Targeted Next-generation Sequencing Identifies Pathogenic Variants in Familial Congenital Heart Disease

Running title: Gene panel identifies pathogenic variants in CHD

Gillian M. Blue, MSc, Edwin P. Kirk, MBBS, PhD, Eleni Giannoulatou, MEng, MPhil, DPhil, Sally L. Dunwoodie, PhD, Joshua WK Ho, PhD, Desiree C.K. Hilton, PhD, Susan M. White, MBBS, Gary F. Sholler, MBBS (Hons1), Richard P. Harvey, PhD, David S. Winlaw, MBBS, MD

*Kids Heart Research, The Children’s Hospital at Westmead, Sydney, Australia.
†The Heart Centre for Children, The Children’s Hospital at Westmead, Sydney, Australia
‡Sydney Medical School, University of Sydney, Australia
§Department of Medical Genetics, Sydney Children’s Hospital, Sydney, Australia
||School of Women’s and Children’s Health, Faculty of Medicine, University of New South Wales, Sydney, Australia
¶Victor Chang Cardiac Research Institute, Darlinghurst, Sydney, Australia
#University of New South Wales, Sydney, Australia
††St Vincent’s Clinical School, Faculty of Medicine, University of New South Wales, Sydney, Australia
‡‡School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia
§§Murdoch Children’s Research Institute, Royal Children’s Hospital, Melbourne, Australia
ǁǁDepartment of Paediatrics, University of Melbourne, Melbourne, Australia

We have no relationships with industry or financial associations to disclose.

ACKNOWLEDGEMENTS: We are grateful to Chain Reaction (www.chain-reaction.org.au) and Justin and Isabel Chan for supporting the development of the CHD gene panel. We also wish to acknowledge A/Prof Nigel Clarke for his contributions towards the development of this study.

Address for correspondence:
David S. Winlaw, MBBS, MD
Heart Centre for Children
The Children’s Hospital at Westmead, Locked Bag 4001
Westmead NSW 2145, Australia.
Phone: +61 2 9635 5298
Fax: +61 2 9845 2163
david.winlaw@health.nsw.gov.au
ABSTRACT

BACKGROUND Many genes have been implicated in the development of congenital heart disease (CHD). Next-generation sequencing (NGS) offers opportunities for genetic testing but is often complicated by logistic and interpretative hurdles.

OBJECTIVES We sought to apply NGS technology to CHD families with multiple affected members using a purpose-designed gene panel to assess diagnostic potential for future clinical application.

METHODS We designed a targeted NGS gene panel for 57 genes previously implicated in CHD. Probands were screened in 16 families with strong CHD histories and in 15 controls. Variants affecting protein-coding regions were classified in silico using prediction programs and filtered according to predicted mode of inheritance, minor allele frequencies (MAF), and presence in databases such as dbSNP and Exome Sequencing Project. Disease segregation studies were conducted in variants identified in CHD cases predicted to be deleterious and with MAF <0.1%.

RESULTS Thirteen potential disease-causing variants were identified in 9 families. Of these, 5 variants segregated with disease phenotype revealing a likely molecular diagnosis in 31% of this cohort. Significant increases in the number of ‘indels, nonsense, and splice’ variants, as well as variants classified as ‘probably damaging’ were identified in CHD cases but not controls. Also, there was a significant increase in the total number of ‘rare’ and ‘low’ frequency variants (MAF <0.05) in the CHD cases.

CONCLUSIONS When multiple relatives are affected by CHD, a gene panel-based approach may identify its cause in up to 31% of families. Identifying causal variants has implications for clinical care and future family planning.

KEY WORDS: congenital heart defects, gene panel, molecular diagnosis

ABBREVIATIONS

CHD = congenital heart disease
CNV = copy number variant
MAF = minor allele frequency
NGS = next-generation sequencing
SNP = single nucleotide polymorphism
The most common noninfectious cause of death in infants, congenital heart disease (CHD) affects approximately 6 to 8 per 1,000 live births (1). The majority of cases (~80%) occur as sporadic events but, in some, multiple family members are affected. The high heritability of CHD, estimated to be between 0.6 and 0.7 (2) suggests a strong genetic component and numerous genes have been linked to syndromic and nonsyndromic forms of CHD (3). The majority of these genes encode transcription factors, although genes in other categories have also been implicated, such as those for structural proteins.

