

A Screening Approach to Identify Clinically Actionable Variants Causing Congenital Heart Disease in Exome Data

Running title: *Szot et al.; Clinically actionable variants causing CHD*

Justin O. Szot, BSc(Adv)(Hons), PhD^{1,2*}, Hartmut Cuny, PhD^{1,3*}, Gillian M. Blue, PhD^{4,5}, David T. Humphreys, BSc(Hons), PhD^{1,3}, Eddie Ip, BSc(Hons)^{1,3}, Katrina Harrison, BSc, MSc, FHGSA⁶, Gary F. Sholler, MBBS(Hons), FRACP, FCSANZ, FACC^{4,5,7}, Eleni Giannoulitou, MEng, MPhil, DPhil^{1,3}, Paul Leo, PhD⁸, Emma L. Duncan, MBBS, PhD^{8,9,10}, Duncan B. Sparrow, BSc(Hons), PhD^{1,3}, Joshua W. K. Ho, BSc(Hons), PhD^{1,3}, Robert M. Graham, MBBS, MD^{1,3}, Nicholas Pachter, MBChB, FRACP, MD^{6,11}, Gavin Chapman, BSc(Hons), PhD^{1,3†}, David S. Winlaw, MBBS(Hons), MD, FRACS^{4,5†}, Sally L. Dunwoodie, BSc(Hons), PhD^{1,2,3†}

¹Victor Chang Cardiac Research Institute; ² Faculty of Science, ³Faculty of Medicine, UNSW; ⁴The Children's Hospital at Westmead, Heart Centre for Children; ⁵Sydney Medical School, University of Sydney, Sydney, NSW; ⁶Genetic Services of Western Australia, Perth, WA; ⁷Sydney Children's Hospitals Network, NSW; ⁸Institute of Health and Biomedical Innovation, Faculty of Health, QUT; ⁹Department of Endocrinology and Diabetes, Royal Brisbane and Women's Hospital; ¹⁰Faculty of Medicine, University of Queensland, Brisbane, QLD; ¹¹School of Paediatrics and Child Health, UWA, Perth, WA, Australia

*contributed equally / † contributed equally

Correspondence:

Sally Dunwoodie, BSc(Hons), PhD

Victor Chang Cardiac Research Institute

Lowy Packer Building

405 Liverpool Street, Darlinghurst

Sydney NSW 2010, Australia

Tel: +61 2 9295 8613

Fax: +61 2 9295 8668, Email: s.dunwoodie@victorchang.edu.au

Journal Subject Terms: Genetics; Developmental biology; Congenital Heart Disease

Abstract:

Background: Congenital heart disease (CHD), structural abnormalities of the heart that arise during embryonic development, are the most common inborn malformations, affecting up to 1% of the population. However, currently only a minority of cases can be explained by genetic abnormalities. The goal of this study was to identify disease-causal genetic variants in 30 families affected by CHD.

Methods and Results: Whole-exome sequencing was performed with the DNA of multiple family members. We utilized a two-tiered whole-exome variant screening and interpretation procedure. Firstly, we manually curated a high-confidence list of 90 genes known to cause CHD in humans, identified predicted-damaging variants in genes on this list, and rated their pathogenicity using ACMG-AMP guidelines. In three families (10%), we found pathogenic variants in known CHD genes *TBX5*, *TFAP2B*, and *PTPN11*, explaining the cardiac lesions. Secondly, exomes were comprehensively analyzed to identify additional predicted-damaging variants that segregate with disease in CHD candidate genes. In ten additional families (33%), likely disease-causal variants were uncovered in *PBX1*, *CNOT1*, *ZFP36L2*, *TEK*, *USP34*, *UPF2*, *KDM5A*, *KMT2C*, *TIE1*, *TEAD2*, and *FLT4*.

Conclusions: The etiology of CHD could be explained using our high-confidence CHD genes list for variant filtering in a subset of cases. Furthermore, our unbiased screening procedure of family exomes implicates additional genes and variants in the pathogenesis of CHD, that suggest themselves for functional validation. This two-tiered approach provides a means of (1) identifying clinically actionable variants and (2) identifying additional disease-causal genes, both of which are essential for improving the molecular diagnosis of CHD.

Key words: congenital heart disease; exome; genetic variation; genetic testing; CHD gene list

Introduction

Congenital heart disease (CHD) is the most common developmental defect in humans and also one of the most severe causes of infant morbidity, affecting approximately 8 per 1000 live births or 1.35 million infants each year worldwide.¹ Moreover, with advancements in prenatal diagnosis and surgical corrections, the number of newborn deaths has declined and more children with CHD survive into adulthood, including those with severe abnormalities of the heart. The recurrence risk is 1-10%, indicative of a genetic basis of disease, depending on the type of CHD and whether the father or mother has CHD.² Therefore, it is necessary to reach a genetic diagnosis in affected families.

Generally, the etiology of CHD is still poorly understood and in only about one third of cases can an inherited or sporadic genetic abnormality be uncovered as the cause of the defect.^{3,4} Single gene mutations can cause some cases of both syndromic and isolated CHD, but many other cases are believed to be multifactorial, implicating multigenic and/or environmental factors in disease causation.^{4,5} Non-syndromic CHD characterized by Mendelian inheritance is frequently caused by a monogenic mutation, but most non-syndromic CHD occurs sporadically.^{3,5}

