Ciliary genes contribute to a complex genetic model of congenital heart disease

by

Kylia Alexandra Williams

B.A, Northwestern University, 2016

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Kylia Alexandra Williams

It was defended on

September 28, 2021

and approved by

Dennis Kostka, PhD, Associate Professor, Department of Developmental Biology

Madhavi Ganapathiraju, PhD, Associate Professor, Department of Biomedical Informatics

Zhao Ren, PhD, Associate Professor, Department of Statistics

Michael Tsang, PhD, Professor, Department of Developmental Biology

Dissertation Director:
Cecilia Lo, PhD, Professor and Chair, Department of Developmental Biology

Ciliary genes contribute to a complex genetic model of congenital heart disease

Kylia Alexandra Williams, PhD

University of Pittsburgh, 2021

Congenital heart disease (CHD) is the most common birth defect, affecting ~1% of infants born each year. While a genetic etiology is strongly supported for CHD, the majority of cases remain unsolved. The genetic heterogeneity, variable expressivity, and incomplete penetrance of CHD suggests that a more complex or non-Mendelian genetic model is involved. A large-scale mouse mutagenesis screen in our lab previously showed an enrichment for cilia related genes among genes causing CHD, including many involved in the ciliogenesis and planar cell polarity network, as well as cilia-transduced cell signaling pathways known to play important roles in cardiovascular development. Importantly, these genes were shown to be part of a tight protein-protein interaction network, and CHD in some mouse lines was observed to have a multigenic etiology. Hence, we hypothesized that a complex genetic model comprising interactions between ciliary genes can contribute to CHD pathogenesis. Here, we investigated the role of protein truncating variants in known CHD genes that are cilia-related in a cohort of 1932 CHD patients and 2602 controls without structural cardiac defects. We show that there is a significantly greater number of CHD cases with variants in more than one cilia-CHD gene than controls. When considering specific subtypes, this is true for patients with left ventricular outflow tract obstructions and laterality defect patients, suggesting that in the context of ciliary variants, a digenic model drives disease in these subtypes. Conversely, there are significantly more conotruncal defect patients with variants in only one cilia-CHD gene compared to controls, suggesting that in ciliary genes, a monogenic model of disease drives this phenotype. We then

identify gene-gene interactions between genes in the cilia-CHD interactome that are statistically significant in cases, but not controls, and use these to identify candidate CHD genes, including ciliary genes that interact or cluster with known CHD genes. Overall, these studies provide evidence for a complex genetic model of CHD involving ciliary genes, identify gene-gene interactions between known CHD genes that may contribute to pathogenesis, and use these interactions to identify novel candidate CHD genes for future study.

Table of Contents

List of Tables

List of Figures

Acknowledgements

For all the people without whom this would not have been possible: First and foremost, my advisor Dr. Cecilia Lo, who has pushed me to become a better scientist and communicator. My thesis committee, and everyone in the Lo Lab and the Dev Bio department, especially Dr. Abha Bais, who helped me so much along the way. Everyone who started with me in the ISB program, with whom I can always discuss what exactly constitutes a sandwich, and Shari Murphy who kept us all on track. The PalPITTations, who helped me sing my way back to sanity every Tuesday night, especially Michelle Spicer, who quite literally scaled a mountain with me. My dog Apollo, who is the cutest ever and was dropped from the heavens exactly when I needed him. My therapist, Susan, who reminds me that I can do hard things and that I do them well. My family, who keep me humble. And last but certainly not least, my mom, Diane, who always wanted me to be a doctor, but not this kind of doctor lol.

Publications

This thesis has been submitted for publication as:

Williams K, Bais A, Zhu W, Lin J, Lo CW. Gene-gene and protein-protein interactions in the ciliary landscape of congenital heart disease. Manuscript in prep.

Williams K, Carson J, Lo C. Genetics of Congenital Heart Disease. *Biomolecules*. 2019; 9(12):879. https://doi.org/10.3390/biom9120879.

Published non-thesis work includes:

Zhu W, Williams K, Young CB, Martin L, Benson DW, Lo CW. Common and rare variants associated with CoA indicate atrial functional dominance contributing to the developmental etiology of coarctation. *European Heart Journal*. Submitted.

Teekakirikul P, Zhu W, Gabriel GC, Young CB, Williams K, Martin LJ, Hill JC, Richards T, Billaud M, Phillippi JA, Wang J, Wu Y, Tan T, Devine W, Lin J, Bais AS, Klonowski J, Moreau de Bellaing A, Saini A, Wang MX, Emerel L, Salamacha N, Wyman SK, Lee C, Li HS, Miron A, Zhang J, Xing J, McNamara DM, Fung E, Kirshbom P, Mahle W, Kochilas LK, He Y, Garg V, White P, McBride KL, Benson DW, Gleason TG, Mital S, Lo CW. Common deletion variants causing protocadherin-α deficiency contribute to the complex genetics of BAV and left-sided congenital heart disease. *Human Genetics and Genomics Advances*. 2021; 2(3). https://doi.org/10.1016/j.xhgg.2021.100037.

1.0 Introduction

Congenital heart disease (CHD) is the most common birth defect[1], affecting about 1% of infants born each year[2,3]. While it remains a leading cause of infant mortality, early surgical intervention has led to greater survival and an increasing number of adults living with CHD[4]. These patients continue to deal with related cardiac issues such as arrhythmias and chronic heart failure, as well as extracardiac disorders including neurodevelopmental defects[5]. To address these rising health concerns, there is urgency to more fully understand the genetic causes of CHD. Knowledge of genetic factors and molecular mechanisms that cause CHD can help calculate the risk of having a child with CHD, facilitate early diagnosis at a finer level of detail than ultrasound, and develop possible interventions and therapeutics for children and adults living with disease.

A genetic underpinning for CHD is strongly supported by the observation of a high recurrence risk and familial forms of the disease[6,7], as well as the well-described association of CHD with chromosomal anomalies[8]. Recently, studies using whole exome sequencing data have allowed us to fully explore the underlying genetics of CHD. However, although our understanding of the causes of CHD have been greatly advanced by studies in animal models, such as mice, and human genetic studies, over half of cases remain unexplained[9,10]. The incomplete penetrance, variable expressivity, and genetic heterogeneity of CHD suggests that these unexplained cases may be explained by a combination of a complex, non-Mendelian genetic model and previously undiscovered causal loci, such as those in ciliary genes, which our lab recently discovered to be associated with cardiac defects[11].

In this dissertation, I explore the role of ciliary genes in human CHD under a complex genetic model of disease. In this first chapter, which has been adapted from Williams et al. 2019[12],

I provide background on CHD, evidence for a complex genetic model of disease, and the discovery of a central role for cilia in congenital heart disease pathogenesis. In chapter 2, I present results supporting the association of the ciliary genes with a complex genetic model of congenital heart disease. In chapter 3, I present results using the cilia-CHD protein-protein interactome to identify gene-gene interactions that may contribute to CHD. Finally, in chapter 4, I give brief conclusions.

1.1 Congenital heart disease

1.1.1 Developmental processes of the four-chambered heart

The heart is one of the first organs to develop during embryogenesis. When functioning properly, deoxygenated blood from the body flows into the right atrium and is pumped from the right ventricle to the lungs, and oxygenated blood from the lungs flows into the left atrium and is then pumped from the left ventricle out through the aorta to the rest of the body. A core group of transcription factors including *NKX2-5* and those from the GATA, T-box (TBX), forkhead box (FOX), and heart and neural crest-derived (HAND) families regulate cell differentiation and proliferation, although their specific downstream targets are not fully elucidated[13–15].

