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

Stephanie LaHaye, BS*; Don Corsmeier, DVM*; Madhumita Basu, PhD; Jessica L. Bowman, MD; Sara Fitzgerald-Butt, MS; Gloria Zender, BS; Kevin Bosse, PhD; Kim L. McBride, MD, MS; Peter White, PhD; Vidu Garg, MD

Background—Congenital heart disease (CHD) is the most common type of birth defect with family- and population-based studies supporting a strong genetic cause for CHD. The goal of this study was to determine whether 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, 2 with patent ductus arteriosus, 2 with tetralogy of Fallot, 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 silico analysis to predict pathogenicity and resulted in the discovery of likely pathogenic mutations in 3 of 9 (33%) families. A GATA4 mutation in the transactivation domain, p.G115W, was identified in familial atrial septal defects and demonstrated decreased transactivation ability in vitro. A p.I263V mutation in TLL1 was identified in an atrial septal defects 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 patent ductus arteriosus 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 allow for a more streamlined approach enabling rapid prioritization and identification of likely pathogenic variants from large WES data sets.

Clinical Trial Registration—URL: https://clinicaltrials.gov; Unique Identifier: NCT0112048. (Circ Cardiovasc Genet. 2016;9:320-329. DOI: 10.1161/CIRCGENETICS.115.001324.)

Key Words: ductus arteriosus, patent ▼ exome ▼ genetic testing ▼ heart septal defects, atrial ▼ tetralogy of Fallot

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 causes of 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 genetic 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.7

Original Article

Received October 6, 2015; accepted June 27, 2016.
From the Center for Cardiovascular Research, The Research Institute (S.L., M.B., S.F.-B., G.Z., K.B., K.L.M., V.G.), The Heart Center (S.L., M.B., J.L.B., S.F.-B., K.L.M., V.G.), and Biomedical Genomics Core and the Center for Microbial Pathogenesis, The Research Institute (D.C., P.W.), Nationwide Children’s Hospital, Columbus, OH; and Department of Molecular Genetics (S.L., V.G.) and Department of Pediatrics (J.L.B., S.F.-B., K.L.M., P.W., V.G.), The Ohio State University, Columbus.

*S. LaHaye and Dr Corsmeier contributed equally to this work and are co-first authors.
The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.115.001324/-/DC1.
Correspondence to Peter White, PhD, or Vidu Garg, MD, Nationwide Children’s Hospital, 700 Children’s Dr, Columbus, Ohio 43205. E-mail Peter.White@nationwidechildrens.org or E-mail Vidu.Garg@nationwidechildrens.org

© 2016 American Heart Association, Inc.

Circ Cardiovasc Genet is available at http://circgenetics.ahajournals.org DOI: 10.1161/CIRCGENETICS.115.001324

320
The majority of CHD occurs as isolated birth defects, termed nonsyndromic, and population- and family-based studies have identified an increased recurrence risk supporting a strong pathogenic genetic component for CHD. Familial cases for numerous forms of nonsyndromic CHD have been reported, and in some cases, where multiple family members were affected, they have led to the identification of CHD-causing genes such as \textit{NKKX2.5} and \textit{GATA4}. In other cases, candidate gene sequencing has been performed on populations of individuals with nonsyndromic CHD and potential disease-causing variants have been identified. Even with these advances, the clinical utility of these findings is not yet clear as genetic testing for nonsyndromic CHD is not routinely performed even when a positive family history exists.

High-throughput (next-generation) sequencing technology has evolved quickly over the past decade, allowing for the rapid sequencing and analysis of human genomes within hours. Because of 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). WES has proven to be an important alternative to single locus-based genetic screening and panel-based sequencing. 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 and inability to expand the candidate gene list without resequencing and previous genotype–phenotype associations.

Here, we demonstrate the utility of WES in familial, nonsyndromic 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 of 9 families. Our study is the first to demonstrate the clinical utility of WES in the genetic diagnosis of familial cases of nonsyndromic 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 with autosomal recessive inheritance, and 1 family (I) had a dysplastic pulmonary valve with autosomal recessive inheritance. Pedigrees (families A–I) with associated phenotype information are shown in Figure I and Table I in the Data Supplement. Informed consent was obtained from study subjects or parents of subjects aged <18 years (assent was obtained from subjects aged 9–17 years) under protocols approved by the Institutional Review Board at Nationwide Children’s Hospital and University of Texas Southwestern. Genomic DNA was isolated from blood samples using the 5’ 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 version 1 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 version 4 probes. For the remaining samples in the study, the SureSelect Human All Exon version 5 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 50× targeted region coverage.