Next-generation sequencing (NGS) enables rapid analysis of large amounts of genetic information. The last few years have produced an explosion of research utilizing NGS technology, especially exome sequencing, for novel gene discoveries in many genetic diseases. While one cannot dispute the advantage of exome sequencing for gene discovery through its unbiased approach, issues relating to coverage, analysis and storage of large amounts of data, and reporting of incidental findings complicate its use in the clinical setting (4). In comparison, limiting the capture regions to only those known to be associated with the disease(s) of interest mitigates some of these issues, making it a valuable and arguably more suitable approach in the diagnostic arena (5). Coverage, both in terms of depth as well as capture of on-target regions, is far greater, and being able to supplement missed NGS regions with Sanger sequencing further ensures high confidence in the results. Furthermore, issues relating to reporting of incidental findings can be avoided almost entirely as only disease-relevant genes are screened. Numerous disease-targeted gene panels are clinically available, including panels for hereditary cancers (6), metabolic disorders (7), cardiomyopathies (8), and aortopathies (9).

NGS technology has been applied to the study of both familial and sporadic forms of CHD (10,11). A recent publication by Zaidi et al. used exome sequencing analysis in parent-offspring trios to compare variants in 362 cases with severe sporadic CHD to 264 controls
They identified a significant excess of protein-altering, deleterious de novo mutations in known heart-expressing genes (odds ratio = 7.5), implicating several hundred genes that collectively may factor into ~10% of sporadic CHD cases. While these findings, as well as similar results relating to the contribution of rare and de novo copy number variants (CNV) (13,14), greatly advance our understanding of sporadic forms of this disease, they do not resolve the cause of most CHD cases. Moreover, this information has little clinical relevance as it is not directly applicable to families with CHD.

In this study, we set out to design the first NGS CHD gene panel comprising genes previously linked to structural heart disease to assess this tool’s diagnostic potential in families with multiple affected individuals and an apparently Mendelian pattern of inheritance. Identifying the possible cause of CHD in such families would have significant clinical relevance in terms of recurrence risk estimates and family planning.

METHODS

STUDY PARTICIPANTS. Ethical approval for this study was obtained from the Sydney Children’s Hospitals Network Human Research Ethics Committee (approval number CHW/2006/123). Individuals with structural CHD and family histories of CHD with an apparently Mendelian inheritance pattern were selected from the Kids Heart Research DNA Bank. Families were excluded if CHD cases were unable to be confirmed via echocardiography and/or they already had a definite or tentative genetic diagnosis. Participating families included a proband with CHD, an immediate family member affected by CHD, and a minimum of one other family member (immediate or extended) with CHD (mean number of affected family members: 4). In most cases, the proband was selected for deoxyribonucleic acid (DNA) analysis but in cases where DNA quantity was inadequate, an immediate family member affected with the same or similar phenotype was selected for
analysis. A total of 16 CHD cases were included in this study, with the number limited by the size of the targeted NGS capture kit used.

**CONTROLS.** To distinguish between possible disease-causing variants and normal population variations, 15 healthy controls from the Kids Heart Research DNA Bank were screened using the CHD panel. All controls had no self-reported history of CHD within 3 generations. Principal component analysis was performed to ensure the controls and CHD cases were ethnically matched using 646 bi-allelic single nucleotide polymorphism (SNP) positions that were present in HapMap samples. After projecting our data onto 415 HapMap samples, no separation was evident between cases and controls, suggesting appropriate matching. *Supplemental Figure 1* shows the results of the principal component analysis.

**CHD PANEL DESIGN.** Harnessing information from various fields, 57 genes were included in the CHD panel. (For a full list of genes included in the panel, see *Supplemental Table 1*.) We used the web application Sure Design (Agilent Technologies, Santa Clara, California) to create a custom SureSelect target enrichment library of the 57 selected genes. Target parameters were manipulated to optimize coverage of all coding exons, 5’ untranslated regions (UTRs) and 3’UTRs, as well as 25 base pair extensions from the 5’ and 3’ ends. A repeat design at reduced stringency was performed to cover missing target regions >80 bp.