Studies in animal models have led to an understanding of heart development in the early embryo. The linear heart tube is comprised of cells from first heart field (FHF) that will give rise to the future LV and part of the atria[16]. In response to endoderm- and ectoderm-derived secreted factors such as bone morphogenetic proteins (BMPs), fibroblast growth factors (FGFs), as well as Wnt pathway signaling, embryonic precursors derived from the mesoderm give rise to cardiac progenitors in the cardiac crescent[17]. These cells migrate and fuse along the midline, generating

the linear heart tube. This is followed by looping of the heart tube, with the outer curvature of the looped heart tube forming the future ventricles, while the venous pole becomes the atrial appendages[16]. When the linear heart tube undergoes looping, bilateral symmetry is broken with the direction of looping reflecting the left-right body axis. This left-right patterning is of critical importance since the heart is one of the most left-right asymmetric organs in the body. Asymmetry is required for efficient oxygenation of blood, establishing circulation from the right side of the heart to circulate blood to the lungs for oxygenation, while the left side pumps oxygenated blood systemically throughout the body. Thus, when left-right patterning is disrupted, such as with randomization of visceral organ situs in heterotaxy, there is invariably complex CHD.

Cells from the second heart field (SHF) migrate into either pole of the linear heart tube, giving rise to the OFT, RV and part of the atria[16]. The conotruncal outflow undergoes septation to generate the aortic and pulmonary arteries. Neural crest cells migrating into the heart play a critical role in regulating outflow septation. Correct alignment of the outflows such that there is proper connection of the aorta with the left ventricle (LV) and pulmonary artery with the right ventricle (RV) is mediated by wedging of the outflows between the cardiac cushions such that there is "mitral to aortic valve continuity"[18].

Formation of the endocardial cushions[19] is mediated via epithelial-to-mesenchymal transition (EMT) of endocardial cells, which is regulated by Notch and vascular endothelial growth factor (VEGF) signaling. The cushions serve as primitive valves early in development, but later remodel to form the mature thin valve leaflets[16]. The atrioventricular (AV) valves are formed from superior and inferior AV cushions that later fuse with the growing muscular septa between the atria and ventricles. The outflow tract cushions give rise to semilunar valves of the aorta and

pulmonary trunk[20]. Cardiogenesis can be disrupted at many points, leading to structural defects that disrupt oxygenation and even lead to embryonic lethality.

1.1.2 Classification and prevalence of congenital heart disease

CHD encompasses a variety of heart defects that are commonly grouped based on the nature of the defect, resulting blood flow, patterns of observed familial recurrence risk, and shared susceptibility genes. Phenotypes are often sorted into major categories of right-sided lesions, left-sided lesions, conotruncal defects, laterality defects, and isolated septal defects[21]. Right-sided lesions include hypoplastic right heart syndrome (HRHS), Ebstein's anomaly, and pulmonary artery atresia. Left-sided lesions include bicuspid aortic valve (BAV), aortic stenosis, coarctation of the aorta (CoA), and hypoplastic left heart syndrome (HLHS). Conotruncal defects include tetralogy of Fallot (TOF), pulmonary atresia, truncus arteriosus, and double outlet right ventricle (DORV) except those with malposed vessels or HLHS. Laterality defects include heterotaxy (HTX), dextrocardia, situs inversus totalis (SIT), malposed vessels, and in many cases, transposition of the great arteries (TGA). Other defects include atrioventricular septal defects (AVSD) and anomalous pulmonary venous return (APVR). Isolated septal defects include atrial septal defects (ASD) and ventricular septal defects (VSD).

A meta-analysis of global birth prevalence of CHD showed that the 'mild lesions' ASD, VSD, and patent ductus arteriosus (PDA) account for 57.9% of CHD burden and that these mild lesions have driven a rise in CHD birth prevalence ~10% every 5 years since 1970[2]. This is likely primarily due to an increase in detection, with a small portion of the increased risk attributable to exposure to genetic or environmental factors. CHD associated with chromosomal abnormalities

represents ~8-10% of all CHD[8] and is believed to have a separate genetic etiology from non-syndromic disease, with a greater proportion driven by copy number and de novo variation[9,22].

1.2 Genetics of model of congenital heart disease

CHD has been shown to have a strong genetic component, with known environmental factors accounting for only ~10% of cases[10]. Investigations into the genetic etiology of CHD have been challenging given the genetic heterogeneity of the disease and the fact that CHD is often sporadic such that family history accounts for only 2.2% of overall CHD risk[23]. Nevertheless, analyses using a human sequencing data have revealed roles for hundreds of genes in the pathogenesis of congenital heart disease. In trio studies, the CHD proband is sequenced along with unaffected parents to identify pathogenic variants that may have arisen de novo. In familial studies, multiple members of a family are clinically assessed for CHD and sequenced to identify variants that are inherited in diseased family members. In cohort studies, a large number of unrelated CHD cases and healthy control samples undergo sequencing to determine if any single gene or set of genes is enriched for variants in the disease samples. Aside from known pathogenic variants in specific genes, studies of de novo and rare inherited variants have revealed a higher overall burden of mutation in variants predicted to be damaging in genes associated with CHD, highly expressed in the heart, or involved in heart development[10]. There is also a high burden of rare copy number variants in CHD patients, which is likely driven by syndromic cases[24].

Our understanding of the genetic causes of CHD has also benefited from studies in mouse models. Inbred mice provide an ideal context to conduct genetic analysis, and importantly, mice have the same four-chamber cardiac anatomy as humans that are susceptible to CHD

pathogenesis[25]. Given this, as well as the rapid advances in reverse genetics for generating gene knockouts, knock-ins, and point mutations, mice have become the model of choice to interrogate the genetic causes of CHD. These have allowed for the rapid verification of CHD candidate genes with disease modeling in vivo, along with in vitro cell and tissue culture studies. The recent use of patient-derived induced pluripotent stem cells (iPSCs) has become especially valuable for mechanistic studies. Using mice, it is also possible to interrogate the genetic etiology of CHD using forward genetics with chemical mutagenesis. Using such forward genetic screening methods with ethylnitrosourea (ENU) mutagenesis, our laboratory has identified over 100 genes causing CHD[11]. Forward genetic screens are advantageous in that they are entirely phenotype-driven, so there is no a priori gene bias, allowing the possibility for discovery of new biology.

1.2.1 Monogenic causes of congenital heart disease

A combination of familial and cohort studies have allowed the identification of many types of variants involved in human CHD pathogenesis. Many syndromes that are associated with CHD are caused by mutations in a single gene that has wide ranging functions during development, such as Kabuki syndrome caused by *KMT2D* or *KDM6A* mutations, CHARGE syndrome caused by *CHD7* mutations, Holt-Oram syndrome caused by *TBX5* mutations, and Alagille syndrome caused by *JAG1* or *NOTCH2* mutations[8]. Mutations in cardiac transcription factors such as *NKX2-5* and *TBX5* are also known to cause isolated CHD and are enriched for de novo and loss of function mutations[15,26]. Proteins with such deleterious mutations displayed changes in transcriptional or synergistic activity, which can interfere with expression of downstream targets, causing the perturbation of cell type specification, and differentiation[27]. Exome sequencing analysis with a trio study design revealed an important role for de novo variants in chromatin modifiers, such as

KMT2D and *KDM6A*, as contributors to diverse CHD phenotypes. The HDAC repressor complex, in particular, plays a key role in many developmental processes, and several proteins that are associated with this complex are associated with CHD[28]. Our lab's mutant mouse screen recovered CHD-causing mutation in *Sap130*, a Sin3A associated protein that is also part of the HDAC repressor complex, mediating left ventricular hypoplasia. The further identification of novel variants and CNVs has emerged from large cohort studies.

