Enrichment for Rare Pathogenic and Highly Damaging Variants in Congenital Heart Disease Patients

by

Jason Christopher Carson

BS, Allegheny College, 2004

MS, University of Pittsburgh, 2009

Submitted to the Graduate Faculty of the

Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Jason Christopher Carson

It was defended on

April 24, 2020

and approved by

Ryan Minster, PhD, MSIS, Assistant Professor, Human Genetics, Graduate School of Public Health, University of Pittsburgh

Dennis Kostka, PhD, Associate Professor, Developmental Biology, School of Medicine, University of Pittsburgh

Hyun Jun Park, PhD, Assistant Professor, Human Genetics, Graduate School of Public Health, University of Pittsburgh

Dissertation Director: Cecilia Lo, PhD, Professor and F. Sargent Cheever Endowed Chair, Developmental Biology, School of Medicine, University of Pittsburgh

Cecilia Lo, PhD

Enrichment for Rare Pathogenic and Highly Damaging Variants in Congenital Heart Disease Patients

Jason Christopher Carson, PhD

University of Pittsburgh, 2020

Abstract

Congenital heart disease (CHD), defined as any structural abnormality of the heart or great vessels, is a leading cause of morbidity and mortality in infants and young children. Even as survival rates for the most severe forms of CHD improve, patients still face a lifetime of monitoring, diagnostic procedures, and possible additional surgeries to address these defects or related issues. Compounding matters is recent evidence suggesting CHD patients may have an increased lifetime risk of developing cancer either due to some of the same genetic mutations that caused the CHD or as a result of other extrinsic factors such as repeated exposure to radiation during diagnostic or interventional procedures. The genetic etiology of CHD is quite complex with both sporadic and familial forms as well as monogenic forms with simple Mendelian inheritance and multigenic forms with complex inheritance patterns. This study investigated whether CHD patients have an increased burden of rare pathogenic or highly damaging variants in genes known to be associated with congenital heart defects or cancer. My results indicate pathogenic and highly damaging variants in some genes with a known association with CHD, but not in genes associated with cancer play a role in the etiology of heart development. Furthermore, these results appear to be consistent with a complex multigenic model of CHD. Finally, they suggest that variants in genes related to response to oxygen levels may serve as a prognostic indicator. However, this study has several limitations, the most notable being a nonrandom sample population and a lack of control for population substructure. In spite of these limitations, I feel

that this research provides a good foundation on which to base future more in-depth analyses with more statistical power. This research is of public health importance as it can help shed light on the genetic etiology of CHD and provide insight into which patients may be at greater risk for later complications which in turn could lead to improved diagnostics, treatment, and preventative measures.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would first like to thank Dr. Cecilia Lo for taking me on as a graduate student in her lab. I would also like the thank Dr. Lo as well as the other members of my committee, Dr. Ryan Minster, Dr. Dennis Kostka, and Dr. HJ Park, for their advice and guidance throughout this project. I would like to thank Omar Khalifa and Bill Devine for their assistance in determining patient phenotypes as well as Abha Bais, Kylia Williams, and Wenjuan Zhu (Hong Kong University) for their assistance in processing the whole exome data and generating the exome spreadsheets. I would also like to thank Dr. Frank Fan for sitting in for Dr. Park for my qualifying exam. Additionally, I would like the thank Dr. Bruce Aronow, a study collaborator at the University of Cincinnati, for his assistance in creating the CHD-ome gene list. Finally, I want to thank the entire Lo Lab and Human Genetics department for the guidance and assistance they have provided. Lastly, I want to thank all those who believed in me and had confidence in my ability to succeed even when I never believed in myself.

Abbreviations

AAD	Acute aortic dissection
ANF	Atrial natriuretic factor
ASD	Atrial septal defect
AVSD	Atrioventricular septal defect
BAF	Brg1/Brm-associated factor
BAV	Bicuspid aortic valve
BH FDR	Benjamini-Hochberg false discovery rate
BMP	Bone morphogenetic protein
CADD	Combined annotation dependent depletion
CHD	Congenital heart disease
CNV	Copy number variation
CoA	Coarctation of the aorta
CPLANE	Ciliogenesis and planar cell polarity effector
DORV	Double outlet right ventricle
ECM	Extracellular matrix
EGF	Epidermal growth factor

EMT	Epithelial-to-mesenchymal transition
ENU	Ethylnitrosourea
FC	Fold change
FGF	Fibroblast growth factor
FHF	First heart field
GO	Gene ontology
GPCR	G protein-coupled receptor
GWAS	Genome wide association study
HLHS	Hypoplastic left heart syndrome
I-SMAD	Inhibitory SMAD
IAA	Interrupted aortic arch
IPA	Ingenuity Pathway Analysis
LVOTO	Left ventricular outflow tract obstruction
NGS	Next generation sequencing
OMIM	Online Mendelian Inheritance in Man
OR	Odds ratio
PCA	Principal component analysis
PCD	Primary ciliary dyskinesia

PDA	Patent ductus arteriosus
PRC2	Polycomb repressive complex 2
PTA	Persistent truncus arteriosus
RVOTO	Right ventricular outflow tract obstruction
SCOOP	Severe childhood onset obesity project
SHF	Second heart field
Shh	Sonic hedgehog
SNP	Single nucleotide polymorphism
TAPVR	Total anomalous pulmonary venous return
TGA	Transposition of the great arteries
TGF-β	Transforming growth factor-beta
TOF	Tetralogy of Fallot
VEGF	Vascular endothelial growth factor
VSD	Ventricular septal defect
WES	Whole exome sequencing

1.0 Introduction

1.1 Congenital Heart Disease: Definition and Prevalence

Congenital heart disease (CHD) refers to any structural abnormality of the heart or great vessels and can range from relatively benign small septal defects or bicuspid aortic valve to highly complex defects involving multiple structures that require open-heart surgery in the immediate neonatal period. Presently the incidence of CHD is estimated at between four and ten per 1000 live births, though this value has increased from an estimated 0.6 per 1000 live births in the 1930s until plateauing at around 9.1 per 1000 around 1995 (Pierpont et al., 2007; van der Linde et al., 2011). This increase has been attributed to improved diagnostic technologies rather than a true increase. Asian populations seem to show a higher incidence of CHD with an increased prevalence of pulmonary outflow defects and fewer left ventricular outflow defects (van der Linde et al., 2011). The overall prevalence of CHD has also been reported to be from 2.5-13.8 per 1,000 births with severe CHD accounting for 0.5-5.1 per 1,000 (Reviewed in van der Bom et al., 2011). A study from Malaysia reported a prevalence of 6.7 per 1000 livebirths with 38% classified as "severe", 15% as "moderate" and 47% as "mild" while also noting a significant increase from 2005 to 2015 (Mat Bah et al., 2018).

An American study reported a slight decrease in the prevalence of severe CHD from 168.9 per 100,000 births in 1999 to 129.3 per 100,00 births in 2008 (A. Egbe, Uppu, Lee, Ho, & Srivastava, 2014). In particular, significant decreases in the prevalence of persistent truncus arteriosus (PTA), tetralogy of Fallot (TOF), and hypoplastic left heart syndrome (HLHS) were noted during this period. This decrease was thought to be related to increased termination of

pregnancy for severe CHD with possible influences of socioeconomic status and access to specialized pre- and perinatal care. Conversely, a German study reported an increased prevalence of severe CHD in recent decades (Pfitzer et al., 2017). This increase was attributed to improvements in detection and treatment leading to increased survival. They also reported ventricular septal defect (VSD), atrial septal defect (ASD), TOF, univentricular heart, and coarctation of the aorta (CoA) were the most commonly seen types of CHD.

There are many different types of structural heart defects that can be grouped into several categories that are not necessarily mutually exclusive including septal defects, right ventricular outflow tract obstruction (RVOTO) defects, left ventricular outflow tract obstruction (LVOTO) defects, conotruncal defects, valve defects, and heterotaxy spectrum defects. Atrial septal defects are the third most common type of CHD with a prevalence of 56-100 per 100,000 live births and can be subdivided into primum, secundum, and sinus venosus defects (Reviewed in Geva, Martins, & Wald, 2014). While secundum defects show a female predominance, primum and sinus venosus defects show near equal prevalence to a slight male predominance. Primum ASD can be considered a variant of common atrioventricular canal in which there is an interatrial communication between the anterior-inferior margin of the foramen ovale and the atrioventricular valve while secundum ASD are defects in the foramen ovale arising from one or more defects in the septum primum. The foramen ovale is a normal interatrial connection during fetal life that normally closes after birth when increased left atrial pressure presses the septum secundum over the septum primum, though it can remain patent in a small percentage of the population. Sinus venosus defects involving the venae cavae or coronary sinus are rare (Thiene & Frescura, 2010). These defects involve a communication between one or more of the right pulmonary veins with the cardiac orifice of the superior vena cava (superior defect) or with the posterior-inferior wall of

the right atrium just above the inferior vena cava orifice (inferior defect) (Geva et al., 2014). Approximately 87% of these defects occur between the upper right pulmonary vein and superior vena cava. Finally, a coronary sinus defect occurs when there is a partial or complete absence of the tissue that normally separates the coronary sinus from the left atrium (unroofed coronary sinus) and can be associated with persistent left superior vena cava.

Atrioventricular septal defects (AVSD) refer to a spectrum of defects characterized by a common atrioventricular connection with interatrial and interventricular connections (Calkoen et al., 2016; Thiene & Frescura, 2010). These defects are seen in 4-5.3 per 10,000 livebirths and comprise 7% of all CHD but approximately 50% of CHD seen in Down syndrome patients (Reviewed in Calkoen et al., 2016). A complete AVSD is defined as co-occurrence of a primum ASD, a non-restrictive VSD, and a common pentacuspid atrioventricular valve with shunting at the atrial and ventricular levels. There are two forms of partial AVSD with the more common variant consisting of an isolated primum ASD with the bridging leaflets of the common atrioventricular valve attached to the ventricular septum and shunting only at atrial level while the less common variant has the bridging leaflets attached to the atrial septum and shunting only at the ventricular level. A transitional AVSD comprises a primum ASD, restrictive VSD, and fusion of the bridging leaflets resulting in two orifices. Finally, an unbalanced AVSD occurs when the common atrioventricular valve is not evenly positioned above the ventricles such that one ventricle receives a dominant portion of the blood flow and can occur concurrently with any of the other three types. Complete AVSD are more common accounting for 56-75% of all AVSD and are more frequently seen in syndromic patients while 8-10% of complete AVSD are of the unbalanced type and are more commonly seen in non-syndromic patients (Calkoen et al., 2016). Malformation of

the endocardial cushions, abnormal formation of the mesenchymal cap or dorsal mesenchymal protrusion, or looping defects may all contribute to the etiology of AVSD.

Ventricular septal defects account for up to 40% of all cardiac anomalies, though the smallest ones may be asymptomatic and close spontaneously making the true incidence difficult to ascertain (Reviewed in Penny & Vick, 2011). They can occur in isolation or in conjunction with and as an integral component of other CHD such as TOF, double outlet right ventricle (DORV), and transposition of the great arteries (TGA). Defects can occur in the membranous septum (perimembranous) or within the inlet, outlet, or apical portion of the muscular septum (Thiene & Frescura, 2010). Incomplete formation of the primitive interventricular septum may contribute to trabecular septal defects while muscular defects are thought to arise from excessive undermining beneath and between trabeculae (Penny & Vick, 2011). Inlet VSD arise from failed fusion of the atrioventricular cushions with each other or with the primitive septum while outlet VSD result from poor development or malalignment of the outlet cushions. Incomplete closure of the region forming the membranous septum coupled with underdevelopment of parts of the muscular septum result in perimembranous defects. Finally, infundibular septal defects occur just below the semilunar valves and may or may not extend into the membranous septum (Thiene & Frescura, 2010).

Right ventricular outflow tract obstruction comprises abnormalities of the outlet portion of the right ventricle, pulmonary valve, and main pulmonary artery. Stenosis or narrowing of the pulmonary outflow can be classified as subvalvar, valvar, or supravalvar and can occur with or without septal defects (Thiene & Frescura, 2010). Valvar pulmonary stenosis is relatively common, accounting for ∽10% of CHD and can have a wide range of presentations from unicuspid to tricuspid and with varying degrees of dysplasia (Reviewed in Kwiatkowski, Hanley, &

Krawczeski, 2016). Pulmonary atresia with intact ventricular septum accounts for less than 10% of cyanotic CHD and is often characterized by membranous pulmonary atresia, hypoplastic pulmonary root, small right ventricle, and abnormal tricuspid valve. Accompanying coronary artery anomalies can lead to right ventricle-dependent coronary circulation (Kwiatkowski et al., 2016).

Left ventricular outflow tract obstruction involves abnormalities of the outlet portion of the left ventricle, aortic valve, ascending aorta, and aortic arch. As with pulmonary stenosis, aortic stenosis can be divided into subvalvar, valvar, and supravalvar types. Subaortic stenosis is present in 6.5% of adults with CHD and is twice as common in males as in females (Reviewed in Aboulhosn & Child, 2015). This defect most commonly arises due to a fibrous membrane or muscular narrowing of the left ventricular outflow tract but can rarely be caused by attachment of accessory mitral chordae to the base of the ventricular septum. Bicuspid aortic valve (BAV) is seen in 25% of patients and other obstructive lesions may also occur such as Shone's complex which includes mitral valve stenosis and coarctation of the aorta. Supravalvar aortic stenosis is defined as a focal or diffuse narrowing of the ascending aorta and, rarely, the aortic arch. It is very rare in the general population, but 50% of patients also have BAV (Aboulhosn & Child, 2015). With a prevalence of 4.6 per 1000, bicuspid aortic valve is the most prevalent CHD and is much more common in males than females. This defect can range from a severely obstructive unicupsid valve to an almost normal appearance with only partial fusion of two leaflets but is often accompanied by dilation of the aortic root. Coarctation of the aorta occurs when the aorta becomes pinched off in the area around the insertion of the ligamentum arteriosusm likely due to excess ductal tissue, though hemodynamics may also play a role, and is associated with BAV in 40-63% of cases (Aboulhosn & Child, 2015; Hanneman, Newman, & Chan, 2017; Thiene & Frescura,

2010). This defect is relatively common accounting for 4-8% of CHD and approximately 4 in 10,000 livebirths overall and is more prevalent in males (1.5:1) than females (Aboulhosn & Child, 2015; Hanneman et al., 2017). Many patients develop collateral vessels connecting upper body vessels to lower body vessels below the level of the coarctation (Hanneman et al., 2017). A pseudocoarctation can arise when an abnormally long segment of the aortic arch at the level of ductus arteriosus collapses or becomes kinked (Hanneman et al., 2017). HLHS is a severe and rare form of CHD occurring in 0.016-0.036% of all livebirths and accounting for 3.8% of all CHD with up to 70% of cases being male (Connor & Thiagarajan, 2007). This defect is characterized by underdevelopment of the left ventricle, aortic valve, and aortic arch accompanied by mitral stenosis or atresia. Interrupted aortic arch (IAA) is a relatively rare (approximately 2 per 100,000 livebirths) defect characterized by a complete disconnection of flow between the ascending and descending aorta (Reviewed in Hanneman et al., 2017). Connection between the upper and lower body is usually via ductus arteriosus with a large accompanying VSD, though rarely it can be through persistence of the 5th pharyngeal arch artery. IAA can be classified into three types based on the location of the interruption with type A having an interruption distal to the left subclavian artery and being the 2nd most common type occurring in 30-40% of cases. Type B IAA is the most common occurring in 50-60% of cases and has an interruption between the left subclavian and left common carotid arteries. This form is often seen with 22q11.2 microdeletion and accompanying extracardiac anomalies and is associated with BAV, outlet VSD, and aortic stenosis. Accounting for less than 5% of cases, IAA-C is the least common form with the interruption occurring between the brachiocephalic and left common carotid arteries (Hanneman et al., 2017).

Conotruncal defects are those involving malformations of the ventricular outlets and arterial trunk with the most common defects including TOF, PTA, TGA, and DORV (T. R.

Johnson, 2010). Persistent truncus arteriosus accounts for 1.1-2.5% of CHD and occurs when the outlet (aortopulmonary) septum fails to form and the aorta and pulmonary arteries persist as a common arterial trunk with a common truncal valve and large VSD (T. R. Johnson, 2010; Thiene & Frescura, 2010). There are several different types based on the branching pattern, but in most cases the truncal valve is abnormal ranging from unicommisural to five cusps (T. R. Johnson, 2010). It sometimes occurs with IAA-B and 35% of patients also have a right aortic arch. Sometimes a small aortopulmonary window may allow communication between an otherwise normally septated aorta and pulmonary artery (Thiene & Frescura, 2010). Double outlet right ventricle accounts for less than 1.5% of CHD and can be classified based on whether the accompanying VSD is subaortic, subpulmonic, or doubly-committed while the great arteries can be normally related, side-by-side, or malposed (Reviewed in T. R. Johnson, 2010). The most common type of DORV is characterized by a subaortic VSD with side-by-side great arteries and a concern for pulmonary stenosis. The Taussig-Bing form has a subpulmonic VSD, side-by-side great arteries, and often subaortic stenosis with a concurrent coarctation or interruption of the arch. The Remote VSD form is rare, but may occur with complete AVSD or heterotaxy or in the setting of straddling atrioventricular valve and hypoplastic ventricle (T. R. Johnson, 2010). Tetralogy of Fallot is the most common form of cyanotic CHD accounting for 3.5-9% of patients and is defined by the four cardinal features of infundibular VSD, overriding aorta, pulmonary stenosis, and right ventricular hypertrophy (T. R. Johnson, 2010; Thiene & Frescura, 2010). The pulmonary valve may be fully atretic resulting in a ductus-dependent pulmonary circulation (Thiene & Frescura, 2010). TOF is often accompanied by other heart defects including right aortic arch (25%), absent pulmonary valve (2.5%), AVSD, or anomalous coronary artery (T. R. Johnson, 2010; Thiene & Frescura, 2010).

In addition to AVSD, a variety of other defects can affect the atrioventricular valves. In tricuspid atresia, the tricuspid valve is atretic with right-to-left shunting at the atrial level and pulmonary circulation dependent on a restrictive VSD (Thiene & Frescura, 2010). Straddling tricuspid valve arises when the chordae tendinae or a papillary muscle of the tricuspid valve straddle or override the ventricular septum (Dodge-Khatami, Mavroudis, Frost, Jacobs, & Mavroudis, 2014). Ebstein's anomaly of the tricuspid valve is a rare defect affecting one in 10,000 infants and characterized by septal and posterior leaflets that are dysplastic and displaced downward into the right ventricle possibly leading to subvalvular pulmonic stenosis or atrialization of the ventricle (Kwiatkowski et al., 2016; Thiene & Frescura, 2010). The majority of patients have a concurrent ASD and it has been associated with left ventricular noncompaction. Double-orifice tricuspid valve is a very rare defect that can be isolated or occur in conjunction with other forms of CHD (Dodge-Khatami et al., 2014). The defect can involve a simple leaflet fenestration (type 1), fusion of leaflet tissue bridging the valve resulting in two orifices but with normal papillary muscles (type 2), or two orifices each supported by their own annulus (type 3).

A cleft mitral valve is commonly seen in conjunction with AVSD, but can also occur in isolation and is characterized by a cleft in the anterior leaflet often accompanied by abnormal attachment of chordae to the membranous septum (Fishbein & Fishbein, 2019). Parachute mitral valve occurs when chordae from both leaflets attach to a single papillary muscle and can cause stenosis. Finally, double-orifice mitral valve involves an abnormal opening in one of the leaflets with chordae attaching the edge of the defect to the underlying papillary muscle (Fishbein & Fishbein, 2019).

Heterotaxy or situs ambiguous refers to the randomized positioning of organs with respect to the right/left axis and is often accompanied by complex CHD. In right isomerism, both atrial

appendages have a right-sided morphology and this is often accompanied by supra- or infradiaphragmatic total anomalous pulmonary venous return (TAPVR) which occurs when the pulmonary veins insert into the vena cava instead of the left atrium. Left isomerism is characterized by both atrial appendages having a left-sided morphology and is often accompanied by an interrupted inferior vena cava with azygous continuation (Thiene & Frescura, 2010). A common atrium due to absence of the primum, secundum, and atrioventricular septa is often seen (Geva et al., 2014; van Praagh, 2006). Discordant atrioventricular connection is often accompanied by a univentricular heart with one ventricle being fully developed and functional and the other being rudimentary or absent (Thiene & Frescura, 2010). Common atrioventricular valve is seen in 69% of patients with asplenia (absent left-sided structures) with normal atrioventricular valves being rare (van Praagh, 2006). TGA accounts for 5-7% of all CHD with 60-70% of cases being male and can be accompanied by VSD, coronary artery anomalies, or patent ductus arteriosus (PDA) (T. R. Johnson, 2010). Dextro-TGA is characterized by concordant atrioventricular connections but discordant ventriculoarterial connection while levo-TGA (congenitally corrected TGA) occurs when both the atrioventricular and ventriculoarterial connections are discordant (Thiene & Frescura, 2010). Asplenia is associated with conotruncal and atrioventricular canal defects, bilateral superior vena cava to the ipsilateral atrium, bilateral venous connections from the lower body and TAPVR (van Praagh, 2006). Pulmonary outflow obstruction is common while DORV/TGA are nearly universal. Polysplenia (absent right-sided structures) is associated with an interrupted inferior vena cava with azygous or hemiazygous continuation and occasional ipsilateral drainage of the pulmonary veins (van Praagh, 2006). Pulmonary outflow obstruction is less common than in asplenia and DORV/TGA occur in less than 50% of cases. There is a high incidence of underdeveloped left ventricle in all heterotaxy

patients. Dextrocardia is also frequent and is thought to be related to an L-looped heart rather than atrial situs inversus (van Praagh, 2006).

A variety of other defects can also occur. There can be a univentricular atrioventricular connection due to tricuspid or mitral atresia or a double-inlet ventricle resulting from misalignment of the atrioventricular valves (Thiene & Frescura, 2010). There can be atresia of the aorta or pulmonary artery or the ductus arteriosus can remain open after birth. Cor triatriatum sinister occurs when an abnormal membrane separates the left atrium into two chambers (Thiene & Frescura, 2010). Coronary artery anomalies such as anomalous origin or high takeoff of a coronary artery can lead to an increased risk of myocardial infarction and sudden death. Wolf-Parkinson-White syndrome occurs due to an abnormal muscle bundle creating a reentrant electrical pathway leading to arrhythmia (Thiene & Frescura, 2010). Finally, there can be a variety of abnormal aortic arch branching patterns such as right aortic arch, double aortic arch, and anomalous origin of the subclavian artery (Hanneman et al., 2017).

1.2 Heart Development

As the most asymmetric organ in the body, heart development is highly complex involving extensive structural rearrangements and tissue remodeling under the control of numerous genes and signaling pathways. This complex process has been extensively reviewed (Bruneau, 2013; Campione & Franco, 2016; Gittenberger-de Groot, Bartelings, Deruiter, & Poelmann, 2005; Hanneman et al., 2017; Ramsdell, 2005; Rochais, Mesbah, & Kelly, 2009; Sylva, van den Hoff, & Moorman, 2014). The first progenitors of cardiac mesoderm become specified around the time of gastrulation in response to high levels of the transforming growth factor (TGF)-β superfamily

transcription factor, *Nodal* (Bruneau, 2013; Ramsdell, 2005). These mesoderm-derived precursors comprising the first heart field (FHF) then unite anteriorly at the midline forming the cardiac crescent which, in response to ectoderm- and endoderm-derived signals that modify bone morphogenetic protein (BMP), fibroblast growth factor (FGF) and wingless-insert (WNT) signaling, begin to differentiate into cardiomyocytes (Sylva et al., 2014). Other molecules that play a role in specification and expansion of the early cardiac mesoderm include the T-box transcription factor *Eomes* which leads to *Mesp1* activation, *Flk1*, *PDGFRA*, Wnt/β-catenin signaling, and BMP signaling (Bruneau, 2013). Newly committed cardiomyocytes then being expressing cardiac-specific genes including *NKX2.5* and *GATA4* (Gittenberger-de Groot et al., 2005)

Embryonic folding brings the lateral edges of the cardiac crescent together at the midline where they fuse to form the linear heart tube with a cranial arterial pole and a caudal venous pole, a process that is dependent on *Mesp1*, *Gata4*, *Furin*, and *Bmp2* (Bruneau, 2013; Sylva et al., 2014). The heart tube then grows through the addition of cells from an adjacent rapidly proliferating pool of cells around the venous pole, the second heart field (SHF), that migrate into the heart and differentiate into cardiomyocytes. Ultimately the FHF will contribute primarily to the left ventricle and parts of the atria in the mature heart while SHF-derived cells will contribute to the mature outflow tract, right ventricle, and venous pole. The SHF as a whole can be molecularly identified by the expression of *Isl2* while its anterior component which gives rise to the outflow tract and right ventricle is identified by the expression of *Fgf10*, an enhancer of *Mef2c* and the sinus venosus is marked by *Tbx18* expression (Bruneau, 2013). Hedgehog signaling is required for the survival and deployment of SHF cells while their recruitment into the linear heart tube and subsequent differentiation into cardiomyocytes is regulated by BMP and FGF signaling (Rochais et al., 2009).

Ultimately, many different transcription factors play a key role in early heart development including *Nkx2.5*, GATA factors, T-box transcription factors, and Hand factors (Bruneau, 2013). An additional layer of regulation is provided by epigenetic factors including chromatin remodeling complexes such as *Baf60c* and the *Brg1/Brm*-associated factor (BAF) complex that interact with transcription factors, histone acetyltransferases, histone deacetylases, and histone methyltransferases such as the polycomb repressive complex 2 (PRC2) (Bruneau, 2013).

The symmetrical heart tube must then be reorganized into the mature four-chambered heart divided by septa and valves. In vertebrate development symmetry is first broken by the generation of leftward fluid flow by motile cilia at the embryonic node which activates the *Nodal/Pitx2/Lefty* signaling cascade leading to asymmetric gene expression (Sylva et al., 2014). In the developing heart, the first molecular signs of asymmetry can be identified in the FHF shortly after gastrulation (Ramsdell, 2005). The first morphological breaking of symmetry occurs with the rightward looping of the ventricles, a process that has been shown to be independent of *Pitx2* but that does require *Cited2* (Sylva et al., 2014). Though not directly involved in looping, expression of the *Pitx2* isoform, *Pitx2c* identifies the sinus venosus, left side of the atrioventricular canal, the ventral portion of the ventricles, and the left-ventral portion of the outflow tract (Campione & Franco, 2016).

Looping results in the formation of an S-shaped heart with an inner curvature and an outer curvature from which the chambers balloon (Sylva et al., 2014). This process is thought to be driven in part by a combination of Nodal signaling in the left myocardium coupled with BMP signaling on the right both interacting with the transcription factor *FOXH1* to coordinate differential cell migration (Campione & Franco, 2016). The atrial chambers form from the dorsolateral side of the heart tube with *Pitx2* playing a role in specifying left vs. right atrium as

well as differentiation of pulmonary vein myocardium (Bruneau, 2013; Sylva et al., 2014). The process of ventricle ballooning from the outer curvature involves a combination of the reinitiation of proliferation and the initiation of chamber-specific gene expression programs (Sylva et al., 2014). Determination of left ventricular identity and the interventricular boundary are dependent on *Tbx5* while *HAND1* and *HAND2* are required for proper formation of the left and right ventricles respectively (C. J. Lin, Lin, Chen, Zhou, & Chang, 2012; Sylva et al., 2014). The outflow tract must undergo elongation and rotation to ensure its proper alignment, a process in which *Pitx2* and *Wnt11* play a role (Campione & Franco, 2016).

T-box transcription factors play an important role in chamber formation and identity. *Tbx5* and *Tbx20* are expressed in most parts of the heart while *Tbx2* and *Tbx3* act to maintain the primary myocardium phenotype of the venous pole, atrioventricular canal, inflow tract, and outflow tract allowing them to ultimately become non-working myocardium while *Tbx1* is expressed at the venous pole (Sylva et al., 2014). In addition to chambers, the heart must also form transition zones that will eventually contribute to the formation of valves, septa, and the conduction system. Looping brings these transition zones together along the inner curvature while the chambers balloon from the outer curvature (Gittenberger-de Groot et al., 2005). Finally, maturation of the chamber myocardium must occur. This involves a process of thickening and trabeculation of the primitive myocardium followed by remodeling and compaction of the trabeculae to form the compact myocardium of the mature heart (Sylva et al., 2014). This process is coordinated through the action of numerous signaling pathways including Notch, Neuregulin/ErbB, EphrinB2/B4, BMP, FGF, semaphorins, retinoic acid, endothelin, and extracellular matrix signaling (Bruneau, 2013).

The primitive heart must then be divided into four chambers through formation of interatrial, atrioventricular, and interventricular septa while the primitive common outflow tract must be divided into aorta and pulmonary arteries through formation of the outflow and aortopulmonary septa. Atrial septation is a complex process and occurs in two phases. The precursors of the atrial septum are established early in a process requiring *Tbx5* (Bruneau, 2013). The septum primum originates right of the midline at the cranio-dorsal wall of the common atrium and grows toward the atrioventricular canal (Sylva et al., 2014). Eventually, the continued growth of the septum primum closes off the foramen primum, however openings form in the septum primum and eventually coalesce to form the ostium secundum which remains open until birth allowing continued interatrial communication. The septum secundum is formed from an infolding of the muscular right atrial wall. Finally, after birth changes in intraatrial pressure presses the primum and secundum septa together closing the ostium secundum and completing the process of atrial septation (Sylva et al., 2014).

The muscular ventricular septum forms at the region of the heart tube between the ballooning ventricles and grows mainly through apposition by the addition of cells from the adjacent left ventricular free wall (Sylva et al., 2014). Several different mechanisms have been proposed for the formation of the muscular ventricular septum; the coalescence of a portion of the ventricular wall interposed between the ballooning ventricles or the active expansion of a "primitive interventricular septum" toward the atrioventricular canal (Penny & Vick, 2011). The membranous portion of the ventricular septum forms from remnants of the fused atrioventricular and outflow cushions which then fuse with the muscular septum to complete ventricular septation (Krishnan et al., 2014; Sylva et al., 2014). *Tbx5, Gata4, Nkx2-5, Sall4*, and Hand transcription factors play an important role in ventricular septation (Bruneau, 2013)

The process of atrioventricular septation begins with the formation of an endocardial cushion at the atrioventricular junction at the looped heart stage (Krishnan et al., 2014). Endocardial cells first undergo epithelial-to-mesenchymal transition (EMT) and then migrate into the cardiac jelly in part through signaling by *Bmp2*, *Tbx2*, *Notch*, and *Vegf* as well as the activation of *Snail1/2* through TGF-β signaling that is essential to confer invasive properties on the migrating cells (Bruneau, 2013; Calkoen et al., 2016). This results in the formation of the inferior, superior, and two lateral atrioventricular cushions. The muscular portion of the septum primum grows toward the atrioventricular canal where its mesenchymal cap ultimately fuses with the superior and inferior cushions followed by their mutual fusion with the dorsal mesenchymal protrusion at the base of the atrial septum (Calkoen et al., 2016). Atrioventricular septation is completed following delamination of the tricuspid valve cusps when the membranous portion of the ventricular septum forms. The inferior and superior cushions ultimately contribute to both the tricuspid and mitral valves while the right lateral cushion contributes to the tricuspid valve and the left to the mitral valve (Calkoen et al., 2016).

Outflow septation begins with the formation of mesenchymal ridges that grow and spiral in a clockwise direction to separate the distal common outflow tract into aorta and pulmonary artery (Krishnan et al., 2014). Septation of the proximal outflow tract begins with the formation of endocardium-derived conal cushions which then fuse to form the conus septum separating the ventricular outflow tracts (C. J. Lin et al., 2012). Cardiac neural crest cells play a critical role in outflow septation and valve formation. These cells originate in rhombomeres 6-8 in the hindbrain and first migrate to a region between pharyngeal arch arteries four and six where they form a mass of condensed mesenchyme and then migrate into and populate the proximal outflow cushions (Gittenberger-de Groot et al., 2005; C. J. Lin et al., 2012). SHF-derived outflow tract myocardium

secretes chemotactic molecules such as *Sema3c* that attract migrating neural crest cells into the outflow cushions while Wnt/β-catenin, *Pitx2*, Notch, TGF, BMP, and Hedgehog pathways also play a role in the migration and survival of neural crest cells and outflow septation (C. J. Lin et al., 2012; Rochais et al., 2009). VEGF signaling plays a more important role in outflow septation than in atrioventricular septation while FGF signaling under the control of *Tbx1* is essential for initiation of a signaling loop that ultimately reaches the adjacent neural crest and endothelial cells (Bruneau, 2013).

Concurrently with septation of the outflow tract, the paired pharyngeal arch arteries are extensively remodeled to form the main pulmonary arteries, aortic arch, and its branches. Formation of the pharyngeal arch arteries begins with the fusion of ventral segments of the primitive paired aortae to form the aortic sac while fusion of the dorsal segments gives rise to the descending aorta (Hanneman et al., 2017). Six pairs of pharyngeal arch arteries develop between the ventral and dorsal aortae and then regress or remodel in a cranial to caudal direction. The first and second arch arteries largely regress, but remnants remain as parts of the maxillary, stapedial, and hyoid arteries. The common carotid artery and the proximal portion of the internal carotid artery form from the third arch. The right fourth arch forms the right subclavian artery while the left fourth arch contributes to the definitive aortic arch. The fifth arch arteries either do not develop or partially develop and then regress. Finally, the right sixth arch artery gives rise to the right pulmonary artery while the left sixth arch gives rise to the main and left pulmonary arteries as well as the ductus arteriosus (Hanneman et al., 2017).

Valve formation begins with the remodeling of cardiac jelly at the venous pole, atrioventricular junction, and outflow tract. Under the regulation of BMPs, TGF-β, and Notch signaling, the overlying endocardium undergoes EMT and migrates into the cardiac jelly (Sylva et

al., 2014). The atrioventricular valves form by detachment of the cushions from the endocardium and remodeling into valve leaflets while the semilunar valves form from intercalated ridges within the outflow tract with a significant contribution from migrating neural crest cells (Gittenberger-de Groot et al., 2005; Sylva et al., 2014). Many genes have key roles in valve formation including extracellular matrix proteins, the connective tissue transcription factor *Sox9*, and the Notch and BMP signaling pathways (Sylva et al., 2014).

1.3 The Genetic Causes of Congenital Heart Disease

Congenital heart disease can occur in isolation or as part of a syndromic condition and it can also occur sporadically or segregate within families. A growing number of genes have been identified as causing or contributing to the etiology of CHD through a variety of different study types including whole exome/genome sequencing, family and linkage studies, genome wide association studies (GWAS), and mouse mutagenesis screens and these have been extensively reviewed (Aburawi, Aburawi, Bagnall, & Bhuiyan, 2015; Andersen, Troelsen Kde, & Larsen, 2014; Azhar & Ware, 2016; Chaix, Andelfinger, & Khairy, 2016; Fahed, Gelb, Seidman, & Seidman, 2013; LaHaye, Lincoln, & Garg, 2014; Lalani & Belmont, 2014; Lebo & Baxter, 2014; Muntean, Toganel, & Benedek, 2017; Pierpont et al., 2007; Prendiville, Jay, & Pu, 2014; Williams, Carson, & Lo, 2019). The following sections will review just some of the primary literature to provide an idea into the scope of research being conducted into the genetic causes of CHD.

1.3.1 Isolated CHD

Atrial septal defects: A candidate gene case-control study identified an extremely rare heterozygous Arg47Trp missense mutation in *CCN1* associated with severe ASD in a patient while an earlier study showed that *Ccn1*-null mice had severe AVSD while haploinsufficiency resulted in delayed ventricular septation and a primum ASD in 20% of mice (Mo & Lau, 2006; Perrot et al., 2015). *CCN1* (Cellular communication network factor 1) encodes a matricellular protein that regulates cell adhesion, migration, proliferation, and survival. Sequencing in 150 probands with isolated ASD identified three mutations in a conserved region of the *NEXN* gene encoding the Nexin F actin-binding protein that were not found in 500 healthy controls (F. Yang et al., 2014). All three of these mutations were found to inhibit *GATA4* expression.

Ventricular septal defects: Sequencing of the coding region of the extracellular matrix protein *HAS2* (Hyaluronan synthase 2) gene in 100 patients with non-syndromic VSD identified a Glu499Val mutation in one patient that was not observed in 200 controls and was shown to diminish enzyme activity (Zhu et al., 2014). An earlier study showed that *Has2*$^{-/-}$ mice have absent endocardial cushions, absent trabeculae, and hypoplasia of the right ventricle and outflow tract (Camenisch et al., 2000). A case-control mutational analysis of 400 patients with isolated VSD plus 400 healthy controls identified two novel heterozygous mutations in highly conserved residues of the *HOMEZ* (Homeodomain leucine zipper-encoding) gene (Xuan et al., 2013). In a study of 467 patients with conotruncal or related CHD, missense mutations in neurotrophic receptor tyrosine kinase 3 (*NTRK3*) were identified in four patients with either conoventricular, posterior malalignment, or conoseptal hypoplasia type VSD (Werner et al., 2014). These variants were predicted to be deleterious and were not present in ethnically matched controls. However, an earlier study had shown decreased *NTRK3* expression and reduced protein levels in cardiac

biopsy samples from patients with TOF compared to isolated VSD or controls (Kong, Liu, & Lu, 2009). An earlier study found that *Ntrk3* null mice had a variety of valve, septal, and outflow tract defects attributed to premature fate restriction of migrating cardiac neural crest cells (Tessarollo et al., 1997). Variants in *FOXH1* (Forkhead factor H1) have been associated with VSD in a Chinese population (B. Wang et al., 2010).

Atrioventricular septal defects: A study of 13 parent-offspring trios as well as 112 unrelated probands identified an excess of rare missense variants in *NR2F2* (Nuclear Receptor Subfamily 2 Group F Member 2) in patients with AVSD (Al Turki et al., 2014). *NR2F2* is a member of the steroid/thyroid hormone nuclear receptor superfamily of transcription factors and was found to be expressed in the atria, coronary vessels, and aorta of fetal human hearts. Of the five variants detected in patients, two were found to have been inherited from unaffected parents suggestive of either incomplete penetrance or that the variants were possibly benign, though all five mutations appeared to have an effect on the transcriptional activator function of *NR2F2* while leaving its repressor function intact. *NR2F2* mutations were also identified in patients with TOF, aortic stenosis with VSD, coarctation of the aorta, and HLHS (Al Turki et al., 2014). In an unrelated study, myocardiocyte-specific knockout of *Nr2f2* in mice was found to result in ventricularized atria (S. P. Wu et al., 2013). A whole exome sequencing (WES) study of 81 unrelated probands with AVSD found significant enrichment for rare and damaging variants in cases relative to controls in a set of 112 candidate genes (D'Alessandro et al., 2016). This enrichment was found to be specific to AVSD probands compared to a separate cohort with TOF and included the *NIPBL*, *CHD7*, *CEP152*, *BMPR1A*, *ZFPM2*, and *MDM4* genes. The authors were then able to confirm these findings in a replication cohort of 81 AVSD probands. Several mouse mutagenesis studies have also identified causative genes for AVSD. An ethylnitrosourea

(ENU) mutagenesis screen identified mutations in dynein axonemal heavy chain 11 (*Dnah11*) that caused AVSD with heterotaxy and mutations in MKS transition zone complex subunit 1 (*Mks1*) that caused AVSD without heterotaxy (Burnicka-Turek et al., 2016). Mice with null mutations in the *Klf2* (kinesin family member 2) gene were found to have disorganized and hypocellular atrioventricular cushions possibly due to defective EMT (Chiplunkar et al., 2013).

Conotruncal defects: *GATA4*, *NKX2.5*, *JAG1*, *FOXC2*, *TBX5*, and *TBX1* are well-characterized causes of TOF (Reviewed in Morgenthau & Frishman, 2018). Sequencing of 231 unrelated probands with TOF or related defects plus 11 patients from families with TOF including eight trios found that loss-of-function variants in *FLT4* (Fms Related Receptor Tyrosine Kinase 4) and *KDR* (Kinase insert domain receptor) are significant contributors to the genetic cause of TOF (Reuter et al., 2018). Furthermore, the authors noted that the presence of variants in VEGF signaling pathway genes were associated with an increased prevalence of absent pulmonary valve. Sequencing of the entire coding sequence and splice junctions of the *BVES* (Blood vessel epicardial substance) gene in 114 unrelated Chinese patients with TOF and 400 controls identified four novel mutations in cases that were not present in controls (M. Wu et al., 2013). A homozygous Arg1299Cys variant in *PLXND1* has been associated with truncus arteriosus (Ta-Shma et al., 2013). Variants in *MEF2C*, *NKX2.5*, *GATA4*, and *GATA6* were identified in individuals with conotruncal heart defects with functional assays showing significant changes in transcriptional or synergistic activity in the proteins harboring these variants (Kodo et al., 2012). Furthermore, the Ala103Val variant found in *MEF2C* (Myocyte enhancer factor 2C) was shown to perturb early cardiogenesis when overexpressed in a zebrafish system. A GWAS case-control study identified 52 significant genotypes including 49 intronic to the *MGAT4C* gene in conotruncal heart defect patients (Agopian et al., 2017).

Left ventricular outflow tract defects: A noncoding variant 151 kilobases from the *GATA4* gene was found to be associated with BAV with a Ser377Gly missense variant trending toward significance in a GWAS of 466 BAV cases plus 4,660 matched controls and replicated in up to 1,326 cases and 8,103 controls (B. Yang et al., 2017). Next-generation sequencing (NGS) of 63 patients with BAV requiring aortic root replacement yielded rare, potentially pathogenic variants in 19 patients with *NOTCH1* variants being the most common and variants also identified in *AXIN1*, *NOS3*, *ELN*, *FBN1*, and *FN1* (Girdauskas et al., 2017). Targeted NGS of 97 candidate genes in 78 unrelated patients with BAV with or without coarctation of the aorta identified 42 rare nonsynonymous exonic variants in 35 of 97 genes (Bonachea et al., 2014). In silico analysis classified 33 of these variants as disease-causing of which 31 variants in 16 patients were confirmed by Sanger sequencing while pathway analysis suggested a role for WNT signaling. A GWAS of an Icelandic population found a rare mutation in myosin heavy chain 6 (*MYH6*) associated with coarctation of the aorta (Bjornsson et al., 2018) . Missense, loss-of-function, and intronic variants in *NOTCH1* have been shown to confer different risks for LVOTO (Helle et al., 2018). Additionally, variants in *NOTCH1* as well as *FOXC2* and *FOXL1* have been found in patients with HLHS (Iascone et al., 2012). Variants in *MIB1* (Mindbomb E3 Ubiquitin Protein Ligase 1), an activator of Notch signaling have been shown to cause CHD with the Thr312Lysfs*55 and Trp271Gly variants shown to significantly reduce *MIB1* function resulting in reduced *JAG1* ubiquitination and induction of Notch signaling (B. Li et al., 2018). A GWAS of multiple cohorts with LVOTO or conotruncal defects identified multiple single nucleotide polymorphisms (SNPs) with a suggestive association with LVOTO including inherited SNPs in in *LRP1B* and *MGATA4* along with maternal SNPs in in *SLC38A3*, *HIVEP3* and *PKD1L2* (Agopian et al., 2017).

Other CHD candidate genes: In a study of 26 patients with TAPVR, one patient was found to have a translocation that disrupted the *ANKRD1* (Ankyrin repeat domain 1) gene and another patient was found with a missense mutation in the same gene (Cinquetti et al., 2008). Mouse in situ hybridization analysis showed this gene to be highly expressed in developing pulmonary veins while functional analysis suggests that the missense mutation ultimately enhances its transcriptional repression activity on atrial natriutretic factor (*ANF*). Sequencing of *DLC1*, a GTPase activating protein highly expressed in the heart, in 151 patients with sporadic CHD revealed 13 nonsynonymous rare variants of which nine were predicted to be damaging (B. Lin et al., 2014). *Dlc1*$^{-/-}$ mice have incomplete and abnormal chamber morphogenesis as well as brain defects while mice homozygous for a gene-trapped *Dlc1*$^{gt/gt}$ allele had enlarged, thin-walled hearts, abnormal yolk sac and placental vasculature, and neural tube defects (Durkin et al., 2005; Sabbir et al., 2010). Variants in the *IRX1* gene, which is important for heart and limb development, have been detected in patients with a variety of CHD but were found to be absent in controls (C. Guo et al., 2017). Iroquois homeobox transcription factors in general have been shown to play a role in the establishment of ventricular chamber properties and of the conduction system (Reviewed in Kim, Rosen, Bruneau, Hui, & Backx, 2012). A *TPM1* (Tropomyosin 1) mutation was identified in a child with severe Ebstein's anomaly and left ventricular noncompaction (Kelle, Bentley, Rohena, Cabalka, & Olson, 2016). Additionally, screening of 380 patients with a variety of CHD identified four variants in four patients with different types of CHD (England et al., 2017). The latter group showed that morpholino knockdown of *tpm1* in zebrafish resulted in defects in looping, atrial septation, and trabeculation while also affecting sarcomere assembly. Variants in *UGDH*, UDP glucose 6 dehydrogenase, were found in patients with cardiac valve defects which could be recapitulated by morpholino knockdown in zebrafish (Hyde et al., 2012). Furthermore, the

morpholino-induced defects could not be rescued by injection of human *UGDH* transcripts carrying the identified mutations. Variants in *MIB1*, an activator of Notch signaling, cause CHD through a decreased level of *JAG1* ubiquitination and reduced Notch signaling induction (B. Li et al., 2018).

Pathways contributing to CHD: A number of signaling pathways have been shown to play key roles in the etiology of CHD. Functionally abnormal gene products have been identified in multiple Nodal pathway genes including *FOXH1*, *CFC1, SMAD2*, *NODAL, GDF1*, and *TDGF1* that are associated with CHD, laterality defects, and holoprosencephaly (Roessler et al., 2008). Abnormal Nodal pathway gene products have more commonly been detected in patients with a small spectrum of conotruncal defects and less frequently in isolated laterality defect and holoprosenccephaly patients.

Wnt/β-catenin signaling also plays a role. β-catenin has been shown to form a signaling gradient that is positively correlated with proliferation of ventricular wall cardiomyocytes with *APC*, a negative regulator of *Wnt*, having the reverse distribution (Ye et al., 2015). Cardiomyocyte-specific deletion of *Apc* was shown to result in activation of canonical Wnt/β-catenin signaling leading to ventricular hypertrophy whereas cardiac deletion of *β-catenin* lead to ventricular hypoplasia. Mutations in *Bcl9/Bcl9l* and *Pygo1/2*, factors engaged in assembly of a Wnt-specific enhanciosome, result in valve, septal, and other cardiac defects (Cantu et al., 2018). Mutant mice broadly maintain canonical Wnt expression but show dysregulated expression of key pathway regulators. *Dkk1/2* interact with *Lrp5/6* to modulate canonical Wnt signaling during development and are expressed in the developing heart (Phillips, Mukhopadhyay, Poscablo, & Westphal, 2011). *Dkk1/2* act to inhibit Wnt signaling to regulate myocardial proliferation and,

though neither single null mutant shows an abnormal phenotype, double null mice show myocardial and epicardial hyperplasia during the early stages of development and VSD later on.

Sonic hedgehog (Shh) signaling has been shown to be required for normal cardiac septation (Goddeeris et al., 2008). *Mef2c-AHF-Cre; Smo$^{flox/-}$* mutants were shown to survive to term but with failure of outflow tract septation and other intracardiac defects similar to those seen with *Shh* null mutants. Loss of Shh signaling was also shown to result in aberrant differentiation and migration of the dorsal mesocardium (Goddeeris et al., 2008). Additionally, *Nkx2.5$^{Cre/+}$; Shh$^{flox/-}$* double mutants were shown to have an unseptated outflow tract as well as a reduced right ventricle, complete AVSD, and incompletely penetrant arch artery defects (Goddeeris, Schwartz, Klingensmith, & Meyers, 2007). Finally, loss of primary cilia in mouse *Ift88$^{cbs/cbs}$* mutant hearts was shown to result in a loss of downstream Shh signaling and a loss of *Isl1* expression in the SHF and reduced or absent BMP signaling in the outflow tract and atrioventricular cushions resulting in atrial and ventricular septal defects and severe outflow tract defects (Willaredt, Gorgas, Gardner, & Tucker, 2012).

1.3.2 Syndromic CHD

Congenital heart disease is a major or minor component of many syndromic conditions of which Down syndrome (AVSD) and DiGeorge syndrome (Conotruncal defects) are probably the most well-known. Down syndrome is the most common chromosomal anomaly and the condition most frequently associated with CHD which is present in about 50% of cases (Reviewed in Versacci, Di Carlo, Digilio, & Marino, 2018). In a study of patients with complete trisomy 21 with or without AVSD, 26 candidate genes were sequenced based on their potential association with AVSD and potentially damaging variants in VEGF-A pathway genes including *COL6A1*,

COL6A2, *CRELD1*, *FBLN2*, *FRZB*, and *GATA5* were found in 20% of cases with a complete balanced AVSD compared to only 3% for patients with structurally normal hearts (Ackerman et al., 2012). Of these genes, only *CRELD1* has previously been associated with AVSD. Furthermore, significant missense variants in *CITED2*, *CTGF*, *TBX20*, and *VTN* were found only in controls and could represent protective alleles. Another study found that no chromosome 21 copy number variations (CNVs) were associated with the risk of AVSD in Down syndrome nor were any individual genes intersected by these CNVs (Rambo-Martin et al., 2018). Additionally, Caucasian cases (Down syndrome with AVSD) had more genes intersected by rare duplications than controls (Down syndrome without AVSD), while in African Americans cases had more bases covered by rare deletions than controls. In a mouse study, overexpression of *Col6a2* and *Dscam* together but not individually resulted in 50% mortality with severe defects including ASD and cardiac hypertrophy (Grossman et al., 2011).

DiGeorge syndrome has a prevalence of one in 3,000-6,000 live births and is the second most common cause of CHD accounting for 52% of patients with IAA type B, 34% of patients with truncus arteriosus, 16% of patients with TOF, and 5-10% of patients with VSD (Reviewed in McDonald-McGinn et al., 2015). Historically DiGeorge syndrome has been lumped together with the phenotypically very similar 22q11.2 deletion syndrome and velocardiofacial syndrome, however more recently it has been used to refer only to those cases with a single gene mutation within the *TBX1* (T-box transcription factor 1) gene. The typical 3 Mb deletion and the less common nested 1.5 and 2 Mb deletions occur within a group of low-copy repeats and arise from non-allelic homologous recombination (McDonald-McGinn et al., 2015)

Though mutations in *TBX1* have long been known to be the major cause of the DiGeorge/22q11.2 deletion syndrome phenotype (Hisato Yagi et al., 2003), more recent work has

largely focused on investigating if other genes within the deletion can produce or modify the phenotype particularly in patients with smaller deletions that do not involve *TBX1*. Copy number analysis of DiGeorge patients without the typical deletion who have normal *TBX1* expression revealed small deletions or duplications in the proximal end of the critical region containing the *PRODH* and *DGCR6* genes (Gao et al., 2015). Attenuation of *Dgcr6* expression in chick had previously been shown to result in cardiovascular system defects similar to those seen in DiGeorge syndrome as well as altering the expression profile of *Tbx1* (Hierck, Molin, Boot, Poelmann, & Gittenberger-de Groot, 2004). Sequencing of the remaining *TBX1* allele in 360 patients with DiGeorge/Velocardiofacial syndrome found an increased number of very rare SNPs in patients with normal hearts compared to patients with CHD suggesting that these SNPs may have protective effects (T. Guo et al., 2011).

