HHS Public Access

Author manuscript

Circ Cardiovasc Genet. Author manuscript; available in PMC 2017 August 01.

Published in final edited form as:

Circ Cardiovasc Genet. 2016 August ; 9(4): 320–329. doi:10.1161/CIRCGENETICS.115.001324.

Utilization of Whole Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease

Stephanie LaHaye, BS[1,2,3,*], **Don Corsmeier, DVM**[4,5,*], **Madhumita Basu, PhD**[1,2], **Jessica L. Bowman, MD**[2,5], **Sara Fitzgerald-Butt, MS**[1,2,5], **Gloria Zender, BS**[1], **Kevin Bosse, PhD**[1], **Kim L. McBride, MD, MS**[1,2,5], **Peter White, PhD**[4,5], and **Vidu Garg, MD**[1,2,3,5]

[1]Center for Cardiovascular Research, The Research Institute at Nationwide Children's Hospital

[2]The Heart Center, Nationwide Children's Hospital

[3]Department of Molecular Genetics, The Ohio State University

[4]Biomedical Genomics Core & the Center for Microbial Pathogenesis, The Research Institute at Nationwide Children's Hospital

[5]Department of Pediatrics, The Ohio State University, Columbus, OH

Abstract

Background—Congenital heart disease (CHD) is the most common type of birth defect with family and population based studies supporting a strong genetic etiology for CHD. The goal of this study was to determine if a whole exome sequencing (WES) approach could identify pathogenic segregating variants in multiplex CHD families.

Methods and Results—WES was performed on 9 kindreds with familial CHD, 4 with atrial septal defects (ASD), 2 with patent ductus arteriosus (PDA), 2 with tetralogy of Fallot (TOF) and 1 with pulmonary valve dysplasia. Rare variants (<1% minor allele frequency) that segregated with disease were identified by WES, and variants in 69 CHD candidate genes were further analyzed. These selected variants were subjected to *in silico* analysis to predict pathogenicity and resulted in the discovery of likely pathogenic mutations in 3/9 (33%) families. A *GATA4* mutation in the transactivation domain, p.G115W*,* was identified in familial ASD and demonstrated decreased transactivation ability *in vitro*. A p.I263V mutation in *TLL1* was identified in an ASD kindred and is predicted to affect the enzymatic functionality of *TLL1*. A disease segregating splice donor site mutation in *MYH11* (c.4599+1delG) was identified in familial PDA and found to disrupt normal splicing of Myh11 mRNA in the affected individual.

Conclusions—Our findings demonstrate the clinical utility of WES to identify causative mutations in familial CHD and demonstrate the successful use of a CHD candidate gene list to

Correspondence: Peter White, PhD, Nationwide Children's Hospital, Center for Microbial Pathogenesis & Biomedical Genomics Core, 700 Children's Drive, Columbus, Ohio 43205, Tel: 614–722-5786, Fax: 614-722-2817, Peter.White@nationwidechildrens.org. Vidu Garg, MD, Nationwide Children's Hospital, Center for Cardiovascular, 700 Children's Drive, Columbus, Ohio 43205, Tel: 614-355-5710, Fax: 614-355-5725, Vidu.Garg@nationwidechildrens.org.

*contributed equally as co-first authors

Disclosures: None

allow for a more streamlined approach enabling rapid prioritization and identification of likely pathogenic variants from large WES data sets.

Clinical Trial Registration—https://clinicaltrials.gov; Unique Identifier: NCT0112048.

Keywords

congenital cardiac defect; genetic testing; exome; GATA4; MYH11; TLL1

Cardiovascular malformations are the most common type of birth defect, affecting ~2% of live births when including bicuspid aortic valve.[1] Advances in the medical and surgical care of these patients have resulted in an increased population prevalence of children and adults with palliated or repaired congenital heart defects (CHD).[2] An increased knowledge of the molecular pathways regulating normal cardiac development by investigations in animal models along with advancements in genetic technologies have aided in the discovery of genetic etiologies for CHD.[3, 4] Even so, there remains a limited application of this new genetic knowledge in clinical practice for the majority of CHD cases.

Historically, the clinical genetics evaluation in CHD focused on those individuals with additional birth defects or developmental delay/intellectual disability, which account for approximately one-quarter of cases.[5] Of these, a significant portion are termed “syndromic” and advances in genetic testing led to the recognition of numerous well-described genetic syndromes, such as 22q11.2 deletion syndrome, which are associated with CHD. More recently, chromosomal microarray (array comparative genome hybridization) has been become increasingly utilized in this population and led to the identification of novel syndromes that are characterized by cardiac along with other birth anomalies.[6] Accordingly, chromosomal microarray has been incorporated into clinical practice for the evaluation of children with multiple birth defects or developmental delay/intellectual disability.[5]

The majority of CHD occurs as isolated birth defects, termed non-syndromic, and population and family-based studies have identified an increased recurrence risk supporting a strong genetic etiologic component for CHD.[7] Familial cases for numerous forms of non-syndromic CHD have been reported and in some cases, where multiple family members were affected, have led to the identification of CHD-causing genes such as *NKX2.5* and *GATA4*.[3, 6, 8] In other cases, candidate gene sequencing has been performed on populations of individuals with non-syndromic CHD and potential disease-causing variants have been identified.[9, 10] Even with these advances, the clinical utility of these findings is not yet clear as genetic testing for non-syndromic CHD is not routinely performed even when a positive family history exists.

High throughput (next generation) sequencing technology has evolved quickly over the last decade allowing for the rapid sequencing and analysis of human genomes within hours.[11,12] Due to limitations in our understanding of the noncoding regions of the genome, much of the current research focuses on the ability to identify disease-causing mutations in protein coding regions of the genome and utilizes whole exome sequencing (WES).[13] WES has proven to be an important alternative to single locus based genetic screening as well as panel based sequencing.[14] Sanger sequencing of individual genes is a relatively inefficient

approach given the considerable amount of time and effort that it involves and while gene panel based approaches overcome this limitation to some extent, they are limited by the *a priori* knowledge of potential disease-causing genes, inability to expand the candidate gene list without resequencing and prior genotype-phenotype associations.

Here, we demonstrate the utility of WES in familial, non-syndromic CHD that implements the use of a CHD gene prioritization strategy. Our strategy allows for the selection of variants occurring in previously implicated CHD-causing genes and results in a more straightforward analysis of variant segregation, allele frequency, pathogenicity prediction, and functional analysis. In our study, we utilized this approach and discovered a likely pathogenic or pathogenic mutation in 3 out of 9 families. Our study is the first to demonstrate the clinical utility of WES in the genetic diagnosis of familial cases of non-syndromic CHD.