Mutations in signaling pathways necessary for heart development have also been associated with CHD. The heart is the most left-right asymmetric organ in the body, so an important signaling pathway in cardiovascular development is the Nodal signaling pathway, which is known to regulate left–right patterning. Nodal expression is restricted to the left side of the developing embryo and initiates a signaling cascade that establishes left–right asymmetry. In CHD patients, there is evidence of the enrichment of heterozygous damaging de novo and loss-of-function mutations in NODAL[26]. Signaling through the Notch pathway regulates cardiac cell fate and morphogenesis of cardiac chambers and valves[29]. Notch regulates EMT of the AV cushion progenitor cells which later contribute to the AV septum[30]. The Vegf signaling pathway is also required for formation of the AV endocardial cushions and their morphogenesis into AV valves[31]. The Wnt/β-catenin pathway has an important role in many different aspects of heart development, including the regulation of cell proliferation in the SHF[32]. The recovery of candidate CHD genes in the Wnt pathway was observed in patients with bicuspid aortic valve (BAV)[33]. Bmp signaling is required for specification and differentiation of the cardiac mesoderm and it regulates Nkx2-5 expression through a negative feedback loop[32,34]. The Ras/Mapk pathway, which regulates proliferation, growth, and other cell processes, is also known to play important roles in CHD. Thus, disruption of the Ras/Mapk pathway results in multiple related disorders collectively termed

RASopathies, the most common of which is Noonan syndrome. Noonan syndrome has the highest incidence of CHD, particularly pulmonary stenosis, among RASopathy patients[8].

Proteins that compose the sarcomere and extracellular matrix are also essential for proper structure and function of cardiac muscle, and mutations in these genes, such as *TTN*, *ELN*, *MYH6*, and *MYH7* are known to cause cardiac defects[8]. Genes that regulate splicing of essential cardiac genes, such as *RBM20*, are also known to cause CHD[42]. Cells must be able to respond and adhere to other cells and the extracellular matrix to maintain structure and transduce intracellular signaling. Our lab's mouse screen identified mutations in *Pcdha9*, encoding a protocadherin cell adhesion protein that is necessary for cell-cell recognition, was shown to have an essential role in valvular morphogenesis, as *Pcdha9* mutation cause bicuspid aortic valve (BAV)[43].

1.2.2 Complex genetic causes of congenital heart disease

Although animal models and familial studies have identified hundreds of genes to cause CHD, the majority of CHD cases cannot be explained when considering coding variants in these known genes under typical dominant or recessive genetic models[10], with a recent study observing known pathogenic small variation in just 5.8% of fetuses with CHD[9]. CHD also exhibits incomplete penetrance with non-Mendelian inheritance patterns within families[7], variable expressivity where one gene is associated with multiple phenotypes[44], and genetic heterogeneity where multiple genes are associated with one phenotype[43]. This suggests that genetic architecture of CHD may be more fully explained by a more complex, non-Mendelian genetic model[45].

That CHD may have a complex oligogenic etiology is suggested by various studies in animal models. One example is the observation that the incidence of CHD in mice heterozygous for the *Nkx2-5* mutations is dependent on genetic background[44]. Also indicative of multigenic

interactions contributing to CHD is our finding of significant enrichment for digenic combinations of mutations in the 100 cilia related CHD genes or cilia transduced cell signaling genes among our CHD mutant mice[11] (Table 1). As these additional mutations were not identified to be disease causing, their enrichment would indicate a secondary role in facilitating disease penetrance by the primary CHD causing mutation. Genes causing CHD recovered from the mouse mutagenesis screen were shown to be part of a tight protein-protein interaction (PPI) network. This PPI enrichment was unexpected given each CHD causing mutation was recovered independently in unrelated mutant lines. Another line of evidence for the complex genetics of CHD is the finding from our mouse screen showing hypoplastic left heart syndrome (HLHS) is a CHD lesion with a requisite multigenic etiology. Of the eight HLHS mouse lines recovered, half of them had mutations in more than one CHD-causing gene. For one HLHS mutant line, Ohia, we showed mutations in two genes, *Sap130* and *Pcdha9*, are required to cause CHD[43]. In fact, statistical modeling of left ventricular outflow tract obstructions in patients estimated that a digenic model of disease is the most likely over a monogenic or more complex genetic model[46].

Table 1. CHD mouse screen exhibits enrichment of CHD gene co-occurrence

Mutations[a]	CHD Screen (n=12,028) [d]	Non-CHD Screen (n=111,862) [d]	Odds Ratio
CHD Gene Mutations[a]	269	1,276	1.96 (1.71-2.24)
	P=1.56E-20		
Digenic Cilia-CHD Mutations[b]	30	101	2.8 (1.84-4.35)
	P=3.45E-06		
Digenic Cell Signaling-CHD Mutations[c]	16	42	3.30 (1.73-6.00)
	P=1.96E-04		

Comparison to immune disorder screen by Dr. Bruce Beutler at UT Southwestern. [a]Number of mutations in any of the 100 CHD genes; [b]CHD genes that are cilia related; [c]CHD genes that are cilia cell signaling related; [d]Number in parenthesis refer to total number of mutations recovered in mutagenized mice from CHD screen or non-CHD screen.

Human studies also point to the complex genetics of congenital heart disease. Chromosomal abnormalities were some of the first identified causes for CHD, and on average, they are estimated to account for 8-10% of cases[8]. Large variants such as these and copy number variants (CNV) affect multiple genes that may be functionally related, as functionally related genes tend to be closer together on chromosomes[47]. CHD is associated with chromosomal syndromes including Down syndrome (trisomy 21), trisomy 18, Williams syndrome, Turner's syndrome, DiGeorge syndrome, and others[8]. A recent study of genetic diagnoses in almost 1000 fetuses with severe CHD found chromosomal abnormalities in 23% and pathogenic copy number variation in 9.9% of fetuses[9].

Our lab recently identified two deletion CNVs, one common and one rare, in the protocadherin-α (PCDHA) cluster that are associated with LVOTO, including bicuspid aortic valve (BAV), coarctation of the aorta (CoA), and hypoplastic left heart syndrome (HLHS)[48]. Reduced expression of *Pcdha* in mutant mice, over 25% of which have BAV at E14.5, suggest that PCDHA haploinsufficiency can result in a maldeveloped aorta and/or aortic valve. These

combined findings suggest that this common deletion in the PCDHA cluster contributes to the high prevalence of BAV, while a secondary mutation is required to cause a rarer and more severe phenotype such as HLHS, as in the *Ohia* mutant mouse line.

Multiple recent studies have shown evidence of complex genetics on a finer grain level than large structural variants. One familial study of TGA identified one patient with mutations in both *ZIC3* and *FOXH1* and another patient with mutations in *ZIC3* and *NKX2-5* that was in the same family but with a different cardiac phenotype[49]. A more recent familial study showed rare heterozygous mutations in three genes, *MYH7*, *MLK2*, *NKX2-5*, are causative of childhood-onset cardiomyopathy[50]. A complex disease model was also suggested in a cohort study where four AVSD patients were observed to have damaging variants in multiple known CHD genes[51]. These studies suggest that in addition to multiple variants being required for disease, the specific genes and variants involved can contribute either modifying or protective effects on phenotype. "Gene-gene" or "digenic" interactions such as these are also referred to as epistatic interactions, where the effects of two variants in combination are statistically different from the additive effects of each variant alone[52].

Significantly, a meta-analysis of known CHD-associated loci from genetic and epidemiological studies showed that risk factors for disease converge in functional networks[53]. This makes sense particularly for an oligogenic model because genes that can cause disease in combination may serve redundant functions or participate in the same multi-protein complexes. Protein-protein interactions (PPIs) are intricately involved in biological functions and disease mechanisms, and it is likely that disturbances of this PPI network are propagated to cause disease.

Analyses of de novo and rare inherited variants conducted by the Pediatric Cardiac Genomics Consortium (PCGC)[54] further showed CHD-genes are enriched for cardiac transcription

factors, signaling pathway genes, ciliary genes, and chromatin modifiers, all of which are known to have multiple and varied roles in heart development[22,26,55,56]. The PPI networks and co-expression networks that encompass CHD genes were shown to have utility in identifying novel CHD candidate genes[53,57,58]. While these and other findings suggest genetic interactions can play a role in the complex genetics of CHD, no systematic study has been performed to investigate protein interaction networks of cilia related genes and genes in cardiovascular disease.