Massively parallel sequencing technology and variant screening have become widely adopted tools for genetic analysis in research and diagnostics, and a growing number of susceptibility genes causing non-syndromic types of CHD have been identified.^{4,6} Unlike past genetic analyses that only focused on single gene loci or a panel of candidate genes, whole-exome sequencing (WES) can provide gene variant frequencies in populations, which in turn give important insights about the true effects of variants, and facilitates the clinical diagnosis of sporadic and Mendelian diseases, respectively.^{7,8} WES offers advantages over whole-genome

sequencing in terms of costs and speed as well as data volume, analysis, and interpretation.^{7,9,10} WES has identified many human genes that cause CHD when mutated.^{11,12} Considering that targeted deletion of more than 500 mouse genes leads to cardiac structural defects (<http://www.informatics.jax.org/>), it is expected that exome and genome sequencing will implicate many more human genes in CHD in the future.¹³ The challenge, however, is to recognize disease-causal variants and distinguish them from numerous benign variants present in every genome. Nevertheless, to quickly obtain diagnostic results for patients or families with CHD, focusing on known disease-causal genes has been reported as a useful approach.^{14,15}

In this study we exome-sequenced a cohort of 30 Australian families with various forms of sporadic or Mendelian CHD. First, we curated a contemporary high-confidence list of genes known to cause human CHD, and we report how many cases can be explained by damaging variants in these genes. We use the updated standards and guidelines for the clinical interpretation of sequence variants with respect to human diseases developed by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) (ACMG-AMP guidelines)¹⁶ to rate pathogenicity and thus identify clinically actionable findings. Second, we comprehensively analyzed whole exomes within families to identify potential disease-causative variants in genes not currently considered causal of human CHD, but plausible based on known gene function and segregation with disease in families. Our study provides an approach that rapidly yields clinically actionable results, and also identifies potentially causal variants that require further information for them to be clinically actionable.

Methods

As this analysis uses patient samples, the data, analytic methods, and study materials will not be made available to other researchers for purposes of reproducing the results or replicating the procedure.

Study participants

Ethical approval for this study was obtained from the Sydney Children's Hospital Network Human Research Ethics Committee (approval number HREC/16/SCHN/73). The cohort consisted of 30 unrelated families of which 18 were recruited at Princess Margaret Hospital for Children, Perth, Australia and 12 were selected from the Kids Heart Research DNA Bank based at Children's Hospital at Westmead, Sydney, Australia. Families originating from Perth were recruited at presentation to a pre-admission clinic prior to cardiac surgery. They included individuals with family histories of CHD and sporadic cases, whereas only families with more than one affected family member were selected from the Kids Heart Research DNA Bank in Sydney. Heart defects in all affected individuals were confirmed using echocardiography. In all families, at the minimum, a trio (parents and proband) was analyzed. Principal component analysis confirmed the self-declared ethnicities of all individuals in the cohort (**Supplemental Figure S1**).

Sample preparation and WES library construction

Blood samples were processed at the Kids Heart Research DNA Bank, Sydney. Whole blood samples originating in Perth were transported to Sydney on dry ice. Genomic DNA was extracted as described previously.¹⁴ Exome libraries were prepared using the Nextera Rapid Capture Exome kit (Illumina, Inc., San Diego, California) and sequenced on Illumina HiSeq2000

and HiSeq2500. Sequencing coverage analysis was performed using Bedtools (<http://bedtools.readthedocs.io/>) in conjunction with custom scripts.

Analysis of sequencing data

Following exome sequencing, DNA sequence reads were aligned to the human reference genome build hg19 using the Burrows-Wheeler Aligner.¹⁷ SNV calling was performed using Genome Analysis Toolkit (GATK) software (Broad Institute, Cambridge, MA) following the germline variation best practices and using the option of HaplotypeCaller. The generated Variant Call Format (VCF) files from multiple samples were then processed together via a Joint Genotyping call with GATK to produce one multi-sample VCF file for all DNA variants. Variants were annotated with various metrics using ANNOVAR (<http://annovar.openbioinformatics.org/en/latest/>). A count for the number of homozygous variants in the Exome Aggregation Consortium (ExAC) database (<http://exac.broadinstitute.org/>)¹⁸ was incorporated with an in-house custom script. The resulting ANNOVAR-annotation file was split into individual families using an in-house custom script.

A control dataset of sequenced exomes was used to help exclude artificial sequencing-derived variants in our samples. The control data consisted of BAM files from 200 reference exome sequences of similar ethnicity, aligned to hg19 using Novoalign (<http://www.novocraft.com>), generated in a previous project using the same sequencing machinery. Sequence variants were identified and a variant frequency table was prepared. We referred to this annotation as the control cohort (CC) number of alleles.

Filtering, prioritization, and validation of variants

A list of 90 high-confidence CHD genes known to cause human CHD when mutated was curated based on stringent criteria and termed high-confidence CHD genes list (hcCHD list) (see

Supplemental Methods and **Supplemental Table S1** for details). The VarSifter tool¹⁹ was used to filter variants based on genomic location, allele frequency, inheritance pattern, and *in silico* predicted-damaging consequence. In the first analysis approach (hcCHD Gene Screen), exomes of affected probands were screened for rare variants in genes on our manually curated hcCHD list. Variants were defined as rare if the minor allele frequency (MAF) was < 0.01 ($< 1\%$) in the ExAC and 1000 Genomes Project databases.²⁰

In the second analysis approach (Comprehensive Screen), inheritance filters were applied to prioritize rare and damaging variants segregating with disease within any gene, and for each family all possible Mendelian inheritance models were tested. MAF thresholds in ExAC and 1000 Genomes were set to < 0.001 ($< 0.1\%$) for *de novo* and autosomal dominant inheritance models, and to < 0.02 (2%) for autosomal recessive and compound heterozygous models. In addition, thresholds for the homozygous variant count in ExAC and heterozygous count in the CC, respectively, were set to < 1 for *de novo* and autosomal dominant, < 2 (ExAC) and < 4 (CC) for compound heterozygous, and < 3 (ExAC) and < 4 (CC) for autosomal recessive inheritance models.

