. Copy-number variation of the glucose transporter gene SLC2A3 and congenital heart defects in the 22q11. 2 deletion syndrome. The American Journal of Human Genetics. 2015 May 7;96(5):753-64.

41. Freeman SB, Taft LF, Dooley KJ, Allran K, Sherman SL, Hassold TJ, Khoury MJ, Saker DM. Population-based study of congenital heart defects in Down syndrome. American journal of medical genetics. 1998 Nov 16;80(3):213-7.

42. Paladini D, Tartaglione A, Agangi A, Teodoro A, Forleo F, Borghese A, Martinelli P. The association between congenital heart disease and Down syndrome in prenatal life. Ultrasound in Obstetrics and Gynecology. 2000 Feb;15(2):104-8.

43. Laursen HB. Congenital heart disease in Down's syndrome. Heart. 1976 Jan 1;38(1):32-8.

44. Kosaki R, Kosaki K, Matsushima K, Mitsui N, Matsumoto N, Ohashi H. Refining chromosomal region critical for Down syndrome-related heart defects with a case of cryptic 21q22. 2 duplication. Congenital anomalies. 2005 Jun;45(2):62-4.

45. Grossman TR, Gamliel A, Wessells RJ, Taghli-Lamallem O, Jepsen K, Ocorr K, Korenberg JR, Peterson KL, Rosenfeld MG, Bodmer R, Bier E. Over-expression of DSCAM and COL6A2 cooperatively generates congenital heart defects. PLoS genetics. 2011 Nov 3;7(11):e1002344.

532 46. Gu H, Smith FC, Taffet SM, Delmar M. High incidence of cardiac malformations in
533 connexin40-deficient mice. Circulation research. 2003 Aug 8;93(3):201-6.

534 47. Theis JL, Niaz T, Sundsbak RS, Fogarty ZC, Bamlet WR, Hagler DJ, et al. CELSR1 Risk
535 Alleles in Familial Bicuspid Aortic Valve and Hypoplastic Left Heart Syndrome. Circ: Genomic
536 and Precision Medicine. 2022;15. doi:10.1161/CIRCGEN.121.003523.

537 48. Pottie L, Adamo CS, Beyens A, Lütke S, Tapaneeyaphan P, De Clercq A, et al. Bi-allelic
538 premature truncating variants in LTBP1 cause cutis laxa syndrome. The American Journal of
539 Human Genetics. 2021;108: 1095–1114. doi:10.1016/j.ajhg.2021.04.016.

540 49. Akasaka T, Ocorr K, Lin L, Vogler G, Bodmer R, Grossfeld P. Overexpression of Kif1A in
541 the Developing Drosophila Heart Causes Valvar and Contractility Defects: Implications for
542 Human Congenital Heart Disease. JCDD. 2020;7: 22. doi:10.3390/jcdd7020022.

543 50. Pannu H, Tran-Fadulu V, Papke CL, Scherer S, Liu Y, Presley C, et al. MYH11 mutations
544 result in a distinct vascular pathology driven by insulin-like growth factor 1 and angiotensin II.
545 Human Molecular Genetics. 2007;16: 2453–2462. doi:10.1093/hmg/ddm201.

546 51. Herman DS, Lam L, Taylor MRG, Wang L, Teekakirikul P, Christodoulou D,
547 et.al. Truncations of titin causing dilated cardiomyopathy. New England Journal of Medicine.
548 2012;366: 619-628. doi: 10.1056/NEJMoa1110186

Supplemental Data

Cohort	Study	Samples	Accession	Microarray
WLS	Wisconsin Longitudinal Study on Aging	8969	Phs001157.v1.pl	Illumina HumanOmniExpress-24 v1.1
HRS	Health and Retirement Study	9426	phs000428.v2.pl	Illumina Human Omni2.5-Quad

S1 Table. Summary of Control Cohorts. Cohort, name of control cohort. Study, study from which genotypes were obtained. Samples, number of control samples in each dataset. Accession, Database of Genotypes and Phenotypes accession number. Microarray, Illumina microarray used for genotyping.