1.3 A central role for cilia in congenital heart disease

A central role for cilia in CHD pathogenesis was discovered by our lab via the use of forward genetics in mice with ENU mutagenesis to recover mutations causing CHD, as mentioned previously[11,59]. Cardiovascular phenotype was assessed using fetal echocardiography, a noninvasive high-throughput phenotyping method that is also highly sensitive for the detection of CHD and allowed the screening of 100,000 fetal mice. While the screen was entirely phenotype-driven, of 99 genes recovered with CHD-causing mutations, surprisingly 67 genes are related to cilia and ciliogenesis. Moreover, 37 of the cilia-related genes are involved in direct protein-protein interactions as part of the Ciliogenesis and PLANar cell polarity Effector (CPLANE) network[60] (Figure 1). The screen also recovered many genes involved in cilia-transduced cell signaling and vesicular trafficking, a cell process critical for ciliogenesis and cilia transduced cell signaling. This enrichment of ciliary genes encoding protein-protein interactors was unexpected given each mutation was recovered independently in unrelated mutant lines.

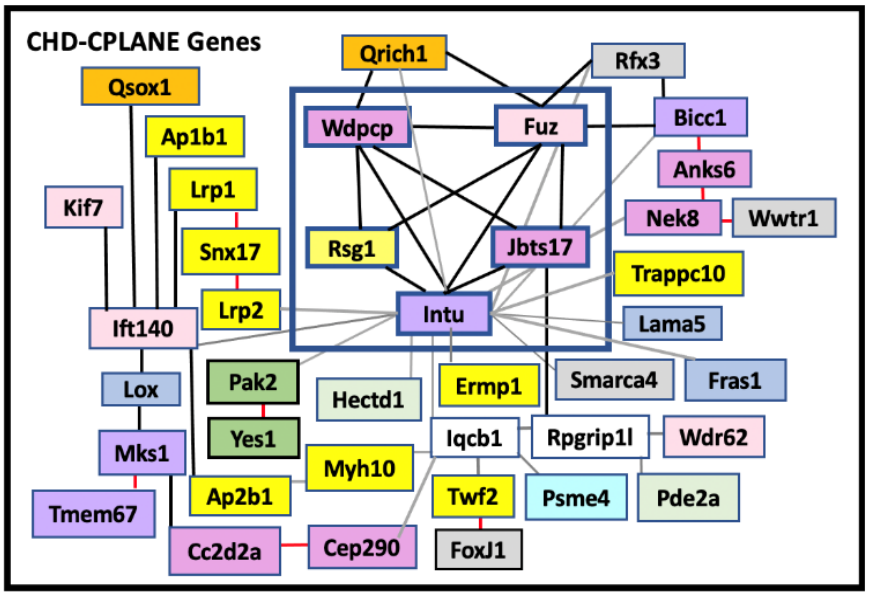

Figure 1. Mouse forward genetic screen recovered cilia related genes causing CHD in the CPLANE network. Black and red lines indicate direct interactors; Grey lines indicate that they exist in a complex. Node colors – grey: transcription, chromatin; orange: thioredoxin and glutamine rich; pink: cilia and ciliogenesis related; yellow: endocytic and vesicular trafficking; blue: extracellular matrix; dark pink: cilia transition zone; purple: basal body centrosome; green: ubiquitin ligase proteosome.

1.3.1 Cilia in the developing heart

Cilia are organelles that project from the cell surface and are found on most cell types. Some cilia are motile, generating extracellular flow or supporting cell motility, such as in the airway where they mediate mucociliary clearance. Other cilia, called primary cilia, are immotile and perform a variety of sensory, cell signaling, and cell cycle control functions. During early development, both motile and primary cilia are found in the embryonic node where they regulate specification of left-right asymmetry. Flow generated by motile cilia leads to activation of the Nodal signaling cascade which establishes left-right asymmetry[61,62]. Ciliary dysfunction at this crucial time can lead to laterality defects such as heterotaxy (randomized left-right patterning). As

the heart is the most left-right asymmetric organ in the body, mice and patients with heterotaxy almost invariably have lethal complex CHD.

Cilia are also found in the embryonic heart – in the epicardium, ventricular trabeculations, the endothelium, and the endocardial cushions that will give rise to the cardiac valves. Cilia are also known to regulate planar cell polarity through regulation of the cytoskeleton. This can affect cell migration behavior and epithelial-mesenchymal cell transitions (EMT). EMT plays an essential role in many of these processes, including deployment of neural crest cells, the epicardially derived cells, and the cardiac cushion mesenchyme. All these developmental processes can be impacted by cilia modulation of cell polarity and cytoskeletal organization, and cilia transduced cell signaling.

1.3.2 Ciliogenesis and planar cell polarity

To form the cilium, centrioles are modified to become the basal body which fuses to the cell membrane[63,64] (Figure 2). The ciliary body is separated from the rest of the cytoplasm by a transition zone, and intraflagellar transport (IFT) machinery moves proteins through the ciliary axoneme to extend the cilium. Vesicular trafficking is necessary to initially dock the basal body to the cell membrane through a ciliary vesicle during ciliogenesis and to move proteins into the ciliary membrane and maintain normal ciliary function (Figure 3)[65–67].

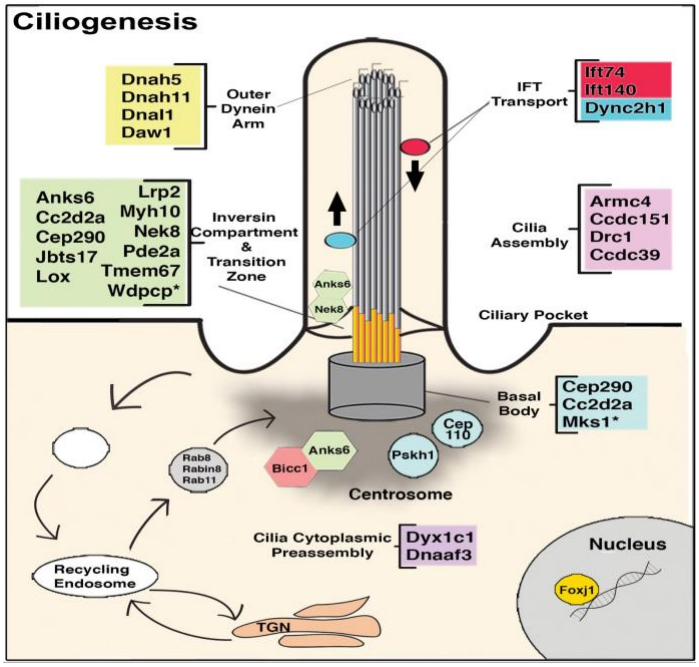

Figure 2. Ciliogenesis

Diagram illustrating the genes recovered from the CHD screen that are required for ciliogenesis. IFT, intraflagellar transport; TGN, trans-golgi network. From Li et al. 2015[11].

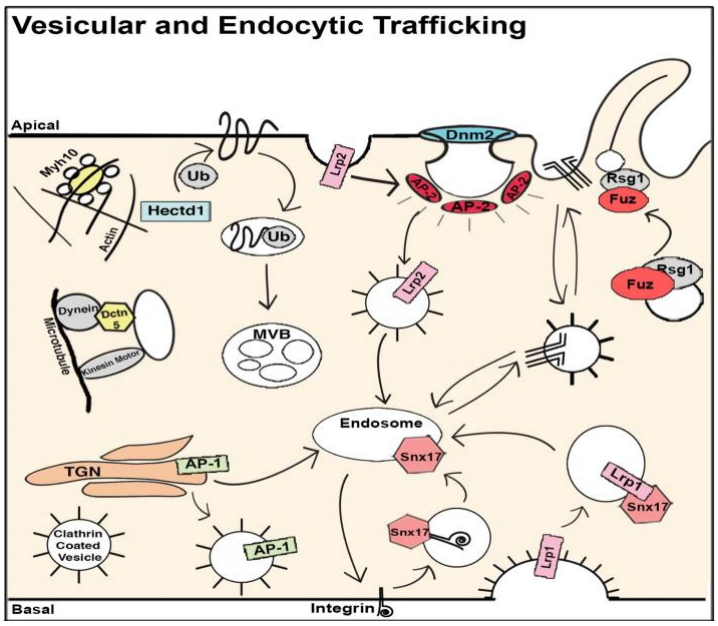

Figure 3. Vesicular and endocytic trafficking through the cilium

Diagram illustrating the biological context of ciliary genes in vesicular and endocytic trafficking. Highlighting denotes recovery from the CHD screen. AP, adaptor protein complex; MVB, multivesicular body; Ub, ubiquitination. From Li et al. 2015[11].