Variants passing population frequency thresholds were checked with the Integrative Genomics Viewer²¹ to confirm the variant call. Variants in highly polymorphic genes which have notoriously high false-positive variant calls (e.g. mucins) were removed irrespective of read quality. Subsequent consideration of protein-damaging variants first removed synonymous calls. Remaining variants were prioritized in VarSifter by their *in silico* predicted impact on protein function using the Combined Annotation-Dependent Depletion (CADD)²² and PolyPhen-2²³ tools. Variants predicted as possibly or probably damaging by PolyPhen-2 (HumVar score ≥ 0.446) and having a scaled CADD score of > 15 were presumed damaging. Variants lacking

PolyPhen-2 and CADD annotations, e.g. indels, were checked with MutationTaster²⁴ and excluded if the variant call was incorrect. Candidacy of top-scoring variants in the Comprehensive Screen was further interrogated with respect to known gene function, available animal models, published *in vitro* evidence for similar gene variants causing protein dysfunction, and tentative associations of previous human gene variants with CHD. Finally, potential disease-causal variants were evaluated according to the ACMG-AMP guidelines.

Results

Clinical characteristics of the cohort

In total, 120 individuals from 30 families underwent exome sequencing. In 14 families, no history of CHD was reported, in which case the trio, consisting of the unaffected parents and affected proband, was sequenced. For the remaining 16 families with apparent Mendelian CHD, trios plus individuals with available DNA were sequenced. Extracardiac anomalies were found in a subset of patients, but the majority (44/57 affected individuals) had isolated CHD (**Table 1**). Pedigrees and individual clinical findings are provided (**Supplemental Figure S2**). In all cases, the sequencing was of sufficient coverage for analysis (**Supplementary Table S2**).

The hcCHD Gene Screen identified clinically actionable variants and other predicted-damaging variants in known CHD-causing genes

Genetic variants passing the initial automated quality control threshold of the GATK software (approx. 50,000 variants per individual) were subjected to a series of filtering steps. Filtering for variants in genes implicated in CHD has been applied previously^{25,26} but these gene lists tend to be too inclusive and often include genes only linked to CHD by knock-out mouse models. By contrast, we curated a contemporary, high-confidence list of genes known to cause CHD in

humans (hcCHD list) by reviewing publicly available CHD gene panels, gene and phenotype databases, as well as published literature for genes that, when mutated, have been repeatedly reported to cause CHD in humans. Genes were only included in the hcCHD list if variants in the respective gene were reported as the monogenic cause of CHD (isolated or in the context of a syndrome) in $\geq$ three independent familial or sporadic cases in $\geq$ two separate publications. In a few cases, single publications reporting multiple individual *de novo* or autosomal dominant cases with monogenic causative variants were deemed sufficient for inclusion of the respective gene. Importantly, rare predicted-damaging variants in genes on the hcCHD list will attract a “pathogenic” rating by the ACMG-AMP guidelines if the variant segregates with disease. The hcCHD list includes 90 genes that met our stringent criteria defining a hcCHD gene (**Supplemental Table S1**). Curation of the hcCHD list will be ongoing with an up-to-date list accessible online (<https://www.victorchang.edu.au/heart-research/embryology>), making this list a valuable resource for diagnostic and research use.

Across the 30 families, 99 rare variants were identified in genes on the hcCHD list. Rare variants, predicted to be damaging by CADD,²² PolyPhen-2,²³ and MutationTaster2,²⁴ were identified for further scrutiny (**Supplemental Table S3**). Evidence from the literature supported the candidature of seven variants that segregated with disease in the respective families (Table 2; **Supplemental Figure S2**). Variants in three genes (*PTPN11*, *TBX5*, *TFAP2B*) are rated pathogenic by the ACMG-AMP guidelines, and are thus considered clinically actionable. A *de novo* missense variant in *PTPN11* (NM_002834.4:c.184T>G, p.Y62D), identified in the proband of Family 13455 with atrial septal defect, patent ductus arteriosus (PDA), and pulmonary valve stenosis (**Supplemental Figure S2**), had been previously documented as pathogenic in ClinVar (SCV000057372.3). A dominantly inherited novel *TBX5* stopgain variant

(NM_000192.3:c.1221C>G, p.Y407*) was identified in Family 00860 with atrial and ventricular septal defects. This variant segregates with disease in all six affected family members, and a frameshift variant 38 amino acids further C-terminal has been reported to cause Holt-Oram syndrome including CHD.²⁷ It is possible that subtle upper limb abnormalities, typical for Holt-Oram syndrome, but only identifiable using radiological imaging, were missed during the clinical examination of Family 00860 (**Supplemental Figure S2**).

We also identified a previously described pathogenic *TFAP2B* variant in Family 02946 (NM_003221.3:c.854G>A, p.R285Q, ClinVar: SCV000028717.2), reported to have a dominant-negative effect on the DNA-binding capability of TFAP2B. This variant was present in an obligate parental carrier, potentially indicative of incomplete penetrance, commonly observed in inherited CHD.¹⁴ Like the previous report, the *TFAP2B* variant was found in individuals with PDA and distinctive facial features of CHAR syndrome (**Supplemental Figure S2**).²⁸

Variants were identified in *NOTCH1*, *TBX20*, *ELN*, and *GATA4* that did not pass PolyPhen-2 and CADD thresholds (**Table 2**) but were present in all respective affected family members and exhibited phenotypic concordance with respect to the literature (**Supplemental Figure S2**). These variants, rated as “variants of uncertain significance” (VUS) according to ACMG-AMP guidelines, may contribute to the overall phenotype as part of digenic or oligogenic scenarios. The remainder of the variants are currently rated VUS or benign as they neither segregate with disease nor pass CADD and PolyPhen-2 thresholds (**Supplemental Table S3**).

