Discovering and Modeling Genetic Causes of Congenital Heart Disease

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Stephanie LaHaye

Graduate Program in Molecular Genetics

The Ohio State University

2017

Dissertation Committee:

Vidu Garg, Advisor
Susan Cole
Helen Chamberlin
Harald Vassein
Joy Lincoln
Abstract

Congenital heart disease (CHD) is the most common birth defect and affects around 2% of live births, when including bicuspid aortic valve (BAV). Although surgical care has significantly improved patient outcome, it remains a major contributor to morbidity and mortality. Population and family based studies have identified a strong genetic component to CHD, however the exact etiology by which CHD occurs is not well understood. The link between genetics and CHD has been strengthened through the study of model organisms and advances in genomic technology and molecular testing, which have led to the advent of key diagnostic tools in cases of syndromic CHD, which occur in the setting of a syndrome. However, the majority of CHD occurs as an isolated phenotype, and clinically there is a lack of genetic diagnostic tools available for this patient population. While whole exome and whole genome sequencing offer potential benefits to the isolated CHD patient population, it has not yet been utilized or tested in a clinical setting. Additionally, pipelines have not yet been developed to systematically analyze the large sequencing datasets that these approaches produce.

As genomic testing continues to improve and become more accessible to clinicians and patients, it is crucial that pipelines are put into place that allow for a
streamlined clinical testing approach. We propose the utilization of whole exome sequencing, with a candidate gene prioritization approach, to allow for the identification of causative mutations in familial CHD. We performed whole exome sequencing on 9 families with apparent Mendelian inherited CHD and prioritized the identified variants utilizing a CHD gene list and further filtered based on segregation, rarity, and predicted pathogenicity. This approach was successful in the identification of potentially pathogenic variants in 3 of the 9 families (33% success rate), and included mutations in GATA4, TLL1, and MYH11. This work supports the use of clinical whole exome sequencing in familial cases of CHD, and offers a pipeline for a streamlined approach to identify high-quality, disease causing mutations.

Identifying disease associated variants is the first step toward understanding the underlying role of genetics in CHD. However, to determine causality and to determine mechanism one must utilize a model system, allowing for the manipulation of the gene of interest and a subsequent disease readout. We utilized this approach, with a mutation previously identified in a family with inherited CHD, and characterized disease development and molecular deficits underlying this phenotype.

We previously reported a family in which a highly penetrant GATA4 mutation segregated with atrial septal defects (ASD) and partially penetrant pulmonary valve stenosis. Mice were utilized that harbored the orthologous G295S disease-causing mutation and recapitulated the human disease
phenotype, exhibiting both ASDs and semilunar valve stenosis. Our goal was to characterize the role of Gata4 in semilunar valve stenosis using Gata4 G295S<sup>ki/wt</sup> mice. As GATA4 is highly expressed in developing semilunar valves, we hypothesized that the Gata4 G295S mutation leads to semilunar valve stenosis due to abnormal valve development. Echocardiographic and histologic examination of adult Gata4 G295S<sup>ki/wt</sup> mice identified functional semilunar valve stenosis, leaflet thickening, and severe disorganization of extracellular matrix (ECM) proteins, including formation of nodules. To determine the onset of disease, 3D-reconstruction was performed on histologic sections of the developing embryonic valves. A reduction in valve leaflet volume was discovered at embryonic day (E)13.5 and morphologic abnormalities were apparent by E15.5, a time-point in which valvular ECM remodeling is occurring. To examine the molecular basis for this phenotype, we performed RNA-seq analysis of E15.5 semilunar valve tissue and identified expression changes in over 1150 genes. Gene Ontology and pathway analysis identified enrichment in the dysregulation of pathways representing ECM organization and WNT signaling; supporting the identification of abnormally remodeled leaflets at this time point. Additionally, several potential direct targets of GATA4 were found to be downregulated in Gata4 G295S<sup>ki/wt</sup> valves. These findings demonstrate a novel role for Gata4 in semilunar valve development and disease through the utilization of a new mouse model for congenital semilunar valve stenosis.
Taken together, these approaches can be used to identify disease associated mutations, determine causality, and lead to a better understanding of the molecular deficits that underlie the disease phenotype. With increased knowledge of disease etiology, therapeutic target discovery and better diagnostics can be developed, which can ultimately lead to improved outcomes in the CHD patient population.
Acknowledgments

I would like to express my deepest appreciation to my advisor Dr. Vidu Garg. Under his mentoring I have developed the ability to think independently and critically, and his training has allowed me to pursue and develop my scientific interests. I would like to also give a special thanks to my thesis committee: Dr. Susan Cole, Dr. Helen Chamberlin, Dr. Harald Vassein, and Dr. Joy Lincoln. Additionally, I would like to thank Dr. Kim McBride for his support and training in human genetics. All six of these individuals have offered countless hours of mentoring, input, guidance, and encouragement. The completion of my thesis work would not be possible without these excellent mentors who have helped guide me through this experience.

I would also like to thank current and past members of the Garg lab, with a special thanks to Sara Koenig, Madhumita Basu, Uddalak Majumdar, Adrianna Matos, James Feller, and Kevin Bosse. Thank you for your daily friendship, advice, and encouragement.

Finally, I would like to thank my family. To my Mom and Dad, thank you for your support as I achieve my goals and continue to pursue my passion in biomedical research. Although the time spent at the bench/computer can be long, I appreciate the encouragement you offer and your understanding when I’m busy working long nights and/or weekends. Last, but certainly not least, I would like to express my
gratitude to my husband, Scott Miller, for being so understanding and supportive over the past 6 years of my PhD training. You have always been there to support me when things have failed, and you have also been there to help me celebrate times of success, for your companionship on this roller coaster ride I am forever grateful.
Vita

2006 ............................................ Tawas Area High School

2011 ............................................. B.S. Genomics and Molecular Genetics,
                                         Michigan State University

2011 ............................................. B.S. Human Biology, Michigan State
                                         University

2011 to present ................................ Graduate Research Fellow, Department
                                         of Molecular Genetics, The Ohio State
                                         University

Publications

   Valve Development and Disease”. In Preparation. 2017.

   for gene-environment interaction in maternal diabetes associated congenital

   B, Garg V. “Notch1 haploinsufficiency in second heart field-derived cells causes

4. LaHaye S, Corsmeier D, Basu M, Bowman JL, Fitzgerald-Butt S, Zender G, Bosse K,
   McBride KL, White P, Garg V. Utilization of Whole Exome Sequencing to Identify
   Causative Mutations in Familial Congenital Heart Disease. Circ Cardiovasc


Fields of Study

Major Field: Molecular Genetics
Table of Contents

Abstract ....................................................................................................................... ii

Acknowledgments ........................................................................................................ vi

Vita ...................................................................................................................................... viii

Publications ................................................................................................................... viii

Fields of Study ............................................................................................................... ix

List of Tables .................................................................................................................. xiv

List of Figures ............................................................................................................... xv

Chapter 1 Introduction ................................................................................................... 1

1.1 OVERVIEW OF CONGENITAL HEART DISEASE ............................................ 1

1.2 DEVELOPMENT OF THE HEART ........................................................................ 3

1.3 CHD ETIOLOGY AND GENETICS ..................................................................... 9

1.4 TECHNOLOGICAL ADVANCES IN GENETIC TESTING ................................. 15

1.5 MODELING OF IDENTIFIED CHD VARIANTS .................................................. 22

1.6 CONCLUDING REMARKS .................................................................................. 30

1.8 FIGURES ............................................................................................................... 32
Chapter 3 Role of Gata4 in Semilunar Valve Development and Disease ..... 70

3.1 RESULTS ............................................................................................................................ 76

3.1.1 Gata4 G295S<sup>ki/wt</sup> mice exhibit progressive functional and structural
valve stenosis phenotypes ........................................................................................................... 76

3.1.2 Developmental deficits underlie aortic valve malformation in Gata4
G295S<sup>ki/wt</sup> mice ...................................................................................................................... 78

3.1.3 Gata4 G295S<sup>ki/wt</sup> mice display differential molecular expression profiles
compared to wildtype littermate controls .................................................................................. 79

3.2 DISCUSSION ....................................................................................................................... 81

3.3 METHODS ........................................................................................................................ 88

3.3.1 Mice .................................................................................................................................. 88

3.3.2 Echocardiography .............................................................................................................. 88

3.3.3 Tissue Fixation and Histology ........................................................................................... 89

3.3.4 Immunostaining .................................................................................................................. 89

3.3.5 AMIRA 3D Reconstruction ................................................................................................. 90

3.3.6 RNA-sequencing ................................................................................................................. 90

3.3.7 Statistics ............................................................................................................................ 91

3.4 FIGURES .......................................................................................................................... 92

Chapter 4 Synopsis .................................................................................................................. 100

4.1 FIGURES ............................................................................................................................ 112

References ............................................................................................................................... 113

Appendix A: Candidate Gene List ......................................................................................... 144
List of Tables

Table 2.1 Potential pathogenic mutations identified by whole exome sequencing. (LaHaye et al., Circ Cardiovasc Genet 2016).......................... 65

Table 2.2 Phenotype information for nine families with familial congenital heart disease. (LaHaye et al., Circ Cardiovasc Genet 2016).................. 66

Table 2.3 In silico analysis of identified sequence variants. (LaHaye et al., Circ Cardiovasc Genet 2016)................................................................. 67

Table 2.4 Primers used for sequencing confirmation of identified variants. (LaHaye et al., Circ Cardiovasc Genet 2016).......................... 68

Table 2.5 Utilization of ACMG standards and guidelines to determine proper sequence variant classification. (LaHaye et al., Circ Cardiovasc Genet 2016)................................................................................. 69

Table A.1 Congenital heart disease candidate gene list........................................... 144
List of Figures

**Figure 1-1** Locations of heart malformations that are usually identified in infancy, and estimated prevalence based on the CONCOR database. (Fahed et al., Circ Res 2013) ................................................................. 32

**Figure 1-2** Mammalian heart development. (Srivastava, Cell 2006) ............... 33

**Figure 1-3** Pathways regulating region specific cardiac morphogenesis. (Srivastava, Cell 2006) .................................................................................................................. 34

**Figure 1-4** Etiology of congenital heart defects. (Nora, Circulation 1968)......... 35

**Figure 1-5** Comparison of traditional sequencing and next-generation sequencing (NGS). (Parikh and Ashley, Circulation 2017) .................................................. 36

**Figure 1-6** Schematic of cardiac lineage differentiation from human PSCs (Doyle et al., Stem Cell Rev and Rep 2015) ................................................................. 37

**Figure 2-1** GATA4 Gly115Trp (G115W) mutation in family C with atrial septal defects (ASDs). (LaHaye et al., Circ Cardiovasc Genet 2016) ..................... 59

**Figure 2-2** TLL1 Ile263Val (I263V) mutation in astacin-like domain of TLL1 in family D with atrial septal defects (ASDs). (LaHaye et al., Circ Cardiovasc Genet 2016) ................................................................. 61

**Figure 2-3** Single nucleotide deletion in MYH11 (c.4599+1delG) in family F with patent ductus arteriosus. (LaHaye et al., Circ Cardiovasc Genet 2016) ...... 62
Figure 2-4 Pedigrees of 9 families with apparent Mendelian inherited CHD. (LaHaye et al., Circ Cardiovasc Genet 2016) ................................................................. 63

Figure 2-5 Whole exome sequencing data analysis work-flow. ......................... 64

Figure 3-1 Gata4 G295S<sup>ki/wt</sup> mice exhibit semilunar valve stenosis. ............ 92

Figure 3-2 Echocardiographic analysis of left ventricular function at one year and two months of age. ............................................................. 94

Figure 3-3 Gata4 G295S<sup>ki/wt</sup> mice exhibit progressive valve disease progression postnatally. ............................................................. 95

Figure 3-4 GATA4 is expressed in the developing outflow tract ...................... 96

Figure 3-5 Gata4 G295S<sup>ki/wt</sup> mice display developmental abnormalities. ....... 97

Figure 3-6 Developing Gata4 G295S<sup>ki/wt</sup> cushions exhibit abnormal molecular expression profiles. ............................................................. 98

Figure 4-1 Timeline of CHD Genetic Discoveries and the Genetic Technologies and Study Designs Used Genetic technologies/study designs are indicated by blue arrows and mark the approximate time when the technology was developed and used. (Blue et al., JACC 2017) ......................................................... 112
Chapter 1 Introduction

1.1 Overview of congenital heart disease

Congenital heart disease (CHD) is the most common birth defect; it affects 12 per 1000 live births globally, equating to one third of all major congenital abnormalities.\(^1\) Additionally, cardiac abnormalities are present in around 30% of all miscarriages.\(^2\) CHD occurs when there is a malformation in a structural component of the heart, a four-chambered organ responsible for pumping oxygenated blood throughout the body.\(^3,4\) Due to the remarkable advances in early diagnosis and surgical care for this patient population, the survival rate has drastically increased. Despite these advances in care, CHD still remains a large contributor to patient morbidity and mortality.\(^5\) Nevertheless, nearly 75% of patients who survive the first year of life go on to survive into adulthood, and current estimates project around 21 million adults, worldwide, are currently living with CHD.\(^4,6-8\) Adults with CHD often require lifelong care, as they frequently exhibit long-term complications, such as arrhythmias, endocarditis, and heart failure.\(^8-10\) Additionally, as this population continues to live into adulthood, the number of adult CHD survivors who are starting families, and potentially passing on genetic contributors to CHD, is also rising. It is therefore essential that we gain insight into the disease etiology and the genetic risks that accompany CHD.
CHDs range in type, severity, and incidence (Figure 1-1). They can manifest themselves in the setting of a syndrome, or can occur in an isolated fashion. Some of the most common CHD phenotypes include septal defects, patent ductus arteriosus (PDA), tetralogy of Fallot (TOF), and semilunar valve stenosis. The most common type of CHD are defects of the septum; which include atrial septal defects (ASD) and ventricular septal defects (VSD). ASD occurs at around 941 per 1 million births, while VSD occurs at around 3570 per one million births.\(^4\) Septal defects occur when there is a hole in the septum that separates either the atria or ventricles, leading to the mixing of oxygenated blood on the left side of the heart and deoxygenated blood on the right side of the heart. PDA occurs in 799 per one million births, and occurs when the embryonic connection between the aorta and pulmonary artery persists after birth, causing deoxygenated blood in the pulmonary artery to mix with oxygenated blood leaving the heart through the aorta.\(^4\) Some phenotypes are more complex, such as cases of TOF, which occurs in 421 per one million births.\(^4\) TOF defects include a constellation of defects. These abnormalities result in oxygen poor blood to be pumped throughout the body. The final of the most common defects is semilunar valve stenosis, which includes aortic valve stenosis (AS) and pulmonary valve stenosis (PS). AS occurs in 401 per one million live births, while PS occurs in 728 per one million live birth.\(^4\) These numbers do not include bicuspid aortic valve (BAV), which occurs when the aortic valve has two leaflets instead of three, predisposing a patient to early onset stenosis and calcification.
BAV is present in around 1%-2% of the population.\textsuperscript{11} BAV phenotypes are variable and often asymptomatic at birth, therefore BAV is often identified later in life. Stenosis of the semilunar valves leads to thickened valve leaflets that are unable to properly open and close, causing the heart to pump harder and eventually leading to heart failure in severe cases. Although these defects range in type and severity, they all arise due to abnormal embryonic development of the cardiovascular structures. It is therefore essential to have a clear understanding of normal cardiovascular structural development to gain proper insight into how these malformations occur.

\textbf{1.2 Development of the heart}

The heart is the first functioning organ to form in vertebrates and it is required to supply the body with oxygen rich blood.\textsuperscript{12,13} The right side of the heart, including the right atrium, ventricle and pulmonary artery, form the pulmonary circuit with the lungs to allow for blood oxygenation. While the left side of the heart, including the left atrium, ventricle, and aorta, is responsible for sending the oxygen rich blood systemically, throughout the body. The walls of the heart are made up of three tissue layers: myocardium, endocardium, and epicardium. The heart relies upon its four valves to ensure unidirectional blood flow (Figure 1-1). The atrioventricular (AV) valves are located between the atrium and ventricle, and include the tricuspid and mitral valves, positioned on the right and left sides of the heart respectively. The AV valves open to allow blood to flow
into the ventricle and close to prevent blood from flowing back into the atria. The semilunar valves are part of the outflow tract and include the pulmonary and aortic valve, located on the right and left sides of the heart respectively. The semilunar valves ensure unidirectional blood flow through the great arteries as the blood pumps out of the ventricles. Finally, the heart muscle is supplied blood through the coronary arteries, which branch off the aorta and supply oxygen rich blood directly to the muscle of the heart.