The forward genetic mouse screen in our lab recovered 4 genes that are core proteins in the CPLANE network to cause CHD, *Wdpcp*, *Fuz*, *Cplane2* (*Rsg1*), and *Cplane1* (*Jbts17*/*C5orf42*). These core genes are deeply conserved throughout evolution[60] and are linked to control of planar cell polarity and thought to recruit intraflagellar transport (IFT) proteins to the base of the cilium and insert complete IFT particles into the ciliary axoneme to elongate the cilium[60]. The mouse screen also identified mutations in *Dnah11*, an axonemal protein, and *Mks1*, a basal body protein, to be associated with CHD, which were also recovered in a separate mouse screen for CHD[41]. The ciliary gene *Ift88* is an intraflagellar transport protein required for cilia formation, and Ift88 null mutant mice exhibited OFT defects. Cilia have also been shown to play a role in aortic valve

disease, such as BAV. Defects in development of the atrioventricular cushions in *Cc2d2a* mutant mice were associated with loss of cilia from the AV cushions.

1.3.3 Cilia-transduced cell signaling pathways

Cell signaling pathways involved in heart development that are cilia-related include Nodal, sonic hedgehog (Shh), and transforming growth factor/bone morphogenesis protein (Tgfb/Bmp) pathways, which are cilia transduced, as well as the non-canonical Wnt pathway, which requires establishment of cell polarity by ciliogenesis and has signaling components that localize to the cilium (Figure 4)[68,69]. Developmental processes can be impacted by cilia and cilia transduced cell signaling further downstream through the modulation of cell polarity and cytoskeletal organization. Hence it is not surprising that ciliary defects can play a central role in the pathogenesis of CHD.

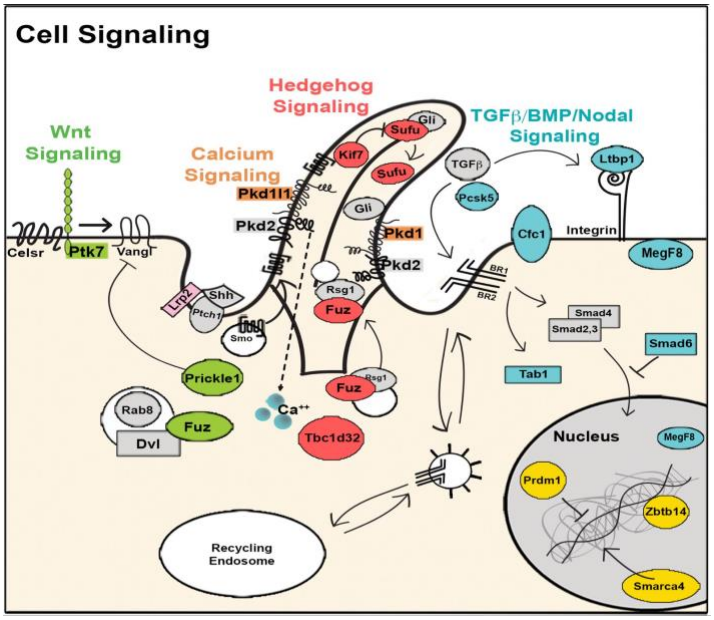

Figure 4. Cilia-transduced cell signaling pathways

Diagram illustrating the biological context of cilia in signaling pathways involved in heart development. Highlighting denotes recovery from the CHD screen. R, receptor. From Li et al. 2015[11].

Sonic hedgehog (SHH) signaling, in particular, is almost entirely transduced through the cilium (Figure 4). SHH signaling has been shown to play an important role in the development of the SHF, outflow tract septation, and proper outflow tract alignment[35,36]. SHH is secreted from the pharyngeal endoderm, and ligand is received by SHF cells, maintaining proliferation of these progenitor cells (Figure 5)[35]. GATA4 was shown to be required for proliferation of SHH-receiving cells and subsequent OFT alignment, and Gata4 mutations in mice cause DORV[37]. Signaling from BMP2 and BMP4 in the outflow tract myocardium, conversely, represses proliferation of SHH-receiving cells, with overexpression leading to premature differentiation of SHF cells and knockout resulting in embryonic lethality (Figure 5)[38]. SHH regulates development of *SIX2*+ progenitor cells, which contribute to the right ventricle, inflow tract, pulmonary trunk and ductus arteriosus[39]. Shh is required for migration of cardiac neural crest cells to the OFT cushion, with

Shh mutations in mice resulting in neural crest cell death and mislocalization (Figure 5)[37]. Shh is also required for atrioventricular septation, with mutations in the pathway causing atrioventricular septal defects in mice[40,41]. The role of SHH in human CHD has not been systematically examined, but the recovery of other regulators of SHH signaling among mutations causing CHD from our large-scale mouse mutagenesis screen would screen would suggest this pathway is likely to play an important role in human CHD.

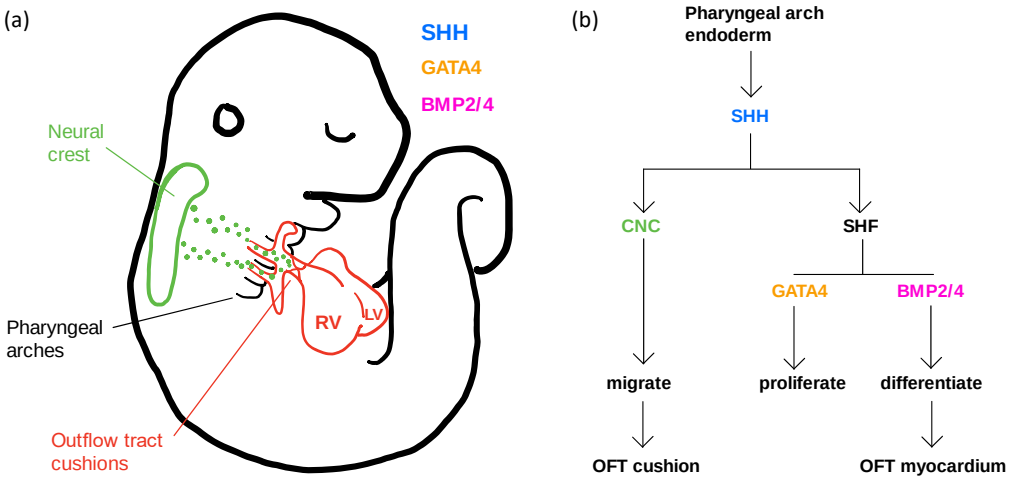

Figure 5. Sonic hedgehog in outflow tract development

SHH (blue) is secreted from the pharyngeal arch endoderm. SHH signaling mediates migration and localization of cardiac neural crest (CNC) cells (green) to the outflow tract (OFT) endocardial cushions (red). SHH-receiving cells expressing GATA4 (orange) proliferate in the SHF, and those receiving signals from BMP2/4 differentiate into OFT myocardium. From Williams et al. 2019[12].

1.3.4 Clinical relevance for cilia genes in congenital heart disease

Researchers are just beginning to appreciate that variants in ciliary genes may also play an important role in the pathogenesis of CHD in patients. Studies by the PCGC have shown an enrichment of genes related to cilia structure and function among those with loss-of-function (LoF) recessive genotypes in CHD probands[26]. They recently observed that 8.9% of CHD probands harbor a damaged ciliary gene and showed that cilia-related genes are enriched for rare, damaging recessive variation[70]. However, there have been no studies addressing the pathogenicity of inherited heterozygous mutations in ciliary genes or an oligogenic model of ciliary genes in human CHD pathogenesis.