The Comprehensive Screen identified clinically actionable variants and other predicted-damaging variants in families with sporadic CHD

In a second approach (Comprehensive Screen), each family was analyzed for potential disease-causal variants across all genes in an unbiased way. In 14 families with no history of CHD (proband/parent trios), *de novo* variants are most likely to cause disease; however, for completeness all inheritance models were applied, identifying additional variants (**Supplemental Table S4**). We identified 18 rare *de novo* variants in 14 probands. Of these variants, 11 were predicted to be damaging (**Supplemental Table S4**), but only 4 had links to heart development (**Table 3** and **Supplemental Figure S2**).

Among them is the *de novo* variant in *PTPN11* also identified in the hcCHD Gene Screen. A missense variant in *PBX1* convincingly explains the cardiac defect, Tetralogy of Fallot with absent pulmonary valve, in proband 13467. *PBX1* is an evolutionarily highly-conserved homeodomain-containing transcription factor, which when deleted in mouse, results in very similar cardiac phenotypes (**Supplemental Figure S2**).²⁹⁻³¹ In four other families, we identified predicted-damaging *de novo* variants in genes that have not yet been linked to human CHD (**Supplemental Table S4**). Of these, two missense variants in *CNOT1* and *ZFP36L2* have supportive evidence in the literature (**Supplemental Figure S2**, Families 13478 and 13481). Accordingly, variants in *PTPN11* and *PBX1* were rated pathogenic, according to ACMG-AMP guidelines, and are thus clinically actionable, while *CNOT1* and *ZFP36L2* were rated “likely pathogenic” (**Table 3**).

We also applied a filter for recessive inheritance and identified four homozygous variants predicted to be damaging. None of the affected genes have a known link to CHD. For a compound heterozygous inheritance model, we identified predicted-damaging candidate variants in seven genes (**Supplemental Table S4**), also none of which have a known link to CHD. Filtering variants for an autosomal dominant inheritance model that assumes incomplete

penetrance in the parental generation returned large variant numbers. In these cases, variants in the hcCHD genes were given priority, but none were compelling candidates (**Supplemental Table S4**).

The Comprehensive Screen identified clinically actionable variants and other predicted-damaging variants in families with a history of CHD

The Comprehensive Screen was also applied to 16 families with a history of structural heart defects. As above, for completeness, all inheritance models were applied, identifying additional variants (**Supplemental Table S5**). We identified 1545 rare autosomal dominant variants in 16 probands. Of these, 428 variants were predicted to be damaging, and top candidates present in all affected family members, with supportive evidence in the literature, are highlighted in **Table 3**. In five families, we identified predicted-damaging variants that segregate with disease. In two families, variants explain the cardiac lesions: the stopgain in *TBX5* as identified with the hcCHD Gene Screen and a known pathogenic missense variant in *TEK*.

TEK (NM_000459.4:c.2744G>A, p.R915H; SCV000044247.1; **Table 3**), previously described to cause VSD and cutaneomucosal venous malformations,³² was identified in a trio within which the mother and proband had VSD. This, and the cardiovascular phenotype of *Tek*-null mice,³³ are strong evidence that *TEK* variant is disease-causal in this trio.

We found compelling candidates that segregate with disease in three families: a missense variant in *USP34*, *UPF2*, and a digenic combination of two missense variants in *KDM5A* and *KMT2C* (**Table 3**). The novel *USP34* variant (NM_014709.3:c.863C>T, p.A288V) segregates with disease and a similar missense variant in this gene was previously reported in a patient with diverse CHDs.¹² Cardiac-specific silencing of this gene in *Drosophila* resulted in significant embryonic lethality, highlighting a possible role of this gene in human heart development.³⁴

The novel *UPF2* variant (NM_080599.2:c.2606T>G, p.I869S) is nearby a previously identified variant E858R that perturbs the UPF2-UPF3b interaction, necessary for nonsense-mediated decay (**Supplemental Figure S2**).³⁵ UPF2 functions as part of the nonsense-mediated mRNA decay surveillance machinery and, although the function of UPF2 has not yet been assessed in the heart, loss of this gene in mice results in embryonic lethality by mid-gestation,³⁶ making this variant a compelling candidate.

KDM5A and *KMT2C*, and their respective variants (NM_001042603.2:c.4399C>T, p.R1467W and NM_170606.2:c.14258A>T, p.Q4753L), may perturb heart development through altered histone methylation (**Supplemental Figure S2**, Family 01179).

Accordingly, variants in *TBX5* and *TEK* were rated pathogenic according to ACMG-AMP guidelines, and are thus clinically actionable, while *USP34* and *UPF2* were rated likely pathogenic.

The Comprehensive Screen revealed four additional likely disease-causal variants in the remaining families with apparent incomplete penetrance of CHD: the *TFAP2B* missense variant identified with the hcCHD Gene Screen, a single-base deletion in *TIE1* leading to a frameshift and truncated protein, a stopgain variant in *TEAD2*, and an indel in *FLT4* (**Table 3**).

TIE1 functionally dimerizes with TEK to mediate development and maintenance of the vascular system. Loss of *TIE1* in mice results in abnormal heart morphologies including a thinning of the atrial wall.^{37,38} In this light, the *TIE1* variant (NM_005424.4:c.1341del p.V448Cfs*9), is a likely cause of the septal defects seen in Family 02529 (**Supplemental Figure S2**). TEAD2 regulates cardiomyocyte proliferation and survival during development through the Hippo signaling pathway and regulation of *Pax3* and *Foxa2* transcription factors.^{39,40} The novel TEAD2 stopgain

(NM_001256659.1:c.1195C>T, p.R399*) is therefore a plausible explanation for the CHDs seen in Family 13430 (**Supplemental Figure S2**).

Accordingly, the variant in *TFAP2B* was rated pathogenic according to ACMG-AMP guidelines, and is thus clinically actionable, while variants in *TIE1*, *FLT4*, and *TEAD2* were rated likely pathogenic.