	EBAV	BAVGWAS	WLS	HRS
PennCNV	6781	73784	58115	163938
cnvPartition	2289	33640	31148	51794
QuantiSNP	1798	21326	14346	85312
Merged	902	7622	21343	14657
Deletions	610	2772	8170	6770
>5 MB	9	22	9830	6114
Rare	84	579	1443	1372
Rare Deletions	59	181	285	394

S2 Table. Comprehensive CNV Summary.EBAV, EBAV Cohort including cases and unaffected family members. BAVGWAS, BAVGWAS cohort. WLS, WLS cohort. HRS, HRS cohort. PennCNV, number of CNV calls detected by PennCNV algorithm after quality control. cnvPartition, number of CNV calls detected by cnvPartition algorithm after quality control. QuantiSNP, number of CNVs detected by QuantiSNP algorithm after quality control. Merged, number of CNV regions after merging initial calls. Deletions, number of CNV regions that are deletions. >5 MB, number of CNV regions that are larger than 5 megabases. Rare, number of large (> 250 kilobases and less than 5 megabases) CNV regions that occur in less than 1 in 1000 samples based on case-control cohort pairs (EBAV and WLS; BAVGWAS and HRS). R. Del., number of large, rare deletions. All values reflect the total CNV calls and regions prior to validation in GenomeStudio.

Chr.	Start BP	Stop BP	Type
1	187296703	187609850	DEL
1	146326373	147340734	DUP
1	79238015	79619893	DEL
1	228625778	228880626	DUP
2	114458921	115208197	DUP
2	4638261	5564549	DUP
3	31901848	32165994	DUP
3	19363589	19813225	DEL
4	84658825	85270309	DUP
5	25468811	25719474	DEL
5	78016365	78286867	DUP
6	95836160	96095769	DEL
8	89353386	89800669	DEL

8	2319555	2585105	DUP
8	4201652	4493979	DUP
8	10111571	10721128	DUP
8	11103895	11856864	DUP
8	9368431	9745798	DUP
8	11448529	11732454	DUP
8	11448529	11808756	DUP
10	134505252	135203544	DEL
12	84108147	84443245	DUP
13	70578273	71593281	DUP
15	32908301	34761123	DEL
16	83302526	84016062	DUP
17	1389	582832	DEL
18	57590566	57955945	DUP
21	41268738	41813285	DUP
21	41268738	41823356	DUP
21	41278694	41823356	DUP
21	41278694	41823356	DUP
22	19580050	20227551	DUP
22	46261909	46931077	DEL
22	48871294	51187440	DEL

S3 Table. Large, Rare Copy Number Variants Identified in the EBAV Cohort. Chr., Chromosome on which CNV is located. Start BP, start basepair of CNV. Stop BP, stop basepair of CNV. Type, denotes if a CNV is a duplication (DUP) or deletion (DEL) event. All CNVs were validated in GenomeStudio.

Gene(s)	Chr.	Start BP	Stop BP	Type
HYDIN2, NBPF12, LOC728989, NBPF13P, PRKAB2, PDIA3P, FM05, CHD1L, LINC00624, BCL9, ACP6,* and *GJA5	1	146326373	147340734	DUP
HYDIN2, NBPF12, LOC728989, NBPF13P, PRKAB2, PDIA3P, FM05, CHD1L, LINC00624, BCL9, ACP6,* and *GJA5*	1	146326373	147229299	DUP
MIR4782, SLC35F5, ACTR3, LOC100499194,* and *LOC440900*	2	114426115	115208197	DUP
MIR4782, SLC35F5, ACTR3, LOC100499194,* and *LOC440900*	2	114614021	114732241	DUP
MIR4782, SLC35F5, ACTR3, LOC100499194,* and *LOC440900*	2	114458921	115208197	DUP
MIR4782, SLC35F5, ACTR3, LOC100499194,* and *LOC440900	2	114458921	115208197	DUP
TTN, AX746670, TTN-AS1,* and *MIR548N	2	179364778	179486671	DUP