**SAMPLE PREPARATION.** All genomic DNA samples from the Kids Heart Research DNA Bank were extracted from whole blood using the QIAamp DNA Blood Midi kit (for 2 ml blood samples; Qiagen GmbH, Hilden, Germany) and the DNA Isolation Kit for Mammalian Blood (for 5 to 10 ml blood samples; Roche Diagnostics, Indianapolis, IN) following the manufacturer’s protocol. The quality and quantity of the DNA samples were measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA) and 3 ug of DNA from each sample was used for analysis.
TARGETED NEXT-GENERATION SEQUENCING. Targeted NGS (including library construction, capture, and sequencing) was carried out at Oxford Gene Technologies (Oxfordshire, United Kingdom). Enrichment of target regions and library preparation was performed using the SureSelectXT2 Custom kit (1kb-499kb,16) according to the SureSelect protocol version 1.2 (Agilent Technologies, Santa Clara, CA). Library concentrations were determined using Agilent’s QPCR NGS Library Quantification Kit (G4880A) with each sample at a final concentration of 10 nM. A HiSeq2000 sequencer was used to sequence the samples using TruSeq v3 chemistry and protocols (Illumina, Inc., San Diego, CA). The CHD cases and controls were analyzed separately with all CHD cases grouped into one enrichment kit and sequencing run and, subsequently, all control samples into another enrichment kit and sequencing run.

DATA ANALYSIS AND FILTERING. Preliminary data analysis (including read alignment, variant calling, and annotation) was carried out by Oxford Gene Technologies (Oxfordshire, United Kingdom). (For detailed data analysis, see Supplemental Materials and Methods: Preliminary data analysis).

All variants affecting protein-coding regions for each sample were categorized into ‘novel’ and ‘known’ variants according to their presence in dbSNP 137. The minor allele frequencies (MAF) of all known variants were reported according to their presence in the National Heart, Lung, and Blood Institute Exome Sequencing Project (ESP) or according to dbSNP 137 if not present in ESP. All variants were subjected to in silico analysis which included prediction programs such as SIFT, PolyPhen2, and Condel, as well as Alamut (Interactive Biosoftware; v.2.3 rev.2) for splicing predictions where appropriate. In addition, literature reviews were conducted on each variant to establish any previous association with CHD.
To compare the number and types of variants between CHD cases and controls, non-synonymous variants were further grouped according to their PolyPhen2 scores and reported MAF in both the 1000 Genomes Project and ESP.

Despite the high coverage of target regions, CNV analysis could not be conducted as no baseline measure could be established due to the higher coverage observed in all of the CHD cases compared with the controls. Furthermore, the number of samples was not large enough to allow CNV calling using CHD cases only.

**STATISTICAL ANALYSIS.** All data relating to the number and types of variants observed in the CHD cases versus controls were analyzed using SPSS version 21 (IBM, Armonk, New York). Descriptive statistics were obtained for each variant group. Poisson log linear regression was applied to all comparative analyses between groups in the CHD cases versus controls. (Poisson regression is used to model count data and as the number of variants in each individual case and control are in ‘count’ format, this was the recommended model of analysis.) The potential disease-causing variants identified in the 5 families following the segregation analyses were removed from the Poisson analysis, thereby only comparing the variants of unknown significance between the CHD cases and controls.

**VARIANT VALIDATION.** Variants warranting further investigation included novel variants predicted to be ‘probably damaging’ according to PolyPhen2 predictions or known variants predicted to be ‘probably damaging’ and with MAF <0.1%. All such variants were confirmed via bidirectional Sanger sequencing and disease segregation studies were carried out on all available family members. Variants considered to be segregating with disease were present in all affected family members and absent in unaffected members. However, as CHD commonly features reduced penetrance, variants present in unaffected family members were considered, provided that the segregation was consistent with inheritance in the family. The
primer sequences used to analyze each variant via polymerase chain reaction are listed in Supplemental Table 2.

RESULTS

CHD PANEL COVERAGE. In CHD cases, a minimum of 99.29% of target regions were covered to a depth of at least 20x with an average read depth of 1,873x in CHD cases. Similarly, in control samples, a minimum of 98.22% of target regions were covered to a depth of at least 20x with an average read depth of 307x. In terms of individual gene coverage, 15 genes were covered 100%. Target regions for 20 genes were covered >99%, target regions in 19 genes were covered >97%, and target regions in 3 genes were covered >92%. The majority of uncovered target regions in individual genes were in the 5’ and 3’ UTRs as well as upstream and downstream regions, with minimal uncovered exonic regions.