In summary, the hcCHD Gene Screen identified disease-causal variants in three cases (one sporadic and two familial), defined as pathogenic by ACMG-AMP guidelines. The Comprehensive Screen also identified these three variants and rated them as top candidates in the respective families. Unsurprisingly, the Comprehensive Screen did not identify additional disease-causal variants in hcCHD genes, but revealed candidate variants in other genes in a further ten cases (three sporadic and seven familial). In total, we were able to uncover credible disease-causal variants in three families with variants in the hcCHD genes list (3/30; 10%) and in ten families with variants in genes not yet considered causal of CHD (10/30; 33%) (**Tables 2, 3** and **Figure 1**). Furthermore, the unbiased Comprehensive Screen identified numerous additional rare and predicted-damaging variants with no previous association to CHD (**Supplemental Tables S4** and **S5**).

Discussion

Here we applied two approaches to identify disease-causal, genetic mutations in 30 Australian families affected with CHD using exome sequencing. The hcCHD Gene Screen focusing on genes known to cause CHD in humans efficiently explained the disease etiology in 10% of cases, and the Comprehensive Screen identifying predicted-damaging variants in the whole coding

region of the genome detected credible gene candidates potentially causing CHD in an additional 33% of cases.

Currently, clinical genetic testing in this field is restricted to cytogenetic screening of patients with multiple congenital abnormalities, copy number variant (CNV) analysis for severe CHDs associated with known syndromes, and single gene testing in a subset of genes with monogenic syndromic association.⁴¹ However, given the complex etiology of CHD, it can be assumed that more genes are involved in heart development and CHD than those currently listed in diagnostic gene panels. To address this unmet need for an up-to-date list defining human CHD genes, we curated the hcCHD list of genes, variants in which will be classified as pathogenic according to ACMG-AMP guidelines if the variant is rare, predicted-damaging, and segregates with disease (**Table 2**).

Performance of family-based WES analysis

The hcCHD Gene Screen identified a gene variant as the definitive cause of CHD in 3 of 30 families (**Figure 1**, left column). Through subsequent application of the Comprehensive Screen, we identified likely causative variants in 10 additional families (**Figure 1**, middle column). Although the latter 10 candidate genes have not yet been unequivocally linked to CHD, there is substantial protein function and/or animal model evidence to indicate a role in human CHD. Moreover, the Comprehensive Screen returned several additional predicted-damaging variants which we do not currently consider disease-causal because of a lack of data relating them to heart development and/or the general lack of knowledge about their function (see **Supplemental Tables S4, S5, and Supplemental Figure S2**). Future functional studies such as analysis of animal models, or assessment of protein function *in vitro* (**Figure 1**), is warranted to determine the actual consequence of these predicted-damaging variants.

To accurately evaluate our yield in comparison to previous reports, it is necessary to consider sporadic and familial cases in our cohort separately. Among the 14 families with sporadic CHD, we identified one disease-causal variant within a known CHD gene (*PTPN11*). With the Comprehensive Screen, we could identify a plausible variant in three additional families, yielding an overall diagnosis rate of 29% (4 of 14). This diagnosis rate for sporadic cases exceeds the maximum expected ~10% contribution by *de novo* variants, as suggested by recent large-scale studies of sporadic CHD.^{12,25}

Our diagnosis rate was higher for the 16 families with a history of CHD. The hcCHD Gene Screen identified disease-causal variants in two known CHD genes (*TBX5* and *TFAP2B*) and the Comprehensive Screen uncovered candidate variants in seven additional families, yielding an overall diagnosis rate of 56% (9 of 16).

Comparable analyses of familial CHD have achieved successes between 33-46%. As these studies, however, only interrogated variants in known CHD genes (57 to 69 genes)^{14,15,42}, this statistic is appropriately below that of our unbiased Comprehensive Screen. The apparent outperformance of these studies compared to our hcCHD yield can be somewhat accounted for, similar to our *de novo* yield, by relative differences in sample sizes, variant filtering, and interpretation, between studies.

In familial cases, monogenic mutations are more likely to be the cause of CHD, especially if the disease occurs across two or more generations in a Mendelian pattern. By contrast, sporadic cases are more likely to be multifactorial, involving multiple genetic and non-genetic factors.^{4,5} Indeed, cumulative data from published reports show that more cases of sporadic CHD remain unresolved by genetic screening than cases of familial CHD.⁴ The absence of convincing damaging recessive or compound heterozygous genotypes in our study is in line

with ~1% incidence previously reported for non-syndromic CHD cases with no parental consanguinity.^{25,26}

In contrast to other cohort-based CHD studies, we reported variant pathogenicity in the context of ACMG-AMP standards. We have shown that predicted-damaging variants identified in genes on the hcCHD list receive a pathogenic classification and, therefore, are considered causative for diagnostic purposes. As the hcCHD list inevitably grows, due to efforts such as our Comprehensive Screen, an increasing number of pathogenic disease-causal variants will be identified. Periodic reinterpretation of clinical exomes in light of new supportive evidence will significantly improve diagnostic success.⁴³ Unlike target panel sequencing, WES data can be reanalyzed when warranted using up-to-date lists such as the hcCHD list to achieve higher diagnostic rates.

Limitations

A general limitation of family-based WES variant filtering and analysis is that it assumes a monogenic cause of disease. In cases where no plausible gene variant can be found, it is possible that the disease is caused by two or more mutations, which individually need not be particularly rare nor damaging. This has been highlighted in recent studies, utilizing unbiased statistical comparisons from large-scale CHD cohorts,^{25,26} showing that non-syndromic CHD patients carry an exome-wide burden of rare inherited protein-truncating variants. These findings recognize a potential oligogenic contribution to non-syndromic CHD and may partially account for observed incomplete penetrance. Another limitation of using WES data concerns cases in which CNVs are the underlying genetic factor. CNVs are estimated to explain up to 10% of sporadic CHD, and even up to 25% of syndromic CHD, but their identification requires high quality whole-genome sequence data or other specific assays.^{4,6} CNV analysis was attempted from whole-exome

sequence data of this cohort, but variant calls were of insufficient quality to be practically interpretable, as is generally known for exome data.⁴⁴ Finally, pathogenic mutations in non-coding genome regions will not be identified with WES. However, the functional relevance of non-coding variants is largely unknown, and since they are difficult to test, it is likely that the identified causes of CHD will be largely limited to the exome for some time.⁹ A future approach to overcome these limitations is to sequence the genome, limit initial analysis to the exome, integrate CNV analysis, and analyze non-coding regions when *in silico* tools to interpret these regions become available.