The development of this highly organized system requires harmonious interaction between several cell lineages and molecular pathways (Figure 1-2, Figure 1-3). The key lineages that make up the heart include the mesodermal cell lineages of the first heart field (FHF) and second heart field (SHF), neural crest cells (CNC), the endocardium, and the pro-epicardium. These lineages contribute to specific components of the heart, and complex spatial and temporal regulation of these developmental processes is required, as changes to this intricate morphogenetic process can lead to developmental malformations of the heart.

The human heart begins to develop at around week two of human gestation, correlating roughly to embryonic day (E)7.5 in mouse, with the formation of the cardiac crescent. The cardiac crescent is made up of differentiating cardiomyocytes and is formed by the coalescence of the FHF and SHF cells along the midline, anterior to the developing neural plate. (Figure 1-2). The FHF is derived from the lateral plate mesoderm and forms the structure
of the early heart tube, subsequently contributing to the left ventricle, portions of the atria, and some of the right ventricle. Additionally, the FHF is marked by transcription factor Hcn4. While the SHF, marked by ISL1, is derived from the medial splanchnic mesoderm adjacent to the pharyngeal endoderm, and contributes to the outflow tract, right ventricle, components of the intraventricular septum, a small portion of the left ventricle, and the atria.

The formation of the cardiac crescent is a tightly regulated process, which requires a complex interplay of positive and negative signaling networks, with many of these signaling events arising from the adjacent endoderm. The initial commitment of cardiac precursors requires signaling from bone morphogenetic protein (BMP), fibroblast growth factor (FGF), and Wingless type (WNT) families of growth factors. The FHF cells require Bmp and Fgf signals to induce expression of crucial cardiac transcription factors Nkx2-5, GATA4 and TBX5; while the SHF requires Wnt signaling to regulate cellular expansion as well as proper expression of ISL1. The SHF also requires FGF, endodermal Sonic hedgehog (SHH), and canonical Wnt signaling to promote proliferation of cardiac progenitor cells; while BMP, NOTCH, and non-canonical WNT play critical roles in the differentiation of these progenitor cells to a cardiac fate. In addition to these genetic factors, epigenetic factors such as histone modifications, DNA methylation, and chromatin remodeling are also involved in cardiac progenitor cell formation and differentiation.
The linear heart tube begins to form by week 3 of human gestation, correlating to E8.0 in mouse, and involves the migration of the FHF and SHF cells to the midline (Figure 1-2). The FHF cells form the scaffolding of the linear heart tube, while the SHF cells position themselves dorsally and migrate into the anterior and posterior ends of the heart tube; these SHF cells will eventually form the right ventricle, conotruncus, and part of the atria. The linear heart tube is made up of an outer myocardium, an inner myocardium, and an acellular extracellular matrix (ECM), which is known as cardiac jelly. The linear heart tube requires regionalized expression of GATA4, NKX2.5, TBX5, and RALDH2; it also utilizes SLIT/ROBO expression under the regulation of T-box factors and NKX2-5 for oriented cell growth. The heart tube begins to contract at approximately day 23 in human development, however anterograde circulation will not ensue until heart looping occurs.

The heart undergoes rightward looping at day 28, E8.5 in mouse (Figure 1-2). At the looping time point, the heart is composed of an outer myocardial lining and an inner endothelial lining. After looping, the myocardium proliferates and differentiates, undergoing segmentation to form the atria and AVC, the ventricle, and the OFT. Due to the segmented regions of gene expression followed by specific patterns of proliferation, the “balloon model” was developed. This model highlights the idea that the different regions of the heart “balloon” out, or undergo directional growth from the heart tube. This model is supported by retrospective cell tracing studies. Several signaling pathways have been
identified as playing key roles at this stage of heart development. Bmp signaling is required to gain left-right polarity, and hyaluronan synthase 2 (HAS2) is required for cardiac progenitors to undergo migration, leading to heart asymmetry.\textsuperscript{47} The heart is the first organ in the developing embryo to exhibit left-right asymmetry, and defects in left-right asymmetry are associated with a range of cardiac defects.\textsuperscript{48} In addition to Has1, Nodal signaling also plays a role in conferring laterality information.\textsuperscript{49} GATA4 works with WNT to control the asymmetric signal propagation, while Wnt/\(\beta\)Catenin negatively regulates the expression of GATA4.\textsuperscript{50} Additionally, Shh plays a role in inducing the left side determinants and activates downstream target GDF1 (Figure 1-3).\textsuperscript{51,52}

At the time of cardiac looping, the cardiac neural crest cells (CNC) invade the outflow tract (Figure 1-2).\textsuperscript{15} CNCs are derived from the dorsal neural tube, which is at the interface of the ectoderm and neural ectoderm.\textsuperscript{53} These cells invade and proliferate within the pharyngeal arches and form the pharyngeal arch arteries, and subsequently go on to become the smooth muscle component of the carotid artery, upper limb artery, aortic arch, and the pulmonary arteries. A subpopulation of CNCs will form the outflow tract, specifically contributing to the septation of the conotruncus and giving rise to the separate aorta and pulmonary artery.\textsuperscript{13,54} CNCs also contribute to the formation of the semilunar valves, giving rise to endocardial cushion interstitial cells and sending key signals for apoptosis and proliferation within the cushions.\textsuperscript{55} WNT, FGF, and BMP signaling play critical roles in the induction of the neural crest.\textsuperscript{56,57} After induction, the CNCs
undergo EMT under the direction of cadherins, Slug/Snail, Rho-kinase, and canonical WNT signaling activity.\textsuperscript{58-62} Early migration of the CNCs occurs under the guidance of FGF, semaphorins, and ephrin signaling; while migration into the outflow tract and arch arteries occurs under the direction of VEGF, SLIT/ROBO, TBX1, Ephrin, BMP, and cadherins.\textsuperscript{63-66} Migration of the CNCs to the proximal outflow tract requires signaling from FGF and MEK/ERK.\textsuperscript{67}

In addition to CNCs, endocardial-derived cells also play crucial roles in the development of the outflow tract.\textsuperscript{55,68} While there is not a clear consensus on the origin of the endocardial cells, it has been proposed that the endothelium may be derived from a common multipotent progenitor within the cardiac mesoderm.\textsuperscript{68-70} The endothelium undergoes endothelial to mesenchymal transition (EMT), giving rise to the EMT-derived interstitial cells of both the semilunar and AV endocardial cushions.\textsuperscript{71} EMT is driven by TGF-β, NOTCH, BMP, and TBX signaling pathways.\textsuperscript{71-75} By week 8 in human, E15.5 in mouse, the structural formation of the heart is complete, however the valves of the heart will continue to develop, remodel, and elongate even after birth.\textsuperscript{76}

A key event that occurs after the looping of the heart is septation of the AVC. AVC septation occurs through the collaborative events of the developing AV cushions, the mesenchymal cap (MC), and the dorsal mesenchymal protrusion (DMP).\textsuperscript{71} The AV cushions and the MC undergo EMT, while the DMP develops from second heart field cells which migrate through the DMP and into the atrial chamber. The fusing of the AV mesenchyme at the atrial canal causes
the divide between the mitral and tricuspid orifices. The formation of the muscular septum occurs when the MC merges with the AV cushions and DMP. After this tissue has been formed it undergoes muscularization to strengthen the septum. Several genes are known to be involved in septation of the heart, including TBX5, GATA4, NKX2.5, SALL4, and the Hand family of transcription factors.

The identification of pathways that play essential roles in cardiovascular development has subsequently led to a better understanding of the underlying role of genetics in CHD. Mutations in many of the transcription factors and signaling molecules described above have also been associated with CHD.

1.3 CHD etiology and genetics

Although the genetic architecture for CHD is not entirely understood, the contribution of genetics to congenital heart disease has been well established. In 1968, James Nora hypothesized that CHD occurred due to the interaction of genetic and environmental factors, and he termed this idea the “multifactorial inheritance model” (Figure 1-4). Nora made this hypothesis based on family recurrence, the increased risk amongst twins, and the homologies to animal CHD models. Although his initial hypothesis was that the etiology for CHD was polygenic, he later revised his hypothesis after the revolutionary work of Ruth Whittemore in 1982. Whittemore published a study that analyzed 373 infants born to 233 mothers with CHD, in which she identified a 16% recurrence risk in offspring and a 60% lesion concordance risk within families. This recurrence
risk is significantly higher than one would expect for polygenic inheritance pattern, and thus invalidated the idea of polygenic inheritance. Further population studies followed, including a CHD study performed in Denmark. Oyen et al published this study in 2009 and found that there was a 3.2% relative risk for any form of CHD to a first degree relative, and that recurrence risks varied based on the type of CHD lesion. When considering the exclusion of chromosomal defects, the population associated risk for anyone with a positive family history of CHD was 4.2%. Additionally, parental consanguinity raises this risk 2-3 fold.

Over the past decade, substantial progress has been made in the study of the molecular genetics of CHD. The first recognized genetic cause of CHD was identified in cases of syndromic CHD. Chromosomal aneuploidy, which occurs when there is an abnormal number of chromosomes, was the first genetic abnormality associated with CHD, and it includes trisomy 21, 18, 13, and Turner syndrome. Around 50% of individuals with trisomy 21 exhibit CHD, with CHD phenotypes ranging from septal defects to atrioventricular canal lesions. In cases of trisomy 18 nearly all affected individuals will display septal defects. Additionally, trisomy 13 has an 80% incidence of CHD, typically in the form of heterotaxy and laterality defects. Around one third of females with Turner syndrome, which occurs when there is only one copy of the X, are born with CHD. Patients with Turner syndrome typically develop defects involving the left ventricular outflow tract. Further study into the chromosomal aberrations and the phenotypes of these patients has uncovered an important role for the appropriate
dosage of specific genes, which when awry can cause cardiovascular malformations.\textsuperscript{4,94,95}

Following the identification of the role of aneuploidy in CHD, copy number variants (CNVs) were also identified as playing a major role in syndromic CHD.\textsuperscript{4,95} A CNV is a large insertion or deletion of a DNA segment, at least 1kb in length, often caused by inappropriate recombination events.\textsuperscript{4} CNVs can alter the dosage of specific genes and can lead to syndromes in a similar way as aneuploidy. For example, one of the most common CHD causing CNVs is 22q11 deletion syndrome, which is caused by a 1.5 or 3-Mb deletion located on chromosome 22q11.2.\textsuperscript{96-98} This CNV arises as a \textit{de novo} change in 93% of cases and affects around 1 per 4,000 births, leading to a syndrome with cardiac malformations, craniofacial abnormalities, thymic or parathyroid hypoplasia, and neurocognitive disabilities.\textsuperscript{99} The most common CHD phenotypes in this syndrome include interrupted aortic arch, TOF, and truncus arteriosus.\textsuperscript{100} Another classic example of CHD caused by CNV is Williams-Beuren syndrome, which is caused by a 1.5-Mb deletion at chromosome 7p11.23.\textsuperscript{101} This syndrome exhibits supravalvular aortic and pulmonary stenosis, elfin facies, developmental delays, and infantile hypercalcemia.\textsuperscript{102}

In addition to chromosome ploidy changes and CNVs, point mutations can also be culprits of syndromic CHD. Point mutations can lead to changes in gene dosage; when this occurs in genes that play central roles in common developmental pathways it can lead to syndromic CHD. One of the first single
gene mutations associated with syndromic CHD was the association of \textit{FBN1} mutations to Marfan Syndrome, which is characterized by aortic root dilation and subsequent increased risk for dissection, skeletal and optic abnormalities.\textsuperscript{103} Another example of this is Alagille syndrome, a syndrome with TOF, pulmonary valve stenosis, skeletal abnormalities, ocular disease, distinctive facies and cholestasis, where over 90\% of the cases are caused by \textit{JAG1} or \textit{NOTCH2} mutations.\textsuperscript{104,105} In both cases, these mutations lead to decreased Notch signaling in Alagille syndrome. Mutations in \textit{TBX5} have been shown to cause Holt-Oram syndrome, a syndrome with septal defects, progressive atrioventricular conduction system disease, and upper limb malformations, in the majority of cases.\textsuperscript{106} Noonan syndrome, which consists of pulmonary valve stenosis, hypertrophic cardiomyopathy, developmental delay, distinctive facies, and bleeding disorders, is caused by mutations in the \textit{RAS} signaling pathway, including \textit{PTPN11, RAF1, SOS1, KRAS, MEK1, MEK2, HRAS, NRAS, SHOC2,} and \textit{CBL}.\textsuperscript{107-110} Mutations in genes within the \textit{RAS} pathway have also been linked to LEOPARD and Costello syndromes, which have similar phenotypes to that of Noonan syndrome.\textsuperscript{111-114}

Nearly 75\% of CHD occurs as an isolated phenotype, not in the setting of a syndrome.\textsuperscript{115} Using large families with inherited CHD, many genes have been found to be implicated in isolated CHD. The overwhelming majority of these identified genes are cardiac transcription factors, including GATA factors, homeobox transcription factors, T-Box transcription factors, as well as additional
transcription factors. Along with transcription factors, mutations have also been identified in genes that play key roles in signal transduction and the formation of structural components of the heart, and they include genes important in Nodal and NOTCH signaling.\(^4\)

GATA4, TBX5, and NKX2-5 are three transcription factors that play critical roles in the development of the cardiovascular system, and are also associated with CHD. These three proteins interact with one another and mutations in each of these genes are associated with overlapping CHD phenotypes. A 2003 study by Garg et al identified a mutation in GATA4 which led to the loss of interaction between GATA4 and TBX5, and resulted in ASD and pulmonary valve stenosis phenotypes.\(^79\) Mutations in TBX5, which are also associated with septal defects, have been associated with the loss of interaction between TBX5 and GATA4.\(^116\) TBX5 was the first gene identified as being associated with CHD, as mutations in TBX5 were found to underlie the majority of Holt-Oram Syndrome cases.\(^106\) Mutations in NKX2-5 are also associated with septal defects, and previous studies have shown that NKX2-5 physically interacts with TBX5 and GATA4.\(^117\) Mutations in NKX2-5 were the first to be associated with isolated cases of familial CHD.\(^118\) Through family based genetic studies, as well as \textit{in vitro} and \textit{in vivo} analyses, these transcription factors have been identified as dosage sensitive, key regulators of cardiovascular development and CHD.

In addition to transcription factors, signaling defects have also been found to cause CHD. The NOTCH signaling pathway is a prime example of how
defective molecular signaling can lead to CHD. In a 2005 study by Garg et al., a mutation in NOTCH1 was found to segregate in a family with outflow tract defects, including bicuspid aortic valve, aortic valve calcification, and other outflow tract defects.119 Furthermore, in a whole exome sequencing screen performed by Preuss et al in 2016, additional variants were identified in NOTCH1 in familial cases of left ventricular outflow tract defects.120 Finally, a recent study published in 2017, found that a patient with hypoplastic left heart syndrome (HLHS) had an inherited mutation associated with BAV on his maternal side and an additional NOTCH1 mutation from his paternal side. Through the utilization of iPSCs and conversion to cardiomyocytes, it was found that NOTCH signaling was significantly dysregulated, as was cardiogenesis due to impaired nitric oxide signaling.121

Several genes have also been identified as having an association to septal defects. Family and population based studies have identified mutations in CRELD1, ALK2, BMP2, and ZIC3 that are associated with atrioventricular septal defects.122-125 Structural genes, MYH6 and ACTC1, have also been associated with ASD.126,127 Additionally, mutations in TLL1, a metalloprotease, and GATA6, a transcription factor, have been found to be associated with ASD.79,128,129

As the ability to identify mutations in patients with CHD continues to improve, it is very likely that genetic analysis will be implemented in the clinical setting. It is predicted that in familial cases of CHD, genetic testing will be able to identify a genetic cause in around 30% of cases.130,131 As technology evolves
and we continue to identify new genes associated with cardiovascular development and disease, this number may increase. Additional studies have identified roles for the non-coding genome and epigenomics, and as we begin to better understand these new horizons in genomic medicine and identify approaches to apply this technology to CHD, we may be able increase genetic testing success rates as well as advancements towards novel therapeutic treatments.

1.4 Technological advances in genetic testing

Identification of the underlying genetic cause for CHD can provide many benefits to patients and families. Genetic technologies can be used as diagnostic tools in cases of a suspected syndrome, and can lead to a tailored medical treatment\textsuperscript{132}. Even in cases of isolated disease, the identification of the genetic cause of disease can lead to important psychological benefits, which include knowledge of the etiology for the birth defect, better understanding of the recurrence risk, and the support of appropriate genetic counseling when it comes to family planning.\textsuperscript{133,134} By identifying genetic contributors to disease, we will gain insight into the differentiation between genetic and environmental risk factors for specific disease phenotypes, and we will also obtain a better grasp on additional risk factors associated with certain genetic abnormalities. Currently, the majority of clinical genetic testing is performed in cases of syndromic CHD, but its utilization in isolated forms is on the horizon.
CHD can manifest itself in the setting of a syndrome, referred to as syndromic CHD, or it can be present as an isolated defect. Syndromic CHD can occur due to chromosomal abnormalities, copy number variations, or single gene mutations. Clinically, diagnosis of syndromic CHD is routinely performed utilizing technologies such as karyotyping, fluorescent in situ hybridization, array comparative genomic hybridization, and Sanger sequencing. Isolated CHD, while more common than syndromic CHD, is not as well understood. Isolated CHD is believed to be caused by either single or multiple gene mutations and clinically does not follow routine genetic testing.