A variety of mouse mutants that model or recapitulate the cardiovascular phenotype in DiGeorge syndrome have also been identified. Inactivation of *Dgcr8* in mouse cardiac neural crest cells produces a wide spectrum of malformations including PTA and VSD (Chapnik, Sasson, Blelloch, & Hornstein, 2012). A significant proportion of these *Dgcr8*-null cells were found to undergo apoptosis resulting in a progenitor pool insufficient for outflow tract remodeling. Loss of *Tbx1* and its transcriptional target *Wnt5a* was found to cause severe hypoplasia of structures dependent on the SHF (L. Chen et al., 2012). *Tbx1* was shown to occupy a T-box binding element within *Wnt5a* and to interact with *Baf60a/Smarcd1* subunit of the BAF chromatin remodeling complex as well as *Setd7* histone H3K4 monomethyltransferase. Mice with heterozygous deletion of *Hic2*, a transcription factor lost in most distal and some typical DiGeorge syndrome deletions, have a partially penetrant late lethality with one-third having overriding aorta and VSD (Dykes et al., 2014). Conditional targeting indicated that this gene is required in cardiac progenitors

expressing *Nkx2.5* and *Mesp2*. Compound heterozygous mutations of *Crkl* and *Tbx1* result in increased penetrance of DiGeroge-like phenotypes compared to either single heterozygote (Guris, Duester, Papaioannou, & Imamoto, 2006). These genes were shown to have a dose-dependent function in pharyngeal arch segmentation and patterning as well as a role in local regulation of retinoic acid metabolism. Reducing the dosage of *β-catenin* was shown to significantly reduce the outflow tract anomalies in conditional *Tbx1*-null mice (Racedo et al., 2017). Tri- or tetrallelic loss of *Spry1/Spry2* results in pharyngeal arch artery defects and other phenotypes reminiscent of DiGeorge syndrome (Simrick et al., 2012). The severity of the defects was significantly exacerbated by heterozygous loss of *Tbx1* which was also correlated with an increase in receptor tyrosine kinase signaling. Finally, mutations in *Six1* and *Eya1*, both of which genetically interact with *Fgf8* and *Tbx1*, were found to recapitulate most features of human 22q11.2 deletion syndrome including outflow tract and aortic arch malformations (C. Guo et al., 2011).

The RASopathies are a collection of syndromes with overlapping features that include cardiac malformations, craniofacial dysmorphism, cutaneous, musculoskeletal, and ocular anomalies, neurocognitive impairment, and a predisposition to certain cancers (Reviewed in Rauen, 2013). These disorders are characterized by mutations in genes within the RAS/MAPK pathway which has a role in regulating the cell cycle and cell growth, differentiation, and senescence. Noonan syndrome is the most common of the RASopathies with a prevalence of one in 1,000-2,000 newborns with approximately 80% having CHD, particularly pulmonary valve stenosis (Rauen, 2013; Romano et al., 2010). Mutations in the *PTPN11* (Protein Tyrosine Phosphatase Non-Receptor Type 11) gene account for about 50% of all cases (Rauen, 2013). More recently, a Gly231Val variant in *MRAS* was identified in a patient with Noonan syndrome and cardiac hypertrophy (Higgins et al., 2017). A sequencing study in several families with Noonan-like

syndrome including one patient with pulmonary valve stenosis identified mutations in *A2ML1* (Alpha-2-Macroglobulin Like 1) (Vissers et al., 2015). Characterization of the identified mutations in zebrafish produced multiple Noonan-like features including heart defects. In a mouse model of Noonan syndrome, creation of an allelic series of *Ptpn11* knock-in mice showed that Noonan syndrome-related CHD including valvular hyperplasia, ASD, VSD, DORV, and myocardial thinning resulted from abnormal expression in endocardium but not myocardium or neural crest (Araki et al., 2009). Penetrance was shown to be affected by the mouse strain background and the specific *Ptpn11* allele and the pathogenic mechanism was shown to involve elongation of the time during which endocardial cells undergo EMT.

Many other causes of syndromic CHD have been identified. Gain-of-function mutations in *ABL1* (ABL Proto-Oncogene 1, Non-Receptor Tyrosine Kinase) that increase phosphorylation activity were found to cosegregate with an autosomal dominant syndrome consisting of CHD including ASD, VSD and aortic dilation, and skeletal malformations (X. Wang et al., 2017). An earlier *Abl1*-null mouse model revealed severe cardiac hypertrophy due to increased cardiomyocyte proliferation indicating that this gene does play a role in heart development (Qiu, Cang, & Goff, 2010). A recessive syndrome consisting of joint dislocation and CHD including BAV and aortic root dilation was attributed to variants in Beta-1,3-Glucuronyltransferase 3 (*B3GAT3)* which is expressed in heart, aorta, and bone (Baasanjav et al., 2011). Variants in the collagen proteins *COL1A1* and *COL1A2* are associated with a form of osteogenesis imperfecta that includes valvular heart disease (Vandersteen et al., 2014). Mutations in *IRX5*, which modulates progenitor migration in branchial arches and directly interacts with *GATA3* and *TRP51*, have been implicated in a syndrome affecting blood, brain, face, heart, and gonad (Bonnard et al., 2012). Mutations in *SMAD2*, a downstream effector in the TGF-β signaling pathway have been linked to

phenotypes ranging from complex CHD with heterotaxy and other extracardiac findings to aortic aneurysm with signs of connective tissue disease (Granadillo et al., 2018). The VATER/VACTERL association of Vertebral Anomalies, Cardiac defects, Tracheo-Esophageal fistula, Renal anomalies, and Limb defects has been linked to mutations in *ZIC3* (Gly17Cys) and *FOXF1* (Gly220Cys) (Hilger et al., 2015). Homozygous missense mutations in *CHST3* (carbohydrate sulfotransferase 3) are associated with Omani-type spondyloepiphyseal dysplasia with valvular heart disease (P. Srivastava, Pandey, Agarwal, Mandal, & Phadke, 2017; Tuysuz et al., 2009).

1.3.3 Familial Occurrence of CHD

Though most CHD is isolated, there is growing evidence that at least some forms can be inherited or run in families. Additionally, as more children with severe CHD are surviving into adulthood and having children of their own, it is important to know how often these defects are inherited by their children. One study found that 727 individuals surviving major CHD produced 393 live offspring, 71 miscarriages, and five termination-of-pregnancy (Burn et al., 1998). A total of sixteen liveborn offspring had CHD for a recurrence risk of 4.1% with CHD occurring in the offspring of affected women more than affected men. A study of 6,640 pregnancies undergoing prenatal echocardiography due to a history of CHD in a parent or sibling found recurrent CHD in 178 (2.7%) cases (Gill, Splitt, Sharland, & Simpson, 2003). The recurrent defects were identical in 37% of cases while 44% of recurrent defects were developmentally similar. Of the families with two or more occurrences, the exact recurrence rate with 55% with the exact recurrence rate being highest for isolated AVSD and laterality defects. Inheritance analysis of probands with LVOTO and their families yielded a relative risk of 36.9 with an estimated heritability of 0.70-

0.90 (McBride et al., 2005). Segregation analysis suggested that one or more minor loci with rare dominant alleles may be segregating in a subset of families while results also suggested significant heritability of several aortic arch and valve measurements.

A Danish population study found that first degree relatives of probands with CHD had a recurrence risk of 79.1 for heterotaxy, 11.7 for conotruncal defects, 24.3 for AVSD, 12.9 for LVOTO, 48.6 for RVOTO, 7.1 for isolated ASD, and 3.4 for isolated VSD (Oyen et al., 2009). The authors noted a recurrence risk ratio of 8.15 for identical defects but only 2.68 for different defects while concluding that, excluding chromosomal defects, only 2.2 percent of CHD in the population is attributable to a family history of CHD in first degree relatives. An epidemiologic study from Norway found that, among sibling pairs, the younger sibling had a 4.1% chance of having CHD if the older sibling had CHD compared to 1.1% if the older sibling did not have CHD for a recurrence risk of 3.6 (Brodwall et al., 2017). The recurrence risk ratio for same-sex twins was reported to be 14.0 while opposite-sex twins had a recurrence risk ratio of 11.9 with 50% of sibling pairs with recurrent CHD having similar types of defects. A large study of 1163 families with 3080 individuals with verified CHD reported reduced penetrance in 32% of families while also noting that a high degree of discordant phenotypes was common though some families with many affected members showing a high concordance rate for certain defects including ASD, VSD, CoA, persistent ductus arteriosus, and TOF (Ellesoe et al., 2018). They reported a rate of first-degree relatedness that was higher for concordant that discordant pairs and showed that specific groups of phenotypes co-occur within families.

A large study estimated that 1.8% of CHD are attributable to rare inherited mutations while also noting a recessive founder mutation in *GDF1* (growth differentiation factor 1) may account for about 5% of severe CHD in the Ashkenazi Jewish community, a recessive *MYH6* phenotype

may account for 11% of Shone's complex cases, and a dominant *FLT4* mutations may account for about 2.3% of TOF cases (Jin et al., 2017). A family-based WES study including 182 individuals from 51 families with multiple members affected by LVOTO-spectrum defects identified disease-associated variants co-segregating within 12 families (Preuss et al., 2016). These variants were found to cluster in the Notch signaling pathway including *NOTCH1*, *ARHGAP31*, *MAML1*, *SMARCA4*, *JARID2*, and *JAG1* and showed significant enrichment for high-pathogenicity variants. Targeted NGS of 57 previously implicated CHD genes in 16 families and 15 controls identified 13 potential causative variants in 9 families of which five co-segregated with the disease (Blue et al., 2014). The authors noted an increased incidence of insertions/deletions, nonsense variants, and splice variants as well as variants classified as "probably damaging" in cases relative to controls as well as an increased number of "rare" and "low-frequency" variants in cases.

In addition to these epidemiological and large-scale studies, causative variants have also been identified in small single-family studies. Intragenic mutations within the *eleastin* (*ELN*) gene have been identified in cases of familial aortic stenosis (Hayano et al., 2019). A dominant and fully penetrant form of familial VSD was found to segregate with a Leu38Pro variant in *CASZ1* which codes for a zinc finger transcription factor essential for cardiogenesis (Huang et al., 2016). A luciferase assay was used to show that the mutant protein had significantly reduced transcriptional activity. A variant in *DCHS1* (Dachsous Cadherin-Related 1) was found to segregate with mitral valve prolapse in several families (Durst et al., 2015). Morpholino knockdown of zebrafish *dchs1* resulted in an atrioventricular canal defect that could be rescued by wildtype human *DCHS1* transcripts but not by transcripts containing the variant found in affected family members. A Gly83Ter mutation in *NR2F2* shown to completely abolish transcriptional activity and synergistic activity with *GATA4* was found to segregate with DORV and VSD in a

family while being absent in 230 ethnically matched healthy controls (Qiao et al., 2018). A frameshift mutation in *NOTCH1* was identified in a patient with HLHS and DORV and found to be inherited from the father who had VSD and pulmonary stenosis (Zahavich, Bowdin, & Mital, 2017). The proband also had a paternal uncle with TOF and paternal grandmother with a history of valve surgery. Finally, a nonsense variant in *PITX2* was found to segregate with disease in a family with 100% penetrant endocardial cushion defect and Axenfeld-Rieger syndrome (C. M. Zhao et al., 2015). This variant was absent in 800 ethnically matched controls and functional analysis showed that the mutant protein had no transcriptional activity and could not synergistically activate transcription with *NKX2.5*.

1.3.4 Mouse Models of CHD

Mouse mutagenesis studies have long been used both to confirm the pathogenicity of variants first identified in human patients and to identify new genes that may contribute to heart development and CHD. There are likely hundreds of such models and what follows is just a small sample to provide an idea of the scope of research being done using mouse models. *Adam19* (ADAM Metallopeptidase Domain 19) is highly expressed in the conotruncus and endocardial cushions during mouse development and mutant mice are perinatal lethal with VSD, abnormal and stenotic semilunar valves, and abnormal cardiac vasculature (Zhou et al., 2004). Mice with null mutations in A-Kinase Anchoring Protein 13 (*Akap13*) had thin-walled hearts and died at E10.5-11.0 while also showing reduced *Mef2c* expression (Mayers et al., 2010). *Arid3b* (AT-Rich Interaction Domain 3B) is expressed in the heart tube, myocardium, and SHF progenitors with mutant mice displaying shortened arterial and venous poles, absent myocardial differentiation, and altered atrioventricular canal patterning with absent EMT (Uribe et al., 2014). Labeling of SHF

progenitors revealed a defect in the addition of cells to the heart. Autophagy Related 13 (*Atg13*) knockout mice are embryonic lethal and show myocardial growth defects (Kaizuka & Mizushima, 2015). *Bmp10*, a member of the TGF-β superfamily of transcription factors, was found to have a crucial role at mid-gestation when the heart transitions from patterning to growth and maturation (H. Chen et al., 2004). Null mice were shown to have ectopic and elevated expression of $p57^{Kip2}$ and a dramatic reduction in cardiomyocyte proliferation activity at E9.0-E9.5. These mice had dysplastic ventricular walls that lacked trabeculae as well as abnormal endocardial cushions. The authors showed that *Bmp10* is expressed in ventricular trabecular myocardium at E9.0-E13.0 but becomes restricted to the atria by E16.0 and concluded that it is required for maintaining normal expression of cardiogenic factors including *Nkx2.5* and *Mef2c* during mid-gestation.

Simultaneous deletion of the executioner caspases *Cas3* and *Cas7* was shown to result in reduced expression of genes that promote DNA replication and cell cycle progression and resulted in mice with hypoplastic hearts at birth (Cardona et al., 2015). Neural crest-specific knockout of *Cdc42* (Cell division cycle 42) was shown to cause embryonic lethality with craniofacial and conotruncal defects (Y. Liu et al., 2013). $Col14a1^{-/-}$ mice were shown to have dysregulated mRNA levels of other collagen types and remodeling enzymes within ventricular myocardium as early as E11.5 (Tao et al., 2012). Though these mice were adult viable, they showed disarrayed collagen fibril organization and defective ventricular morphogenesis resulting in cardiac dysfunction. The ENU-induced $Ddx11^{cetus}$ mutant as well as $Ddx11^{KO}$ DEAD/H-box helicase 11 mutants result in a severe phenotype with developmental arrest at E7.5-E8.5 and cardiogenesis arrested at the linear heart tube stage as well as failure of somitogenesis and failure of chorioallantoic fusion (Cota & Garcia-Garcia, 2012). Endothelium-specific knockout of *Furin*, a proprotein convertase, results in neonatal lethality with VSD and/or valve defects (W. Kim et al., 2012).

Galnt1 which encodes a glycosyltransferase involved in initiating mucin-type O-glycosylation was shown to be essential for normal valve development and cardiac function (Tian et al., 2015). Mutant embryos showed aberrant valve formation as a result of increased cell proliferation within the outflow cushions with the developing valves showing reduced levels of *Adamts1* and *Adamts5*, decreased versican cleavage, increased levels of other extracellular matrix (ECM) proteins, and increased BMP and MAPK signaling. Adult null mice have compromised heart function including semilunar valve stenosis or regurgitation, an altered ejection fraction, and dilation of the heart.

Heterozygous *Gata4*$^{+/-}$ mice were shown to have outflow tract defects including DORV and VSD (J. Liu et al., 2019). They showed that *Gata4* is required by Hedgehog-responsive progenitors within the SHF for normal outflow tract development. In support of this, *Gata4*$^{+/-}$*/Smo*$^{+/-}$ mice displayed a more severe phenotype including PTA. Furthermore, there was enhanced *Gata6* expression in *Gata4*$^{+/-}$ SHF due to a lack of repression by *Gata4* which binds repressor sites upstream of the Gata6 transcription start site. These results led the authors to conclude that outflow tract development is controlled by a *Gata4*-*Gata6*-Hedgehog regulatory network. Creation of *Gata4*$^{-/-}$*/Gata6*$^{-/-}$ embryos through a tetraploid embryo complementation strategy to circumvent their requirement in extra-embryonic endoderm resulted in acardiac embryos that only generated SHF progenitors (R. Zhao et al., 2008).

Ventricular expansion has been shown to be regulated by *Hand1* and *Hand2* in a dose-dependent manner (McFadden et al., 2005). *Hand1* knock-in mice were shown to have increased expansion of the outer curvature of both ventricles with absent formation of the interventricular groove and septum (Togi et al., 2004). They showed that expression of *Hand1* within the entire ventricle altered the expression pattern of *Chisel*, *ANF*, and *Hand2*, but not *Tbx5*. Furthermore,

the interventricular groove formed in transgenic mice that overexpressed *Hand1* in the right ventricle but not the border region. An earlier study showed that *Hand1* null mice were embryonic lethal between E8.0-E9.5 with heart development arresting at the looping stage (Firulli, McFadden, Lin, Srivastava, & Olson, 1998).

Deletion of *Ino80*, a chromatin remodeling protein, in vascular endothelial cells was shown to prevent ventricular compaction and was also associated with defective coronary angiogenesis (Rhee et al., 2018). Furthermore, specific deletion in the sinus venosus and endocardium which are major contributors to the coronary vasculature produced an intermediate phenotype. In vitro experiments revealed that endothelial cells exert an effect on myocardial expansion that is dependent on *Ino80* but independent of blood flow.

$Nipbl^{+/-}$ and $Nipbl^{FLEX/+}$ NIPBL cohesion loading factor mutant mice generated as a model for human Cornelia-de Lange syndrome have ASD with delayed ventricular septation and small hearts (Santos et al., 2016). Expression analysis showed dysregulation of *Hand1*, *Pitx2c*, and *c-Myc*. Additionally, lineage-specific knockout indicated that cardiomyocytes, non-cardiac mesoderm, and endoderm lineages, but not neural crest contribute to the heart defects seen in these mutants. The authors thus concluded that there is a complex interaction between multiple tissue types where the risk conferred by a genetic abnormality in one lineage can be modified by the genetic status of another lineage in a non-additive manner.

Cardiac-selective overexpression of *Nexn*, an F-actin binding protein that functions as a negative regulator of cardiac contractile markers by inhibiting *Gata4*, was shown to result in ASD (F. Yang et al., 2014). Cardiomyocyte-specific conditional knockdown of *Nr2f2* results in ventricularized atria (S. P. Wu et al., 2013). In vitro modeling of the human $NKX2.5^{R142C/+}$ mutation in mouse ($Nkx2.5^{141C/+}$) embryonic stem cells revealed that this mutation caused altered

expression of genes essential for normal cardiac development (Zakariyah et al., 2018). Reduced cardiomyogenesis was found to correlate with reduced *Nkx2.5* nuclear import and downregulation of BMP and Notch signaling. Finally, expression analysis at the early stages of cardiac differentiation revealed a profound dysregulation of genes involved in differentiation and function. Mice with anterior SHF-specific deletion of *Rac1*, a small GTPase, had outflow tract and aortic valve defects (Leung et al., 2015). Disrupted Slit/Robo signaling, which likely acts as a regulator of Notch signaling, results in membranous VSD and BAV (Mommersteeg, Yeh, Parnavelas, & Andrews, 2015). The proper differentiation of cardiac neural crest cells in the outflow tract has been shown to be dependent on *Sox4* and *Sox11* (Paul, Harvey, Wegner, & Sock, 2014). They are thought to participate in regulation of cytoskeletal cell adhesion and ECM molecules and deletion of both in mesoderm and/or cardiac neural crest leads to outflow tract abnormalities ranging from DORV to PTA. Deletion of *Yap1* in cardiac and smooth muscle was found to result in multiple CHD including hypoplastic myocardium, membranous VSD, DORV, and severe vascular anomalies (Y. Wang et al., 2014). Mutations in the retinaldehyde reductase gene, *Dhrs3*, result in outflow tract defects as well as defects in atrial and ventricular septation (Billings et al., 2013).

1.4 The Role of Cilia in Heart Development and CHD

Cilia are microtubule-based projections from the cell surface that can be either motile or nonmotile (primary) that have a variety of functions including signaling, detection of different types of stimuli, and movement of fluid. Defects in cilia formation, structure, and/or function can lead to a variety of diseases collectively terms the ciliopathies. The functions and related pathologies of both primary and motile cilia have been extensively reviewed (Ishikawa, 2017;

Malicki & Avidor-Reiss, 2014; Mitchison & Valente, 2017; Nachury, 2018; Pala, Alomari, & Nauli, 2017; Reiter & Leroux, 2017). Motile cilia at the embryonic node generate a leftward fluid flow that initiates a left-sided signaling cascade leading to specification of the left-right axis. Mutations in genes involved in the formation, structure, and/or function of these nodal cilia lead to abnormal left/right patterning and heterotaxy-spectrum CHD including abnormal heart looping, atrial isomerism, and other complex defects (Barratt, Glanville-Jones, & Arkell, 2014; Boskovski et al., 2013; Del Viso et al., 2016; Field et al., 2011; Francis, Christopher, Devine, Ostrowski, & Lo, 2012; Hartill et al., 2018; Kamura et al., 2011; Kinzel et al., 2010; Larkins, Long, & Caspary, 2012; Robson et al., 2019; Simms et al., 2012).

A variety of studies have looked at the contribution of cilia genes to CHD. One study found that heterotaxy patients with CHD are enriched for genes known to cause primary ciliary dyskinesia (PCD) (Nakhleh et al., 2012). Heterotaxy patients have been found to have a two-fold excess of CNVs compared to controls with these CNVs including cilia genes such as *NEK2*, *ROCK2*, *GALNT11*, and *NUP188* which all disrupt left/right determination and *Pitx2* expression when knocked down in *Xenopus* (Fakhro et al., 2011). A phenotype-driven mouse mutagenesis screen with the goal of identifying genes associated with CHD recovered a large proportion of cilia-related genes (Klena, Gibbs, & Lo, 2017). Of the 91 recessive mutations recovered in 61 genes, 34 were cilia-related, 16 were involved in cilia-transduced cell signaling, and 10 were involved in vesicular trafficking. The cilia-associated CHD genes had a variety of functions including signal transduction, vesicle trafficking, and left/right patterning. Finally, a recessive mouse forward screen showed that cilia and cilia-transduced cell signaling play a significant role in the etiology of CHD (Y. Li et al., 2015).

Among PCD patients, 46% have situs solitus while 47.7% have situs inversus and 6.3% have heterotaxy (Kennedy et al., 2007). While defects in the inner dynein arms and central apparatus are more prevalent in PCD patients with situs solitus, PCD patients with heterotaxy or situs inversus more commonly have outer dynein arm defects. Another study found that 50% of PCD patients have laterality defects of whom 3.5-6% have cardiovascular malformations with 2.6% of those being complex CHD (Harrison, Shapiro, & Kennedy, 2016). Additionally, 2.3% of patients with situs ambiguous also have complex CHD while 4.6% of patients had no CHD (Shapiro et al., 2014). Ultimately, it is thought that the prevalence of CHD with heterotaxy is 200-fold higher in PCD patients than in the general population (Kennedy et al., 2007).

Mutations in a variety of genes needed for cilia assembly, motility, or transduction of the Nodal signaling cascade have been shown to result in laterality defects and associated CHD in model organisms. Mutations in *Zic2*, a component of a regulatory network that drives ciliation of nodal cells, have been associated with randomized cardiac situs and loss of asymmetric gene expression in the early somite stage node and lateral plate mesoderm (Barratt et al., 2014). Though not expressed in these tissues, *Zic2* is transiently expressed in the embryonic node during mid to late gastrulation and mutants have short, abnormal nodal cilia along with downregulated expression of genes required for ciliogenesis (*Noto*, *Rfx3*, *Foxj1*) and cilia function (*Pkd1l1*). *Galnt11*-mediated Notch signaling mediates the distribution of motile to immotile cilia at the *Xenopus tropicalis* left/right organizer (Boskovski et al., 2013). Depletion of *galnt11* or *notch1* was shown to increase the ratio of motile cilia and produces laterality defects reminiscent of those seen in *Pkd2* mutants while Notch overexpression decreases the ratio and produces a phenotype reminiscent of PCD. Inner ring nucleoporins *nup93* and *nup188* have been shown to localize to the ciliary base in Xenopus with their depletion leading to loss of *pitx2* expression and loss of

asymmetric *coco* (*dand5*, *cerl2*) expression at the left-right organizer and reduced cilia density leading to abnormal heart looping (Del Viso et al., 2016). *DNAAF1* (Dynein axonemal assembly factor 1) is required for dynein heavy chain assembly and mutations cause PCD with or without heterotaxy and associated CHD in humans while zebrafish mutants display randomized heart looping (Hartill et al., 2018; van Rooijen et al., 2008). Pitchfork (*Pifo*) has been shown to associate with ciliary targeting complexes and accumulate at basal bodies during cilia disassembly with haploinsuffiency shown to cause duplicated node cilia and left-right patterning defects in mice (Kinzel et al., 2010). The authors also reported finding mutations in this gene in a fetus with situs inversus totalis and renal and liver cysts as well as a patient with DORV. Mice harboring the *Arl13b*^{hnm} mutation in ADP Ribosylation Factor Like GTPase 13B have short node cilia leading to an inability to maintain asymmetric Nodal expression ultimately downregulating *Cerl2* expression at the node leading to heterotaxy and abnormal heart looping (Larkins et al., 2012). Mutations in the zebrafish ortholog of *AHI1* (Abelson helper integration site 1), a cause of Joubert syndrome in humans, result in absent cilia in Kupffer's vesicle leading to cardiac looping defects (Simms et al., 2012)

Mice harboring mutations in *Dnaic1* (Dynein axonemal intermediate chain 1) have immotile nodal cilia with half showing situs solitus or inversus and the other half showing heterotaxy including bias towards an L-looped heart (Francis et al., 2012). Mice with normal or mirror-image situs had normal hearts or only mild defects such as azygous continuation or septal defects whereas heterotaxy mice all had complex CHD. Mutations in Xenopus *rnf20*, a member of the H2Bub1 complex that regulates inner dynein arm assembly, result in abnormal cilia motion (Robson et al., 2019). Mutations in murine *Pkd1l1* and *Pkd2*, encoding polycystin 1-like 1 and polycystin 2 respectively, result in failed activation of asymmetric gene expression at the node

and right isomerism (Field et al., 2011). These proteins are thought to form a complex that serves as a "nodal flow detector" with Medaka embryos harboring the *abecobe* mutant (*pkd1l1*) displaying defective asymmetric expression of s*outhpaw*, *lefty*, and *charon* in the context of normal nodal flow (Kamura et al., 2011).

Two patients have been reported with a heterozygous non-synonymous 455G>A mutation in the functional domain of *DAND5*, a master regulator of Nodal signaling (Cristo et al., 2017). The first patient had left isomerism with VSD, overriding aorta, and pulmonary atresia while the second patient had a VSD, overriding aorta, pulmonary atresia, and hypertrophy of the right ventricle. Functional analysis showed the mutant protein had significantly decreased activity compared to wild type. A Lys132Ter nonsense variant in *HAND1* was identified in a patient with DORV (L. Li et al., 2017). This variant was not found in 600 controls and was shown to have no transcriptional activity and disrupted synergistic activation with *GATA4*. Mutations in *MEGF8* (Multiple EGF-like domains 8) have been associated with a form of Carpenter syndrome (Acrocephalopolysyndactyly type II) that includes laterality defects including situs inversus totalis, dextrocardia, and TGA (Twigg et al., 2012). *ZIC3* is one of the better known causes of heterotaxy and CHD accounting for 5.2% of sporadic cases in males (Cowan, Tariq, & Ware, 2014). A study of probands with TGA and a family history of concordant or discordant CHD identified sequence variants in *FOXH1*, *ZIC3*, and *NODAL* (De Luca et al., 2010).

Primary cilia have been known to exist in the embryonic and adult heart for over 40 years (Myklebust, Engedal, Saetersdal, & Ulstein, 1977; Rash, Shay, & Biesele, 1969). In the rat, primary cilia have been identified from all stages of development from embryo to subadult (Kaur, McGlashan, & Ward, 2018). Monocilia have been noted on most regions of the heart except for the atrial septum and their presence has been correlated with the amount of shear stress with high

shear stress noted to induce deciliation and EMT (Egorova et al., 2011; Van der Heiden et al., 2006). Cilia have been shown to be present on non-cardiomyocyte cells in embryonic and adult mouse heart, but absent from cardiomyocytes at all stages instead being exclusive to cardiac fibroblasts where they have a key role in the cardiac fibrosis pathway (Villalobos et al., 2019). Another study noted that cilia are observed infrequently on the mitral valve endocardium but are found on most interstitial cells of the embryonic, fetal, and neonatal mitral valve and that their presence correlates with the type of extracellular matrix being produced (Toomer et al., 2019).

It is now known that primary, but not motile, cilia play a key role in the later stages of heart development after the left/right axis has been specified. One study showed that cilia motility genes are not expressed in SHF whereas genes needed for cilia structure and function are expressed while another study showed that knockdown of primary cilia in P19 CL6 cells using *Ift88/20* siRNA or cyclopamine blocks hedgehog signaling and prevents these cells from differentiating into beating cardiomyocytes (Burnicka-Turek et al., 2016; Clement et al., 2009). At E9.5 of mouse development cilia line the endothelium of the developing atria and the ventricular trabeculations and less prominently on the compact myocardium and are also present within the cushion mesenchyme and on the atrial side of the overlying endothelium (Slough, Cooney, & Brueckner, 2008). Primary cilia are found at distinct regions of the developing atria and ventricles at E10.5-E12.5 (Gerhardt, Lier, Kuschel, & Ruther, 2013). From E11.5 to E13.5 primary cilia are present on interstitial cells of the developing outflow cushions but rare on valve endocardium (Toomer et al., 2017). Cilia expression is temporally regulated and restricted to specific extracellular matrix zones of the developing aortic valve where they grow to a maximum length of 2.5-3 μm by E17.5 before gradually shortening postnatally being mostly lost from the interstitial cells of adult atrioventricular valves. Cilia-mediated Hedgehog signaling through *desert hedgehog* (*Dhh*) which

is robustly expressed in cushion endocardium is thought to play an important role in aortic valve development (Toomer et al., 2017). It is thought that cilia aid in restraining valve size and modulate the production of components of the extracellular matrix. Primary cilia have been shown to frequently occur on the atrioventricular and conotruncal cushions at E12.5 (Willaredt et al., 2012). In these tissues, mesenchymal cells were shown to have randomly oriented cilia while endocardial cells oriented their cilia towards the lumen of the inflow or outflow tract. Cilia remain present on the atrial endocardium at E12.5, though not as prominently as at E9.5 (Slough et al., 2008). Cilia also persist on the ventricular surface of trabeculations within both ventricles as well as being prominent on the epicardium extending toward the pericardial space.

Mutations in a number of genes affecting both primary and motile cilia, some of which are known to cause ciliopathies, have been linked to CHD in humans. A patient with PDA, VSD, dextroversion, pulmonary artery hypoplasia, and other extracardiac findings was found to have compound heterozygous variants in *CRBS2*, a Crumbs family protein implicated in the planar cell polarity pathway and needed for proper cilia alignment, and a heterozygous variant in *TTCB21* (Jaron et al., 2016). Orofaciodigital syndrome type VI, caused by mutations in the ciliogenesis and planar polarity effector (CPLANE) genes *INTU* and *WDPCP*, can have TOF and coarctation of the aorta (Toriyama et al., 2016). Mutations in the actin proteins *ACTB* and *ACTG1* are associated with Baraitser-Winter cerebrofrontofacial syndrome which can include BAV, VSD, ASD, PDA, and mitral regurgitation (Verloes et al., 2015). *INVS* and *NPHP3* mutations, though more commonly associated with nephonophthisis, can also be associated with CHD (Wolf, 2015). Ellis-van Crevald syndrome caused by mutations in the *EVC1* or *EVC2* genes has been associated with common atrium, persistent left superior vena cava, and atrioventricular canal defects (Digilio et al., 1999). *Evc1/2* have been shown to be expressed in the developing atrioventricular cushions

and outflow tract around the time of atrioventricular septation in mouse studies (Sund, Roelker, Ramachandran, Durbin, & Benson, 2009). Patients with AVSD have been shown to be enriched for rare variants in centrosomal protein 152 (*CEP152*) compared to controls (D'Alessandro et al., 2016). A GWAS of patients with BAV revealed significant SNPs in or near genes involved in regulation of ciliogenesis through the exocyst (Fulmer et al., 2019). Disruption of the exocyst in zebrafish resulted in defective ciliogenesis and cilia signaling as well as a variety of heart defects while mouse mutants had aortic valve defects including BAV, stenosis, and calcification. Mutations in *DZIP1*, a regulator of ciliogenesis and/or cilia signaling, have been shown to segregate with mitral valve prolapse in a large family and have also been found in several unrelated individuals with isolated mitral valve prolapse (Toomer et al., 2019). *Dzip1*$^{S14R/+}$ neonatal mice have a mitral valve phenotype associated with cilia that are shorter in length compared to wild type littermates.

A number of cilia genes have also been found to cause heart defects in mice. Loss of *Ftm* (*Rpgrip1l*) results in shortened cilia and reduced proliferation in ciliated areas ultimately leading to reduced ventricular wall thickness and VSD (Gerhardt et al., 2013). The authors hypothesized that cilia-mediated hedgehog and PDGFR signaling regulate ventricular proliferation. Examination of E9.5 *Lrd*$^{-/-}$ (immotile nodal cilia), *Pkd2*$^{-/-}$ (defective ciliary mechanosensation), and *Kif3a*$^{-/-}$ (absent cilia) mice revealed decreased cellularity of the endocardial cushions and thinning of the compact myocardium with increasing severity (Slough et al., 2008). The authors only examined D-looped hearts to control for defective left/right determination and hypothesized that cardiac cilia function as flow sensors with a vital role in the regional control of EMT and may sense myocardial function in ventricles and/or epicardium to regulate proliferation during development.

Ift88^{tm1Rpw} mutant mice (a null allele) lack cilia and have dilated ventricles with decreased trabeculation and abnormal outflow tract development at E11.5 (Clement et al., 2009). Additionally, knockdown of *Ift20* and *Ift88* by siRNA or cyclopamine was found to block hedgehog signaling. Conditional knockout of *Ift88* in cardiac endothelium results in shortening or loss of primary cilia and a highly penetrant BAV phenotype (Toomer et al., 2017). Homozygous *cobblestone* (*cbs*) mutants, encoding a hypomorphic *Ift88* allele, result in a variety of heart defects including PTA with malposition of the common trunk, abnormal arch artery patterning, ASD, VSD, and a hypoplastic myocardium (Willaredt et al., 2012). Mutant hearts showed a substantial loss of primary cilia in the atrioventricular cushions and pericardium as well as loss of *Isl1* expression in SHF, downregulation of downstream targets of Shh signaling and reduced *Bmp2/4* expression.

Pkd1$^{del17\text{-}21\beta geo}$-/- mice at E13.5-14.5 have CHD including DORV, disorganized myocardium, and abnormal atrioventricular septation (Boulter et al., 2001). *Pkd1* is expressed in the aorta at E9.5 later becoming expressed in all major vessels and throughout the heart, particularly in the outflow tract, endocardial cushions, and valve leaflets by E10.5. *Pkd1* is also expressed at high levels in the adult mouse aortic outflow tract and atrial appendages as well as expression in vascular endothelial and smooth muscle cells of the major vessels (Boulter et al., 2001). *Pkd2*$^{-/-}$ mice exhibit completely penetrant VSD with some also having ASD or AVSD (G. Wu et al., 2000). A mouse ENU screen identified *Dnah11* and *Mks1* as causing recessive AVSD phenotypes (Burnicka-Turek et al., 2016). Mutations in *Dnah11*, an axonemal protein required for cilia motility, were shown to not disrupt SHF Hedgehog signaling and caused AVSD with concurrent heterotaxy whereas mutation in *Mks1*, encoding a basal body protein, disrupted Hedgehog signaling within the SHF and caused isolated AVSD without heterotaxy. Finally, the

mouse *Beetlejuice Prickle1* mutant has defective planar cell polarity leading to outflow tract shortening and malalignment as well as motile and primary cilia defects (Gibbs et al., 2016).

1.5 Public Health Significance and Objectives

Congenital heart disease is one of the most common birth defects with an estimated prevalence of 4-10 per 1000 live births (Pierpont et al., 2007). These defects can range from simple and sometimes asymptomatic septal defects or bicuspid aortic valve to highly complex malformations involving multiple structures. Sequencing and genome association studies in humans as well as mutagenesis studies in mice and other model organisms have identified a growing number of genes that either directly cause or modify the risk or phenotype of congenital heart disease. With improvements in the diagnosis and surgical management of these defects, the rate of long-term survival has steadily increased. One review reported that mortality rate for CHD decreased by 40% from 1979-1997 and by a further 24% by 2006 (Mazor Dray & Marelli, 2015). This review also notes that the average age of patients with severe CHD has increased from 11 in 1985 to 17 in 2000 and 25 in 2010. A Norwegian study reported that survival to age 16 in children with severe CHD increased from 62% from 1971-89 to 87% from 1999-2011 (Erikssen et al., 2015). By 2000 the global population of adults with CHD was estimated at 40,000 per 10,000,000 with this number expected to increase (Marelli, Mackie, Ionescu-Ittu, Rahme, & Pilote, 2007). Despite the increased survival, CHD patients still face a lifetime of follow-up that may include various diagnostic procedures or possible repeat surgeries as well as the risk of late complications such as atrial arrhythmias, pulmonary arterial hypertension, infectious endocarditis, and heart failure (Bouma & Mulder, 2017; Mazor Dray & Marelli, 2015). In addition, there is increasing

evidence that some of the same genetic defects that cause CHD as well as extrinsic factors such as radiation exposure from interventional procedures and chronic hypoxia may increase the lifetime risk of developing cancer for these patients (Botto et al., 2013; Fisher et al., 2012; Gurvitz et al., 2016; Moore, 2009; Nishi, Miyake, Takeda, & Hatae, 2000; Olsen et al., 2014; Opotowsky et al., 2015). With these factors in mind, a better understanding of the genetic etiology is of public health relevance as knowledge of the genetic causes of CHD may lead to improved preventative and interventional strategies as well as better risk assessment for early and late complications. To this end, this study was untaken with the following objective in mind.

1) Show that CHD patients have an increased burden of rare pathogenic and highly damaging variants in genes known to be associated with heart development.
2) Examine if CHD patients have an increased burden of genes known to be associated with cancer
3) Identify novel genotype-phenotype correlations in our patient cohort

2.0 Enrichment for Rare Pathogenic and Highly Damaging Variants

2.1 Introduction

Recent evidence has increasingly pointed to a complex origin for CHD that goes beyond simple Mendelian inheritance with variants in multiple genes often acting to modify expressivity and penetrance of phenotypes. A prime example of this comes from a recently published case that describes a child and fetus with severe left ventricular noncompaction who inherited a variant in *MYH7*, known to cause cardiomyopathies, from their father who was only mildly affected as well as an *NKX2.5* variant from their unaffected mother (Gifford et al., 2019). The authors propose that the *MYH7* variant which arose de novo in the father drives the noncompaction phenotype while the *NKX2.5* variant inherited from the mother dramatically increased the severity. This evidence was supported by the observation of left ventricular noncompaction in mice that were triply heterozygous for the *MYH7* and *NKX2.5* variants as well as a variant in *MKL2* which is involved in transcriptional regulation of cardiac genes and was present in the father and children as well as the paternal grandfather who had a history of arrhythmia but no structural abnormalities.

A WES study of patients with left-sided heart defects detected a great deal of genetic heterogeneity as well as complex inheritance involving dominant, recessive, and X-linked models (A. H. Li et al., 2017). The authors also found that a majority of the detected heterozygous loss-of-function variants were inherited from presumably unaffected parents, though it was noted that cardiac examinations were not performed on the parents so subclinical defects such as BAV could not be ruled out. Another WES study found that patients with severe CHD had a significant excess of protein-altering variants in genes expressed in the heart (Zaidi et al., 2013). A large study of

parent-child trios of syndromic and non-syndromic CHD cases reported a significant enrichment for de novo but not inherited protein-truncating variants in syndromic cases but conversely an excess of protein-truncating variants inherited from unaffected parents in non-syndromic cases (Sifrim et al., 2016). A GWAS of patients with 22q11.2 deletion identified SNPs intronic to *GPR198* associated with TOF and that may have a regulatory role with the nearby *MEF2C* gene (T. Guo et al., 2017). Another GWAS of Han Chinese cases and controls identified SNPs in *TBX15* and *MAML3* that were associated with risk of CHD (Hu et al., 2013).

A study of parent-child trios of patients with AVSD found that 32% of trios carried at least 1 disease-causing variant across 19 loci suggesting AVSD can be caused by a heterogeneous group of genes (Priest et al., 2016). In a study of event-free survival in a cohort of patients with acute aortic dissection (AAD), patients with three or more variants in genes known to be associated with AAD had shorter event-free survival than patients with fewer variants (Zheng et al., 2018). Another study showed that variants in *PLOD1* may predispose BAV patients to aortic dilation due to impaired collagen cross-linking (Wagsater et al., 2013). A multi-level study found that isolated TOF is caused by a combination of deleterious, rare, and private mutations in genes essential for apoptosis, cell growth, sarcomere assembly, neural crest, and SHF (Grunert et al., 2014). Nonsynonymous variants in the transcription factor *PBX3* was found to be present in 5.2% of CHD patients compared with 1.3% of controls suggesting they may act as modifiers or risk alleles for CHD, especially outflow tract defects (Arrington, Dowse, Bleyl, & Bowles, 2012). A study found that different missense, loss-of-function, and intronic variants within *NOTCH1* confer different risks for LVOTO (Helle et al., 2018). Finally, it has been observed that CHD patients frequently have ciliary dysfunction and are enriched for de novo mutations in cilia-related pathways (Klena et al., 2017).

Several studies in mice have also found evidence of the complex genetic nature of CHD. A mouse forward genetic screen found a multigenic etiology for HLHS with mutations in *Sap130* driving left ventricular hypoplasia and mutations in *Pcdha9* increasing penetrance of aortic valve defects while also finding enrichment for variants in 10 human chromosomal regions hinting at complex inheritance (X. Liu et al., 2017; H. Yagi et al., 2018). Intercrossing different strains of mice with a heterozygous *Nkx2.5*$^{+/-}$ mutation produced phenotypes with differing severity based on co-inheritance of different quantitative trait loci from different strain backgrounds (Akhirome et al., 2019). Finally, several studies have indicated that mutations in genes that interact with *Tbx1* can modify the severity of CHD in mouse models of DiGeorge syndrome (L. Chen et al., 2012; C. Guo et al., 2011).

Adding another layer to this complexity is the finding that epigenetics, particularly chromatin modifiers, play an important role in CHD. One study found that CHD patients had a significant excess of mutations in genes involved with histone H3 lysine 4 (H3K4) methylation and H2BK120 ubiquitination (Zaidi et al., 2013). Another study found significant enrichment for loss-of-function and protein altering variants in genes encoding components of the H2Bub1 complex, a regulator of tissue-specific expression of cilia genes through histone H2B monoubiquitination, in CHD patient with laterality defects but not controls (Robson et al., 2019). The authors, through knockdown of complex protein *rnf20* in zebrafish and Xenopus, showed that laterality defects resulted from impaired motile cilia motion and that this gene regulates assembly of inner dynein arms. Rare variants in *SMYD4*, a lysine methyltransferase that interacts with *HDAC1* and may also modulate histone acetylation, were found in a cohort of CHD patients (Xiao et al., 2018). Knockout of *smyd4* in zebrafish was found to cause heart defects including abnormal left/right patterning, abnormal looping, and hypoplastic ventricles.

In mice, inactivation of *CHD7* encoding an ATP-dependent nucleosome remodeling factor impairs multiple BMP signaling-dependent cardiogenic processes (Y. Liu et al., 2014). The authors showed that *CHD7* interacts with canonical BMP signaling pathway nuclear mediators *Smad1/5/8* as well as associating with *Nkx2.5* in a BMP-dependent manner. *Fam60a* interacts with components of the Sin3a HDAC transcriptional corepressor complex and *Fam60a*$^{-/-}$ mice have hypoplastic right ventricles, TGA, DORV, and other extracardiac anomalies reminiscent of laterality defects albeit with normal *Pitx2* expression (Nabeshima et al., 2018). Mutations in *Sap130*, a component of the Sin3a HDAC corepressor complex cause HLHS in conjunction with mutations in *Pcdha9* (H. Yagi et al., 2018). Mice with mutations in *Brg1*, part of the BAF chromatin remodeling complex, have a variety of heart defects as well as dysregulated expression of multiple cardiac genes including *Nppa*, *Tbx5*, and *Bmp10* (Takeuchi et al., 2011). Finally, knockdown of zebrafish and Xenopus orthologs of the DNA methyl transferases *DNMT1* and *DNMT3B* result in abnormal heart jogging and heterotaxy (L. Wang et al., 2017).

Considering these factors, the aim of this portion of the study was to determine if CHD patients have an enrichment of any particular type of variants or in any particular class of gene.

2.2 Materials and Methods

2.2.1 Study Design and Objectives

This research was undertaken as part of a study titled "Analysis of the Genetic Basis of Congenital Heart Disease" (Pitt Protocol STUDY19030229) with the primary goal of investigating whether ciliary dysfunction plays a role in CHD. For this study, 1500 CHD patients aged 0 to 65

including 500 with heterotaxy were to be recruited from the University of Pittsburgh Medical Center and the Children's National Medical Center in Washington D.C. Patients without heart disease but with confirmed or suspected PCD or symptoms consistent with a ciliopathy were also recruited. Some relatives, both affected and unaffected, of patients recruited at the University of Pittsburgh Medical Center were also recruited. Additionally, specimens from 29 patients undergoing heart transplant for hypoplastic left heart syndrome were obtained from the University of California at San Diego and a small number of samples were obtained from the University of Minnesota. For patients recruited at the University of Pittsburgh a detailed medical history including the nature of the structural heart defects, presence of extracardiac defects, and family history were obtained through patient questionnaires administered by a clinical research coordinator and review of medical records by a cardiac pathologist. One or more cardiac "phenotype constellations" were then assigned to each patient by the cardiac pathologist based on detailed review of medical records.

2.2.2 Characteristics of the Study Cohort

To date over 800 participants including probands and family members both affected and unaffected have been recruited for this study of whom good whole exome sequencing data has been obtained for 651. For the purposes of this project, only 586 sequenced probands with structural heart defects were included while relatives of probands were excluded regardless of CHD status. The study population included 362 males and 224 females (ratio 1.6161) of whom the vast majority (505) identified as White. These patients had a wide range of structural heart defects ranging from relatively simple isolated septal defects or bicuspid aortic valve to highly complex defects involving multiple structures (Appendix Table 1). It should be noted that only

the patients recruited from Children's Hospital of Pittsburgh received the questionnaire and had their records reviewed by a cardiac pathologist and that not all patient medical records had been reviewed at the time of writing this dissertation. For the non-Children's patients, phenotypes were assigned based only on the information that was available.

2.2.3 Whole Exome Sequencing and Data Cleaning

Whole exome sequencing was performed using the Illumina HiSeq2000 platform (Illumina, San Diego, CA) and Agilent SureSelect V4 and V5 exon capture kits (Agilent Technologies, Santa Clara, CA) at 100x coverage. Alignment of Fastq files to hg19 was carried out using BWA 0.5.9-r16 (H. Li & Durbin, 2009). For quality assessment, FREEMIX (Jun et al., 2012) was used to check for contamination and SAMtools 1.9. (H. Li et al., 2009) was used to check for proper alignment. Single nucleotide and insertion deletion variant calling, genotyping, and filtering of BAM files at population level utilizing the GATK 4.0.8.1 best practices pipeline (DePristo et al., 2011; Van der Auwera et al., 2013). Bcftools 1.9 was used for additional site-specific filtering such that SNPs within 10 base pairs of an indel, clusters of indels within 5 base pairs of each other, sites with missing rate (genotype quality < 20) greater than 10%, sites with a Hardy Weinberg p-value of less than 10e-6, and sites less available for sequencing and mapping based on ExAC (Lek et al., 2016) are excluded. Variant annotation was carried out using Ensembl (Hunt et al., 2018) VEP26 with predicted transcript and protein changes in the Ensembl known gene table (hg19 accessed 24 March 2019), allele frequency in GnomAD (Karczewski et al., 2019), and predicted deleteriousness based on Combined Annotation-Dependent Depletion (CADD) score (Kircher et al., 2014).

Tab-delimited .txt files were generated with additional filters applied such that calls with genome quality < 20 and depth < 20 as well as variants with a GnomAD (version r2.1.1) allele frequency > 0.01, a CADD score < 10, or those that were present in > 20% of our samples were excluded. Patient information including unique identifier, race, sex, and cardiac defects were added. Because demographic information was not available for some of the non-Pitt samples, this information was calculated based on principal component analysis (PCA) of the exome data for all patients. Additional annotations from ClinVar (Landrum et al., 2016) (v.20190422), dbNSFP (X. Liu, Wu, Li, & Boerwinkle, 2016) (v3.5a for hg19) and additional damage metric scores were downloaded and integrated. Annotations from the Human Gene Mutation Database (HGMD) (Stenson et al., 2014) v.2019.3 were added manually using the batch search feature. The phenotypes associated with variants curated in ClinVar and HGMD were later used as part of the data analysis.

2.2.4 Gene Lists

Several lists of genes were used to prioritize candidate genes. The CHD100 list is a set of 100 genes that have been identified to cause congenital heart defects in mice though mutagenesis screens carried out in the Lo Lab (Y. Li et al., 2015 and unpbulished data). The CHD-ome list is a collection of genes that have been associated with congenital heart defects in humans, mice, or other model organisms. This list includes genes from the CHD100 as well as genes identified to cause heart defects in human from two large studies (Jin et al., 2017; Sifrim et al., 2016), a set of genes associated with heart development generated though gene ontology (GO) search on the Mouse Genome Informatics database (Eppig et al., 2017) carried out by a study collaborator, and extensive literature review. It should be noted that not all CHD-ome genes have been reported as

causing heart defects in the literature, particularly some of those obtained from the GO search, and that some associations may be dubious.

The Ciliome gene list is a set of genes that have been associated with ciliogenesis, cilia function, or cilia-mediated signaling. This list was compiled from multiple sources including an ongoing mouse mutagenesis study being carried out in the Lo Lab, the LoFullCiliome list of cilia-associated genes generated through a past literature search, a list of 40 genes known to cause primary ciliary dyskinesia (Zariwala, Knowles, & Leigh, 2007) the Syscilia list of cilia-associated genes (van Dam, Wheway, Slaats, Huynen, & Giles, 2013), and more recent literature searches.