WES Analysis

Primary analysis consisted of using Illumina’s Real-Time Analysis software to perform base calling and quality scoring from the raw intensity files. The resulting base call format files were then converted and demultiplexed using Illumina’s bc12fastq2 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. Churchill utilizes the Burrows–Wheeler Aligner to align sequence data to the reference genome (UCSC build hg19). Duplicate sequence reads were removed using PicardTools (version 1.104). Local realignment was performed on the aligned sequence data using the Genome Analysis Toolkit (version 3.3-0). Churchill’s own deterministic implementation of base quality score recalibration was used. The GATK’s HaplotypeCaller was used to call variants. To maximize sensitivity, variant calling was performed across all samples in the study. The use of the GATK’s variant quality score recalibration was excluded in favor of using Churchill own quality-based variant filtering algorithm.

ANNOVAR (ANNOtate VARIation), a software tool to annotate genetic variation, was used along with custom-in house scripts to provide mutation and gene information, protein functional predictions and population allele frequencies. Commonly used heuristic filtering methods based on these annotations were applied. Common variation occurring at >1% minor allele frequency in the population was excluded. Variants outside of coding regions (defined as >4 base pairs from an exon splice site) and 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 AD model of inheritance. Families G, H, and I were additionally analyzed using homozygous recessive and compound heterozygous models. As shown in Figure I in the Data Supplement, 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 on 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 on our literature review, we identified 69 candidate CHD-causing genes (Table II in the Data Supplement). Variants identified by WES were filtered
by this CHD gene list and then were subsequently analyzed for segregation among available affected family members, when possible. WES heterozygote calls (and homozygote for families G, H, and I) that segregated with the affected family members subsequently underwent 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 use 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 (Tables III and IV in the Data Supplement). For family C, adult human DNA (50-181-348; Biochain Institute, Inc.) was used for control because 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 American College of Medical Genetics and Genomics Standards and Guidelines' recommendations: pathogenic, likely pathogenic, uncertain significance, likely benign, benign (Table V in the Data Supplement).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).17,18 Luciferase activity was measured 48 hours 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 t test. P<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-approved protocol. Patient and control dermal fibroblasts (ATCC human adult primary dermal fibroblasts, PCS-201-012) were cultured, and 10 ng/mL of recombinant TGFβ-1 ligand was added to the dermal fibroblast media for 48 hours, after 15 hours of serum starvation, to increase the expression of MYH11. Total mRNA was isolated from these cells with TRIzol (Thermo Scientific 15596026) followed by total RNA purification (Norgen 17200). mRNA (1 μg) was used to synthesize cDNA with the Superscript VILO cDNA synthesis kit (Thermo Scientific 115470). Reverse transcription polymerase chain reaction was performed with the following primers: F: 5’-CAAGAAGAAACCCGGCAGAAGCTCAACGTG-3’ and R: 5’-AAAGATCTCATCTGGAOGCCAGGGCATC-3’, to generate a 1064 bp fragment of MYH11 (NM_001040113). Polymerase chain reaction 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 (Figures I and II and Table I in the Data Supplement). Six variants in 4 of 9 families (Table) were identified that met the following criteria: determined to be rare (<1% minor allele frequency in population), predicted damaging (in at least 4 of 7 in silico functional algorithms), and segregated with disease in affected individuals (Table; Table II in the Data Supplement). In 3 families (C, D, and 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. Although 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 5 members affected with AD inheritance of ASD (Figure 1A and 1B). Individual (II-5) is a 14-year-old boy who presented to clinic after an episode of chest pain and an abnormal ECG that was significant for left-axis deviation and biventricular hypertrophy. A heart murmur was noted on physical examination, 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, which had spontaneously closed. He underwent percutaneous device closure of the ASD shortly after diagnosis. His family

| Table. Potential Pathogenic Mutations Identified by Whole Exome Sequencing |
|--------------------------|----------|---------|-----------------|
| Family | Phenotype | Gene | Mutation | Variant Classification |
| C      | Atrial septal defect | GATA4 | p.G115W | Likely pathogenic |
| D      | Atrial septal defect | EVC2 | p.R875W | Uncertain significance |
| D      | Patent ductus arteriosus | MYH11 | c.4599+1delG | Pathogenic |
| H      | Tetralogy of Fallot | MYBPC3 | p.V321M | Uncertain significance |
| H      |                | SOS1 | p.K1241E | Uncertain significance |

Variant classification is determined based on the 2015 American College of Medical Genetics and Genomics standards and guidelines for the interpretation of sequence variants.16
LaHaye et al  Whole Exome Sequencing in Congenital Heart Disease

323

history is significant for a brother (II-7) diagnosed with a large secundum ASD at 10 years of age who underwent percutaneous device closure, 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 the age of 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). 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 nonsyndromic 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).