Syndromic CHD makes up around 20% of all CHD cases. Syndromic CHD can occur due to aneuploidy, which is an abnormal number of chromosomes. Testing for aneuploidy can be performed through the utilization of karyotyping with Giemsa banding (G-banding). This cytogenetic approach produces a visible karyotype with AT-rich regions stained darker and producing a band, allowing for the identification of chromosomal abnormalities or ploidy changes. A limit of G-banding is its inability to detect anything smaller than 5-10Mb chromosomal abnormality.

Another common cytogenetic approach is the utilization of fluorescent in situ hybridization (FISH). FISH typically utilizes fluorescently labeled probes which detect specific DNA sequences, and is often used in patients who are predicted to have 22q11 deletion syndrome. While technological advances in FISH allow for down to a 3Mb CNV, it is not scalable to the genome-wide level.
A major technological advancement for cytogenetics occurred with the advent of array comparative genomic hybridization (array CGH) in 1999. Array CGH is a technique that utilizes microarray technology to identify copy number variations through the employment of fluorescently labeled probes which are hybridized to metaphase chromosomes. Array CGH is used to detect submicroscopic CNVs and advances in this technology have allowed identification of CNVs down to the single gene level. Array CGH has better yields than G-banding and FISH and is often used in cases of suspected syndromic CHD.

Although chromosomal abnormalities and CNVs often underlie syndromic CHD, single gene mutations have also been found to cause certain types of syndromic CHD. For example, the majority of Holt Oram syndrome cases are caused by mutations in TBX5. Clinicians who suspect specific syndromes known to be caused by mutations to certain genes often will utilize Sanger sequencing of such genes. While specific mutations can be identified with this method, it is both time and cost intensive. Additionally, clinicians need to know which genes to order testing on ahead of time.

Evolving genetic technologies have aided in the ability to perform diagnostic testing in cases of syndromic CHD. However, the majority of CHD occurs in patients who do not have any other phenotypes. The role of genetics in isolated CHD has been well established through the high heritability within families and through the identification of large families with segregating, disease
causing mutations. However, there remains a paucity in genetic testing options for this patient population. The advent of next generation sequencing offers hope, however there are many questions as to how to apply this technology and how to analyze and interpret the mass amounts of data that are produced.

Before the dawn of next generation sequencing, scientists utilized linkage analysis to identify disease causing mutations in inherited forms of CHD. The utilization of this approach allows for the detection of the chromosomal location of disease causing mutations, and was used for the latter half of the twentieth century to map Mendelian and complex traits. Linkage analysis refers to the observation that alleles are connected to each other on chromosomes, and crossing over must occur for these loci to be separated. Therefore, loci that are near one another remain physically “linked” during meiosis. Recombination fractions can be assigned between two loci, where decreasing scores are associated with a closer location. A score of 0 means that the location never undergoes crossing over; whereas, a score of 0.5 would indicate random allele recombination during meiosis and that these two alleles are not linked. Additionally, a scoring system referred to as logarithm of the odds score (LOD) is used, which is a statistical method used to determine the likelihood that a gene and a disease are linked. Generally, a LOD score of 3 or more, which is about 1,000 to 1 odds that this occurs due to chance, suggests that the gene and disease are linked. The higher the LOD score, the more likely the gene of interest
is linked to the phenotype. Initial genetic linkage was performed with southern blots; however, the dawn of polymerase chain reaction (PCR) led to the identification of polymorphic short tandem repeats (STRs). \(^{140}\) These microsatellite STRs are easy to detect and analyze through simple PCR amplification and can be resolved by gel electrophoresis. Subsequently, these markers are now able to be analyzed in a high throughput manner using differential fluorochrome labeling referred to as a single nucleotide polymorphism (SNP) array. Linkage can also be performed using common SNPs with next generation sequencing data. These different types of analyses allow for the linkage of chromosomal regions to inherited disease, this method was used to identify many of the genes found to be associate with CHD: \(\text{NKX2-5}, \text{TBX5}, \text{GATA4}, \text{PTPN11}, \text{NOTCH1}, \) and others.\(^{79,106,118,119,141}\)

“Next generation” sequencing refers to high throughput sequencing technologies which emerged in 2005, replacing the “first generation” instruments which utilized capillary electrophoresis and Sanger di-deoxy sequencing (Figure 1-5).\(^{142}\) Solexa released the first next generation sequencer, known as the “Genome Analyzer”. This piece of equipment was the first to utilize the short read, massively parallel sequencing technology and could produce around 1 Gb of data per run. Current technologies utilize this same approach and can produce up to 3,000 Gb per run. Costs have dropped dramatically over the years, with illumina’s newest sequencing platform “Novaseq” promising costs as low as $100 a genome upon its release in 2018.
Next generation sequencing utilizes the incorporation of fluorescently labeled deoxyribonucleotide triphosphates (dNTPs) through PCR (Figure 1-5)\textsuperscript{142}. The nucleotide incorporation is read by fluorophore excitation and is preformed across millions of fragments in a massively parallel fashion, denoting the name “massively paralleled sequencing”. There are four key steps to next generation sequencing: library preparation, cluster generation, sequencing, and data analysis\textsuperscript{143}. Library preparation involves the fusion of DNA molecules to adapters, which are then amplified by PCR. Cluster generation occurs when the library is loaded onto the sequencing flow cell, and captured onto surface bound oligonucleotides that are complimentary to the adapter sequences used in the library preperation. Bridge amplification causes the generation of clonal clusters. After the clusters are generated, they are sequenced using a terminator based fluorescence method, TIFF images are generated by the sequencer, illumina converts these files to .bcl files which include the A,T,C, and G nucleotide calls along with quality information, and then these files are demultiplexed and converted to fastq file formats. In the data analysis phase, these fastq files are aligned to a references genome, and different bioinformatics software programs are utilized to either identify SNPs, CNVs, or metagenomics analysis. Additionally, different segregation analysis software can be utilized depending on the inheritance of the CHD\textsuperscript{142}.

In addition to the illumina approach of massive parallel sequencing of short reads, Pacific Biosciences (PacBio) and Oxford Nanopore Technologies
have released real-time long-read sequencing technologies. The PacBio single-molecule real-time (SMRT) approach is the most commonly used long read technology. The SMRT approach allows for reads that are ~3kb in length, which aids in the ability to perform de novo genome assembly, removing the need for a reference genome.\textsuperscript{144} There are clear benefits to this approach, especially when considering the ability to identify CNVs and mutations in tandem repeat regions, where illumina encounters difficulty in mapping repetitive elements. In addition to SMRT, Oxford Nanopore Technologies has released the MinION, which is the size of a cell-phone and utilizes temporal tracing of a template DNA strand, not requiring any amplification. This temporal tracing is referred to as “squiggle space” and it utilizes a motor protein that identifies the changes in current that are associated with specific nucleotides as the DNA sequence passes through the nanopore.\textsuperscript{144} The downfall to both types of long-read sequencing technologies is the increased number of errors per read as well as the increase in cost to gain a higher coverage of the genome; however one can consider incorporating both short and long read technologies into one approach to achieve maximum read coverage, length, and confidence in variants called.

The application of next generation sequencing to CHD analysis is vast, and can include panel based sequencing, whole exome sequencing, and whole genome sequencing.\textsuperscript{131,142} Each of these approaches requires different library generations and different downstream analysis. In 2014 Blue et al. published an article in JACC in which they utilized panel based sequencing on 57 genes.
previously found to be associated with CHD and identified potentially causative mutations in 31% of families with isolated CHD. Additionally, several studies have been performed on isolated CHD that occurs sporadically, and the findings indicate that patients with sporadic CHD have a higher burden of mutations in CHD or cardiac development associated genes. While panel sequencing has supported the utilization of next generation sequencing in clinical settings, the utilization of whole exome sequencing and whole genome sequencing has not been examined. This is likely due to the intense bioinformatic analysis that must be executed. However, as the costs continue to drop for next generation sequencing and as our ability continues to progress on large data analysis, next generation sequencing will become accessible to more patients and clinicians. Whole exome and whole genome sequencing may prove to be important genetic testing tools for patients with isolated CHD in the future, and someday it could be as easily ordered and analyzed as syndromic patients who routinely undergo karyotyping or array-CGH.

1.5 Modeling of identified CHD variants

Human genetic studies utilize statistical association to link a variant to a disease, however modelling must be performed to determine the effect of a mutation on disease phenotype. Additionally, in order to develop novel therapies and diagnostics, it is essential to understand disease mechanism, which can be accomplished through the utilization of a model system. Many model systems
have been employed to study the development and disease of the cardiovascular system, and while each model organism has its own advantages and benefits as well as its own sets of downfalls, these models are necessary to truly understand disease etiology.

The most commonly used model organism to study CHD is the mouse model. The mouse is the choice model due to the similarity of the cardiovascular system structure, development, and physiology between humans and mice. Additionally, mice serve as an excellent model organism due to their well annotated genetic code. Mice share approximately 70% of their exomes with humans, with many of the conserved genes sharing up 99% homology.\textsuperscript{150,151} This conservation can allow for modeling of human disease in mice through orthologous mutations. Recent advances in genetic editing have made this model organism even more accessible, as CRISPR/Cas9 allows for fast and efficient generation of mice with genetic mutations/gene knockouts.\textsuperscript{152} Additionally, mice are kept in inbred lines, which leads to mice that are genetically homologous with one another, allowing for the avoidance of heterogeneity that occurs in human studies.

The most common approach for mouse modeling in CHD is the analysis of candidate genes, either by transgenic, knockout, or knock-in mouse models. Transgenic models employ the injection of DNA containing the sequence of a gene of interest, and allows for this sequence to be inserted randomly into any place into the genome. Transgenic approaches do not affect the endogenous
gene expression. Gene targeting approaches include the utilization of homologous recombination in embryonic stem cells, or in more recent cases the implementation of transcription activator-like effector nucleases (TALEN) or clustered regularly interspaced short palindromic repeats (CRISPR)/cas9. These approaches allow for the insertion of tags, sequence variants, or the deletion of portions of the genome. Additionally, these approaches can be employed to insert inducible cre-lox systems, allowing for spatial and temporal specificity of gene activation/inactivation. In addition to the previously mentioned “reverse genetic” approaches, large scale “forward genetic” screens can also be implanted through mutagenesis techniques such as N-Ethyl-N-Nitrosourea (ENU).

One of the most thorough forward CHD mouse screens was performed by Cecilia Lo’s group in 2015. This study utilized a systems analysis approach in which inbred mice were mutagenized by ENU injections and subsequently bred to recover recessive mutations. This approach led to the identification of 218 CHD mutant mouse lines. 30% of these newly identified lines exhibited heterotaxy, which is the randomization of left right patterning. In addition to laterality defects, lines were identified with double outlet right ventricle (DORV) and septal defects. Whole exome sequencing was performed on these lines and led to the identification of 91 pathogenic mutations in 61 different genes, 27 which had not previously been associated with CHD. Interestingly, 34 of the identified genes were found to be cilia related. This study identified a new role
for cilia related genes, and has subsequently led to many publications identifying important roles for cilia genes in cardiovascular development and CHD.

Mouse modeling has been instrumental in the understanding of the role of genetics in CHD, and has proven to be an informational tool to better understand syndromic as well as isolated CHD. Modeling of syndromic CHD has generated several informative mouse models, including DiGeorge syndrome, Holt-Oram syndrome, Marfan syndrome, Williams syndrome, Noonan syndrome, Alagille syndrome, and Down syndrome.\textsuperscript{155-160} Additionally, mouse modeling has played in integral role in understanding the role of specific genes in isolated cases of CHD. Through the utilization of mouse models, complex signaling pathways that play a role in chamber patterning and formation, and include Nkx2.5, GATA, Tbx, Bmp, and serum response factor (Srf).\textsuperscript{80,156,161,162} Additionally, mouse modeling has allowed for a clearer understanding of outflow tract development and disease through the study of genes such as Tbx1, Gata5, Acvr1, Hand2, Hey2, Prdm1, Tcfap2a, and Foxc1.\textsuperscript{163-170} Septation defects have been modeled through the study of genes including Tbx5, Nkx2.5, Gata4, Hoxb1, Hoxa1, Hoxa3, Shh, Tgf-β2, Bmp4, and Pdgfra.\textsuperscript{80,156,161,171-175}

While the mouse is the favored organism in CHD modeling, this model does have caveats and limitations. It is essential that investigators consider the method of genome editing in this model. Transgenic mice often lead to inappropriately high expression of the inserted gene, which can lead to artefactual phenotypes. Additionally, some genomic editing approaches in which
genes of interest are “knocked out” actually have expressed of protein fragments or hypomorphic alleles. Several early models of gene knockouts were generated in a way that also led to a decreased expression in neighboring loci. Additionally, not all knock out mice are created equal, for example 3 separate knockouts were made for Mrf4, with 3 very different phenotypes described for this knockout model.\textsuperscript{176} The majority of these problems are amenable due to the current genome editing technology, but they remain a point to be considered. Additionally, there are fundamental differences in the cardiac physiologies between mouse and human. Mice also have different cardiac output, calcium flux, hemodynamics, and electrophysiology. Finally, noncoding regions in human are not always conserved in mice, making analysis of promoter and enhancer region mutations difficult. Although there are limitations to this model, it does not negate the exceptional amounts of information and the fundamental role that mice have played in our understanding of cardiovascular development and disease, however it does impress the fact that this data should be treated with caution and consideration.

Another important CHD model is the Drosophila. Although the Drosophila model has been utilized to study human genetics and disease for nearly a century, in the past few decades it has gained traction and emerged as an important model to study CHD. Drosophila offer many advantages, which include sophisticated genetic techniques, a multitude of available transgenic RNAi lines, a short life cycle and reproductive process, and a well-defined cardiovascular
Drosophila offer unique advantages including the commonly used Gal4-upstream activating sequence (UAS) system, which allows for ectopic gene expression. Additionally, even though drosophila have been found to contain a homolog to 80% of human disease associated genes, there genomes are only 5% to that of humans, due to multiple spliced isoforms and different promoter start sites. This genome size makes forward genetic screens much less time and labor intensive, aiding in the ability to identify disease associated genes. Although the drosophila cardiovascular is simple compared to that of humans, its utilization in the study of CHD has yielded many important discoveries. The drosophila adult heart is composed of a single layer of cardiomyocyte like cells and contains ostia cells, which function as valves. The tube-like heart contains rostral and caudal pacemakers, which control the retrograde and anterograde calcium transients leading to the inflow of hemolymph through the ostia and subsequent outflow through the aorta, located in the thorax. Drosophila studies have led to key findings in cardiovascular research, including the identification of tinman, which we refer to as NKX2-5 in humans. Drosophila studies have also allowed for a better understanding of genes such as Dpp (BMP), Hand (HAND1/2), dMef (MEF2), pannier (GATA2), eve (EVX1/2), and pointed (ETS1/2). Additional models that can be used to study CHD include Xenopus laevis and zebrafish (Danio rerio), both of which have a single ventricle and spiral valve to separate the atria from the ventricle. In both species, the cardiovascular
system becomes functional at around 24 hours post fertilization. Both zebrafish and Xenopus are translucent during embryonic stages of development, which allows for non-invasive monitoring and visualizations. Zebrafish remain translucent, even as adults, leading to a continued ability to perform non-invasive analysis throughout their lifespan. Morpholinos are an important genetic tool that can be implemented in both zebrafish and Xenopus, which lead to knocking down of the gene target of interest\textsuperscript{186}. More recently however, TALENS and CRISPR/cas9 technologies have proven to be very useful in these organisms as they lead to less off target effects and more specificity of genetic manipulation.\textsuperscript{187} The utilization of Xenopus have led to advancements in the understanding the role of \textit{Nkx} genes in cardiovascular development as well as how the left/right axis is established under the control of TGF-\textit{β}.\textsuperscript{188,189} Zebrafish studies on the other hand have led to a better understanding of cardiac regeneration and are often used in large scale screening efforts.\textsuperscript{190}