For this dissertation, attention will be paid to the CHD-ome and Ciliome as a whole plus the CHD100 subset within the CHD-ome. The CHD-ome consists of 1405 genes and loci with the CHD100 nested completely within it while the Ciliome list consists of 1280 genes. Figure 1, generated using Venny (Oliveros, 2007-2015), illustrates the degree of overlap. There are 277 genes in common between the CHD-ome and Ciliome with 65 of them also being CHD100 genes while 35 CHD-ome genes are in the CHD100 but not the Ciliome.

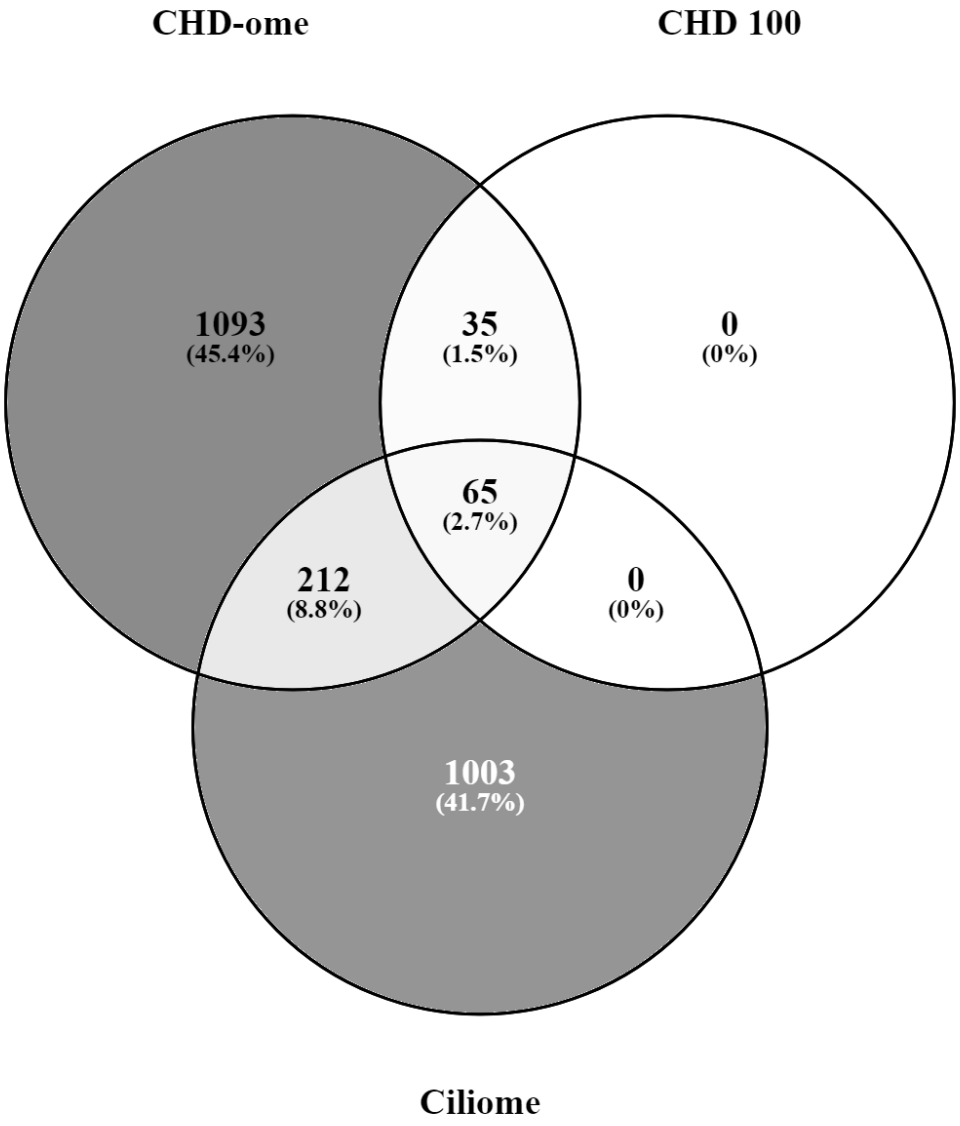


Figure 1 Overlap of gene sets used in this dissertation

2.2.5 Control Cohort

For a control cohort, datasets were downloaded from the UK10K consortium (The UK10K Consortium et al., 2015). This dataset comprises three different study populations; whole exome

data from ~3,000 participants in a study on schizophrenia and autism spectrum disorder, whole exome data from ~1,000 individuals from the Severe Childhood Onset Obesity Project (SCOOP) as well as obese adults from several population cohorts, and whole genome data from the Avon Longitudinal Study of Parents and Children (ALSPAC). For this dissertation, the SCOOP/obesity dataset, henceforth referred to as SCOOP, was used for a control population. The VCF files for the SCOOP cohort and our patient cohort, henceforth referred to as CHD, were combined and processed as a single dataset to ensure that calls were over the same intervals and that the data was all identically filtered.

2.2.6 "Pathogenic" and "Highly Damaging" Variants

Pathogenic variants were defined as those being curated as "Likely pathogenic" or "Pathogenic" by ClinVar and as "DM" (Disease-associated mutation) by HGMD. As ClinVar is a freely accessible database, variant data including "clinical significance" which provides a pathogenicity rating as well as associated phenotypes was downloaded from the database and incorporated into the spreadsheets. HGMD is a subscription-based service with a substantial fee and the database files were not available for download. Therefore, it was necessary to manually look up variants by hg19 coordinates using the 'batch search' feature which can look up 500 variants at a time.

Highly damaging variants were defined as those with a CADD score ≥ 30 corresponding to the top 0.1% most damaging variants. In short, CADD scores are generated by contrasting the annotations of fixed or nearly fixed alleles with simulated variants (Kircher et al., 2014; Rentzsch, Witten, Cooper, Shendure, & Kircher, 2019). While fixed alleles will be depleted for variants with a severely negative impact on organismal fitness by natural selection, this will not occur in

simulated variants. Scores are generated through a machine learning algorithm that integrates information from multiple annotations. CADD scores are represented in a phred-like log10 scaled C-score based on the C-score rank of each individual variant relative to the 8.6 billion possible variants such that C-scores ≥ 10 represent the top 10% of damaging variants, scores ≥ 20 represent the top 1%, and so forth. Nonsense variants tend to have the highest C-scores followed by missense and canonical splice variants. The advantages of CADD include the ability to merge information from multiple annotations into a single value while also being able to accommodate expansions to these annotations as well as combining the generality of conservation metrics with the specificity of functional metrics. The main weaknesses of CADD scores include the inability to account for differences in local mutations rates and related phenomena, an inability to reflect the strength with which a site is being selected, difficulty with intergenic sites, and an inability to directly correlate CADD score with the likelihood that a variant is truly pathogenic (Kircher et al., 2014; Rentzsch et al., 2019)

To better identify and categorize curated phenotype terms that were related to the cardiovascular system, the International Classification of Diseases (ICD)-10 version 2016 code was obtained either directly from the WHO website ("ICD-10-CM the Complete Official Codebook," 2015), or from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, Sato, Furumichi, Morishima, & Tanabe, 2019), or Orphanet (Pavan et al., 2017). Google searches and Wikipedia were also occasionally used for terms whose codes could not be found elsewhere. Cardiovascular system phenotypes received codes I00-I99 (Diseases of the Circulatory System) or Q20-Q28 (Congenital Malformations of the Circulatory System). These were further subdivided into structural CHD terms which encompassed Q20-Q28 and a few from I00-I99, and non-structural terms which encompassed the bulk of I00-I99.

2.2.7 Data Analysis and Statistics

Summary statistics were calculated in Microsoft Excel and a two-sample t-test assuming unequal variance was carried out using the Analysis ToolPak add-in to compare the mean number of calls between cases and controls. Allele counts were obtained for each cohort at the variant, gene, and phenotype level and for overall number of alleles, number of highly damaging alleles, and number of pathogenic alleles. Allele counts were then normalized to number of associated alleles per 100,000 alleles in the population. Fold change (FC) was then calculated based on normalized allele counts and relative to the SCOOP cohort such that $FC > 1$ indicates more alleles in the CHD cases and $FC < -1$ indicates more alleles in the SCOOP controls.

A binomial test was carried out to evaluate the probability of observing at least as many alleles as were present in the CHD cases by chance assuming the allele frequency in the SCOOP controls is representative of the population frequency. This test was done at the variant, gene and phenotype level. The test was carried out in Microsoft Excel (2016) using the binom.dist function in the form *1-binom.dist()* to produce the upper tail probability. The arguments for the binom.dist function are as follows.

1) *Number_s* is the number of successes defined as the number of associated alleles in the CHD cases
2) *Trial* is the number of trials defined as the number of possible associated alleles in the CHD cases equivalent to the number of samples times two alleles per locus times the number of associated variants.
3) *Probability_s* is the probability of success defined as the allele frequency of associated alleles in the SCOOP controls.
4) *Cumulative* when set to ‘TRUE performs the cumulative distribution function.

The Benjamini-Hochberg false discovery rate (BH FDR) was then calculated in Excel using the following method.

1) Raw p-values as calculated by the binomial test were ordered from smallest to largest and ranked giving equal p-values the same rank
2) The Benjamini-Hochberg critical value was calculated using the formula *(i/m)Q* where *i* is the rank of the raw p-value, *m* is the total number of tests, and *Q* is the desired false discovery rate equal to 0.05 in this case.

Odds ratios, defined as the odds of having a minor allele in the CHD cases over the odds of having a minor allele in the SCOOP controls where "minor allele" refers to alleles in any variant present in the filtered dataset, were calculated at the variant, gene, and phenotype level using Microsoft Excel. Fisher's exact test and Benjamini-Hochberg FDR were then calculated using the R statistical package (R Development Core Team, 2013) and a slightly modified version of a freely available script found on an online discussion forum (Appendix D).

2.2.8 Enrichment Analysis

Gene lists were made of genes in each category (overall, pathogenic, highly damaging) with the largest absolute fold change (Appendix C.1) These lists were run through functional enrichment analysis on ToppGene (J. Chen, Bardes, Aronow, & Jegga, 2009) and Variant Effects Analysis on Ingenuity Pathway Analysis (IPA) (Kramer, Green, Pollard, & Tugendreich, 2014). ToppGene outputs lists of significant terms divided into multiple categories derived from various sources and calculates a p-values and a Benjamini-Hochberg false discovery rate (q-value) with the significance cutoff set at 0.05. IPA Variant Effects Analysis produces a variety of different forms of data including a list of "Canonical pathways" and a list of "Diseases and Bio Functions"

terms that are significantly enriched in the given gene set with significance reported in p-values. Lists of selected options and parameters for ToppGene and IPA analyses can be found in Appendixes C.2 and C.3, respectively. Additionally, IPA has a "comparison analysis" feature that allows direct comparison of results from multiple analyses such that the analyses for CHD datasets (positive FC) could be compared to the lists from the SCOOP dataset (negative fold change). For the comparison analysis IPA converts the p-values of significant terms to $-\log_{10}$(p-values). 'Deltas' were calculated as the difference in $-\log_{10}$(p-value) with positive deltas indicating more significance in the CHD cases and negative deltas indicating more significance in the SCOOP controls.

2.3 Results

2.3.1 Summary Statistics

Summary statistics for the CHD and SCOOP cohorts can be found in Table 1. Overall there were a combined 36,913 variants in 12,711 genes between the two datasets with a great deal of overlap as 35,327 (95.55%) variants were present in both datasets leaving only 181 variants present in the CHD cases alone and 1405 variants present in the SCOOP controls alone. It should be noted that non-CHD probands were only removed after processing and filtering and the degree of overlap between the CHD and SCOOP datasets was even greater in the original version that included data from all 651 study participants who have been sequenced with 199 and 299 variants unique to the CHD and SCOOP cohorts respectively.

The CHD cases consisted of 586 samples while the SCOOP controls consisted of 982 samples, thus the total number of variants and calls is larger in the SCOOP controls. While the average number of calls overall and for damaging and pathogenic calls is higher in the CHD cases, it should be noted that the cohorts are not stratified by race as the racial makeup of the SCOOP controls was not immediately available. In looking at the CHD cases alone and stratifying by race as determined by PCA, Africans had far more calls than Whites while Asians had fewer calls than Whites and mixed-race patients had numbers intermediate between their component races (Appendix Figure 1). CHD cases had a significantly higher mean number of calls overall and in calls in CHD-ome and Ciliome genes (Table 1). Cases also had significantly more calls in damaging variants overall compared to controls while the excess in damaging calls in CHD-ome and Ciliome genes in cases compared to controls was trending toward significance. The mean number of calls in variants curated at pathogenic were not significantly different between cases and controls overall or when restricting to CHD-ome or Ciliome genes. Though all samples in both cases and controls had at least two calls in highly damaging variants, there were many case and control samples with zero or one calls in curated pathogenic variants. (Table 1).

Table 1: Summary Statistics

	Overall		Highly Damaging		Pathogenic	
Variants	**CHD**	**SCOOP**	**CHD**	**SCOOP**	**CHD**	**SCOOP**
Total	35508	36732	3515	3683	298	309
Calls	**CHD**	**SCOOP**	**CHD**	**SCOOP**	**CHD**	**SCOOP**
Total	125728	181074	10807	16484	906	1451
Average per sample	214.5529	184.3931	18.4420	16.7862	1.5461	1.4776
	p = 6.74E-09		p = 1.11E-08		p = 0.2898	
CHD-ome	22.8481	19.8737	1.5375	1.4165	0.2389	0.2108
	p = 9.08E-08		p = 0.0578		p = 0.2694	
Ciliome	21.1297	18.2953	1.7918	1.6487	0.2201	0.2016
	p = 2.14E-06		p = 0.0526		p = 0.4587	
COSMIC CGC	10.3959	9.3391	0.6519	0.5448	0.0546	0.0550
	p = 8.72E-06		p == 0.0122		p = 0.9747	

Table 1 Continued

Minimum	53	79	5	3	0	0
Maximum	804	962	46	48	7	6
Standard Deviation	113.9552	65.2577	5.9050	4.7690	1.2885	1.1507
Variance	12985.7827	4258.5691	34.8693	22.7432	1.6603	1.3242
Samples with 0 calls	0	0	0	0	138	200
Samples with 1 call	0	0	0	0	180	359
Samples with 2+ calls	586	982	586	982	268	423
p-values are the 2-tailed p-values for the 2-sample t-test assuming unequal variance						

2.3.2 Enrichment for Rare Variants

Rare variants were those with a GnomAD allele frequency ≤ 0.01. Basic filtering had removed variants with a GnomAD allele frequency > 0.01 and with a CADD score < 10 leaving only rare (GnomAD allele frequency ≤ 0.01) damaging (CADD scores ≥ 10) variants. Overall, there were 7029 genes with a normalized allele count FC > 1 including 19 genes with variants only present in CHD cases while 5682 genes had an FC < −1 including 159 with variants present only in SCOOP controls. Of the genes with FC > 1, 37 were CHD100 genes, 539 were CHD-ome genes including two (*AGTR2*, *NSHDL*) with variants only present in CHD cases, and 504 Ciliome genes. Though a larger number of CHD100 genes had negative fold changes than positive fold changes (50 versus 37), fewer CHD-ome (468) and Ciliome (414) genes had negative FC than positive. However, all three gene sets had more genes that had variants present in SCOOP controls only than genes with variants present in CHD cases only. Of the 87 CHD100 genes present in the dataset, eight had FC ≥ 2 and none had FC ≥ 5 while 143/1007 CHD-ome genes and 133/918 Ciliome genes had FC ≥ 2. There were 20 CHD-ome and 16 Ciliome genes with FC ≥ 5.

Of the 1047 genes with an odds ratio greater than 1 and significant raw p-value, only 5 (*SMAD6*, *CCDC39*, *C50RF42*, *MEGF8*, and *EVL*) were CHD100 genes, but none remained significant after adjustment. For the CHD-ome 96 genes had significant p-values before

adjustment, but only five (*PEPD*, *NCOR2*, *ABCC3*, *CASQ2*, *G6PD*) remained significant after adjustment. Of the 1223 genes with a significant excess of alleles in CHD cases overall, there were six CHD100 genes (*SMAD6*, *CCDC39*, *C5ORF42*, *MEGF8*, *EVL*, *HECTD1*) and 101 CHD-ome genes with 21 of the latter group also being Ciliome genes (Appendix Table 2).

2.3.2.1 ToppGene Analysis

ToppGene enrichment analysis of the genes with FC ≥ 5 plus those with variants only in CHD cases returned few results, though there were multiple terms related to both G-protein-coupled receptors and olfaction. (Appendix table 3) "GPCR downstream signaling" also came up as a significant term when looking at the set of genes with homozygous calls only in CHD cases. Among the genes associated with the GPCR-related terms were four CHD-ome genes (*ACKR3*, *APLNR*, *HTR2B*, *S1PR1*) of which *ACKR3* has been reported as a candidate gene accounting for the TGA seen in patients with dup(2)(q37.3) (Costain et al., 2016). A corresponding analysis of the genes with FC ≤ −5 plus those with variants present only in SCOOP controls did not produce any significant results.

2.3.2.2 Ingenuity Pathway Analysis

For IPA analyses, the largest $-\log_{10}$(p-value) tended to be cancer terms for both sets as were the terms with the largest deltas. Though there were few bona fide structural CHD terms, significant terms related to the cardiovascular system tended to have positive deltas and encompassed a number of CHD-ome genes including *AGTR2*, *CASQ2*, *CFLAR*, *DAG1*, *EGFR*, *G6PD*, *HTR2B*, *MYLK2*, *S1PR1*, and *ZIC3*. Additionally, there were a number of terms related to apoptosis (Appendix table 4).

2.3.2.3 Analysis by curated phenotype

At the phenotype level, among the 13 phenotype terms with significant odds ratios after Benjamini-Hochberg FDR adjustment were four terms related to cardiomyopathy and one term related to catacholaminergic ventricular tachycardia. Among the phenotype terms with a significant excess of alleles in CHD cases were two terms for different forms of Loeys-Dietz syndrome, Atrioventricular septal defect 2, left ventricular noncompaction, 2 terms for nonspecific heart defects and the very general term 'cardiovascular phenotype'. Additionally, 'Hypoplastic left heart syndrome 2" (BH FDR = 0.0617) was close to significance. Finally, there were six significant cilia-related phenotypes (Appendix Table 5)

2.3.3 Enrichment for Rare Pathogenic Variants

Pathogenic variants were defined as those being curated as "(likely) pathogenic" by ClinVar and "DM" by HGMD. Only 231 genes had pathogenic variants in at least one of the cohorts, though this included 5 CHD100 genes, 35 CHD-ome genes, and 32 Ciliome genes of which zero, seven, and four had allele FC $\geq$ 2, respectively. Only six genes had significant odds ratios before adjustment, two of which had OR < 1, and no genes remained significant after adjustment. However, of the four genes with significant odds ratios > 1, three (*G6PD, ZIC3*, *TTC7A*) are CHD-ome genes. Those three genes plus another CHD-ome gene, *AIP*, were among the eight genes with a significant excess of alleles in CHD cases after FDR adjustment (Table 2). The association between *ZIC3* and heterotaxy spectrum CHD has been well documented (Andersen et al., 2014; Azhar & Ware, 2016; Cowan et al., 2014; De Luca et al., 2010; Deng, Xia, & Deng, 2015; Fahed et al., 2013; Finsterer, Stollberger, & Towbin, 2017; Hilger et al., 2015; Lalani & Belmont, 2014; Lebo & Baxter, 2014; A. H. Li et al., 2019; McCulley & Black, 2012;

Pierpont et al., 2007; Prendiville et al., 2014; Richards & Garg, 2010; Sutherland & Ware, 2009; Wessels & Willems, 2010; Xie et al., 2017). Mutation in the mouse ortholog of Tetratricopeptide repeat domain 7A (*TTC7A*) gene (*Ttc7*) have been associated with an enlarged heart or decreased heart weight (Beamer, Pelsue, Shultz, Sundberg, & Barker, 1995; Takabayashi & Katoh, 2005). Male mice hemizygous for a *G6pd* (Glucose-6-phosphate dehydrogenase) mutation have developmental arrest with distended hearts by E7.5 while heterozygous females are either normal or die by E12.5 with an abnormal placenta depending on the parental origin of the mutant X chromosome (Longo et al., 2002).

Table 2: Genes with significant enrichments for pathogenic variants

	Gene List			Allele Count			Odds Ratio		
Gene	**CHD100**	**CHD-ome**	**Ciliome**	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
G6PD	FALSE	TRUE	FALSE	14.4020	4.86E-11	5.61E-09	16.9706	0.0004	0.0809
NOBOX	FALSE	FALSE	FALSE	5.7608	5.81E-08	4.47E-06	6.7907	0.0009	0.0993
ZIC3	FALSE	TRUE	TRUE	5.0407	3.5E-05	0.0020	5.9455	0.0160	0.9042
TTC7A	FALSE	TRUE	FALSE	2.3403	0.0004	0.0170	2.7564	0.0234	1.0000
AIP	FALSE	TRUE	FALSE	5.7608	0.0004	0.0177	6.7807	0.0669	1.0000
ABCB4	FALSE	FALSE	FALSE	5.7608	0.0004	0.0177	6.7807	0.0669	1.0000
CLCN1	FALSE	FALSE	FALSE	5.7608	0.0004	0.0177	6.7807	0.0669	1.0000
CDHR1	FALSE	FALSE	FALSE	2.8804	0.0008	0.0269	3.3969	0.0671	1.0000

Gene List indicates if the gene is present in either of the 3 focus gene lists. For allele count, FC is based on normalized allele counts and the p-value is the upper tail probability of the binomial test. For OR, the p-value was obtained from Fisher's exact test

2.3.3.1 ToppGene Analysis

ToppGene analysis of the set of genes with pathogenic variants with allele FC ≥ 2 or with variants only present in CHD cases produced a number of significant terms related to structural CHD that were not significant when carrying out a ToppGene analysis of the set of genes with pathogenic variants with allele FC ≤ -2 or with variants present only in SCOOP controls (Table

3). However, most of the significant structural CHD terms (BH FDR = 2.22E-03) were associated with *ZIC3* alone. Supporting the link between cilia and CHD, the term "9+0 nonmotile cilium" (BH FDR = 3.36E-05) was significant only in the enriched set. Additionally, a number of terms related to pituitary adenoma, all associated with *AIP* (BH FDR = 2.22E-03) and multiple muscular dystrophy-related terms were also significant only in the enriched set. The terms with the smallest Benjamini-Hochberg q-values encompassed four CHD-ome genes (*AIP*, *C21ORF2*, F*KRP*, *ZIC3*) and three ciliome genes (*C21ORF2*, *FAM161A*, *ZIC3*). Mutations in *AIP* have been associated with DORV and other CHD in mice (B. C. Lin et al., 2007; Obler, Juraszek, Smoot, & Natowicz, 2008). In a mouse model of limb-girdle muscular dystrophy related to *FKRP* mutations, *P144Lneo*- mice were found to have left ventricular hypertrophy and dysfunction (Q. Yu et al., 2018). *C21ORF2* (*CFAP410*) was reported to be a candidate CHD gene in patients with del(21)(q22.13-q22.3) in a chromosome microarray study that mentioned previous evidence based on a mouse knockout (Shanshen, Rosenberg, & Van Bergen, 2018), however the article did not provide a citation or source for the mouse knockout and no corroborating evidence could be found.

Table 3: ToppGene analysis of genes enriched for pathogenic variants

Name	Genes in query	p-value	q-value FDR B&H
Becker muscular dystrophy	4 (ANO5, CLCN1, FKRP, ALDH7A1)	1.26E-06	5.14E-04
9+0 non-motile cilium	4 (CERKL, FAM161A, CDHR1, CFAP410)	8.36E-05	7.40E-03
Photoreceptor cell cilium	4 (CERKL, FAM161A, CDHR1, CFAP410)	7.88E-05	7.40E-03
Congenital heart defects, multiple types, 1, x-linked	1 (ZIC3)	2.22E-03	1.30E-02
Heterotaxy, visceral, 1, x-linked	1 (ZIC3)	2.22E-03	1.30E-02
Heterotaxy, visceral, 1, x-linked; htx1	1 (ZIC3)	2.22E-03	1.30E-02
Heterotaxy, visceral, X-linked	1 (ZIC3)	2.22E-03	1.30E-02
Multiple ventricular septal defects	1 (ZIC3)	2.22E-03	1.30E-02
VACTERL association, x-linked, with or without hydrocephalus	1 (ZIC3)	2.22E-03	1.30E-02
Obscure African cardiomyopathy	1 (CLCN1)	2.22E-03	1.30E-02

Table 3 Continued

Acromegaly due to pituitary adenoma 1	1 (AIP)	2.22E-03	1.30E-02
Adenocarcinoma of liver	1 (CYP24A1)	2.22E-03	1.30E-02
Pituitary adenoma 1, multiple types	1 (AIP)	2.22E-03	1.30E-02
Pituitary adenoma, prolactin-secreting	1 (AIP)	2.22E-03	1.30E-02
Prolactinoma, familial	1 (AIP)	2.22E-03	1.30E-02

2.3.3.2 Ingenuity Pathway Analysis

There were 24 significant cardiovascular system-related terms with positive deltas after IPA comparison analysis of the sets of genes with allele FC ≥ 2 and FC ≤ -2 for pathogenic variants, all of which were significant only in CHD cases (Table 4). These terms encompassed 11 genes of which four (*AIP*, *FKRP*, *G6PD*, *ZIC3*) are CHD-ome genes with *ZIC3* also being a ciliome gene. Additionally, IPA identified several significant disease and function networks of which the top two were "Cardiovascular System Development and Function, Cellular Assembly and Organization, Cellular Development" (Network 1) with 10 of 35 network molecules being in the enriched gene set and "Cardiovascular Disease, Inflammatory Response, Organismal Injury and Abnormalities" (Network 2) with 9 of 35 network molecules being in the enriched gene set (Figure 2). In the figure Network 1 connections are shown in black while Network 2 connections are shown in pink with molecules in the query gene set being shaded gray. There is a substantial amount of overlap between these two networks with a large number of connections appearing to converge on *TP53*, *CCND1*, *TGFB1*, and *TGFBR2*. Though none of these genes were in the query list, only *CCND1* is not a CHD-ome gene. TGFB1 is associated with vascular anomalies in humans and valve defects in mice (Mattassi et al., 2018; D. Srivastava & Olson, 2000) while TGFBR2 has been associated with a variety of phenotypes involving the great vessels as well as heterotaxy and VSD (Azhar & Ware, 2016; Deng et al., 2015; LaHaye et al., 2014; Lalani & Belmont, 2014; Lebo & Baxter, 2014; C. J. Lin et al., 2012; Mattassi et al., 2018; Neeb, Lajiness,

Bolanis, & Conway, 2013; Pierpont et al., 2018; Richards & Garg, 2010; Zheng et al., 2018). Mice with a TP53 knock-in allele ($p53^{25,26,53,54/+}$) have atrioventricular cushion and outflow tract defects caused by inappropriately triggering cell cycle arrest or apoptosis (Van Nostrand et al., 2014). Overexpression of *Ccnd1*, encoding Cyclin D1, in mice has been shown to result in abnormal multinucleate cardiomyocytes (Soonpaa et al., 1997). Pathogenic mutations to genes in the query list could thus alter the expression or transcriptional activity of these central genes leading to altered regulation of growth, proliferation, and apoptosis during heart development.

Table 4: IPA analysis of genes enriched for pathogenic alleles

Diseases and Bio Functions	GS5	GS5b	GS5b Genes
Inflammation of vessel	4.251281	4.21414	ABCB4, ADA2, PCSK9
Cardiac allograft vasculopathy	2.800796	2.788561	PCSK9
Inferior vena cava thrombosis	2.800796	2.788561	VWF
Quantity of trabecular myocardium	2.800796	2.788561	ZIC3
VACTERL with hydrocephalus, X-linked	2.800796	2.788561	ZIC3
X-linked multiple types congenital heart defects type 1	2.800796	2.788561	ZIC3
Inflammation of artery	2.738459	2.714131	ADA2, PCSK9
Closure of ductus venosus	2.500099	2.487875	AIP
Fragmentation of cardiac muscle	2.500099	2.487875	TRIM63
Persistence of ductus venosus	2.500099	2.487875	AIP
Structural integrity of m-bands	2.500099	2.487875	TRIM63
Structural integrity of myosin filaments	2.500099	2.487875	TRIM63
Rheumatic valvular heart disease	2.324342	2.312127	VWF
Degeneration of cardiomyocytes	1.803798	1.791651	HADHA
Abnormal morphology of myocardium	1.797615	1.774342	HADHA, TRIM63
Abnormal morphology of dilated vasculature	1.762738	1.750602	G6PD
Outgrowth of blood vessel	1.601675	1.589588	G6PD
Abnormal morphology of dilated heart	1.551188	1.53912	G6PD
Familial congenital heart disease	1.547929	1.525214	AIP, ZIC3
Contractility of cardiomyocytes	1.506096	1.494047	TRIM63
Dextrocardia	1.394146	1.382156	ZIC3
Primary cardiomyopathy	1.39616	1.373886	FKRP, TRIM63
Morphology of cardiovascular system	1.415838	1.369887	FKRP, G6PD, HADHA, TRIM63, VWF

Terms related to the cardiovascular system that were significant only in the CHD gene sets. GS5 and GS5b are the gene sets that were analyzed (Appendix C.1) and the numbers are $-\log_{10}$(p-values) generated from the IPA comparison analysis.

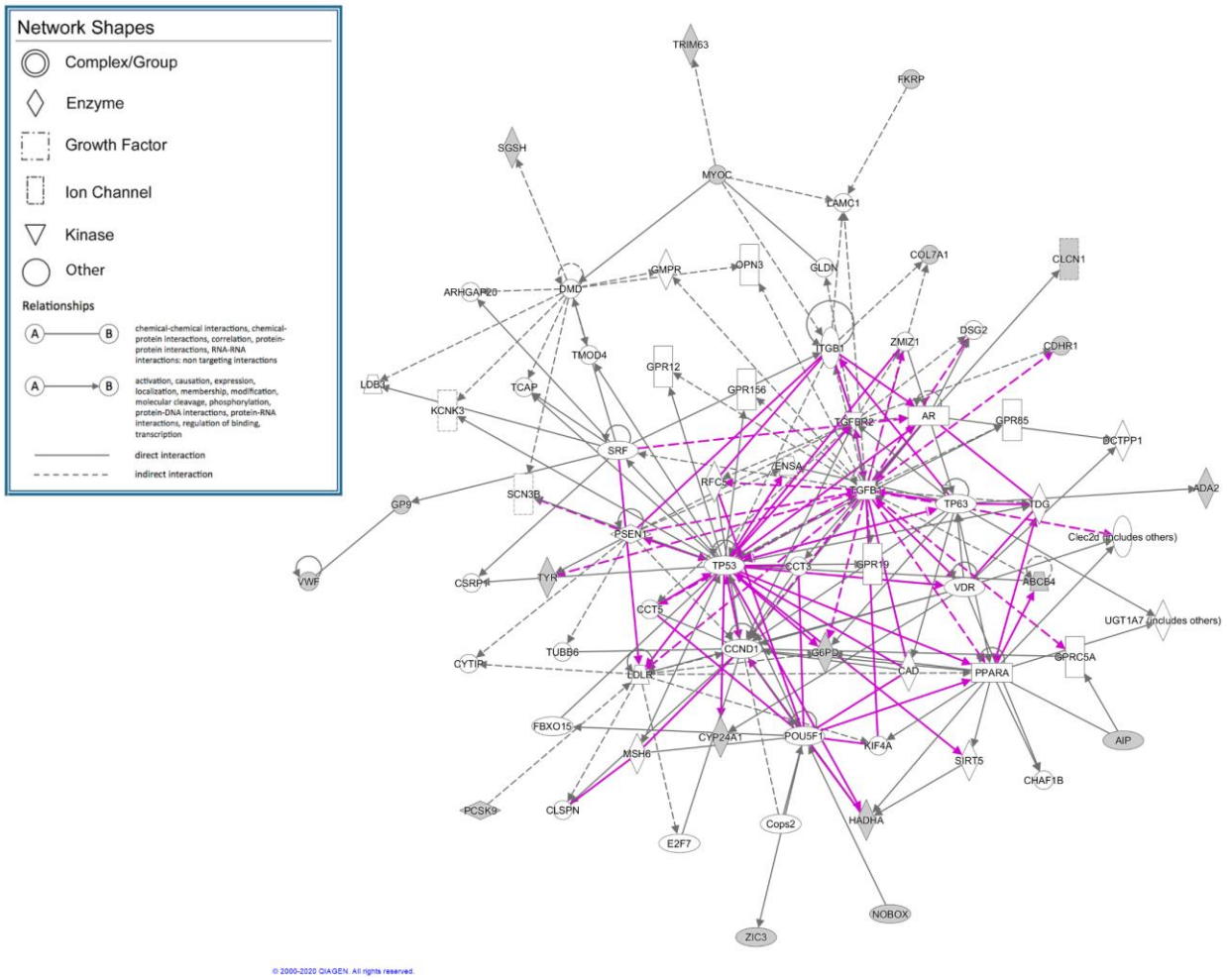


Figure 2: IPA network of genes enriched for alleles in pathogenic variants

2.3.3.3 Analysis by curated phenotype

At the phenotype level 32 phenotypes had significant raw p-values for odds ratio, but none remained significant after FDR adjustment, however 19 had a significant excess of pathogenic alleles in CHD cases after FDR adjustment. There were only two significant structural CHD terms (Table 5), “Cardiac malformation” and “Congenital heart defects 1 nonsyndromic 1”, plus one cilia-related term that can include CHD, “Heterotaxy visceral X-linked”, and two terms related to VATER/VACTERL association, all with a BH FDR of 0.0059 and associated with *ZIC3* alone.

The aforementioned two terms were the only terms out of 10 structural CHD terms to be significant, though "Noncompaction, left ventricular" had a significant excess of pathogenic alleles before adjustment and 8 of 10 terms had OR > 1. In addition to the 10 structural CHD terms, there were 26 terms for non-structural heart disease (cardiomyopathies, channelopathies), though none were significant. Overall, the two terms with the most significant excess of pathogenic alleles in CHD cases both related to glucose-6-phosphate dehydrogenase (G6PD) deficiency) (BH FDR = 2.86E-08). In fact, five of the 19 significant phenotypes related to G6PD with three additional phenotypes related to ovarian failure and associated with *NOBOX*.

Table 5: Significant phenotypes with pathogenic alleles

	Allele Count			Odds Ratio		
Phenotype (genes)	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
Glucose-6-phosphate dehydrogenase deficiency (G6PD)	14.4020	4.86E-11	2.86E-08	16.8251	0.0004	0.1880
Cardiac malformation (ZIC3)	5.0407	3.5E-05	0.0059	5.8944	0.0160	0.8129
Congenital heart defects 1, nonsyndromic, 1 (ZIC3)	5.0407	3.5E-05	0.0059	5.8944	0.0160	0.8129
Heterotaxy, visceral, X-linked (ZIC3)	5..0407	3.50E-05	0.0059	5.8944	0.0160	0.8129
VACTERL association with hydrocephaly, X-linked (ZIC3)	5..0407	3.50E-05	0.0059	5.8944	0.0160	0.8129
VATER/VACTERL association (ZIC3)	5..0407	3.50E-05	0.0059	5.8944	0.0160	0.8129
Noncompaction, left ventricular (HADHA)	2.8804	0.0228	0.9559	3.3556	0.5598	1.0000
For allele count, FC is based on normalized allele counts and the p-value is the upper tail probability of the binomial test. For OR, the p-value was obtained from Fisher's exact test						

2.3.4 Enrichment for Rare Highly Damaging Variants

Highly damaging variants were defined as those with a CADD score ≥ 30. Of the 2864 genes with highly damaging variants, 32 were CHD100 genes, 241 were CHD-ome genes, and 258 were Ciliome genes. There were 11 genes with highly damaging variants only in CHD cases, though none were CHD-ome genes and only one, *DNALI1*, is a Ciliome gene. Of the 32 CHD100 genes with highly damaging variants, 22 had allele count FC ≥ 2 compared to only seven with FC

≤ -2. Additionally, 36/241 CHD-ome and 183/258 Ciliome genes with highly damaging variants had FC ≥ 2. Though 64 genes had significant odds ratios before adjustment, none remained significant after FDR adjustment. Of the 93 genes with significantly more highly damaging alleles in CHD cases, four (*CCDC39*, *NRG1*, *SMAD6*, *TRIM32*) are CHD-ome genes with *CCDC39* and *TRIM32* also being ciliome genes (Table 6). A further nine Ciliome genes were also significant. Variants in *SMAD6* have been associated with primarily left-sided obstructive lesions in humans while causing a variety of valvular and great artery defects in mice (Andersen et al., 2014; Chaix et al., 2016; Fahed et al., 2013; LaHaye et al., 2014; Y. Li et al., 2015; C. J. Lin et al., 2012; X. Liu et al., 2017; Prendiville et al., 2014; D. Srivastava & Olson, 2000). *Nrg1* mutations in mice cause hypoplastic endocardial cushions and absent myocardial trabeculation (C. J. Lin et al., 2012; D. Srivastava & Olson, 2000). Mutations in *CCDC39* are a cause of primary ciliary dyskinesia with situs inversus or heterotaxy in humans while mouse mutants display AVSD or complex CHD with heterotaxy (Burnicka-Turek et al., 2016; Deng et al., 2015; Y. Li et al., 2015). Variants in *TRIM32* cause Bardet-Biedl syndrome with VSD or dextrocardia in humans (Lalani & Belmont, 2014).

Table 6: Genes significantly enriched for alleles in variants with CADD ≥ 30

	Gene List			Allele Count			Odds Ratio		
Gene	**CHD100**	**CHD-ome**	**Ciliome**	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
CCDC39	TRUE	TRUE	TRUE	8.6412	5.02E-10	3.59E-07	10.2365	0.0003	0.1969
EFCAB6	FALSE	FALSE	TRUE	11.5216	1.51E-08	6.17E-06	13.6083	0.0022	0.4503
MOGAT2	FALSE	FALSE	TRUE	6.4809	5.41E-07	0.0001	7.6315	0.0033	0.4503
PKD1L2	FALSE	FALSE	TRUE	6.4809	5.41E-07	0.0001	7.6315	0.0033	0.4503
SMAD6	TRUE	TRUE	FALSE	4.3206	1.39E-06	0.0003	5.1130	0.0029	0.4503
ACAD10	FALSE	FALSE	TRUE	8.6412	3.13E-06	0.0006	10.1886	0.0125	0.8685
TRIM32	FALSE	TRUE	TRUE	7.2010	3.73E-05	0.0045	8.4831	0.0293	1.0000
KIF14	FALSE	FALSE	TRUE	7.2010	3.73E-05	0.0045	8.4831	0.0293	1.0000
TARS	FALSE	FALSE	TRUE	5.0407	3.5E-05	0.0046	5.9455	0.0160	0.8685
NRG1	FALSE	TRUE	FALSE	3.8405	0.0001	0.0106	4.5315	0.0240	1.0000
DNAH1	FALSE	FALSE	TRUE	2.2258	0.0002	0.0160	2.6180	0.0170	0.8785

Table 6 Continued

LRRC48	FALSE	FALSE	TRUE	3.2405	0.0002	0.0174	3.8248	0.0214	1.0000
CCDC170	FALSE	FALSE	TRUE	5.7608	0.0004	0.0296	6.7807	0.0669	1.0000
Gene List indicates if the gene is present in either of the 3 focus gene lists. For allele count, FC is based on normalized allele counts and the p-value is the upper tail probability of the binomial test. For OR, the p-value was obtained from Fisher's exact test									

2.3.4.1 ToppGene Analysis

ToppGene analysis of the set of genes with a highly damaging variant allele FC ≥ 2 or with highly damaging variants only in CHD cases produced few results, though three terms were related to fibronectin and two were related to EF-hand domains (Appendix Table 6). Genes encompassed by these terms included eight CHD-ome genes (*ARHGEF4*, *FLT4*, *MYBPC3*, *NRG1*, *RYR2*, *TRIM32*, *TSC2*, *ZFYVE16*) of which two (*RYR2*, *TRIM32*) are also Ciliome genes. *ARHGEF4* has been identified as a candidate gene accounting for the TOF seen in del(2)(q22.1) patients (Silversides et al., 2012). Mutations in *FLT4* have been linked to TOF and vascular anomalies in humans with the gene also being a candidate CHD gene for copy number variants at 5q35.3 as well as being associated with vascular remodeling defects in mice (Dumont et al., 1998; Mattassi et al., 2018; Reuter et al., 2018; Xie et al., 2017). *MYBPC3* variants have been linked to left ventricular noncompaction as well as atrial and ventricular septal defects (Lebo & Baxter, 2014; van Waning et al., 2018; Wessels & Willems, 2010). *RYR2* has also been associated with left ventricular noncompaction (van Waning et al., 2018). In addition to causing tuberous sclerosis, *TSC2* variants have been associated with left ventricular obstruction defects (Jin et al., 2017). Finally, *ZFYVE16* mutations have been potentially associated with situs anomalies (A. H. Li et al., 2019) These analyses were also done with an FC ≥ 3 to reduce the number of genes, but these analyses produced no results in ToppGene. Additionally, an earlier analysis that did not include

genes with highly damaging variants only in CHD cases produced some additional results related to muscle fibers in the FC ≥ 2 set.

2.3.4.2 Ingenuity Pathway Analysis

Ingenuity Pathway Analysis of the FC ≥ 2 (GS9 and GS9c) and FC ≥ 3 (GS9b and GS9d) sets both produced many terms related to the cardiovascular system, though there was virtually no overlap between significant terms for each set (Table 7). Significant terms for the FC ≥ 2 set tended to be more general or related to cardiomyopathies or channelopathies while significant terms for the FC ≥ 3 set tended to be more specific with terms related to the aortic valve and atrioventricular cushions. There were also a large number of significant terms related to the cardiovascular system for the FC ≤ −2 and FC ≤ −3 sets, but these terms had no overlap with the positive FC sets. The cardiovascular system-related terms with positive deltas encompassed 39 genes of which 14 (*ABI1*, *BCAR1*, *DTNA*, *FLT4*, *HAND1*, *HEYL*, *MYBPC3*, *NRG1*, *PKP2*, *PPARGC1B*, *RYR2*, *SMAD6*, *TNNI1*, *TSC2*) were CHD-ome genes. *ABI1* variants have been associated with left ventricular outflow tract obstruction with extracardiac defects (Jin et al., 2017). Some cases of TOF have been linked with *BCAR1* (Reuter et al., 2018). Variants in *DTNA* have been associated with left ventricular noncompaction, persistent left superior vena cava, PDA, and HLHS (Lalani & Belmont, 2014; Lebo & Baxter, 2014). *HAND1* mutations have been associated with a variety of defects in both humans and mice including septal defects, DORV, and HLHS (Azhar & Ware, 2016; Firulli et al., 1998; L. Li et al., 2017; McCulley & Black, 2012; McFadden et al., 2005; Togi et al., 2004; Wessels & Willems, 2010). Global deletion of *Hey1* and *HeyL* in mice results in VSD as well as atrioventricular and semilunar valve defects (C. J. Lin et al., 2012). *Pkp2*$^{-/-}$ mice have hearts with thin-walled atria, reduced trabeculations, and blood leakage into the pericardium (Grossmann et al., 2004). *Ppargc1a*$^{-/-}$*/Ppargc1b*$^{-/-}$ double mutant mice have small

hearts with reduced output and arrest of mitochondrial biogenesis during the late fetal period (L. Lai et al., 2008). Finally, *TNNI1*, which was curated as a CHD-ome gene through GO search, does not appear to have any structural phenotype but may be associated with sudden arrhythmic cardiac death or Sudden Infant Death Syndrome in humans (Shafaattalab et al., 2019). Finally, the term "Apoptosis of endothelial progenitor cells", associated with *NRG1*, was significant only the sets that included genes with highly damaging variants only in the CHD set (GS9b, GS9d). "Apoptosis of microvascular endothelial cells", associated with *CD36*, *COMP*, and *ITSN1*, none of which are CHD-ome genes, was significant in all 4 CHD genes sets but none of the SCOOP gene sets.

Table 7: IPA analysis of genes with highly damaging variants

Diseases and Bio Functions	GS9	GS9b	GS9c	GS9d	GS9c-9d genes
Abnormal morphology of myocardium	1.8009	0	1.7516	0	ABI1, BCAR1, HADHA, MYBPC3, PKP2, PPARGC1B, RYR2
Morphology of myocardium	2.0519	0	1.9954	0	ABI1, BCAR1, HADHA, MYBPC3, PKP2, PPARGC1B, RYR2, TSC2
Failure of heart	1.8435	0	1.7656	0	CACNA1S, CD36, CYP2C19, FLT4, GRIN3B, HAND1, MYBPC3, MYOM1, PDE4C, PPARGC1B, PTGIR, RASSF1, RYR2, XDH
Dysfunction of heart	1.8527	0	1.7857	0	CACNA1S, CD36, DMD, DTNA, LAMA4, MYBPC3, PKP2, PTGIR, RYR2, SEMA3A, XDH
Left heart disease	1.9157	0	1.8519	0	CACNA1S, DMD, DTNA, HAND1, LAMA4, MYBPC3, PKP2, PTGIR, RYR2, XDH
Degeneration of heart	1.9544	0	1.9206	0	DTNA, HADHA, LAMA4, MYBPC3
Abnormality of atrium	1.7360	0	1.6829	0	DTNA, HDAC7, LAMA4, MYBPC3, PDE4C, PKP2, SEMA3A, SIGMAR1
Familial left ventricular noncompaction	3.4085	0	3.3706	0	DTNA, MYBPC3, PKP2, RYR2
Morphology of cardiovascular tissue	1.8770	0	1.8321	0	FLT4, HDAC7, NPHS2, S1PR3, SLC4A11, TKT
Degeneration of cardiomyocytes	1.7388	0	1.7199	0	HADHA, LAMA4
Morphogenesis of cardiac muscle tissue in heart ventricle	2.5544	0	2.5184	0	HAND1, MYBPC3, PKP2, TNNI1
Congenital heart defects, dysmorphic facial features and intellectual developmental disorder	0	2.2113	0	2.1746	CDK13
Morphogenesis of aortic valve	0	2.0594	0	1.9890	HEYL, SMAD6
Apoptosis of endothelial progenitor cells	0	1.9116	0	1.8750	NRG1

Table 7 Continued

Expansion of atrioventricular canal cushion	0	1.9116	0	1.8750	NRG1
Expansion of ventricular compact zone	0	1.9116	0	1.8750	NRG1
Loeys-Dietz syndrome type 3	0	1.9116	0	1.8750	SMAD6
Congenital heart disease with extracardiac congenital anomalies	0	1.6634	0	1.5951	B4GALT7, CDK13
Aortic valve disease type 2	0	1.9116	0	1.8750	SMAD6
Bardet-Biedl syndrome type 11	0	2.2113	0	2.1746	TRIM32
Meckel syndrome type 12	0	2.2113	0	2.1746	KIF14
Primary ciliary dyskinesia type 14	0	1.9116	0	1.8750	CCDC39
Terms related to the cardiovascular system that were significant only in the CHD gene sets. GS9-9d are the gene sets that were analyzed (Appendix C.1) and the numbers are $-\log_{10}$(p-values) generated from the IPA comparison analysis.					

An IPA analysis of the set of genes with highly damaging variants with homozygous calls only present in the CHD dataset produced a number of terms related to heart development that were only significant in CHD cases when compared to the set of genes with homozygous calls in highly damaging variants only present in SCOOP controls (Table 8). Of the four genes associated with the terms in Table 8, two (*DOCK6*, *NRG1*) are CHD-ome genes while *ANK3* is a ciliome gene. Mutations in *DOCK6* are a cause of Adams-Oliver syndrome which can include CHD (Hassed, Li, Mulvihill, Aston, & Palmer, 2017).

This analysis also produced two overlapping disease and function networks related to the cardiovascular system (Figure 3). These were "Cardiovascular System Development and Function, Cell Death and Survival, Skeletal and Muscular Disorders" (Network 1, black lines) for which 16 of 35 network molecules were in the input dataset and "Cardiovascular System Development and Function, Cell Death and Survival, Cell Morphology" (Network 2, pink lines) for which one of two molecules in the network were in the input dataset (*MRPL13*). These networks overlap and appear to converge on *CTNNB1*, *CDKN1A*, *VEGFA*, and the *ERK1/2* complex. *CTNNB1* and *VEGFA* are integral components of the WNT/β-catenin and VEGF signaling pathways respectively and mutations in these genes are both known to cause a variety of heart defects in

humans and mice (Fahed et al., 2013; C. J. Lin et al., 2012; Lincoln & Garg, 2014; Neeb et al., 2013; Prendiville et al., 2014; Reuter et al., 2018; Wessels & Willems, 2010). Some patients with an ~1 Mb deletion in distal 22q11.2 that includes the *ERK2* (*MAPK1*) gene show VSD, PTA, and RAA while conditional knockout in mouse neural crest causes conotruncal defects (Newbern et al., 2008). Loss of *ERK1* (*MAPK3*) has been proposed to underlie the TOF seen in patients with dup(16)(p11.2) (Silversides et al., 2012). An MGI query for *Cdkn1a*, encoding cyclin dependent kinase inhibitor 1A, indicated that mutations in this gene are associated with "abnormal heart morphology" based on data uploaded from the International Mouse Phenotyping Consortium (IMPC) (Dickinson et al., 2016). This points to a model in which homozygous damaging mutations in query list genes alter transcriptional activity of core genes to cause CHD.

Table 8: IPA analysis of genes with homozygous calls in highly damaging variants only in the CHD cohort

Diseases and Bio Functions	GS11	GS11 genes
Adams-Oliver syndrome type 2	2.6636	DOCK6
Apoptosis of endothelial progenitor cells	2.3631	NRG1
Adams-Oliver syndrome type 1	2.3631	DOCK6
Expansion of atrioventricular canal cushion	2.3631	NRG1
Expansion of ventricular compact zone	2.3631	NRG1
Trabeculation of heart ventricle	2.1874	NRG1
Dilated cardiomyopathy 3B	2.0629	DMD
Differentiation of endocardial cells	1.8878	NRG1
Apoptosis of ventricular myocytes	1.8213	NRG1
Contraction of cardiac muscle	1.7509	ANK3, DMD
Myocardial fibrosis	1.6678	DMD
Terms significant only in the CHD gene set. G11 is the analyzed gene set (Appendix C.1) and the numbers are -$\log_{10}$(p-values) generated from the IPA comparison analysis.		