Conclusions

Filtering WES data using the hcCHD list of genes known to cause CHD quickly identifies variants that are pathogenic according to the ACMG-AMP guidelines and thus are clinically actionable. The hcCHD list, as it is up-to-date, will be invaluable in a diagnostic setting. Therefore, in probands with CHD requiring surgery, with no identified pathogenic CNVs, and their parents, we propose exome sequence analysis of variants in hcCHD genes as a first-line diagnostic approach. The Comprehensive Screen, however, significantly increases the rate of identifying CHD-causing variants, and also implicates additional candidate genes that are likely to be included in the hcCHD list in the future, ultimately improving the rate of patient diagnosis.

Acknowledgments: We are grateful to the families who participated in this research, to Dr. David Andrews (Princess Margaret Hospital, Perth), and the individuals and agencies that supported this research.

Sources of Funding: This work was supported by National Health and Medical Research Council (NHMRC) [Fellowships ID514900, ID1042002 to S.L.D., and ID1105271 to J.H., and

Program Grant ID1074386 to S.L.D. and R.M.G.]; Australian National Heart Foundation [Fellowship ID100848 to J.H. and ID101204 to E.G.]); The Kirby Foundation to J.H.; Office of Health and Medical Research NSW Government to S.L.D.; Chain Reaction (The Ultimate Corporate Bike Challenge) to S.L.D., D.S.W.; Channel 7 Telethon to S.L.D., R.M.G., and The Key Foundation to S.L.D.

Disclosures: None.

References:

1. van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ, et al. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol*. 2011;58:2241-2247.
2. Blue GM, Kirk EP, Sholler GF, Harvey RP, Winlaw DS. Congenital heart disease: current knowledge about causes and inheritance. *Med J Aust*. 2012;197:155-159.
3. Akhirome E, Walton NA, Nogee JM, Jay PY. The Complex Genetic Basis of Congenital Heart Defects. *Circ J*. 2017;81:629-634.
4. Blue GM, Kirk EP, Giannoulatou E, Sholler GF, Dunwoodie SL, Harvey RP, et al. Advances in the Genetics of Congenital Heart Disease: A Clinician's Guide. *J Am Coll Cardiol*. 2017;69:859-870.
5. Gelb BD, Chung WK. Complex genetics and the etiology of human congenital heart disease. *Cold Spring Harb Perspect Med*. 2014;4:a013953.
6. Muntean I, Toganel R, Benedek T. Genetics of Congenital Heart Disease: Past and Present. *Biochem Genet*. 2017;55:105-123.
7. Petersen BS, Fredrich B, Hoepfner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017;18:14.
8. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Koster J, Marais A, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet*. 2017;25:176-182.
9. Lelieveld SH, Veltman JA, Gilissen C. Novel bioinformatic developments for exome sequencing. *Hum Genet*. 2016;135:603-614.
10. Sun Y, Ruivenkamp CA, Hoffer MJ, Vrijenhoek T, Kriek M, van Asperen CJ, et al. Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat*. 2015;36:648-655.

11. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*. 2015;350:1262-1266.
12. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013;498:220-223.
13. Andersen TA, Troelsen Kde L, Larsen LA. Of mice and men: molecular genetics of congenital heart disease. *Cell Mol Life Sci*. 2014;71:1327-1352.
14. Blue GM, Kirk EP, Giannoulatou E, Dunwoodie SL, Ho JW, Hilton DC, et al. Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. *J Am Coll Cardiol*. 2014;64:2498-2506.
15. LaHaye S, Corsmeier D, Basu M, Bowman JL, Fitzgerald-Butt S, Zender G, et al. Utilization of Whole Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease. *Circ Cardiovasc Genet*. 2016;9:320-329.
16. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405-424.
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
18. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285-291.
19. Teer JK, Green ED, Mullikin JC, Biesecker LG. VarSifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics*. 2012;28:599-600.
20. The Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68-74.
21. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24-26.
22. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310-315.
23. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248-249.

24. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11:361-362.
25. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet*. 2017;49:1593-1601.
26. Sifrim A, Hitz MP, Wilsdon A, Breckpot J, Turki SH, Thienpont B, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet*. 2016;48:1060-1065.
27. Posch MG, Perrot A, Berger F, Ozcelik C. Molecular genetics of congenital atrial septal defects. *Clin Res Cardiol*. 2010;99:137-147.
28. Zhao F, Weismann CG, Satoda M, Pierpont ME, Sweeney E, Thompson EM, et al. Novel TFAP2B mutations that cause Char syndrome provide a genotype-phenotype correlation. *Am J Hum Genet*. 2001;69:695-703.
29. Slavotinek A, Risolino M, Losa M, Cho MT, Monaghan KG, Schneidman-Duhovny D, et al. De novo, deleterious sequence variants that alter the transcriptional activity of the homeoprotein PBX1 are associated with intellectual disability and pleiotropic developmental defects. *Hum Mol Genet*. 2017:ddx363-ddx363.
30. Chang CP, Stankunas K, Shang C, Kao SC, Twu KY, Cleary ML. Pbx1 functions in distinct regulatory networks to pattern the great arteries and cardiac outflow tract. *Development*. 2008;135:3577-3586.
31. Stankunas K, Shang C, Twu KY, Kao SC, Jenkins NA, Copeland NG, et al. Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ Res*. 2008;103:702-709.
32. Wouters V, Limaye N, Uebelhoer M, Irrthum A, Boon LM, Mulliken JB, et al. Hereditary cutaneomucosal venous malformations are caused by TIE2 mutations with widely variable hyper-phosphorylating effects. *Eur J Hum Genet*. 2010;18:414-420.
33. Puri MC, Partanen J, Rossant J, Bernstein A. Interaction of the TEK and TIE receptor tyrosine kinases during cardiovascular development. *Development*. 1999;126:4569-4580.
34. Zhu JY, Fu Y, Nettleton M, Richman A, Han Z. High throughput in vivo functional validation of candidate congenital heart disease genes in *Drosophila*. *Elife*. 2017;6:e22617.
35. Boehm V, Haberman N, Ottens F, Ule J, Gehring NH. 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep*. 2014;9:555-568.

36. Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Monch K, Thoren LA, Nielsen FC, et al. NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.* 2008;22:1381-1396.
37. Patan S. TIE1 and TIE2 receptor tyrosine kinases inversely regulate embryonic angiogenesis by the mechanism of intussusceptive microvascular growth. *Microvasc Res.* 1998;56:1-21.
38. Sato TN, Tozawa Y, Deutsch U, Wolburg-Buchholz K, Fujiwara Y, Gendron-Maguire M, et al. Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature.* 1995;376:70-74.
39. Landin-Malt A, Benhaddou A, Zider A, Flagiello D. An evolutionary, structural and functional overview of the mammalian TEAD1 and TEAD2 transcription factors. *Gene.* 2016;591:292-303.
40. Zhang Y, Del Re DP. A growing role for the Hippo signaling pathway in the heart. *J Mol Med (Berl).* 2017;95:465-472.
41. Ito S, Chapman KA, Kisling M, John AS. Appropriate Use of Genetic Testing in Congenital Heart Disease Patients. *Curr Cardiol Rep.* 2017;19:24.
42. Jia Y, Louw JJ, Breckpot J, Callewaert B, Barrea C, Sznajer Y, et al. The diagnostic value of next generation sequencing in familial nonsyndromic congenital heart defects. *Am J Med Genet A.* 2015;167A:1822-1829.
43. Nambot S, Thevenon J, Kuentz P, Duffourd Y, Tisserant E, Bruel AL, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet Med.* 2017.
44. Hong CS, Singh LN, Mullikin JC, Biesecker LG. Assessing the reproducibility of exome copy number variations predictions. *Genome Med.* 2016;8:82.

Table 1. Clinical characteristics of the affected study participants

Characteristics		Affected individuals (<i>n</i> = 57)	
Sex		<i>n</i>	%
Male		24	(42)
Female		33	(58)
Ethnicity*			
European		33	(58)
East Asian		2	(4)
South Asian		1	(2)
African		3	(5)
Other		6	(11)
Not recorded		12	(21)
Cardiac lesions[†]			
Septal defect		21	(37)
Septal defect with minor abnormalities		10	(18)
Malformation of the outflow tract		7	(12)
AVSD & variants		3	(5)
Obstructive lesion		4	(7)
Functional single ventricle		2	(4)
Other		10	(18)
Extracardiac anomalies			
Developmental delay		7	(12)
Dysmorphic features		5	(9)
Failure to thrive		2	(4)
ECG abnormalities		2	(4)
Duodenal atresia		1	(2)
Congenital hip dysplasia		1	(2)

Tracheomalacia, tongue-tie	1	(2)
Abdominal <i>situs inversus</i> with right sided spleen	1	(2)
Tracheoesophageal fistula	1	(2)
Premature birth	1	(2)
None recorded	44	(77)

Family history of CHD

Yes, other family members affected	43	(75)
No, isolated sporadic case	14	(25)

*Self-declared ethnicity and based on family history.

†The primary cardiac lesions are listed. ‘Other’ includes all CHD subtypes not covered by any of the categories, such as patent ductus arteriosus, patent foramen ovale, right-sided aortic arch, or aberrant subclavian artery.

AVSD = atrioventricular septal defect

Table 2. Top variants identified in the hcCHD Gene Screen

Family ID	Gene	Model*	Genomic position [†]	cDNA change	Amino acid change	ExAC MAF [‡]	PP2 HVAR [§]	CADD	ACMG class. [#]	CHD phenotype(s)**
13455	<i>PTPN11</i>	DN	12:112888168	T>G	Y62D	0	PD	26.9	P	ASD, PS, PDA
00860	<i>TBX5</i>	AD	12:114793673	C>G	Y407*	0	–	40	P	ASD, VSD, MV regurgitation, LV hypoplasia
02946	<i>TFAP2B</i>	AD ^{††}	6:50805720	G>A	R285Q	0	PD	35	P	PDA
01285	<i>NOTCH1</i>	AD	9:139409016	A>G	N718S	0	B	18.92	VUS	MV stenosis, VSD
13430	<i>ELN</i>	AD ^{††}	7:73469100	c.1150+1G>A	intron	5E-05	–	12.06	VUS	AVSD, PDA, PFO
13397	<i>TBX20</i>	AD	7:35288426	C>G	S136S	0	–	–	VUS	ASD, VSD, PDA
02669	<i>GATA4</i>	AD ^{††}	8:11615804	G>A	T383T	5E-05	–	–	VUS	AS, ASD, AVSD PFO

White: causal variants in hcCHD genes; grey: variants of uncertain significance. The four VUS are highlighted because they might contribute to an oligogenic scenario based on presence in all affected family members and concordance with previously observed phenotypes.

***Model:** Inheritance model. DN: *de novo*; AD: autosomal dominant.

[†]**Genomic position:** Refers to the human reference genome build hg19.

[‡]**ExAC MAF:** Minor allele frequency in the ExAC database.