TTN, AX746670, TTN-AS1,* and *MIR548N*	2	179395466	179517632	DUP
GATA4, C8orf49, NEIL2, FDFT1,* and *CTSB	8	11506208	11786255	DUP
GATA4, C8orf49, NEIL2, FDFT1,* and *CTSB	8	11103895	11856864	DUP
GATA4, C8orf49, NEIL2, FDFT1,* and *CTSB	8	11448529	11808756	DUP
GATA4, C8orf49, NEIL2, FDFT1,* and *CTSB	8	11448529	11732454	DUP
PARD3	10	35107733	35284461	DUP
PARD3	10	35107733	35271898	DUP
KLHL1* and *ATXN8OS	13	70578273	71593281	DUP
KLHL1* and *ATXN8OS*	13	70589082	71548725	DUP
KLHL1* and *ATXN8OS*	13	70730307	70773605	DEL
NECAB2	16	83302526	84016062	DUP
NECAB2*	16	83303915	83999565	DUP
PCP4, DSCAM, MIR4760,* and *DSCAM-AS1	21	41278694	41823356	DUP
PCP4, DSCAM, MIR4760,* and *DSCAM-AS1	21	41268738	41813285	DUP
PCP4, DSCAM, MIR4760,* and *DSCAM-AS1	21	41278694	41823356	DUP
PCP4, DSCAM, MIR4760,* and *DSCAM-AS1	21	41278694	41813285	DUP
PCP4, DSCAM, MIR4760,* and *DSCAM-AS1	21	41268738	41823356	DUP
TBX1, GNB1L, C22orf29, TXNRD2, COMT, MIR4761, ARVCF, TANGO2, MIR185, DGCR8, MIR3618, MIR1306, TRMT2A, RANBP1, ZDHHC8, LOC388849, LOC284865,* and *LINC00896	22	19580050	20227551	DUP
TBX1, GNB1L, C22orf29, TXNRD2, COMT, MIR4761, ARVCF, TANGO2, MIR185, DGCR8, MIR3618, MIR1306, TRMT2A, RANBP1, ZDHHC8, LOC388849, LOC284865, LINC00896, RTN4R,* and *MIR1286	22	18877787	21461607	DUP

S4 Table. Rare CNVs Enriched in EBAV Cohort. Gene(s), genes intersected by CNV. Chr, chromosome on which each CNV is on. Start BP, start basepair of each CNV. Stop BP, stop basepair of each CNV. Type, denotes if a CNV was a duplication (DUP) or deletion (DEL) event.

568 * Indicates the call was from an unaffected family member.
569

Gene(s)	Chr	Start BP	Stop BP	Type
LOC100507334	2	110852875	111406073	DUP
LOC100507334	2	110982530	112007875	DUP
MIR128-2	3	35775249	35938795	DUP
MIR128-2	3	35775249	35938795	DUP
MIR128-2	3	35785608	35936616	DUP
TMPRSS11E, UGT2B17, UGT2B15, UGT2B10	4	69599357	69712995	DUP
AHRR, C5orf55, EXOC3, FLJ00157, AK023178, PP7080, BC013821, LOC100996325,* and *CEP72	5	323965	889536	DUP
AHRR, C5orf55, EXOC3, FLJ00157, AK023178, PP7080,* and *BC013821	5	287907	602256	DUP
AHRR, C5orf55, EXOC3, FLJ00157, AK023178, PP7080,* and *BC013821	5	310925	548342	DUP
AHRR, C5orf55, EXOC3, FLJ00157, AK023178, PP7080, BC013821, LOC100996325,* and *CEP72	5	426109	673408	DUP
AHRR, C5orf55, EXOC3, FLJ00157, AK023178, PP7080, BC013821, LOC100996325,* and *CEP72	5	589727	701920	DUP
AHRR, C5orf55, EXOC3, FLJ00157, AK023178, PP7080, BC013821, LOC100996325,* and *CEP72	5	589727	701920	DUP
NIPBL	5	36764235	37046626	DUP
NIPBL	5	36805679	37046626	DUP
NIPBL	5	36898424	37046626	DUP
NIPBL	5	36911625	37052624	DUP
SGK223, CLDN23,* and *MFHAS1	8	8064756	11143272	DUP
SGK223, CLDN23,* and *MFHAS1	8	8064756	11882065	DUP
SGK223, CLDN23,* and *MFHAS1	8	8064756	8655355	DUP
SGK223, CLDN23,* and *MFHAS1	8	8114228	8627839	DUP
SGK223, CLDN23,* and *MFHAS1	8	8202294	8674049	DUP
SGK223, CLDN23,* and *MFHAS1	8	8221088	8650456	DUP
CUL5	11	107755731	107965390	DUP
NANOG* and *NANOGNB	12	7893437	8101326	DUP