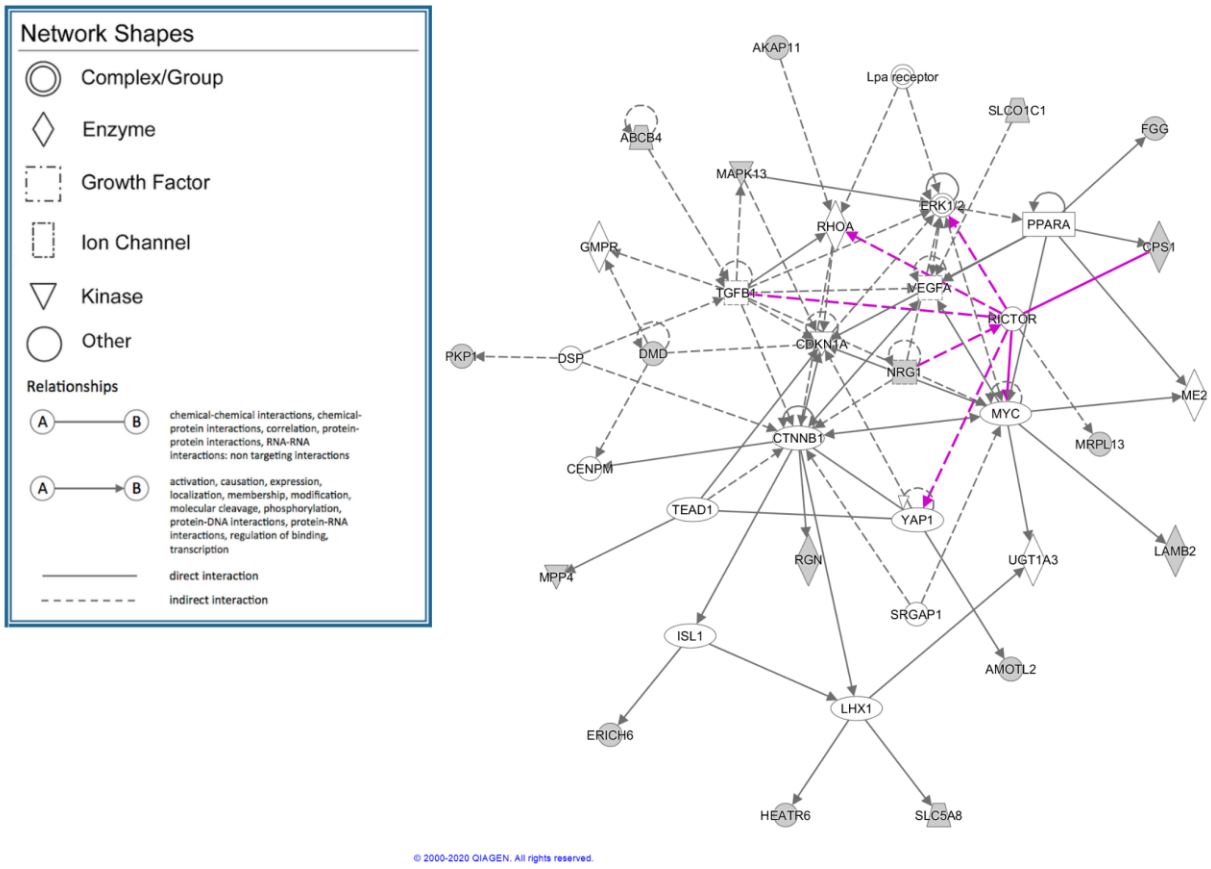


Figure 3: IPA network of genes with homozygous calls in highly damaging variants only in the CHD cohort

2.3.4.3 Analysis by curated phenotype

At the phenotype level, 40 phenotypes had a significant raw p-value for odds ratio including 31 with OR > 1, however none of them remained significant after FDR adjustment. There were 291 phenotypes with a significant excess of highly damaging alleles in CHD cases before adjustment which was reduced to 34 after adjustment. The majority of the most significant terms were not bona fide disease terms, but rather terms indicating "susceptibility" or "association with". In fact, 12 of these terms were all associated with a single *CD36* variant. Though there were 35 structural CHD terms with highly damaging alleles, only 10 of these had an OR > 1 and there was a significant excess of highly damaging alleles in only 7 before adjustment (Table 9) and one, "Loeys-Dietz syndrome 3" (FDR = 0.0006) after adjustment. This term was associated

with a single variant in *SMAD6*. Additionally, there were two cilia-related terms with a significant excess of highly damaging alleles; "Bardet-Biedl syndrome 11" (FDR = 0.0074) associated with *TRIM32* and "Ciliary dyskinesia, primary 37" (FDR = 0.0195) associated with *DNAH1*. Of those two genes, TRIM32 is a CHD-ome gene. Finally, 'Zellweger syndrome' associated with *PEX6* also showed an excess of highly damaging alleles. Variants in *PEX6* have also been associated with TOF (Grunert et al., 2014).

Table 9: Phenotypes Enriched for Alleles in Highly Damaging Variants

	Alleles			Odds Ratio		
Phenotype (genes)	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
Loeys-Dietz syndrome 3 (SMAD6)	4.3206	1.39E-06	0.0006	5.0690	0.0030	0.3107
Bardet-Biedl syndrome 11 (TRIM32)	7.2010	3.73E-05	0.0074	8.4105	0.0300	1.0000
Ciliary dyskinesia, primary 37 (DNAH1)	2.2258	0.0002	0.0195	2.5956	0.0172	1.0000
Zellweger syndrome (PEX6)	5.7608	0.0004	0.0380	6.7226	0.0682	1.0000
Hypoplastic left heart syndrome 1 (HAND1)	2.8804	0.0075	0.3087	3.3596	0.2048	1.0000
Left ventricular noncompaction 1 (DTNA))	2.8804	0.0075	0.3087	3.3596	0.2048	1.0000
Left ventricular noncompaction 10 (MYBPC3)	2.8804	0.0075	0.3087	3.3596	0.2048	1.0000
Thoracic aortic aneurysm, phenotype modifier (ADCK4)	1.9203	0.0112	0.3704	2.242	0.1654	1.0000
Marfan/Thoracic aortic aneurysms and dissections (CBS)	1.8003	0.0347	0.7555	2.0994	0.3080	1.0000
Noncompaction, left ventricular (ACADVL, HADHA, LAMA4)	1.9203	0.0357	0.6895	2.2358	0.4362	1.0000
For allele count, FC is based on normalized allele counts and the p-value is the upper tail probability of the binomial test. For OR, the p-value was obtained from Fisher's exact test						

2.4 Discussion

Though a variety of CHD-ome genes showed significant excesses of alleles for different analyses, there were several genes that seemed to pop up frequently. As previously mentioned, *SMAD6* and *ZIC3* are already well-known causes of CHD. *SMAD6* is one of two inhibitory SMAD

(I-SMAD) proteins along with *SMAD7* that act as inhibitors of TGF-β signaling through a variety of mechanisms (Reviewed in Miyazawa & Miyazono, 2017). It acts to repress BMP signaling during normal development of the heart valves and outflow tract. *ZIC3* encodes a multifunctional zinc finger transcription factor that can interact with components of multiple developmental signaling pathways including Hedgehog, Wnt, Planar Cell Polarity, and TGF-β (Reviewed in Bellchambers & Ware, 2018).

Mutations in *C21ORF2*, also known as *CFAP410* (Cilia and flagella associated protein 410), have been linked with several ciliopathy phenotypes including cone/rod dystrophy, axial spondylometaphyseal dysplasia, and Jeune asphyxiating thoracic dystrophy (McInerney-Leo et al., 2017; Reiter & Leroux, 2017). Shanshen et al. (2018) identified this gene as a CHD candidate gene in patients with del(21)(q22.13-q22.3) based on genetic analysis of patients and cross-referencing with the Online Mendelian Inheritance in Man (OMIM) database which they say indicted a mouse knockout model. However, they did not explicitly cite the source and I was unable to find any evidence for a mouse knockout model with a cardiac phenotype through multiple searches, so this genes inclusion in the CHD-ome may be dubious. Data suggests this gene plays a role in cilia formation and/or maintenance (C. K. Lai et al., 2011). Like *ZIC3* and *C21ORF2*, *CCDC39* (Coiled-coil domain containing 39) is a ciliome as well and CHD-ome gene but is also present in the CHD100 set and has been associated with PCD, situs inversus, and heterotaxy in humans while mouse mutants have displayed AVSD, complex CHD, and heterotaxy (Burnicka-Turek et al., 2016; Deng et al., 2015; Y. Li et al., 2015). *CCDC39* is necessary for proper motion of motile cilia as it plays an essential role in the construction of inner dynein arms and dynein regulatory complexes (Merveille et al., 2011). *FKRP* (Fukutin related protein) mutations cause dystroglycanopathy/muscular dystrophy in humans with left ventricular

noncompaction and ventricular dysfunction observed in a mouse model (Q. Yu et al., 2018). The *FKRP* protein in involved in glycosylation of alpha-dystroglycan which is essential for proper sarcolemma function (Brockington et al., 2001). Neuregulin 1 (*NRG1*) encodes an epidermal growth factor (EGF) family ligand and when mutated in mice results in absent ventricular trabeculation and endocardial cushion defects with deficient mesenchymal cells (Lee et al., 1995; C. J. Lin et al., 2012; Meyer & Birchmeier, 1995; D. Srivastava & Olson, 2000). Its Erb-b2 receptor tyrosine kinase binding partners, *ERBB2*, *ERBB3*, and *ERBB4*, are themselves CHD-ome genes which similarly cause absent trabeculation or cushion hypoplasia when mutated in mice (Lee et al., 1995; Meyer & Birchmeier, 1995)

The significant enrichment for *G6PD* variants is interesting. There are conflicting reports on whether deficiency of *G6PD*, which catalyzes the rate-determining step in the pentose phosphate pathway and synthesizes NADH, increase the risk for cardiovascular disease or have a protective effect (Reviewed in Hecker, Leopold, Gupte, Recchia, & Stanley, 2013). Some mouse studies have indicated a protective effect through decreases in cholesterol synthesis, superoxide production, and reductive stress while other studies have indicated that mutant mice are more susceptible to ventricular dilation post-myocardial infarct or pressure overload and that they have a poorer recovery from ischemia-reperfusion injury. In mice, hemizygous male embryos developmentally arrest by E7.5 and have distended hearts while heterozygous females are either normal or die by E12.5 with a placentation defect depending on parental inheritance of the mutant X chromosome. It could be interesting to investigate whether patients harboring a *G6PD* mutation had a higher incidence of adverse outcome than patients without a mutation. At the very least, *G6PD* status could potentially be used as a prognostic factor to identify patients at higher risk of adverse outcome.

ToppGene analyses produced several terms related to G-protein coupled receptor (GPCR) activity and signaling and though the majority of genes associated with those terms tended to be related to olfaction, they did include four CHD-ome genes. *ACKR3* (Atypical chemokine receptor 3) has been identified as a candidate gene for the dextro-transposition of the great arteries seen in patients with dup(2)(q37.3) (Costain et al., 2016). This gene, also known as *Cxcr7*, encodes a chemokine receptor that plays a key role in valve development and has been shown to cause semilunar valve defects with occasional malalignment of the aorta when mutated in mice (Sierro et al., 2007; S. Yu, Crawford, Tsuchihashi, Behrens, & Srivastava, 2011). Mutations in *Aplnr*, encoding the Apelin receptor, have been associated with a variety of vasculogenesis and cardiac defects in mouse mutants (Charo et al., 2009; Freyer et al., 2017; Kang et al., 2013). *Htr2b*, encoding the 5-hydroxytryptamine (serotonin) receptor 2B, has been linked to reduced ventricular trabeculation and weight and disorganized sarcomeres when mutated in mice while overexpression results in cardiac hypertrophy (Nebigil et al., 2000; Nebigil et al., 2003). In mice, mutations in the sphingosine-1-phophate receptor 1 (*S1pr1*) gene result in defective formation and patterning of the aorta (Allende, Yamashita, & Proia, 2003; Gaengel et al., 2012). A recent review (Nebigil & Desaubry, 2019) discusses the importance of GPCRs in epicardial epithelial-to-mesenchymal and mesenchymal-to-epithelial transformations as well as noting that these pathways could be therapeutic targets. Additionally, GPCRs have a major role in cilia signaling including the modulation of Hedgehog signaling (Reviewed in Mykytyn & Askwith, 2017).

That the genes and pathways that came up frequently included such a wide range of functions including multiple signaling pathways, cilia formation and maintenance, and sarcomere stability is in agreement with the highly complex nature of heart development (Bruneau, 2013; Klena et al., 2017; Sylva et al., 2014; Williams et al., 2019). The fact that a large number of CHD-

ome and Ciliome genes were both enriched and depleted in CHD cases could be interpreted multiple ways. First, it could be that the genes that are enriched in our cases are playing more of a role in the etiology of CHD, at least among our cohort. Second, it could be that the lack of rare pathogenic or damaging variants in the genes that are depleted in CHD cases could have a protective effect with wildtype alleles of the depleted genes acting to offset some of the deleterious effects of the variants in the enriched genes. Though only eight genes showed a significant excess of pathogenic alleles in CHD cases, the fact that four of those genes were CHD-ome genes seems significant. Additionally, eight of the 37 genes with a significant excess of alleles in CHD cases before adjustment were CHD-ome genes. This agrees with the hypothesis that CHD patients are enriched for pathogenic variants in genes associated with heart development.

Though there were far more genes with highly damaging variants than with pathogenic variants, only two CHD100 and an additional two CHD-ome genes had a significant excess of highly damaging alleles in CHD cases after adjustment. The small proportion of significant CHD-ome genes compared to the total number of significant genes could indicate that highly damaging variants in genes involved in heart development are not well tolerated. Though very few structural CHD phenotypes were significant, 8/10 structural phenotypes with pathogenic allele had odds ratios > 1 compared to only 10/32 with highly damaging variants. This again suggests variants associated with CHD tend to be less damaging in CHD patients. Overall, these observations may suggest a complex genetic model with multiple components. CHD patients may carry moderately damaging variants in multiple genes affecting multiple pathways involved in heart development. Individually, these variants may not cause disease, but together they can significantly perturb critical signaling pathways to the point of causing abnormal development. If even one of these moderately damaging variants were to be replaced with a highly damaging variant, then this would

amplify the disruptions in key developmental pathways to the point they would no longer be compatible with life. Conversely, healthy controls with highly damaging variants would have fewer damaging variants in other genes involved in heart development such that other partially redundant pathways would be able to sufficiently compensate for the single highly damaging variant and allow normal development. Protective alleles in other genes that lessen the impact of damaging variants may also play a role on both sides of this model.

That CHD patients showed significant difference in the mean number of calls in damaging variants overall and a difference in the mean number of calls in damaging variants in CHD-ome and Ciliome genes that was trending towards significance compared to SCOOP controls while the means number of calls in curated pathogenic variants between cases and controls were not significant across the board seems to add another wrinkle to this model. It would be necessary to conduct a more detailed analysis to determine if certain genes are driving the significance at the global level. As there was a large number of genes with highly damaging variants, even if very few genes showed a significant excess of alleles in CHD patients at the gene level, the cumulative effect of many genes with a non-significant excess of alleles may be enough to drive the global difference in the mean number of calls toward significance. The small number of genes with curated pathogenic variants may have prevented such a phenomenon from occurring in this category.

This study has several limitations. Firstly, our study population is not a random sample of consecutive CHD patients but shows a bias towards heterotaxy spectrum CHD and HLHS. This may lead to overrepresentation of variants associated with these conditions and underrepresentation of variants associated with other types of CHD. Additionally, our study population contains a mix of isolated and syndromic CHD cases which may have different genetic

etiologies as has previously been demonstrated (Sifrim et al., 2016). Specifically, our study population contained several patients each with Down syndrome and DiGeorge syndrome. This study is also limited by the fact that the exome data is not controlled for population stratification. Though the racial makeup of the CHD cases is known, the racial makeup of the SCOOP controls was not immediately available and, based on the distribution of calls by range, there appears to be a significantly higher percentage of non-whites among the CHD cases than the SCOOP controls. This would result in inflating the apparent enrichment for variants that have a higher allele frequency in non-whites. In the absence of control for race, it would be beneficial to perform a statistical test to estimate to what extent data from non-white cases is influencing the results.

It is also possible that my definition of pathogenicity was too stringent as indicated by the fact that there were so few genes with "pathogenic" alleles defined as those curated as both "(Likely) pathogenic" by ClinVar and "DM" by HGMD. There were a large number of variants that were curated as either "(Likely) pathogenic" or "DM", but not both, and these were left out of this part of the analysis. Using an "or" definition of pathogenicity such that variants that were curated as "(Likely) pathogenic" by ClinVar" or "DM" by HGMD would have included a much larger number of variants in the analysis (Appendix Table 2). Another issue in this area is that the phenotypes were only those that have been associated with the variants in our dataset. There are certainly other phenotypes associated with variants that were not present in the dataset. For example, though some CHD100 and CHD-ome genes were not associated with any cardiac phenotypes in our dataset, they could very well be known to cause heart defects in humans, but no variants curated as such were present in our dataset.

There were also issues with the data analysis methods. The initial analysis only calculated fold changes in normalized allele counts and the cutoffs for generating the gene lists to be run in

ToppGene and IPA were chosen arbitrarily to obtain lists that were not too long or too short. It might have made more sense to have calculated odds ratios and then generated the lists for enrichment analysis based on those genes with significant odds ratios. However, though many genes had significant raw p-values after carrying out Fisher's exact test, implementing a Benjamini-Hochberg false discovery rate eliminated virtually all of them. This was likely due to the very large difference in number of mutant alleles vs. reference alleles in the study cohorts. This method also could not take zygosity into account. No adjustment was made for hemizygosity of X chromosome genes in males so that hemizygous males were treated as homozygous with two alleles. Additionally, our study cohort had several patients each with trisomy 21 and monosomy X which were also not taken into account.

Ingenuity Pathway Analysis also produced some difficulties. Though this software can be used for enrichment analysis, it is more designed to look at expression data. Though there are a large number of data sources that gene sets are queried against which are all selected by default, when they are left selected the "Disease and Function" output will be almost entirely cancer-related terms. Leaving the "Cell Lines" option selected in the "Tissues and Cell Lines Activated" section will probably also contribute to the excess of cancer-related terms as the cell lines are almost exclusively cancer cell lines. Further compounding this issue is that outputs seem to be limited to 500 terms. The parameters for these analyses would thus have to be carefully selected so as not to overload the output with cancer-related terms. Despite these limitations, this research provides additional support for the complex genetic etiology of congenital heart disease as well as a framework on which to base future research.

3.0 Cancer Genes in CHD Patients

3.1 Introduction

Down syndrome is not only the condition most frequently associated with CHD which is present in about 50% of cases (Reviewed in Versacci et al., 2018), but also has a well-documented association with leukemia (Botto et al., 2013; Fisher et al., 2012; Mili, Khoury, Flanders, & Greenberg, 1993; Mili, Lynch, Khoury, Flanders, & Edmonds, 1993; Nishi et al., 2000; Windham, Bjerkedal, & Langmark, 1985). The links between developmental genes and cancer have also been reviewed (Bellacosa, 2013; Moore, 2009). Some of the same complex regulatory networks critical for embryogenesis may also play roles in the later development of cancer, and their role may exceed that of genes involved in later fetal growth and maturation (Moore, 2009). Among these developmental pathways and regulators are the Hedgehog and Wnt/β-catenin signaling pathways, the Polycomb repressive complex, and other chromatin modifiers and epigenetic factors. Additionally, multiple classes of disease have been associated with both developmental defects and increased cancer risk (Bellacosa, 2013). Among these disease classes are the cohesinopathies such as Cornelia-de Lange syndrome, the RASopathies such as Noonan syndrome, the phakomatoses such as Neurofibromatosis, and the overgrowth syndromes such as Beckwith-Wiedemann syndrome. It has been hypothesized that childhood cancers could arise from perturbations of the same developmental pathways that result in birth defects (Botto et al., 2013). Similarly, it is thought that pediatric solid tumors arise from mutations in early development genes while hematopoietic malignancies result from mutations in genes active later in development (Narod, Hawkins, Robertson, & Stiller, 1997). Fisher et al. (2012) proposed a similar hypothesis

while also theorizing that leukemias mainly arise from single gene defects that alter clonal leukocyte proliferation.

Numerous studies have documented the relationship between birth defects and childhood cancer. A statistically significant relationship was first recognized at least 35 years ago with children with birth defects also noted to develop cancer at a younger age than those without birth defects (Windham et al., 1985). A large-scale literature review has documented a general positive relationship between birth defects and cancer as well as relationships between specific birth defects and cancer types (K. J. Johnson et al., 2017). Additional studies noted that patients with malformations tended to develop cancer more frequently as a whole (Nishi et al., 2000; Norwood et al., 2017). A pair of related studies found that children with birth defects are significantly more likely to develop cancer before the age of 15 (Mili, Khoury, et al., 1993; Mili, Lynch, et al., 1993). These studies found an excess of leukemia, brain tumors, neuroblastoma, Wilms tumor, and retinoblastoma. Another study found that children with minor birth defects are more likely to have cancer and, conversely, that children with cancer are more likely to have minor birth defects (Durmaz et al., 2011). Botto, Flood, et al. (2013) reported that children with birth defects had a 2.9-fold increased risk of cancer compared to a control population without birth defects. Finally, it has been reported that children with non-chromosomal birth defects are at increased risk for solid tumors including lymphomas, central nervous system tumors, neuroblastoma, and non-central nervous system germ cell tumors, but not leukemias (Fisher et al., 2012).

A Canadian study reported a 1.6-2 times higher prevalence of cancer in CHD patients compared to the general population while a study from Western Australia reported a cancer hazard ratio of 1.74 among children with cardiovascular defects (Dawson, Charles, Bower, de Klerk, & Milne, 2015; Gurvitz et al., 2016). It has been reported that children with Wilms tumor have a

higher incidence of TGA, aortic stenosis, and VSD than the general population while an excess of cardiac septal defects has also been reported in children with several types of cancer (Botto et al., 2013; Narod et al., 1997). CHD patients have been reported to develop brain tumors and Wilms tumor more frequently than expected (Nishi et al., 2000). A study examining late causes of death in CHD patients reported a greater likelihood of dying from neoplastic disorders than the general population mainly attributed to genetic factors and interventional exposures (Raissadati, Nieminen, Haukka, Sairanen, & Jokinen, 2016). Another study reported that children with CHD had an elevated cancer hazard risk and seemed at particular risk for lymphoma (Fisher et al., 2012). Children with defects of the great arteries, valves, and atrioventricular septum were found to have a higher cancer risk, especially for lymphatic and hematopoietic cancers (Sun, Overvad, & Olsen, 2014). Finally, heart anomalies have been associated with hepatoblastoma, neuroblastoma and other unspecified malignancies (Norwood et al., 2017).

The RASopathies are a group of related disorders caused by dysregulation of the RAS/MAPK pathway and characterized by a variety of birth defects including CHD such as pulmonary stenosis and septal defects as well as an increased risk of certain cancers (Kratz et al., 2015; Kratz, Rapisuwon, Reed, Hasle, & Rosenberg, 2011; Rauen, 2013). Among the RASopathies, Noonan syndrome is the most common single gene cause of CHD with over 80% of patients having a cardiovascular malformation, particularly pulmonary stenosis, while patients are also at increased risk of developing neuroblastoma, acute leukemia, low-grade glioma, and rhabdomyosarcoma (Kratz et al., 2011; Romano et al., 2010). Costello syndrome is associated with an increased risk of neuroblastoma, rhabdomyosarcoma, and bladder cancer while the cancer types associated with cardiofaciocutaneous syndrome and Noonan syndrome with multiple lentigines overlap those of Noonan syndrome and Costello syndrome respectively (Kratz et al.,

2011). Finally, neurofibromatosis can include atrioventricular valve regurgitation, mitral valve prolapse, aortic valve regurgitation, pulmonary valve stenosis, and ASD, or VSD in addition to the eponymous neurofibromas and other tumors (Incecik, Herguner, Alinc Erdem, & Altunbasak, 2015).

Several extrinsic factors have also been proposed to contribute to the increased risk of cancer in CHD patients. Raissadati et al. (2016) had noted that CHD patients have an increased rate of neoplasia in radiation-sensitive tissues. It has been hypothesized that adult CHD patients are at an increased risk for cancer due to interventional radiation exposure, genetics, and other environmental factors (Gurvitz et al., 2016). One study found that the risk of developing cancer among CHD patients undergoing catheterization was elevated but not to a statistically significant level while another study reported that young CHD patients incur a small but significant risk of developing radiation-induced cancers following interventional catheterization (Olsen et al., 2014; Yakoumakis et al., 2013). Finally, it has been observed that CHD patients who develop cancer had a higher rate of undergoing procedures that involved low-dose exposure to ionizing radiation than did CHD patients who did not develop cancer (Cohen et al., 2018). Another hypothesis involves chronic activation of the hypoxia pathway in patients with cyanotic CHD leading to an increased risk for cancer such as pheochromocytoma or paraganglioma (Opotowsky et al., 2015). Finally, it has been noted that CHD patients undergoing univentricular palliation leading to the Fontan procedure can develop hepatic dysfunction secondary to chronic pulmonary hypertension or vascular congestion and right heart failure which can in turn lead to an increased risk of hepatocellular carcinoma as soon as two years post-Fontan (Asrani, Warnes, & Kamath, 2013; Augustyn, Peng, Singal, & Yopp, 2015; A. C. Egbe et al., 2018).

Over the course of reviewing WES data from study participants, it was noticed that these patients appeared to have a large number of rare variants in genes known to be associated with cancer (Lo, 2016-2019). To this end, I investigated if there is a real excess of rare variants in genes known to be associated with cancer in our CHD patients compared to a control cohort.

3.2 Materials and Methods

3.2.1 Gene Sets

Two partially overlapping gene sets were used in this analysis. The first set was the set of genes with variants present in CHD cases and/or SCOOP controls that are associated with a cancer phenotype as curated by ClinVar or HGMD. Cancer phenotypes were defined as those with the ICD-10 version 2016 codes C00-C97 (Malignant neoplasms), D00-D09 (In situ neoplasms), D10-D36 (Benign neoplasms), D37-D48 (Neoplasms of uncertain or unknown behavior), and Q85 (Phakomatoses not elsewhere classified) as well as Z15.09 (Genetic susceptibility to other malignant neoplasms) which was found through a Google search for ICD-10 codes related to cancer susceptibility.

The second gene set was obtained from the Catalogue Of Somatic Mutations In Cancer (COSMIC) Cancer Gene Census (CGC) (Forbes et al., 2015; Sondka et al., 2018; Tate et al., 2019) which is a list of genes for which germline and/or somatic mutations have been linked to cancer. The CGC contains 717 genes and six loci of which 680 genes have somatic mutations associated with cancer, 107 have germline mutations, and 64 have both somatic and germline mutations. Figure 2 illustrates the degree of overlap between the two cancer gene sets and the CHD-ome and

Ciliome gene sets including only genes present in our dataset. The darkest gray indicates the area where all four lists intersect, the medium gray indicates the intersection of three lists, the light gray indicates the intersection of two lists, and the white indicates no intersection.

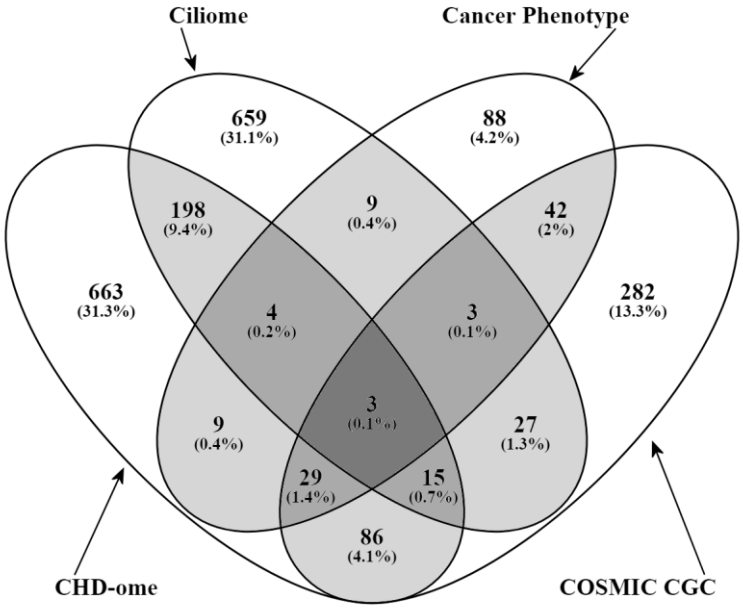


Figure 4 Overlap of CHD-ome, Ciliome, and cancer gene sets

3.2.2 Data Analysis and Statistics

Data analysis was carried out in the same manner as described earlier except restricted to genes with variants in CHD cases and/or SCOOP controls associated with a cancer phenotype or genes in the COSMIC CGC.

3.3 Results

3.3.1 Enrichment for Genes with Cancer Phenotypes

3.3.1.1 Overall enrichment

Overall, variants in 187 genes had at least one cancer phenotype as curated by ClinVar or HGMD. Of those genes, 13 had significant odds ratios before adjustment, seven with OR > 1 and six with OR < 1, but none remained significant after adjustment. There were 10 genes with a significant excess of alleles in CHD cases after adjustment of which five (*AIP*, *ACKR3*, *EGFR*, *ENG*, *PTCH1*) are CHD-ome genes (Table 10). In mice, global deletion of EGFR, encoding the epidermal growth factor receptor gene, results in hypoplastic outflow valves (C. J. Lin et al., 2012). Endoglin (*ENG*) variants have been associated with vascular anomalies in humans (Mattassi et al., 2018). PTCH1 (Patched 1) mutations have been associated with CHD including BAV with or without coarctation of the aorta in humans (Bonachea et al., 2014; Zaidi & Brueckner, 2017).

Table 10: Significant genes with cancer phenotypes overall

	Gene List			Allele Count			Odds Ratio		
Gene	**CHD100**	**CHD-ome**	**Ciliome**	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
AIP	FALSE	TRUE	FALSE	4.0326	6.09E-07	6.56E-05	4.7552	0.0015	0.0996
SPINK1	FALSE	FALSE	FALSE	4.8007	2.92E-06	0.0002	5.6540	0.0065	0.2224
CD27	FALSE	FALSE	FALSE	5.7608	4.55E-06	0.0004	6.8007	0.0073	0.2236
ACKR3	FALSE	TRUE	FALSE	5.0507	3.53E-05	0.0019	5.9305	0.0161	0.3177
WWOX	FALSE	FALSE	FALSE	1.6386	0.0002	0.0089	1.9266	0.0118	0.2836
EGFR	FALSE	TRUE	FALSE	5.7608	0.0004	0.0127	6.7807	0.0669	0.5869
CHEK2	FALSE	FALSE	FALSE	5.7608	0.0004	0.0127	6.7807	0.0669	0.5869
ENG	FALSE	TRUE	FALSE	1.6907	0.0005	0.0143	1.9918	0.0181	0.3296
MSH2	FALSE	FALSE	FALSE	2.5924	0.0011	0.0271	3.0463	0.0501	0.5267
PTCH1	FALSE	TRUE	TRUE	2.0803	0.0013	0.0321	2.4446	0.0451	0.5146

Gene List indicates if the gene is present in either of the 3 focus gene lists. For allele count, FC is based on normalized allele counts and the p-value is the upper tail probability of the binomial test. For OR, the p-value was obtained from Fisher's exact test

There was a total of 361 curated cancer phenotypes with 21 having a significant odds ratio before adjustment, nine with OR > 1 and 12 with OR < 1, though none remained significant after adjustment. There were 80 cancer phenotypes with an excess of alleles in CHD cases before adjustment with 35 remaining significant after adjustment (Appendix Table 7).

3.3.1.1.1 ToppGene analysis

ToppGene analysis of the set of genes with cancer phenotype allele FC ≥ 2 overall produced multiple significant structural CHD phenotype terms that were significant only in that set and not in the set of genes with cancer phenotype allele FC ≤ −2 (Appendix Table 8). However, there were a number of structural CHD terms as well as terms related to heart development and morphogenesis that were significant only in the set of genes with cancer phenotype allele FC ≤ −2. Additionally, there were also a number of significant terms related to cardiac neoplasms. There was little overlap in significant terms between sets with only two terms related to the cardiovascular system, "Neoplasm of the heart" and "Angiofibromas" out of 104 being significant in both sets. There were also a number of terms related to apoptosis, nitric oxide signaling, and cilia/hedgehog signaling that were significant in the set of genes with allele FC ≥ 2 for variants associated with cancer phenotypes, though there were also a number of such significant terms in the set of genes with allele FC ≤ −2.

Of particular interest were the results of a ToppGene analysis of the genes appearing in Table 10 which produced multiple terms related to apoptosis, nitric oxide synthesis, ERBB4 signaling, and transcriptional regulation by AP-2 (TFAP2) family transcription factors (Table 11). Many of these or similar terms were also significant in ToppGene analysis of the set of genes with curated cancer phenotype allele FC ≥ 2 overall. Apoptosis has been shown to play a pivotal role in heart development, particularly with respect to the endocardial cushions and atrioventricular canal (Abdelwahid, Pelliniemi, & Jokinen, 2002; Grunert et al., 2014; Jerse & Zidar, 2011; Poelmann, Molin, Wisse, & Gittenberger-de Groot, 2000; Sanchis, Llovera, Ballester, & Comella, 2008; Vicente Steijn et al., 2018; D. Yang et al., 2012; Z. Zhao & Rivkees, 2000). The AP-2

(TFAP2) signaling pathway has been shown to play a critical role in neural crest and outflow tract development in mice as well as being linked with familial isolated patent ductus arteriosus in humans (Brewer, Feng, Huang, Sullivan, & Williams, 2004; Brewer, Jiang, Donaldson, Williams, & Sucov, 2002; Ji et al., 2014). Evidence has suggested that defective nitric oxide synthesis and signaling play a role in the etiology of a variety of cardiac defects including abnormal cardiac situs, septal defects, BAV, and HLHS (Hrstka, Li, Nelson, & Pipeline, 2017; Y. Liu & Feng, 2012; Siamwala et al., 2019). *ERBB4* signaling plays an important role in valve development and myocardial trabeculation (Iwamoto & Mekada, 2006; D. Srivastava & Olson, 2000).

Table 11: ToppGene analysis of genes in Table 10

Name	Genes in Query	p-value	q-value FDR B&H
apoptotic signaling pathway	5 (CHEK2, MSH2, CD27, WWOX, ACKR3)	8.73E-06	1.88E-03
intrinsic apoptotic signaling pathway	4 (CHEK2, MSH2, WWOX, ACKR3)	1.17E-05	2.16E-03
intrinsic apoptotic signaling pathway by p53 class mediator	3 (CHEK2, MSH2, WWOX)	7.93E-06	1.88E-03
intrinsic apoptotic signaling pathway in response to DNA damage	3 (CHEK2, MSH2, ACKR3)	2.11E-05	3.00E-03
negative regulation of response to stimulus	7 (CHEK2, SPINK1, EGFR, ENG, WWOX, ACKR3, PTCH1)	6.73E-06	1.88E-03
nitric-oxide synthase activity	2 (EGFR, ENG)	4.23E-04	1.90E-02
regulation of apoptotic process	7 (CHEK2, EGFR, MSH2, ENG, CD27, WWOX, ACKR3)	3.91E-06	1.88E-03
regulation of cell death	7 (CHEK2, EGFR, MSH2, ENG, CD27, WWOX, ACKR3)	7.75E-06	1.88E-03
regulation of nitric oxide mediated signal transduction	2 (SPINK1, EGFR)	1.46E-05	2.36E-03
regulation of programmed cell death	7 (CHEK2, EGFR, MSH2, ENG, CD27, WWOX, ACKR3)	4.41E-06	1.88E-03
Signaling by ERBB4	2 (EGFR, WWOX)	4.13E-04	4.85E-02
Transcriptional regulation by the AP-2 (TFAP2) family of transcription factors	2 (EGFR, WWOX)	3.57E-04	4.85E-02

3.3.1.1.2 Ingenuity Pathway Analysis

IPA comparison analysis of the sets of genes with allele FC ≥ 2 and FC ≤ −2 for variants with cancer phenotypes overall produced only seven structural CHD terms of which six were present only in the set with FC ≥ 2 and having positive deltas while the seventh term, "cardiogenesis" was present in both sets and had a negative delta (−1.7107). The six terms with positive deltas encompassed only five genes (*ACKR3*, *AIP*, *CDK4*, *EGFR*, *PTCH1*). Overall, the set of genes with FC ≤ −2 for variants with cancer phenotypes produced a number of significant cardiovascular terms, many of which related to formation of the heart. Additionally, there were several terms related to apoptosis significant only in the CHD set, however most of them were not relevant to the heart (Table 12). There were also multiple apoptosis-related terms significant only the in the SCOOP set and one term, "apoptosis" significant in both sets (delta = 0.0743).

Table 12: IPA analysis of genes with allele FC ≥ 2 for variants with cancer phenotypes

Diseases and Bio Functions	GS19	Genes
Exit from apoptosis of B lymphocytes	2.7430	MSH2
Closure of ductus venosus	2.7430	AIP
Persistence of ductus venosus	2.7430	AIP
Abnormal morphology of cardiac valve	2.8200	ACKR3, CDK4
Apoptosis of germinal center	3.0438	MSH2
Apoptosis of pleural cells	3.0438	EGFR
Cardiogenesis	3.5356	ACKR3, CASP10, CDK4, EGFR, PTCH1
Formation of semilunar valve	3.6037	ACKR3, EGFR
Ventricular septal defect	3.6389	ACKR3, AIP, CDK4
Congenital heart disease	3.6459	ACKR3, AIP, CDK4, PTCH1
Apoptosis of T lymphocytes	3.8467	CASP10, CD27, CHEK2, ITK
Apoptosis	4.1573	AIP, CASP10, CD27, CHEK2, EGFR, ITK, LIG4, MITF, MSH2, PTCH1
Apoptosis of lymphocytes	4.6845	CASP10, CD27, CHEK2, ITK, MSH2
Apoptosis of blood cells	5.0994	AIP, CASP10, CD27, CHEK2, ITK, MSH2
Terms significant only in the CHD gene set. G19 is the analyzed gene set (Appendix C.1) and the numbers are -$\log_{10}$(p-values) generated from the IPA comparison analysis.		

3.3.1.2 Enrichment for pathogenic and highly damaging variants

Pathogenic variants are those curated as "(likely) pathogenic" by ClinVar and "DM" by HGMD while highly damaging variants are those with a CADD score ≥ 30. Only 19 genes had variants curated as pathogenic by both ClinVar and HGMD that were associated with cancer phenotypes with none having a significant unadjusted odds ratio p-value. Only five genes had a significant excess of curated pathogenic alleles in CHD cases unadjusted of which only one, AIP, remained significant after adjustment (BH FDR = 0.0177). There were 71 genes with cancer phenotypes with highly damaging variants though, as with pathogenic variants, none had a significant unadjusted odds ratio p-value. There were 15 genes with a significant excess of highly damaging alleles in CHD cases unadjusted, but none remained significant after adjustment.

Of the 93 cancer phenotypes with pathogenic alleles, none had significant odds ratios unadjusted and only 13 had a significant excess of alleles in CHD cases before adjustment with none remaining significant after adjustment. However, six terms related to pituitary or parathyroid adenoma were trending towards significance (BH FDR = 0.0562), though all these terms were associated with a single variant in the *AIP* gene. There were 107 cancer phenotypes with highly damaging variants, but only one had a significant OR before adjustment and 29 had a significant excess of alleles in CHD cases, but nothing remained significant after adjustment. The term closest to significance after adjustment was "Hereditary cancer predisposition syndrome" (BH FDR = 0.0728) associated with variants in *APC*, *ATM*, *BLM*, *BMPR1A*, *BRCA2*, *BRIP1*, *CDH1*, *NBN*, *POLD1*, *POLE*, *SUFU*, and *TSC2*.

3.3.1.2.1 ToppGene Analysis

Because the sets of genes with cancer phenotypes enriched for pathogenic and highly damaging variants were individually so small, they were combined and run through enrichment analysis as a single set. No structural CHD phenotype terms were significant in the set of genes with pathogenic or highly damaging allele FC ≥ 2, though there were several terms related to vascular anomalies (Table 13). The most significant term related to the cardiovascular system for this analysis was "Atrial fibrillation (BH FDR = 5.93E-03) associated with *ROS1*, *CYP2C19*, and *ATM* while there were also three significant terms related to response to oxygen levels all of which were associated with *TSC2*, *BRIP1*, and *ATM*. Also, there were again a number of terms related to apoptosis. Finally, there were multiple structural CHD terms that were significant in the sets of genes with allele FC ≤ -2 both overall and when restricting to pathogenic or highly damaging variants.

Table 13: ToppGene analysis of genes with pathogenic/highly damaging variants associated with cancer phenotypes

Name	Hit in Query List	p-value	q-value FDR B&H
increased sensitivity to induced cell death	4 (TYR, BRIP1, ATM, ERCC6)	3.98E-05	3.07E-03
Cardiac rhabdomyoma	1 (TSC2)	7.37E-04	6.93E-03
Abnormal induced cell death	4 (TYR, BRIP1, ATM, ERCC6)	1.50E-04	7.36E-03
Abnormality of the vasculature	8 (FANCM, CDH1, POLD1, TSC2, BRIP1, ATM, AIP, ERCC6)	9.41E-04	1.18E-02
response to hypoxia	3 (TSC2, BRIP1, ATM)	1.37E-03	2.51E-02
Abnormal cell death	7 (TYR, CDH1, TSC2, BRIP1, ATM, AIP, ERCC6)	1.27E-03	2.58E-02
response to decreased oxygen levels	3 (TSC2, BRIP1, ATM)	1.50E-03	2.59E-02
response to oxygen levels	3 (TSC2, BRIP1, ATM)	1.78E-03	2.59E-02
intrinsic apoptotic signaling pathway in response to DNA damage	2 (ATM, ERCC6)	2.05E-03	2.81E-02
Mitochondrial cardiomyopathy	1 (ROS1)	1.17E-02	3.02E-02
Abnormal carotid artery morphology	2 (FANCM, BRIP1)	5.43E-03	3.50E-02

Table 13 Continued

Abnormal vascular morphology	6 (FANCM, CDH1, TSC2, BRIP1, AIP, ERCC6)	5.42E-03	3.50E-02
Angiofibroma	1 (TSC2)	1.54E-02	3.51E-02
Cardiovascular Diseases	3 (ROS1, CYP2C19, ATM)	1.53E-02	3.51E-02
Aortic Diseases	1 (TSC2)	1.61E-02	3.59E-02
Ventricular Dysfunction, Left	1 (ATM)	1.61E-02	3.59E-02
Abnormal apoptosis	6 (TYR, CDH1, TSC2, BRIP1, ATM, AIP)	2.62E-03	3.66E-02
abnormal vascular development	4 (TYR, CDH1, TSC2, AIP)	3.34E-03	4.02E-02
patent ductus venosus	1 (AIP)	3.36E-03	4.02E-02
Angiomyolipoma	1 (TSC2)	1.97E-02	4.09E-02
Coronary Stenosis	1 (ATM)	2.55E-02	4.71E-02
Increased trophectoderm apoptosis	1 (CHD1)	5.59E-03	4.93E-02

3.3.1.2.2 Ingenuity Pathway Analysis

When restricting to variants with pathogenic or highly damaging alleles, "Hypoxia signaling in the cardiovascular system" was one of the Canonical Pathway terms that was significant only in the set with FC ≥ 2 (Table 14). This analysis produced only 15 total significant 'Disease and Bio Function' terms related to the cardiovascular system of which four were significant in the set with FC ≥ 2 (Table 14) and 11 were significant in the set with FC ≤ −2 with no terms being significant in both sets. The significant cardiovascular terms for the set with pathogenic or highly damaging allele FC ≥ 2 encompassed only four genes (*AIP*, *ATM*, *CYP2C19*, *TSC2*). These results seem to suggest that pathogenic and highly damaging variants in genes with curated cancer phenotypes do not play a major role in the etiology of CHD though rare variants overall may play a role, especially those in genes in the apoptosis pathway. Additionally, they suggest that variants in genes related to hypoxia signaling or response to oxygen levels may serve as a prognostic indicator. As for the term "Major adverse cardiovascular event", variants in CYP2C19 have been associated with a higher incidence of coronary heart disease and adverse outcomes following treatment (Zhang et al., 2019).

Table 14: IPA analysis of genes with pathogenic or highly damaging variants with cancer phenotypes

Canonical Pathways	**GS21**	
Hypoxia Signaling in the Cardiovascular System	1.3703	
Diseases and Bio Functions	**GS21**	**Genes**
Familial cardiovascular disease	3.3493	AIP, ATM, CYP2C19, TSC2
Apoptosis of epithelial cell lines	3.2570	CDH1, TSC2
Dysfunction of vascular endothelial cells	3.2309	ATM
Apoptosis of squamous epithelial cells	3.2309	CHD1
Persistence of ductus venosus	2.9300	AIP
Major adverse cardiovascular event	2.6292	CYP2C19
Terms significant only in the CHD gene set. G21 is the analyzed gene set (Appendix C.1) and the numbers are -$\log_{10}$(p-values) generated from the IPA comparison analysis.		

3.3.2 Enrichment for COSMIC Cancer Gene Census Genes

3.3.2.1 Overall Enrichment

Overall, 487 COSMIC CGC genes were present in our dataset including eight CHD100, 133 CHD-ome, 48 Ciliome, and 77 with variants curated as being associated with cancer phenotypes. CHD patients had significantly more calls in CGC genes overall and in calls in highly damaging CGC gene variants but the average number of calls per sample in CGC gene variants curated as pathogenic were nearly identical and not significantly different (Table 1). There were 34 CGC genes with a significant odds ratio before adjustment including 26 with OR > 1, but only one (*NCOR2*) remained significant after adjustment. There were 35 CGC genes with significantly more alleles in CHD cases after adjustment including 13 CHD-ome genes (*NCOR2*, *ACKR3*, *USP44*, *FSTL3*, *NRG1*, *EGFR*, *PIK3CB*, *ARID1A*, *SPEN*, *PTCH1*, *USP8*) (Table 14). In mice *Arid1a* (AT-rich interaction domain 1A) has been shown to be required for cardiac neural crest development with mutants showing outflow tract and pharyngeal arch artery defects as well as VSD and absent myocardial trabeculation (Chandler et al., 2013; Chandler & Magnuson, 2016). *FSTL3* (Follistatin-like 3) has been identified as a candidate gene accounting for the TOF seen in patients with del(19)(p13.3) (Silversides et al., 2012). *Ncor2* (Nuclear receptor corepressor 2) null mice exhibit hypoplastic ventricle walls and VSD (Jepsen, Gleiberman, Shi, Simon, & Rosenfeld, 2008). *PIK3CB*, encoding Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta, has been identified as a candidate gene accounting for the CHD seen in patients with del(3)(q22.3) (Xie et al., 2017). SPEN family transcriptional repressor (*SPEN*) variants have been associated with HLHS in humans while mouse mutants exhibit defects in septation and myocardium formation (Kuroda et al., 2003; X. Liu et al., 2017). Variants in ubiquitin-specific peptidase 44 (*USP4*) have been associated with CHD in humans (Zaidi et al., 2013). Finally, mice with

mutations in *Usp8* (Ubiquitin-specific peptidase 8) show abnormal heart development (Niendorf et al., 2007).

Table 15: Significant CHD-ome genes in the CGC

	Gene List			Allele Count			Odds Ratio		
Gene	**CHD100**	**CHD-ome**	**Ciliome**	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
NCOR2	FALSE	TRUE	FALSE	1.0110	5.96E-09	1.46E-06	1.2026	2.84E-06	0.0029
ACKR3	FALSE	TRUE	FALSE	5.0407	3.53E-05	0.0019	5.9305	0.0161	0.3177
USP44	FALSE	TRUE	FALSE	2.8805	9.38E-05	0.0042	3.3886	0.0133	0.2905
FSTL3	FALSE	TRUE	FALSE	3.8405	0.0001	0.0046	4.5315	0.0240	0.3688
NRG1	FALSE	TRUE	FALSE	1.7831	0.0003	0.0092	2.0967	0.0147	0.3092
EGFR	FALSE	TRUE	FALSE	5.7608	0.0004	0.0127	6.7807	0.0669	0.5869
PIK3CB	FALSE	TRUE	FALSE	2.8804	0.0008	0.0227	3.3886	0.0674	0.5869
ARID1A	FALSE	TRUE	FALSE	1.8833	0.0011	0.0289	2.2132	0.0361	0.4534
SPEN	FALSE	TRUE	FALSE	2.1603	0.0014	0.0320	2.5376	0.0392	0.4820
PTCH1	FALSE	TRUE	TRUE	2.0803	0.0013	0.0321	2.4446	0.0451	0.5146
USP8	FALSE	TRUE	FALSE	3.6005	0.0015	0.0340	4.2324	0.1099	0.7091

Gene List indicates if the gene is present in either of the 3 focus gene lists. For allele count, FC is based on normalized allele counts and the p-value is the upper tail probability of the binomial test. For OR, the p-value was obtained from Fisher's exact test

3.3.2.1.1 ToppGene Analysis

ToppGene analysis of the sets of CGC genes with FC ≥ 2 and FC ≤ −2 overall produced 107 terms potentially related to the cardiovascular system of which 39 were significant in the enriched set and 71 in the depleted set with three terms being significant in both. Aside from several terms related to valvular dysplasia, all associated with *FLNA* alone (Table 16), most bona-fide structural CHD terms were significant only in SCOOP controls. Though there were few bona-fide structural CHD terms significant in CHD cases, there were several syndromes that can include CHD including Noonan syndrome 10 (*LZTR1*) (BH FDR = 1.48E-02) and Cornelia-de Lange syndrome *(MYC*, *AFF4*) (BH FDR = 1.48E-02). Other significant terms of interest included "abnormal oxygen level" (BH FDR = 1.03E-02) associated with *MYC*, *FLNA*, *PRRX1*, *EGFR*, *NFIB*, *ACKR3*, *RSPO2*, and *SDHB* as well as several terms related to apoptosis. Overall, the terms significant in CHD cases only encompassed one CHD100 gene (*DNM2*) and a further nine CHD-ome genes (*ACKR3*, *AFF4*, *EGFR*, *FLNA*, *LZTR1*, *MYOD1*, *PIK3CB*, *PTCH1*, *SOX2*).