[§]**PP2 HVAR:** PolyPhen-2 HumVar predictive score. PD: Probably Damaging (score ≥ 0.909); PS: Possibly Damaging ($0.447 \leq$ and ≤ 0.909); B, benign (≤ 0.446); –: not applicable.

^{||}**CADD:** scaled CADD score; –: not applicable.

[#]**ACMG class.:** Interpretive classification according to the updated guidelines by the ACMG-AMP. P: Pathogenic; LP: Likely Pathogenic; VUS: Variant of Uncertain Significance.

****CHD phenotype(s):** Summarized main phenotypes in the family. For detailed phenotype descriptions, see Supplemental Figure S2.

^{††} Variants with presumed incomplete penetrance

Table 3. Top predicted-damaging variants identified in the Comprehensive Screen

Family ID	Gene	Model *	Genomic position [†]	cDNA change	Amino acid change	ExAC MAF [‡]	PP2 HVAR [§]	CADD	ACMG class. #	CHD phenotype(s)**
13455	<i>PTPN11</i>	DN	12:112888168	T>G	Y62D	0	PD	26.9	P	ASD, PS, PDA
00860	<i>TBX5</i>	AD	12:114793673	C>G	Y407*	0	–	40	P	ASD, VSD, MV regurgitation, LV hypoplasia
02946	<i>TFAP2B</i>	AD ^{††}	6:50805720	G>A	R285Q	0	PD	35	P	PDA
13467	<i>PBX1</i>	DN	1:164768976	G>C	R184P	0	PD	33	P	TOF
13478	<i>CNOT1</i>	DN	16:58612645	T>G	H514Q	0	PD	20.8	LP	AVSD, DORV, PS
13481	<i>ZFP36L2</i>	DN	2:43452311	G>A	R211H	0	PD	16.53	LP	AS
13397	<i>TEK</i>	AD	9:27212762	G>A	R915H	0	PD	35	P	ASD, VSD, PDA
02944	<i>USP34</i>	AD	2:61607455	C>T	A288V	0	PD	23.9	LP	ASD, PS
02567	<i>UPF2</i>	AD	10:11997475	T>G	I869S	0	PD	22.7	LP	PDA, PFO
01179	<i>KDM5A</i>	AD	12:404795	C>T	R1467W	0	PD	20.1	VUS	PFO, TOF, VSD
	<i>KMT2C</i>	AD	7:151841883	A>T	Q4753L	0	PS	16.91	VUS	
02529	<i>TIE1</i>	AD ^{††}	1:43777344	delC	V448Cfs*9	0	–	–	LP	ASD, PAPVC, PLSVC
13430	<i>TEAD2</i>	AD ^{††}	19:49845739	C>T	R399*	0	–	31	LP	AVSD, PDA, PFO, VSD
03293 [‡]	<i>FLT4</i>	AD ^{††}	5:180047670	delGGT insT	G781Vfs *18	0	–	–	LP	TOF

White: pathogenic variants in hcCHD genes; grey: likely causal variants in genes with supportive evidence in the literature (see **Supplemental Figure 2**)). ***Model**: Inheritance model. DN: *de novo*; AD: autosomal dominant.

[†]**Genomic position**: Refers to the human reference genome build hg19.

[‡]**ExAC MAF**: Minor allele frequency in the ExAC database.

[§]**PP2 HVAR**: PolyPhen-2 HumVar predictive score. PD: Probably Damaging (score ≥ 0.909); PS: Possibly Damaging ($0.447 \leq$ and ≤ 0.909); B: benign (≤ 0.446); –: not applicable.

^{||}**CADD**: scaled CADD score; –: not applicable.

[#]**ACMG class.**: Interpretive classification according to the updated guidelines by the ACMG-AMP. P, Pathogenic; LP, Likely Pathogenic; VUS: Variant of Uncertain Significance.

^{**}**CHD phenotype(s)**: Summarized main phenotypes of the family. For detailed phenotype descriptions, see Supplemental Figure S2.

^{††} Variants with presumed incomplete penetrance

^{‡‡} Further details about Family 03293 will be submitted for publication elsewhere

Figure Legends:

Figure 1. Comprehensive two-tiered whole-exome sequencing analysis to identify disease-causal genetic factors in 30 families with sporadic and familial CHD, respectively. Variant filtration utilizing a list of 90 genes known to cause CHD identified three disease-causal variants in three families (one sporadic and two familial cases). With a comprehensive family-based variant screening procedure, ten additional potentially causal variant candidates in genes not yet linked to CHD in humans could be identified (three sporadic and seven familial cases).

Clinical Perspective

Providing a genetic diagnosis helps anticipate patient outcomes and improves clinical care. In the context of congenital heart disease (CHD), the majority of sporadic cases are not offered genetic testing and only limited genetic counselling, partly because of the perception that the yield is low. To address the unmet need for efficient and actionable genetic diagnosis in a clinical setting, we curated a dynamic high-confidence list of genes (hcCHD genes list), which when mutated in humans result in CHD. Applying this list in the context of the standards and guidelines for clinical interpretation of sequence variants as defined by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology, our genetic diagnostic approach streamlines clinical interpretation, efficiently identifying clinically actionable causes of disease, which may be directly relayed to families for reproductive planning, disease management and prognostic purposes. We utilized this approach on exome data from a cohort of 30 Australian families with sporadic or familial CHD. We identified a causative genetic variant in 3/30 cases. To further empower the molecular diagnosis of CHD, we additionally analyzed entire exomes comprehensively within a defined framework. We addressed rarity, predicted-pathogenicity and segregation with disease, thereby implicating a likely disease-causal variant in an additional 10/30 families. These complementary approaches highlight the utility of exome analysis in CHD cases with clinically relevant findings identified in 13/30 (43%). This study highlights the utility of exome sequencing in CHD populations who would not be conventional candidates for such investigations.