Methods

Study Population

The study cohort included a total of 9 families with apparent Mendelian inheritance of CHD. Four families (A, B, C, and D) had atrial septal defects (ASD) with autosomal dominant (AD) inheritance, 2 families (E and F) had patent ductus arteriosus (PDA) with AD inheritance, 2 families (G and H) had tetralogy of Fallot (TOF) with autosomal recessive (AR) inheritance, and 1 family (I) had a dysplastic pulmonary valve with AR inheritance. Pedigrees (Families A-I) with associated phenotype information are shown in Supplemental Figure 1 and Supplemental Table 1. Informed consent was obtained from study subjects or parents of subjects less than 18 years of age (assent was obtained from subjects 9–17 years of age) under protocols approved by the Institutional Review Board (IRB) at Nationwide Children's Hospital and University of Texas Southwestern. Genomic DNA was isolated from blood samples using the 5 PRIME DNA extraction kit (Thermo Fisher Scientific, Pittsburgh, PA).

Exome Sequencing Library Construction

Exome libraries for family D were constructed using Illumina TruSeq Exome Enrichment Kit with v1 capture probes (Illumina, CA). All other genomic DNA samples in the study were processed using Agilent SureSelectXT Target Enrichment System for Illumina Paired End Sequencing Protocol (Agilent Technologies, CA). DNA libraries for families B, C, and E were captured with SureSelect Human All Exon v4 probes. For the remaining samples in the study, the SureSelect Human All Exon v5 kit was used. Paired-end 100 base pair reads were generated for exome-enriched libraries sequenced on the Illumina HiSeq 2000 to a minimum depth of 50X targeted region coverage.

Whole Exome Sequencing Analysis

Primary analysis consisted of using Illumina's Real-Time Analysis (RTA) software to perform base calling and quality scoring from the raw intensity files. The resulting base call (BCL) format files were then converted and demultiplexed using Illumina's bcl2fastq2

Conversion Software into the standard FASTQ file format appropriate for secondary analysis.

Secondary analysis was performed using Churchill, a pipeline developed in house for the discovery of human genetic variation that implements a best practices workflow for variant discovery and genotyping (https://www.broadinstitute.org/gatk/guide/best-practices).[11] Churchill utilizes the Burrows-Wheeler Aligner (BWA) to align sequence data to the reference genome (UCSC build hg19). Duplicate sequence reads were removed using PicardTools (v1.104). Local realignment was performed on the aligned sequence data using the Genome Analysis Toolkit (v3.3–0). Churchill's own deterministic implementation of base quality score recalibration was used. The GATK's HaplotypeCaller (HC) was used to call variants. To maximize sensitivity, variant calling was performed across all samples in the study. Use of the GATK's variant quality score recalibration (VQSR) was excluded in favor of using Churchill's own quality-based variant filtering algorithm.

ANNOVAR and custom in-house scripts were used to annotate the variant call set with mutation and gene information, protein functional predictions, and population allele frequencies.[15] Commonly used heuristic filtering methods based on these annotations were applied. Common variation occurring at greater than 1% minor allele frequency (MAF) in the population was excluded. Variants outside of coding regions (defined as greater than 4 base pairs from an exon splice site) as well as exonic variants coding for synonymous single nucleotide polymorphisms were also dropped. Variants were further filtered, when applicable, based on the pattern of inheritance expected from examination of the pedigree. All families were considered for autosomal dominant model of inheritance. Families G, H, and I were additionally analyzed using homozygous recessive and compound heterozygous models. As shown in Supplemental Figure 1, we did not have DNA samples from all family members, therefore analysis using segregation was not performed in all circumstances. In such cases, we identified variants present in the affected individuals and then sequenced these genes in either unaffected family members or control DNA, to ensure that the variants were not present. We also took into consideration potential environmental risk factors that may lead to phenocopies, as in family F, and allowed for inheritance patterns that excluded such individuals.

CHD Candidate Gene List

A candidate CHD gene list approach was implemented to identify potentially pathogenic mutations in genes that had been previously published to cause CHD based upon a literature review. The following criteria were used to construct our candidate CHD gene list: 1) Published report of identified mutations in the candidate gene in at least 3 sporadic cases with a similar CHD phenotype 2) Published demonstration of disease-segregating mutation in the candidate gene within a family. Based upon our literature review, we identified 69 candidate CHD-causing genes (Supplemental Table 2). Variants identified by WES were filtered by this CHD gene list, and then were subsequently analyzed for segregation amongst available affected family members, when possible. WES "heterozygote" calls (and "homozygote" for Families G, H, I) that segregated with the affected family members subsequently underwent *in silico* analysis to predict pathogenicity of the sequence variant.

In Silico Functional Analysis of Identified CHD Gene Variants

In silico analysis was performed using algorithms to predict pathogenicity of identified sequence variants. The following prediction software was used to analyze the rare variants in candidate CHD genes identified through WES: SIFT, GERP++, Polyphen2 Complex, Polyphen2 Mendelian, PhyloP, FATHMM, and SiPhy. These different prediction software programs utilize algorithms to calculate the potential damage caused by a nucleotide variant by determining the likelihood of the substituted amino acid to affect protein function. Variants that were predicted to be damaging in at least 4 of the 7 algorithms were verified by bidirectional Sanger sequencing and considered for further analysis (Supplemental Tables 3 and 4). For Family C, adult human DNA (#50-181-348, Biochain Institute, Inc.) was used for control since no unaffected family members were available. Based on the *in silico* findings and available previously published literature, these mutations were then classified with the following terms based on ACMG Standards and Guidelines' recommendations: "pathogenic"; "likely pathogenic"; "uncertain significance"; "likely benign"; "benign" (Supplemental Table 5).[16]

Plasmid Construction and Site-Directed Mutagenesis

An expression construct was generated for the murine Gata4 G115W recombinant protein. The G115W point mutation was introduced into the orthologous mouse Gata4 cDNA (NM_008092.3) expression vector containing a FLAG-tag and verified by sequencing using previously published methodology (QuikChange II Site-Directed Mutagenesis Kit, Agilent 200523).[1718]

Luciferase Assays

HeLa cells were transfected using Lipofectamine 2000 (Invitrogen 11668027) with 300 ng of either alpha-myosin heavy chain (α-MHC) or atrial natriuretic factor (ANF) luciferase reporter, and 300 ng of wildtype Gata4 or 300ng of Gata4 G115W, as previously described.[18] Luciferase activity was measured 48hr after transient transfection. Immunoblots were used to verify appropriate protein expression. Three independent experiments were performed in duplicate and statistical comparison was performed using a Student's t-test. P-value < 0.05 was considered statistically significant.