In addition to the utilization of different model organisms, human cells are an excellent tool to model CHD. Recent advances in the ability to convert human cells to a pluripotent state, referred to as induced pluripotent stem cells (iPSCs), have allowed for a better understanding of how cardiac cells are derived as well as the role of mutations in cellular activity, morphology, and molecular expression profiles\textsuperscript{191}. The utilization of iPSCs allows for the collection of patient somatic tissue, which is more readily available than cardiac tissue, and subsequent conversion to a pluripotent state, similar to embryonic stem cells (ESCs). iPSCs
can be reprogrammed to distinctive terminally differentiated cell types, including cardiomyocytes and endothelial cells, allowing for the modeling of different CHD phenotypes (Figure 1-6). An important feature of iPSCs is that they contain the same genetic information as the patient, so mutations present in the patient will be present within the iPSCs as well as the terminally differentiated cells. Through the utilization of this concept and CRISPR/cas9 genome editing technology, one can repair suspected mutations and determine if this genetic changes rescues the CHD phenotype. Additionally, in a similar approach mutations can be added to iPSCs using CRISPR/cas9 to determine if a mutation is sufficient to cause disease. There are many benefits to utilizing iPSCs, including the ability to test therapeutics on patient cells, to insert or repair mutations, and the ability to analyze non-coding regions of the genome, which are not well conserved when considering other organisms. However, there are still limitations, one of the most difficult hurdles is how to model complex disease with cells. Recent work by the Srivastava lab has shown that iPSC-derived cardiomyocytes can be used to better understand the biochemical deficits that exist due to patient mutations. While iPSCs offer a great deal of hope in their ability to model human disease, the limitations of how to appropriately use these cells to model CHD and to truly understand the read out of “normal” and “disease states” must be first overcome.
1.6 Concluding remarks

CHD affects 1% of live births, yet we do not have a clear understanding of the disease etiology, nor do we have therapeutic treatment options or genetic diagnostics available for the majority of this patient population. While there is a clear contribution of genetics to CHD, gaining a complete understanding of the etiology of CHD has remained an elusive task. Analysis of familial CHD has led to the identification of many key genes responsible for CHD phenotypes. However, the identification of these variants has still not led to the advent of clinical genetic testing in non-syndromic familial CHD, nor has it led to a complete understanding of disease etiology. Evolving technologies utilizing massively paralleled sequencing, including whole exome sequencing and whole genome sequencing, offer potential clinical application and the possibility to fill the gap that exists with genetic testing for isolated CHD. However, even with the ability sequence and identify variants in these patients, the task of identifying causal mutations remains an immense undertaking. Even with the utilization of statistical approaches, predictive algorithms, or in vitro functional testing on identified variants, one must remain skeptical, as these methods are unable to fully explain disease mechanism and cannot ascertain disease causality. To truly understand the biochemical deficits that underlie CHD and in order to develop novel therapeutics, it is essential that models be generated in order to understand disease mechanism and confirm mutation causality. Here, we utilize whole exome sequencing to identify causal mutations in cases of familial
congenital heart disease and subsequently demonstrate the utilization of a mouse model harboring the Gata4 G295S mutant allele to model disease and lead to a better understanding of disease mechanism.
1.8 Figures

Figure 1-1 Locations of heart malformations that are usually identified in infancy, and estimated prevalence based on the CONCOR database. (Fahed et al., Circ Res 2013)

Numbers indicate the birth prevalence per million live births. AS indicates aortic stenosis; ASD, atrial septal defect; AVSD, atrioventricular septal defect; CoA, coarctation of the aorta; Ebstein, Ebstein anomaly; HLH, hypoplastic left heart; MA, mitral atresia; PDA, patent ductus arteriosus; PS, pulmonary stenosis; PTA, persistent truncus arteriosus; TA, tricuspid atresia; TGA, transposition of the great arteries; SV, single ventricle; TOF, tetralogy of Fallot; and VSD, ventricular septal defect.
Oblique views of whole embryos and frontal views of cardiac precursors during human cardiac development are shown. (First panel) First heart field (FHF) cells form a crescent shape in the anterior embryo with second heart field (SHF) cells medial and anterior to the FHF. (Second panel) SHF cells lie dorsal to the straight heart tube and begin to migrate (arrows) into the anterior and posterior ends of the tube to form the right ventricle (RV), conotruncus (CT), and part of the atria (A). (Third panel) Following rightward looping of the heart tube, cardiac neural crest (CNC) cells also migrate (arrow) into the outflow tract from the neural folds to septate the outflow tract and pattern the bilaterally symmetric aortic arch arteries (III, IV, and VI). (Fourth panel) Septation of the ventricles, atria, and atrioventricular valves (AVV) results in the four-chambered heart. V, ventricle; LV, left ventricle; LA, left atrium; RA, right atrium; AS, aortic sac; Ao, aorta; PA, pulmonary artery; RSCA, right subclavian artery; LSCA, left subclavian artery; RCA, right carotid artery; LCA, left carotid artery; DA, ductus arteriosus.

**Figure 1-2** Mammalian heart development. (Srivastava, *Cell* 2006)
Figure 1-3  Pathways regulating region specific cardiac morphogenesis. (Srivastava, Cell 2006)

A partial list of transcription factors, signaling proteins, and miRNAs that can be placed in pathways that influence the formation of regions of the heart is shown. Positive influences are indicated by arrowheads, and negative effects by bars. Physical interactions are indicated by dashed lines between factors. In some cases relationships of proteins are unknown. Pathways regulating neural crest cells have been reviewed elsewhere (Stoller and Epstein, 2005). FHF, first heart field; SHF, second heart field. 

Up the normal anatomic relationships in the heart, seen most rightward in this panel. Red shaded areas represent first heart field derivatives, yellow shaded areas represent areas that are most likely derived from second heart field and blue shaded areas represent areas that are derived from proepicardium. Schematic representation of gene regulatory interactions between known signaling pathways (green boxes) and transcription factor interactions in the first heart field (left side) and second heart field (right side). A number of signaling molecules and transcription factors play overlapping roles in the two cardiac progenitor cell populations.
Figure 1-4 Etiology of congenital heart defects. (Nora, Circulation 1968)

Schema for the etiology of congenital heart diseases, emphasizing major importance of multifactorial inheritance and the genetic-environmental interaction, and the lesser role of chromosomal and single gene.
In traditional sequencing, DNA is replicated in the presence of fluorescently labeled bases, yielding differently sized strands with different terminal bases. These terminal bases are identified to assemble the entire sequence and compare it to the reference genome. In NGS, DNA is simultaneously sequenced into billions of overlapping short reads that are then assembled by alignment to the reference genome. Here, each diploid genome is represented by black and orange copies. Green sequence represents reference genome. Purple boxes illustrate heterozygous variants, the black box illustrates the problem of uneven coverage, and the red box indicates the inability of short reads to accurately sequence large areas of repeated sequence.
Figure 1-6 Schematic of cardiac lineage differentiation from human PSCs (Doyle et al., Stem Cell Rev and Rep 2015)

The three primary stages of in vitro CM differentiation from hiPSCs are indicated: induction of cardiac mesoderm, specification of CPCs and differentiation of CMs. Factors involved in directing differentiation of pluripotent stem cells to mesodermal progenitor cells and subsequent cardiovascular lineage cells are indicated. Signaling molecules are in yellow boxes. Transcription factors (within cells) and cell surface markers (below cells) expressed by each cell type are indicated. Genes (structural proteins and cell surface markers) expressed by cardiomyocytes, endothelial cells, smooth muscle cells and fibroblasts are also indicated (below images)
Chapter 2 Utilization of Whole Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease

This chapter was previously published in Circulation: Cardiovascular Genetics and has been reproduced with permission. Copyright is held by Wolterz Kluwer. “Stephanie LaHaye, Don Corsemeier, Madhumita Basu, Jessica L. Bowman, Sara Fitzgerald-Butt, Gloria Zender, Kevin Bosse, Kim L. McBride, Peter White, Vidu Garg. Utilization of whole exome sequencing to identify causative mutations in familial congenital heart disease. Circulation: Cardiovascular Genetics. 2016. CIRCGENETICS.115.001324.”

1.1 Introduction

Cardiovascular malformations are the most common type of birth defect, affecting ~2% of live births when including bicuspid aortic valve.\textsuperscript{193} 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).\textsuperscript{194} 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. 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. 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. Accordingly, chromosomal microarray has been incorporated into clinical practice for the evaluation of children with multiple birth defects or developmental delay/intellectual disability.

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 NKX2.5 and 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\textsuperscript{95,198}. 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\textsuperscript{199,200}. 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)\textsuperscript{201}. WES has proven to be an important alternative to single locus-based genetic screening and panel-based sequencing\textsuperscript{202}. 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.

2.1 Results

2.1.1 Identification of disease causing variants

Individuals from 9 families with Mendelian inherited forms of CHDs were analyzed with WES (Figures 2-4 and 2-5, Table 2.2). Six variants in 4 of 9 families (Table 2.1) 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 2.1, Table 2.3). 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.
2.1.2 GATA4 Gly115Trp mutation in Familial ASD

Family C had 5 members affected with AD inheritance of ASD (Figure 2-1 A,B). 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 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 2.1). 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. 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. 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 2-1 C,D). The G115W mutation is located near the transactivation domain of GATA4, a transcription factor required for development of the heart (Figure 2-1F). The glycine residue at position 115 is highly conserved in mammals (Figure 2-1 F; Table 2.3). The SIFT, Polyphen2 Complex, Polyphen2 Mendelian, and FATHMM algorithms predict this mutation to be damaging (Table 2.3). We generated the mutant Gata4 G115W protein expression construct and examined its ability to activate transcription of downstream target genes in vitro using α-myosin heavy chain and atrial natriuretic factor luciferase reporters that contain Gata-dependent cardiac enhancers. Using both the α-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 (G,H Figure 2-1).

2.1.3 TLL1 Ile263Val in Familial ASD

Similar to family C, family D had apparent AD inheritance of familial ASD (Figure 2-2 A,B). 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 2.1). 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. 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 2-2 A-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 2-2E, Table 2.3). This variant was predicted to be damaging by Polyphen2 Complex, Polyphen2 Mendelian, GERP++, PhyloP, FATHMM, and SiPhy (Table 2.3). This variant has been previously reported in the ExAC database in 1 of 121 300 alleles and has a 0.000008244 minor allele frequency.
2.1.4 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 2-3 A,B). The proband (IV-4) presented for cardiac evaluation of a heart murmur at 7 weeks of age, and an echo-cardiogram 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 2-3, Table 2.1).

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_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.208 (Figure 2-3 A-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.209 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 2-3F). 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.210

Discussion 2.2

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. The 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 (ECM). Mice with targeted deletion of *Tll1* display a range of cardio-vascular malformations including atrioventricular septal defects, double outlet right ventricle, and dysplastic heart valves with associated embryonic lethality. *TLL1* has not been that well studied in human
CHD, but heterozygous mutations have been reported in patients with ASD.\textsuperscript{128} The Ile263Val (I263V) variant identified in \textit{TLL1} is considered likely pathogenic based on several lines of evidence. It was identified in the only available affected individual and not in 1 unaffected individual in family D. It is very rare in the general population, identified in only 1 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 to 348, which forms the catalytic cleft responsible for accommodating substrates to be cleaved by \textit{TLL1}. Isoleucines located in this position are highly conserved within the zinc-dependent metalloprotease BMP1/TLD-like subfamily and are also conserved across species.\textsuperscript{206,207} Previous findings have suggested that alteration of this highly conserved amino acid sequence in this region is associated with lower enzymatic efficacy\textsuperscript{128}. Additional investigation will be required to determine the mechanisms by which mutations in \textit{TLL1} disrupt normal septation of the heart.

\textit{MYH11} belongs to the myosin heavy chain family and is a major contractile protein in smooth muscle cells.\textsuperscript{210} \textit{Myh11} knockout mice are lethal at postnatal day 3 and exhibit several smooth muscle anomalies including a defect in the closure of the ductus arteriosus.\textsuperscript{223} Mutations in \textit{MYH11} have been identified in families with inherited PDA and thoracic aortic aneurysms and dissections.\textsuperscript{210,224} 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 bro-
blasts 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. 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. 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 2-3A). 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 out ow 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. 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.\textsuperscript{227} 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.\textsuperscript{228} 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.\textsuperscript{229}

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 consortiums, including the Pediatric Cardiac Genomics Consortium, as compared with targeted sequencing approaches that have been proposed.\textsuperscript{145,230} 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.231

2.3 Methods

2.3.1 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).

2.3.2 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.

2.3.3 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 les were then converted and demultiplexed using Illumina’s bcl2fastq2 Conversion Software into the standard FASTQ le 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 work flow for variant discovery and genotyping (https://www.broadinstitute.org/gatk/guide/best-practices). 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.

ANNовар (АНОтate 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 2-4, 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.

2.3.4 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 (Appendix A). 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 in silico analysis to predict pathogenicity of the sequence variant.

2.3.5 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 (Table 2.3 and Table 2.4. 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 2.5).233

2.3.6 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).217,234
2.3.7 Luciferase assays

HeLa cells, an immortalized cell line derived from cervical cancer cells, were transfected using Lipofectamine 2000 (Invitrogen 11668027) with 300 ng of either α-myosin heavy chain or atrial natriuretic factor luciferase reporter and 300 ng of wild-type Gata4 or 300 ng of Gata4 G115W, as previously described.217 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.

2.3.8 MYH11 expression etudies 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 \( TGF\beta-1 \) 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 \( \mu \)g) was used to synthesize cDNA with the Superscript VILO cDNA synthesis kit (Thermo Scientific 11754).
Reverse transcription polymerase chain reaction was performed with the following primers: F: 5′-CAAGAAGAAACCGCGCAGAGCTCAACGTG-3′ and R: 5′-AAAGATCTCATCTCTGGAGGCACGGGCATC-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.
2.4 Figures

Figure 2-1 GATA4 Gly115Trp (G115W) mutation in family C with atrial septal defects (ASDs). (LaHaye et al., Circ Cardiovasc Genet 2016)

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.
**Figure 2-2** TLL1 Ile263Val (I263V) mutation in astacin-like domain of TLL1 in family D with atrial septal defects (ASDs). (LaHaye et al., *Circ Cardiovasc Genet* 2016)

**A)** Pedigree of autosomal dominant inheritance of ASD in family D. **B)** The table shows phenotypes of affected family members. **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 when compared with unaffected family member II-6 (D). **E)** Cross-species alignment of protein sequence of TLL1 demonstrating highly conserved isoleucine at position 263 (arrow). National Center for Biotechnology Information accession numbers that were utilized for TLL1 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, and zebrafish: NP_571085.1. **F)** Ile263Val (*) is located in the astacin-like metalloprotease domain of TLL1. CUB indicates complement C1r/C1s, Uegf, Bmp1 domain; EGF, epidermal growth factor domain; and ZnMc, astacin-like metalloprotease domain.
Figure 2-3 Single nucleotide deletion in MYH11 (c.4599+1delG) in family F with patent ductus arteriosus. (LaHaye et al., Circ Cardiovasc Genet 2016)

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.
Figure 2-4 Pedigrees of 9 families with apparent Mendelian inherited CHD. (LaHaye et al., Circ Cardiovasc Genet 2016)

(A-D) Kindreds with autosomal dominant familial ASD. (E-F) kindreds with autosomal dominant familial PDA. (G-H) Kindreds with autosomal recessive TOF. (I) Kindreds with autosomal recessive dysplastic pulmonary valve. Individuals with an asterisk after their number underwent whole exome sequencing.
Figure 2-5 Whole exome sequencing data analysis work-flow. (LaHaye et al., Circ Cardiovasc Genet 2016)

Sample preparation: blue, primary analysis: green, secondary analysis: purple, tertiary analysis: orange, functional work: red.
2.5 Tables

<table>
<thead>
<tr>
<th>Family</th>
<th>Phenotype</th>
<th>Gene</th>
<th>Mutation</th>
<th>Variant Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Atrial Septal Defect</td>
<td>GATA4</td>
<td>p.G115W</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EVC2</td>
<td>p.R875W</td>
<td>Uncertain Significance</td>
</tr>
<tr>
<td>D</td>
<td>Atrial Septal Defect</td>
<td>TLL1</td>
<td>p.I263V</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>F</td>
<td>Patent Ductus Arteriosus</td>
<td>MYH11</td>
<td>c.4599+1delG</td>
<td>Pathogenic</td>
</tr>
<tr>
<td>H</td>
<td>Tetralogy of Fallot</td>
<td>MYBPC3</td>
<td>p.V321M</td>
<td>Uncertain Significance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SOS1</td>
<td>p.K1241E</td>
<td>Uncertain Significance</td>
</tr>
</tbody>
</table>

**Table 2.1** Potential pathogenic mutations identified by whole exome sequencing. (LaHaye et al., *Circ Cardiovasc Genet* 2016)

Variant classification determined based on the 2015 American College of Medical Genetics and Genomics standards and guidelines for the interpretation of sequence variants.233
Table 2.2 Phenotype information for nine families with familial congenital heart disease. (LaHaye et al., Circ Cardiovasc Genet 2016)

U=Unknown
Table 2.3 In silico analysis of identified sequence variants. (LaHaye et al., Circ Cardiovasc Genet 2016)