As described above, our forward genetic mouse screen identified many ciliary genes that are protein-protein interactors, and a significant number of mouse lines from our forward genetic mouse screen contained heterozygous variants in secondary CHD-causing genes[11,43]. With regards to cilia biology, we note ciliogenesis and cilia transduced cell signaling are well described to involve large multiprotein complexes, such as the BBsome[71] and intraflageller transport (IFT) machinery[72,73]. These large multiprotein complexes regulate protein trafficking into and out of the cilia and play essential roles in cilia biology important for cardiovascular development and thus have significant impact on the pathogenesis of CHD. Cilia are necessary for many functions throughout development, and it is likely that redundancy among members of the ciliary network requires damage in multiple genes for strong effects to propagate throughout the network and cause disease. This is demonstrated in a recent study of ciliopathy, where it was shown that secondary variants in BBS complex genes contribute to Bardet-Biedl syndrome[74]. Thus, gene-gene interactions such as these within the ciliary interactome network may provide the genomic framework for the complex genetics of CHD.

2.0 Protein truncating variants in ciliary genes are associated with a complex genetic model of congenital heart disease

2.1 Background

Our lab's recessive forward genetic mouse screen previously identified a central role for cilia in the pathogenesis of congenital heart disease[11]. We identified 100 genes to cause CHD, and surprisingly, 69 of them were cilia-related, while others were related to cilia-transduced cell signaling pathways and vesicular trafficking, a core component of ciliogenesis. Studies by the PCGC in a cohort of human CHD patients have supported the involvement of cilia in human CHD, showing an enrichment of genes related to cilia structure and function among those with loss-of-function (LoF) recessive genotypes[70]. They also recently observed that 8.9% of CHD probands harbor a damaged ciliary gene and showed that cilia-related genes are enriched for rare, damaging recessive variation[70]. However, there have been no studies addressing the contribution of inherited heterozygous mutations in ciliary genes to CHD pathogenesis and whether ciliary variants may be pathogenic under a dominant or complex genetic model of disease.

It is widely believed that most congenital heart disease cases have a complex genetic etiology, requiring variants in multiple genes to cause disease[8,10,12,45,46,75,76]. CHD exhibits non-Mendelian inheritance patterns, and this is further complicated by variable disease penetrance such that not all individuals with a pathogenic mutation exhibit disease[7], variable expressivity where one gene is associated with multiple phenotypes[44], and genetic heterogeneity where multiple genes are associated with one phenotype[43]. A complex etiology is supported by familial studies indicating oligogenic inheritance[50] and clinical studies where patients have damaging variants in

multiple known CHD genes[51]. A digenic etiology, in particular, has been shown to have the best fit in statistical modeling of genetic linkage data obtained for a cohort of patients with left ventricular outflow tract obstructive (LVO) lesions[46].

There is evidence that genes associated with the cilium can cause CHD under a digenic model of disease. We noted that a significant number of the CHD mutant mice recovered in our screen harbored not only a single pathogenic mutation but were further enriched for additional secondary incidental mutations in other known CHD-cilia genes, supporting a complex genetic model of disease[11,43]. Importantly, this included the significant enrichment for digenic pairing of mutations in cilia-CHD genes (OR=2.8, p =3.3e-6) and in cilia transduced cell signaling genes (OR=3.5, p=2.0e-4). Moreover, a recent study showed an enrichment of patients with Bardet-Biedl syndrome, a ciliopathy, that have mutations in multiple disease-causing genes[74], supporting the notion that damage to multiple ciliary genes is necessary for disease pathogenesis. Here, we aimed to determine whether protein truncating variants, both homozygous and heterozygous, in ciliary genes are associated with human CHD and whether a digenic model of disease could explain this association of ciliary genes with CHD.

2.2 Methods

2.2.1 Whole exome sequencing

All data access requests and human studies were approved by Institutional Review Board of the University of Pittsburgh School of Medicine (STUDY20010180) and the Children's Hospital of Pittsburgh (CHP). We obtained written informed consent from all participants and/or

parents for children. Personal identities of the study participants were encrypted and secured in accordance with approved guidelines and regulations. This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided.

2.2.1.1 Germline variant calling

For 656 subjects recruited at the Children's Hospital of Pittsburgh UPMC (CHP), whole exome sequencing (WES) sequencing was carried out on Illumina HiSeq2000 with 100 paired-end reads at 80-100X coverage using Agilent V4 or V5 exome capture kit. For 2425 samples obtained from the Pediatric Cardiac Genomics Consortium[54] (PCGC, dbGaP phs001194.v2.p2) and 5140 healthy control samples obtained from the Alzheimer's Disease Sequencing Project[77] (ADSP, NG00067.v2) (with no personal or family history of dementia-related disease), SRA files were downloaded from the NCBI SRA database and converted to FASTQ files using sra-toolkit. BWA-MEM[78] was used to align reads in fastq files to human reference genome GRCh38. BAM files were further processed using GATK4 Best Practices workflows[79]. The intersection of the WES capture kit intervals used to sequence each cohort was taken, and single nucleotide variants (SNVs) and small indels (InDels) were detected individually using GATK HaplotypeCaller and jointly called using GATK GenotypeGVCFs. Further quality filtering was applied using bcftools 1.9 and qctool 2.0.6. High quality variants were recovered that: 1) have excess heterozygosity p-value > 3.4e-6; 2) passed GATK Variant Score Quality Recalibration (VSQR) with 99.95% sensitivity; 3) have genotype quality ≥ 20; 4) are SNVs or InDels not within 10bp or 5bp of an indel, respectively; 5) have missing rate < 10% and differential missingness p-value > 10e-6; and 6) have control HWE p-value > 10e-6 (Figure 6).

2.2.1.2 Variant annotation and filtering

Variants were annotated using Ensembl VEP[80] v102 with variant identifiers, gene symbol in NCBI RefSeq[81] v109, variant consequence of the most severely affected transcript, allele frequency in gnomAD[82] exomes v1.2.2, ClinVar[83] significance, and variant deleteriousness predictors such as SIFT[84] and PolyPhen[85]. Phred-scaled CADD[86] scores and GERP[87] RS scores were obtained from CADD v1.6. Protein truncating variants (PTV) were used for this analysis, as PTVs are expected to have the most severe impact on biological activity. That includes variants with a consequence of start loss, stop loss, frameshift insertion-deletion, splice acceptor site, and splice donor site.

2.2.1.3 Sample-level quality control

Samples with a FREEMIX[88] score of greater than 0.075 were considered contaminated and removed before filtering. Samples with missingness of greater than 10% and that are outliers in the number of variants present were removed before analysis. To remove pairs with cryptic relatedness, one sample was removed for each pair found to be related by pedigree or KING kinship analysis[89] (PLINK, cutoff=0.09375 for second degree relatives), and samples with 5 or more relationships were removed. Principal component analysis (PCA) was performed using genotypes of common variants with AF > 0.05 in Plink 1.9[90] to determine samples with European ancestry similar to CHP in PCGC and ADSP cohorts. 481 Pitt cases, 1451 PCGC cases, and 2602 ADSP controls passed sample-level filtering.