Table 16: ToppGene Analysis of COSMC CGC genes with FC ≥ 2 overall

Name	Hit in Query List	p-value	q-value FDR B&H
canonical Wnt signaling pathway involved in negative regulation of apoptotic process	2 (MYC, MITF)	8.38E-05	4.58E-03
abnormal oxygen level	8 (MYC, FLNA, PRRX1, EGFR, NFIB, ACKR3, RSPO2, SDHB)	1.59E-04	1.03E-02
Neoplasms, Radiation-Induced	2 (MYC, PTCH1)	2.35E-03	1.48E-02
Cornelia De Lange Syndrome	2 (MYC, AFF4)	3.11E-03	1.48E-02
Cardiac Paraganglioma	1 (SDHB)	3.64E-03	1.48E-02
NOONAN SYNDROME 10	1 (LZTR1)	3.64E-03	1.48E-02
Congenital dysplasia of cardiac valve	1 (FLNA)	3.64E-03	1.48E-02
abnormal response to radiation	4 (ERCC2, CHEK2, MSH2, PTCH1)	3.37E-04	1.62E-02
Costello syndrome (disorder)	2 (PIK3CB, CDK4)	4.59E-03	1.78E-02
cyanosis	7 (MYC, FLNA, PRRX1, EGFR, NFIB, ACKR3, RSPO2)	4.50E-04	1.87E-02

Table 16 Continued

cardiovascular system development	9 (RSPO3, FLNA, CYSLTR2, PRRX1, PIK3CB, ACKR3, TNFAIP3, DDIT3, DNM2)	1.12E-03	2.06E-02
Radiation-Related Angiosarcoma	1 (MYC)	7.23E-03	2.32E-02
Abnormal heart morphology	19 (SOX2, MYC, TPM3, CNBP, FLNA, USP8, PRRX1, LZTR1, WAS, CANT1, PAX8, CHEK2, MSH2, ATIC, PTCH1, AFF4, SDHB, DNM2, CDK4)	1.97E-03	3.21E-02
Abnormality of cardiovascular system morphology	22 (SOX2, MYC, TPM3, CNBP, FLNA, USP8, PRRX1, ERCC2, LZTR1, WAS, SH2B3, CANT1, PAX8, CD28, CHEK2, MSH2, ATIC, PTCH1, AFF4, SDHB, DNM2, CDK4)	2.16E-03	3.37E-02
Notch-mediated HES/HEY network	3 (MYOD1, NCOA1, SPEN)	1.10E-03	3.43E-02
Notch signaling pathway	4 (SOX2, MYC, EGFR, SPEN)	3.28E-03	3.57E-02
nitric-oxide synthase binding	2 (EGFR, DNM2)	3.21E-03	3.63E-02
canonical Wnt signaling pathway involved in positive regulation of apoptotic process	1 (MYC)	5.85E-03	4.37E-02

3.3.2.1.2 Ingenuity Pathway Analysis

An IPA analysis of the sets of CGC genes with FC ≥ 2 and FC ≤ −2 overall produced 18 terms related to the cardiovascular system of which only six were significant in CHD cases with only two of those terms, "Enlargement of heart" ($-\log_{10}$ p-value = 4.3330) and "Congenital anomaly of the cardiovascular system" ($-\log_{10}$ p-value = 4.0832) being significant only in CHD cases. The six terms significant in CHD cases encompassed 24 genes including one CHD100 gene (*DNM4*) and an additional 11 CHD-ome genes (*ABI1*, *ACKR3*, *EGFR*, *FLNA*, *FSTL3*, *MYOD1*, *PIK3CB*, *PTCH1*, *SPEN*, *USP44*, *USP8*). There were 16 terms significant in the set of genes with FC ≤ −2 with many of them related to heart morphology/morphogenesis and a few related to conduction.

3.3.2.2 Enrichment for pathogenic and highly damaging variants

Only nine CGC genes had curated pathogenic variants with none having significant raw odds ratio p-values and only one (*ZEB1*) having a significant excess of alleles in CHD cases before adjustment but losing significance after adjustment. None of the nine genes were CHD100 or CHD-ome genes, though three (*VHL*, *TSHR*, *NBN*) were Ciliome genes. There were only three genes whose odds ratios for highly damaging alleles were significant before adjustment, but none remained significant after adjustment. Alleles in highly damaging variants were significantly more frequent in 29 genes before adjustment, however only two genes, *RNF213*, (BH FDR = 0.0001) and *NRG1* (BH FDR = 0.0106), remained significant after adjustment. An additional 10 CHD-ome genes (*ABI1*, *SETBP1*, *SPEN*, *ATM*, *PER1*, *ATIC*, *ARID1B*, *FLT4*, *MYOD1*, *TSC2*) had significant raw p-values.

3.3.2.2.1 ToppGene analysis

ToppGene analysis of the set of CGC genes with highly damaging alleles FC ≥ 2 did not produce any significant bona-fide structural CHD terms, though two significant terms were "Abnormal cardiovascular development" associated with *RNF213*, *NRG1*, *TSC2*, *FLT4*, *ABI1*, *SPEN* and *CDH1* and "Fenestration (morphologic abnormality)" associated with *FLT4*, a CHD-ome gene (Table 16). Additionally, there were four significant terms related to VEGF signaling, a term related to WNT signaling, as well as two terms with potential prognostic value for CHD patients, "Radiation-Related Angiosarcoma" associated with *FLT4* and "response to oxygen levels" associated with *TSC2*, *BRIP1*, *ATM*, and *MYOD1*. ToppGene analysis of the set of CGC genes with highly damaging allele FC ≤ −2 produced a number of terms related to cardiovascular system morphogenesis and development.

Table 17: ToppGene analysis of CGC genes with highly damaging variants

Name	Hit Count in Query List	p-value	FDR B&H
Cardiac rhabdomyoma	1 (TSC2)	1.48E-03	1.51E-02
MOYAMOYA DISEASE 2	1 (RNF213)	1.48E-03	1.51E-02
Moyamoya disease 2	1 (RNF213)	1.48E-03	1.51E-02
Radiation-Related Angiosarcoma	1 (FLT4)	2.95E-03	2.14E-02
Fenestration (morphologic abnormality)	1 (FLT4)	8.81E-03	3.48E-02
abnormal cardiovascular development	7 (RNF213, NRG1, TSC2, FLT4, ABI1, SPEN, CDH1)	6.26E-04	3.81E-02
negative regulation of Wnt signaling pathway	3 (RNF213, TSC2, CDH1)	2.20E-03	3.86E-02
response to oxygen levels	4 (TSC2, BRIP1, ATM, MYOD1)	1.49E-03	3.86E-02
abnormal endocardium morphology	2 (TSC2, ABI1)	1.49E-03	4.04E-02
Nuclear signaling by ERBB4	2 (NCOR1, NRG1)	6.48E-04	4.09E-02
Signaling by VEGF	4 (JAK1, NRG1, FLT4, ABI1)	1.55E-03	4.09E-02
altered vascular endothelial growth factor signaling	1 (FLT4)	1.53E-03	4.09E-02
Signaling by ERBB4	2 (NCOR1, NRG1)	1.92E-03	4.27E-02
vascular endothelial growth factor-activated receptor activity	1 (FLT4)	1.12E-02	4.28E-02

Table 17 Continued

vascular endothelial growth factor binding	1 (FLT4)	1.24E-02	4.53E-02
dihydrofolate metabolic process	1 (ATIC)	4.76E-03	4.84E-02

3.3.2.2.2 Ingenuity Pathway Analysis

An IPA analysis of the sets of CGC genes with highly damaging allele FC ≥ 2 or FC ≤ −2 showed that significant terms related to the cardiovascular system were only present in SCOOP controls. There were also no significant terms related to apoptosis in the FC ≥ 2 set. It is possible that the analysis settings for this set were incorrect as most of the results were cancer-related terms and over 400 of the 500 total significant terms in the FC ≥ 2 set were not significant in the FC ≤ −2 set

3.4 Discussion

Only a small proportion of the total gene set had variants associated with a cancer phenotype by ClinVar or HGMD within our study population. There are certainly other genes in our dataset that have variants associated with cancer, however cancer-associated variants in those genes were not represented in our dataset. This small sample size of genes probably makes it difficult to arrive at any solid conclusions. However, it is worth noting that five of the 10 genes with cancer phenotypes having a significant excess of alleles in CHD cases after adjustment were CHD-ome genes (*AIP*, *ACKR3*, *EGFR*, *ENG*, *PTCH1*). This observation is in agreement with the hypothesis that genes associated with cancer are playing a role in the etiology of CHD. Additionally, the finding of several terms related to apoptosis in a ToppGene analysis of these 10 genes is significant as this process plays several key roles in heart development. Only one gene, *AIP*, had a significant excess of pathogenic alleles associated with a cancer phenotype after adjustment while there were no significant genes with highly damaging variants associated with a

cancer phenotype after adjustment. No genes had a significant raw odds ratio p-value either overall or for genes with pathogenic or highly damaging variants associated with a cancer phenotype.

Similarly, though there were a large number of cancer phenotypes overall, none had a significant odds ratio unadjusted while few had a significant excess of overall alleles in CHD cases and nothing remained significant after adjustment for either pathogenic or highly damaging alleles. Though there were five terms approaching significance for pathogenic alleles, these were all associated with the same variant in *AIP*. This illustrates a common problem when performing an analysis based on curated phenotypes in that many variants have multiple curated phenotypes and, in many cases, that variant may be the only variant associated with many of those phenotypes.

ToppGene analysis of the sets of genes with overall allele FC ≥ 2 or FC ≤ -2 produced a variety of significant terms related to CHD or heart/cardiovascular system morphology and development in each set with little overlap. IPA comparison analysis of the gene sets revealed several terms related to structural CHD that were significant only in the CHD set, however multiple terms related to cardiovascular system morphology and development were also significant in the SCOOP set. This could suggest that variants in these genes are providing a protective effect in the SCOOP controls. The small number of genes with cancer phenotypes that had an allele FC ≥ 2 or FC ≤ -2 for pathogenic or highly damaging variants necessitated combining them into a single set for enrichment analysis. ToppGene analysis of the set with allele FC ≥ 2 did not produce any significant structural CHD terms but did produce a number of terms related to response to hypoxia all associated with *ATM*, *BRIP1*, and *TSC2*. Variants in these genes could thus provide a prognostic indicator of how well patients with cyanotic CHD will tolerate long-term decreased oxygen saturation, though the terms do not indicate in what direction the response is modified. These terms could also bear significance in relation to the possible increased cancer risk secondary

to chronic activation of the hypoxia pathway (Oleaga-Alday, Goni-Goicoechea, Calles-Romero, Perez de Ciriza-Cordeu, & Paja-Fano, 2015; Opotowsky et al., 2015). There were also multiple terms related to tumors originating in the heart or vascular tissue for which variants in the associated genes may influence cancer risk in CHD patients. A number of terms related to cardiovascular system morphology and development were significant only in the SCOOP set. This could suggest that pathogenic or highly damaging variants in these genes that are associated with cancer phenotypes do not play a major role in the etiology of CHD in our study population. Another possibility is that these damaging variants are actually playing a protective role in the SCOOP population, perhaps by counterbalancing perturbations in developmental pathways caused by variants in other genes. Both scenarios are consistent with a multigenic model in which mildly damaging variants in multiple genes affecting multiple pathways act synergistically to cause CHD in the first scenario or variants in some genes acting to counteract the variants in other genes in the second scenario.

IPA analysis of these gene sets revealed "Hypoxia Signaling in the Cardiovascular System" as a canonical pathway term that was significant only in CHD cases. This again suggests that variants in genes involved in hypoxia signaling and response to oxygen levels could be used as a prognostic indicator. There were few significant 'Disease and Bio Function' terms related to the cardiovascular system including just four that were significant in CHD cases. However, two of these terms, "Dysfunction of vascular endothelial cells" associated with *ATM* and "Major adverse cardiovascular event" associated with *CYP2C19* may suggest variants in these genes may serve as prognostic indicators of patient outcome.

Though *NCOR2* was the only CGC gene with a significant odds ratio, its true significance may be questionable as the numbers were driven by a single common variant present in the

majority of samples from both cohorts that was inadvertently left in the data after filtering. The allele frequency for this variant was however much higher in CHD cases (99.6%) compared to SCOOP controls (85.1%). Additionally, there are 13 *NCOR2* variants in our dataset with 12 of them being enriched in CHD cases. *NCOR2*, (nuclear receptor corepressor 2) encodes a corepressor protein that interacts with *FOXP1*, itself a CHD-ome gene, to drive cardiac growth and mutant mice were shown to have hypoplastic ventricles and VSD (Jepsen et al., 2008). With this in mind, this gene may still be worth investigating further for potential modifier effects. With the exception of several terms related to valvular dysplasia, all associated with *FLNA*, ToppGene analysis of the genes with overall allele FC ≥ 2 produced no bona fide structural CHD terms. Mutations in *FLNA* (filamin A) have been associated with isolated valve defects, particularly involving the mitral valve, as well as Melnick-Needles syndrome, the most severe of the otopalatodigital syndromes (Azhar & Ware, 2016; LaHaye et al., 2014; Lalani & Belmont, 2014; Prendiville et al., 2014; Wessels & Willems, 2010). This analysis did produce several significant terms for which investigation of the associated genes and variants could provide prognostic information. Among these terms, "Cyanosis" and "Abnormal oxygen level" could indicate genes in which variants may provide information as to how well chronic hypoxia will be tolerated in cyanotic CHD patients. Variants in *MYC* and *PTCH1* associated with "Neoplasms, Radiation-Induced" may identify patients at higher risk for developing cancers secondary to interventional radiation exposure (Cohen et al., 2018; Yakoumakis et al., 2013). Most significant terms related to the cardiovascular system produced from an IPA analysis were significant only in SCOOP controls including several terms related to cardiovascular system development and morphology. This could suggest that variants in the associated genes primarily play a modifier role such that

they only cause CHD when combined with variants in other genes or that these genes are exerting a protective effect on subjects in the SCOOP cohort who carry them.

Only nine CGC genes have variants curated as "(Likely) pathogenic" by ClinVar and "DM" by HGMD with none being significant. This could suggest that pathogenic variants in these genes do not play a role in CHD. As there were only nine CGC genes with pathogenic variants with only one of those genes having an allele FC ≥ 2 or ≤ -2, enrichment analysis was not carried out for this set. *MUTYH* was the lone gene with a large fold change (-2.0830).

Though many more CGC genes had highly damaging variants than pathogenic variants, *NRG1* was the only CHD-ome gene with a significant excess of such alleles after correction. In mice, mutations in *Nrg1* have been associated with hypoplastic endocardial cushions and defective trabeculation (C. J. Lin et al., 2012; D. Srivastava & Olson, 2000). It may however be worth noting that 10 of 29 (34.5%) CGC genes with significantly more alleles in CHD cases before adjustment were CHD-ome genes when considering that only 27.3% (133/487) of the CGC genes in the entire dataset were CHD-ome genes. ToppGene analysis of the set of genes with allele FC ≥ 2 for highly damaging variants did not produce any significant bona fide structural CHD terms however, two terms, "Radiation-Related Angiosarcoma" associated with *FLT4* and "response to oxygen levels" associated with *TSC2, BRIP1*, *ATM*, and *MYOD1*, could provide possible prognostic factors. Additionally, there were four significant terms related to VEGF signaling. Damaging variants in genes in this pathway have been linked to TOF as well as Down syndrome-related AVSD (Ackerman et al., 2012; Reuter et al., 2018). Multiple terms related to structural CHD were only significant in ToppGene analysis of genes with an allele FC ≤ -2 for highly damaging variants. Similarly, IPA analyses of the genes sets with allele FC ≥ 2 or FC ≤ -2 for highly damaging variants produced significant terms related to the cardiovascular system only in SCOOP controls.

Overall, these results suggest that pathogenic and/or highly damaging variants in genes associated with cancer are not playing a major role in the etiology of CHD in our study population at least in terms of direct causation, but that they may have modifier effects. However, it did identify genes associated with response to low oxygen levels and with radiation-induced neoplasms as possible prognostic and risk factors.

One question that arises when considering this data is how the focus on and pre-selection of cancer genes impacts the null model of no enrichment for cancer-related genes. Though this analysis was specifically looking at genes with curated cancer phenotypes and genes in the CGC, the p-values obtained were still the same as those obtained from the analysis in Chapter 2. In this sense, the p-values show which cancer-related genes are significant among the entire set of genes overall or the entire set of genes with highly damaging variants or the entire set of genes with curated pathogenic variants. However, the gene sets used for enrichment analysis were those with large fold changes and with associated cancer phenotypes or those with large fold changes and present in the CGC. This may have had the effect of artificially increasing the significance of terms related to inherent properties of cancer cells such as invasiveness, proliferation, apoptosis, and angiogenesis. However, as many of these properties are also inherent to cells in the developing embryo, these terms could still be relevant. Additionally, many of these terms could also be found as significant terms in the un-focused genes sets based on the entire set of genes in our study population.

As this study was a subset of the larger study discussed in Chapter 2, it suffers from the same weaknesses including a non-random sample population and failure to correct for population stratification. In addition to these, there is the issue of somatic versus germline mutations in cancer genes. Though the CGC does indicate whether identified mutations in each gene are somatic,

germline, or both, the type of mutation was not taken into consideration and all genes were used in the analysis. Ideally, it would have been better to use only those genes for which germline mutations have been identified as it is entirely possible that somatic mutations found only in cancer cells could be so deleterious as to be incompatible with embryonic development if they were to occur as a germline mutations present in all cells of a developing embryo. However, neither ClinVar nor HGMD annotations indicate whether variants associated with cancer are found in a germline or somatic state so it is quite possible that the analysis based on genes associated with cancer phenotypes also contains a mix of genes with somatic and germline mutations. In this respect, the inclusion of genes with only somatic mutations in the CGC analysis may not have mattered and may have in fact been the correct approach. It was also interesting to note that of the 187 genes with variants associated with a cancer phenotype in our dataset, only 77 of these genes were present in the CGC. This could reflect differences in the criteria used to identify a gene as being associated with cancer. Finally, the use of ICD-10 codes to identify cancer phenotypes may have missed some syndromic conditions that include increased risk of developing cancer if the cancer susceptibility is not the primary phenotype. For instance, xeroderma pigmentosum which is well known to confer an increased risk of skin cancer has the ICD-10 code Q82.1 placing it in the subsection “Other congenital malformations of the skin”. Ultimately, despite its limitations, this study provides possible insights into genes and variants that may modify risk of cancer secondary to interventional radiation exposure or to chronic activation of the hypoxia pathway.

4.0 Correlating Patient Phenotypes

4.1 Introduction

Over the course of the preceding analyses, there were multiple genes that came up frequently. The purpose of this section is to summarize some of these genes and examine the phenotypes of patients carrying variants in these genes to determine if any patterns are present.

4.2 Materials and Methods

This section will mainly just focus on reporting the phenotypes of patients who carry variants in genes of interest to determine if there are any malformation patterns in patients carrying these variants.

4.3 Results and Discussion

4.3.1 AIP

AIP (Aryl hydrocarbon receptor interacting protein) showed an excess of alleles in the CHD cohort both overall (BH FDR = 6.56E-05) and when restricting to variants curated as pathogenic (BH FDR = 0.0176). Our dataset contained two missense variants; a c.47G>A p.Arg16His variant present in 10 CHD and four SCOOP samples as well as a C.911G>A

p.Arg304Gln variant present in four CHD and one SCOOP sample. Neither variant is highly damaging by the definition used in this dissertation (CADD scores of 25 and 25.4 respectively) (table 18). Though the Arg16His variant showed a significant excess of alleles in CHD controls, the Arg304Gln variant did not. However it was the Arg304Gln variant that stood out as three out of four of our patients carrying this variant had DORV while only two out of 10 patients with the Arg16His variant had DORV. Preliminary data has shown a significant odds of having DORV among patients with an *AIP* variant in our study population (OR = 3.6631, χ^2 p-value = 0.0245). *Aip* null mice have been shown to present DORV while mice harboring a hypomorphic allele have shown a patent ductus venosus (B. C. Lin et al., 2007; Obler et al., 2008). The authors hypothesized that this protein may act as a chaperone that plays a role in folding and localization of other proteins and that the loss of *AIP* causes DORV through the aberrant function of one of these "client" proteins.

Table 18: Summary of patients with *AIP* variants

Variant	11:67250676:G:A	11:67258382:G:A
HGVS	c.47G>A, p.Arg16His	c.911G>A, p.Arg304Gln
CADD PHRED	25	24.4
GnomAD AF	0.0020	0.0017
ClinVar	Conflicting interpretations, Likely benign, Uncertain significance): Hereditary cancer-predisposing syndrome, Somatotroph adenoma, Familial Isolated Pituitary Adenomas	Conflicting interpretations (Drug response, Likely benign, Pathogenic, Uncertain significance): Dopamine agonist response, Hereditary cancer-predisposing syndrome, Pituitary dependent hypercortisolism, Somatotroph adenoma, Familial Isolated Pituitary Adenomas
HGMD	DM?: Pituitary adenoma|Cushing's disease	DM: Pituitary adenoma, Macroadenoma, Parathyroid adenoma, Acromegaly & somatotroph adenoma
Samples (HOM)	10 (0)	4 (0)
OR (raw/adj. p)	4.2169 (0.0115/0.4496)	6.7226 (0.0682/0.8192)
Binomial p (adj.)	3.98E-05 (0.0249)	0.0004 (0.1787)
Patients	**7022:** AVSD-partial, HLV, PAPVR, hemiazygous continuation, heterotaxy-polysplenia **7064:** AVSD	**7238:** DORV, VSD, persistent LSVC **7517:** ASD-secundum, hypoplastic ascending aorta, PDA **7528:** DORV, HRV, CoA

Table 18 Continued

	7105: DORV, VSD-subpulmonic, HRV, HAA **7106:** DORV, mitral atresia, VSD, HLHS **7277:** ASD-secundum **7539:** CoA, BAV **7557:** TOF with pulmonary atresia, Possible compound heterozygote **7659:** DILV, TGA, VSD **7748:** TOF **7765:** BAV, CoA	**7557:** TOF with pulmonary atresia, possible compound heterozygote

4.3.2 ACKR3

Atypical chemokine receptor 3 (*ACKR3*) was one of four CHD-ome genes along with *APLNR*, *HTR2B*, and *S1PR1* that were among the multiple terms associated with GPCR signaling that came up in a ToppGene analysis of genes with an overall allele FC ≥ 5. Though variants in the other 3 genes did not show any clear phenotype pattern, three of six patients harboring a c.892C>T p.His298Tyr variant had PTA with two of them also having persistent left superior vena cava (Table 19). However, one of these three patients has extracardiac abnormalities and could thus represent an undiagnosed syndromic condition. There was an additional c.1007C>T p.Ala336Val variant present in a single CHD patient with TOF. Neither variant had a significant OR, but the His292Tyr variant showed a significant excess of alleles in the CHD cases (BH FDR = 0.0029). This gene is significantly enriched in the CHD cases overall (BH FDR = 0.0019) and preliminary data indicated an odds ratio of 47.5 for having PTA given an *ACKR3* variant (χ^2 p-value = 1.73E-14). *ACKR3* has been identified as a candidate gene for the dextro-transposition of the great arteries seen in patients with dup(2)(q37.3) (Costain et al., 2016). This gene, also known as *Cxcr7*, encodes a chemokine receptor that plays a key role in valve development and has been shown to cause semilunar valve defects with occasional malalignment of the aorta when mutated in mice, though defects in aortopulmonary septation were not seen (Sierro et al., 2007; S. Yu et al., 2011)

Table 19: Summary of patients with *ACKR3* variants

Variant	2:237490000:C:T	2:237490115:C:T
HGVS	c.892C>T; p.His298Tyr	c.1007C>T; p.Ala336Val
CADD PRED	27.5	23.4
GnomAD AF	0.00223276	0.0005329
ClinVar	Not curated	Not curated
HGMD	DM? (Renal cell carcinoma, susceptibility)	Not curated
Samples (HOM)	6 (0)	1 (0)
OR (raw/adj. p)	10.10 (0.0129/0.4496)	1.6763 (1/1)
Binomal p (adj.)	3.13E-06 (0.0029)	0.1208 (1)
Patients	**7266** (White F): PTA, persistent left SVC, ASD-secundum, product of diabetic mother **7288** (White M): PTA, persistent left SVC, genitourinary defects **7320** (White M): c-TGA, ASD-secundum **7535** (White F): Ebstein's malformation **7587** (White M): VSD, PDA, multiple extracardiac malformations **7740** (White F): Truncus arteriosus, S/P heart transplant	**7231** (White F): TOF

4.3.3 ZFYVE16

ToppGene analysis of the set of genes with allele FC ≥ 2 in highly damaging variants produced terms that encompassed eight CHD-ome genes including *ZFYVE16* which was associated with the term "Protein phosphatase 1 regulatory subunits". Of the four patients harboring a c.4055C>T p.Thr1352Ile variant in *ZFYVE16* (Zinc finger FYVE-type containing 16), three have TGA (Table 20). There was a significant excess of *ZFYVE16* alleles in CHD cases overall (BH FDR = 0.0157) and a significant excess of highly damaging alleles before adjustment (p = 0.0075). At the variant level, neither variant had a significant odds ratio, nor did they have a significant excess of allele in CHD cases after adjustment. Preliminarily, a χ^2 test indicated that the odds of having TGA given a *ZFYVE16* variant were not significant (p = 0.1315). Additionally, one of the three patients with c.127G>A p.Val43Ile has heterotaxy. Variants in this gene have

been associated with situs abnormalities in humans (A. H. Li et al., 2019). The FYVE-type domain containing zinc finger proteins have been implicated in the regulation of TGF-β signaling through SMADs (Moustakas & Heldin, 2002).

Table 20: Summary of patients with *ZFYVE16* variants

Variant	5:79732631:G:A	5:79768610:C:T
HGVS	c.127G>A, p.Val43Ile	c.4055C>T, p.Thr1352Ile
CADD PHRED	23.5	32
GnomAD AF	0.0014	0.0010
ClinVar	Not curated	Not curated
HGMD	Not curated	Not curated
Samples (HOM)	3 (0)	4 (0)
OR (raw/adj. p)	5.0376 (0.1501/1)	3.3596 (0.2048/1)
Binomial p (adj.)	0.0033 (1)	0.0075 (1)
Patients	**6014:** AVSD-complete, PAPVR, Interrupted IVC, Hemiazygous continuation, Heterotaxy **7758:** AVSD, Down syndrome **7766:** DILV, Pulmonary stenosis, HRV, VSD	**7364:** c-TGA, VSD-muscular, right-sided aortic arch **7482:** VSD-perimembranous, pulmonary stenosis, ASD-secundum **7593:** TGA, VSD-muscular **7660:** AVSD-unbalanced, DORV, c-TGA, PAPVR

4.3.4 ABI1

Variants in *ABI1* (Abl interactor 1) have been associated with LVOTO defects with extracardiac anomalies in a large study (Jin et al., 2017). There were three patients in our study cohort each heterozygous for the same highly damaging *ABI1* variant of whom two had LVOTO defects (Table 21). This variant showed a significant excess of highly damaging alleles overall ($p = 0.0033$), but lost significance after adjustment. For IPA enrichment analysis of genes with allele FC ≥ 2 for genes with highly damaging analysis, this gene came up associated with the terms "Abnormal morphology of myocardium" and "Morphology of myocardium". In mice, *Abi1* has been shown to play a critical role in reorganization of the actin cytoskeleton and cell migration

through the WAVE2 complex with null embryos displaying cardiac and other defects (Dubielecka et al., 2011).

Table 21: Summary of patients with ABI1 variant

Variant	10:27066073:G:A
HGVS	c.383C>T; p.Ala128Val
CADD PHRED	30
GnomAD AF	0.000354
ClinVar	Not curated
HGMD	Not curated
Samples (HOM)	3 (0)
OR (raw/adj. p)	5.0376 (0.1501, 1)
Binomial p (adj.)	0.0033 (1)
Patients	**3003:** Single ventricle, TAPVR, AVSD-complete, PCD, heterotaxy **7351:** HLHS, Shone's complex **7428: P**arachute mitral valve, hypoplastic aortic arch, aortic stenosis with BAV, ASD-secundum, Shone's complex

4.3.5 ARHGEF4

Rho guanine nucleotide exchange factor 4 (*ARHGEF4*) showed a significant enrichment for highly damaging alleles before adjustment ($p = 0.0228$), but lost significance after adjustment. This gene also came up in a ToppGene analysis of genes with allele FC ≥ 2 for highly damaging variants associated with the term "Rho guanine nucleotide exchange factors". Overall, there were five variants in the combined CHD and SCOOP exome data, but one of these variants was only present in SCOOP controls. Though this gene has been identified as a candidate gene accounting for the TOF seen in patients with del(2)(q22.1) (Silversides et al., 2012), none of the five patients with *ARHGEF4* variants had TOF, though they all have left-sided lesions (Table 22). Though not associated with structural heart defects in mice, *Arhgef4* (also known as *Asef*) has been shown to

be required for vascular endothelial cell migration mediated by FGF and VEGF signaling (Kawasaki et al., 2010).

Table 22: Summary of patients with *ARHGEF4* variants

Variant	2:131688627:G:C	2:131688828:A:G	2:131704208:G:A	2:131803637:C:T
HGVS	c.97G>C; p.Asp33His	c.298A>G; p.Lys100Glu	c.427G>A; p.Ala143Thr	c.1948C>T; p.Arg650Cys
CADD PHRED	11.24	14.06	12.77	34
GnomAD AF	0.0015	0.0005	0.0003	0.0007
ClinVar	Not curated	Not curated	Not curated	Not curated
HGMD	Not curated	Not curated	Not curated	Not curated
Samples (HOM)	1 (0)	1 (0)	1 (0)	1 (0)
OR (raw/adj. p)	0.8377 (1/1)	0.3346 (0.4208/1)	0.8377 (1/1)	3.3556 (0.5600/1)
Binomial p (adj.)	0.3351 (1)	0.7988 (1)	0.3351 (1)	0.0228 (1)
Patients	**7221:** Shone's complex, aortic stenosis, CoA, PAPVR	**7112:** BAV	**7004:** DILV, TGA, IAA, VSD	**7302:** HLHS, ASD-secundum **7669:** TGA

4.3.6 BCAR1

Mutations in the *BCAR1* scaffold protein, Cas family member gene have been linked to some cases of TOF (Reuter et al., 2018). In our study cohort, there were 11 total BCAR1 variants with one of those being highly damaging. One of the three patients carrying this highly damaging variant had TOF as did one patient each carrying one of two other variants (Table 23). *BCAR1* showed a significant enrichment for alleles overall (BH FDR = 0.0005), but only unadjusted for alleles in highly damaging variants (p = 0.0331). Similar to the previously mentioned *ABI1*, this gene came up in an IPA enrichment analysis of genes with allele FC ≥ 2 for genes with highly damaging analysis, this gene came up associated with the terms "Abnormal morphology of myocardium" and "Morphology of myocardium". *Bcar1* (also known as *Crkas*) encodes the

p130^{Cas} protein and plays a role in actin filament assembly with mutant mice displaying poorly developed hearts with disorganized myofibrils and disrupted Z-disks (Honda et al., 1998)

Table 23: Summary of patients with *BCAR1* variants

Variant	16:75268977:C:T	16:75269046:C:T	16:75269697:G:A
HGVS	c.1958G>A; p.Arg653Gln	c.1889G>A; p.Arg630Gln	c.1238C>T; p.Pro413Leu
CADD PHRED	27.3	34	23.4
GnomAD AF	0.0053	0.0009	0.0029
ClinVar	Not curated	Not curated	Not curated
HGMD	Not curated	Not curated	Not curated
Samples (HOM)	11 (0)	3 (0)	6 (0)
OR (raw/adj. p)	1.1535 (0.6951/1)	2.5175 (0.3691, 1)	1.6792 (0.3822 (1)
Binomial p (adj.)	0.2525 (1)	0.0331 (1)	0.0714 (1)
Patients	**6016:** Tricuspid atresia, VSD, HRV, pulmonary stenosis, persistent left SVC **7130:** VSD, aortic stenosis with BAV, CoA **7219:** TOF **7244:** TGA, tricuspid atresia, VSD, ASD-secundum **7271:** DILV, c-TGA, VSD, HRV **7357:** Ebstein's malformation, ASD-secundum **7447:** DILV, straddling tricuspid valve, ASD-secundum, HRV **7484:** VSD **7610:** DORV, VSD-subpulmonic, Double orifice mitral valve, IAA **7661:** Dextrocardia, AVSD-unbalanced, Right-sided aortic arch, Pulmonary atresia, Right atrial isomerism **7684:** Aortic stenosis with BAV, CoA	**7096:** TGA **7104:** TOF **7275:** Aortic stenosis with BAV	**7083:** DILV, aortopulmonary collaterals **7264:** TOF **7386:** HLHS, CoA, ASD-secundum **7400:** HLHS, ASD-secundum **7544:** Superior sinus venosus defect, PAPVR **7708:** Dextrocardia, TGA, HRV, Pulmonary atresia

4.3.7 C21ORF2

A study by Shanshen et al. (2018) reported that *C21ORF2*, also known as *CFAP410* (Cilia and flagella associated protein 410) was a candidate gene to explain the CHD seen in patients with del(21)(q22.13-q22.3) based on a review and the presence of a mouse knockout model, however my own search was unable to turn up any evidence of a mouse knockout that displayed any structural CHD phenotypes. Still, in looking over our patient data, there seems to be a high prevalence of aortic valve and aortic arch phenotypes among patients carrying one of the three variants present in our dataset (Table 24). Though the function of *C21ORF2* is not fully understood, a small study suggested possible roles in ciliogenesis and hedgehog signaling (Wheway et al., 2015). In our patient population, there is a significant excess of *C21ORF2* alleles overall (BH FDR = 0.0008), however significance for an excess of pathogenic alleles is lost after adjustment (p = 0.0228). This gene came up associated with a variety of significant terms including several non-motile cilia-related terms in a ToppGene analysis of genes with an allele FC ≥ 2 for variants curated as pathogenic.

Table 24: Summary of patients with *C21ORF2* variants

Variant	21:45750380:G:A	21:45751814:C:T	21:45753071:C:G
HGVS	c.965C>T; p.Ala322Val	c.457G>A: p.Gly153Ser	c.218G>C; p.Arg73Pro
CADD PHRED	16.71	21.8	23.9
GnomAD AF	0.0040	0.0062	0.0003
ClinVar	Not curated	Not curated	Pathogenic/Likely pathogenic: Retinitis pigmentosa; Cone/cone-rod dystrophy; Spondylometaphyseal dysplasia axial
HGMD	Not curated	Not curated	DM: Jeune syndrome; Spondylometaphyseal dysplasia, axial; Retinitis pigmentosa, simplex; Cone dystrophy; Retinal dystrophy

Table 24 Continued

Samples (HOM)	4 (0)	7 (0)	1 (1)
OR (raw/adj. p)	6.7226 (0.0682, 0.8192)	2.3541 (0.1449, 1)	3.3556 (0.5600 (1)
Binomial p (adj.)	0.0004 (0.1787)	0.0114 (1)	0.0228 (1)
Patients	**7002:** TGA, VSD, ASD-secundum **7198:** aortic stenosis with BAV, CoA, Short stature, Turner syndrome ***7380:** VSD-inlet, aortic stenosis, CoA, PDA **7405:** TAPVR, hypoplastic aortic arch	**6065:** AVSD-unbalanced, interrupted IVC, ASD **7091:** TOF, pulmonary atresia, right-sided aortic arch **7113:** DORV, AVSD-unbalanced, parachute mitral valve, HLV **7446:** HRV (PA/IVS), Scoliosis **7504:** AVSD-partial **7672:** CoA **7766:** DILV, Pulmonary stenosis, HRV, VSD	**7505:** Aortic stenosis with BAV
*Possible compound heterozygote with 21:45753117:C:A (c.172G>T; p.Val58Leu) present in 23 total samples			

4.3.8 FLNA

Mutations in filamin A (*FLNA*) have been associated with X-linked valvular dysplasia, mitral valve disease, and Melnick-Needles syndrome, the most severe form of the otopalatodigital syndromes which can include TOF (Azhar & Ware, 2016; LaHaye et al., 2014; Lalani & Belmont, 2014; Prendiville et al., 2014; Wessels & Willems, 2010). There were two variants in this gene present in our dataset, both present as possible compound heterozygotes in three female patients as well as two SCOOP samples. The variants are located about 15 kb apart and could represent either a haplogroup or some sort of linkage. This along with the variants being present in both cases and controls suggests it may be benign, though it is noteworthy that all three CHD patients have valve defects as well as extracardiac defects that may represent subtle features of the otopalatodigital syndromes (Table 25). Overall, there was a significant excess of alleles in CHD cases before adjustment ($p = 0.0112$), however this significance was lost after adjustment. This

gene had initially shown up in a ToppGene analysis of CGC genes associated with several terms related to valvular dysplasia as well as the term "abnormal oxygen level". This latter association likely refers to its regulation and modification by the oxygen sensor *PHD2* as it relates to the control of dendritic spines and synapses (Segura et al., 2016).

Table 25: Summary of patients with *FLNA* variants

Variant	X:153578465:G:A	X:153593616:G:A
HGVS	c.7267C>T; p.Pro2423Ser	c.1579C>T; p.Arg527Cys
CADD PHRED	19.18	28
GnomAD AF	0.000435	0.000441
ClinVar	Benign/likely benign - Connective tissue disorder; Melnick-Needles syndrome; Oto-palato-digital syndrome, type II; Periventricular nodular heterotopia 1; Frontometaphyseal dysplasia; Cardiovascular phenotype	Conflicting interpretations: Benign; Likely benign; Uncertain significance - Connective tissue disorder; Melnick-Needles syndrome; Oto-palato-digital syndrome, type II; Periventricular nodular heterotopia 1; Frontometaphyseal dysplasia; Cardiovascular phenotype
HGMD	Not curated	Not curated
Samples (HOM)	3 (0)	3 (0)
OR (raw/adj. p)	2.5175 (0.3691/1)	2.5175 (0.3691/1)
Binomial p (adj.)	0.0331 (1)	0.0331 (1)
Patients	7146: HLHS, aortic atresia, mitral atresia, hypoplastic aortic arch, ankyloglossia, hip dysplasia 7223: VSD, pulmonary stenosis, Dysmorphic features 7744: VSD, Subpulmonic stenosis, Genitourinary defect, underdeveloped thumbs	

4.3.9 SMAD6

Our patient cohort showed an excess of highly damaging variants in SMAD6 (BH FDR = 0.0003) as well as an excess of highly damaging alleles associated with the phenotype "Loeys-Dietz syndrome 3" caused by *SMAD6* mutations. Variants in *SMAD6* (SMAD family member 6) have been associated with a variety of mainly left-sided lesions in both humans and mice

(Andersen et al., 2014; Chaix et al., 2016; Fahed et al., 2013; LaHaye et al., 2014; Y. Li et al., 2015; C. J. Lin et al., 2012; X. Liu et al., 2017; Prendiville et al., 2014; D. Srivastava & Olson, 2000). Contruncal defects (TOF, DORV, TGA) were common among the 12 patients in our study cohort carrying a SMAD6 variant (Table 26). *SMAD6* is one of two inhibitory SMAD (I-SMAD) proteins along with *SMAD7* that act as inhibitors of TGF-β signaling through a variety of mechanisms (Reviewed in Miyazawa & Miyazono, 2017). It acts to repress BMP signaling during normal development of the heart valves and outflow tract.

Table 26: Summary of patients with *SMAD6* variants

Variant	15:66995657:G:A
HGVS	c.61G>A; p.Asp21Asn
CADD PHRED	30
GnomAD AF	0.005063
ClinVar	Benign: Loeys-Dietz syndrome 3
HGMD	Not curated
Samples (HOM)	12 (0)
OR (raw/adj. p)	5.0690 (0.0030, 0.3073)
Binomial p (adj.)	1.39E-06 (0.0014)
Patients	**7086:** TOF, RAA **7226:** TOF with pulmonary atresia, aortopulmonary collateral, DiGeorge syndrome **7323:** TGA **7378:** TOF with pulmonary atresia **7435:** TGA **7442:** TGA **7453:** HLHS, coronary artery fistula, ASD-secundum **7513:** DORV, pulmonary stenosis, right-sided aortic arch, Goldenhar syndrome **7539:** CoA, BAV **7559:** DORV, mitral atresia, VSD-perimembranous, HLV, Dysmorphic features **7562:** TGA **7675:** IAA, VSD, subaortic stenosis

4.3.10 TRIM32

CHD patients had a significant excess of alleles in rare highly damaging variants (BH FDR = 0.00045) in the Tripartite motif-containing 32 (*TRIM32*) gene. It also came up in a ToppGene analysis of genes with an allele FC ≥ 2 for rare damaging variants associated with the term "striated muscle myosin thick filament". Variants in this gene have been associated with the ciliopathy, Bardeet-Biedl syndrome, which can occasionally include VSD or dextrocardia (Lalani & Belmont, 2014). *TRIM32* has also been shown to be involved in cardiomyocyte hypertrophy as well as being implicated as a key regulator of cardiomyocyte viability and apoptosis (Borlepawar et al., 2017). Of the four variants present in our dataset, only the damaging Arg408Cys variant was significant though one patient with this variant and one with the Gln186His variant had dextrocardia (Table 27).

Table 27: Summary of patients with *TRIM32* variants

Variant	9:119460398:C:T	9:119460542:C:T	9:119460579:G:C	9:119461243:C:T
HGVS	c.377C>T; p.Ala126Val	c.521C>T; p.Ser174Phe	c.558G>C; p.Gln186His	c.1222C>T; p.Arg408Cys
CADD PHRED	21.3	25.9	23.2	33
GnomAD AF	1.25E-05	0.0002	0.0019	0.0015
ClinVar	Not curated	Uncertain significance – Bardet-Biedl syndrome	Likely benign – Bardet-Biedl syndrome	Conflicting interpretations: Likely benign; Likely pathogenic; Uncertain significance – Bardet-Biedl syndrome; Bardet-Biedl syndrome 11
HGMD	Not curated	Not curated	Not curated	DM?: Usher syndrome
Samples (HOM)	1 (0)	1 (0)	2 (0)	5 (0)
OR (raw/adj. p)	0.8377 (1, 1)	1.6763 (1, 1)	0.8376 (1, 1)	8.4105 (0.0300, 0.6083)
Binomial p (adj.)	0.3350 (1)	0.1208 (1)	0.4270 (1)	3.73E-05 (0.0242)
Patients	**7351:** HLHS	**7562:** TGA	**7386:** HLHS, CoA, ASD-secundum	**7276:** BAV **7330:** BAV

Table 27 Continued

			7635: Dextrocardia, AVSD-unbalanced, DORV, Azygous continuation, Heterotaxy-asplenia	**7384:** Hypoplastic right ventricle (pulmonary atresia with intact ventricular septum) **7563:** TOF, VSD, Right aortic arch **7661:** Dextrocardia, AVSD-unbalanced, Right aortic arch, Pulmonary atresia, Right atrial isomerism, Heterotaxy-asplenia

4.3.11 ZIC3

The association between *ZIC3* (Zinc finger protein of cerebellum 3) and heterotaxy spectrum CHD has been well documented (Andersen et al., 2014; Azhar & Ware, 2016; Cowan et al., 2014; De Luca et al., 2010; Deng et al., 2015; Fahed et al., 2013; Finsterer et al., 2017; Hilger et al., 2015; Lalani & Belmont, 2014; Lebo & Baxter, 2014; A. H. Li et al., 2019; McCulley & Black, 2012; Pierpont et al., 2007; Prendiville et al., 2014; Richards & Garg, 2010; Sutherland & Ware, 2009; Wessels & Willems, 2010; Xie et al., 2017). This gene encodes a multifunctional zinc finger transcription factor that can interact with components of multiple developmental signaling pathways including Hedgehog, Wnt, Planar Cell Polarity, and TGF-β (Reviewed in Bellchambers & Ware, 2018). There was a significant excess of curated pathogenic ZIC3 alleles among CHD patients (BH FDR = 0.0020) while also being associated with several phenotypes and enrichment analysis terms related to cardiac malformations, heterotaxy, and VATER/VACTERL association. Of the five patients carrying the single *ZIC3* variant in our dataset, three had diagnosed syndromic conditions that could account for their CHD while the other two had heterotaxy consistent with variants in this gene (Table 28). Of the two male patients, the one with Down syndrome was reported as being homozygous in the exome data, though the other male patient was reported as being heterozygous. As the exome data did not consider hemizygosity for

X chromosome loci in males and instead treated them as homozygous calls, it would have been expected that all male patients would have been homozygous for all X chromosome loci. This could perhaps be explained by low quality sites in some male patients with different reads reporting different nucleotides.

Table 28: Summary of patients with *ZIC3* variants

Variant	X:136648899:G:T
HGVS	c.49G>T; p.Gly17Cys
CADD PHRED	24.9
GnomAD AF	0.0026
ClinVar	Conflicting interpretations: Benign; Likely benign; Pathogenic - VACTERL association with hydrocephaly, X-linked; Heterotaxy, visceral, X-linked; Congenital heart defects 1, nonsyndromic, 1
HGMD	DM: Cardiac malformation; VATER/VACTERL association
Samples (HOM)	5 (2)*
OR (raw/adj. p)	5.8944 (0.0165, 0.4893)
Binomial p (adj.)	3.5E-05 (0.0231)
Patients	**2007 (M)::** TOF, AVSD, Down syndrome **7007 (M):** DORV, AVSD-unbalanced, Azygous continuation, Bilateral SVC, Left atrial isomerism, Heterotaxy-polysplenia **7198 (F):** Aortic stenosis with BAV, CoA, Short stature, Turner syndrome **7283 (F)**: AVSD-unbalanced, Dextrocardia, Heterotaxy-asplenia **7719 (F):** Hypoplastic right ventricle, CHARGE syndrome
*The 2 homozygous calls were a presumably hemizygous male with Down syndrome and a female with Turner syndrome (monosomy X), though the other male patient with a ZIC3 variant was reported as heterozygous.	

5.0 Final Summary and Conclusions

The goal of this study was to investigate whether patients with congenital heart disease have an increased burden of rare pathogenic or highly damaging variants in genes known to be associated with heart development and genes known to be associated with cancer. To this end, there were a number of CHD-ome genes that came up frequently including *ACKR3*, *AIP*, *C21ORF2*, *CCDC39*, *FKRP*, *FLT4*, *G6PD*, *NRG1*, *SMAD6*, and *ZIC3*. Of these genes, *SMAD6* and *ZIC3* are well-described causes of congenital heart disease in humans while most of the other genes have been linked with human CHD or heart defects in mouse models. While only eight genes showed a significant excess of alleles in pathogenic variants in CHD cases, four of those genes were CHD-ome genes whereas while 93 genes have a significant excess of alleles in highly damaging variants in CHD cases, only six of them were CHD-ome genes. This seems to suggest that highly damaging variants in genes associated with heart development are not well tolerated in the CHD population.

These results hint at a complex model where multiple variants of low to moderate damaging impact combine to cause CHD in patients. In these cases, multiple variants of low to moderate impact may be sufficient to disturb the complex developmental pathway involved in heart development to the point of causing CHD. If just one of those variants were to be replaced by a highly damaging variant, this could then disrupt these developmental pathways to such an extent as to be incompatible with life. Highly damaging variants in CHD-ome genes may be more tolerated in a control population due to a lack of additional variants in other CHD-ome genes such that other pathways could compensate for the effect of a single highly damaging variant. Alternatively, the presence of significant terms related to heart development in the SCOOP gene

sets could indicate that variants in these genes are somehow exerting a protective effect perhaps by counterbalancing the effects of variants in other developmental genes. It would thus be interesting to investigate whether there is a difference in the number of variants in CHD-ome genes between cases and controls and if there is a difference in the average functional impact of the variants present. It may also be worthwhile to further subdivide the cases into mild CHD and complex/severe CHD groups.

A next logical step would have been to investigate whether or not any of the genes that came up repeatedly seem to be associated with any particular type or class of CHD in the study cohort. Preliminary work did in fact suggest links between *AIP* and DORV, *ACKR3* and PTA, and *ZFYVE16* and TGA. These variants would first need to be validated by Sanger sequencing in our patients and then tested against a replication cohort to investigate if the same relationships occur. If validated and replicated, these variants could be generated in the orthologous mouse genes through CRISPR-*Cas9* gene editing to determine if they cause the same or similar defects in mice. A dataset was obtained from the Pediatric Cardiac Genomics Consortium (Jin et al., 2017) to use as a replication cohort, however processing of the data was not completed in time to carry out the analysis for this dissertation.

Though only 10 genes with curated cancer phenotypes in our dataset had a significant excess of alleles in CHD cases, five of those genes were CHD-ome genes which agrees with the hypothesis that genes associated with cancer play a significant role in heart development. Furthermore, ToppGene analysis of these 10 genes suggests that genes involved in the apoptosis pathway in particular are playing a role. Of the 35 CGC genes with an overall excess of alleles in CHD cases, 11 were CHD-ome genes which again suggests genes involved in cancer are playing a role in heart development. Results pointed toward those cancer genes involved with apoptosis

as playing a significant role in the etiology of CHD. However, despite the overall significance, there were very few genes with variants with curated cancer phenotypes or CGC genes that showed an excess of pathogenic or highly damaging alleles. I feel these results again hint at the previously described complex etiology of CHD. Pathway enrichment analysis of these cancer genes produced several terms related to oxygen levels and radiation-related neoplasms which could both be of interest as chronic hypoxia and radiation exposure during interventional treatments have both been linked to cancer (Cohen et al., 2018; Oleaga-Alday et al., 2015; Opotowsky et al., 2015; Yakoumakis et al., 2013).

This study was limited by several factors, the most notable of which were a non-random study population and a lack of control for population stratification. Due to the selection criteria for study participants, there is a bias toward heterotaxy spectrum CHD and HLHS which in turn could skew results to show enrichment for variants in genes associated with these classes of CHD and not for other classes. Because the demographic information for the SCOOP controls was not immediately available, this was not controlled for. However, the distribution of average number of calls per subject between cohorts suggests that there is a much larger proportion of non-whites among the CHD cases than the SCOOP controls. The SCOOP cohort itself is not an ideal control cohort because the subjects all have severe obesity and are thus not 'normal'. Furthermore, the presence of subclinical CHD such as BAV or small septal defects cannot be excluded from these subjects. Data was obtained from the Framingham Heart Study (Eicher et al., 2017; Tsao & Vasan, 2015) to use as a control, however this data was acquired too late in the process to redo the analysis.

There were also a number of limitations to the data analysis methods used beyond the lack of control for population stratification. Enrichment analysis was done based on fold change of normalized allele counts with the cutoff chosen somewhat arbitrarily to obtain lists that were not

too long or too short. It might have made more sense to create these lists based on significant odds ratios, however these calculations were not carried out until after the enrichment analysis had been done based on fold changes. It would have been necessary to perform these analyses using raw p-values as adjusted p-values essentially eliminated all significance. This loss of significance was likely related to very large difference between number of mutant and wild type alleles in the cohorts. Though the tests for significance for pathogenic and highly damaging alleles were limited to the set of genes with such alleles, the analysis of different subsets was not restricted in this way. For instance, the analysis of genes with pathogenic variants curated as causing cancer phenotypes was based on the analysis of the entire set of genes with curated pathogenic variants regardless of phenotypes. This may have artificially reduced the number of significant genes; however, it could also suggest that genes associated with cancer phenotypes are not particularly enriched in our study population relative to all genes with curated phenotypes. Finally, the exact and binomial tests cannot adequately control for population stratification and other covariates inherent in genetic analysis and it would have been preferable to use a more advanced methodology such as the sequence kernel association test (Ionita-Laza, Lee, Makarov, Buxbaum, & Lin, 2013; M. C. Wu et al., 2011; Yan et al., 2014) for these analyses.