Figure 1. GATA4 Gly115Trp (G115W) mutation in family C with atrial septal defects (ASDs). A, Pedigree of family C with autosomal dominant inheritance of ASD. B, The table shows 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 when compared with an unaffected, unrelated control subject (D). E, Cross-species alignment of GATA4 protein sequence demonstrating highly conserved glycine at codon 115 (arrow). National Center for Biotechnology Information 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, and zebrafish: NP_571311.2. F, Gly115Trp is located adjacent to the first GATA4 transactivation domain (TAD1). G, Decreased luciferase activity in HeLa cells transfected with Gata4 G115W plasmid when compared with wild-type 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 wild-type or G115W mutant protein. GAPDH is shown as a loading control. Four independent experiments were performed, and statistical comparisons were done utilizing the Student t test. CZf indicates c-terminal zinc finger; NLS, nuclear localization sequence; NZf, n-terminal zinc finger; and TAD2, transactivation domain 2. *P<0.05. VSD indicates ventricular septal defect.
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 of \textit{GATA4}, a transcription factor required for development of the heart (Figure 1F).\textsuperscript{22,23} The glycine residue at position 115 is highly conserved in mammals (Figure 1E; Table III in the Data Supplement). The SIFT, Polyphen2 Complex, Polyphen2 Mendelian, and FATHMM algorithms predict this mutation to be damaging (Table III in the Data Supplement).

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 and atrial natriuretic factor luciferase reporters that contain Gata-dependent cardiac enhancers.\textsuperscript{20} Using both the \(\alpha\)-myosin heavy chain and atrial natriuretic factor luciferase reporters, we found that the G115W mutant protein had significantly decreased transcriptional activity when compared with the wild-type Gata4 (Figure 1G and 1H).

**TLL1 Ile263Val in Familial ASD**

Similar to family C, family D had apparent AD inheritance of familial ASD (Figure 2A and 2B). Among the 4 affected family members, individual (III-3) is an infant who presented with a heart murmur at 1 month of age and an echocardiogram showed a large secundum ASD. He was followed up 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) who underwent surgical closure of an ASD as an adult.

Using our candidate gene-based methodology, we identified a rare heterozygous variant in TLL1 on WES of individual II-1 (Table). 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. 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 through 2D). This mutation causes an amino acid change that occurs at a highly conserved residue in the metallocprotease active domain, which is required for TLL1’s protease activity (Figure 2E and 2F; Table III in the Data Supplement). This variant was predicted to be damaging by Polyphen2 Complex, Polyphen2 Mendelian, GERP++, PhyloP, FATHMM, and SiPhy (Table

Figure 3. Single nucleotide deletion in MYH11 (c.4599+1delG) in family F with patent ductus arteriosus. A, Pedigree showing autosomal dominant inheritance of patent ductus arteriosus in family F. B, The table shows 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 when compared with 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 when compared with control. Myosin Head indicates myosin head motor domain; Myosin Tail, myosin coiled-coil rod-like tail domain; and SH3, SR3 homology domain.
III in the Data Supplement). This variant has been previously reported in the ExAC database in 1 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 4 members with PDA inherited in an AD manner (Figure 3A and 3B). The proband (IV-4) presented for cardiac evaluation of a heart murmur at 7 weeks of age, and an echocardiogram showed a moderate to large PDA. His birth history is significant for being born premature at 33 weeks of gestation. The PDA was occluded percutaneously at 15 months of age because of left atrial and left ventricular dilation. The family history is significant for multiple family members who 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; Table I in the Data Supplement).

WES was performed on 3 affected family members (II-2, II-4, and IV-4) and 2 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_0010401113) 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. (Figure 3A through 3D) 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. We extracted RNA from the control and patient II-2 dermal fibroblast cell lines and performed reverse transcription polymerase chain reaction to analyze the MYH11 transcript. We found an in-frame deletion of exon 33 in dermal fibroblasts of patient II-2 when compared with control (Figure 3F). Exon 33 encodes 71 amino acids of the coiled-coil domain of MYH11 that spans from amino acid 844 to 1934 and is important for protein function.