MYH11 Expression Studies in Human Dermal Fibroblasts

Dermal fibroblasts were collected from patient II-2 of Family F by performing a skin biopsy, in accordance with the policies outlined in the Nationwide Children's Hospital Institutional Review Board (IRB) approved protocol. Patient and control dermal fibroblasts (ATCC human adult primary dermal fibroblasts, PCS-201-012) were cultured and 10ng/ml of recombinant *TGFβ-1* ligand was added to the dermal fibroblast media for 48 hours, after 15 hours of serum starvation, to increase expression of *MYH11*. Total mRNA was isolated from these cells with TRIzol (Thermo Scientific 15596026) followed by total RNA purification (Norgen 17200). 1 μg of mRNA was used to synthesize cDNA with the Superscript VILO cDNA synthesis kit (Thermo Scientific 11754). RT-PCR was performed with the following primers: F: 5′-CAAGAAGAAACCCGGCAGAAGCTCAACGTG-3′ and R: 5′-AAAGATCTCATCTCTGGAGGCACGGGCATC-3′, to generate a 1064 bp fragment of

MYH11 (NM_001040113). PCR products were excised from a 1.5% agarose gel and Sanger sequencing was performed.

Results

Identification of Disease-Causing Variants

Individuals from 9 families with Mendelian inherited forms of CHDs were analyzed with WES (Supplemental Figure 1, Supplemental Figure 2 and Supplemental Table 1). Six variants in 4 out of 9 families (Table 1) were identified that met the following criteria: determined to be rare (<1% MAF in population), predicted damaging (in at least 4 out of 7 *in silico* functional algorithms), and segregated with disease in affected individuals (Table 1) (Supplemental Table 2). In 3 families (C, D, F), we identified potentially pathogenic mutations in genes that had been previously implicated in published human genetic studies with similar cardiac phenotypes and these findings are discussed in detail below. While rare, damaging sequence variants were also identified in potential candidate genes in family H, the associated cardiac phenotypes within the family were not consistent with the reported literature. This limited our ability to conclude that these sequence variants were contributing to cardiac phenotypes within these families. No segregating variants were identified in Families A, B, E, G, and I using this approach.

GATA4 Gly115Trp mutation in familial ASD

Family C had five members affected with autosomal dominant inheritance of ASD (Figure 1A, B). Individual (II-5) is a 14 year old male who presented to clinic after an episode of chest pain and an abnormal electrocardiogram which was significant for left axis deviation and biventricular hypertrophy. A heart murmur was noted on physical exam and an echocardiogram showed a large secundum ASD with moderate right atrial and right ventricular enlargement. By report, he was born with a membranous ventricular septal defect (VSD), which had spontaneously closed. He underwent percutaneous device closure of the ASD shortly after diagnosis. His family history is significant for a brother (II-7) diagnosed with a large secundum atrial septal defect at 10 years of age who underwent percutaneous device closure and 2 sisters (II-1 and II-2) with secundum ASDs that “spontaneously” closed as infants and a mother (I-2) with secundum ASD surgically repaired at age 5 years.

Using the candidate gene based approach to the WES data generated from 2 affected children (II-5 and II-7), segregating variants in *GATA4* and *EVC2* were found in Family C (Table 1). Although both individuals II-5 and II-7 carried the *EVC2* variant, mutations in *EVC2* have been linked to Ellis-van Creveld syndrome, which is not consistent with the phenotype of affected family members.[19] As mutations in *GATA4* have been shown to cause non-syndromic ASD in a familial and sporadic cases, we chose to focus on the variant in *GATA4*.[20, 21] This heterozygous G to T transversion at nucleotide position 343, which predicted an amino acid change from glycine to tryptophan at residue 115 (*GATA4* Gly115Trp, NM_002052), segregated with the available affected family members (II-5 and II-7), and is not present in control DNA, 1000 Genomes, or the ExAC database (Figure 1C and 1D). The G115W mutation is located near the transactivation domain (TAD) of *GATA4*, a transcription factor required for development of the heart (Figure 1F).[22, 23] The glycine

residue at position 115 is highly conserved in mammals (Figure 1E, Supplemental Table 3). The SIFT, Polyphen2 Complex, Polyphen2 Mendelian, and FATHMM algorithms predict this mutation to be damaging (Supplemental Table 3).

We generated the mutant Gata4 G115W protein expression construct and examined its ability to activate transcription of downstream target genes *in vitro* using alpha-myosin heavy chain (α-MHC) and atrial natriuretic factor (ANF) luciferase reporters which contain Gata-dependent cardiac enhancers.[20] Using both the α-MHC and ANF luciferase reporters, we found that the G115W mutant protein had significantly decreased transcriptional activity as compared to wildtype Gata4 (Figure 1G, H).

TLL1 Ile263Val in familial ASD

Similar to Family C, Family D had apparent autosomal dominant inheritance of familial ASD (Figure 2A, B). Among the four affected family members, (III-3) is a male who presented with a heart murmur at one month of age and an echocardiogram showed a large secundum atrial septal defect (ASD). He was followed until 6 years of age, at which time an echocardiogram continued to demonstrate a moderate-sized secundum ASD with associated right atrial and right ventricular enlargement, and he underwent percutaneous device closure in the cardiac catheterization laboratory. The father of the proband (II-1) was diagnosed as an adult with a large secundum ASD and had also undergone percutaneous device closure. The older sister of the proband (III-1) underwent surgical closure of an ASD early in life. There is also a paternal grandmother (I-2) that underwent surgical closure of an ASD as an adult.

Using our candidate gene based methodology, we identified a rare heterozygous variant in *TLL1* upon WES of individual II-1 (Table 1). *TLL1* encodes the astacin-like, zinc-dependent, metalloprotease Tolloid-like protein 1, a gene in which mutations have previously been identified in sporadic cases of ASD.[24] An A to G transition at nucleotide position 787 that predicted an amino acid change from an isoleucine to a valine at residue 263 (TLL1 Ile263Val, NM_001204760) was identified in an affected member (II-1) and not in an unaffected member, II-6 (Figure 2A–D). This mutation causes an amino acid change that occurs at a highly conserved residue in the metalloprotease active domain, which is required for TLL1's protease activity (Figure 2E and 2F, Supplemental Table 3).[25, 26] This variant was predicted to be damaging by Polyphen2 Complex, Polyphen2 Mendelian, GERP ++, PhyloP, FATHMM, and SiPhy (Supplemental Table 3). This variant has been previously reported in the ExAC database in 1 out of 121,300 alleles and has a 0.000008244 minor allele frequency.