Despite these limitations, this research still provides a foundation on which to guide future research. This research has provided more evidence of a complex, multigenic etiology for at least some CHD as well as preliminarily identifying some disease-gene associations. Furthermore, genes associated with response to oxygen levels and radiation-induced cancers were identified as possible prognostic factors. Variants in genes related to response to oxygen levels could provide information as to how well patients with cyanotic CHD will tolerate prolonged cyanosis as well as to their increased risk of cancer due to chronic activation of the hypoxia pathway. Variants in

genes associated with radiation-induced cancers could serve as risk factors to identify patients at increased risk of such cancers from repeated interventional exposures. Together, these results are of public health significance as they provide a basis for the further understanding of the etiology of CHD, for the development of novel treatment strategies, and for the assessment of risks of late complications.

Appendix A Tables

Appendix Table 1 Structural Cardiac Defects in Our Study Population

Structural Cardiac Defect	Number of Patients
Transposition of the great arteries	84
VSD (unspecified)	76
Secundum ASD	75
Coarctation of the aorta	73
Hypoplastic left heart syndrome	70
Double outlet right ventricle	65
Pulmonary stenosis	64
Tetralogy of Fallot	60
Bicuspid aortic valve	47
Aortic stenosis with bicuspid aortic valve	36
Right-sided aortic arch	31
Congenitally corrected transposition of the great arteries	30
Ebstein's malformation of the tricuspid valve	28
Hypoplastic aortic arch	25
Pulmonary atresia	25
AVSD-unbalanced	24
Aortic stenosis	22
Tricuspid atresia	19
VSD-perimembranous	19
Hypoplastic right ventricle (Pulmonary atresia with intact septum)	18
TOF with pulmonary atresia	18
Double inlet left ventricle	17
Patent ductus arteriosus	17
Hypoplastic right ventricle	16
Total anomalous pulmonary venous return	16
Mitral atresia	15
Hypoplastic left ventricle	14
Aortopulmonary collaterals	13
AVSD (unspecified)	13
Dextrocardia	13
Partial anomalous pulmonary venous return	12
Persistent left superior vena cava	12
Subaortic stenosis	12
Truncus arteriosus (common arterial trunk)	12
VSD-muscular	11

Appendix Table 1 Continued

Aortic atresia	10
Interrupted aorta	10
Mitral stenosis	10
ASD (unspecified)	8
AVSD-partial	8
Interrupted inferior vena cava	8
Multiple unspecified VSDs	8
Azygous continuation	7
Coronary artery anomaly	7
Coronary artery fistula	7
Parachute mitral valve	7
AVSD-balanced	6
AVSD-complete	6
Tricuspid stenosis	6
Bilateral superior vena cava	5
Left isomerism of the atrial appendages	5
Hemiazygous continuation	4
Juxtaposition of the atrial appendages to the right	4
Mitral hypoplasia	4
Subpulmonic VSD	4
Mesocardia	3
Shone's complex	3
Single ventricle	3
Bicuspid pulmonary valve	2
Dilated left ventricle	2
Hypertrophy of the right ventricle	2
Pulmonary atresia with intact ventricular septum	2
Subpulmonic stenosis	2
Doubly committed VSD	2
Inlet VSD	2
Outlet VSD	2
Ssubaortic VSD	2
Sinus venosus ASD	1
Criss-cross atrioventricular valves	1
Dextroversion	1
Dilated aortic root	1
Distal origin of the right subclavian artery	1
Double orifice mitral valve	1
Double orifice tricuspid valve	1
Hypoplastic right heart syndrome	1
Hypertrophy of the left ventricle	1
Hypoplastic ascending aorta	1

Appendix Table 1 Continued

Malposition of great vessels	1
Overriding and straddling tricuspid valve	1
Overriding tricuspid valve	1
Right isomerism of the atrial appendages	1
Straddling tricuspid valve	1
Superior sinus venosus defect	1
Apical trabecular VSD	1
VSD-noncommitted	1

Appendix Table 2: Significant CHD-ome and Ciliome genes overall

	Gene Set			Allele Count			Odds Ratio		
SYMBOL	**CHD100**	**CHD-ome**	**Ciliome**	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
PEPD	FALSE	TRUE	FALSE	4.1766	2.15E-12	3.04E-09	4.9178	2.85E-06	0.0029
G6PD	FALSE	TRUE	FALSE	14.4020	4.86E-11	4.11E-08	16.9706	0.0004	0.0489
CASQ2	FALSE	TRUE	FALSE	6.1209	1E-10	5.8E-08	7.2136	5.18E-05	0.0151
NCOR2	FALSE	TRUE	FALSE	1.0110	5.96E-09	1.46E-06	1.2026	2.84E-06	0.0029
ABCC3	FALSE	TRUE	FALSE	1.7471	1.37E-08	3.05E-06	2.0560	3.44E-05	0.0151
HTR2B	FALSE	TRUE	FALSE	7.2010	5.68E-08	9.03E-06	8.5156	0.0014	0.0996
MYLK2	FALSE	TRUE	FALSE	10.0814	2.3E-07	3.02E-05	11.8970	0.0052	0.2003
UQCRC2	FALSE	TRUE	FALSE	10.0814	2.3E-07	3.02E-05	11.8970	0.0052	0.2003
AIP	FALSE	TRUE	FALSE	4.0326	6.09E-07	6.56E-05	4.7552	0.0015	0.0996
SHANK3	FALSE	TRUE	FALSE	3.6005	1.21E-06	0.0001	4.2359	0.0023	0.1328
SMAD6	TRUE	TRUE	FALSE	4.3206	1.39E-06	0.0001	5.1130	0.0029	0.1540
RYR2	FALSE	TRUE	TRUE	1.9941	2.48E-06	0.0002	2.3449	0.0009	0.0768
WNK3	FALSE	TRUE	FALSE	8.6412	3.16E-06	0.0002	10.1648	0.0125	0.2836
DAG1	FALSE	TRUE	FALSE	5.7608	4.55E-06	0.0004	6.8007	0.0073	0.2236
PEX6	FALSE	TRUE	FALSE	2.7364	4.73E-06	0.0004	3.2174	0.0031	0.1565
PPARD	FALSE	TRUE	FALSE	2.3043	7.1E-06	0.0005	2.7402	0.0023	0.1344
BCAR1	FALSE	TRUE	FALSE	1.6802	7.25E-06	0.0005	1.9756	0.0015	0.0996
ECE2	FALSE	TRUE	FALSE	2.2503	8.01E-06	0.0005	2.6460	0.0030	0.1565
CCNH	FALSE	TRUE	FALSE	3.9606	7.87E-06	0.0005	4.6654	0.0058	0.2103
UBR2	FALSE	TRUE	FALSE	3.9606	7.92E-06	0.0005	4.6596	0.0058	0.2103
C21orf2	FALSE	TRUE	TRUE	1.8517	1.2E-05	0.0008	2.1823	0.0026	0.1443
PCDHA9	FALSE	TRUE	FALSE	4.3206	1.8E-05	0.0011	5.1023	0.0125	0.2836
MTHFD1	FALSE	TRUE	TRUE	2.0163	2E-05	0.0012	2.3729	0.0040	0.1719
MYBPC3	FALSE	TRUE	FALSE	2.0466	2.11E-05	0.0012	2.4051	0.0034	0.1585
BBS2	FALSE	TRUE	TRUE	3.1204	2.32E-05	0.0013	3.6771	0.0074	0.2236
CASP5	FALSE	TRUE	FALSE	3.1204	2.34E-05	0.0013	3.6720	0.0074	0.2239
IFT88	FALSE	TRUE	TRUE	2.8804	3.22E-05	0.0018	3.3852	0.0108	0.2836
ZIC3	FALSE	TRUE	TRUE	5.0407	3.5E-05	0.0019	5.9455	0.0160	0.3177
NR0B1	FALSE	TRUE	FALSE	5.0407	3.5E-05	0.0019	5.9455	0.0160	0.3177
CFLAR	FALSE	TRUE	FALSE	5.0407	3.5E-05	0.0019	5.9455	0.0160	0.3177
ACKR3	FALSE	TRUE	FALSE	5.0407	3.53E-05	0.0019	5.9305	0.0161	0.3177
VANGL1	FALSE	TRUE	FALSE	5.0407	3.53E-05	0.0019	5.9305	0.0161	0.3177
ALG13	FALSE	TRUE	FALSE	7.2010	3.73E-05	0.0020	8.4831	0.0293	0.3929
BVES	FALSE	TRUE	FALSE	7.2010	3.73E-05	0.0020	8.4831	0.0293	0.3929
S1PR1	FALSE	TRUE	FALSE	7.2010	3.73E-05	0.0020	8.4831	0.0293	0.3929
ACE	FALSE	TRUE	FALSE	1.4638	3.9E-05	0.0021	1.7203	0.0027	0.1540

Appendix Table 2 Continued

PRKAG2	FALSE	TRUE	FALSE	3.6005	4.04E-05	0.0021	4.2394	0.0112	0.2836
PRKDC	FALSE	TRUE	FALSE	1.8785	4.15E-05	0.0022	2.2077	0.0042	0.1762
CCDC39	TRUE	TRUE	TRUE	1.5711	4.55E-05	0.0023	1.8482	0.0034	0.1585
ALPK3	FALSE	TRUE	FALSE	2.3567	5.68E-05	0.0028	2.7703	0.0069	0.2236
SLC25A46	FALSE	TRUE	FALSE	2.0332	5.76E-05	0.0028	2.3919	0.0057	0.2103
BBS12	FALSE	TRUE	TRUE	2.1123	6.05E-05	0.0029	2.4860	0.0062	0.2203
SPAG1	FALSE	TRUE	TRUE	3.1684	6.68E-05	0.0031	3.7313	0.0164	0.3177
PCDHA2	FALSE	TRUE	FALSE	3.1684	6.71E-05	0.0031	3.7270	0.0165	0.3177
C5orf42	TRUE	TRUE	TRUE	1.5362	7.79E-05	0.0035	1.8056	0.0050	0.2003
USP44	FALSE	TRUE	FALSE	2.8804	9.38E-05	0.0042	3.3886	0.0133	0.2905
FSTL3	FALSE	TRUE	FALSE	3.8405	0.0001	0.0046	4.5315	0.0240	0.3688
PDLIM3	FALSE	TRUE	FALSE	3.8405	0.0001	0.0046	4.5315	0.0240	0.3688
KCNAB1	FALSE	TRUE	FALSE	3.8405	0.0001	0.0046	4.5193	0.0241	0.3688
FPR1	FALSE	TRUE	FALSE	2.5204	0.0001	0.0058	2.9657	0.0139	0.2980
MEGF8	TRUE	TRUE	TRUE	1.5510	0.0002	0.0063	1.8235	0.0078	0.2328
SOX4	FALSE	TRUE	FALSE	1.7282	0.0002	0.0066	2.0372	0.0110	0.2836
CRELD1	FALSE	TRUE	TRUE	3.2405	0.0002	0.0072	3.8102	0.0216	0.3682
ETV2	FALSE	TRUE	FALSE	3.2405	0.0002	0.0072	3.8138	0.0216	0.3682
ADAMTS10	FALSE	TRUE	FALSE	1.5637	0.0002	0.0089	1.8396	0.0105	0.2824
NRG1	FALSE	TRUE	FALSE	1.7831	0.0003	0.0092	2.0967	0.0147	0.3092
PTBP1	FALSE	TRUE	FALSE	1.8903	0.0004	0.0125	2.2295	0.0165	0.3177
ZNF423	FALSE	TRUE	TRUE	1.6706	0.0004	0.0126	1.9646	0.0158	0.3177
EGFR	FALSE	TRUE	FALSE	5.7608	0.0004	0.0127	6.7807	0.0669	0.5869
ABCC6	FALSE	TRUE	FALSE	5.7608	0.0004	0.0127	6.7807	0.0669	0.5869
GP1BB	FALSE	TRUE	FALSE	5.7608	0.0004	0.0127	6.7807	0.0669	0.5869
EDNRA	FALSE	TRUE	FALSE	2.4689	0.0004	0.0131	2.9173	0.0293	0.3929
PCDHA12	FALSE	TRUE	FALSE	1.9203	0.0004	0.0132	2.2567	0.0214	0.3682
EVC2	FALSE	TRUE	TRUE	1.5202	0.0004	0.0138	1.7867	0.0154	0.3177
TXNRD2	FALSE	TRUE	FALSE	1.9546	0.0004	0.0138	2.2991	0.0186	0.3362
CHRNA1	FALSE	TRUE	FALSE	2.2403	0.0005	0.0142	2.6331	0.0285	0.3929
DOT1L	FALSE	TRUE	FALSE	2.2403	0.0005	0.0142	2.6331	0.0285	0.3929
ENG	FALSE	TRUE	FALSE	1.6907	0.0005	0.0143	1.9918	0.0181	0.3296
TCOF1	FALSE	TRUE	FALSE	1.3168	0.0005	0.0150	1.5496	0.0120	0.2836
EVL	TRUE	TRUE	FALSE	3.3605	0.0005	0.0157	3.9616	0.0458	0.5146
GJC1	FALSE	TRUE	FALSE	3.3605	0.0005	0.0157	3.9616	0.0458	0.5146
RIPPLY3	FALSE	TRUE	FALSE	3.3605	0.0005	0.0157	3.9527	0.0460	0.5146
ZFYVE16	FALSE	TRUE	FALSE	3.3605	0.0005	0.0157	3.9527	0.0460	0.5146
MYO18B	FALSE	TRUE	FALSE	1.3859	0.0005	0.0157	1.6287	0.0145	0.3075
GPR126	FALSE	TRUE	FALSE	1.7145	0.0006	0.0165	2.0153	0.0211	0.3682
XIRP2	FALSE	TRUE	FALSE	1.7282	0.0006	0.0179	2.0313	0.0193	0.3465
PIK3CB	FALSE	TRUE	FALSE	2.8804	0.0008	0.0227	3.3886	0.0674	0.5869

Appendix Table 2 Continued

CCDC40	FALSE	TRUE	TRUE	1.7791	0.0008	0.0227	2.0912	0.0278	0.3929
TP53BP2	FALSE	TRUE	FALSE	1.2624	0.0008	0.0228	1.4875	0.0179	0.3279
CNTN3	FALSE	TRUE	TRUE	1.5362	0.0009	0.0237	1.8060	0.0245	0.3728
TLDC1	FALSE	TRUE	FALSE	1.5395	0.0010	0.0259	1.8098	0.0229	0.3688
SGCZ	FALSE	TRUE	FALSE	2.5924	0.0011	0.0271	3.0463	0.0501	0.5267
HECTD1	TRUE	TRUE	TRUE	2.5924	0.0011	0.0271	3.0476	0.0501	0.5267
PTPRJ	FALSE	TRUE	FALSE	1.8833	0.0011	0.0288	2.2125	0.0361	0.4534
ARID1A	FALSE	TRUE	FALSE	1.8833	0.0011	0.0289	2.2132	0.0361	0.4534
TTN	FALSE	TRUE	FALSE	-1.0209	0.0011	0.0290	1.1500	0.0118	0.2836
CCDC148	FALSE	TRUE	FALSE	2.4003	0.0012	0.0302	2.8222	0.0659	0.5869
TRPS1	FALSE	TRUE	FALSE	2.4003	0.0012	0.0302	2.8222	0.0659	0.5869
NEK2	FALSE	TRUE	TRUE	2.2632	0.0013	0.0317	2.6639	0.0485	0.5266
BASP1	FALSE	TRUE	FALSE	2.0163	0.0013	0.0318	2.3718	0.0358	0.4521
SPEN	FALSE	TRUE	FALSE	2.1603	0.0014	0.0320	2.5376	0.0392	0.4820
PTCH1	FALSE	TRUE	TRUE	2.0803	0.0013	0.0321	2.4446	0.0451	0.5146
SIRT1	FALSE	TRUE	FALSE	2.0803	0.0013	0.0321	2.4495	0.0449	0.5146
CLUAP1	FALSE	TRUE	TRUE	2.0803	0.0013	0.0321	2.4468	0.0450	0.5146
TNFSF11	FALSE	TRUE	FALSE	2.0803	0.0013	0.0321	2.4468	0.0450	0.5146
MYOCD	FALSE	TRUE	FALSE	2.1603	0.0013	0.0321	2.5393	0.0391	0.4818
SETDB2	FALSE	TRUE	FALSE	3.6005	0.0014	0.0340	4.2394	0.1097	0.7091
SIN3B	FALSE	TRUE	FALSE	3.6005	0.0014	0.0340	4.2394	0.1097	0.7091
USP8	FALSE	TRUE	FALSE	3.6005	0.0015	0.0340	4.2324	0.1099	0.7091
FLNB	FALSE	TRUE	FALSE	1.2268	0.0020	0.0447	1.4414	0.0277	0.3929
NPHP4	FALSE	TRUE	TRUE	1.4023	0.0020	0.0450	1.6474	0.0316	0.4157

Appendix Table 3: ToppGene analysis of genes with allele FC ≥ 5 overall

Name	Genes in query	p-value	q-value FDR B&H
GPCR downstream signaling	33 (PDE7A, OR6K2, MCF2, CCL27, F2, CXCR5, OR8A1, OR10A4, ACKR3, AGTR2, OR1G1, TAAR9, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, GNGT2, OR9A2, OR2AG1, OR5P2, HTR2B, HTR4, OR5L1, OR5AS1, OR5K3, OR5H14, S1PR1, DGKK, OR2Y1, OR51V1, GPSM1, NMS)	4.20E-06	5.06E-03
Signaling by GPCR	38 (PDE7A, OR6K2, PSMA5, MCF2, PSMC6, CCL27, F2, CXCR5, OR8A1, OR10A4, ACKR3, AGTR2, OR1G1, TAAR9, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, GNGT2, OR9A2, OR2AG1, OR5P2, HTR2B, HTR4, OR5L1, OR5AS1, WNT8B, OR5K3, OR5H14, CNKSR2, S1PR1, EGFR, DGKK, OR2Y1, OR51V1, GPSM1, NMS)	1.78E-05	9.22E-03
Olfactory transduction	18 (OR6K2, OR8A1, OR10A4, OR1G1, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, OR9A2, OR2AG1, OR5P2, OR5L1, OR5AS1, OR5K3, OR5H14, OR2Y1, OR51V1)	2.30E-05	9.22E-03
Olfactory Signaling Pathway	18 (OR6K2, OR8A1, OR10A4, OR1G1, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, OR9A2, OR2AG1, OR5P2, OR5L1, OR5AS1, OR5K3, OR5H14, OR2Y1, OR51V1)	3.77E-05	1.13E-02
Lupus Erythematosus, Systemic	28 (DCAF1, IL10RB, IL12A, NAT2, PDE7A, PSMA5, DAG1, TNIP1, NCL, F2, SOCS2, CXCR5, OR10A4, ACKR3, BANK1, PNP, OR2AG1, HNRNPDL, PLXNA3, TIMD4, S1PR1, ABCC6, TNFRSF11B, EGFR, EBI3, CD27, CD36, TNFRSF1A)	7.64E-06	2.17E-02
G protein-coupled receptor signaling pathway	41 (KLK6, PDE7A, OR6K2, MCF2, CCL27, F2, CXCR5, OR8A1, DEFB1, OR10A4, ACKR3, AGTR2, OR1G1, GPSM2, TAAR9, CCL17, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, GNGT2, OR9A2, CA2, OR2AG1, OR5P2, GPR3, HTR2B, HTR4, OR5L1, OR5AS1, OR5K3, OR5H14, S1PR1, DGKK, OR2Y1, MRGPRX3, MRAP2, OR51V1, GPSM1, NMS)	5.33E-06	2.20E-02
olfactory receptor activity	18 (OR6K2, OR8A1, OR10A4, OR1G1, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, OR9A2, OR2AG1, OR5P2, OR5L1, OR5AS1, OR5K3, OR5H14, OR2Y1, OR51V1)	5.84E-05	2.76E-02
transmembrane signaling receptor activity	38 (IL10RB, CRLF1, OR6K2, TNFRSF10A, CXCR5, OR8A1, OR10A4, ACKR3, AGTR2, OR1G1, TAAR9, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, OR9A2, GP1BB, OR2AG1, OR5P2, PLXNA3, GPR3, HTR2B, HTR4, OR5L1, OR5AS1, OR5K3, GRIA3, OR5H14, CLDN3, S1PR1, EGFR, EBI3, CD27, OR2Y1, MRGPRX3, TNFRSF1A, OR51V1)	1.22E-04	2.76E-02
odorant binding	8 (OR8A1, OR8G5, OR8D4, OR5P2, OR5L1, OR5AS1, OR5K3, OR5H14)	1.31E-04	2.76E-02
signaling receptor activity	43 (IL10RB, CRLF1, OR6K2, TNFRSF10A, F2, CXCR5, OR8A1, OR10A4, AMOT, ACKR3, AGTR2, NR0B1, OR1G1, TAAR9, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, OR9A2, GP1BB, OR2AG1, OR5P2, PLXNA3, GPR3, HTR2B, HTR4, OR5L1, OR5AS1, OR5K3, GRIA3, OR5H14, CLDN3, S1PR1, TNFRSF11B, EGFR, EBI3, CD27, CD36, OR2Y1, MRGPRX3, TNFRSF1A, OR51V1)	1.65E-04	2.76E-02
G protein-coupled receptor activity	27 (OR6K2, CXCR5, OR8A1, OR10A4, ACKR3, AGTR2, OR1G1, TAAR9, OR8G5, OR10G8, OR6C6, OR8D4, OR7G2, OR9A2, OR2AG1, OR5P2, GPR3, HTR2B, HTR4, OR5L1, OR5AS1, OR5K3, OR5H14, S1PR1, OR2Y1, MRGPRX3, OR51V1)	1.83E-04	2.76E-02

Appendix Table 3 Continued

Tumor necrosis factor receptor superfamily	4 (TNFRSF10A, TNFRSF11B, CD27, TNFRSF1A)	2.71E-04	4.23E-02

Appendix Table 4: IPA analysis of genes with allele FC ≥ 5 overall

Diseases and Bio Functions	GS1	GS1b	Molecules
Apoptosis of dermal cells	3.55061	3.398593	CFLAR, EGFR, PIAS4, PRDX6, TNFRSF10A, TNFRSF1A
Apoptosis of synovial cells	3.203127	3.096519	CFLAR, S1PR1, TNFRSF10A, TNFRSF11B
Apoptosis of keratinocytes	2.931953	2.806023	EGFR, PIAS4, PRDX6, TNFRSF10A, TNFRSF1A
Apoptosis of fibroblast-like synoviocytes	2.91593	2.833972	S1PR1, TNFRSF10A, TNFRSF11B
Beat of cells	2.855215	2.798741	CA2, CASQ2
Progressive cardiomyopathy	2.855215	2.798741	DAG1, MED30
Apoptosis of microvascular endothelial cells	2.598588	2.517939	CD36, TNFRSF10A, TNFRSF1A
Apoptosis of epithelial cells	2.439064	2.241002	CFLAR, EGFR, F2, NCL, PIAS4, PRDX6, S1PR1, TNFRSF10A, TNFRSF1A, TNIP1
Apoptosis of connective tissue cells	2.301996	2.108166	CD36, CFLAR, CRADD, DCAF1, EGFR, S1PR1, TNFRSF10A, TNFRSF11B, TNFRSF1A, TNIP1
Survival of vascular endothelial cells	2.224656	2.146084	CFLAR, PRDX6, S1PR1
Abnormal morphology of dilated vasculature	2.135477	2.080411	AMOT, G6PD
Apoptosis of bladder cancer cell lines	2.111823	2.034024	CFLAR, DCAF1, EGFR
Apoptosis of endothelial cells	2.061799	1.930709	CD36, CFLAR, PRDX6, TNFRSF10A, TNFRSF11B, TNFRSF1A
Apoptosis of CD8+ T lymphocyte	2.059721	2.004888	CD27, CFLAR
Apoptosis of peripheral B lymphocytes	1.921618	1.893077	CFLAR
Apoptosis of pleural cells	1.921618	1.893077	EGFR
Lipoapoptosis of hepatocytes	1.921618	1.893077	TNFRSF10A
Quantity of trabecular myocardium	1.921618	1.893077	ZIC3
Vacterl with hydrocephalus, X-linked	1.921618	1.893077	ZIC3
X-linked multiple types congenital heart defects type 1	1.921618	1.893077	ZIC3
X-linked visceral heterotaxy 1	1.921618	1.893077	ZIC3
Hypertrophy of heart	1.886004	2.562005	AGTR2, CASQ2, CCDC134, CD36, EGFR, HTR2B, KCNJ11, MYLK2, PFKFB1, RRAD, TCF15, TNFRSF1A
Cardiac death	1.880552	0	CA2, F2, KCNJ11, SCNN1A
Failure of heart	1.674018	0	CA2, CD36, CFLAR, F2, HTR2B, KCNJ11, PDE7A, RRAD, SCNN1A

Appendix Table 5: Significant cardiovascular and cilia phenotypes overall

		Allele count			Odds ratio		
Phenotype	**Variants**	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
Cardiomyopathy, non-compaction, left ventricular	1	15.8422	2.33E-12	3.5E-09	18.7602	0.0001	0.0488
Ventricular tachycardia, catecholaminergic polymorphic, 2	3	6.1209	1E-10	5.02E-08	7.2136	5.18E-05	0.0380
Dilated cardiomyopathy 1JJ	5	3.4565	5.14E-09	1.78E-06	4.0691	0.0001	0.0487
Hypertrophic cardiomyopathy	126	1.1275	1.07E-08	3.02E-06	1.3243	8.03E-06	0.0132
Dilated Cardiomyopathy, Dominant	139	1.0677	8.13E-07	0.0001	1.2539	0.0001	0.0487
Loeys-Dietz syndrome 3	1	4.3206	1.39E-06	0.0002	5.1130	0.0029	0.1547
Loeys-Dietz syndrome 2	6	2.3043	7.77E-06	0.0008	2.7102	0.0024	0.1370
Joubert syndrome 17	8	1.7367	1.6E-05	0.0013	2.0428	0.0025	0.1394
Cardiac malformation	1	5.0407	3.5E-05	0.0023	5.9455	0.0160	0.3555
Congenital heart defects 1, nonsyndromic, 1	1	5.0407	3.5E-05	0.0023	5.9455	0.0160	0.3555
Heterotaxy, visceral, X-linked	1	5.0407	3.5E-05	0.0023	5.9455	0.0160	0.3555
Bardet-Biedl syndrome 11	1	7.2010	3.73E-05	0.0024	8.4831	0.0293	0.4791
Ciliary dyskinesia	119	1.0416	4.05E-05	0.0025	1.2231	0.0015	0.1204
Ciliary dyskinesia, primary, 28	2	3.1684	6.68E-05	0.0036	3.7313	0.0164	0.3572
Cardiovascular phenotype	383	-1.0546	0.0002	0.0072	1.1131	0.0022	0.1253
Atrioventricular septal defect 2	3	3.2405	0.0002	0.0084	3.8102	0.0216	0.4284
Noncompaction, left ventricular	17	1.3157	0.0002	0.0097	1.5462	0.0076	0.2591
Ciliary dyskinesia, primary, 37	23	1.2326	0.0003	0.0131	1.4482	0.0092	0.2883
Ciliary dyskinesia, primary, 33	7	1.5302	0.0007	0.0224	1.7992	0.0210	0.4284
Hypoplastic left heart syndrome 1	2	2.8804	0.0025	0.0617	3.3865	0.0861	0.7513

Appendix Table 6: ToppGene analysis of genes with allele FC ≥ 2 for highly damaging variants

Name	Genes in Query List	p-value	q-value FDR B&H
Solute carriers	18 (SLC8B1, SLC30A10, SLC25A24, SLC7A14, SLC2A2, SLC6A13, SLC25A10, SLC4A4, SLC16A13, SLC17A7, SLC35E4, SLC27A3, SLC7A8, SLC16A3, SLC7A5, SLC14A2, SLC16A8, SLC4A11)	3.67E-04	2.70E-02
EF-hand domain containing	11 (CABP5, MCC, RYR2, RYR3, SLC25A24, EFCAB6, ITSN1, DUOX2, EFHC2, CHP2, SDF4)	2.37E-03	4.76E-02
Protein phosphatase 1 regulatory subunits	10 (TSC2, SH3RF2, CAMSAP3, SLC7A14, NCOR1, PPP1R12C, ZFYVE16, AKAP11, RIMBP2, MPHOSPH10)	1.85E-03	4.17E-02
Fibronectin type III domain containing|I-set domain containing|Immunoglobulin like domain containing	10 (NRG1, ROBO4, FLT4, MYBPC2, MYBPC3, MYOM1, PTPRD, ADAMTSL1, PAPLN, KIRREL1)	7.63E-04	3.37E-02
Receptor Tyrosine Kinases|Fibronectin type III domain containing|V-set domain containing|Immunoglobulin like domain containing	12 (ROBO4, FANK1, MYBPC2, MYBPC3, MYOM1, ITGB4, FNDC7, PTPRC, PTPRD, PTPRH, ROS1, EPHA8)	3.86E-05	8.52E-03
Rho guanine nucleotide exchange factors	6 (ARHGEF19, ITSN1, ARHGEF4, ARHGEF11, ARHGEF17, FARP2)	1.28E-03	4.03E-02
Keratins, type I	4 (KRT26, KRT24, KRT36, KRT28)	1.58E-03	4.17E-02
DNA polymerases	4 (POLD1, POLL, DNTT, POLM)	7.33E-04	3.37E-02
ATP binding cassette subfamily A	3 (ABCA7, ABCA10, ABCA8)	1.89E-03	4.17E-02
Fibronectin type III domain containing|I-set domain containing|Myosin binding proteins	3 (MYBPC2, MYBPC3, MYOM1)	3.15E-04	2.70E-02
striated muscle myosin thick filament	3 (TRIM32, MYBPC3, MYOM1)	4.99E-05	3.04E-02
Ryanodine receptors|EF-hand domain containing	2 (RYR2, RYR3)	9.84E-04	3.63E-02

Appendix Table 7: Significant caner phenotypes overall

		Allele Count			Odds Ratio		
Phenotype	**Variants**	**FC**	**p-value**	**BH FDR**	**OR**	**p-value**	**BH FDR**
Familial Isolated Pituitary Adenomas	2	4.0326	6.09E-07	9.45E-05	4.7552	0.0015	0.1204
Pituitary adenoma	2	4.0326	6.09E-07	9.45E-05	4.7552	0.0015	0.1204
Somatotroph adenoma	2	4.0326	6.09E-07	9.45E-05	4.7552	0.0015	0.1204
Renal cell carcinoma, susceptibility	1	8.6412	3.13E-06	0.0004	10.1886	0.0125	0.3159
Lymphoproliferative syndrome 2	1	5.7608	4.55E-06	0.0005	6.8007	0.0073	0.2543
Adenomatous polyposis coli, attenuated	1	7.2010	3.73E-05	0.0024	8.4831	0.0293	0.4791
Hereditary nonpolyposis colorectal cancer type 8	1	7.2010	3.73E-05	0.0024	8.4831	0.0293	0.4791
Subcutaneous panniculitis-like T cell lymphomas with haemophagocytic lymphohistiocytic syndrome	2	2.2258	0.0002	0.0077	2.6240	0.0169	0.3586
Acromegaly & somatotroph adenoma	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Adrenocortical carcinoma	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Bladder, breast, colon, endometrial, male breast, ovarian, pilocytic astrocytoma & renal cancer, melanoma & colon polyps	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Breast and colorectal cancer, susceptibility to	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Breast cancer, and colorectal cancer association with	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Cancer in pancreatic cancer syndrome, association with	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Cancer of multiple types, susceptibility to	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Colorectal adenocarcinoma, endometrial carcinoma, esophagogastric adenocarcinoma, melanoma and other cancers	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Cowden/Cowden-like syndrome	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Differentiated thyroid carcinoma, increased risk	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Gastric cancer, association with	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Gastric cancer, increased risk, association with	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Gastrointestinal carcinoma	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
HNPCC-related colorectal cancer, association with	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802

Appendix Table 7 Continued

Leukemia risk	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Li-Fraumeni syndrome 2	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Li-Fraumeni syndrome, increased risk	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Lung cancer squamous, reduced risk	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Macroadenoma	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Multi-organ cancers	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Papillary thyroid carcinoma, association with	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Papillary thyroid carcinoma, increased risk	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Polycythaemia vera, increased risk	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Prostate cancer and additional primary cancers	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Thrombocythaemia, increased risk, association with	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Thyroid cancer, increased risk	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802
Uterine serous carcinoma	1	5.7608	0.0004	0.0140	6.7807	0.0669	0.6802

Appendix Table 8: ToppGene analysis of genes with allele FC ≥ 2 in variants with cancer phenotypes

Name	Genes in Query List	p-value	q-value FDR B&H
Neoplasm of the heart	6 (CHEK2, CASP10, LIG4, SDHB, CDK4, PTCH1)	3.18E-06	1.26E-04
regulation of apoptotic process	9 (CHEK2, EGFR, CD27, CASP10, LIG4, ACKR3, MSH2, CDK4, MITF)	2.18E-05	4.89E-03
regulation of programmed cell death	9 (CHEK2, EGFR, CD27, CASP10, LIG4, ACKR3, MSH2, CDK4, MITF)	2.52E-05	4.89E-03
regulation of cell death	9 (CHEK2, EGFR, CD27, CASP10, LIG4, ACKR3, MSH2, CDK4, MITF)	4.93E-05	5.48E-03
regulation of nitric oxide mediated signal transduction	2 (SPINK1, EGFR)	6.14E-05	5.92E-03
Cardiac fibroma	1 (PTCH1)	1.17E-03	6.10E-03
Cardiac Paraganglioma	1 (SDHB)	1.17E-03	6.10E-03
calcified aortic valve	2 (EGFR, ACKR3)	1.19E-04	6.31E-03
Heart Neoplasm	1 (SDHB)	2.33E-03	9.24E-03
intrinsic apoptotic signaling pathway in response to DNA damage	3 (CHEK2, ACKR3, MSH2)	1.92E-04	9.86E-03
apoptotic signaling pathway	5 (CHEK2, CD27, CASP10, ACKR3, MSH2)	4.07E-04	1.44E-02
nitric oxide mediated signal transduction	2 (SPINK1, EGFR)	4.28E-04	1.44E-02
negative regulation of apoptotic process	6 (EGFR, CD27, LIG4, ACKR3, MSH2, MITF)	4.03E-04	1.44E-02
negative regulation of programmed cell death	6 (EGFR, CD27, LIG4, ACKR3, MSH2, MITF)	4.53E-04	1.44E-02
Atrial myxoma	1 (MSH2)	5.82E-03	1.62E-02
aortic valve stenosis	2 (EGFR, ACKR3)	6.04E-04	1.62E-02
heart left ventricle outflow tract stenosis	2 (EGFR, ACKR3)	6.04E-04	1.62E-02
Abnormal venous morphology	5 (CHEK2, SPINK1, LIG4, UGT1A10, CDK4)	3.00E-03	1.65E-02
abnormal induced cell death	4 (CHEK2, TYR, LIG4, MSH2)	8.27E-04	1.79E-02
thick pulmonary valve	2 (EGFR, ACKR3)	8.46E-04	1.79E-02
Angiofibromas	2 (MSH2, SDHB)	3.51E-03	1.85E-02
abnormal cell death	9 (CHEK2, TYR, CD27, FANCB, LIG4, MSH2, ITK, AIP, PTCH1)	1.11E-03	1.94E-02
thick aortic valve	2 (EGFR, ACKR3)	1.13E-03	1.94E-02
negative regulation of cell death	6 (EGFR, CD27, LIG4, ACKR3, MSH2, MITF)	8.00E-04	1.99E-02
intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator	2 (CHEK2, MSH2)	1.03E-03	2.08E-02
abnormal apoptosis	8 (TYR, CD27, FANCB, LIG4, MSH2, ITK, AIP, PTCH1)	1.31E-03	2.15E-02
Abnormal vascular morphology	9 (CHEK2, SPINK1, CASP10, FANCB, LIG4, UGT1A10, SDHB, AIP, CDK4)	5.52E-03	2.55E-02
Myxoma of heart	1 (EGFR)	1.51E-02	2.99E-02

Appendix Table 8 Continued

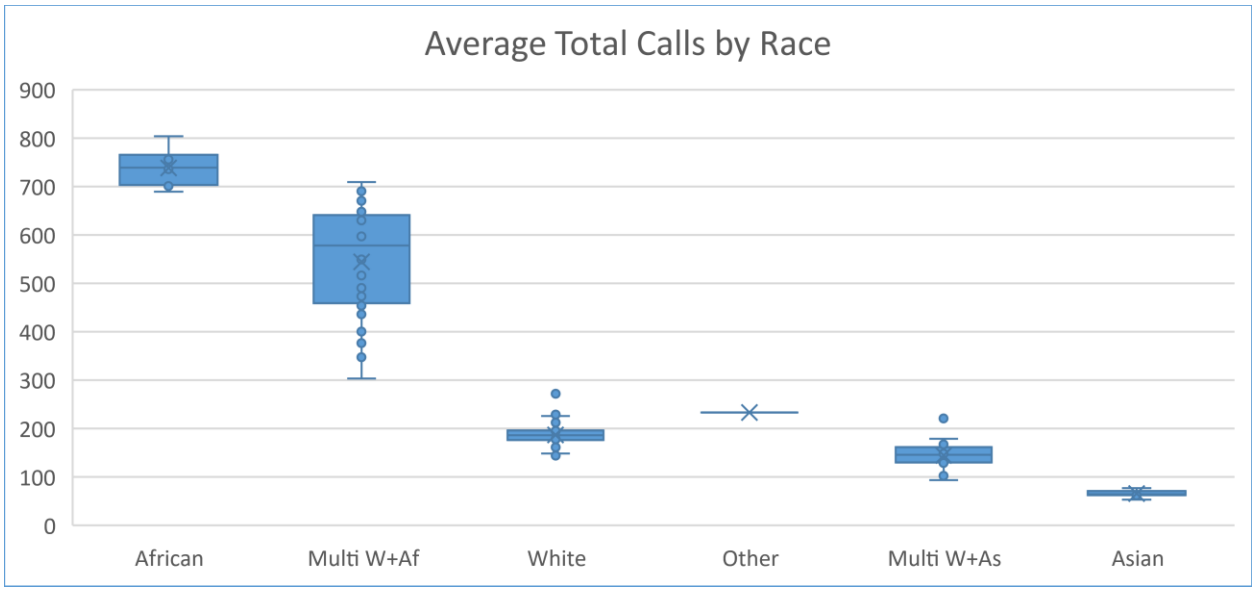
Myxoma of the Endocardium	1 (EGFR)	1.74E-02	3.27E-02
increased sensitivity to induced cell death	3 (TYR, LIG4, MSH2)	3.52E-03	3.63E-02
negative regulation of nitric oxide mediated signal transduction	1 (SPINK1)	2.98E-03	3.67E-02
intrinsic apoptotic signaling pathway by p53 class mediator	2 (CHEK2, MSH2)	3.04E-03	3.71E-02
hedgehog receptor activity	1 (PTCH1)	3.12E-03	3.84E-02
smoothened binding	1 (PTCH1)	3.12E-03	3.84E-02
Neoplasms, Radiation-Induced	1 (PTCH1)	2.31E-02	3.92E-02
intrinsic apoptotic signaling pathway	3 (CHEK2, ACKR3, MSH2)	3.61E-03	3.98E-02
abnormal heart valve morphology	3 (EGFR, ACKR3, CDK4)	4.43E-03	4.25E-02
hedgehog family protein binding	1 (PTCH1)	5.19E-03	4.28E-02
canonical Wnt signaling pathway involved in negative regulation of apoptotic process	1 (MITF)	4.95E-03	4.31E-02
positive regulation of nitric oxide mediated signal transduction	1 (EGFR)	4.95E-03	4.31E-02
Abnormal heart morphology	10 (CHEK2, CASP10, FANCB, LIG4, UGT1A10, MSH2, SDHB, AIP, CDK4, PTCH1)	1.16E-02	4.35E-02
patent ductus venosus	1 (AIP)	5.03E-03	4.42E-02
Endocardial Cushion Defects	1 (EGFR)	2.88E-02	4.55E-02
patched binding	1 (PTCH1)	8.29E-03	4.80E-02
Bicuspid aortic valve	1 (EGFR)	3.11E-02	4.80E-02
Cardiac Lymphoma	1 (CD27)	3.22E-02	4.91E-02

Appendix Table 9: Breakdown of variants by ClinVar and HGMD pathogenicity

ClinVar Clinical Significance	HGMD Pathogenicity						
	DM	DM|DM?	DM?	DP	DFP	FP	NA
Pathogenic	311	0	95	8	7	6	58
Pathogenic/Likely pathogenic	43	0	4	0	0	0	4
Likely pathogenic	126	0	42	4	6	1	29
Benign	318	1	806	58	28	33	3109
Benign/Likely benign	65	0	228	24	7	2	728
Likely benign	293	1	655	48	22	17	2267
Conflicting interpretations	229	1	342	16	10	6	868
Affects	2	0	0	0	0	4	2
association	0	0	1	0	1	0	0
drug response	6	0	0	0	3	4	0
protective	2	0	1	1	1	0	0
risk factor	17	0	10	4	8	0	0
other	4	0	3	0	0	3	0
Uncertain significance	369	1	482	21	13	14	2230
not provided	9	0	11	1	2	2	78
NA	180	0	468	122	37	123	29476

Orange cells are variants that were included using the "and" definition of pathogenicity while yellow cells are variants that would have been considered pathogenic using the "or" definition. Note that many variants had multiple significance assertions in ClinVar so the numbers are not exclusive. DM = Disease-associated mutation, DM? = Disease associated mutation without strong evidence, DP = Disease-associated polymorphism, DFP = Disease-associated functional polymorphism, FP = Functional polymorphism, NA = Not applicable.

Appendix B Figures



Appendix Figure 1: Average number of calls by race for the CHD cohort

Breaking down the number of calls by sample by race as determined by principal component analysis of exome data showed that Africans had far more calls than Whites and Asians with the mixed races being intermediate between their component races. The cutoff for assigning one race exclusively was 85% and the one sample marked 'other' was found to be 84% white and 8% each of African and Asian.

Appendix C Lists

Appendix C.1 Gene Sets for Enrichment Analysis

The following is a list of gene sets (GS) used for enrichment analysis in ToppGene and IPA. Initially, these analyses were done only with genes for which a numeric fold change could be calculated, i.e. those with alleles in both the CHD and SCOOP cohorts. The analyses were then redone to include "unique" genes that only had variants in that category (overall, pathogenic, CADD ≥ 30) in one cohort such that for the sets with FC ≥, "unique" refers to the CHD cohort and for sets with FC ≤, "unique" refers to the SCOOP cohort. Also listed are the total number of genes in each set. Gene sets 7 and 8 were to be genes with FC ≥ 1 or FC ≤ −1 for pathogenic alleles, but these sets were not looked at. For the Cosmic Gene Census (CGC) gene sets there were no genes with variants only in the CHD cohort, so gene sets 15, 17, and 19 did not have "b" sets. Additionally, only nine CGC genes had pathogenic alleles so no enrichment analyses were done for these.

- GS1: Set of genes with FC ≥ 5 (n = 278)
    - GS1b: Set of genes with FC ≥ 5 plus unique (n = 297)
- GS2: Set of genes with FC ≤ −5 overall (n = 58)
    - GS2b: Set of genes with FC ≤ −5 overall plus unique (n = 217)
- GS3: Genes with FC ≥ 2 and homozygous calls only in CHD (n = 247)
- GS4: Genes with FC ≤ −2 and homozygous calls only in SCOOP (n = 46)
- GS5: Genes with FC ≥ 2 for pathogenic alleles (n = 35)
    - GS5b: Genes with FC ≥ 2 plus unique for pathogenic alleles (n = 36)

- GS6: Genes with FC ≤ −2 for pathogenic alleles (n = 38)
    - GS6b: Genes with FC ≤ −2 plus unique for pathogenic alleles (n = 45)
- GS9: Genes with FC ≥ 2 for alleles in variants with CADD ≥ 30 (n = 478)
    - GS9b: Genes with FC ≥ 3 for alleles in variants with CADD ≥ 30 (n = 140)
    - GS9c: Genes with FC ≥ 2 plus unique for alleles in variants with CADD ≥ 30 (n = 489)
    - GS9d: Genes with FC ≥ 3 plus unique for alleles in variants with CADD ≥ 30 (n = 151)
- GS10: Genes with FC ≤ −2 for alleles in variants with CADD ≥ 30 (n = 370)
    - GS10b: Genes with FC ≤ −3 for alleles in variants with CADD ≥ 30 (n = 111)
    - GS10c: Genes with FC ≤ −2 plus unique for alleles in variants with CADD ≥ 30 (n = 481)
    - GS10d: Genes with FC ≤ −3 plus unique for alleles in variants with CADD ≥ 30 (n = 222)
- GS11: Genes with homozygous calls only in CHD for variants with CADD ≥ 30 (n = 50)
- GS12: Genes with homozygous calls only in SCOOP for variants with CADD ≥ 30 (n = 43)
- GS13: Genes with FC ≥ 2 for loss-of-function (LOF) alleles (n = 117)
    - GS13b: Genes with FC ≥ 2 + unique for LOF alleles (n = 120)
- GS14: Genes with FC ≤ −2 for LOF alleles (n = 82)
    - GS14b: Genes with FC ≤ −2 plus unique for LOF alleles (n = 105)
- GS15: CGC genes with FC ≥ 2 for variants with CADD ≥ 30 (n = 24)
- GS16: CGC genes with FC ≤ −2 for variants with CADD ≥ 30 (n = 12)

    - GS16b: CGC genes with FC ≤ −2 plus unique for variants with CADD ≥ 30 (n = 16)
- GS17: CGC genes with FC ≥ 2 overall (n = 60)
- GS18: CGC genes with FC ≤ −2 overall (n = 41)
    - GS18b CGC genes with FC ≤ −2 plus unique overall (n = 49)
- GS19: Genes with cancer phenotype and FC ≥ 2 plus unique (n = 20)
- GS20: Genes with cancer phenotype and FC ≤ −2 plus unique (n = 24)
- GS21: Genes with cancer phenotype and FC ≥ 2 plus unique for variants curated as pathogenic or with CADD ≥ 30 (n = 3 + 11 – 1 duplicate = 14)
- GS22: Genes with cancer phenotype and FC ≤ −2 plus unique for variants curated as pathogenic or with CADD ≥ 30 (n = 4 + 18 – 2 duplicates = 20)

Appendix C.2 ToppGene Parameters

ToppGene provides a variety of analysis and filtering options allowing you to modify the types of results you get. The p-value method can be set as either probability density function or cumulative distribution function while offering Bonferroni, Benjamini-Hochberg, and Benjamini-Yekutieli corrections. The p-value cutoff can also be adjusted as can the maximum and minimum number of genes associated with output terms. The "Feature options" settings provide a list of databases against which your gene set will be queried to search for enrichment, some of which have subcategories. By default, all of these are checked, however some of these were unchecked as they were deemed not relevant or did not produce any relevant results. Below are the parameters that were used for all ToppGene Analyses in this dissertation.

- Calculations
    - p-value method: Probability density function
    - Correction: FDR (Benjamini-Hochberg)
    - p-value cutoff: 0.05
    - Gene limits: $1 \leq n \leq 2000$
- Feature options
    - GO: Molecular Function
    - GO: Biological Process
    - GO: Cellular Component
    - Human Phenotype
    - Mouse Phenotype
    - Pathways
        - BioSystems: BIOCYC
        - BioSystems: KEGG
        - BioSystems: Pathway Interaction Database
        - BioSystems: REACTOME
        - GenMAPP
        - MSigDB C2 BIOCARTA (v6.0)
        - PantherDB
        - Pathway Ontology
        - SMPDB
    - Transcription Factor Binding Site
    - Gene Family

    - Disease
        - Clinical Variations
        - DisGeNET BeFree
        - DisGeNET Curated
        - GWAS
        - OMIM

Appendix C.3 IPA Parameters

Ingenuity Pathway Analysis (IPA) provides a large number of options to modify your parameters within each of the analysis types. For this project, which was only looking at gene set enrichment, the Variant Effect Analysis was used. General settings allow you to choose the reference set and what types of relationships to consider. "Network interactions" allow you to define the size and number of networks that will be generated as well as how they will be scored. "Node type" defines the types of molecules that will be included in networks. A large number of data sources are also available against which input gene sets will be queried for enrichment. There are also options to select the confidence level, restrict the species for which gene sets are compared, select the tissue types and cell lines to be considered, and the types of mutations to be considered. The parameters used for IPA analyses in this study are as follows.

- General Settings
    - Reference Set: Ingenuity Knowledge Base (Genes Only)
    - Relationships to Consider: Direct and Indirect
- Network Interactions

    - Interaction networks
        - Include endogenous chemicals: No
        - Molecules per network: 35
        - Networks per analysis: 25
    - Causal networks
        - Score using causal paths only: Yes
- Node types
    - Canonical pathways
    - Complex
    - Disease
    - Enzyme
    - Function
    - G-protein-coupled receptor
    - Group
    - Growth factor
    - Ion channel
    - Kinase
    - Ligand-dependent nuclear receptor
    - Peptidase
    - Phosphatase
    - Transcription regulator
    - Translation regulator
    - Transmembrane receptor

    - Transporter
    - Other
- Data sources
    - Ingenuity-supported 3rd party information
        - Additional sources
            - An open access database of Genome-Wide Association Results
            - BioGRID: An Online Interaction Respository With Data Compiled Through Comprehensive Curation Efforts
            - COSMIC
            - ClinVar
            - Gene Ontology (GO)
            - HumanCyc: Curated database of human metabolic pathways, enzymes, metabolites, and reactions
            - IntAct: Freely available, open source database system and analysis tools for molecular interaction data
            - Mouse Genome Database (MGD)
            - Obesity Gene Map Database
            - OMIM
    - Ingenuity expert information
        - Ingenuity Expert Findings
        - Ingenuity ExpertAssist Findings
- Confidence: Experimentally observed
- Species: Human, Mouse

- Tissues and Cell Lines Activated
    - Tissues and primary cells: Check all
    - Cell lines: Uncheck all
- Mutation
    - Functional effect
        - Gain of function
        - Knockout
        - Loss of function
        - Null
    - Inheritance model
        - Dominant
        - Recessive
        - X-linked
        - Y-linked
    - Translation impact
        - Frameshift
        - In frame
        - Nonsense
        - Missense
        - Silent
    - Unidentified mutation
    - Zygosity
        - Hemizygous

        - Homozygous
        - Heterozygous
    - Wild type

Appendix D Scripts

Below is the original unmodified script ManyFishersTest.R found on an online message board and freely available. This script allows the calculation of many Fisher's exact tests from data arranged in a large table such that the values for each cell of the 2×2 table constructed for each individual test corresponds to a column in the large table.

```
# Load necessary packages
library(purrr)
library(broom)
library(dplyr)
library(tidyr)

#create example data
## I created this to look like the data you have a screenshot of.
genes.table <- data.frame(gene = paste0("gene", 1:10),
                          mut.cont = runif(10, 0, 10) %>% round(),
                          WT.cont = runif(10, 0, 10) %>% round(),
                          mut.dis = runif(10, 0, 10) %>% round(),
                          WT.dis = runif(10, 0, 10) %>% round())

# Tidy the data
## this makes it easier to convert to 2x2 tables later, which are needed for fisher.test
genes.tidy <- genes.table %>%
  gather(-gene, key = treatment, value = expression) %>% #gathers treatments into one column
```

```
        separate(treatment, into = c("genotype", "treatment")) #splits into a
genotype and treatment column



      # Split by gene
      ## this creates a list of dataframes
      gene.list <- genes.tidy %>%
        split(.$gene)

      # Create tables
      ## This makes 2x2 tables for each gene
      gene.tables <- map(gene.list, ~xtabs(expression ~ genotype + treatment,
data = .))



      # Do fisher test and get p.value
      output <- map_dfr(gene.tables,
                        ~fisher.test(.) %>% tidy(), .id = "gene") %>%
        select(gene, p.value) %>%
        mutate(p.adj = p.adjust(p.value, "fdr"))

      output
```

This script was modified into a script called PolyExact.R by removing the block to create example data and changing the variable names to match my data.