NUMBER AND TYPES OF VARIANTS IDENTIFIED IN CHD CASES AND CONTROLS. In total, 91 variants affecting protein coding regions were identified, 17 of which were novel – with 10 variants unique to CHD cases, 6 unique to controls, and 1 present in several CHD cases as well as a single control. Many known variants were present in multiple CHD cases and controls, including 25 known variants predicted to be ‘probably damaging’ according to various prediction programs. Supplemental Table 3 lists all the variants affecting protein-coding regions identified in this study cohort along with the observed frequencies in CHD cases versus controls, as well as their reported frequencies according to their ESP and PolyPhen2 scores and predictions. The variants affecting protein-coding regions identified in each CHD case and control are presented in Supplemental Tables 4 and 5 respectively.

We calculated the mean number of variants for each CHD case and control (Table 1). An indel variant was identified in 2 controls (2/15 = 0.13 as per Table 1); however, no
splicing or nonsense variants were identified in any controls compared with 2 nonsense variants in 2 CHD cases, 1 splicing variant in a single case, and 9 indels in 8 CHD cases (12/16 = 0.75 as per Table 1). All variants were further classified according to their MAF and nonsynonymous variants according to their PolyPhen2 classification.

**COMPARISONS BETWEEN GROUPS OF VARIANTS IN CHD CASES AND CONTROLS.** Poisson log linear regression revealed a significant difference in the ‘indels, nonsense, and splice’ variants with 4.69 times as many of these variants in the CHD cases (95% confidence interval [CI]: 1.03 to 21.36; p = 0.05). We also identified significantly more ‘probably damaging’ variants with 1.63 times as many in the CHD cases than controls (95% CI: 1.02 to 2.62; p = 0.04). There was a significant difference in the total number of ‘rare’ and ‘low’ frequency variants (MAF <0.05) with 1.71 times as many of these variants in the CHD cases than controls (95% CI: 1.12 to 2.61; p = 0.01). Figure 1 shows the comparison in mean number of variants per individual using the different variant classifications for CHD cases and controls.

**LIKELY PATHOGENIC VARIANT ANALYSES.** Thirteen variants warranting further investigations were identified in 9 of the 16 CHD cases. All 13 variants were predicted to be ‘probably damaging’ according to the various prediction programs and with PolyPhen2 scores above 0.985. Of these, 6 variants were novel and not present in dbSNP 137 or ESP. The remaining 7 all had MAF <0.1% aside from one variant, *NKX2-5 R25C*, which had a MAF = 0.95%. As this variant has previously been reported to be associated with CHD (15-17), it was included for further analysis.

Disease segregation studies in the 13 variants revealed 5 variants that cosegregated with disease and are likely pathogenic variants. The diagnostic performance of the CHD panel is therefore 31% (5 out of 16 families). Table 2 lists the 13 variants for which disease segregation studies were done, as well as the respective MAF and segregation study results;
Figure 2 illustrates the pedigrees of the 5 families segregating the likely pathogenic variants as well as details on these variants. Pedigrees of the remaining families in which segregation studies were conducted but no likely pathogenic variants identified are illustrated in Supplemental Figure 2.

DISCUSSION

The CHD gene panel is the first tool of its kind to offer families affected by Mendelian forms of CHD a reasonable chance at a genetic diagnosis by efficiently and effectively interrogating a large number of genes linked to CHD through NGS technology utilization. We consider gene panels more clinically applicable than exome sequencing because of the quicker turnaround times (due to reduced sequencing volume and associated data analysis), higher and more reliable coverage, and ability to avoid incidental findings. We applied the CHD gene panel to families with strong histories of structural heart disease in an attempt to identify the variants responsible for their disease (Central Illustration). The CHD panel was able to identify variants that are likely the cause of CHD in 31% of the families analyzed. These results importantly provide a clear explanation for the CHD in the family and confirm the presumed mechanism of inheritance. The knowledge opens up options for future pregnancies for those found to carry the familial mutations (e.g., pre-implantation genetic diagnosis). Furthermore, finding a causative mutation in certain genes, such as in TBX5, NKX2-5, or TBX20, would suggest screening for conduction abnormalities or adult-onset cardiomyopathies in mutation-positive family members, as conduction defects and cardiomyopathies can occur in the absence of structural CHD (18,19).