Single nucleotide deletion in *MYH11* at exon 33 splice site in familial PDA

Family F had four members with patent ductus arteriosus (PDA) inherited in an autosomal dominant manner (Figure 3A, 3B). The proband (IV-4) presented for cardiac evaluation of a heart murmur at 7 weeks of age and an echocardiogram that showed a moderate to large PDA. His birth history is significant for being born premature at 33 weeks gestation. The PDA was occluded percutaneously at 15 months of age due to left atrial and left ventricular dilation. The family history is significant for multiple family members that required surgical

ligation of a PDA including his mother (III-3), maternal uncle (III-2), maternal grandmother (II-2), and maternal great aunt (II-4) (Figure 3 and Supplemental Table 1).

WES was performed on three affected family members (II-2, II-4 and IV-4) and two unaffected family members (II-1 and II-3). A heterozygous single nucleotide deletion of an intronic +1 splice donor site at exon 33 (c.4599+1delG) in *MYH11 (*NM_001040113) was found to segregate with available affected family members except affected family member, II-4, who did not carry this mutation. Interestingly, the mother (I-2) of II-4 had a rubella infection during her pregnancy, which is a known environmental risk factor for PDA.[27](Figure 3A–D) This mutation has not been previously identified in any public databases.

As the nucleotide deletion affected a splice donor site before exon 33, we predicted that it would lead to a defect in splicing by deletion of exon 33 in affected patients. To test this, we obtained dermal fibroblasts from affected patient II-2. Dermal fibroblasts from the affected patient and a control human dermal fibroblast cell line were cultured in the presence of *TGFβ-1* to increase expression of smooth muscle genes.[28] We extracted RNA from the control and patient II-2 dermal fibroblast cell lines and performed reverse transcription polymerase chain reaction (RT-PCR) to analyze the MYH11 transcript. We found an in frame deletion of exon 33 in dermal fibroblasts of patient II-2 as compared to control (Figure 3F). Exon 33 encodes 71 amino acids of the coiled-coil domain of MYH11 that spans from amino acid 844-1934 and is important for protein function.[29]

Discussion

Here, we investigated the genetic etiology of 9 familial cases of CHD with Mendelian inheritance using a WES approach and are the first to successfully demonstrate that this approach can identify a likely pathogenic or pathogenic mutation in 33% percent of cases that were analyzed. Employment of our unique CHD gene prioritization strategy in conjunction with WES allowed for rapid identification of potentially pathogenic mutations. In two of the cases, in vitro analysis of identified sequence variants was consistent with the mutations causing congenital cardiac malformations. Additionally, our findings demonstrate the effective utilization of WES to identify causative mutations in familial CHD even when there is limited availability of affected individuals. For each of the three families in which we identified likely causative mutations, there is substantial evidence supporting the etiologic role of these mutations in CHD.

GATA4 encodes a cardiac transcription factor that is required for proper cardiovascular development in multiple species and mutations in *GATA4* have previously been shown to cause ASD in familial and sporadic cases.[20, 22, 23, 30–33] Additional evidence supporting the role of GATA4 in ASD was demonstrated by the generation of murine models harboring the *Gata4* G295S or *Gata4* M310V mutations, both of which recapitulated the human ASD phenotype.[34, 35] The Gly115Trp (G115W) mutation occurs near one of the transactivation domains of GATA4, a region in which prior mutations have been identified in patients with ASD.[36] In addition, *in vitro* transactivation studies demonstrate similar loss of function deficits as those identified with other disease-causing mutations.[18, 20] This mutation

highlights the ability of our WES methodology to identify likely pathogenic mutations in familial CHD cases when there is limited availability of affected family members for analysis.

TLL1 is member of the peptidase M12A family of metalloproteases and plays a role in matrix deposition through its procollagen C-proteinase activity, which cleaves C-propeptides of procollagens I-III and converts them into fibrous components of the extracellular matrix (ECM).[26, 37–40] Mice with targeted deletion of *Tll1* display a range of cardiovascular malformations including atrioventricular septal defects, double outlet right ventricle (DORV), and dysplastic heart valves with associated embryonic lethality.[41] *TLL1* has not been that well studied in human CHD, but heterozygous mutations have been reported in patients with ASD.[24] The Ile263Val (I263V) variant identified in *TLL1* is considered likely pathogenic based on several lines of evidence. It was identified in the only available affected individual and not in one unaffected individual in family D. It is very rare in the general population, identified in only 1 out of 60,650 individuals in the ExAC database, and is predicted to be pathogenic by bioinformatic analysis. Finally, the I263V variant is located in a highly conserved region of the astacin like protease domain, located at amino acids 148–348, which forms the catalytic cleft responsible for accommodating substrates to be cleaved by *TLL1*. Isoleucines located in this position are highly conserved within the zinc-dependent metalloprotease BMP1/TLD-like subfamily, and are also conserved across species.[25, 26] Previous findings have suggested that alteration of this highly conserved amino acid sequence in this region are associated with lower enzymatic efficacy.[24] Additional investigation will be required to determine the mechanisms by which mutations in *TLL1* disrupt normal septation of the heart.

MYH11 belongs to the myosin heavy chain family and is as a major contractile protein in smooth muscle cells.[29] *Myh11* knock-out mice are lethal at postnatal day 3, and exhibit several smooth muscle anomalies including a defect in the closure of the ductus arteriosus[42]. Mutations in *MYH11* have been identified in families with inherited PDA and thoracic aortic aneurysms and dissections (TAAD).[29, 43] We identified a single nucleotide deletion at a +1 splice donor site at exon 33 in Family F and analysis of an affected patient's dermal fibroblasts demonstrated that this led to loss of exon 33 predicting a 71 amino acid in frame deletion in the C-terminal region of *MYH11*. Interestingly this deletion is in the same position as one of the families in the original report by Zhu et al. who reported a single nucleotide change (c.4599+1G→T) that prevented proper splicing of exon 33 in *MYH11*. This loss of exon 33 is predicted to have the same coiled-coil domain defects identified by the COILS *in silico* software in the French kindred.[44] We obtained an echocardiogram on II-2 and there was no aortic dilation but III-2 and III-3 have not been examined. Interestingly, there were two individuals within this family who had a PDA that may have been from a non-genetic cause. Individual I-2 had a rubella infection while pregnant for individual II-4 and rubella infections during pregnancy have a high incidence of PDA in the newborn infant.[27] Also, individual IV-4 was born prematurely at 33 weeks gestation and prematurity is another known risk factor for PDA. Our WES approach was able to demonstrate that the cause of II-4's PDA is likely environmental not genetic while IV-4 harbored the disease-causing mutation (Figure 3A). This family is a prime example of the importance of obtaining an accurate history and proper phenotyping, and highlights the

complexities that can arise when analyzing the segregation of mutant alleles in families where phenocopies may exist. Our functional work confirms the pathogenicity of the MYH11:c.4599+1delG mutation and we predict, in agreement with previously published findings, that this affects the protein's coiled-coil domain, impairing protein functionality and resulting in a PDA phenotype.