Bibliography

Abdelwahid, E., Pelliniemi, L. J., & Jokinen, E. (2002). Cell death and differentiation in the development of the endocardial cushion of the embryonic heart. *Microsc Res Tech, 58*(5), 395-403. doi:10.1002/jemt.10159

Aboulhosn, J., & Child, J. S. (2015). Echocardiographic evaluation of congenital left ventricular outflow obstruction. *Echocardiography (Mount Kisco, N.Y.), 32 Suppl 2*, S140-S147. doi:10.1111/echo.12181

Aburawi, E. H., Aburawi, H. E., Bagnall, K. M., & Bhuiyan, Z. A. (2015). Molecular insight into heart development and congenital heart disease: An update review from the Arab countries. *Trends Cardiovasc Med, 25*(4), 291-301. doi:10.1016/j.tcm.2014.11.007

Ackerman, C., Locke, A. E., Feingold, E., Reshey, B., Espana, K., Thusberg, J., . . . Maslen, C. L. (2012). An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects. *Am J Hum Genet, 91*(4), 646-659. doi:10.1016/j.ajhg.2012.08.017

Agopian, A. J., Goldmuntz, E., Hakonarson, H., Sewda, A., Taylor, D., & Mitchell, L. E. (2017). Genome-Wide Association Studies and Meta-Analyses for Congenital Heart Defects. *Circ Cardiovasc Genet, 10*(3), e001449. doi:10.1161/circgenetics.116.001449

Akhirome, E. O., Regmi, S. D., Magnan, R., Ugwu, N., Qin, Y., Schulkey, C. E., . . . Jay, P. Y. (2019). The fitness cost of a congenital heart defect shapes its genetic architecture. *bioRxiv*, 531988. doi:10.1101/531988

Al Turki, S., Manickaraj, A. K., Mercer, C. L., Gerety, S. S., Hitz, M. P., Lindsay, S., . . . Hurles, M. E. (2014). Rare variants in NR2F2 cause congenital heart defects in humans. *Am J Hum Genet, 94*(4), 574-585. doi:10.1016/j.ajhg.2014.03.007

Allende, M. L., Yamashita, T., & Proia, R. L. (2003). G-protein-coupled receptor S1P1 acts within endothelial cells to regulate vascular maturation. *Blood, 102*(10), 3665-3667. doi:10.1182/blood-2003-02-0460

Andersen, T. A., Troelsen Kde, L., & Larsen, L. A. (2014). Of mice and men: molecular genetics of congenital heart disease. *Cell Mol Life Sci, 71*(8), 1327-1352. doi:10.1007/s00018-013-1430-1

Araki, T., Chan, G., Newbigging, S., Morikawa, L., Bronson, R. T., & Neel, B. G. (2009). Noonan syndrome cardiac defects are caused by PTPN11 acting in endocardium to enhance endocardial-mesenchymal transformation. *Proc Natl Acad Sci U S A, 106*(12), 4736-4741. doi:10.1073/pnas.0810053106

Arrington, C. B., Dowse, B. R., Bleyl, S. B., & Bowles, N. E. (2012). Non-synonymous variants in pre-B cell leukemia homeobox (PBX) genes are associated with congenital heart defects. *Eur J Med Genet, 55*(4), 235-237. doi:10.1016/j.ejmg.2012.02.002

Asrani, S. K., Warnes, C. A., & Kamath, P. S. (2013). Hepatocellular carcinoma after the Fontan procedure. *N Engl J Med, 368*(18), 1756-1757. doi:10.1056/NEJMc1214222

Augustyn, A., Peng, L., Singal, A. G., & Yopp, A. C. (2015). Surveillance for hepatocellular carcinoma secondary to cardiogenic cirrhosis in patients with congenital heart disease. *Clin Res Cardiol, 104*(5), 446-449. doi:10.1007/s00392-015-0809-4

Azhar, M., & Ware, S. M. (2016). Genetic and Developmental Basis of Cardiovascular Malformations. *Clin Perinatol, 43*(1), 39-53. doi:10.1016/j.clp.2015.11.002

Baasanjav, S., Al-Gazali, L., Hashiguchi, T., Mizumoto, S., Fischer, B., Horn, D., . . . Hoffmann, K. (2011). Faulty initiation of proteoglycan synthesis causes cardiac and joint defects. *Am J Hum Genet, 89*(1), 15-27. doi:10.1016/j.ajhg.2011.05.021

Barratt, K. S., Glanville-Jones, H. C., & Arkell, R. M. (2014). The Zic2 gene directs the formation and function of node cilia to control cardiac situs. *Genesis, 52*(6), 626-635. doi:10.1002/dvg.22767

Beamer, W. G., Pelsue, S. C., Shultz, L. D., Sundberg, J. P., & Barker, J. E. (1995). The flaky skin (fsn) mutation in mice: map location and description of the anemia. *Blood, 86*(8), 3220-3226. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7579418

Bellacosa, A. (2013). Developmental disease and cancer: biological and clinical overlaps. *Am J Med Genet A, 161a*(11), 2788-2796. doi:10.1002/ajmg.a.36267

Bellchambers, H. M., & Ware, S. M. (2018). ZIC3 in Heterotaxy. *Adv Exp Med Biol, 1046*, 301-327. doi:10.1007/978-981-10-7311-3_15

Billings, S. E., Pierzchalski, K., Butler Tjaden, N. E., Pang, X. Y., Trainor, P. A., Kane, M. A., & Moise, A. R. (2013). The retinaldehyde reductase DHRS3 is essential for preventing the formation of excess retinoic acid during embryonic development. *FASEB J, 27*(12), 4877-4889. doi:10.1096/fj.13-227967

Bjornsson, T., Thorolfsdottir, R. B., Sveinbjornsson, G., Sulem, P., Norddahl, G. L., Helgadottir, A., . . . Stefansson, K. (2018). A rare missense mutation in MYH6 associates with non-syndromic coarctation of the aorta. *Eur Heart J*. doi:10.1093/eurheartj/ehy142

Blue, G. M., Kirk, E. P., Giannoulatou, E., Dunwoodie, S. L., Ho, J. W., Hilton, D. C., . . . Winlaw, D. S. (2014). Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. *J Am Coll Cardiol, 64*(23), 2498-2506. doi:10.1016/j.jacc.2014.09.048

Bonachea, E. M., Zender, G., White, P., Corsmeier, D., Newsom, D., Fitzgerald-Butt, S., . . . McBride, K. L. (2014). Use of a targeted, combinatorial next-generation sequencing approach for the study of bicuspid aortic valve. *BMC Med Genomics, 7*, 56. doi:10.1186/1755-8794-7-56

Bonnard, C., Strobl, A. C., Shboul, M., Lee, H., Merriman, B., Nelson, S. F., . . . Reversade, B. (2012). Mutations in IRX5 impair craniofacial development and germ cell migration via SDF1. *Nat Genet, 44*(6), 709-713. doi:10.1038/ng.2259

Borlepawar, A., Rangrez, A. Y., Bernt, A., Christen, L., Sossalla, S., Frank, D., & Frey, N. (2017). TRIM24 protein promotes and TRIM32 protein inhibits cardiomyocyte hypertrophy via regulation of dysbindin protein levels. *J Biol Chem, 292*(24), 10180-10196. doi:10.1074/jbc.M116.752543

Boskovski, M. T., Yuan, S., Pedersen, N. B., Goth, C. K., Makova, S., Clausen, H., . . . Khokha, M. K. (2013). The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality. *Nature, 504*(7480), 456-459. doi:10.1038/nature12723

Botto, L. D., Flood, T., Little, J., Fluchel, M. N., Krikov, S., Feldkamp, M. L., . . . Romitti, P. A. (2013). Cancer risk in children and adolescents with birth defects: a population-based cohort study. *PLoS One, 8*(7), e69077. doi:10.1371/journal.pone.0069077

Boulter, C., Mulroy, S., Webb, S., Fleming, S., Brindle, K., & Sandford, R. (2001). Cardiovascular, skeletal, and renal defects in mice with a targeted disruption of the Pkd1 gene. *Proc Natl Acad Sci U S A, 98*(21), 12174-12179. doi:10.1073/pnas.211191098

Bouma, B. J., & Mulder, B. J. (2017). Changing Landscape of Congenital Heart Disease. *Circ Res, 120*(6), 908-922. doi:10.1161/circresaha.116.309302

Brewer, S., Feng, W., Huang, J., Sullivan, S., & Williams, T. (2004). Wnt1-Cre-mediated deletion of AP-2alpha causes multiple neural crest-related defects. *Dev Biol, 267*(1), 135-152. doi:10.1016/j.ydbio.2003.10.039

Brewer, S., Jiang, X., Donaldson, S., Williams, T., & Sucov, H. M. (2002). Requirement for AP-2alpha in cardiac outflow tract morphogenesis. *Mech Dev, 110*(1-2), 139-149. doi:10.1016/s0925-4773(01)00579-2

Brockington, M., Blake, D. J., Prandini, P., Brown, S. C., Torelli, S., Benson, M. A., . . . Muntoni, F. (2001). Mutations in the fukutin-related protein gene (FKRP) cause a form of congenital muscular dystrophy with secondary laminin alpha2 deficiency and abnormal glycosylation of alpha-dystroglycan. *Am J Hum Genet, 69*(6), 1198-1209. doi:10.1086/324412

Brodwall, K., Greve, G., Leirgul, E., Tell, G. S., Vollset, S. E., & Oyen, N. (2017). Recurrence of congenital heart defects among siblings-a nationwide study. *Am J Med Genet A, 173*(6), 1575-1585. doi:10.1002/ajmg.a.38237

Bruneau, B. G. (2013). Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb Perspect Biol, 5*(3), a008292. doi:10.1101/cshperspect.a008292

Burn, J., Brennan, P., Little, J., Holloway, S., Coffey, R., Somerville, J., . . . Hunter, A. S. (1998). Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study. *Lancet, 351*(9099), 311-316. doi:10.1016/s0140-6736(97)06486-6

Burnicka-Turek, O., Steimle, J. D., Huang, W., Felker, L., Kamp, A., Kweon, J., . . . Moskowitz, I. P. (2016). Cilia gene mutations cause atrioventricular septal defects by multiple mechanisms. *Hum Mol Genet*. doi:10.1093/hmg/ddw155

Calkoen, E. E., Hazekamp, M. G., Blom, N. A., Elders, B. B. L. J., Gittenberger-de Groot, A. C., Haak, M. C., . . . Jongbloed, M. R. M. (2016). Atrioventricular septal defect: From embryonic development to long-term follow-up. *Int J Cardiol, 202*, 784-795. doi:10.1016/j.ijcard.2015.09.081

Camenisch, T. D., Spicer, A. P., Brehm-Gibson, T., Biesterfeldt, J., Augustine, M. L., Calabro, A., Jr., . . . McDonald, J. A. (2000). Disruption of hyaluronan synthase-2 abrogates normal cardiac morphogenesis and hyaluronan-mediated transformation of epithelium to mesenchyme. *J Clin Invest, 106*(3), 349-360. doi:10.1172/JCI10272

Campione, M., & Franco, D. (2016). Current Perspectives in Cardiac Laterality. *J Cardiovasc Dev Dis, 3*(4). doi:10.3390/jcdd3040034

Cantu, C., Felker, A., Zimmerli, D., Prummel, K. D., Cabello, E. M., Chiavacci, E., . . . Mosimann, C. (2018). Mutations in Bcl9 and Pygo genes cause congenital heart defects by tissue-specific perturbation of Wnt/beta-catenin signaling. *Genes Dev, 32*(21-22), 1443-1458. doi:10.1101/gad.315531.118

Cardona, M., Lopez, J. A., Serafin, A., Rongvaux, A., Inserte, J., Garcia-Dorado, D., . . . Sanchis, D. (2015). Executioner Caspase-3 and 7 Deficiency Reduces Myocyte Number in the Developing Mouse Heart. *PLoS One, 10*(6), e0131411. doi:10.1371/journal.pone.0131411

Chaix, M. A., Andelfinger, G., & Khairy, P. (2016). Genetic testing in congenital heart disease: A clinical approach. *World J Cardiol, 8*(2), 180-191. doi:10.4330/wjc.v8.i2.180

Chandler, R. L., Brennan, J., Schisler, J. C., Serber, D., Patterson, C., & Magnuson, T. (2013). ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Mol Cell Biol, 33*(2), 265-280. doi:10.1128/MCB.01008-12

Chandler, R. L., & Magnuson, T. (2016). The SWI/SNF BAF-A complex is essential for neural crest development. *Dev Biol, 411*(1), 15-24. doi:10.1016/j.ydbio.2016.01.015

Chapnik, E., Sasson, V., Blelloch, R., & Hornstein, E. (2012). Dgcr8 controls neural crest cells survival in cardiovascular development. *Dev Biol, 362*(1), 50-56. doi:10.1016/j.ydbio.2011.11.008

Charo, D. N., Ho, M., Fajardo, G., Kawana, M., Kundu, R. K., Sheikh, A. Y., . . . Quertermous, T. (2009). Endogenous regulation of cardiovascular function by apelin-APJ. *Am J Physiol Heart Circ Physiol, 297*(5), H1904-1913. doi:10.1152/ajpheart.00686.2009

Chen, H., Shi, S., Acosta, L., Li, W., Lu, J., Bao, S., . . . Shou, W. (2004). BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development, 131*(9), 2219-2231. doi:10.1242/dev.01094

Chen, J., Bardes, E. E., Aronow, B. J., & Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res, 37*(Web Server issue), W305-311. doi:10.1093/nar/gkp427

Chen, L., Fulcoli, F. G., Ferrentino, R., Martucciello, S., Illingworth, E. A., & Baldini, A. (2012). Transcriptional control in cardiac progenitors: Tbx1 interacts with the BAF chromatin remodeling complex and regulates Wnt5a. *PLoS Genet, 8*(3), e1002571. doi:10.1371/journal.pgen.1002571

Chiplunkar, A. R., Lung, T. K., Alhashem, Y., Koppenhaver, B. A., Salloum, F. N., Kukreja, R. C., . . . Lloyd, J. A. (2013). Kruppel-like factor 2 is required for normal mouse cardiac development. *PLoS One, 8*(2), e54891. doi:10.1371/journal.pone.0054891

Cinquetti, R., Badi, I., Campione, M., Bortoletto, E., Chiesa, G., Parolini, C., . . . Acquati, F. (2008). Transcriptional deregulation and a missense mutation define ANKRD1 as a candidate gene for total anomalous pulmonary venous return. *Hum Mutat, 29*(4), 468-474. doi:10.1002/humu.20711

Clement, C. A., Kristensen, S. G., Mollgard, K., Pazour, G. J., Yoder, B. K., Larsen, L. A., & Christensen, S. T. (2009). The primary cilium coordinates early cardiogenesis and hedgehog signaling in cardiomyocyte differentiation. *J Cell Sci, 122*(Pt 17), 3070-3082. doi:10.1242/jcs.049676

Cohen, S., Liu, A., Gurvitz, M., Guo, L., Therrien, J., Laprise, C., . . . Marelli, A. J. (2018). Exposure to Low-Dose Ionizing Radiation From Cardiac Procedures and Malignancy Risk in Adults With Congenital Heart Disease. *Circulation, 137*(13), 1334-1345. doi:10.1161/circulationaha.117.029138

Connor, J. A., & Thiagarajan, R. (2007). Hypoplastic left heart syndrome. *Orphanet J Rare Dis, 2*, 23. doi:10.1186/1750-1172-2-23

Costain, G., Lionel, A. C., Ogura, L., Marshall, C. R., Scherer, S. W., Silversides, C. K., & Bassett, A. S. (2016). Genome-wide rare copy number variations contribute to genetic risk for transposition of the great arteries. *Int J Cardiol, 204*, 115-121. doi:10.1016/j.ijcard.2015.11.127

Cota, C. D., & Garcia-Garcia, M. J. (2012). The ENU-induced cetus mutation reveals an essential role of the DNA helicase DDX11 for mesoderm development during early mouse embryogenesis. *Dev Dyn, 241*(8), 1249-1259. doi:10.1002/dvdy.23810

Cowan, J., Tariq, M., & Ware, S. M. (2014). Genetic and functional analyses of ZIC3 variants in congenital heart disease. *Hum Mutat, 35*(1), 66-75. doi:10.1002/humu.22457

Cristo, F., Inacio, J. M., de Almeida, S., Mendes, P., Martins, D. S., Maio, J., . . . Belo, J. A. (2017). Functional study of DAND5 variant in patients with Congenital Heart Disease and laterality defects. *BMC Med Genet, 18*(1), 77. doi:10.1186/s12881-017-0444-1

D'Alessandro, L. C., Al Turki, S., Manickaraj, A. K., Manase, D., Mulder, B. J., Bergin, L., . . . Mital, S. (2016). Exome sequencing identifies rare variants in multiple genes in atrioventricular septal defect. *Genet Med, 18*(2), 189-198. doi:10.1038/gim.2015.60

Dawson, S., Charles, A. K., Bower, C., de Klerk, N. H., & Milne, E. (2015). Risk of cancer among children with birth defects: a novel approach. *Birth Defects Res A Clin Mol Teratol, 103*(4), 284-291. doi:10.1002/bdra.23364

De Luca, A., Sarkozy, A., Consoli, F., Ferese, R., Guida, V., Dentici, M. L., . . . Dallapiccola, B. (2010). Familial transposition of the great arteries caused by multiple mutations in laterality genes. *Heart, 96*(9), 673-677. doi:10.1136/hrt.2009.181685

Del Viso, F., Huang, F., Myers, J., Chalfant, M., Zhang, Y., Reza, N., . . . Khokha, M. K. (2016). Congenital Heart Disease Genetics Uncovers Context-Dependent Organization and Function of Nucleoporins at Cilia. *Dev Cell, 38*(5), 478-492. doi:10.1016/j.devcel.2016.08.002

Deng, H., Xia, H., & Deng, S. (2015). Genetic basis of human left-right asymmetry disorders. *Expert Rev Mol Med, 16*, e19. doi:10.1017/erm.2014.22

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet, 43*(5), 491-498. doi:10.1038/ng.806

Dickinson, M. E., Flenniken, A. M., Ji, X., Teboul, L., Wong, M. D., White, J. K., . . . Murray, S. A. (2016). High-throughput discovery of novel developmental phenotypes. *Nature, 537*(7621), 508-514. doi:10.1038/nature19356

Digilio, M. C., Marino, B., Ammirati, A., Borzaga, U., Giannotti, A., & Dallapiccola, B. (1999). Cardiac malformations in patients with oral-facial-skeletal syndromes: clinical similarities with heterotaxia. *Am J Med Genet, 84*(4), 350-356. doi:10.1002/(sici)1096-8628(19990604)84:4<350::aid-ajmg8>3.0.co;2-e

Dodge-Khatami, A., Mavroudis, C. D., Frost, J., Jacobs, J. P., & Mavroudis, C. (2014). Repairing the tricuspid valve in congenital heart diseases other than Ebstein's. *Cardiol Young, 24*(6), 1077-1087. doi:10.1017/S104795111400198X

Dubielecka, P. M., Ladwein, K. I., Xiong, X., Migeotte, I., Chorzalska, A., Anderson, K. V., . . . Kotula, L. (2011). Essential role for Abi1 in embryonic survival and WAVE2 complex integrity. *Proc Natl Acad Sci U S A, 108*(17), 7022-7027. doi:10.1073/pnas.1016811108

Dumont, D. J., Jussila, L., Taipale, J., Lymboussaki, A., Mustonen, T., Pajusola, K., . . . Alitalo, K. (1998). Cardiovascular failure in mouse embryos deficient in VEGF receptor-3. *Science, 282*(5390), 946-949. doi:10.1126/science.282.5390.946

Durkin, M. E., Avner, M. R., Huh, C. G., Yuan, B. Z., Thorgeirsson, S. S., & Popescu, N. C. (2005). DLC-1, a Rho GTPase-activating protein with tumor suppressor function, is essential for embryonic development. *FEBS Lett, 579*(5), 1191-1196. doi:10.1016/j.febslet.2004.12.090

Durmaz, A., Durmaz, B., Kadioglu, B., Aksoylar, S., Karapinar, D., Koturoglu, G., . . . Cogulu, O. (2011). The Association of minor congenital anomalies and childhood cancer. *Pediatr Blood Cancer, 56*(7), 1098-1102. doi:10.1002/pbc.23049

Durst, R., Sauls, K., Peal, D. S., deVlaming, A., Toomer, K., Leyne, M., . . . Slaugenhaupt, S. A. (2015). Mutations in DCHS1 cause mitral valve prolapse. *Nature, 525*(7567), 109-113. doi:10.1038/nature14670

Dykes, I. M., van Bueren, K. L., Ashmore, R. J., Floss, T., Wurst, W., Szumska, D., . . . Scambler, P. J. (2014). HIC2 is a novel dosage-dependent regulator of cardiac development located within the distal 22q11 deletion syndrome region. *Circ Res, 115*(1), 23-31. doi:10.1161/CIRCRESAHA.115.303300

Egbe, A., Uppu, S., Lee, S., Ho, D., & Srivastava, S. (2014). Changing prevalence of severe congenital heart disease: a population-based study. *Pediatr Cardiol, 35*(7), 1232-1238. doi:10.1007/s00246-014-0921-7

Egbe, A. C., Poterucha, J. T., Warnes, C. A., Connolly, H. M., Baskar, S., Ginde, S., . . . Veldtman, G. R. (2018). Hepatocellular Carcinoma After Fontan Operation. *Circulation, 138*(7), 746-748. doi:10.1161/circulationaha.117.032717

Egorova, A. D., Khedoe, P. P., Goumans, M. J., Yoder, B. K., Nauli, S. M., ten Dijke, P., . . . Hierck, B. P. (2011). Lack of primary cilia primes shear-induced endothelial-to-mesenchymal transition. *Circ Res, 108*(9), 1093-1101. doi:10.1161/circresaha.110.231860

Eicher, J. D., Chen, M. H., Pitsillides, A. N., Lin, H., Veeraraghavan, N., Brody, J. A., . . . Johnson, A. D. (2017). Whole exome sequencing in the Framingham Heart Study identifies rare variation in HYAL2 that influences platelet aggregation. *Thromb Haemost, 117*(6), 1083-1092. doi:10.1160/TH16-09-0677

Ellesoe, S. G., Workman, C. T., Bouvagnet, P., Loffredo, C. A., McBride, K. L., Hinton, R. B., . . . Larsen, L. A. (2018). Familial co-occurrence of congenital heart defects follows distinct patterns. *Eur Heart J, 39*(12), 1015-1022. doi:10.1093/eurheartj/ehx314

England, J., Granados-Riveron, J., Polo-Parada, L., Kuriakose, D., Moore, C., Brook, J. D., . . . Loughna, S. (2017). Tropomyosin 1: Multiple roles in the developing heart and in the formation of congenital heart defects. *J Mol Cell Cardiol, 106*, 1-13. doi:10.1016/j.yjmcc.2017.03.006

Eppig, J. T., Smith, C. L., Blake, J. A., Ringwald, M., Kadin, J. A., Richardson, J. E., & Bult, C. J. (2017). Mouse Genome Informatics (MGI): Resources for Mining Mouse Genetic, Genomic, and Biological Data in Support of Primary and Translational Research. *Methods Mol Biol, 1488*, 47-73. doi:10.1007/978-1-4939-6427-7_3

Erikssen, G., Liestol, K., Seem, E., Birkeland, S., Saatvedt, K. J., Hoel, T. N., . . . Lindberg, H. L. (2015). Achievements in congenital heart defect surgery: a prospective, 40-year study of 7038 patients. *Circulation, 131*(4), 337-346; discussion 346. doi:10.1161/CIRCULATIONAHA.114.012033

Fahed, A. C., Gelb, B. D., Seidman, J. G., & Seidman, C. E. (2013). Genetics of congenital heart disease: the glass half empty. *Circ Res, 112*(4), 707-720. doi:10.1161/circresaha.112.300853

Fakhro, K. A., Choi, M., Ware, S. M., Belmont, J. W., Towbin, J. A., Lifton, R. P., . . . Brueckner, M. (2011). Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci U S A, 108*(7), 2915-2920. doi:10.1073/pnas.1019645108

Field, S., Riley, K. L., Grimes, D. T., Hilton, H., Simon, M., Powles-Glover, N., . . . Norris, D. P. (2011). Pkd1l1 establishes left-right asymmetry and physically interacts with Pkd2. *Development, 138*(6), 1131-1142. doi:10.1242/dev.058149

Finsterer, J., Stollberger, C., & Towbin, J. A. (2017). Left ventricular noncompaction cardiomyopathy: cardiac, neuromuscular, and genetic factors. *Nat Rev Cardiol, 14*(4), 224-237. doi:10.1038/nrcardio.2016.207

Firulli, A. B., McFadden, D. G., Lin, Q., Srivastava, D., & Olson, E. N. (1998). Heart and extra-embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nat Genet, 18*(3), 266-270. doi:10.1038/ng0398-266

Fishbein, G. A., & Fishbein, M. C. (2019). Mitral Valve Pathology. *Current cardiology reports, 21*(7), 61-61. doi:10.1007/s11886-019-1145-5

Fisher, P. G., Reynolds, P., Von Behren, J., Carmichael, S. L., Rasmussen, S. A., & Shaw, G. M. (2012). Cancer in children with nonchromosomal birth defects. *J Pediatr, 160*(6), 978-983. doi:10.1016/j.jpeds.2011.12.006

Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., . . . Campbell, P. J. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res, 43*(Database issue), D805-811. doi:10.1093/nar/gku1075

Francis, R. J., Christopher, A., Devine, W. A., Ostrowski, L., & Lo, C. (2012). Congenital heart disease and the specification of left-right asymmetry. *Am J Physiol Heart Circ Physiol, 302*(10), H2102-2111. doi:10.1152/ajpheart.01118.2011

Freyer, L., Hsu, C. W., Nowotschin, S., Pauli, A., Ishida, J., Kuba, K., . . . Hadjantonakis, A. K. (2017). Loss of Apela Peptide in Mice Causes Low Penetrance Embryonic Lethality and Defects in Early Mesodermal Derivatives. *Cell Rep, 20*(9), 2116-2130. doi:10.1016/j.celrep.2017.08.014

Fulmer, D., Toomer, K., Guo, L., Moore, K., Glover, J., Moore, R., . . . Lipschutz, J. H. (2019). Defects in the Exocyst-Cilia Machinery Cause Bicuspid Aortic Valve Disease and Aortic Stenosis. *Circulation*. doi:10.1161/CIRCULATIONAHA.119.038376

Gaengel, K., Niaudet, C., Hagikura, K., Lavina, B., Muhl, L., Hofmann, J. J., . . . Betsholtz, C. (2012). The sphingosine-1-phosphate receptor S1PR1 restricts sprouting angiogenesis by regulating the interplay between VE-cadherin and VEGFR2. *Dev Cell, 23*(3), 587-599. doi:10.1016/j.devcel.2012.08.005

Gao, W., Higaki, T., Eguchi-Ishimae, M., Iwabuki, H., Wu, Z., Yamamoto, E., . . . Eguchi, M. (2015). DGCR6 at the proximal part of the DiGeorge critical region is involved in conotruncal heart defects. *Hum Genome Var, 2*, 15004. doi:10.1038/hgv.2015.4

Gerhardt, C., Lier, J. M., Kuschel, S., & Ruther, U. (2013). The ciliary protein Ftm is required for ventricular wall and septal development. *PLoS One, 8*(2), e57545. doi:10.1371/journal.pone.0057545

Geva, T., Martins, J. D., & Wald, R. M. (2014). Atrial septal defects. *Lancet, 383*(9932), 1921-1932. doi:10.1016/S0140-6736(13)62145-5

Gibbs, B. C., Damerla, R. R., Vladar, E. K., Chatterjee, B., Wan, Y., Liu, X., . . . Lo, C. W. (2016). Prickle1 mutation causes planar cell polarity and directional cell migration defects associated with cardiac outflow tract anomalies and other structural birth defects. *Biol Open, 5*(3), 323-335. doi:10.1242/bio.015750

Gifford, C. A., Ranade, S. S., Samarakoon, R., Salunga, H. T., de Soysa, T. Y., Huang, Y., . . . Srivastava, D. (2019). Oligogenic inheritance of a human heart disease involving a genetic modifier. *Science, 364*(6443), 865-870. doi:10.1126/science.aat5056

Gill, H. K., Splitt, M., Sharland, G. K., & Simpson, J. M. (2003). Patterns of recurrence of congenital heart disease: an analysis of 6,640 consecutive pregnancies evaluated by

detailed fetal echocardiography. *J Am Coll Cardiol, 42*(5), 923-929. Retrieved from https://www.sciencedirect.com/science/article/pii/S0735109703008532?via%3Dihub

Girdauskas, E., Geist, L., Disha, K., Kazakbaev, I., Gross, T., Schulz, S., . . . Kurth, I. (2017). Genetic abnormalities in bicuspid aortic valve root phenotype: preliminary results. *Eur J Cardiothorac Surg, 52*(1), 156-162. doi:10.1093/ejcts/ezx065

Gittenberger-de Groot, A. C., Bartelings, M. M., Deruiter, M. C., & Poelmann, R. E. (2005). Basics of cardiac development for the understanding of congenital heart malformations. *Pediatr Res, 57*(2), 169-176. doi:10.1203/01.Pdr.0000148710.69159.61

Goddeeris, M. M., Rho, S., Petiet, A., Davenport, C. L., Johnson, G. A., Meyers, E. N., & Klingensmith, J. (2008). Intracardiac septation requires hedgehog-dependent cellular contributions from outside the heart. *Development, 135*(10), 1887-1895. doi:10.1242/dev.016147

Goddeeris, M. M., Schwartz, R., Klingensmith, J., & Meyers, E. N. (2007). Independent requirements for Hedgehog signaling by both the anterior heart field and neural crest cells for outflow tract development. *Development, 134*(8), 1593-1604. doi:10.1242/dev.02824

Granadillo, J. L., Chung, W. K., Hecht, L., Corsten-Janssen, N., Wegner, D., Nij Bijvank, S. W. A., . . . Shinawi, M. (2018). Variable cardiovascular phenotypes associated with SMAD2 pathogenic variants. *Hum Mutat, 39*(12), 1875-1884. doi:10.1002/humu.23627

Grossman, T. R., Gamliel, A., Wessells, R. J., Taghli-Lamallem, O., Jepsen, K., Ocorr, K., . . . Bier, E. (2011). Over-expression of DSCAM and COL6A2 cooperatively generates congenital heart defects. *PLoS Genet, 7*(11), e1002344. doi:10.1371/journal.pgen.1002344

Grossmann, K. S., Grund, C., Huelsken, J., Behrend, M., Erdmann, B., Franke, W. W., & Birchmeier, W. (2004). Requirement of plakophilin 2 for heart morphogenesis and cardiac junction formation. *J Cell Biol, 167*(1), 149-160. doi:10.1083/jcb.200402096

Grunert, M., Dorn, C., Schueler, M., Dunkel, I., Schlesinger, J., Mebus, S., . . . Sperling, S. R. (2014). Rare and private variations in neural crest, apoptosis and sarcomere genes define the polygenic background of isolated Tetralogy of Fallot. *Hum Mol Genet, 23*(12), 3115-3128. doi:10.1093/hmg/ddu021

Guo, C., Sun, Y., Zhou, B., Adam, R. M., Li, X., Pu, W. T., . . . Li, X. (2011). A Tbx1-Six1/Eya1-Fgf8 genetic pathway controls mammalian cardiovascular and craniofacial morphogenesis. *J Clin Invest, 121*(4), 1585-1595. doi:10.1172/JCI44630

Guo, C., Wang, Q., Wang, Y., Yang, L., Luo, H., Cao, X. F., . . . Lu, C. (2017). Exome sequencing reveals novel IRXI mutation in congenital heart disease. *Mol Med Rep, 15*(5), 3193-3197. doi:10.3892/mmr.2017.6410

Guo, T., McDonald-McGinn, D., Blonska, A., Shanske, A., Bassett, A. S., Chow, E., . . . International Chromosome 22q, C. (2011). Genotype and cardiovascular phenotype correlations with TBX1 in 1,022 velo-cardio-facial/DiGeorge/22q11.2 deletion syndrome patients. *Hum Mutat, 32*(11), 1278-1289. doi:10.1002/humu.21568

Guo, T., Repetto, G. M., McDonald McGinn, D. M., Chung, J. H., Nomaru, H., Campbell, C. L., . . . Morrow, B. E. (2017). Genome-Wide Association Study to Find Modifiers for Tetralogy of Fallot in the 22q11.2 Deletion Syndrome Identifies Variants in the GPR98 Locus on 5q14.3. *Circ Cardiovasc Genet, 10*(5). doi:10.1161/circgenetics.116.001690

Guris, D. L., Duester, G., Papaioannou, V. E., & Imamoto, A. (2006). Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. *Dev Cell, 10*(1), 81-92. doi:10.1016/j.devcel.2005.12.002

Gurvitz, M., Ionescu-Ittu, R., Guo, L., Eisenberg, M. J., Abrahamowicz, M., Pilote, L., & Marelli, A. J. (2016). Prevalence of Cancer in Adults With Congenital Heart Disease Compared With the General Population. *Am J Cardiol, 118*(11), 1742-1750. doi:10.1016/j.amjcard.2016.08.057

Hanneman, K., Newman, B., & Chan, F. (2017). Congenital Variants and Anomalies of the Aortic Arch. *Radiographics : a review publication of the Radiological Society of North America, Inc, 37*(1), 32-51. doi:10.1148/rg.2017160033

Harrison, M. J., Shapiro, A. J., & Kennedy, M. P. (2016). Congenital Heart Disease and Primary Ciliary Dyskinesia. *Paediatr Respir Rev, 18*, 25-32. doi:10.1016/j.prrv.2015.09.003

Hartill, V. L., van de Hoek, G., Patel, M. P., Little, R., Watson, C. M., Berry, I. R., . . . Johnson, C. A. (2018). DNAAF1 links heart laterality with the AAA+ ATPase RUVBL1 and ciliary intraflagellar transport. *Hum Mol Genet, 27*(3), 529-545. doi:10.1093/hmg/ddx422

Hassed, S., Li, S., Mulvihill, J., Aston, C., & Palmer, S. (2017). Adams-Oliver syndrome review of the literature: Refining the diagnostic phenotype. *Am J Med Genet A, 173*(3), 790-800. doi:10.1002/ajmg.a.37889

Hayano, S., Okuno, Y., Tsutsumi, M., Inagaki, H., Fukasawa, Y., Kurahashi, H., . . . Kato, T. (2019). Frequent intragenic microdeletions of elastin in familial supravalvular aortic stenosis. *Int J Cardiol, 274*, 290-295. doi:10.1016/j.ijcard.2018.09.032

Hecker, P. A., Leopold, J. A., Gupte, S. A., Recchia, F. A., & Stanley, W. C. (2013). Impact of glucose-6-phosphate dehydrogenase deficiency on the pathophysiology of cardiovascular disease. *Am J Physiol Heart Circ Physiol, 304*(4), H491-500. doi:10.1152/ajpheart.00721.2012

Helle, E., Cordova-Palomera, A., Ojala, T., Saha, P., Potiny, P., Gustafsson, S., . . . Priest, J. R. (2018). Loss of function, missense, and intronic variants in NOTCH1 confer different risks for left ventricular outflow tract obstructive heart defects in two European cohorts. *Genet Epidemiol*. doi:10.1002/gepi.22176

Hierck, B. P., Molin, D. G., Boot, M. J., Poelmann, R. E., & Gittenberger-de Groot, A. C. (2004). A chicken model for DGCR6 as a modifier gene in the DiGeorge critical region. *Pediatr Res, 56*(3), 440-448. doi:10.1203/01.PDR.0000136151.50127.1C

Higgins, E. M., Bos, J. M., Mason-Suares, H., Tester, D. J., Ackerman, J. P., MacRae, C. A., . . . Ackerman, M. J. (2017). Elucidation of MRAS-mediated Noonan syndrome with cardiac hypertrophy. *JCI Insight, 2*(5), e91225. doi:10.1172/jci.insight.91225

Hilger, A. C., Halbritter, J., Pennimpede, T., van der Ven, A., Sarma, G., Braun, D. A., . . . Hildebrandt, F. (2015). Targeted Resequencing of 29 Candidate Genes and Mouse Expression Studies Implicate ZIC3 and FOXF1 in Human VATER/VACTERL Association. *Hum Mutat, 36*(12), 1150-1154. doi:10.1002/humu.22859

Honda, H., Oda, H., Nakamoto, T., Honda, Z., Sakai, R., Suzuki, T., . . . Hirai, H. (1998). Cardiovascular anomaly, impaired actin bundling and resistance to Src-induced transformation in mice lacking p130Cas. *Nat Genet, 19*(4), 361-365. doi:10.1038/1246

Hrstka, S. C. L., Li, X., Nelson, T. J., & Pipeline, W. P. G. (2017). NOTCH1-Dependent Nitric Oxide Signaling Deficiency in Hypoplastic Left Heart Syndrome Revealed Through Patient-Specific Phenotypes Detected in Bioengineered Cardiogenesis. *Stem Cells, 35*(4), 1106-1119. doi:10.1002/stem.2582

Hu, Z., Shi, Y., Mo, X., Xu, J., Zhao, B., Lin, Y., . . . Shen, H. (2013). A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat Genet, 45*(7), 818-821. doi:10.1038/ng.2636

Huang, R. T., Xue, S., Wang, J., Gu, J. Y., Xu, J. H., Li, Y. J., . . . Yang, Y. Q. (2016). CASZ1 loss-of-function mutation associated with congenital heart disease. *Gene, 595*(1), 62-68. doi:10.1016/j.gene.2016.09.044

Hunt, S. E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., . . . Cunningham, F. (2018). Ensembl variation resources. *Database (Oxford), 2018*. doi:10.1093/database/bay119

Hyde, A. S., Farmer, E. L., Easley, K. E., van Lammeren, K., Christoffels, V. M., Barycki, J. J., . . . Simpson, M. A. (2012). UDP-glucose dehydrogenase polymorphisms from patients with congenital heart valve defects disrupt enzyme stability and quaternary assembly. *J Biol Chem, 287*(39), 32708-32716. doi:10.1074/jbc.M112.395202

Iascone, M., Ciccone, R., Galletti, L., Marchetti, D., Seddio, F., Lincesso, A. R., . . . Zuffardi, O. (2012). Identification of de novo mutations and rare variants in hypoplastic left heart syndrome. *Clin Genet, 81*(6), 542-554. doi:10.1111/j.1399-0004.2011.01674.x

ICD-10-CM the Complete Official Codebook. (2015). Retrieved from https://icd.who.int/browse10/2019/en. (ISBN 10: 1622022122

ISBN 13: 9781622022120). Retrieved September 2, 2016, from American Medical Association https://icd.who.int/browse10/2019/en

Incecik, F., Herguner, O. M., Alinc Erdem, S., & Altunbasak, S. (2015). Neurofibromatosis type 1 and cardiac manifestations. *Turk Kardiyol Dern Ars, 43*(8), 714-716. doi:10.5543/tkda.2015.27557

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., & Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet, 92*(6), 841-853. doi:10.1016/j.ajhg.2013.04.015

Ishikawa, T. (2017). Axoneme Structure from Motile Cilia. *Cold Spring Harbor perspectives in biology, 9*(1), a028076. doi:10.1101/cshperspect.a028076

Iwamoto, R., & Mekada, E. (2006). ErbB and HB-EGF signaling in heart development and function. *Cell Struct Funct, 31*(1), 1-14. doi:10.1247/csf.31.1

Jaron, R., Rosenfeld, N., Zahdeh, F., Carmi, S., Beni-Adani, L., Doviner, V., . . . Levy-Lahad, E. (2016). Expanding the phenotype of CRB2 mutations - A new ciliopathy syndrome? *Clin Genet, 90*(6), 540-544. doi:10.1111/cge.12764

Jepsen, K., Gleiberman, A. S., Shi, C., Simon, D. I., & Rosenfeld, M. G. (2008). Cooperative regulation in development by SMRT and FOXP1. *Genes Dev, 22*(6), 740-745. doi:10.1101/gad.1637108

Jerse, M., & Zidar, N. (2011). Apoptosis in the developing human heart resembles apoptosis in epithelial tissues. *Cell Tissue Res, 343*(3), 537-543. doi:10.1007/s00441-010-1121-2

Ji, W., Benson, M. A., Bhattacharya, S., Chen, Y., Hu, J., & Li, F. (2014). Characterization of transcription factor AP-2 beta mutations involved in familial isolated patent ductus arteriosus suggests haploinsufficiency. *J Surg Res, 188*(2), 466-472. doi:10.1016/j.jss.2014.01.015

Jin, S. C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S. R., . . . Brueckner, M. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet, 49*(11), 1593-1601. doi:10.1038/ng.3970

Johnson, K. J., Lee, J. M., Ahsan, K., Padda, H., Feng, Q., Partap, S., . . . Druley, T. E. (2017). Pediatric cancer risk in association with birth defects: A systematic review. *PLoS One, 12*(7), e0181246. doi:10.1371/journal.pone.0181246

Johnson, T. R. (2010). Conotruncal cardiac defects: a clinical imaging perspective. *Pediatric Cardiology, 31*(3), 430-437. doi:10.1007/s00246-010-9668-y

Jun, G., Flickinger, M., Hetrick, K. N., Romm, J. M., Doheny, K. F., Abecasis, G. R., . . . Kang, H. M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet, 91*(5), 839-848. doi:10.1016/j.ajhg.2012.09.004

Kaizuka, T., & Mizushima, N. (2015). Atg13 Is Essential for Autophagy and Cardiac Development in Mice. *Mol Cell Biol, 36*(4), 585-595. doi:10.1128/MCB.01005-15

Kamura, K., Kobayashi, D., Uehara, Y., Koshida, S., Iijima, N., Kudo, A., . . . Takeda, H. (2011). Pkd1l1 complexes with Pkd2 on motile cilia and functions to establish the left-right axis. *Development, 138*(6), 1121-1129. doi:10.1242/dev.058271

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res, 47*(D1), D590-D595. doi:10.1093/nar/gky962

Kang, Y., Kim, J., Anderson, J. P., Wu, J., Gleim, S. R., Kundu, R. K., . . . Chun, H. J. (2013). Apelin-APJ signaling is a critical regulator of endothelial MEF2 activation in cardiovascular development. *Circ Res, 113*(1), 22-31. doi:10.1161/CIRCRESAHA.113.301324

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., . . . MacArthur, D. G. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-offunction intolerance across human protein-coding genes. *bioRxiv*. doi:10.1101/531210

Kaur, S., McGlashan, S. R., & Ward, M. L. (2018). Evidence of primary cilia in the developing rat heart. *Cilia, 7*, 4. doi:10.1186/s13630-018-0058-z

Kawasaki, Y., Jigami, T., Furukawa, S., Sagara, M., Echizen, K., Shibata, Y., . . . Akiyama, T. (2010). The adenomatous polyposis coli-associated guanine nucleotide exchange factor Asef is involved in angiogenesis. *J Biol Chem, 285*(2), 1199-1207. doi:10.1074/jbc.M109.040691

Kelle, A. M., Bentley, S. J., Rohena, L. O., Cabalka, A. K., & Olson, T. M. (2016). Ebstein anomaly, left ventricular non-compaction, and early onset heart failure associated with a de novo alpha-tropomyosin gene mutation. *Am J Med Genet A, 170*(8), 2186-2190. doi:10.1002/ajmg.a.37745

Kennedy, M. P., Omran, H., Leigh, M. W., Dell, S., Morgan, L., Molina, P. L., . . . Knowles, M. R. (2007). Congenital heart disease and other heterotaxic defects in a large cohort of patients with primary ciliary dyskinesia. *Circulation, 115*(22), 2814-2821. doi:10.1161/circulationaha.106.649038

Kim, K. H., Rosen, A., Bruneau, B. G., Hui, C. C., & Backx, P. H. (2012). Iroquois homeodomain transcription factors in heart development and function. *Circ Res, 110*(11), 1513-1524. doi:10.1161/CIRCRESAHA.112.265041

Kim, W., Essalmani, R., Szumska, D., Creemers, J. W., Roebroek, A. J., D'Orleans-Juste, P., . . . Prat, A. (2012). Loss of endothelial furin leads to cardiac malformation and early postnatal death. *Mol Cell Biol, 32*(17), 3382-3391. doi:10.1128/MCB.06331-11

Kinzel, D., Boldt, K., Davis, E. E., Burtscher, I., Trumbach, D., Diplas, B., . . . Lickert, H. (2010). Pitchfork regulates primary cilia disassembly and left-right asymmetry. *Dev Cell, 19*(1), 66-77. doi:10.1016/j.devcel.2010.06.005

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet, 46*(3), 310-315. doi:10.1038/ng.2892

Klena, N. T., Gibbs, B. C., & Lo, C. W. (2017). Cilia and Ciliopathies in Congenital Heart Disease. *Cold Spring Harb Perspect Biol*. doi:10.1101/cshperspect.a028266

Kodo, K., Nishizawa, T., Furutani, M., Arai, S., Ishihara, K., Oda, M., . . . Yamagishi, H. (2012). Genetic analysis of essential cardiac transcription factors in 256 patients with non-syndromic congenital heart defects. *Circ J, 76*(7), 1703-1711. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22498567 https://www.jstage.jst.go.jp/article/circj/76/7/76_CJ-11-1389/_pdf

Kong, B., Liu, Y. L., & Lu, X. D. (2009). Decreased expression of neurotrophic tyrosine receptor kinase 3 is associated with the outflow tract defect of human tetralogy of Fallot. *Chin Med J (Engl), 122*(2), 153-157.

Kramer, A., Green, J., Pollard, J., Jr., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics, 30*(4), 523-530. doi:10.1093/bioinformatics/btt703

Kratz, C. P., Franke, L., Peters, H., Kohlschmidt, N., Kazmierczak, B., Finckh, U., . . . Zenker, M. (2015). Cancer spectrum and frequency among children with Noonan, Costello, and cardio-facio-cutaneous syndromes. *Br J Cancer, 112*(8), 1392-1397. doi:10.1038/bjc.2015.75

Kratz, C. P., Rapisuwon, S., Reed, H., Hasle, H., & Rosenberg, P. S. (2011). Cancer in Noonan, Costello, cardiofaciocutaneous and LEOPARD syndromes. *Am J Med Genet C Semin Med Genet, 157c*(2), 83-89. doi:10.1002/ajmg.c.30300

Krishnan, A., Samtani, R., Dhanantwari, P., Lee, E., Yamada, S., Shiota, K., . . . Lo, C. W. (2014). A detailed comparison of mouse and human cardiac development. *Pediatr Res, 76*(6), 500-507. doi:10.1038/pr.2014.128

Kuroda, K., Han, H., Tani, S., Tanigaki, K., Tun, T., Furukawa, T., . . . Honjo, T. (2003). Regulation of Marginal Zone B Cell Development by MINT, a Suppressor of Notch/RBP-J Signaling Pathway. *Immunity, 18*(2), 301-312. doi:10.1016/s1074-7613(03)00029-3

Kwiatkowski, D. M., Hanley, F. L., & Krawczeski, C. D. (2016). Right Ventricular Outflow Tract Obstruction: Pulmonary Atresia With Intact Ventricular Septum, Pulmonary Stenosis, and Ebstein's Malformation. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 17*(8 Suppl 1), S323-S329. doi:10.1097/PCC.0000000000000818

LaHaye, S., Lincoln, J., & Garg, V. (2014). Genetics of valvular heart disease. *Curr Cardiol Rep, 16*(6), 487. doi:10.1007/s11886-014-0487-2

Lai, C. K., Gupta, N., Wen, X., Rangell, L., Chih, B., Peterson, A. S., . . . Scales, S. J. (2011). Functional characterization of putative cilia genes by high-content analysis. *Mol Biol Cell, 22*(7), 1104-1119. doi:10.1091/mbc.E10-07-0596

Lai, L., Leone, T. C., Zechner, C., Schaeffer, P. J., Kelly, S. M., Flanagan, D. P., . . . Kelly, D. P. (2008). Transcriptional coactivators PGC-1alpha and PGC-lbeta control overlapping programs required for perinatal maturation of the heart. *Genes Dev, 22*(14), 1948-1961. doi:10.1101/gad.1661708

Lalani, S. R., & Belmont, J. W. (2014). Genetic basis of congenital cardiovascular malformations. *Eur J Med Genet, 57*(8), 402-413. doi:10.1016/j.ejmg.2014.04.010

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., . . . Maglott, D. R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res, 44*(D1), D862-868. doi:10.1093/nar/gkv1222

Larkins, C. E., Long, A. B., & Caspary, T. (2012). Defective Nodal and Cerl2 expression in the Arl13b(hnn) mutant node underlie its heterotaxia. *Dev Biol, 367*(1), 15-24. doi:10.1016/j.ydbio.2012.04.011

Lebo, M. S., & Baxter, S. M. (2014). New molecular genetic tests in the diagnosis of heart disease. *Clin Lab Med, 34*(1), 137-156, vii-viii. doi:10.1016/j.cll.2013.11.011

Lee, K. F., Simon, H., Chen, H., Bates, B., Hung, M. C., & Hauser, C. (1995). Requirement for neuregulin receptor erbB2 in neural and cardiac development. *Nature, 378*(6555), 394-398. doi:10.1038/378394a0

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285-291. doi:10.1038/nature19057

Leung, C., Liu, Y., Lu, X., Kim, M., Drysdale, T. A., & Feng, Q. (2015). Rac1 Signaling Is Required for Anterior Second Heart Field Cellular Organization and Cardiac Outflow Tract Development. *J Am Heart Assoc, 5*(1). doi:10.1161/JAHA.115.002508

Li, A. H., Hanchard, N. A., Azamian, M., D'Alessandro, L. C. A., Coban-Akdemir, Z., Lopez, K. N., . . . Belmont, J. W. (2019). Genetic architecture of laterality defects revealed by whole exome sequencing. *Eur J Hum Genet*. doi:10.1038/s41431-018-0307-z

Li, A. H., Hanchard, N. A., Furthner, D., Fernbach, S., Azamian, M., Nicosia, A., . . . Belmont, J. W. (2017). Whole exome sequencing in 342 congenital cardiac left sided lesion cases reveals extensive genetic heterogeneity and complex inheritance patterns. *Genome Med, 9*(1), 95. doi:10.1186/s13073-017-0482-5

Li, B., Yu, L., Liu, D., Yang, X., Zheng, Y., Gui, Y., & Wang, H. (2018). MIB1 mutations reduce Notch signaling activation and contribute to congenital heart disease. *Clin Sci (Lond), 132*(23), 2483-2491. doi:10.1042/cs20180732

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, L., Wang, J., Liu, X. Y., Liu, H., Shi, H. Y., Yang, X. X., . . . Yang, Y. Q. (2017). HAND1 loss-of-function mutation contributes to congenital double outlet right ventricle. *Int J Mol Med, 39*(3), 711-718. doi:10.3892/ijmm.2017.2865

Li, Y., Klena, N. T., Gabriel, G. C., Liu, X., Kim, A. J., Lemke, K., . . . Lo, C. W. (2015). Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature, 521*(7553), 520-524. doi:10.1038/nature14269

Lin, B., Wang, Y., Wang, Z., Tan, H., Kong, X., Shu, Y., . . . Hu, L. (2014). Uncovering the rare variants of DLC1 isoform 1 and their functional effects in a Chinese sporadic congenital heart disease cohort. *PLoS One, 9*(2), e90215. doi:10.1371/journal.pone.0090215

Lin, B. C., Sullivan, R., Lee, Y., Moran, S., Glover, E., & Bradfield, C. A. (2007). Deletion of the aryl hydrocarbon receptor-associated protein 9 leads to cardiac malformation and embryonic lethality. *J Biol Chem, 282*(49), 35924-35932. doi:10.1074/jbc.M705471200

Lin, C. J., Lin, C. Y., Chen, C. H., Zhou, B., & Chang, C. P. (2012). Partitioning the heart: mechanisms of cardiac septation and valve development. *Development, 139*(18), 3277-3299. doi:10.1242/dev.063495

Lincoln, J., & Garg, V. (2014). Etiology of Valvular Heart Disease. *Circulation Journal, 78*(8), 1801-1807. doi:10.1253/circj.CJ-14-0510

Liu, J., Cheng, H., Xiang, M., Zhou, L., Wu, B., Moskowitz, I. P., . . . Xie, L. (2019). Gata4 regulates hedgehog signaling and Gata6 expression for outflow tract development. *PLoS Genet, 15*(5), e1007711. doi:10.1371/journal.pgen.1007711

Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat, 37*(3), 235-241. doi:10.1002/humu.22932

Liu, X., Yagi, H., Saeed, S., Bais, A. S., Gabriel, G. C., Chen, Z., . . . Lo, C. W. (2017). The complex genetics of hypoplastic left heart syndrome. *Nat Genet*. doi:10.1038/ng.3870

Liu, Y., & Feng, Q. (2012). NOing the heart: role of nitric oxide synthase-3 in heart development. *Differentiation, 84*(1), 54-61. doi:10.1016/j.diff.2012.04.004

Liu, Y., Harmelink, C., Peng, Y., Chen, Y., Wang, Q., & Jiao, K. (2014). CHD7 interacts with BMP R-SMADs to epigenetically regulate cardiogenesis in mice. *Hum Mol Genet, 23*(8), 2145-2156. doi:10.1093/hmg/ddt610

Liu, Y., Jin, Y., Li, J., Seto, E., Kuo, E., Yu, W., . . . Peng, X. (2013). Inactivation of Cdc42 in neural crest cells causes craniofacial and cardiovascular morphogenesis defects. *Dev Biol, 383*(2), 239-252. doi:10.1016/j.ydbio.2013.09.013

Lo, C. W. (2016-2019).