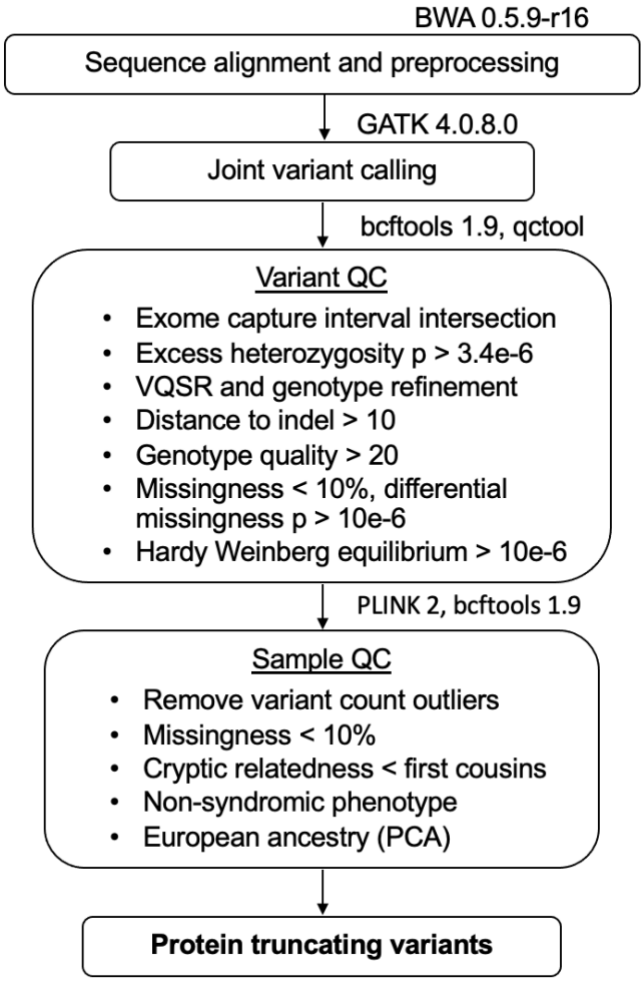

Figure 6. Germline variant calling and quality control pipeline

2.2.2 Cardiac phenotyping

CHD subjects were placed into mutually exclusive groups by major lesion. Cardiac phenotypes were divided into categories based on Botto et al. 2007[21]: conotruncal defect (CTD), left ventricular outflow tract obstruction (LVOTO), laterality defect, or Other. CTD phenotypes include truncus arteriosus, interruption of the aortic arch type B, conoventricular ventricular septal defects, Tetralogy of Fallot (TOF), and double-outlet right ventricle (DORV). LVOTO phenotypes

include hypoplastic left heart syndrome (HLHS), coarctation of the aorta (CoA), aortic stenosis (AS), bicuspid aortic valve (BAV), and interruption of the aortic arch type A (IAA-A). Laterality defects includes heterotaxy, dextrocardia, situs inversus, left or right isomerism (LAI, RAI), asplenia or polysplenia, l-transposition of the great arteries (L-TGA), and d-transposition of the great arteries (d-TGA). Other cardiac defects include atrioventricular septal defect, anomalous pulmonary venous return, pulmonary atresia, tricuspid atresia, pulmonary stenosis, Ebstein's anomaly, and isolated ventricular septal defect and atrial septal defect. Controls were included if they had no personal or family history of dementia-related disease. No cardiac phenotype information is available for the control cohort, but the subjects are primarily over 65[91], and in their infancy, CHD was largely not survivable[5].

2.2.3 Association analysis

2.2.3.1 Gene list curation

Genes were identified that are known to be associated with CHD from our forward genetic mouse screen[11], the PCGC[22,70], MGI[92], CHDGene[93,94], and the literature[12]. Of these, cilia-genes were identified from known ciliopathy genes[63], the CiliaCarta database[95], the PCGC[70], the CPLANE network[60], and the literature[60,96–101]. In total, 150 genes were considered to be cilia-CHD related.

2.2.3.2 Genome-wide gene level association analysis

We performed single-gene association analysis of case-control sequencing data using Multi-marker Analysis of GenoMic Annotation (MAGMA)[102], obtaining effect sizes and p-values for significance of each gene. MAGMA collapses variants into genes using principal component

analysis (PCA) and then performs linear regressions using the PCs as predictors for the phenotype, testing the null hypothesis that none of the PCs have an effect on phenotype. This method reduces the number of variables being tested and corrects for linkage disequilibrium (LD) between variants in a gene. The Benjamini-Hochberg correction was applied to account for the number of tests performed, setting the false discovery rate to 0.05.

2.2.3.3 Gene set association analysis

MAGMA[102] was used to perform competitive gene set analysis, which uses the regression framework to determine whether the joint effect of genes in a gene set is greater than that of other genes in the genome. To test for a monogenic model, the one-sided Fisher's exact test in R was used to determine whether there was an enrichment of cases compared to controls with only one gene containing a protein truncating variant in a gene set, as opposed to having zero, two, or three genes in the gene set containing a protein truncating variant. To test for a digenic model, the one-sided Fisher's exact test in R was used to determine whether there was an enrichment of cases compared to controls with protein truncating variants in two or more genes in the gene set as opposed to zero or one. To further validate significant results ($p<0.05$), permutation testing was used to assess significance of the association in relation to 10,000 random size-matched gene sets.

2.2.3.4 Enrichment analysis

Gene set enrichment analysis was performed using Metascape[103]. Protein-protein interactions (PPIs) were identified from StringDB[104]. Spatial transcriptomics data from fetal heart tissue of embryos at 4.5-5, 6.5, and 9 post-conception weeks were used to determine whether genes were expressed in the same cell types and/or heart regions[105].

2.3 Results

2.3.1 Cohort and variant selection

We analyzed whole exome sequencing (WES) data from 481 CHD patients from the Children's Hospital of Pittsburgh UPMC (Pitt), 1451 CHD patients from the Pediatric Cardiac Genomics Consortium (PCGC)[54], and 2602 controls from the Alzheimer's Disease Sequencing Project (ADSP)[77] with European ancestry (Table 2). To investigate phenotype-specific effects, CHD patients were grouped into subtypes including conotruncal defects (CTD), left ventricular outflow tract obstructions (LVOTO), and laterality defects[21]. To reduce noise introduced by variants that are not damaging, variants were filtered for protein-truncating variants (PTVs), as these are most likely to lead to a functional effect on the protein.

Table 2. Summary of study participants and phenotypes

	Pitt (% Male)	PCGC (% Male)	Total (% Male)
CTD	104 (59.6)	308 (57.5)	412 (58.0)
LVOTO	144 (74.3)	501 (66.9)	645 (68.5)
Laterality	152 (62.5)	245 (66.9)	397 (65.2)
Other	81 (56.8)	397 (49.6)	478 (50.8)
Total CHD	**481 (64.4)**	**1451 (60.2)**	**1932 (61.2)**
Total Control			**2602 (40.9)**

2.3.2 Gene-level analysis of protein truncating variants

First, we wanted to determine if PTVs in any individual cilia-CHD genes are significantly associated with the either the CHD cohort overall or CHD subtypes compared to controls. We performed genome-wide, gene-level association analysis using MAGMA[102] and looked to see if the list of statistically significant genes contained any known cilia-related genes in the CiliaCarta database[95], the CPLANE interactome network[60], curated PCGC cilia gene list[70], or known ciliopathy gene list[63]. Although no cilia-related genes were statistically significant after multiple testing correction when considering all CHD cases together, further analysis revealed that PTVs in multiple cilia-related genes are associated with specific subtypes. *NOTCH1* is significantly associated with LVOTO (MAGMA p=1.27e-5, Table 3), and this association of *NOTCH1* with LVOTO has been previously shown for predicted LoF and intronic variants[106], as well as an association of rare or likely pathogenic variants with BAV requiring aortic root replacement[107] and of de novo and rare variants with HLHS[108,109]. Cilia genes *CELSR3* and *C20orf85*, as well as two cilia genes previously shown to cause laterality defects, *CCDC103*[110] and *DNAAF4*[111], are significantly associated with CTD (Table 3). Five cilia genes (*CELSR3*, *CUX1*, *IFT81*, *DYNC2I2*, *CELSR2*) are associated with laterality defects (Table 3).

We performed gene set enrichment analysis using Metascape to determine if this is a significant enrichment of cilia genes among those that are statistically significant and competitive gene set analysis in MAGMA to compare the effect sizes of the cilia-CHD gene set to the rest of the genome. Neither test showed that these cilia-related genes with low p-values are statistically significant in cases compared to controls, which suggests that individually, cilia-related genes have relatively small contributions to CHD. However, neither of these methods considers possible interactions between variants, where a variant in more than one gene is required to cause disease.