Utilization of a CHD gene list functions to allow for a more straightforward approach to WES data analysis, and accordingly, a resultant increase in the confidence of the identified mutations. Without the ability to prioritize variants in this manner, the list of WES variants is often too large to systematically analyze the pathogenicity of the identified variants. Even though we used an extensive list of *in silico* prediction algorithms to predict the variant pathogenicity, we recognize that the algorithms often have shared criteria to determine if sequence variants are deleterious and this ultimately limits their overall utility. By focusing on known CHD genes, we were able to prioritize variants in genes that previously have been shown to lead to CHD when mutated in humans and have known functions in cardiac development in animal models. Ultimately, we plan to expand our WES approach to identify novel genes for CHD, potentially by examining a larger list of genes that have been implicated in cardiovascular development in animal models but have not yet been implicated in human disease.

While our study was relatively small, we demonstrated a success rate of 33% in identifying likely pathogenic and pathogenic mutations. A potential explanation is that we focused on non-left ventricular outflow tract obstruction (non-LVOTO) CHDs. LVOTO malformations include bicuspid aortic valve, aortic valve stenosis, coarctation of the aorta, mitral valve stenosis, and hypoplastic left heart syndrome, and are well known to occur in families and are considered to have a strong genetic component. [45] However, WES in LVOT families has not yet been successful in identification of causative monogenic mutations in these cases and we have similar unpublished results.[46] Also, our families had highly concordant cardiac phenotypes within affected members as opposed to more pleiotropic CHD and this may have led to our greater ability to identify a potentially causative mutation.[47] The reason for the lower success rates in LVOT and pleiotropic familial CHD maybe related to an oligogenic etiology for these types of familial CHD.[48]

In summary for all three potentially pathogenic mutations, there exists substantial data, which encompasses our functional data and published human and mouse studies supporting that these mutations are disease-causing within these families. Based upon our work, we propose that a subset of familial cases of CHD, characterized by non-LVOT, concordant phenotypes and autosomal dominant inheritance may result in higher success rate for WES if this is adapted to a more clinical setting. Given the current success rate of other clinical testing methods, utilization of WES with a CHD gene prioritization strategy offers substantial benefit to cases of familial CHD. In light of increasing research efforts to identify the genetic basis for CHD, a WES approach also allows for adjustments of the CHD gene list based upon new discoveries, which are likely to occur from large government-funded consortiums, including the Pediatric Cardiac Genomics Consortium, as compared to targeted sequencing approaches which have been proposed.[49, 50] In accordance with the ACMG statement on use of WES in clinical settings, CHD cases that appear to have a Mendelian

inheritance pattern would be appropriate candidates for WES using a CHD gene prioritization strategy.[51]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

K.L.M., P.W. and V.G. jointly supervised this work as senior authors. We thank the participants and their families for their involvement in this study.

Funding Sources: This work was supported by funding from the National Institutes of Health (R01 HL109758) to K.L.M., P.W. and V.G., K.B. was supported by National Institutes of Health (NHLBI) training grant (T32 HL098039), and M.B. was supported by a American Heart Association Postdoctoral Fellowship (Great Rivers Affiliate).

References

1. Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol. 2002; 39:1890–1900. [PubMed: 12084585]
2. Marelli AJ, Mackie AS, Ionescu-Ittu R, Rahme E, Pilote L. Congenital heart disease in the general population: changing prevalence and age distribution. Circulation. 2007; 115:163–172. [PubMed: 17210844]
3. Andersen TA, de Troelsen KL, Larsen LA. Of mice and men: molecular genetics of congenital heart disease. Cell Mol Life Sci. 2014; 71:1327–1352. [PubMed: 23934094]
4. Srivastava D. Making or breaking the heart: from lineage determination to morphogenesis. Cell. 2006; 126:1037–1048. [PubMed: 16990131]
5. Miller A, Riehle-Colarusso T, Alverson CJ, Frias JL, Correa A. Congenital heart defects and major structural noncardiac anomalies, Atlanta, Georgia, 1968 to 2005. J Pediatr. 2011; 159:70–78. e2. [PubMed: 21329942]
6. Gelb BD, Chung WK. Complex genetics and the etiology of human congenital heart disease. Cold Spring Harb Perspect Med. 2014; 4:a013953. [PubMed: 24985128]
7. Oyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M. Recurrence of congenital heart defects in families. Circulation. 2009; 120:295–301. [PubMed: 19597048]
8. McBride KL, Garg V. Impact of Mendelian inheritance in cardiovascular disease. Ann N Y Acad Sci. 2010; 1214:122–137. [PubMed: 20958326]
9. Wessels MW, Willems PJ. Genetic factors in non-syndromic congenital heart malformations. Clin Genet. 2010; 78:103–123. [PubMed: 20497191]
10. Richards AA, Garg V. Genetics of congenital heart disease. Curr Cardiol Rev. 2010; 6:91–97. [PubMed: 21532774]
11. Kelly BJ, Fitch JR, Hu Y, Corsmeier DJ, Zhong H, Wetzel AN, et al. An ultra-fast, deterministic, highly scalable and balanced parallelization strategy for the discovery of human genetic variation in clinical and population-scale genomics. Genome Biol. 2015; 16:6. [PubMed: 25600152]
12. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011; 470:198–203. [PubMed: 21307932]
13. El-Metwally S, Hamza T, Zakaria M, Helmy M. Next-generation sequence assembly: four stages of data processing and computational challenges. PLoS Comput Biol. 2013; 9:e1003345. [PubMed: 24348224]
14. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet. 2014; 59:5–15. [PubMed: 24196381]
15. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]