NANOG* and *NANOGNB	12	7918339	8109412	DUP
NANOG* and *NANOGNB	12	7942473	8109412	DUP
NANOG* and *NANOGNB	12	7942945	8123777	DUP
NANOG* and *NANOGNB	12	7942945	8105015	DUP
NANOG* and *NANOGNB	12	7945559	8101326	DUP
NANOG* and *NANOGNB	12	7945559	8105015	DUP
NANOG* and *NANOGNB	12	7945559	8105015	DUP
NANOG* and *NANOGNB	12	7945559	8109412	DUP
NANOG* and *NANOGNB	12	7945559	8130958	DUP
UBE2MP1*, *LOC283914*, *LOC146481*, and *LOC100130700	16	34355747	34740580	DUP
LOC283914* and *LOC146481	16	34428972	34723621	DUP
LOC283914	16	34433468	34663346	DUP
FAM101B*, *VPS53*, and *FAM57A	17	1389	641023	DUP
FAM101B*, *VPS53*, *FAM57A*, *GEMIN4*, *DQ581337*, and *DBIL5P	17	225778	906268	DEL
FAM101B*, *VPS53*, *FAM57A*, and *GEMIN4	17	225778	649766	DUP
FAM101B*, *VPS53*, *FAM57A*, and *GEMIN4	17	238906	650372	DUP
FAM101B*, *VPS53*, *FAM57A*, *GEMIN4*, *DQ581337*, and *DBIL5P	17	284614	831667	DUP
RYR1*, *MAP4K1*, and *EIF3K	19	38683266	39116961	DUP
RYR1*, *MAP4K1*, and *EIF3K	19	38976659	39116961	DUP
RYR1*, *MAP4K1*, and *EIF3K	19	38993142	39116961	DUP
RYR1*, *MAP4K1*, and *EIF3K	19	38993142	39116961	DUP
CYP2A7*, *CYP2G1P*, *CYP2B7P1*, and *CYP2B6	19	41349732	41508557	DUP
CYP2A7*, *CYP2G1P*, *CYP2B7P1*, and *CYP2B6	19	41350509	41600054	DUP
CYP2A7*, *CYP2G1P*, *CYP2B7P1*, and *CYP2B6	19	41354458	41588347	DUP
CYP2A7*, *CYP2G1P*, *CYP2B7P1*, and *CYP2B6	19	41386035	41522338	DUP
CYP2A7*, *CYP2G1P*, *CYP2B7P1*, and *CYP2B6	19	41386814	41531705	DUP
CYP2A7*, *CYP2G1P*, *CYP2B7P1*, and *CYP2B6	19	41386814	41519306	DUP

S5 Table. Rare CNVs Enriched in BAVGWAS Cohort. Gene(s), genes intersected by CNV. Chr, chromosome on which each CNV is on. Start BP, start basepair of each CNV. Stop BP, stop basepair of each CNV. Type, denotes if a CNV was a duplication (DUP) or deletion (DEL) event.

Principal Gene/Regions	Chr.	Start BP	Stop BP	Type
KIF1A	2	241640262	241678528	DUP
KIF1A	2	241640262	241678528	DUP
KIF1A	2	241652252	241678528	DUP
KIF1A*	2	241626057	241702124	DUP
KIF1A*	2	241607616	241702124	DUP
KIF1A*	2	241644718	241709924	DUP
LTBP1	2	32639775	33331219	DUP
LTBP1	2	32775984	33331219	DUP
LTBP1*	2	32633925	33331219	DUP
LTBP1*	2	32633925	33331219	DUP
LTBP1*	2	32639775	33331219	DUP
RAF1	3	12599717	12803792	DUP
FLT4*	5	180019198	180056863	DEL
MICA	6	31360255	31453029	DEL
MICA	6	31360255	31485928	DEL
MICA	6	31360255	31487876	DEL
MICA	6	31360255	31457633	DUP
MICA	6	31361397	31453029	DUP
MICA*	6	31360255	31453029	DEL
MICA*	6	31360255	31453029	DEL
MICA*	6	31360255	31453029	DEL
MICA*	6	31360255	31485928	DEL
MICA*	6	31360255	31485928	DEL
MICA*	6	31360255	31485928	DEL
MICA*	6	31383960	31485928	DEL
MICA*	6	31355260	31453029	DEL
GATA4*	8	11506208	11786255	DUP
GATA4*	8	11506208	11999394	DUP
MUC5B	11	1078312	1300406	DUP
NANOG*	12	7945559	8123777	DUP
MYH11	16	14975292	16295863	DUP
MYH11	16	15484868	18309593	DUP
MAPK3	16	27977483	30174024	DUP
NCOR1	17	15976558	16012829	DUP
DSCAM*	21	41278161	41856480	DUP
DSCAM*	21	41278694	41813285	DUP
22q11*	22	19698129	19883189	DEL
22q11*	22	19682627	19755127	DEL
22q11	22	19701341	19776365	DEL
22q11	22	19701341	19808938	DEL