*variants in splicing regions cannot undergo these predictive algorithms
<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATA4</td>
<td>G115W</td>
<td>5'-TCCACGCATATTATCGTTGTTG-3'</td>
<td>5'-ACTCGAAGGTAGTGAAGCACCCTC-3'</td>
</tr>
<tr>
<td>EVC2</td>
<td>R875W</td>
<td>5'-TTGCGTGGGATTGGGTTG-3'</td>
<td>5'-CACAGACAGGACGGGTTG-3'</td>
</tr>
<tr>
<td>TLI1</td>
<td>I236V</td>
<td>5'-ACCACCTTCTCCACCCATC-3'</td>
<td>5'-CCCATCTAGCCTAAACCA-3'</td>
</tr>
<tr>
<td>MYH11</td>
<td>IVS32+1del</td>
<td>5'-GCCTGGTACAAATGGAGA-3'</td>
<td>5'-CTCTCATTCTGTTCGCTCCC-3'</td>
</tr>
<tr>
<td>MYBPC3</td>
<td>V321M</td>
<td>5'-TCTCTGACCTCTCCCTGCA-3'</td>
<td>5'-TTTCTGCTTTTGCTCCCAGG-3'</td>
</tr>
<tr>
<td>SOS1</td>
<td>K1241E</td>
<td>5'-CCACACTGCTACCCAGCCA-3'</td>
<td>5'-GTTGTTTGGAGGAGGAGGTGG-3'</td>
</tr>
</tbody>
</table>

**Table 2.4** Primers used for sequencing confirmation of identified variants. (LaHaye et al., *Circ Cardiovasc Genet* 2016)
Table 2.5 Utilization of ACMG standards and guidelines to determine proper sequence variant classification. (LaHaye et al., *Circ Cardiovasc Genet* 2016)

*Refer to ACMG Standards and Guidelines for explanation of terminology

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Evidence of Pathogenicity</th>
<th>Evidence of Benign Impact</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Very Strong</td>
<td>Strong</td>
<td>Moderate</td>
</tr>
<tr>
<td>GATA4</td>
<td>G115W</td>
<td>PM1, PM2</td>
<td>PP1</td>
<td>PP2</td>
</tr>
<tr>
<td>EVC2</td>
<td>R875W</td>
<td>PM1</td>
<td>PP3, PP4</td>
<td></td>
</tr>
<tr>
<td>TLL1</td>
<td>I236V</td>
<td>PM1, PM2</td>
<td>PP3</td>
<td>PP4</td>
</tr>
<tr>
<td>MYH11</td>
<td>IVS32+1DEL</td>
<td>PSV1</td>
<td>PP1</td>
<td>PP2</td>
</tr>
<tr>
<td>MYBPC3</td>
<td>V321M</td>
<td>PM1, PM2</td>
<td>PP3</td>
<td>PP4</td>
</tr>
<tr>
<td>SOS1</td>
<td>K1241E</td>
<td>PP3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PSV1**: null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon)  
**PS1**: Same amino acid change as a previously established pathogenic variant regardless of nucleotide change  
**PM1**: Located in a mutational hot spot and/or critical and well-established functional domain  
**PM2**: Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing project, 1000 Genomes Project, or ExAC  
**PM4**: Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants  
**PP1**: Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease  
**PP2**: Missense variant in a gene that has low rate of benign missense variation, in which missense variants are a common mechanism of disease  
**PP3**: Multiple lines of computational evidence support deleterious effect on gene/gene product (conservation, evolutionary, splicing impact, etc)  
**PP4**: Patient’s phenotype or family history is highly specific for a disease with a single genetic etiology
Chapter 3 Role of Gata4 in Semilunar Valve Development and Disease

Congenital heart disease (CHD) is the most common birth defect, affecting ~1% of live births, excluding bicuspid aortic valve (BAV), with semilunar valve malformations accounting for ~10% CHD cases. Semilunar valve malformations, which result in stenosis and regurgitation, are associated with a significant risk of morbidity and mortality. The semilunar valves include the pulmonary and aortic valves, located at the base of the pulmonary artery and aorta respectively. Congenital semilunar valve stenosis is a progressive disease, which without intervention results in ventricular hypertrophy and ultimately heart failure. Limited treatment options exist to treat valve stenosis, and surgical intervention and valve replacement are ultimately required to treat late stages of disease. The etiology of semilunar valve stenosis is not entirely understood, however population based studies support a strong role for genetic contributors. Although it has been established that genetics play a key role, there has been a lack of identified human gene mutations associated with congenital semilunar valve stenosis. In 2003, a role for GATA4 in congenital cardiac septation defects and pulmonary valve stenosis in humans was identified. The segregating mutation was a highly conserved glycine to serine change at codon 296. While all of the patients with the G296S mutation exhibited the septal defect,
a completely penetrant phenotype, only a subset of patients with the mutation displayed pulmonary valve stenosis, a partially penetrant phenotype. The link between GATA4 and septal defects has previously been established, however the link between GATA4 and semilunar valve stenosis is unknown.80

GATA4 belongs to the GATA family of transcription factors, which contain two class IV zinc fingers that preferentially bind the consensus sequence 5’-WGATAR-3’ of downstream target gene promoters. The GATA4 G296S mutation is located near the second zinc finger binding domain, and in vitro analysis of this mutation identified deficits in DNA binding and transactivation of direct targets, in addition to loss of interaction with TBX5.79 GATA4 plays a critical role in cardiovascular morphogenesis, as it is required for ventralization and formation of the heart tube.38 Complete loss of GATA4 results in developmental delay by E7.5 and subsequent embryonic lethality by E9.5.38 GATA4 has also been shown to play an essential role in the development of the atrioventricular (AV) valves, as endothelial loss of GATA4 prevents endothelial to mesenchymal transition (EMT), resulting in hypocellular AV valves and ultimate embryonic lethality between E11.5-12.5.240 However, GATA4 has not been studied in the setting of semilunar valve development and its link to semilunar valve disease is unknown.

To determine the mechanism by which the G296S mutation causes CHD, an in vivo mouse model was generated using homologous recombination with the orthologous G295S mutation knocked into the endogenous Gata4 locus.80 Although the homozygous Gata4 G295Ski/ki mice were embryonic lethal at
embryonic day (E)11.5, the heterozygous \textit{Gata4 G295S}^{ki/wt} mice were viable and born in proper Mendelian ratios. The \textit{Gata4 G295S}^{ki/wt} mice exhibited ASDs, which were attributed to the diminished expression of GATA4 target genes and functional deficits in cardiomyocyte proliferation.\textsuperscript{80} Partially penetrant semilunar valve stenosis was also reported, but not fully characterized. Here, we aim to describe the congenital semilunar valve stenosis phenotype in the \textit{Gata4 G295S}^{ki/wt} mice and to determine the developmental deficits that lead to disease.

Development of the cardiac valves is an intricate process involving the several cell lineage that are tightly controlled by multiple genetic programs.\textsuperscript{239,241,242} The AV valves are the first to develop, preceding the development of the semilunar valves, at around human embryonic day (E)31-E35 and in the mouse around E9.5.\textsuperscript{241} Both the AV and the semilunar valves begin with the formation of endocardial cushions.\textsuperscript{243} The endocardial cushions start as swellings of acellular cardiac jelly, within the wall of the linear heart tube, of the AV and outflow tract (OFT) regions. The AV cushions give rise to the mitral and tricuspid leaflets, while the semilunar cushions give rise to the aortic and pulmonary valves.

The development of the endocardial cushions begins with the crucial event of endothelial to mesenchymal transition (EMT), which is initiated by signals from the adjacent myocardium.\textsuperscript{239} These signals promote a subset of endothelial cells, within the lining of the cushion, to lose cell-cell contacts and transform into mesenchymal cells, which invade the cardiac jelly of the
cushions. These mesenchymal cells contribute to the progenitor cell pool, which give rise to the mature valve apparatus structures. Previous studies have shown that members of the TGF-β family are required for EMT, and secretion of BMP2 from the myocardium acts synergistically with endothelial derived TGF-β to enhance mesenchymal cell formation. \(^{244-246}\) In addition to these growth factors, the NOTCH and RAS signaling pathways have also been implicated in endocardial cushion development of the AVC and OFT. \(^{74,247}\)

Following active stages of EMT, newly transformed mesenchymal cells within the AV region continue to proliferate resulting in expansion of the endocardial cushions and elongation into valve primordia. \(^{241,248}\) The expanding endocardial cushions divide the common atrioventricular canal, first by growth of the superior (anterior) and inferior (posterior) endocardial cushions and then by growth of the 2 lateral cushions. \(^{249}\) Finally, the AV canal is divided into the left and right AV valve orifices when the superior and inferior cushions fuse. \(^{249}\) In addition to endothelial-derived cells, there is evidence that cells from the epicardium populate the cushions and parietal leaflets of the valve primordia and cells from the dorsal mesenchymal protrusion, a derivative of the second heart field, contribute to atrioventricular septation. \(^{250-252}\) The final morphogenetic steps that lead to formation of the mature atrioventricular septum and valve leaflets is less defined, but NFAT signaling has been shown to be important. \(^{253,254}\)

Development of the OFT cushions involves not only EMT, but also requires contribution from a population of migratory neural crest cells, referred to
as the cardiac neural crest (CNC), which arise from the neural tube between the optic placode and third somite.\textsuperscript{249} The CNC cells populate the aortic sac after migrating through the pharyngeal arches. In the aortic sac, the CNC cells are required for the proper septation of the common OFT (truncus arteriosus) into the aorta and pulmonary artery. The CNC cells also contribute to the development of the semilunar valve leaflets and the superior aspect of the ventricular septum.\textsuperscript{239}

As semilunar valve development occurs in a synchronized manner with OFT development, the CNC cells contribute to formation of the truncal swellings. A small eminence on each swelling, along with a third precursor region, will grow to form the valve cusps in the aorta and pulmonary artery.\textsuperscript{249} The OFT can be divided into two portions, the distal (d)OFT and the proximal p(OFT). The mesenchymal cells that contribute to the pOFT are mostly of EMT descent, while the dOFT is comprised mostly of neural crest derived mesenchymal cells.\textsuperscript{71}

Individual valve leaflets are initially evident in human fetuses at around weeks 5 and 6 of development and in mice at around E13.5.\textsuperscript{242} The pulmonary and aortic valves each contain three leaflets. The pulmonary valves are made up of the anterior cusp (AC), the right cusp (RC), and the left cusp (LC).\textsuperscript{71} The aortic valve is composed of the right coronary cusp (RCC), the left coronary cusp (LCC), and the non-coronary cusp (NCC), named due to the location of the coronary arteries which supply the heart with oxygenated blood.\textsuperscript{71} The valve primordia will continue to develop by a process of thinning, reshaping, and elongation, as well as remodeling of the extracellular matrix (ECM). During the
remodeling of the valves, a process referred to as “excavation” occurs, which causes the cushions to undergo thinning and sculpting, producing the semilunar shape of the outflow tract valves and creating the coronary sinuses of the valve. Valve remodeling occurs well into postnatal development, until the valve has become well organized with a stratified trilaminar structure made up of organized ECM layers, which contains elastin, which makes up the atrialis of the AV and the ventricularis of the OFT, fibrillar collagen that form the fibrosa, and proteoglycans, which make up the spongiosa.

Several mouse models have been developed that are deficient in valve development, resulting in semilunar valve disease. Mouse models of deficient valve remodeling been identified, including the Scleraxis−/− mouse model which displays thickened valves by E17.5 and exhibits abnormal ECM deposition as well as aberrant precursor cell lineage differentiation. Additionally, mice with loss of Adamts5 exhibit improper versican cleavage, leading to a loss of excavation, resulting in valve thickening and stenosis. The endocardial deficient Brg1 mouse is an example of how disrupted EMT can result stenotic semilunar valves. In this model EMT is slightly disrupted, leading to a potential compensation by another cell type, and results in a stenotic valve. Models of severely disrupted EMT, including the endothelial deficient Gata4 mouse model, ultimately lead to early lethality due to hypocellular valves. Although many models of valve stenosis have been generated, a model of valve disease has not
yet been generated that harbors an orthologous disease causing mutation that results in congenital valve disease phenotype recapitulation.

Here, we aim to characterize the Gata4 G295S<sup>ki/wt</sup> mouse model, which is a novel model of congenital semilunar valve stenosis. Through the utilization of echocardiography and histology we describe functionally and structurally stenotic adult valves, which exhibit thickened, dysmorphic leaflets that are functionally stenotic by 2 months of age and develop calcific nodules by one year of age. Histologic analysis and 3D reconstruction established a developmental origin of disease, while molecular profiling identified key molecular pathways and processes that were dysregulated during development. Our work is the first to link GATA4 with semilunar valve disease and is also the first to model congenital semilunar valve stenosis utilizing an orthologous human disease causing mutation.

3.1 Results

3.1.1 Gata4 G295S<sup>ki/wt</sup> mice exhibit progressive functional and structural valve stenosis phenotypes

Previously, our group generated the Gata4 G295S<sup>ki/wt</sup> mouse model, which recapitulated the human CHD phenotype, displaying ASD and apparent partially penetrant semilunar valve stenosis in adult mice. The ASD phenotype was analyzed and was attributed to defects in cardiomyocyte proliferation, however the semilunar valve phenotype was not studied. To confirm the
semilunar valve stenosis phenotype in the \textit{Gata4 G295S}^{\text{ki/wt}} mouse model, echocardiography was performed on \textit{Gata4 G295S}^{\text{ki/wt}} and wildtype littermate control mice at one year of age and at two months of age. Echocardiograms showed increased aortic and pulmonic velocities in a subset of the \textit{Gata4 G295S}^{\text{ki/wt}} mice (Figure 3-1 A-D, Figure 3-2). At 2 months of age and 1 year of age, around 60% of the \textit{Gata4 G295S}^{\text{ki/wt}} mice exhibit aortic valve stenosis while around 10% exhibit pulmonary valve stenosis. Analysis of left ventricular ejection fraction did not identify ventricular cardiac dysfunction in either \textit{Gata4 G295S}^{\text{ki/wt}} or wildtype littermate control mice.

Histologic analysis was performed on the aortic valves, as the aortic valve was more commonly diseased. At one year of age, \textit{Gata4 G295S}^{\text{ki/wt}} mice exhibit thickened and abnormally shaped aortic valve leaflets containing hypertrophic-like cells, compared to wildtype littermate controls. (Figure 3-1 E-H). Russel-Movat’s Pentachrome identified a disorganization of collagen fibers and proteoglycans throughout the leaflets. Alizarin red identified calcific nodules in the aortic valves of one year old \textit{Gata4 G295S}^{\text{ki/wt}} (Figure 3-1 M-P). Histologic analysis at two months of age displayed thickened and dysmorphic aortic valve leaflets, compared to wildtype littermate controls (Figure 3-1 Q-T). Russel-Movat’s Pentachrome identified a disorganization of collagens and proteoglycans, as well as the formation of proteoglycan rich nodules in the \textit{Gata4 G295S}^{\text{ki/wt}} mice (Figure 3.4.1 U-X).
To determine if this disease was congenital, as it was in humans, we analyzed the progression of the aortic valve stenosis phenotype in postnatal Gata4 G295S\(^{ki/wt}\) mice. At postnatal day (P)21, the wildtype littermate control aortic valves have approached the completion of their remodeling, elongation, and thinning, however the Gata4 G295S\(^{ki/wt}\) mice exhibit distal tip thickening and increased levels of proteoglycans throughout the distal portion of the leaflet, (Figure 3-3 A-D). At P10 the wildtype littermate control valves are still undergoing postnatal remodeling (Figure 3-3 E-H), however the Gata4 G295S\(^{ki/wt}\) mice exhibit abnormally shaped valves, with clearly increased levels of collagen and proteoglycans (Figure 3-3 E-H). Taken together, this data suggests that the Gata4 G295S\(^{ki/wt}\) mice display congenital aortic valve stenosis, with dysmorphic valves containing disorganized ECM that eventually calcify by one year of age.

3.1.2 Developmental deficits underlie aortic valve malformation in Gata4 G295S\(^{ki/wt}\) mice

The valve stenosis phenotype exhibited by the patients with the Gata4 G296S mutation was congenital, therefore to determine if the semilunar valve stenosis phenotype in our model is due to the disruption of the development of the valves, causing congenital stenosis, we analyzed developmental time points. GATA4 is expressed throughout the development of the semilunar valves (Figure 3-4), and is known to play key roles in EMT in the development of the atrioventricular valves, however Gata4 has not been studied in the setting of
79

to determine if EMT was disrupted in this model, we analyzed H&E stained sections of the OFT at E11.5. Histologic sections identified outflow tract cushions that appeared smaller and less dense in the Gata4 G295S<sup>ki/wt</sup> embryos compared to wildtype littermate controls (Figure A-E), indicative of potential defects in EMT. In addition to this, at E13.5 the total volume of the aortic cushions is significantly smaller in the Gata4 G295S<sup>ki/wt</sup> embryos compared to wildtype littermate controls (Figure 3-5 F-J'). However, the right coronary cusp (RCC) and left coronary cusp (LCC) are not significantly smaller. Interestingly, the non-coronary cusp is significantly smaller in the Gata4 G295S<sup>ki/wt</sup> embryos compared to wildtype littermate controls (Figure 3-5 F-J', P). By E15.5, while the NCC remains significantly smaller in the mutants, the RCC and LCC appear to be of similar volume in the mutants and wildtype littermate controls. However, the RCC and LCC exhibit obvious morphologic abnormalities (K-O',P Figure 3.5). This data suggests that the mutant valves start out small due to a potential reduction in EMT, and although the RCC and LCC leaflets are able to recover in size, they are unable to compensate for the initial cellular depletion, and this may ultimately lead to the disease phenotype.