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. [PubMed: 25741868]
17. Chang SW, Mislankar M, Misra C, Huang N, Dajusta DG, Harrison SM, et al. Genetic abnormalities in FOXP1 are associated with congenital heart defects. Hum Mutat. 2013; 34:1226–1230. [PubMed: 23766104]
18. Schluterman MK, Krysiak AE, Kathiriya IS, Abate N, Chandalia M, Srivastava D, et al. Screening and biochemical analysis of GATA4 sequence variations identified in patients with congenital heart disease. Am J Med Genet A. 2007; 143A:817–823. [PubMed: 17352393]
19. Galdzicka M, Patnala S, Hirshman MG, Cai JF, Nitowsky H, Egeland JA, et al. A new gene, EVC2, is mutated in Ellis-van Creveld syndrome. Mol Genet Metab. 2002; 77:291–295. [PubMed: 12468274]
20. Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature. 2003; 424:443–447. [PubMed: 12845333]
21. Rajagopal SK, Ma Q, Obler D, Shen J, Manichaikul A, Tomita-Mitchell A, et al. Spectrum of heart disease associated with murine and human GATA4 mutation. J Mol Cell Cardiol. 2007; 43:677–685. [PubMed: 17643447]
22. Kuo CT, Morrisey EE, Anandappa R, Sigrist K, Lu MM, Parmacek MS, et al. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes Dev. 1997; 11:1048–1060. [PubMed: 9136932]
23. Molkentin JD, Lin Q, Duncan SA, Olson EN. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. Genes Dev. 1997; 11:1061–1072. [PubMed: 9136933]
24. Stanczak P, Witecka J, Szydlo A, Gutmajster E, Lisik M, Augusciak-Duma A, et al. Mutations in mammalian tolloid-like 1 gene detected in adult patients with ASD. Eur J Hum Genet. 2009; 17:344–351. [PubMed: 18830233]
25. Stocker, W., Gomis-Ruth, FX. Astacins: Proteases in Development and Tissue Differentiation. In: Brix, K., Stocker, W., editors. Proteases: Structure and Function. Vienna: Springe; 2013. p. 235-263.
26. Ge G, Zhang Y, Steiglitz BM, Greenspan DS. Mammalian tolloid-like 1 binds procollagen C-proteinase enhancer protein 1 and differs from bone morphogenetic protein 1 in the functional roles of homologous protein domains. J Biol Chem. 2006; 281:10786–10798. [PubMed: 16507574]
27. Gibson S, Lewis KC. Congenital heart disease following maternal rubella during pregnancy. AMA Am J Dis Child. 1952; 83:317–319. [PubMed: 14902080]
28. Mia MM, Boersema M, Bank RA. Interleukin-1beta attenuates myofibroblast formation and extracellular matrix production in dermal and lung fibroblasts exposed to transforming growth factor-beta1. PloS One. 2014; 9:e91559. [PubMed: 24622053]
29. Zhu L, Vranckx R, Khau Van Kien P, Lalande A, Boisset N, Mathieu F, et al. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. Nat Genet. 2006; 38:343–349. [PubMed: 16444274]
30. Holtzinger A, Evans T. Gata4 regulates the formation of multiple organs. Development. 2005; 132:4005–4014. [PubMed: 16079152]
31. Sarkozy A, Conti E, Neri C, D'Agostino R, Digilio MC, Esposito G, et al. Spectrum of atrial septal defects associated with mutations of NKX2.5 and GATA4 transcription factors. J Med Genet. 2005; 42:e16. [PubMed: 15689439]
32. Liu XY, Wang J, Zheng JH, Bai K, Liu ZM, Wang XZ, et al. Involvement of a novel GATA4 mutation in atrial septal defects. Int J Mol Med. 2011; 28:17–23. [PubMed: 21373748]
33. Yang YQ, Wang J, Liu XY, Chen XZ, Zhang W, Wang XZ. Mutation spectrum of GATA4 associated with congenital atrial septal defects. Arch Med Sci. 2013; 9:976–83. [PubMed: 24482639]

34. Misra C, Sachan N, McNally CR, Koenig SN, Nichols HA, Guggilam A, et al. Congenital heart disease-causing Gata4 mutation displays functional deficits in vivo. PLoS Genet. 2012; 8:e1002690. [PubMed: 22589735]
35. Han H, Chen Y, Liu G, Han Z, Zhao Z, Tang Y. GATA4 transgenic mice as an in vivo model of congenital heart disease. Int J Mol Med. 2015; 35:1545–1553. [PubMed: 25873328]
36. Kodo K, Yamagishi H. GATA transcription factors in congenital heart defects: a commentary on a novel GATA6 mutation in patients with tetralogy of Fallot or atrial septal defect. J Hum Genet. 2010; 55:637–638. [PubMed: 20739941]
37. Kessler E, Takahara K, Biniaminov L, Brusel M, Greenspan DS. Bone morphogenetic protein-1: the type I procollagen C-proteinase. Science. 1996; 271:360–362. [PubMed: 8553073]
38. Li SW, Sieron AL, Fertala A, Hojima Y, Arnold WV, Prockop DJ. The C-proteinase that processes procollagens to fibrillar collagens is identical to the protein previously identified as bone morphogenic protein-1. Proc Natl Acad Sci U S A. 1996; 93:5127–5130. [PubMed: 8643539]
39. Panchenko MV, Stetler-Stevenson WG, Trubetskoy OV, Gacheru SN, Kagan HM. Metalloproteinase activity secreted by fibrogenic cells in the processing of prolysyl oxidase. Potential role of procollagen C-proteinase. J Biol Chem. 1996; 271:7113–7119. [PubMed: 8636146]
40. Pappano WN, Steiglitz BM, Scott IC, Keene DR, Greenspan DS. Use of Bmp1/Tll1 doubly homozygous null mice and proteomics to identify and validate in vivo substrates of bone morphogenetic protein 1/tolloid-like metalloproteinases. Mol Cell Biol. 2003; 23:4428–4438. [PubMed: 12808086]
41. Clark TG, Conway SJ, Scott IC, Labosky PA, Winnier G, Bundy J, et al. The mammalian Tolloid-like 1 gene, Tll1, is necessary for normal septation and positioning of the heart. Development. 1999; 126:2631–2642. [PubMed: 10331975]
42. Morano I, Chai GX, Baltas LG, Lamounier-Zepter V, Lutsch G, Kott M, et al. Smooth-muscle contraction without smooth-muscle myosin. Nat Cell Biol. 2000; 2:371–375. [PubMed: 10854329]
43. Harakalova M, van der Smagt J, de Kovel CG, Van't Slot R, Poot M, Nijman IJ, et al. Incomplete segregation of MYH11 variants with thoracic aortic aneurysms and dissections and patent ductus arteriosus. Eur J Hum Genet. 2013; 21:487–493. [PubMed: 22968129]
44. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991; 252:1162–1164. [PubMed: 2031185]
45. McBride KL, Garg V. Heredity of bicuspid aortic valve: is family screening indicated? Heart. 2011; 97:1193–1195. [PubMed: 21459929]
46. Martin LJ, Pilipenko V, Kaufman KM, Cripe L, Kottyan LC, Keddache M, et al. Whole exome sequencing for familial bicuspid aortic valve identifies putative variants. Circ Cardiovasc Genet. 2014; 7:677–683. [PubMed: 25085919]
47. Arrington CB, Bleyl SB, Matsunami N, Bonnell GD, Otterud BE, Nielsen DC, et al. Exome analysis of a family with pleiotropic congenital heart disease. Circ Cardiovasc Genet. 2012; 5:175–182. [PubMed: 22337856]
48. McBride KL, Marengo L, Canfield M, Langlois P, Fixler D, Belmont JW. Epidemiology of noncomplex left ventricular outflow tract obstruction malformations (aortic valve stenosis, coarctation of the aorta, hypoplastic left heart syndrome) in Texas, 1999–2001. Birth Defects Res A Clin Mol Teratol. 2005; 73:555–561. [PubMed: 16007587]
49. 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. [PubMed: 25500235]
50. Gelb B, Brueckner M, Chung W, Goldmuntz E, Kaltman, et al. Pediatric Cardiac Genomics C. The Congenital Heart Disease Genetic Network Study: rationale, design, and early results. Circ Res. 2013; 112:698–706. [PubMed: 23410879]
51. ACMG Board of Directors. Points to Consider in the Clinical Application of Genomic Sequencing. Genet Med. 2012; 14:759–761. [PubMed: 22863877]