PATHOGENIC VARIANTS. The variant identified in family 12946, TFAP2β R285Q, is a known pathogenic variant associated with Char syndrome, an autosomal dominant disorder affecting the heart, hands, and face. This variant is found in the basic domain, a highly
conserved region critical for DNA binding (20). Affected individuals usually have 3 distinctive features: a patent ductus arteriosus (PDA), distinctive facial appearance, and aplasia or hypoplasia of the middle phalanges of the fifth fingers. No hand abnormalities are present in family 12946, but these may be subtle and only detectable using radiological imaging. Some individuals in the family had toe syndactyly, a less common feature of Char syndrome (21). There is a suggestion of incomplete penetrance in family 12946 with some mutation-positive individuals having no history of PDA or other cardiac disease.

Variants in \( \text{TBX5} \) were identified in 2 families presenting with dominant forms of secundum atrial septal defect and/or ventricular septal defect. Mutations in \( \text{TBX5} \) cause Holt-Oram syndrome, a highly penetrant, autosomal dominant disorder characterized by upper limb abnormalities and structural heart disease, particularly cardiac septal defects (22). \( \text{TBX5} \) plays a critical role in cardiac development and interacts with other transcription factors, including \( \text{GATA4} \) and \( \text{NKX2-5} \), to regulate cardiac chamber septation and development of the conduction system (23,24).

Most of the mutations reported occurred within the highly conserved T-box domain, which ranges from amino acid residues 55 to 238. Interestingly, the position of the mutation within the T-box seems to affect phenotypic expression with mutations at the N-terminal of the T-box resulting in a predominately cardiac phenotype, while mutations at the C-terminal end result in a predominantly skeletal phenotype (25). The first variant identified in this study, \( \text{TBX5} \ D166G \), which is in the T-box, was present in all affected family members and absent in the unaffected mother. Cross et al. reported a variant in close proximity to the one identified in our study, \( \text{G169R} \), in a family displaying primarily cardiac defects and only subtle skeletal abnormalities (26); thus it is possible that mild skeletal abnormalities have been overlooked in this family. Further studies on the \( \text{G169R} \) variant revealed that while there was little or no effect on DNA binding affinity, this variant resulted in a complete loss
of synergistic transcription activity between \( TBX5 \) and \( NKX2-5 \) \( (27) \). The second \( TBX5 \) variant \( (TBX5\ Y407X) \) was identified in a family with septal defects and conduction defects, another common feature of Holt-Oram syndrome. This particular variant is a stop-gain resulting in a truncated protein product at position 407. Termination occurs in the last coding exon of the gene and would therefore be predicted to escape nonsense-mediated decay. However, reports of a stop-gain at position 456, known to cause Holt-Oram syndrome, supports the pathogenicity of this variant \( (28) \). Segregation of this variant with the cardiac phenotype provides additional support for its pathogenic status. While this family has undergone clinical genetic assessment, subtle skeletal features such as distally placed thumbs or mildly hypoplastic clavicles might have been missed without radiological imaging.

A splicing variant, \( c.950-3C>G \), in the elastin \( (ELN) \) gene was identified in family 12637 with supravalvular aortic stenosis (SVAS). Williams syndrome is caused by deletion of a region at \( 7q11.23 \), which includes \( ELN \) and several other genes. SVAS is present in 60\% to 84\% of affected individuals and is attributed to haploinsufficiency for \( ELN \) \( (29) \). Mutations in \( ELN \) have been reported in nonsyndromic forms of SVAS, too, and many of these are splice-site variants \( (30,31) \). The splice-site variant identified in the proband of family 12637 affects a conserved residue in the 3’ acceptor splice site in intron 17, resulting in premature splicing according to multiple splicing prediction programs. Incorrect splicing produces an altered amino acid sequence for exon 18 prior to termination at position 323. Segregation studies further support pathogenicity as all 3 affected family members (and no unaffected individuals) are heterozygous for this variant.