3.1.3 Gata4 G295S<sup>ki/wt</sup> mice display differential molecular expression profiles compared to wildtype littermate controls

To better understand the molecular profile differences that exist during development between the diseased Gata4 G295S<sup>ki/wt</sup> OFT and wildtype littermate
controls, RNA-seq analysis was performed on total RNA from microdissected outflow tracts from E15.5 embryos. 1158 genes were found to be significantly differentially expressed between the Gata4 G295S<sup>ki/wt</sup> and wildtype control littermates (p <0.05), of these, 445 were differentially expressed by at least a 1.5 fold change (Figure 3-6 A-B). Interestingly, when considering this expression dataset as a whole, it is skewed towards an overall decrease in gene expression in the Gata4 G295S<sup>ki/wt</sup>, as shown by the –log10(p-value) vs. log2 fold change volcano plot in Figure 3-6 B. Gene ontology analysis was performed on the differentially expressed genes, with a p value <0.05, and identified enrichment in GO terms involved in ECM, cell adhesion, and WNT signaling (Figure 3-6 C,D). Additionally, KEGG pathway analysis was performed on this same data set, and identified WNT as the most enriched pathway (Figure 3-6 E,F). WNT signaling is an important pathway for development of the outflow tract and is known to play key roles in EMT and the migration of the cardiac neural crest during development. Additionally, As the Gata4 G295S mutation is known to be a loss of function mutation, we identified several potential direct targets of Gata4 in the downregulated gene list using Genomatix software. Within this dataset, 8 of genes are known to be associated with EMT and are denoted with a red asterisk (Figure 3-6 G). Taken together, this data supports a state of disease in the E15.5 outflow tract and identifies changes in ECM and cell adhesion, as well as a dysregulation of WNT signaling, a key pathway in the development of the outflow tract.
3.2 Discussion

Valve disease, including both congenital and acquired cases, affects nearly 2.5% of the population, yet we do not have a clear understanding of disease etiology nor are there any pharmaceutical treatment options available to prevent or reverse disease progression. Our model is the first to utilize an orthologous human mutation in a mouse model of congenital semilunar valve stenosis. This is the first mouse model to utilize a human disease mutation that causes congenital semilunar valve stenosis. This model allows for a better understanding of semilunar valve development and also offers insight into deficits that can lead to valve disease. Here, we show that the Gata4 G295S mutation causes functional semilunar valve stenosis in a subset of mice, with morphologic abnormalities and disorganized ECM, progressing to the accumulation of nodules positive for calcification by one year of age. Gata4 G295S<sup>wt</sup> mice exhibit ECM defects as early as E11.5, with leaflets remaining small at E13.5, and progressing to morphologically abnormal by E15.5. The molecular profiles of the E15.5 OFTs suggest a dysregulation in the development and remodeling of the ECM as well as defects in WNT signaling, a pathway known to play key roles in valve development, including EMT and CNC migration. Through the utilization of this model we have identified a novel role for Gata4 in semilunar valve development and disease, and have also provided a new model to study congenital semilunar valve stenosis.
The development of the acellular endocardial cushion into the mature valve leaflet is a complex process involving EMT, CNC invasion, and remodeling of the leaflet. Multiple studies utilizing lineage tracing have identified a complicated interaction between EMT and CNC lineages to pattern the semilunar valves. While EMT derived cells are the first to invade the mesenchymal space, the neural crest lineage is also required for proper formation of the valve architecture. Our work has identified potential deficits in EMT in the Gata4 G295S wt compared to control, as shown by histologic sections at E11.5 and 3D reconstruction at E13.5 and E15.5. Gata4 plays an essential role in the process of EMT in the AV valves, as endocardial loss of Gata4 leads to a complete loss of EMT and subsequent hypocellularity of the AV cushions. Due to the valve hypocellularity, these embryos are lethal between E11.5-E12.5. While it has been noted that these embryos exhibit a loss of EMT in the pOFT, the ability to study the affect this has on the development of the OFT is limited, due to early lethality and changes in flow over the OFT as a result of AV cushion defects. Our work is the first to suggest that Gata4 plays an important role in the EMT of semilunar valve development. To confirm these findings, subsequent work should be performed that includes 3D reconstruction of the E11.5 valves, including volumetric measurements and quantification, as well as tracing of the Tie2-lineage with a reporter mouse, such as the Rosa reporter mouse: Gt(ROSA)26Sor^{m4(ACTB-tdTomato,-EGFP)Luo}J. Lineage tracing of the Tie2 derived cells will allow for proper identification and quantification of valvular cells that are
derived from the EMT lineage and will lead to a better understanding of potential EMT defects that arise due to the Gata4 G295S mutation.

Molecular profiling of the E15.5 OFTs identified several pathways that were dysregulated, at the RNA level, in the Gata4 G295S<sup>wt</sup> OFTs. The most significantly dysregulated pathway, identified by KEGG pathway analysis, was the Wnt signaling pathway (Figure 3-6 E-F). The WNT signaling pathway has multiple essential roles during cardiac development, including the regulation of EMT and proliferation, as well as CNC migration, and later roles in valve maturation and homeostasis<sup>71,73,262,263</sup>. Our data suggests that the canonical pathway, the WNT/Ca<sup>2+</sup> pathway, and the planar cell polarity pathway (PCP) are all dysregulated in the diseased Gata4 G295S<sup>wt</sup> OFT (Figure 3-6 F). In the OFT, canonical WNT signaling has been shown to play important roles in the in second heart field (SHF) and CNC. Canonical WNT signaling, which signals through β-catenin, is required for EMT in the proximal outflow tract (pOFT)<sup>263</sup>. Canonical WNT signaling has also been shown to activate TCF-dependent transcription, required for CNCs to undergo G1/S transition and subsequent migration to the OFT<sup>63</sup>. Future work analyzing β-catenin protein expression in the mutant valves should be considered to determine the dysregulation of this pathway at the protein level. The WNT/Ca<sup>2+</sup> pathway has also been closely associated with the development of the valve, as NFATC1 is required for the remodeling of the OFT cushions.<sup>264,265</sup> Interestingly, we identified a decrease in Nfatc1 in the Gata4 G295S<sup>wt</sup> at the RNA level. Additionally, we have identified a predicted Gata4
binding site in the promoter of *Nfatc1*, which suggests that it may be a direct target of GATA4 that may be downregulated due to loss of function of GATA4; further work is needed to confirm this finding, including protein expression analysis of NFATC1. The WNT-PCP pathway also plays an important role in valve development and has been shown to be involved in contact inhibition, as CNCs require appropriate localized expression of *disheveled* to the point of cell-cell contact. This localized expression allows the cell to gain directionality and is followed by appropriate cellular locomotion. CNCs require appropriate directionality to ensure proper migration into the cushions. The PCP pathway works in harmony with ephrin signaling, which is expressed by CNCs and was found to be disrupted in our model. To determine if CNC migration is disrupted in the *Gata4 G295S<sup>kiw</sup>* mouse model, lineage tracing of the Wnt1 lineage in the *Gata4 G295S<sup>kiw</sup>* embryo is required. By combining this data with the previously proposed Tie2 lineage tracing, one will be able to gain a clearer understanding of any cell lineage deficits that exist within the *Gata4 G295S<sup>kiw</sup>* mouse model.

The original *GATA4* mutation was identified in a large family with inherited ASD and partially penetrant pulmonary valve stenosis. *GATA4* mutations have been associated with numerous cardiac defects, including ASD, ventricular septal defect (VSD), atrioventricular septal defect (AVSD), tetralogy of fallot (TOF), persistant truncus arteriosus (PTA), patent ductus arteriosus (PDA), partial anomalous pulmonary venous return (PAPVR), as well as pulmonary stenosis in the setting of ASD. Mutations in *GATA4* that are associated with
pulmonary stenosis in the setting of an ASD have been reported in 3 separate families, 2 families with G296S mutations and in a family with a K319E mutation.\textsuperscript{79,268} Additionally, in 2013 Wang et al identified a GATA4 A74D variant in a sporadic patient with isolated pulmonary valve stenosis. However, information on predicted damaging score of the A74D variant was not given and functional studies were not performed; upon further investigation this variant is not present in the gnomAD database.\textsuperscript{269} Due to the lack of available information associating GATA4 mutations to semilunar valve defects, we sequenced 52 patients with semilunar valve stenosis for nonsynomous variants in GATA4. We did not identify any variants within the coding sequence of GATA4 in this patient population (data not shown), however this could be due to small sample size which could be rectified through analysis of a larger cohort of patients. Additionally, it may be worthwhile to consider testing genes that were found to be dysregulated in the Gata4 G295S\textsuperscript{kiwt} developing embryos, as dysregulation of these genes ultimately leads to valve stenosis in the mouse model. Additionally, these genes can be stratified according to EMT association, CNC migration association, or even WNT signaling association. Further characterization of the underlying deficiencies within the mouse model will lead to a clearer understanding of potential lineage or pathway deficits, which will guide the focus of future stratification approaches. Additionally, querying of large CHD cohorts, such as what has been collected and sequenced by the Pediatric Cardiac Genomics Consortium (PCGC), would be an ideal approach to testing for GATA4
variants and other potential variants in candidate genes identified in the RNA-seq analysis, given the vast number of patient samples as well as phenotype information available.²³⁰

Although our model recapitulates the semilunar valve stenosis phenotype, some questions remain when comparing mouse phenotype to human phenotype. One of the key differences is in the affected valves; in the patients the pulmonary valve is always affected, whereas in the mouse it is usually the aortic valve. Although there is no specific or correct answer as to why this occurs, it is most likely due to the differences in physiology and anatomy that exist between humans and mice. The blood flow to organs, such as the brain, is different in humans than it is in mice, which could lead to changes in the hemodynamics and pressure that the semilunar valves experience. Additionally, heart rate and cardiac output differ between species. Although the mouse may not exhibit the exact same phenotype as the patients, it is still a valuable model of congenital heart disease, as it is the first model of a human mutation causing congenital valve stenosis in a mouse. Additionally, as the semilunar valves are made from the same cells and undergo very similar molecular signaling patterns, these studies are valuable to the central understanding of the role of GATA4 in semilunar valve development and disease.

Our work characterizing the congenital semilunar valve disease phenotype in the Gata4 G295S<sup>kiwt</sup> model supports the idea of incorporating human valve disease causing mutations into the mouse model. This approach permits in vivo
validation of mutation pathogenicity, while supporting the molecular biology approaches of identifying disease mechanism, which can lead to therapeutic target discovery. Most current mouse models of valve disease utilize complete gene knockouts and this is not an appropriate approach when considering biological relevance, as most human disease causing mutations are hypomorphic in nature and do not cause a complete loss of function. Due to the advent of genomic editing techniques, such as CRISPR/cas9, future mouse models of valve disease should consider incorporation of orthologous human mutations. Additionally, by utilizing comprehensive and fastidious techniques, such as 3D reconstruction, one can identify the subtle defects that can occur during development, in the setting of a hypomorphic allele. As we have shown, the *Gata4* G295S allele has partial functionality and the developmental defects are subtle. However, this slightly imperfect valve is unable to maintain structure and function when under the hemodynamic stress and pressure of blood flow, and what starts as a slightly disorganized ECM develops into an extremely dysmorphic and stenotic adult valve that can become calcified.

Nearly 2.5% of the population suffers from valve disease, and there are currently no pharmacologic treatment options to stop or reverse valve disease, with late stage treatment requiring surgical intervention. This work offers a potential translational impact through the identification of genes dysregulated during abnormal valve development. These genes can be utilized as potential
biomarkers to identify early stages of disease, additionally these targets may be used as future therapeutic treatment options.

3.3 Methods

3.3.1 Mice

Animal use was approved and monitored on protocol AR09-00040 by the Institutional Animal Care and Use Committee at the Research Institute at Nationwide Children's Hospital. *Gata4* $Gata4^{kis/wt}$ mice were kept on a $C57BL/6J$ background and genotyped using an allelic discrimination assay, (Applied Biosystems, CA), which is able to detect the G295S point mutation utilizing fluorescently labeled VIC and FAM labeled probes: GATA-G295SFAM: 6FAMTGTAATGCCTGCAGCCMGBNFQ and GATA-G295SVIC: VICATGCCTGCGCTMGBNFQ, and primer sequences FWD 5’ CATCCACCTCACCCCATGGA 3’ and REV 5’ CACGCTGTGGCGTCGTAAT 3’ as previously described.80

3.3.2 Echocardiography

Transthoracic echocardiography was performed with the Visualsonics VEVO 2100 Ultrasound System, as previously described.80 The mice were sedated with 3% isoflurane and then titrated to 1-2% isoflurane for maintenance of a heart rate at around 500 beats per minute. The aortic and pulmonic
velocities were measured utilizing the pulse-wave doppler across the valves. Analysis was performed in a genotype blind fashion.

3.3.3 Tissue Fixation and Histology

For embryonic and adult time points, tissues were collected and fixed for at least 24 hours in 10% formalin. For calcification studies, tissues were fixed in 4% paraformaldehyde. Tissues were subsequently embedded in paraffin and serial sectioning was performed at a thickness of 6 microns. Staining was performed using Hematoxylin and Eosin (Sigma Aldrich), following standard procedure. Russel-Movat’s pentachrome (American MasterTech, KTRMP) was performed utilizing following standard protocol. Calcification staining was performed utilizing the Alizarin Red Solution (Millipore, 2003999).

3.3.4 Immunostaining

Immunohistochemistry was performed utilizing the Gata4 antibody (Santa Cruz Biotechnology, #SC-1237) at a 1:1000 dilution, followed with a secondary anti-goat antibody (Vectorlabs) at a 1:1000 dilution. Washing was performed with 1XPBST, ABC was performed with Vectastain ABC HRP kit and subsequent DAB was performed with the Vector DAB Peroxidase (HRP) Substrate kit.
3.3.5 AMIRA 3D Reconstruction

To analyze valve morphology at a 3-dimensional level, AMIRA 3D reconstruction software (version 5.5.0) was utilized. Briefly, 6 um serial sections were collected that contained the entire aortic valve, these sections were stained with hematoxylin and eosin, histologic sections were imaged, aligned with the AMIRA software, and then regions were selected with the software for reconstruction. Volumetric analysis was collected utilizing the measure function and was collected seperately for each leaflet.

3.3.6 RNA-sequencing

RNA-sequencing was performed at the Ohio State University genomic shared resource at the James Comprehensive Cancer Center. Differential gene expression analysis was performed in collaboration with OceanRidge Biosciences, NY. RNA-seq was performed on E15.5 outflow tracts that were microdissected in an RNase free setting and total RNA was collected using the Total RNA Purification kit (Norgen Biotek Corp, 17200). Libraries were generated using stranded total RNA TruSeq Stranded Total RNA LT w/ Ribo-Zero Gold Set A (illumina, RS-122-2301). 50 base paired end reads were performed on an illumina high seq 2500. Fastq files were aligned to the mouse genome (mm9) using TopHat. Differential gene expression was identified using Bioconductor easyRNASEq, with annotations from ensemble version 75mmu38.2. Data was filtered and the adjusted gene RPKMs were identified and subsequent statistical
analysis was performed. Principal component analysis identified a potential batch effect in sample KI2, it was therefore removed from the study. A heatmap was generated from the 580 genes with the lowest ANOVA P values (<0.05) (Figure 3-6 A). Expression intensity was log-2-transformed and a heatmap was generated using Gene Cluster 3.0, with the intensity values adjusted by centering genes. Clustering was performed using centered correlation as a distance measure and average linkage as method. Intensity is red on a red to green scale with representative log 2 units. The volcano plot was generated using R Studio version 1.0.136 and utilized ggplots2 to generate a graph of log2fold change (x axis) vs. -10log10 pvalue (y axis) with red color labeling to identify genes with an FDR<0.05, orange to denote expression change >1.5 fold, and green to identify those genes that have both an FDR<0.05 and a fold change >1.5 fold. Gene Ontology (GO) and KEGG Pathway analysis was performed using DAVID version 6.8 with the list of 1158 statistically significantly different genes, p value <0.05. GO Direct was used for Biological Process and Cellular Component enrichment analysis, providing a GO mapping that is directly annotated by the source database, with the exclusion of parent terms.