22q11	22	20742450	21461607	DEL

S5 Table. EBAV CNVs intersecting with Genes of Interest. Gene/Region, Principal gene or region of interest intersected by CNV. Chr, chromosome on which each CNV is on. Start BP, start basepair of each CNV. Stop BP, stop basepair of each CNV. Type, denotes if a CNV was a duplication (DUP) or deletion (DEL) event.
* Indicates the call was from an unaffected family member.
** Indicates the call was from an affected family member from a multiplex family.

Principal Gene/Regions	Chr.	Start BP	Stop BP	Type
GJA5	1	145723645	148343177	DUP
GJA5	1	145723739	148343177	DUP
GJA5	1	145801230	147824365	DUP
GJA5	1	147166377	147308112	DUP
TMEM87B/FBLN7	2	110982530	113103748	DUP
TMEM87B/FBLN8	2	111399346	113103748	DEL
TMEM87B/FBLN9	2	111404636	113215796	DUP
KIF1A	2	241623458	241697884	DUP
KIF1A	2	241623458	241697884	DUP
KIF1A	2	241623458	241698298	DUP
KIF1A	2	241623458	241724479	DUP
KIF1A	2	241626057	241689833	DUP
KIF1A	2	241626057	241689833	DUP
KIF1A	2	241626057	241689833	DUP
KIF1A	2	241626057	241689833	DUP
KIF1A	2	241626057	241689833	DUP
KIF1A	2	241626057	241702124	DUP
KIF1A	2	241626057	241702124	DUP
KIF1A	2	241640262	241689833	DUP
KIF1A	2	241640262	241697773	DUP
LTBP1	2	32619581	33299434	DUP
LTBP1	2	32619581	33331219	DUP
LTBP1	2	32633925	33302342	DUP
LTBP1	2	32633925	33302342	DUP
LTBP1	2	32633925	33331219	DUP
LTBP1	2	32633925	33331219	DUP
LTBP1	2	32633925	33331219	DUP
LTBP1	2	32633925	33369552	DUP
LTBP1	2	32689829	33331219	DUP
RAF1	3	12645681	12739194	DUP
TGFBR2	3	29993977	31273870	DEL
SOX7/GATA4	8	8064756	11882065	DUP
SOX7/GATA4	8	8064756	11882065	DUP
SOX7/GATA4	8	8064756	11882065	DUP

SOX7/GATA4	8	8064756	12009597	DUP
SOX7/GATA4	8	10109379	11987960	DUP
SOX7	8	10587741	10683929	DEL
GATA4	8	10914233	11853596	DUP
GATA4	8	11349186	11821835	DUP
GATA4	8	11385469	11882065	DUP
TGFBR1	9	101861767	102092282	DUP
MYH11	16	14761719	16281154	DUP
MYH11	16	14761719	16315360	DUP
MYH11	16	14975292	16299148	DEL
MYH11	16	14975292	16308351	DUP
MYH11	16	14975292	16308351	DUP
MYH11	16	14975292	16308351	DUP
MYH11	16	14975292	16308351	DUP
MYH11	16	14975292	16308351	DUP
MYH11	16	14975292	16308351	DUP
MYH11	16	14975292	16315360	DUP
MYH11	16	15092120	16291933	DUP
MYH11	16	15125441	16292128	DUP
MYH11	16	15240816	18584353	DUP
MAPK3	16	29647342	30199713	DUP
MAPK4	16	29647342	30199713	DUP
MAPK5	16	29647342	30199713	DUP
DSCAM	21	41254102	41516071	DUP
DSCAM	21	41254456	41536215	DUP
22q11	22	16874656	20241436	DEL
22q11	22	17818807	19002159	DUP
22q11	22	18644702	21726191	DUP
22q11	22	18877787	21461607	DUP
22q11	22	18877787	21461607	DUP
22q11	22	18877787	21028007	DEL
22q11	22	18877787	21804903	DEL
22q11	22	19062020	20264937	DUP
22q11	22	19667336	20329526	DEL
22q11	22	19682627	20233865	DEL
22q11	22	19682627	20262166	DEL
22q11	22	19693418	20264937	DEL
22q11	22	19701341	20300738	DEL
22q11	22	19724224	20300738	DEL
22q11	22	19951816	24298181	DUP
22q11	22	20719325	21726191	DEL