**Discussion**

Here, we investigated the genetic cause 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% of cases that were analyzed. The use of our unique CHD gene prioritization strategy in conjunction with WES allowed for rapid identification of potentially pathogenic mutations. In 2 of the cases, in vitro analysis of identified sequence variants was consistent with the mutations causing congenital cardiac malformations. In addition, 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 3 families in which we identified likely causative mutations, there is substantial evidence supporting the pathogenic 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. 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. This Gly115Trp (G115W) mutation occurs near one of the transactivation domains of GATA4, a region in which previous mutations have been identified in patients with ASD. In addition, in vitro transactivation studies demonstrate similar loss of function deficits as those identified with other disease-causing mutations. 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 a 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. (Figure 3A through 3D) 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. We extracted RNA from the control and patient II-2 dermal fibroblast cell lines and performed reverse transcription polymerase chain reaction to analyze the MYH11 transcript. We found an in-frame deletion of exon 33 in dermal fibroblasts of patient II-2 when compared with control (Figure 3F). Exon 33 encodes 71 amino acids of the coiled-coil domain of MYH11 that spans from amino acid 844 to 1934 and is important for protein function.
family F, and analysis of an affected patient’s dermal fibroblasts demonstrated that this led to the 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 2 individuals within this family, who had a PDA that may have been from a nongenetic 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 of gestation, and prematurity is an 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, whereas 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 whether 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.

Although 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 (LVOT) obstruction CHDs. LVOT obstruction 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.37 The reason for the lower success rates in LVOT and pleiotropic familial CHD maybe related to an oligogenic cause for these types of familial CHD.48

In summary, for all 3 potentially pathogenic mutations, there exist substantial data, which encompasses our functional data and published human and mouse studies supporting that these mutations are disease causing within these families. Based on our work, we propose that a subset of familial cases of CHD, characterized by non-LVOT, concordant phenotypes and AD 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 on new discoveries, which are likely to occur from large government-funded consortia, including the Pediatric Cardiac Genomics Consortium, as compared with targeted sequencing approaches that have been proposed.49,50 In accordance with the American College of Medical Genetics and Genomics statement on the use of WES in clinical settings, CHD cases that seem to have a Mendelian inheritance pattern would be appropriate candidates for WES using a CHD gene prioritization strategy.51

Acknowledgments

Drs McBride, White, and Garg jointly supervised this work as senior authors. We thank the participants and their families for their involvement in this study. The genomic data from this project will be deposited in dbGaP under the project entitled “Genetics Testing of Individuals and Families with Congenital Heart Disease” (accession number pending).

Sources of Funding

This work was supported by funding from the National Institutes of Health (R01 HL109758) to Drs McBride, White, and Garg. Dr Bosse was supported by National Institutes of Health training grant (T32 HL098039), and Dr Basu was supported by a American Heart Association Postdoctoral Fellowship (Great Rivers Affiliate).

Disclosures

None.

References

1. Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol. 2002;39:1980–1900.
2. Marelli AJ, Mackie AS, Ionesco-Itna R, Rahme E, Pilote L. Congenital heart disease in the general population: changing prevalence and age distribution. Circulation. 2007;115:163–172. doi: 10.1161/CIRCULATIONAHA.106.627224.
3. Andersen TS, Troelsen Kde L, Larsen LA. Of mice and men: molecular genetics of congenital heart disease. Cell Mol Life Sci. 2014;71:1327–1352. doi: 10.1007/s00018-013-1430-1.
4. Srivastava D. Making or breaking the heart: from lineage determination to morphogenesis. Cell. 2006;126:1037–1048. doi: 10.1016/j.cell.2006.09.003.
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 and intellectual disabilities) as part of a syndrome. Although 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 to identify the cause of isolated CHD in families containing multiple affected individuals with apparent Mendelian inheritance. We studied 9 families with atrial septal defects, patent ductus arteriosus, or tetralogy of Fallot. Utilizing a CHD candidate gene list and bioinformatics approaches, we identified a likely disease-causing mutation in 3 of 9 families. Our approach demonstrates a high yield in familial CHD indicating whole exome sequencing 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 on the cause of the birth defect, provides an improved recurrence risk for genetic counseling, and can differentiate genetic from environmental causes (as reported in our patent ductus arteriosus family). Finally, 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 query the whole exome sequencing data.