Longo, L., Vanegas, O. C., Patel, M., Rosti, V., Li, H., Waka, J., . . . Luzzatto, L. (2002). Maternally transmitted severe glucose 6-phosphate dehydrogenase deficiency is an embryonic lethal. *EMBO J, 21*(16), 4229-4239. doi:10.1093/emboj/cdf426

Malicki, J., & Avidor-Reiss, T. (2014). From the cytoplasm into the cilium: Bon voyage. *Organogenesis, 10*(1), 138-157.

Marelli, A. J., Mackie, A. S., Ionescu-Ittu, R., Rahme, E., & Pilote, L. (2007). Congenital heart disease in the general population: changing prevalence and age distribution. *Circulation, 115*(2), 163-172. doi:10.1161/CIRCULATIONAHA.106.627224

Mat Bah, M. N., Sapian, M. H., Jamil, M. T., Abdullah, N., Alias, E. Y., & Zahari, N. (2018). The birth prevalence, severity, and temporal trends of congenital heart disease in the middle-income country: A population-based study. *Congenit Heart Dis, 13*(6), 1012-1027. doi:10.1111/chd.12672

Mattassi, R., Manara, E., Colombo, P. G., Manara, S., Porcella, A., Bruno, G., . . . Bertelli, M. (2018). Variant discovery in patients with Mendelian vascular anomalies by next-generation sequencing and their use in patient clinical management. *J Vasc Surg, 67*(3), 922-932 e911. doi:10.1016/j.jvs.2017.02.034

Mayers, C. M., Wadell, J., McLean, K., Venere, M., Malik, M., Shibata, T., . . . Segars, J. H. (2010). The Rho guanine nucleotide exchange factor AKAP13 (BRX) is essential for cardiac development in mice. *J Biol Chem, 285*(16), 12344-12354. doi:10.1074/jbc.M110.106856

Mazor Dray, E., & Marelli, A. J. (2015). Adult Congenital Heart Disease: Scope of the Problem. *Cardiology clinics, 33*(4), 503-vii. doi:10.1016/j.ccl.2015.07.001

McBride, K. L., Pignatelli, R., Lewin, M., Ho, T., Fernbach, S., Menesses, A., . . . Belmont, J. W. (2005). Inheritance analysis of congenital left ventricular outflow tract obstruction

malformations: Segregation, multiplex relative risk, and heritability. *Am J Med Genet A, 134A*(2), 180-186. doi:10.1002/ajmg.a.30602

McCulley, D. J., & Black, B. L. (2012). Transcription factor pathways and congenital heart disease. *Curr Top Dev Biol, 100*, 253-277. doi:10.1016/b978-0-12-387786-4.00008-7

McDonald-McGinn, D. M., Sullivan, K. E., Marino, B., Philip, N., Swillen, A., Vorstman, J. A., . . . Bassett, A. S. (2015). 22q11.2 deletion syndrome. *Nat Rev Dis Primers, 1*, 15071. doi:10.1038/nrdp.2015.71

McFadden, D. G., Barbosa, A. C., Richardson, J. A., Schneider, M. D., Srivastava, D., & Olson, E. N. (2005). The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development, 132*(1), 189-201. doi:10.1242/dev.01562

McInerney-Leo, A. M., Wheeler, L., Marshall, M. S., Anderson, L. K., Zankl, A., Brown, M. A., . . . Duncan, E. L. (2017). Homozygous variant in C21orf2 in a case of Jeune syndrome with severe thoracic involvement: Extending the phenotypic spectrum. *Am J Med Genet A, 173*(6), 1698-1704. doi:10.1002/ajmg.a.38215

Merveille, A. C., Davis, E. E., Becker-Heck, A., Legendre, M., Amirav, I., Bataille, G., . . . Amselem, S. (2011). CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nat Genet, 43*(1), 72-78. doi:10.1038/ng.726

Meyer, D., & Birchmeier, C. (1995). Multiple essential functions of neuregulin in development. *Nature, 378*(6555), 386-390. doi:10.1038/378386a0

Mili, F., Khoury, M. J., Flanders, W. D., & Greenberg, R. S. (1993). Risk of childhood cancer for infants with birth defects. I. A record-linkage study, Atlanta, Georgia, 1968-1988. *Am J Epidemiol, 137*(6), 629-638.

Mili, F., Lynch, C. F., Khoury, M. J., Flanders, W. D., & Edmonds, L. D. (1993). Risk of childhood cancer for infants with birth defects. II. A record-linkage study, Iowa, 1983-1989. *Am J Epidemiol, 137*(6), 639-644. doi:10.1093/oxfordjournals.aje.a116721

Mitchison, H. M., & Valente, E. M. (2017). Motile and non-motile cilia in human pathology: from function to phenotypes. *The Journal of pathology, 241*(2), 294-309. doi:10.1002/path.4843

Miyazawa, K., & Miyazono, K. (2017). Regulation of TGF-beta Family Signaling by Inhibitory Smads. *Cold Spring Harb Perspect Biol, 9*(3). doi:10.1101/cshperspect.a022095

Mo, F. E., & Lau, L. F. (2006). The matricellular protein CCN1 is essential for cardiac development. *Circ Res, 99*(9), 961-969. doi:10.1161/01.RES.0000248426.35019.89

Mommersteeg, M. T., Yeh, M. L., Parnavelas, J. G., & Andrews, W. D. (2015). Disrupted Slit-Robo signalling results in membranous ventricular septum defects and bicuspid aortic valves. *Cardiovasc Res, 106*(1), 55-66. doi:10.1093/cvr/cvv040

Moore, S. W. (2009). Developmental genes and cancer in children. *Pediatr Blood Cancer, 52*(7), 755-760. doi:10.1002/pbc.21831

Morgenthau, A., & Frishman, W. H. (2018). Genetic Origins of Tetralogy of Fallot. *Cardiol Rev, 26*(2), 86-92. doi:10.1097/CRD.0000000000000170

Moustakas, A., & Heldin, C. H. (2002). From mono- to oligo-Smads: the heart of the matter in TGF-beta signal transduction. *Genes Dev, 16*(15), 1867-1871. doi:10.1101/gad.1016802

Muntean, I., Toganel, R., & Benedek, T. (2017). Genetics of Congenital Heart Disease: Past and Present. *Biochem Genet, 55*(2), 105-123. doi:10.1007/s10528-016-9780-7

Myklebust, R., Engedal, H., Saetersdal, T. S., & Ulstein, M. (1977). Primary 9 + 0 cilia in the embryonic and the adult human heart. *Anat Embryol (Berl), 151*(2), 127-139. doi:10.1007/bf00297476

Mykytyn, K., & Askwith, C. (2017). G-Protein-Coupled Receptor Signaling in Cilia. *Cold Spring Harbor perspectives in biology, 9*(9), a028183. doi:10.1101/cshperspect.a028183

Nabeshima, R., Nishimura, O., Maeda, T., Shimizu, N., Ide, T., Yashiro, K., . . . Hamada, H. (2018). Loss of Fam60a, a Sin3a subunit, results in embryonic lethality and is associated with aberrant methylation at a subset of gene promoters. *Elife, 7*. doi:10.7554/eLife.36435

Nachury, M. V. (2018). The molecular machines that traffic signaling receptors into and out of cilia. *Current opinion in cell biology, 51*, 124-131. doi:10.1016/j.ceb.2018.03.004

Nakhleh, N., Francis, R., Giese, R. A., Tian, X., Li, Y., Zariwala, M. A., . . . Lo, C. W. (2012). High prevalence of respiratory ciliary dysfunction in congenital heart disease patients with heterotaxy. *Circulation, 125*(18), 2232-2242. doi:10.1161/CIRCULATIONAHA.111.079780

Narod, S. A., Hawkins, M. M., Robertson, C. M., & Stiller, C. A. (1997). Congenital anomalies and childhood cancer in Great Britain. *Am J Hum Genet, 60*(3), 474-485. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1712528/pdf/ajhg00003-0012.pdf

Nebigil, C. G., Choi, D. S., Dierich, A., Hickel, P., Le Meur, M., Messaddeq, N., . . . Maroteaux, L. (2000). Serotonin 2B receptor is required for heart development. *Proc Natl Acad Sci U S A, 97*(17), 9508-9513. doi:10.1073/pnas.97.17.9508

Nebigil, C. G., & Desaubry, L. (2019). The role of GPCR signaling in cardiac Epithelial to Mesenchymal Transformation (EMT). *Trends Cardiovasc Med, 29*(4), 200-204. doi:10.1016/j.tcm.2018.08.007

Nebigil, C. G., Jaffre, F., Messaddeq, N., Hickel, P., Monassier, L., Launay, J. M., & Maroteaux, L. (2003). Overexpression of the serotonin 5-HT2B receptor in heart leads to abnormal mitochondrial function and cardiac hypertrophy. *Circulation, 107*(25), 3223-3229. doi:10.1161/01.CIR.0000074224.57016.01

Neeb, Z., Lajiness, J. D., Bolanis, E., & Conway, S. J. (2013). Cardiac outflow tract anomalies. *Wiley Interdiscip Rev Dev Biol, 2*(4), 499-530. doi:10.1002/wdev.98

Newbern, J., Zhong, J., Wickramasinghe, R. S., Li, X., Wu, Y., Samuels, I., . . . Landreth, G. E. (2008). Mouse and human phenotypes indicate a critical conserved role for ERK2 signaling in neural crest development. *Proc Natl Acad Sci U S A, 105*(44), 17115-17120. doi:10.1073/pnas.0805239105

Niendorf, S., Oksche, A., Kisser, A., Lohler, J., Prinz, M., Schorle, H., . . . Knobeloch, K. P. (2007). Essential role of ubiquitin-specific protease 8 for receptor tyrosine kinase stability and endocytic trafficking in vivo. *Mol Cell Biol, 27*(13), 5029-5039. doi:10.1128/MCB.01566-06

Nishi, M., Miyake, H., Takeda, T., & Hatae, Y. (2000). Congenital malformations and childhood cancer. *Med Pediatr Oncol, 34*(4), 250-254.

Norwood, M. S., Lupo, P. J., Chow, E. J., Scheurer, M. E., Plon, S. E., Danysh, H. E., . . . Mueller, B. A. (2017). Childhood cancer risk in those with chromosomal and non-chromosomal congenital anomalies in Washington State: 1984-2013. *PLoS One, 12*(6), e0179006. doi:10.1371/journal.pone.0179006

Obler, D., Juraszek, A. L., Smoot, L. B., & Natowicz, M. R. (2008). Double outlet right ventricle: aetiologies and associations. *J Med Genet, 45*(8), 481-497. doi:10.1136/jmg.2008.057984

Oleaga-Alday, A., Goni-Goicoechea, F., Calles-Romero, L., Perez de Ciriza-Cordeu, M., & Paja-Fano, M. (2015). Paraganglioma and cyanotic congenital heart disease: The role of tisular hipoxia. *Endocrinol Nutr, 62*(8), 413-414. doi:10.1016/j.endonu.2015.05.004

Oliveros, J. C. (2007-2015). Venny. An interactive tool for comparing lists with Venn's diagrams. Retrieved from https://bioinfogp.cnb.csic.es/tools/venny/index.html

Olsen, M., Garne, E., Svaerke, C., Sondergaard, L., Nissen, H., Andersen, H. O., . . . Videbaek, J. (2014). Cancer risk among patients with congenital heart defects: a nationwide follow-up study. *Cardiol Young, 24*(1), 40-46. doi:10.1017/s1047951112002144

Opotowsky, A. R., Moko, L. E., Ginns, J., Rosenbaum, M., Greutmann, M., Aboulhosn, J., . . . Vaidya, A. (2015). Pheochromocytoma and paraganglioma in cyanotic congenital heart disease. *J Clin Endocrinol Metab, 100*(4), 1325-1334. doi:10.1210/jc.2014-3863

Oyen, N., Poulsen, G., Boyd, H. A., Wohlfahrt, J., Jensen, P. K., & Melbye, M. (2009). Recurrence of congenital heart defects in families. *Circulation, 120*(4), 295-301. doi:10.1161/CIRCULATIONAHA.109.857987

Pala, R., Alomari, N., & Nauli, S. M. (2017). Primary Cilium-Dependent Signaling Mechanisms. *International journal of molecular sciences, 18*(11), 2272. doi:10.3390/ijms18112272

Paul, M. H., Harvey, R. P., Wegner, M., & Sock, E. (2014). Cardiac outflow tract development relies on the complex function of Sox4 and Sox11 in multiple cell types. *Cell Mol Life Sci, 71*(15), 2931-2945. doi:10.1007/s00018-013-1523-x

Pavan, S., Rommel, K., Mateo Marquina, M. E., Hohn, S., Lanneau, V., & Rath, A. (2017). Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. *PLoS One, 12*(1), e0170365. doi:10.1371/journal.pone.0170365

Penny, D. J., & Vick, G. W., 3rd. (2011). Ventricular septal defect. *Lancet, 377*(9771), 1103-1112. doi:10.1016/S0140-6736(10)61339-6

Perrot, A., Schmitt, K. R., Roth, E. M., Stiller, B., Posch, M. G., Browne, E. N., . . . Ozcelik, C. (2015). CCN1 mutation is associated with atrial septal defect. *Pediatr Cardiol, 36*(2), 295-299. doi:10.1007/s00246-014-1001-8

Pfitzer, C., Helm, P. C., Ferentzi, H., Rosenthal, L. M., Bauer, U. M. M., Berger, F., & Schmitt, K. R. L. (2017). Changing prevalence of severe congenital heart disease: Results from the National Register for Congenital Heart Defects in Germany. *Congenit Heart Dis, 12*(6), 787-793. doi:10.1111/chd.12515

Phillips, M. D., Mukhopadhyay, M., Poscablo, C., & Westphal, H. (2011). Dkk1 and Dkk2 regulate epicardial specification during mouse heart development. *Int J Cardiol, 150*(2), 186-192. doi:10.1016/j.ijcard.2010.04.007

Pierpont, M. E., Basson, C. T., Benson, D. W., Jr., Gelb, B. D., Giglia, T. M., Goldmuntz, E., . . . American Heart Association Congenital Cardiac Defects Committee, C. o. C. D. i. t. Y. (2007). Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation, 115*(23), 3015-3038. doi:10.1161/CIRCULATIONAHA.106.183056

Pierpont, M. E., Brueckner, M., Chung, W. K., Garg, V., Lacro, R. V., McGuire, A. L., . . . Precision, M. (2018). Genetic Basis for Congenital Heart Disease: Revisited: A Scientific Statement From the American Heart Association. *Circulation, 138*(21), e653-e711. doi:10.1161/CIR.0000000000000606

Poelmann, R. E., Molin, D., Wisse, L. J., & Gittenberger-de Groot, A. C. (2000). Apoptosis in cardiac development. *Cell Tissue Res, 301*(1), 43-52. doi:10.1007/s004410000227

Prendiville, T., Jay, P. Y., & Pu, W. T. (2014). Insights into the genetic structure of congenital heart disease from human and murine studies on monogenic disorders. *Cold Spring Harb Perspect Med, 4*(10). doi:10.1101/cshperspect.a013946

Preuss, C., Capredon, M., Wunnemann, F., Chetaille, P., Prince, A., Godard, B., . . . Andelfinger, G. (2016). Family Based Whole Exome Sequencing Reveals the Multifaceted Role of Notch Signaling in Congenital Heart Disease. *PLoS Genet, 12*(10), e1006335. doi:10.1371/journal.pgen.1006335

Priest, J. R., Osoegawa, K., Mohammed, N., Nanda, V., Kundu, R., Schultz, K., . . . Ashley, E. A. (2016). De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. *PLoS Genet, 12*(4), e1005963. doi:10.1371/journal.pgen.1005963

Qiao, X. H., Wang, Q., Wang, J., Liu, X. Y., Xu, Y. J., Huang, R. T., . . . Yang, Y. Q. (2018). A novel NR2F2 loss-of-function mutation predisposes to congenital heart defect. *Eur J Med Genet, 61*(4), 197-203. doi:10.1016/j.ejmg.2017.12.003

Qiu, Z., Cang, Y., & Goff, S. P. (2010). c-Abl tyrosine kinase regulates cardiac growth and development. *Proc Natl Acad Sci U S A, 107*(3), 1136-1141. doi:10.1073/pnas.0913131107

R Development Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Racedo, S. E., Hasten, E., Lin, M., Devakanmalai, G. S., Guo, T., Ozbudak, E. M., . . . Morrow, B. E. (2017). Reduced dosage of beta-catenin provides significant rescue of cardiac outflow tract anomalies in a Tbx1 conditional null mouse model of 22q11.2 deletion syndrome. *PLoS Genet, 13*(3), e1006687. doi:10.1371/journal.pgen.1006687

Raissadati, A., Nieminen, H., Haukka, J., Sairanen, H., & Jokinen, E. (2016). Late Causes of Death After Pediatric Cardiac Surgery: A 60-Year Population-Based Study. *J Am Coll Cardiol, 68*(5), 487-498. doi:10.1016/j.jacc.2016.05.038

Rambo-Martin, B. L., Mulle, J. G., Cutler, D. J., Bean, L. J. H., Rosser, T. C., Dooley, K. J., . . . Zwick, M. E. (2018). Analysis of Copy Number Variants on Chromosome 21 in Down Syndrome-Associated Congenital Heart Defects. *G3 (Bethesda), 8*(1), 105-111. doi:10.1534/g3.117.300366

Ramsdell, A. F. (2005). Left-right asymmetry and congenital cardiac defects: getting to the heart of the matter in vertebrate left-right axis determination. *Dev Biol, 288*(1), 1-20. doi:10.1016/j.ydbio.2005.07.038

Rash, J. E., Shay, J. W., & Biesele, J. J. (1969). Cilia in cardiac differentiation. *J Ultrastruct Res, 29*(5), 470-484. doi:10.1016/s0022-5320(69)90067-7

Rauen, K. A. (2013). The RASopathies. *Annu Rev Genomics Hum Genet, 14*, 355-369. doi:10.1146/annurev-genom-091212-153523

Reiter, J. F., & Leroux, M. R. (2017). Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol, 18*(9), 533-547. doi:10.1038/nrm.2017.60

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res, 47*(D1), D886-D894. doi:10.1093/nar/gky1016

Reuter, M. S., Jobling, R., Chaturvedi, R. R., Manshaei, R., Costain, G., Heung, T., . . . Bassett, A. S. (2018). Haploinsufficiency of vascular endothelial growth factor related signaling genes is associated with tetralogy of Fallot. *Genet Med*. doi:10.1038/s41436-018-0260-9

Rhee, S., Chung, J. I., King, D. A., D'Amato, G., Paik, D. T., Duan, A., . . . Red-Horse, K. (2018). Endothelial deletion of Ino80 disrupts coronary angiogenesis and causes congenital heart disease. *Nat Commun, 9*(1), 368. doi:10.1038/s41467-017-02796-3

Richards, A. A., & Garg, V. (2010). Genetics of congenital heart disease. *Curr Cardiol Rev, 6*(2), 91-97. doi:10.2174/157340310791162703

Robson, A., Makova, S. Z., Barish, S., Zaidi, S., Mehta, S., Drozd, J., . . . Brueckner, M. (2019). Histone H2B monoubiquitination regulates heart development via epigenetic control of cilia motility. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1808341116

Rochais, F., Mesbah, K., & Kelly, R. G. (2009). Signaling pathways controlling second heart field development. *Circ Res, 104*(8), 933-942. doi:10.1161/circresaha.109.194464

Roessler, E., Ouspenskaia, M. V., Karkera, J. D., Velez, J. I., Kantipong, A., Lacbawan, F., . . . Muenke, M. (2008). Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. *Am J Hum Genet, 83*(1), 18-29. doi:10.1016/j.ajhg.2008.05.012

Romano, A. A., Allanson, J. E., Dahlgren, J., Gelb, B. D., Hall, B., Pierpont, M. E., . . . Noonan, J. A. (2010). Noonan syndrome: clinical features, diagnosis, and management guidelines. *Pediatrics, 126*(4), 746-759. doi:10.1542/peds.2009-3207

Sabbir, M. G., Wigle, N., Loewen, S., Gu, Y., Buse, C., Hicks, G. G., & Mowat, M. R. (2010). Identification and characterization of Dlc1 isoforms in the mouse and study of the biological function of a single gene trapped isoform. *BMC Biol, 8*, 17. doi:10.1186/1741-7007-8-17

Sanchis, D., Llovera, M., Ballester, M., & Comella, J. X. (2008). An alternative view of apoptosis in heart development and disease. *Cardiovasc Res, 77*(3), 448-451. doi:10.1093/cvr/cvm074

Santos, R., Kawauchi, S., Jacobs, R. E., Lopez-Burks, M. E., Choi, H., Wikenheiser, J., . . . Calof, A. L. (2016). Conditional Creation and Rescue of Nipbl-Deficiency in Mice Reveals Multiple Determinants of Risk for Congenital Heart Defects. *PLoS Biol, 14*(9), e2000197. doi:10.1371/journal.pbio.2000197

Segura, I., Lange, C., Knevels, E., Moskalyuk, A., Pulizzi, R., Eelen, G., . . . Carmeliet, P. (2016). The Oxygen Sensor PHD2 Controls Dendritic Spines and Synapses via Modification of Filamin A. *Cell Rep, 14*(11), 2653-2667. doi:10.1016/j.celrep.2016.02.047

Shafaattalab, S., Li, A. Y., Lin, E., Stevens, C. M., Dewar, L. J., Lynn, F. C., . . . Tibbits, G. F. (2019). In vitro analyses of suspected arrhythmogenic thin filament variants as a cause of sudden cardiac death in infants. *Proc Natl Acad Sci U S A, 116*(14), 6969-6974. doi:10.1073/pnas.1819023116

Shanshen, E., Rosenberg, J., & Van Bergen, A. H. (2018). Identification of Novel Congenital Heart Disease Candidate Genes Using Chromosome Microarray. *Pediatr Cardiol, 39*(1), 148-159. doi:10.1007/s00246-017-1741-3

Shapiro, A. J., Davis, S. D., Ferkol, T., Dell, S. D., Rosenfeld, M., Olivier, K. N., . . . Leigh, M. W. (2014). Laterality defects other than situs inversus totalis in primary ciliary dyskinesia: insights into situs ambiguus and heterotaxy. *Chest, 146*(5), 1176-1186. doi:10.1378/chest.13-1704

Siamwala, J. H., Kumar, P., Veeriah, V., Muley, A., Rajendran, S., Konikkat, S., . . . Chatterjee, S. (2019). Nitric Oxide Reverses the Position of the Heart during Embryonic Development. *International journal of molecular sciences, 20*(5). doi:10.3390/ijms20051157

Sierro, F., Biben, C., Martinez-Munoz, L., Mellado, M., Ransohoff, R. M., Li, M., . . . Mackay, F. (2007). Disrupted cardiac development but normal hematopoiesis in mice deficient in the second CXCL12/SDF-1 receptor, CXCR7. *Proc Natl Acad Sci U S A, 104*(37), 14759-14764. doi:10.1073/pnas.0702229104

Sifrim, A., Hitz, M. P., Wilsdon, A., Breckpot, J., Turki, S. H., Thienpont, B., . . . Hurles, M. E. (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet*. doi:10.1038/ng.3627

Silversides, C. K., Lionel, A. C., Costain, G., Merico, D., Migita, O., Liu, B., . . . Bassett, A. S. (2012). Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS Genet, 8*(8), e1002843. doi:10.1371/journal.pgen.1002843

Simms, R. J., Hynes, A. M., Eley, L., Inglis, D., Chaudhry, B., Dawe, H. R., & Sayer, J. A. (2012). Modelling a ciliopathy: Ahi1 knockdown in model systems reveals an essential role in brain, retinal, and renal development. *Cell Mol Life Sci, 69*(6), 993-1009. doi:10.1007/s00018-011-0826-z

Simrick, S., Szumska, D., Gardiner, J. R., Jones, K., Sagar, K., Morrow, B., . . . Basson, M. A. (2012). Biallelic expression of Tbx1 protects the embryo from developmental defects caused by increased receptor tyrosine kinase signaling. *Dev Dyn, 241*(8), 1310-1324. doi:10.1002/dvdy.23812

Slough, J., Cooney, L., & Brueckner, M. (2008). Monocilia in the embryonic mouse heart suggest a direct role for cilia in cardiac morphogenesis. *Dev Dyn, 237*(9), 2304-2314. doi:10.1002/dvdy.21669

Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., & Forbes, S. A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer, 18*(11), 696-705. doi:10.1038/s41568-018-0060-1

Soonpaa, M. H., Koh, G. Y., Pajak, L., Jing, S., Wang, H., Franklin, M. T., . . . Field, L. J. (1997). Cyclin D1 overexpression promotes cardiomyocyte DNA synthesis and multinucleation in transgenic mice. *J Clin Invest, 99*(11), 2644-2654. doi:10.1172/JCI119453

Srivastava, D., & Olson, E. N. (2000). A genetic blueprint for cardiac development. *Nature, 407*(6801), 221-226. doi:10.1038/35025190

Srivastava, P., Pandey, H., Agarwal, D., Mandal, K., & Phadke, S. R. (2017). Spondyloepiphyseal dysplasia Omani type: CHST3 mutation spectrum and phenotypes in three Indian families. *Am J Med Genet A, 173*(1), 163-168. doi:10.1002/ajmg.a.37996

Stenson, P. D., Mort, M., Ball, E. V., Shaw, K., Phillips, A. D., & Cooper, D. N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human Genetics, 133*(1), 1-9. doi:10.1007/s00439-013-1358-4

Sun, Y., Overvad, K., & Olsen, J. (2014). Cancer risks in children with congenital malformations in the nervous and circulatory system-A population based cohort study. *Cancer Epidemiol, 38*(4), 393-400. doi:10.1016/j.canep.2014.04.001

Sund, K. L., Roelker, S., Ramachandran, V., Durbin, L., & Benson, D. W. (2009). Analysis of Ellis van Creveld syndrome gene products: implications for cardiovascular development and disease. *Hum Mol Genet, 18*(10), 1813-1824. doi:10.1093/hmg/ddp098

Sutherland, M. J., & Ware, S. M. (2009). Disorders of left-right asymmetry: heterotaxy and situs inversus. *Am J Med Genet C Semin Med Genet, 151c*(4), 307-317. doi:10.1002/ajmg.c.30228

Sylva, M., van den Hoff, M. J., & Moorman, A. F. (2014). Development of the human heart. *Am J Med Genet A, 164a*(6), 1347-1371. doi:10.1002/ajmg.a.35896

Ta-Shma, A., Pierri, C. L., Stepensky, P., Shaag, A., Zenvirt, S., Elpeleg, O., & Rein, A. J. (2013). Isolated truncus arteriosus associated with a mutation in the plexin-D1 gene. *Am J Med Genet A, 161A*(12), 3115-3120. doi:10.1002/ajmg.a.36194

Takabayashi, S., & Katoh, H. (2005). A mutant mouse with severe anemia and skin abnormalities controlled by a new allele of the flaky skin (fsn) locus. *Exp Anim, 54*(4), 339-347. doi:10.1538/expanim.54.339

Takeuchi, J. K., Lou, X., Alexander, J. M., Sugizaki, H., Delgado-Olguin, P., Holloway, A. K., . . . Bruneau, B. G. (2011). Chromatin remodelling complex dosage modulates transcription factor function in heart development. *Nat Commun, 2*, 187. doi:10.1038/ncomms1187

Tao, G., Levay, A. K., Peacock, J. D., Huk, D. J., Both, S. N., Purcell, N. H., . . . Lincoln, J. (2012). Collagen XIV is important for growth and structural integrity of the myocardium. *J Mol Cell Cardiol, 53*(5), 626-638. doi:10.1016/j.yjmcc.2012.08.002

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., . . . Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res, 47*(D1), D941-D947. doi:10.1093/nar/gky1015

Tessarollo, L., Tsoulfas, P., Donovan, M. J., Palko, M. E., Blair-Flynn, J., Hempstead, B. L., & Parada, L. F. (1997). Targeted deletion of all isoforms of the trkC gene suggests the use of alternate receptors by its ligand neurotrophin-3 in neuronal development and implicates trkC in normal cardiogenesis. *Proc Natl Acad Sci U S A, 94*(26), 14776-14781. doi:10.1073/pnas.94.26.14776

The UK10K Consortium, Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., . . . Soranzo, N. (2015). The UK10K project identifies rare variants in health and disease. *Nature, 526*(7571), 82-90. doi:10.1038/nature14962

Thiene, G., & Frescura, C. (2010). Anatomical and pathophysiological classification of congenital heart disease. *Cardiovasc Pathol, 19*(5), 259-274. doi:10.1016/j.carpath.2010.02.006

Tian, E., Stevens, S. R., Guan, Y., Springer, D. A., Anderson, S. A., Starost, M. F., . . . Tabak, L. A. (2015). Galnt1 is required for normal heart valve development and cardiac function. *PLoS One, 10*(1), e0115861. doi:10.1371/journal.pone.0115861

Togi, K., Kawamoto, T., Yamauchi, R., Yoshida, Y., Kita, T., & Tanaka, M. (2004). Role of Hand1/eHAND in the dorso-ventral patterning and interventricular septum formation in the embryonic heart. *Mol Cell Biol, 24*(11), 4627-4635. doi:10.1128/MCB.24.11.4627-4635.2004

Toomer, K. A., Fulmer, D., Guo, L., Drohan, A., Peterson, N., Swanson, P., . . . Norris, R. A. (2017). A role for primary cilia in aortic valve development and disease. *Dev Dyn, 246*(8), 625-634. doi:10.1002/dvdy.24524

Toomer, K. A., Yu, M., Fulmer, D., Guo, L., Moore, K. S., Moore, R., . . . Norris, R. A. (2019). Primary cilia defects causing mitral valve prolapse. *Sci Transl Med, 11*(493). doi:10.1126/scitranslmed.aax0290

Toriyama, M., Lee, C., Taylor, S. P., Duran, I., Cohn, D. H., Bruel, A. L., . . . Wallingford, J. B. (2016). The ciliopathy-associated CPLANE proteins direct basal body recruitment of intraflagellar transport machinery. *Nat Genet, 48*(6), 648-656. doi:10.1038/ng.3558

Tsao, C. W., & Vasan, R. S. (2015). Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol, 44*(6), 1800-1813. doi:10.1093/ije/dyv337

Tuysuz, B., Mizumoto, S., Sugahara, K., Celebi, A., Mundlos, S., & Turkmen, S. (2009). Omani-type spondyloepiphyseal dysplasia with cardiac involvement caused by a missense mutation in CHST3. *Clin Genet, 75*(4), 375-383. doi:10.1111/j.1399-0004.2009.01167.x

Twigg, S. R., Lloyd, D., Jenkins, D., Elcioglu, N. E., Cooper, C. D., Al-Sannaa, N., . . . Wilkie, A. O. (2012). Mutations in multidomain protein MEGF8 identify a Carpenter syndrome subtype associated with defective lateralization. *Am J Hum Genet, 91*(5), 897-905. doi:10.1016/j.ajhg.2012.08.027

Uribe, V., Badia-Careaga, C., Casanova, J. C., Dominguez, J. N., de la Pompa, J. L., & Sanz-Ezquerro, J. J. (2014). Arid3b is essential for second heart field cell deployment and heart patterning. *Development, 141*(21), 4168-4181. doi:10.1242/dev.109918

van Dam, T. J., Wheway, G., Slaats, G. G., Huynen, M. A., & Giles, R. H. (2013). The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia, 2*(1), 7. doi:10.1186/2046-2530-2-7

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., . . . DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics, 43*, 11.10.11-11.10.33. doi:10.1002/0471250953.bi1110s43

van der Bom, T., Zomer, A. C., Zwinderman, A. H., Meijboom, F. J., Bouma, B. J., & Mulder, B. J. (2011). The changing epidemiology of congenital heart disease. *Nat Rev Cardiol, 8*(1), 50-60. doi:10.1038/nrcardio.2010.166

Van der Heiden, K., Groenendijk, B. C., Hierck, B. P., Hogers, B., Koerten, H. K., Mommaas, A. M., . . . Poelmann, R. E. (2006). Monocilia on chicken embryonic endocardium in low shear stress areas. *Dev Dyn, 235*(1), 19-28. doi:10.1002/dvdy.20557

van der Linde, D., Konings, E. E., Slager, M. A., Witsenburg, M., Helbing, W. A., Takkenberg, J. J., & Roos-Hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol, 58*(21), 2241-2247. doi:10.1016/j.jacc.2011.08.025

Van Nostrand, J. L., Brady, C. A., Jung, H., Fuentes, D. R., Kozak, M. M., Johnson, T. M., . . . Attardi, L. D. (2014). Inappropriate p53 activation during development induces features of CHARGE syndrome. *Nature, 514*(7521), 228-232. doi:10.1038/nature13585

van Praagh, S. (2006). Cardiac malpositions and the heterotaxy syndromes. In J. F. Keane, D. C. Fyler, & J. E. Lock (Eds.), *Nadas' Pediatric Cardiology* (2nd ed., pp. 675-695). Philadelphia: Saunders.

van Rooijen, E., Giles, R. H., Voest, E. E., van Rooijen, C., Schulte-Merker, S., & van Eeden, F. J. (2008). LRRC50, a conserved ciliary protein implicated in polycystic kidney disease. *J Am Soc Nephrol, 19*(6), 1128-1138. doi:10.1681/ASN.2007080917

van Waning, J. I., Caliskan, K., Hoedemaekers, Y. M., van Spaendonck-Zwarts, K. Y., Baas, A. F., Boekholdt, S. M., . . . Majoor-Krakauer, D. (2018). Genetics, Clinical Features, and Long-Term Outcome of Noncompaction Cardiomyopathy. *J Am Coll Cardiol, 71*(7), 711-722. doi:10.1016/j.jacc.2017.12.019

Vandersteen, A. M., Lund, A. M., Ferguson, D. J., Sawle, P., Pollitt, R. C., Holder, S. E., . . . Pope, F. M. (2014). Four patients with Sillence type I osteogenesis imperfecta and mild bone fragility, complicated by left ventricular cardiac valvular disease and cardiac tissue fragility caused by type I collagen mutations. *Am J Med Genet A, 164A*(2), 386-391. doi:10.1002/ajmg.a.36285

Verloes, A., Di Donato, N., Masliah-Planchon, J., Jongmans, M., Abdul-Raman, O. A., Albrecht, B., . . . Pilz, D. T. (2015). Baraitser-Winter cerebrofrontofacial syndrome: delineation of the spectrum in 42 cases. *Eur J Hum Genet, 23*(3), 292-301. doi:10.1038/ejhg.2014.95

Versacci, P., Di Carlo, D., Digilio, M. C., & Marino, B. (2018). Cardiovascular disease in Down syndrome. *Curr Opin Pediatr, 30*(5), 616-622. doi:10.1097/MOP.0000000000000661

Vicente Steijn, R., Sedmera, D., Blom, N. A., Jongbloed, M., Kvasilova, A., & Nanka, O. (2018). Apoptosis and epicardial contributions act as complementary factors in remodeling of the atrioventricular canal myocardium and atrioventricular conduction patterns in the embryonic chick heart. *Dev Dyn, 247*(9), 1033-1042. doi:10.1002/dvdy.24642

Villalobos, E., Criollo, A., Schiattarella, G. G., Altamirano, F., French, K. M., May, H. I., . . . Hill, J. A. (2019). Fibroblast Primary Cilia Are Required for Cardiac Fibrosis. *Circulation, 139*(20), 2342-2357. doi:10.1161/circulationaha.117.028752

Vissers, L. E., Bonetti, M., Paardekooper Overman, J., Nillesen, W. M., Frints, S. G., de Ligt, J., . . . den Hertog, J. (2015). Heterozygous germline mutations in A2ML1 are associated with a disorder clinically related to Noonan syndrome. *Eur J Hum Genet, 23*(3), 317-324. doi:10.1038/ejhg.2014.115

Wagsater, D., Paloschi, V., Hanemaaijer, R., Hultenby, K., Bank, R. A., Franco-Cereceda, A., . . . Eriksson, P. (2013). Impaired collagen biosynthesis and cross-linking in aorta of patients with bicuspid aortic valve. *J Am Heart Assoc, 2*(1), e000034. doi:10.1161/JAHA.112.000034

Wang, B., Yan, J., Mi, R., Zhou, S., Xie, X., Wang, J., & Ma, X. (2010). Forkhead box H1 (FOXH1) sequence variants in ventricular septal defect. *Int J Cardiol, 145*(1), 83-85. doi:10.1016/j.ijcard.2009.05.030

Wang, L., Liu, Z., Lin, H., Ma, D., Tao, Q., & Liu, F. (2017). Epigenetic regulation of left-right asymmetry by DNA methylation. *EMBO J, 36*(20), 2987-2997. doi:10.15252/embj.201796580

Wang, X., Charng, W. L., Chen, C. A., Rosenfeld, J. A., Al Shamsi, A., Al-Gazali, L., . . . Yang, Y. (2017). Germline mutations in ABL1 cause an autosomal dominant syndrome characterized by congenital heart defects and skeletal malformations. *Nat Genet, 49*(4), 613-617. doi:10.1038/ng.3815

Wang, Y., Hu, G., Liu, F., Wang, X., Wu, M., Schwarz, J. J., & Zhou, J. (2014). Deletion of yes-associated protein (YAP) specifically in cardiac and vascular smooth muscle cells reveals a crucial role for YAP in mouse cardiovascular development. *Circ Res, 114*(6), 957-965. doi:10.1161/CIRCRESAHA.114.303411

Werner, P., Paluru, P., Simpson, A. M., Latney, B., Iyer, R., Brodeur, G. M., & Goldmuntz, E. (2014). Mutations in NTRK3 suggest a novel signaling pathway in human congenital heart disease. *Hum Mutat, 35*(12), 1459-1468. doi:10.1002/humu.22688

Wessels, M. W., & Willems, P. J. (2010). Genetic factors in non-syndromic congenital heart malformations. *Clin Genet, 78*(2), 103-123. doi:10.1111/j.1399-0004.2010.01435.x

Wheway, G., Schmidts, M., Mans, D. A., Szymanska, K., Nguyen, T. T., Racher, H., . . . Johnson, C. A. (2015). An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes. *Nat Cell Biol, 17*(8), 1074-1087. doi:10.1038/ncb3201

Willaredt, M. A., Gorgas, K., Gardner, H. A., & Tucker, K. L. (2012). Multiple essential roles for primary cilia in heart development. *Cilia, 1*(1), 23. doi:10.1186/2046-2530-1-23

Williams, K., Carson, J., & Lo, C. (2019). Genetics of Congenital Heart Disease. *Biomolecules, 9*(12). doi:10.3390/biom9120879

Windham, G. C., Bjerkedal, T., & Langmark, F. (1985). A population-based study of cancer incidence in twins and in children with congenital malformations or low birth weight, Norway, 1967-1980. *Am J Epidemiol, 121*(1), 49-56.

Wolf, M. T. (2015). Nephronophthisis and related syndromes. *Curr Opin Pediatr, 27*(2), 201-211. doi:10.1097/MOP.0000000000000194

Wu, G., Markowitz, G. S., Li, L., D'Agati, V. D., Factor, S. M., Geng, L., . . . Somlo, S. (2000). Cardiac defects and renal failure in mice with targeted mutations in Pkd2. *Nat Genet, 24*(1), 75-78. doi:10.1038/71724

Wu, M., Li, Y., He, X., Shao, X., Yang, F., Zhao, M., . . . Zhou, L. (2013). Mutational and functional analysis of the BVES gene coding region in Chinese patients with non-syndromic tetralogy of Fallot. *Int J Mol Med, 31*(4), 899-903. doi:10.3892/ijmm.2013.1275

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet, 89*(1), 82-93. doi:10.1016/j.ajhg.2011.05.029

Wu, S. P., Cheng, C. M., Lanz, R. B., Wang, T., Respress, J. L., Ather, S., . . . Tsai, S. Y. (2013). Atrial identity is determined by a COUP-TFII regulatory network. *Dev Cell, 25*(4), 417-426. doi:10.1016/j.devcel.2013.04.017

Xiao, D., Wang, H., Hao, L., Guo, X., Ma, X., Qian, Y., . . . Ma, D. (2018). The roles of SMYD4 in epigenetic regulation of cardiac development in zebrafish. *PLoS Genet, 14*(8), e1007578. doi:10.1371/journal.pgen.1007578

Xie, H. M., Werner, P., Stambolian, D., Bailey-Wilson, J. E., Hakonarson, H., White, P. S., . . . Goldmuntz, E. (2017). Rare copy number variants in patients with congenital conotruncal heart defects. *Birth Defects Res, 109*(4), 271-295. doi:10.1002/bdra.23609

Xuan, C., Jia, K. G., Wang, B. B., Bai, X. Y., Gao, G., Yang, Q., . . . He, G. W. (2013). Identification of two novel mutations of the HOMEZ gene in Chinese patients with isolated ventricular septal defect. *Genet Test Mol Biomarkers, 17*(5), 390-394. doi:10.1089/gtmb.2012.0435

Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., Minoshima, S., . . . Matsuoka, R. (2003). Role of TBX1 in human del22q11.2 syndrome. *The Lancet, 362*(9393), 1366-1373. doi:10.1016/s0140-6736(03)14632-6

Yagi, H., Liu, X., Gabriel, G. C., Wu, Y., Peterson, K., Murray, S. A., . . . Lo, C. W. (2018). The Genetic Landscape of Hypoplastic Left Heart Syndrome. *Pediatr Cardiol*. doi:10.1007/s00246-018-1861-4

Yakoumakis, E., Kostopoulou, H., Makri, T., Dimitriadis, A., Georgiou, E., & Tsalafoutas, I. (2013). Estimation of radiation dose and risk to children undergoing cardiac catheterization for the treatment of a congenital heart disease using Monte Carlo simulations. *Pediatr Radiol, 43*(3), 339-346. doi:10.1007/s00247-012-2510-3

Yan, Q., Tiwari, H. K., Yi, N., Lin, W. Y., Gao, G., Lou, X. Y., . . . Liu, N. (2014). Kernel-machine testing coupled with a rank-truncation method for genetic pathway analysis. *Genet Epidemiol, 38*(5), 447-456. doi:10.1002/gepi.21813

Yang, B., Zhou, W., Jiao, J., Nielsen, J. B., Mathis, M. R., Heydarpour, M., . . . Willer, C. J. (2017). Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve. *Nat Commun, 8*, 15481. doi:10.1038/ncomms15481

Yang, D., Lai, D., Huang, X., Shi, X., Gao, Z., Huang, F., . . . Geng, Y. J. (2012). The defects in development and apoptosis of cardiomyocytes in mice lacking the transcriptional factor Pax-8. *Int J Cardiol, 154*(1), 43-51. doi:10.1016/j.ijcard.2010.08.057

Yang, F., Zhou, L., Wang, Q., You, X., Li, Y., Zhao, Y., . . . Cao, H. (2014). NEXN inhibits GATA4 and leads to atrial septal defects in mice and humans. *Cardiovasc Res, 103*(2), 228-237. doi:10.1093/cvr/cvu134

Ye, B., Hou, N., Xiao, L., Xu, Y., Boyer, J., Xu, H., & Li, F. (2015). APC controls asymmetric Wnt/beta-catenin signaling and cardiomyocyte proliferation gradient in the heart. *J Mol Cell Cardiol, 89*(Pt B), 287-296. doi:10.1016/j.yjmcc.2015.10.018

Yu, Q., Morales, M., Li, N., Fritz, A. G., Ruobing, R., Blaeser, A., . . . Spurney, C. F. (2018). Skeletal, cardiac, and respiratory muscle function and histopathology in the P448Lneo-mouse model of FKRP-deficient muscular dystrophy. *Skelet Muscle, 8*(1), 13. doi:10.1186/s13395-018-0158-x

Yu, S., Crawford, D., Tsuchihashi, T., Behrens, T. W., & Srivastava, D. (2011). The chemokine receptor CXCR7 functions to regulate cardiac valve remodeling. *Dev Dyn, 240*(2), 384-393. doi:10.1002/dvdy.22549

Zahavich, L., Bowdin, S., & Mital, S. (2017). Use of Clinical Exome Sequencing in Isolated Congenital Heart Disease. *Circ Cardiovasc Genet, 10*(3). doi:10.1161/CIRCGENETICS.116.001581

Zaidi, S., & Brueckner, M. (2017). Genetics and Genomics of Congenital Heart Disease. *Circ Res, 120*(6), 923-940. doi:10.1161/circresaha.116.309140

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J. D., . . . Lifton, R. P. (2013). De novo mutations in histone-modifying genes in congenital heart disease. *Nature, 498*(7453), 220-223. doi:10.1038/nature12141

Zakariyah, A. F., Rajgara, R. F., Horner, E., Cattin, M. E., Blais, A., Skerjanc, I. S., & Burgon, P. G. (2018). In Vitro Modeling of Congenital Heart Defects Associated with an NKX2-5 Mutation Revealed a Dysregulation in BMP/Notch-Mediated Signaling. *Stem Cells, 36*(4), 514-526. doi:10.1002/stem.2766

Zariwala, M. A., Knowles, M. R., & Leigh, M. W. (2007, 2019 Dec 5). Primary Ciliary Dyskinesia. *GeneReviews [Internet]*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK1122/

Zhang, Y. Y., Zhou, X., Ji, W. J., Liu, T., Ma, J., Zhang, Y., & Li, Y. M. (2019). Association between CYP2C19*2/*3 Polymorphisms and Coronary Heart Disease. *Curr Med Sci, 39*(1), 44-51. doi:10.1007/s11596-019-1998-2

Zhao, C. M., Peng, L. Y., Li, L., Liu, X. Y., Wang, J., Zhang, X. L., . . . Yang, Y. Q. (2015). PITX2 Loss-of-Function Mutation Contributes to Congenital Endocardial Cushion Defect and Axenfeld-Rieger Syndrome. *PLoS One, 10*(4), e0124409. doi:10.1371/journal.pone.0124409

Zhao, R., Watt, A. J., Battle, M. A., Li, J., Bondow, B. J., & Duncan, S. A. (2008). Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev Biol, 317*(2), 614-619. doi:10.1016/j.ydbio.2008.03.013

Zhao, Z., & Rivkees, S. A. (2000). Programmed cell death in the developing heart: regulation by BMP4 and FGF2. *Dev Dyn, 217*(4), 388-400. doi:10.1002/(sici)1097-0177(200004)217:4<388::Aid-dvdy6>3.0.Co;2-n

Zheng, J., Guo, J., Huang, L., Wu, Q., Yin, K., Wang, L., . . . Cheng, J. (2018). Genetic diagnosis of acute aortic dissection in South China Han population using next-generation sequencing. *Int J Legal Med, 132*(5), 1273-1280. doi:10.1007/s00414-018-1890-9

Zhou, H. M., Weskamp, G., Chesneau, V., Sahin, U., Vortkamp, A., Horiuchi, K., . . . Blobel, C. P. (2004). Essential role for ADAM19 in cardiovascular morphogenesis. *Mol Cell Biol, 24*(1), 96-104. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14673146 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC303363/pdf/1089.pdf

Zhu, X., Deng, X., Huang, G., Wang, J., Yang, J., Chen, S., . . . Wang, B. (2014). A novel mutation of Hyaluronan synthase 2 gene in Chinese children with ventricular septal defect. *PLoS One, 9*(2), e87437. doi:10.1371/journal.pone.0087437