\( NOTCH1 \) has predominantly been linked to left-sided cardiac lesions \( (32) \); however, in family 11756 variant \( G200R \) was identified in 2 second cousins displaying right-sided lesions. Second cousins only share one-thirty-second (3.13\%) of their DNA, so the observation that these affected cousins share this novel, ‘probably pathogenic’ variant (as
predicted by all 3 prediction programs used) suggests a role in disease mechanism. Yet the apparently incomplete penetrance in the family members linking the 2 second cousins makes it difficult to be certain of the role of this variant as the cause of CHD in this family.

The reported NKX2-5 variant, R25C, identified in family 13041, has previously been linked to a number of CHD lesions, including tetralogy of Fallot, interrupted aortic arch, and hypoplastic left heart syndrome (15-17). The significance of this variant, however, remains unclear as it has been detected in healthy controls (33) and is present in ESP with a frequency of ~1%. Functional analysis of this variant demonstrated a slight impairment in dimerization in the mutant protein versus wild-type protein (34), suggesting a possible role in disease development. Still, the fact that the unaffected mother in family 13041 is homozygous for the R25C variant suggests this variant is likely benign.

To summarise, in three of the five families with pathogenic variants, the heart defect could be associated with a previously undiagnosed syndrome, albeit with minimal extra-cardiac manifestations in most affected individuals. However irrespective of whether the eventual diagnosis is syndromic or non-syndromic CHD, this information has the potential to make a profound impact on these families. In some families, promising causal variants were identified, such as the TBX20 R420X variant in family 12944, however segregation analyses were not supportive of pathogenicity. While a stop-gain variant is hard to ignore, the fact that this variant is present in ESP (albeit very rare), is near the end of the protein product as opposed to within the highly conserved T-box, together with non-segregation with disease in the family, points towards it not being a sole contributor to disease phenotype.

**STUDY LIMITATIONS.** Our findings highlight the importance of variant segregation analyses in determining variant pathogenicity, even if this approach did not identify pathogenic variants in the majority of families. This may have occurred because these families may carry mutations in genes not targeted by this study given that we chose to
examine the most highly implicated genes in CHD causation. Alternatively, these families may carry mutations in these genes but they lie in regions other than those captured and resequenced here. Also, while there was good reason to think these were families with Mendelian inheritance with CHD due to mutations in a single gene per family, in some of the families, more than one pathogenic variant may have been segregating, resulting in CHD due to di- or oligogenic inheritance, as suggested in the study by Granados-Riveron et al. (35). This could cause misleading results in segregation studies, so that a genuinely pathogenic variant is wrongly rejected due to apparent non-segregation with phenotype. Lastly, it is possible that one or more of the families, despite appearing to show Mendelian inheritance, actually had CHD due to multifactorial causation. Future exome sequencing may be able to resolve some of these possibilities.

**ADDITIONAL VARIANTS.** We have demonstrated the success of the CHD panel in identifying pathogenic variants in families with apparently Mendelian forms of CHD. Our efforts raise further questions, however, regarding the additional protein-altering variants detected in the CHD cases and their relative contribution to disease phenotype. We identified a significant increase in ‘probably damaging’ variants, ‘indels, nonsense, and splice’ variants, and a number of ‘rare’ and ‘low’ frequency variants (MAF <0.05) in CHD cases versus controls. At present, we do not fully understand the significance and contribution of these additional variants towards the cardiac phenotype, but they may be genetic modifiers of disease. A larger cohort would be required to clarify their significance and contribution towards disease.

**CONCLUSIONS**

Families with Mendelian forms of CHD can be considered candidates for genetic screening using a gene panel-based approach. The value of being able to offer families a chance at
obtaining a genetic diagnosis cannot be underestimated both from a clinical and psychosocial perspective. The ability to identify pathogenic variants in families has significant implications for future family planning as well as assessments for associated conditions, such as conduction defects or cardiomyopathies in certain genes. We have demonstrated a promising diagnostic yield in familial forms of CHD; however, further research and development of this tool, particularly in a clinical setting, is required to assess its practical application.
PERSPECTIVES

COMPETENCY IN PATIENT CARE: In up to 31% of families in which multiple relatives are affected by congenital heart disease (CHD), a genetic basis can be identified, and rapid screening for disease-related genes can be facilitated using advanced sequencing technologies.