22q11	22	21246902	22702508	DEL
22q11	22	21424414	22015771	DUP
CESLR1	22	45236935	48193505	DEL
CESLR1	22	46751367	47159028	DUP
CESLR1	22	46924254	46931077	DEL

S7 Table. BAVGWAS CNVs intersecting with Genes of Interest. Gene/Region, Principal gene or region of interest intersected by CNV. Chr, chromosome on which each CNV is on. Start BP, start basepair of each CNV. Stop BP, stop basepair of each CNV. Type, denotes if a CNV was a duplication (DUP) or deletion (DEL) event.

Chr.	Start BP	Stop BP	Type	Description
2	138066736	143331537	DUP	Mosaic LOH
2	183476298	189945752	DUP	Mosaic LOH
3	143040791	168814375	DUP	Mosaic LOH
3	66206	7768285	DEL	Constitutional
6	148301116	156618923	DEL	Constitutional
7	101355402	106892492	DEL	Mosaic
8	6970806	12525566	DUP	Constitutional
8	170692	11987960	DUP	Constitutional
14	101350298	107283150	DUP	Mosaic LOH
14	71135027	107283150	DUP	Mosaic LOH
15	80465431	88497147	DUP	Mosaic LOH
15	93593528	102150818	DUP	Mosaic LOH
15	22761722	28540261	DEL	Constitutional
17	15175570	22234751	DUP	Mosaic
18	67445173	78010620	DEL	Constitutional
20	31265482	50716159	DEL	Mosaic
20	61098	25829977	DEL	Mosaic
20	31240778	48292606	DEL	Mosaic
20	50320079	62960292	DUP	Constitutional
21	14359894	48099610	DUP	Trisomy 21
21	14359894	48099610	DUP	Trisomy 21

S8 Table. Large Genomic Events in BAVGWAS Chr., Chromosome CNV on which CNV is located. Start BP, start base pair of CNV. Stop BP, stop base pair of CNV. Type, denotes if a CNV is a duplication (DUP) or deletion (DEL) event. Description, denotes if the CNV was a mosaic loss of heterozygosity (Mosaic LOH), loss of heterozygosity (LOH), mosaic (Mosaic), constitutional (constitutional), or trisomy 21 (Trisomy 21) event.

PROBAND	GENE	SEGREGATES?	WITH CNV	NO CNV	SEX
BAV064	*GATA4*	Yes	Father*, Paternal Grandfather*	Paternal Grandmother	Female
BAV475	*DSCAM*	Yes	Sister*	Father	Female
BAV787	*CELSR1*	Yes	None	Daughter	Female

BAV330	*KIF1A*	Yes	None	Father	Female
BAV478	*KIF1A*	Yes	None	Father	Male
BAV829	*LTBP1*	No	Son, Father	Mother, Brother	Female

S9 Table. Pedigree Information for CNVs that Segregated with Disease. Proband, identification number of proband with CNV intersecting with gene of interest. Gene, gene of interest intersected by CNV. Segregates?, indicates if the CNV segregated with disease. Family With CNV, family members of proband that were found to have a CNV intersecting with the respective gene. Family Without CNV, family members of proband who were not found to have a CNV intersecting with the respective gene. Family members are listed if their genotype was available for the study. Sex, sex of the proband.
*Indicates family members who also have BAV.