Clinical Perspective

Genomic medicine has rapidly advanced over the past decade, and in some fields allows clinicians the ability to identify genetic contributors for a disease that can be used to tailor personal treatment regimens and provide accurate genetic counseling. Genetic causes have been identified for many congenital heart diseases (CHD) associated with additional problems (other birth defects, intellectual disabilities) as part of a syndrome. While recurrence rates for isolated CHDs in families is increased, few genetic causes have been found, limiting the ability to provide personalized medicine for individuals with isolated CHDs. Here, we investigated the utility of whole exome sequencing (WES) to identify the cause of isolated CHD in families containing multiple affected individuals with apparent Mendelian inheritance. We studied nine families with atrial septal defects, patent ductus arteriosus (PDA), or tetralogy of Fallot. Utilizing a CHD candidate gene list and bioinformatics approaches, we identified a likely disease-causing mutation in 3/9 families. Our approach demonstrates a high yield in familial CHD indicating WES should be considered in clinical practice for these families. Identification of a pathogenic variant carries an important psychological impact for the family that comes with the knowledge regarding the etiology of the birth defect, provides an improved recurrence risk for genetic counseling, and can differentiate genetic from environmental causes (as reported in our PDA family). Lastly, our approach is superior to panel based testing, as more genes associated with CHD are identified, clinical providers can update the CHD candidate gene list and re-query the WES data.

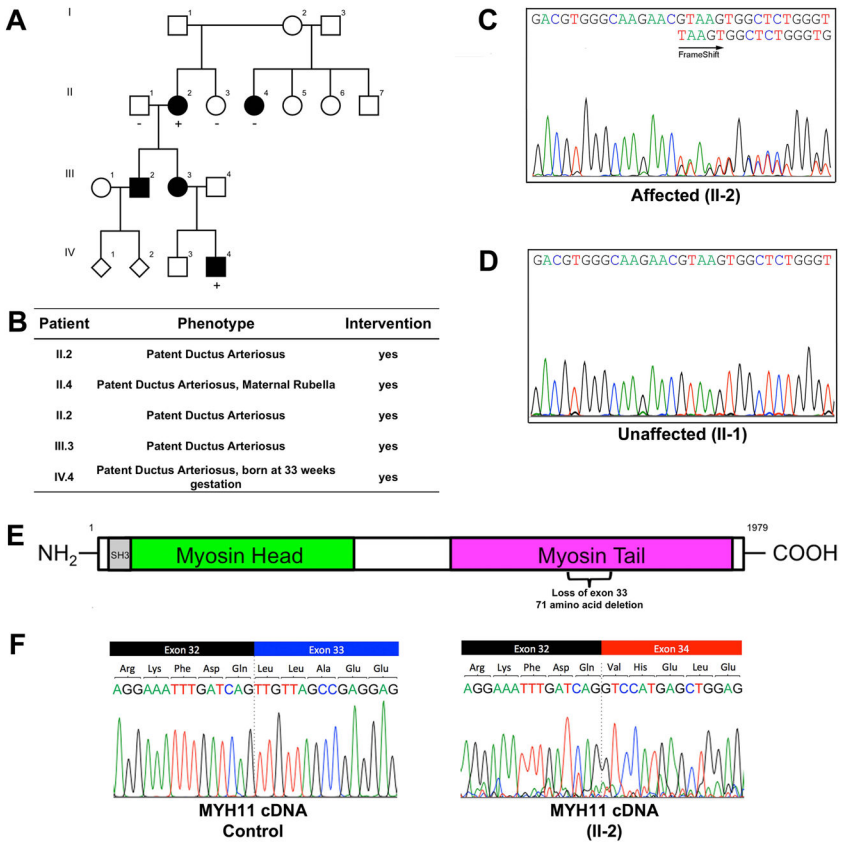

Figure 1.

GATA4 Gly115Trp (G115W) mutation in Family C with atrial septal defects. (A) Pedigree of family C with autosomal dominant inheritance of ASD. (B) Table showing cardiac phenotypes of affected family members. (C) Sequence chromatogram of *GATA4* exon 1 in affected individual II-5 displays a heterozygous nucleotide change 343G>T, causing a glycine to tryptophan change at amino acid residue 115 as compared to an unaffected, unrelated control subject (D). (E) Cross-species alignment of GATA4 protein sequence demonstrating highly conserved glycine at codon 115 (arrow). NCBI accession numbers that were utilized for GATA4 alignment are as follows: Human: NP_001295022.1, Cow: NP_001179806.1, Rat: NP_653331.1, Mouse: NP_032118.2, Chicken: NP_001280035.1, Frog: NP_001084098.1, Zebrafish: NP_571311.2. (F) Gly115Trp is located adjacent to the first GATA4 transactivation domain (TAD1). TAD2, transactivation domain 2; NZf, n-terminal zinc finger; CZf, c-terminal zinc finger; NLS, nuclear localization sequence. (G) Decreased luciferase activity in HeLa cells transfected with Gata4 G115W plasmid when compared to wildtype Gata4. Similar results were noted with both alpha myosin heavy chain (α-MHC) and atrial natriuretic factor (ANF) luciferase reporters. (H) Western blot showing expression of Gata4 wildtype or G115W mutant protein. GAPDH is shown as a loading control. Four independent experiments were performed and statistical comparisons were done utilizing the student's t-test. Two tailed p-value <0.05 was considered statistically significant. * p value <0.05.