TRANSLATIONAL OUTLOOK: While the diagnostic yield of CHD gene panels appears promising in identifying familial forms of CHD, further research is needed to assess its utility in clinical practice.
REFERENCES


FIGURE LEGENDS

Central Illustration. Next generation sequencing gene panel study to identify inherited variants for structural heart disease.

A gene panel covering 57 genes previously implicated in congenital heart disease (CHD) identified 13 potential disease-causing variants in 9 families. Of these, pathogenic variants were seen in 5 (31%) families.

FIGURE 1 Comparisons of Variants of Unknown Significance

Poisson log linear regression was used to compare the groups of variants in the congenital heart disease (CHD) cases and controls. Variants are presented as means per individual across the various classifications. Interpretation of comparisons between groups should be taken in the context of the size of the cohort (n = 31): variants are classified according to PolyPhen2 predictions (A); variants are classified according to MAF (B); and variants with MAF <0.05 are classified according to PolyPhen2 (C). Error bars show 95% confidence intervals.

*\( p \leq 0.05 \); **\( p \leq 0.01 \).

MAF = minor allele frequency.

FIGURE 2 Family Pedigrees and Variant Details for Families with Disease-segregating Variants

Family 12946 segregating the \( TFAP2\beta R285Q \) variant (A); family 10861 segregating the \( TBX5 Y407X \) variant (B); family 10006 segregating the \( TBX5 D166G \) variant (C); family 12637 segregating the \( ELN \) (c.950-3C>G) variant (D); and family 11756 segregating the \( NOTCH1 G200R \) variant. + = mutation positive; − = mutation negative; ▼ = proband.
TABLE 1 Mean Number of Variants Identified

<table>
<thead>
<tr>
<th>Type of Variant</th>
<th>Mean Number per CHD</th>
<th>Mean Number per Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case</td>
<td>Control</td>
</tr>
<tr>
<td>Total coding variants</td>
<td>16.13 (2.31)</td>
<td>13.27 (2.55)</td>
</tr>
<tr>
<td>Novel</td>
<td>0.94 (0.77)</td>
<td>0.47 (0.52)</td>
</tr>
<tr>
<td>‘Indels/nonsense/splice site’</td>
<td>0.75 (0.17)*</td>
<td>0.13 (0.35)</td>
</tr>
<tr>
<td>Classification according to PolyPhen2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Probably damaging’</td>
<td>3.13 (1.02)*</td>
<td>1.80 (1.21)</td>
</tr>
<tr>
<td>‘Possibly damaging’</td>
<td>1.31 (0.95)</td>
<td>1.07 (0.88)</td>
</tr>
<tr>
<td>‘Benign’</td>
<td>10.81 (2.37)</td>
<td>10.27 (2.46)</td>
</tr>
<tr>
<td>Classification according to MAF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Rare’ (MAF &lt;0.01)</td>
<td>2.44 (1.15)</td>
<td>1.60 (1.35)</td>
</tr>
<tr>
<td>‘Low frequency’ (MAF = 0.01–0.05)</td>
<td>1.63 (0.96)**</td>
<td>0.60 (0.91)</td>
</tr>
<tr>
<td>‘Common’ (MAF &gt;0.05)</td>
<td>12.00 (1.93)</td>
<td>11.07 (2.22)</td>
</tr>
</tbody>
</table>

Values are n (SD).

CHD = congenital heart disease; MAF = minor allele frequency; SD = standard deviation.

*p < 0.05; **p < 0.01.
Novel protein-coding variants
Indel, nonsense & splice-site variants
Possibly damaging protein-coding variants
Benign protein-coding variants
Total protein-coding variants

Variant type according to PolyPhen2

Novel protein-coding variants
Indel, nonsense & splice-site variants
Possibly damaging protein-coding variants
Benign protein-coding variants
Total protein-coding variants

Variant type according to MAF

Participants
CHD
Controls

Participants
CHD
Controls

Participants
CHD
Controls

Variant type according to MAF

Participants
CHD
Controls

Participants
CHD
Controls

Totals with MAF<0.05 classified according to PolyPhen2

Participants
CHD
Controls

Participants
CHD
Controls

Participants
CHD
Controls