3.3.7 Statistics

Statistical analysis was performed using student’s t-test, where a p-value <0.05 was considered significant. An n of at least 3 was used for all experiments.
Figure 3-1 *Gata4* G295S<sup>Wt/Wt</sup> mice exhibit semilunar valve stenosis.
Figure 3-1 continued

(A-D) Echocardiography identified functional stenosis at one year (A,B) and two months of age (C,D) in Gata4 G295S<sup>ki/wt</sup> mice compared to controls. Stenotic valves indicated in red. Histologic sections of aortic valve of Gata4 G295S<sup>ki/wt</sup> mice exhibit thickened, dysmorphic leaflets (I,J) and abnormal ECM (K,L) shown with Movat’s Pentachrome staining at one year of age compared to wildtype control valves (E-H). (F,H,J,L) are high magnification images of (E,G,I,K) respectively. Alizarin red staining identified apparent calcific nodules (black arrow) at one year of age (O,P) compared to wild type controls (M,N). (P) and (N) are high magnification images of (O) and (M), respectively. Similarly, aortic valve sections at 2 months of age demonstrate thickened, dysmorphic leaflets (S,T) and abnormal ECM (W,X) compared to wildtype control valves (Q,R,U,V). White arrow highlights proteoglycan rich nodule formation. (R,T,V,X) are high magnification images of (Q,S,U,W). White arrow signifies region of collagenous nodule formation. Scale bar represents 100um.
Figure 3-2 Echocardiographic analysis of left ventricular function at one year and two months of age.

Left ventricular function of Gata4 G295S<sup>+/−</sup> and wildtype littermate controls at 2 months (A-C) and one year (D-F) of age. A) and D) represent left ventricular ejection fraction, B) and E) represent left ventricular end systolic volume, and C) and F) represent left ventricular end diastolic volume.
Figure 3-3 *Gata4* G295S<sup>ki/wt</sup> mice exhibit progressive valve disease progression postnatally.

(A-D) P21 aortic valves sections stained with Movat’s Pentachrome stain identify thickened leaflets with ECM disorganization and an enrichment of proteoglycans (blue) in the *Gata4* G295S<sup>ki/wt</sup> valves (C-D) compared to control (A-B). B) and D) are high magnification images of A) and C). E-H P10 aortic valve sections stained with Movat’s Pentachrome stain exhibit dysmorphic valves with abnormal ECM distributions in the *Gata4* G295S<sup>ki/wt</sup> valves (G-H) compared to control (E-F). F) and H) are high magnification images of E) and G). Scale bar represents 100µm.
Figure 3-4 GATA4 is expressed in the developing outflow tract

Immunohistochemistry identifies GATA4 expression in the aortic (A,B,E,F) and pulmonic (C,D,G,H) valve at E13.5 (A-D) and E18.5 (E-H) time points. B,D,F,H are high magnification images of A,C,E,G. Scale bar represents 100um.
Figure 3-5 *Gata4 G295S*<sub>ki/wt</sub> mice display developmental abnormalities. 

(A-E) H&E stained E11.5 histologic sections display smaller, less dense cushions in *Gata4 G295S*<sub>ki/wt</sub> outflow tract cushions (C-E) compared to wildtype littermate control (A-B). F-J are 3D reconstructed images of E13.5 H&E stained histologic sections of *Gata4 G295S*<sub>ki/wt</sub> (H'-J') and wildtype control littermates (F',G'). *Gata4 G295S*<sub>ki/wt</sub> reconstructed E13.5 valves (H-J) display smaller noncoronary cusps (NCC) compared to wildtype littermate controls (F-G), volumetric quantification shown in P. K-O are 3D reconstructed images of E15.5 H&E stained histologic sections of *Gata4 G295S*<sub>ki/wt</sub> (M'-O') and wildtype control littermates (K',L'). *Gata4 G295S*<sub>ki/wt</sub> reconstructed E15.5 valves (M-O) display smaller NCC cusps and abnormally shaped right coronary cusps (RCC) and left coronary cusps (LCC) compared to wildtype littermate controls (K,L), volumetric quantification shown in Q. *= pvalue<0.05. Scale bar represents 100um.
Figure 3-6 Developing *Gata4 G295S*\(^{ki/wt}\) cushions exhibit abnormal molecular expression profiles.
Figure 3-6 Continued

(A) Heatmap demonstrating differential gene expression of the 580 genes with the lowest ANOVA P values (<0.05) between Gata4 G295S<sup>ki/wt</sup> and wildtype littermate control OFTs at E15.5. The expression intensity is displayed as log-2-transformed and plotted on a red (upregulated) to green (downregulated) scale. (B) Volcano plot demonstrating the differential gene expression of the Gata4 G295S<sup>ki/wt</sup> OFTs compared to control. Red: FDR<0.05, orange: expression change >1.5 fold, and green: FDR<0.05 and a fold change >1.5 fold. (C-E) Gene Ontology (GO) Direct Pathway Analysis and KEGG pathway analysis were performed on the 1158 differentially expressed genes using DAIVD software. (C) GO Biological Process analysis identified an enrichment of many key heart development pathways including ECM organization, cell adhesion, and Wnt pathway regulation. (D) GO Cellular Component analysis identified an enrichment in ECM components. (E) KEGG Pathway analysis identified changes in key heart development pathways, including the most enriched pathway: Wnt signaling. (F) Many components of the Wnt signaling pathway are affected, including the canonical, planar cell polarity (PCP), and Wnt/Ca<sup>2+</sup> pathways. Affected components have a red asterisk, red arrow up indicates an increase in expression, while a green arrow down indicates a decrease in expression. (G) Of the genes that were significantly downregulated by at least 1.5 fold, Genomatix software identified several potential direct targets of Gata4. Those targets known to be associated with EMT are denoted by a red asterisk. (H) Wnt signaling KEGG pathway exhibits multiple components differentially expression in the Gata4 G295S<sup>ki/wt</sup> OFTs compared to wildtype littermate control. Green arrows down indicate decreased expression and red arrows up indicate increased expression in Gata4 G295S<sup>ki/wt</sup>. 
Chapter 4 Synopsis

Congenital heart disease (CHD) is the most common birth defect, affecting nearly 2% of live births. CHD is associated with increased morbidity and mortality, however the etiology by which it occurs is not well understood. Genetics have been well established as a key component to CHD, both through familial and population based studies. Although there is an established genetic component, genetic testing is not utilized clinically in patients with isolated CHD. Employment of genetic testing in the clinical setting can allow for a better understanding of mutation prevalence among patient populations and has clear benefits for patients. A caveat to the identification of these mutations, is the lack of clarity in the understanding of if these mutations are truly causal, in addition to the mechanism by which these mutations cause disease. A mutation is causal if it is sufficient to cause disease, therefore if one were to put this suspected variant into a model it should recapitulate disease if it is causal. By identifying the causative mutation and taking the next step of generating and analyzing models, we will be able to gain insight into disease etiology. This deeper understanding of how mutations cause CHD will open the door to a future of better diagnostics and potentially life-saving therapeutics. The purpose of this work is to utilize whole
exome sequencing to identify disease causing mutations and to apply mouse modeling to better understand the mechanism by which mutations can lead to CHD.

In chapter 2, we aimed to utilize a whole exome sequencing approach to identify disease causing mutations in patients with familial CHD. We focused on a patient population with concordant, non-left ventricular outflow tract (LVOT) phenotypes. We utilized an approach which prioritized variants that met a stringent criteria and were clearly associated with CHD. The variants were subsequently filtered based on rarity (<1% mean allele frequency) and *in silico* predicted pathogenicity. We identified 3 variants that were considered to be damaging or potentially damaging based upon the American College of Medical Genetics and Genomics (ACMG) Standards and Guidelines' recommended for determining variant pathogenicity. These variants were found in *GATA4*, *TLL1*, and *MYH11*. We subsequently went on to perform functional testing, when applicable, to determine *in vitro* deficits in protein functionality due to these mutations. Our findings support a role for whole exome sequencing in concordant cases of familial non-LVOT CHD and offer an approach that is practical and accessible for a clinical setting.

In chapter 3, we sought to utilize a mouse model to determine the molecular deficits that underlie a mutation identified in familial congenital valve stenosis. We employed the *Gata4 G295S*^<sub>ki/wt</sub> mouse model, which encompasses an orthologous disease causing mutation identified in a large family with inherited
semilunar valve stenosis and atrial septal defects. We utilized echocardiography and found that a subset of the \textit{Gata4} G295S\textsuperscript{ki/wt} mice exhibit high aortic and pulmonary valve velocities in comparison to control, at two months and one year of age. Additionally, through the utilization of histologic approaches we identified structurally dysmorphic valves with disorganized extracellular matrix (ECM), that are present and progress postnatally to a more severe phenotype by two months of age. This phenotype included collagenous nodules that calcified by one year of age. These mice exhibited postnatal valve disease, which suggested a congenital disease phenotype.

To determine the developmental onset of disease, we utilized histologic sections and 3D reconstruction with Amira software. We identified leaflets that were smaller at E11.5, likely due to a hypomorphic GATA4 allele known to play key roles in EMT in the atrioventricular valves. By E13.5 the left and right coronary cusps are of similar size, while the non-coronary cusp remained significantly smaller. At E15.5 the NCC remained significantly smaller in the mutant, while the RCC and LCC were comparable volumes between mutant and wildtype littermate controls. However, morphologically the \textit{Gata4} G295S\textsuperscript{ki/wt} cushions were extremely abnormal, specifically the left and right coronary cusp. At this time, point molecular profiling identified an enrichment of differentially expressed ECM genes and important developmental signaling pathways. The Wnt signaling pathway was the most enriched differentially expressed KEGG pathway. Taken together, this data supports the notion that the \textit{Gata4} G295S\textsuperscript{ki/wt}
allele is a hypomorph that leads to deficiencies in EMT. Although the leaflets are eventually able to compensate in volume, dysregulated molecular expression profiling shows that that cells are unable to function properly, ultimately causing valve stenosis.

To truly understand disease etiology, one must take a systems approach and combine data from human genetic studies to identify mutations (chapter 2) and follow them up with molecular biology approaches by utilizing mutation modeling (chapter 3). By taking this dual approach we will gain insight not only into what mutations are present in the population, but also how we can better diagnose and target these mutations with novel therapies. Although we were unable to model the specific mutations identified in our whole exome sequencing screen of familial CHD, due to time constraints, we were able to model a previously identified mutation in GATA4, which led to a better understanding of the molecular deficits that underlie the CHD phenotype. The next step for the mutations identified in chapter 2 is to generate mouse models. Although some in vitro analysis was performed in chapter 2, the incorporation of mouse modeling will determine if the variants are just contributing to or are causing disease. In the case of each of the 3 variants we identified, the amino acid sequences are conserved to mouse. This conservation makes these variants ideal candidates for mouse generation with an orthologous mutation. Additionally, with the advent of CRISPR/cas9 genome engineering, these mice can be generated in a time and cost considerate manner. While the molecular analysis is not trivial, and
there are some caveats to mouse generation, this approach is ideal for identifying the causality of mutations in CHD.

Although we successfully identified mutations in 33% of the familial cases of CHD that we analyzed, we also identified variants of uncertain significance according to the ACMG Standard and Guidelines in two families. First, in family C we identified a R875W variant in EVC2. This family had atrial septal defects (ASD) and was the same family that we identified the likely pathogenic G115W mutation in GATA4. Mutations in EVC2 are typically associated with Ellis-van Creveld syndrome, a syndrome characterized by dwarfism, polydactyly, dental abnormalities, and CHD. CHD occurs in 50% of Ellis-van Creveld syndrome cases, with common atria (CA) being the most frequently associated phenotype, occurring due to a failure of the atria to septate. The family members of family C did not display any phenotypes that would match that of Ellis-van Creveld syndrome, outside of the CHD phenotype. Interestingly, ASD and CA are the result of abnormalities in atrial septation. Mutations in genes associated with syndromes have been shown to cause isolated cases of CHD, so there is a possibility that this mutation could be contributory. However, given what is known about GATA4 and its association to ASD, it is much more likely that this phenotype is due to the G115W mutation in GATA4. By generating mouse models with each of these mutations it can be determined if either is sufficient to cause ASD. Finally, family H displayed tetralogy of Fallot (TOF), and while the ACMG Standards and Guidelines did not identify any likely pathogenic variants, a
variant V321M in *MYBPC3* and variant K1241E in *SOS1* were identified. These variants were filtered in an autosomal recessive manner, meaning the variant was kept for further analysis if it was present in one of the parents, even though neither parent had TOF. No clearly autosomal recessive variant was identified in this family, but variants that met the earlier criteria and that were present in a parent and the affected offspring were kept. This filtering approach made it difficult to determine if a variant was truly contributory or causative. One potential idea is that one mutation is inherited from the father and a second mutation is inherited from the mother, leading to a combinatorial effect that causes disease. However, in this specific case, both the *MYBPC3* and K1241E mutation are present in the father, so if these two mutations were sufficient to cause disease, he should also exhibit TOF. As the data so far are inconclusive, further analysis, including potential mouse modeling, is required to determine how these variants may be contributing to disease. Subsequent analysis includes setting less stringent filtering criteria and utilizing a larger prioritization list, taking into consideration complementing mutations coming from each parent that may cause disease, and considering that this could be an oligogenic disease.

We were unable to identify pathogenic mutations in 6 of the 9 families. The approach taken in chapter 2 specifically looked for variants existing within a small CHD candidate gene list, only considered a single gene mode of inheritance, and did not consider non-coding or epigenomic changes. By focusing specifically on the 69 genes that are known to be associated with CHD
we were unable to identify variants that may have existed in important genes known to be expressed during cardiovascular development, that when knocked out in mouse lead to a CHD phenotype, and that have not yet been identified. We used this approach to allow for the straightforward analysis of a complex dataset to lead to the identification of variants in genes that we know cause CHD, however we were aware that things may be missed with our approach. We were able to identify likely causal mutations based on our strict criteria, however by incorporating modifications to our pipeline, one can take our basic approach and apply it to oligogenic approaches. Additionally, our approach can be modified to increase a larger list of candidate genes that will allow for the identification of genes not previously associated with CHD. The approach we have developed is modifiable due to the “virtual” nature of our sequencing panel, which is amenable and can allow for re-querying of the obtained sequencing data.

It is predicted that the majority of CHD is genetically complex and most likely due to an oligogenic disease etiology, including multiple genetic contributors.94 By taking an oligogenic disease approach, we predict that in 2 of the 9 families discussed in chapter 2, that this may be occurring, specifically in family H and I which appear to have recessively inherited disease. We have taken an approach in which we consider the affected individuals as needing multiple “gene hits” to push them over the threshold of “normally developed heart” to “CHD,” which is considered an oligogenic disease etiology. In theory, these variants can be inherited from an unaffected parent; however the affected
offspring must contain both (or more) variants, which are deemed necessary to cause disease. Future experiments to better study this potential disease model include the integration of pathway network analysis to identify potential genetic interactors, followed by subsequent work modeling these potential interactors to determine \textit{in vivo} synergy.

Additional players in congenital heart disease include environment and epigenetics. As we found in family F, environment can play an important role in the development of CHD. It is essential that when performing genetic analysis of patients for CHD, that proper medical records and patient phenotyping are implemented. If a patient was exposed to a teratogen during development, such as maternal illness, it can lead to a phenocopy resulting in issues when analyzing variant segregation. One way that the environment can lead to disease is through changes to the epigenome. As our knowledge of the mechanisms of epigenomics continues to increase, we will be faced with new tasks of how we can translate this information to clinical settings. Unfortunately, testing for epigenomic changes is difficult, as it requires diseased tissue and subsequent cell sorting and separation to better understand what chromatin changes are occurring. As single cell ATAC-seq and new technologies continue to evolve, this technique could become more clinically useful in the future.