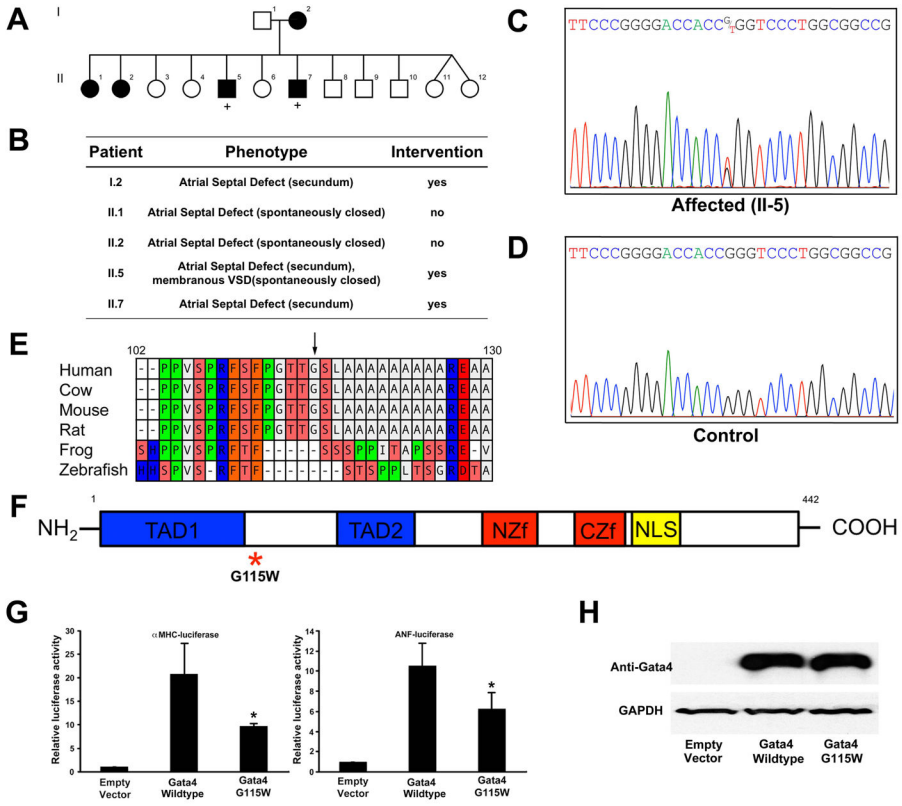

Figure 2.

TLL1 Ile263Val (I263V) mutation in Astacin-Like Domain of TLL1 in family D with atrial septal defects. (A) Pedigree of autosomal dominant inheritance of ASD in family D. (B) Phenotypes of affected family members are shown in Table. (C) Sequence chromatogram of affected patient II-1 displays a heterozygous nucleotide change 787A>G in TLL1, predicting a isoleucine to valine mutation at amino acid position 263, as compared to unaffected family member II-6 (D). (E) Cross-species alignment of protein sequence of TLL1 demonstrating highly conserved isoleucine at position 263 (arrow). NCBI accession numbers that were utilized for GATA4 alignment are as follows: Human:NP_036596.3, Cow:NP_001180043.1, Rat:NP_001099551.1, Mouse:NP_033416.2, Chicken:NP_990034.2, Frog:NP_001083894.1, Zebrafish: NP_571085.1. (F) Ile263Val (*) is located in the astacin-like metalloprotease domain of TLL1. ZnMc: astacin-like metalloprotease domain; CUB: Complement C1r/C1s, Uegf, Bmp1 domain; EGF: epidermal growth factor domain.

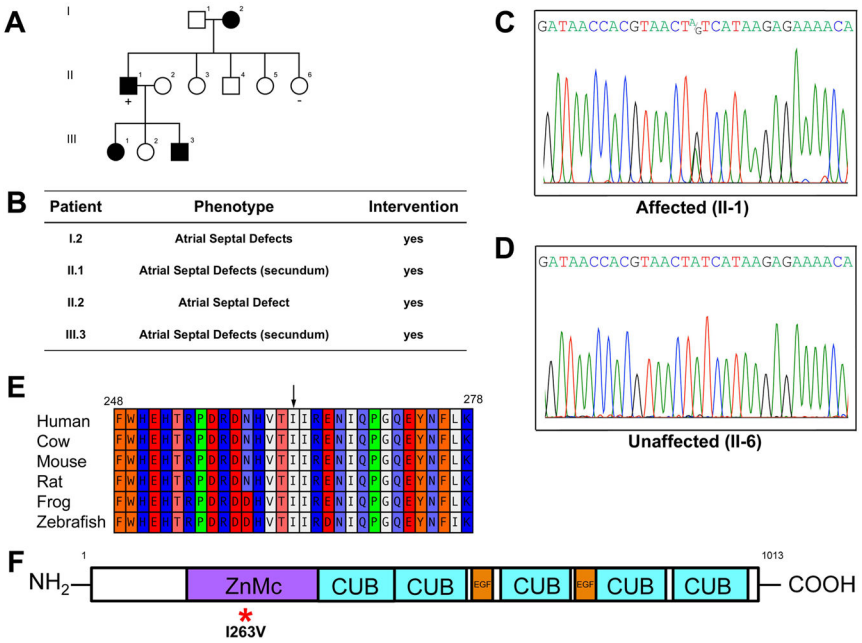

Figure 3.

Single Nucleotide Deletion in *MYH11* (c.4599+1delG) in Family F with patent ductus arteriosus. (A) Pedigree showing autosomal dominant inheritance of PDA in family F. (B) Table showing phenotypes of affected family members. (C) Sequence chromatogram of affected family member II-2 shows a heterozygous deletion of the +1 splice site of exon 33, leading to a frameshift mutation as compared to unaffected individual, II-1 (D). (E) Schematic representation of the MYH11 protein and with deletion of 71 amino acids within myosin tail. (F) Sequence chromatogram of cDNA obtained from dermal fibroblasts of affected individual II-2 that shows loss of exon 33 as compared to control. SH3: SR3 homology domain, Myosin Head: myosin head motor domain, Myosin Tail: myosin coiled-coil rod like tail domain.

Table 1

Potential Pathogenic Mutations Identified by Whole Exome Sequencing

Family	Phenotype	Gene	Mutation	Variant Classification
C	Atrial Septal Defect	*GATA4*	p.G115W	Likely Pathogenic
		EVC2	p.R875W	Uncertain Significance
D	Atrial Septal Defect	*TLL1*	p.I236V	Likely Pathogenic
F	Patent Ductus Arteriosus	*MYH11*	c.4599+1delG	Pathogenic
H	Tetralogy of Fallot	*MYBPC3*	p.V321M	Uncertain Significance
		SOS1	p.K1241E	Uncertain Significance

Variant Classification determined based on the 2015 American College of Medical Genetics and Genomics standards and guidelines for the interpretation of sequence variants.[16]