An additional problem that we run into with whole exome sequencing in low coverage. One infamously difficult region of the genome to capture in whole exome sequencing is the region containing \textit{GATA4}, this is one way that panels
surpass whole exome. However, as capture kits continue to improve, this is something that should be resolved. This can also be resolved by the utilization of whole genome sequencing. Additionally, one must consider that with whole exome sequencing we are only considering regions of the genome that code for protein, or “the exome”. At this time, data analysis techniques are able to analyze the exome in a more efficient and appropriate way than what can be performed for the entire genome. For example, we have many in silico pathogenicity tests that we can perform on variants within the exome, and equivalents just do not exist for non-coding variants, even those that occur in well conserved enhancers or promoters. A lot of time an energy has been put into better understanding the non-coding genome, a huge impact has been made by the ENCODE project, which has led to the enlightenment of the role of many enhancers. Additionally, projects focusing on understanding the role of enhancers specifically in the heart, such as the recent work from Len Pennachio’s lab, which identified thousands of active enhancers in the heart and modeled several of them in vivo. Identification of these enhancers and the interesting overlap to genome wide association studies of patients with cardiovascular disease has led to the identification of specific haplotypes occurring within these newly identified enhancers. It is only a matter of time before this data can be easily applied to whole genome sequencing data. The caveat is how we will model variants identified in these regions, as many of these noncoding regions are not conserved to mouse. In such case, determining pathogenicity of a variant occurring in an unconserved
noncoding region will be difficult. Of course, this can be resolved through the utilization of human induced pluripotent stem cells (iPSCs), however the enigma of how one truly models disease in a cell line remains. Additionally, one must consider the specific cell type that the iPSCs would need to be differentiated to, which when considering complex CHD is not trivial. For example, when one considers valve disease, what would be considered the appropriate terminally differentiated cell to focus the study on? One could answer endothelial cells and one could answer valve interstitial cells, however given the different lineages that contribute to the interstitial cells does it matter if that cell being studied goes through a neural crest lineage transformation or an EMT transformation? Additionally, one should consider the possibility that both cell types are required for appropriate signaling to occur. These are questions that are yet to be answered, however these questions are sure to be tackled as the technology of iPSCs continues to improve and as we continue to identify variants in regions of the genome that are not conserved to other models.

In chapter 2 we focus on right sided heart disease that presents in a concordant fashion, however, we know that this isn’t always the case. In cases of left ventricular outflow tract (LVOT) lesions, it is known that the phenotypes tend to be more variable and it has been difficult to pinpoint monogenic causes of familial LVOT disease\textsuperscript{273,274}. While our current approach may not be suitable for patients with LVOT, our lab, in collaboration with others, are working towards a pipeline that can accurately identify disease contributing variants. This analysis
will include predicted pathogenicity as well as the potential incorporation of pathway network analysis, associating the variants to the developmental pathways they may be disrupting. This will allow for a more systematic approach, in a similar way to how we have described the workflow of our oligogenic disease etiology pipeline. In this way, we may be able to accurately identify disease contributing variants.

As the technologies that allow us to discover and model mutations continue to evolve, our ability to gain an understanding of disease etiology also evolves. As shown in Figure 4-1, with the advent of next generation sequencing, the past decade has led to an exponential increase in our knowledge behind the role of genetics in CHD. Better sequencing, such as the touted $100 per genome capabilities of the NovaSeq, and better ways to model disease, including genome editing with CRISPR/cas9 and TALENs, have led us to an acceleration in the identification of genetic causes of CHD. Additionally, new databases such as the Broad Institute’s “gnomAD” genome aggregate database, which boasts data from 123,136 whole-exome sequences and 15,496 whole-genome sequences, allows for a better understanding of population frequency of identified variants. While we, and others, have only been able to accurately predict disease causing mutations in a subset of patients, by utilizing the approaches we have listed in Chapters 2 and 3, and by incorporating the ideas of oligogenic disease etiology, modifications to prioritization lists, incorporation of noncoding regions of the genome, and utilization of modern tools in modelling, we predict that the next
decade will lead to more advances in the understanding of the genetic contribution to CHD. These advancements can translate to clinical application and lead to a positive impact on the lives of patients with CHD.
4.1 Figures

Figure 4-1 Timeline of CHD Genetic Discoveries and the Genetic Technologies and Study Designs Used. Genetic technologies/study designs are indicated by blue arrows and mark the approximate time when the technology was developed and used. (Blue et al., JACC 2017)

Genetic technologies/study designs are indicated by blue arrows and mark the approximate time when the technology was developed and used. AVSD = atrioventricular septal defect; BWIS = Baltimore-Washington Infant Study; CHD = congenital heart disease; CMA = chromosomal microarray; CNV = copy-number variation; ES = exome sequencing; GWAS = genome-wide association studies; MPS = massively parallel sequencing; NDD = neurodevelopmental disabilities; nsCHD = nonsyndromic CHD; RR = recurrence risk; sCHD = syndromic CHD.
References


75. Shirai, M., Imanaka-Yoshida, K., Schneider, M.D., Schwartz, R.J. & Morisaki, T. T-box 2, a mediator of Bmp-Smad signaling, induced


173. Sanford, L.P. et al. TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. Development 124, 2659-70 (1997).


175. Bax, N.A. et al. Cardiac malformations in Pdgfalpha mutant embryos are associated with increased expression of WT1 and Nkx2.5 in the second heart field. Dev Dyn 239, 2307-17 (2010).


### Appendix A: Candidate Gene List

**Table A.1:** Congenital heart disease candidate gene list

<table>
<thead>
<tr>
<th>Gene</th>
<th>Ensemble ID</th>
<th>Associated Congenital Heart Defect</th>
<th>Syndrome With Associated Congenital Heart Defect</th>
<th>References</th>
<th>Clinical Testing Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTC1</td>
<td>ENSG00000159251</td>
<td>Atrial Septal Defect, Left Ventricular noncompaction</td>
<td>Khassen et al., Circulation 2008</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>ACVR1</td>
<td>ENSG0000011170</td>
<td>Aortic/ventricular septal defects</td>
<td>Smith et al., Circulation 2009</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>ACVR2B</td>
<td>ENSG00000114730</td>
<td>Dextrocardia, Single Atrium, Single Ventricle, Transposition of the Great Arteries, Aneurysm Pulmonary Valve Return, Aortic/ventricular Canal Defect, Coarctation of the Aorta, Pulmonary Valve Atresia, Pulmonary Valve Stenosis, Double Outlet Right Ventricle</td>
<td>Heterozygosity</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>ALDH1A2</td>
<td>ENSG00000128919</td>
<td>Tetralogy of Fallot</td>
<td>Parvan et al., BMC Med Genet 2009</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>ANKR1D1</td>
<td>ENSG00000148677</td>
<td>Total anomalous pulmonary venous return</td>
<td>Chequetti et al., Hum Mutat 2008</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>BRAF</td>
<td>ENSG00000157766</td>
<td>Noonan Syndrome, Omenn-like/syndromic features</td>
<td>Sankoy et al., Hum Mutat 2009</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>CFC1</td>
<td>ENSG00000139690</td>
<td>Dextrocardia, Single Atrium, Single Ventricle, Transposition of the Great Arteries, Aneurysm Pulmonary Valve Return, Aortic/ventricular Canal Defect, Coarctation of the Aorta, Pulmonary Valve Atresia, Pulmonary Valve Stenosis, Double Outlet Right Ventricle</td>
<td>Heterozygosity</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>CM2</td>
<td>ENSG00000171313</td>
<td>Tetralogy of Fallot, Double Outlet Right Ventricle, Ventricular Septal Defect, Atrial Septal Defect, Patent Ductus Arteriosus, Tetralogy of Fallot, Pulmonary Valve Stenosis, Mitral Tricuspid Valve Stenosis</td>
<td>CHARGE Syndrome</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>CTED2</td>
<td>ENSG00000164442</td>
<td>Atrial Septal Defect, Ventricular Septal Defect</td>
<td>Speling et al., Hum Mutat 2009</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>CREBBP</td>
<td>ENSG0000006338</td>
<td>Atrial Septal Defect, Ventricular Septal Defect, Patent Ductus Arteriosus, Coarctation of the Aorta, Pulmonary Stenosis, Bicuspid Aortic Valve</td>
<td>Rubinstein-Taybi syndrome</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>CRL1D1</td>
<td>ENSG00000126703</td>
<td>Dextrocardia, Single Atrium, Single Ventricle, Transposition of the Great Arteries, Aneurysm Pulmonary Valve Return, Aortic/ventricular Canal Defect, Coarctation of the Aorta, Pulmonary Valve Atresia, Pulmonary Valve Stenosis, Double Outlet Right Ventricle</td>
<td>Heterozygosity</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>ELN</td>
<td>ENSG00000035840</td>
<td>Aortic Valve Stenosis</td>
<td>Curr et al., Cell 1992</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EP300</td>
<td>ENSG00000103393</td>
<td>Atrial Septal Defect, Ventricular Septal Defect, Patent Ductus Arteriosus, Coarctation of the Aorta, Pulmonary Stenosis, Bicuspid Aortic Valve</td>
<td>Rubinstein-Taybi syndrome</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EVC</td>
<td>ENSG00000072840</td>
<td>Atrial Septal Defect, Aortic/ventricular Canal Defect, Common Atrium, Left Superior Vena Cava, Ventricular Septal Defect, Pulmonary Valve Stenosis, Transposition of the Great Arteries</td>
<td>Ellis-van Creveld Syndrome</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EVC2</td>
<td>ENSG00000170340</td>
<td>Atrial Septal Defect, Aortic/ventricular Canal Defect, Common Atrium, Left Superior Vena Cava, Ventricular Septal Defect, Pulmonary Valve Stenosis, Transposition of the Great Arteries</td>
<td>Ellis-van Creveld Syndrome</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table A.1 Continued

<table>
<thead>
<tr>
<th>Gene</th>
<th>Full Reference</th>
<th>Full Description</th>
<th>PATJ Syndrome</th>
<th>Application</th>
<th>Gene Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBN1</td>
<td>ENSG00000168147</td>
<td>Bicuspid Aortic Valve, Aortic Regurgitation, Mitral Valve Prolapse, Aortic Dissection, Aortic Anteryism</td>
<td>Marfan syndrome</td>
<td>Yes</td>
<td>Dez et al., Nature 1991</td>
</tr>
<tr>
<td>FLNA</td>
<td>ENSG00000199224</td>
<td>Cardiac Valvular Dysplasia</td>
<td></td>
<td>Yes</td>
<td>Boniouch et al., Am J Hum Genet 1997</td>
</tr>
<tr>
<td>FOG2</td>
<td>ENSG00000199486</td>
<td>Tetralogy of Fallot, Double Outlet Right Ventricle</td>
<td></td>
<td>Yes</td>
<td>Rituz et al., Hum Mutat 2003</td>
</tr>
<tr>
<td>FOXH1</td>
<td>ENSG00000160073</td>
<td>Tetralogy of Fallot, Ventricular Septal Defect</td>
<td></td>
<td>Yes</td>
<td>Wang et al., Int J Cardio 2010</td>
</tr>
<tr>
<td>GATA4</td>
<td>ENSG00000130574</td>
<td>Atrial Septal Defect, Pulmonary Valve Stenosis, Tetralogy of Fallot, Ventricular Septal Defect</td>
<td></td>
<td>Yes</td>
<td>Garg et al., Nature 2003</td>
</tr>
<tr>
<td>GATA6</td>
<td>ENSG00000141448</td>
<td>Atrial Septal Defect, Atrioventricular Septal Defect, Persistent Tricuspid Arteriosus, Tetralogy of Fallot</td>
<td></td>
<td>Yes</td>
<td>Kado et al., PHAS 2009</td>
</tr>
<tr>
<td>GDF1</td>
<td>ENSG00000130283</td>
<td>Heterotaxy, Transposition of Great Arteries, Tetralogy of Fallot, Double Outlet Right Ventricle</td>
<td></td>
<td>Yes</td>
<td>Lin et al., J Hum Genet 2010</td>
</tr>
<tr>
<td>JAG1</td>
<td>ENSG00000152861</td>
<td>Atrial Septal Defect, Hypoplastic Left Heart Syndrome</td>
<td></td>
<td>Yes</td>
<td>Dasgupta et al., Mutat Res 2001</td>
</tr>
<tr>
<td>HAND1</td>
<td>ENSG00000164107</td>
<td>Atrial Septal Defect, Hypoplastic Left Heart Syndrome</td>
<td></td>
<td>No</td>
<td>Reardon-Buettner et al., Hum Molec Genet 2009, Reardon-Buettner et al., Hum Molec Genet 2009</td>
</tr>
<tr>
<td>LEFTY2</td>
<td>ENSG00000143766</td>
<td>Aortic Stenosis, Single Artery, Single Ventricle, Transposition of the Great Arteries, Atrioventricular Canal Defect, Coarctation of the Aorta, Pulmonary Valve Atrioes, Pulmonary Valve Stenosis, Double Outlet Right Ventricle</td>
<td></td>
<td>Yes</td>
<td>Heterotaxy</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>MAP2K1</th>
<th>ENSG00000169032</th>
<th>146</th>
<th>286</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP2K2</td>
<td>ENSG00000126634</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>MAFR3</td>
<td>ENSG0000015479</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>MED12</td>
<td>ENSG00000123006</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>MED13L</td>
<td>ENSG00000123066</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>MYBP3</td>
<td>ENSG00000134571</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>MYH11</td>
<td>ENSG00000133392</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>MYH6</td>
<td>ENSG00000197618</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>MYH7</td>
<td>ENSG0000092654</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>NIPR1</td>
<td>ENSG00000164190</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>NOG2-5</td>
<td>ENSG00000183072</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>NOG2-6</td>
<td>ENSG00000180553</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>NOGAL</td>
<td>ENSG00000158574</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>NOTCH1</td>
<td>ENSG00000146800</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>NOTCH2</td>
<td>ENSG00000134250</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>NRAS</td>
<td>ENSG00000213281</td>
<td>306</td>
<td></td>
</tr>
<tr>
<td>PDGFRA</td>
<td>ENSG00000134853</td>
<td>306</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1 Continued

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Condition</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicuspid Aortic Valve, Aortic Coarctation, Patent Ductus Arteriosus</td>
<td></td>
<td>Quintino-Rivera et al., Hum Mol Genet 2015</td>
</tr>
<tr>
<td>Transposition of Great Arteries</td>
<td>Mucinoid, Circulation 2003</td>
<td></td>
</tr>
<tr>
<td>Left ventricular noncompaction</td>
<td></td>
<td>Probst et al., Circ Cardiovasc Genet 2011</td>
</tr>
<tr>
<td>Atrial Septal Defect</td>
<td></td>
<td>Zhu et al., Hum Mol Genet 2007</td>
</tr>
<tr>
<td>Left ventricular noncompaction, Ebstein Anomaly, Atrial Septal Defects</td>
<td></td>
<td>Ching et al., Nature Genet 2005</td>
</tr>
<tr>
<td>Atrial Septal Defect, Ventricular Septal Defect, Pulmonary Stenosis</td>
<td></td>
<td>Postma et al., Clin Cardiovasc Genet 2011</td>
</tr>
<tr>
<td>Atrial Septal Defect, Ventricular Septal Defect, Pulmonary Stenosis</td>
<td></td>
<td>Engelen et al., Neth Heart J 2013</td>
</tr>
<tr>
<td>Atrial Septal Defect, Congenital Heart Malformations, Hypoplastic Left Heart Syndrome, Tetralogy of Fallot, Ventricular Septal Defect</td>
<td></td>
<td>Borok et al., J Med Genet 2004</td>
</tr>
<tr>
<td>Congenital Heart Malformations, Pulmonary Venous Return</td>
<td></td>
<td>Cleland et al., J Clin Invest 1999</td>
</tr>
<tr>
<td>Deoxycardia, Single Atrium, Single Ventricle, Transposition of the Great Arteries, Anomalous Pulmonary Venous Return, Atrioventricular Canal Defect, Coarctation of the Aorta, Pulmonary Valve Atelectasis, Pulmonary Valve Stenosis, Double Outlet Right Ventricle</td>
<td></td>
<td>Goldmuntz et al., Circulation 2001</td>
</tr>
<tr>
<td>Aortic Valve Disease, Bicuspid Aortic Valve, Left Ventricular Outflow Tract Obstruction</td>
<td></td>
<td>McGinn et al., J Am Coll Cardiol 2003</td>
</tr>
<tr>
<td>Tetrology of Fallot, Pulmonary Valve Stenosis, Atrial Septal Defect, Ventricular Septal Defect</td>
<td></td>
<td>Stokhuysen et al., Clin Genet 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table A.1 Continued

<table>
<thead>
<tr>
<th>Gene</th>
<th>Entrez Gene ID</th>
<th>Syndrome Description</th>
<th>Diagnosis Method</th>
<th>Reference</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAH1</td>
<td>ENSG00000108557</td>
<td>Pulmonary Valve Stenosis, Atroventricular Septal Defect, Tricuspid Stenosis, Mitral Stenosis, Aortic Stenosis, Pulmonic Stenosis, Tetralogy of Fallot</td>
<td>Smith-Magenis syndrome</td>
<td>Seranski et al., Genomics 1999; Seranski et al., Gene 2001; Sager et al., Nature Genet 2003</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Gene</th>
<th>Ensembl ID</th>
<th>Severe Cardiac Malformations</th>
<th>Related Conditions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLL1</td>
<td>ENSG00000038395</td>
<td>Atrial Septal Defect</td>
<td></td>
<td>Stanczak et al., Europ J Med Genet 2009</td>
</tr>
<tr>
<td>VEGF4</td>
<td>ENSG00000112715</td>
<td>Left Ventricular Outflow Tract Obstruction</td>
<td></td>
<td>Zhao et al., Biochem Biophys Res Comm 2010</td>
</tr>
</tbody>
</table>