Gene set enrichment analysis specifically of the significantly associated cilia genes in Table 3 using Metascape[103] show that these are enriched for the GO biological process cilium assembly (P = 3.63e-10.

Table 3. Significantly associated genes that are cilia-related

Phenotype	Gene	pLI score	N SNPs	Z-stat	MAGMA p	Source
LVOTO	NOTCH1	1	9	4.2121	1.265200e-05	Watkins
CTD	CELSR3	0.99984	4	4.9036	4.704400e-07	CiliaCarta
CTD	C20orf85	1.94e-5	2	4.1400	1.736600e-05	CiliaCarta
CTD	CCDC103	0.00448	1	3.4925	2.392900e-04	CiliaCarta
CTD	DNAAF4	6.88e-11	2	3.3710	3.744600e-04	CiliaCarta
Laterality	CUX1	1	4	5.0008	2.855200e-07	CPLANE
Laterality	CELSR3	0.99984	3	4.2904	8.916400e-06	CiliaCarta
Laterality	IFT81	5.86e-12	2	3.4409	2.898500e-04	CiliaCarta
Laterality	DYNC2I2	0.0028131	2	3.4409	2.898500e-04	CiliaCarta
Laterality	CELSR2	0.99983	2	3.4409	2.898500e-04	CiliaCarta

2.3.3 Presence of protein truncating variants in cilia-CHD genes

Next, we sought to determine whether cilia-related genes as a whole are associated with CHD under a more complex genetic model of disease. To investigate this question of genetic model, we identified 150 genes that are already known to cause CHD and are associated with the cilium, which comprise the cilia-CHD gene set (Table 5). Of 667 known CHD genes curated, this is a significant enrichment of cilia-related genes among CHD causing genes (p=1.74e-7). To determine the portion of CHD cases to which cilia-CHD PTVs can contribute, we looked at the distribution of PTVs per individual (Figure 7). 19.3% of CHD cases have at least one PTV in one

of these genes, and 62.7% of these genes have variants in the CHD case cohort. When considering each subtype separately, 21.4% of CTD patients, 17.7% of LVOTO patients, and 21.7% of laterality defect patients have at least one PTV in a cilia-CHD gene. In the control cohort, 16.4% have at least one PTV in a cilia-CHD gene, and 53.3% of cilia-CHD genes have a variant in a control subject (Figure 7). This is an enrichment of PTV genotypes in cilia-CHD genes among cases compared to controls (Fisher's exact p=0.0029). The greater number of cilia-CHD genes with PTVs in the CHD case cohort is surprising given that the sample size is smaller than the control cohort.

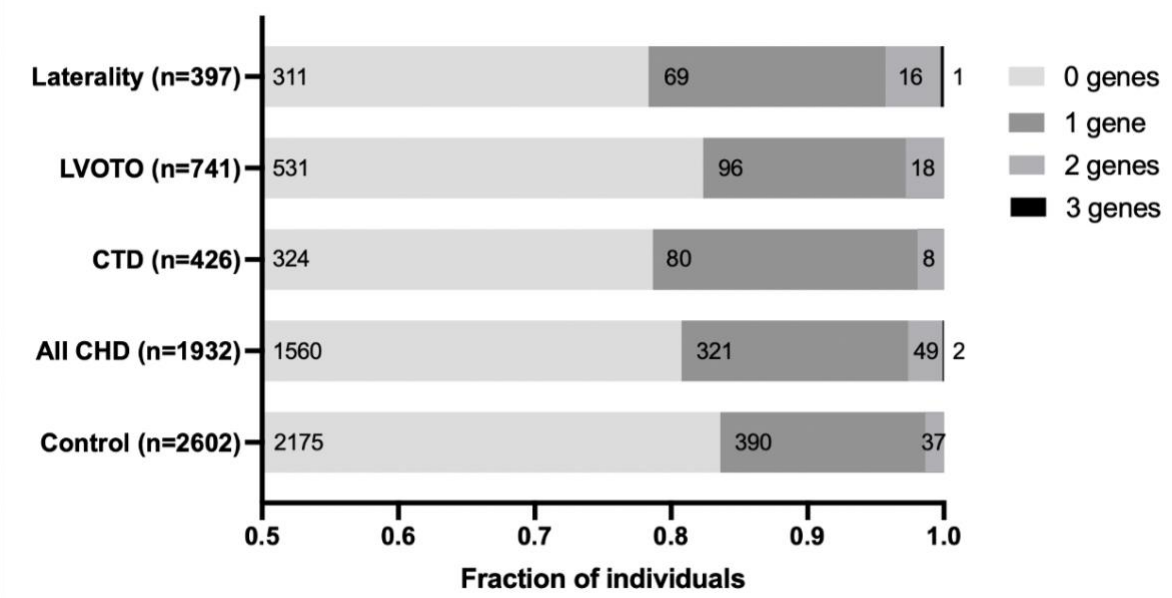

Figure 7. Distribution of individuals with cilia-CHD PTVs

Of the 150 cilia-CHD genes, 30 have PTVs in only cases, and these are enriched for genes that are intolerant to loss-of-function[82] (7 genes with gnomAD pLI score > 0.9, p=0.03) and not enriched for genes that are intolerant to gain of function[112] (Fisher's exact OR=1.80, p=0.22). Only 16 genes have PTVs in only controls, and none of these have a pLI score > 0.9. Interestingly, 5

LAT patients have multiple PTVs in one cilia-CHD gene (Fisher's exact OR=8.28, p=0.0032, Table 4). Although given the lack of parent data it is not possible to distinguish compound heterozygous genotypes, this observation aligns with previous studies suggesting that a recessive model of disease may be more appropriate for laterality phenotypes[70]. There are 31 CHD patients with known pathogenic or likely pathogenic variants in cilia-CHD genes, but this is not a significant enrichment over controls (Fisher's exact OR=1.45, p=0.19), and 143 patients have variants with a consequence that leads to nonsense mediated decay, as annotated by Ensembl VEP[80], but this is not significantly different than controls (Fisher's exact OR=0.96, p=0.73).

Table 4. Recessive genotypes in patients with laterality defects

Sample ID	Variant	GT	gnomAD AF	CADD Phred	Gene	Consequence	Phenotype
PITT_7160	chr5:13865675:G:A	1/1	3.99e-6	48	DNAH5	Stop gain	SIT, PCD
PITT_7469	chr5:13721014:G:A	0/1	.	41	DNAH5	Stop gain	SIT
PITT_7469	chr5:13727638:TTTTCTCA:T	0/1	.	35	DNAH5	Frameshift	SIT
PITT_7713	chr15:56446886:CCTTT:C	1/1	.	28.5	MNS1	Frameshift	SIT, DORV
PITT_7751	chr14:58432438:AG:A	0/1	3.05e-3	33	KIAA0586	Stop gain	TGA
PITT_7751	chr14:58474727:C:T	0/1		38	KIAA0586	Stop gain	TGA
PCGC_7397	chr20:62311007:CA:C	0/1	.	33	LAMA5	Frameshift	HTX
PCGC_7397	chr20:62332701:TG:T	0/1	.	30	LAMA5	Frameshift	HTX

Variants in red are known pathogenic variants causing ciliary dyskinesia in ClinVar.

2.3.4 Testing a complex genetic model of disease

Next, we wanted to determine if there is an enrichment of PTVs in cases compared to controls under both a monogenic and a digenic model of disease. There is a significant enrichment of all CTD patients (Fisher's exact OR=1.40, p=0.01194) with a PTV in only one cilia-CHD gene compared to controls (Figure 8a). There is a significant enrichment of cases with a PTV in two or

more cilia-CHD genes when considering all CHD patients (Fisher's exact OR=1.88, p=0.004396), LVOTO patients (Fisher's exact OR=1.99, p=0.02496), and laterality patients (Fisher's exact OR=3.10, p=0.0003633) (Figure 8b). Considering that the proportion of males in the CHD case cohort (61.2%) is significantly greater than that of the control cohort (40.9%, p=2.2e-16), we also performed this analysis stratified by gender. Many of the results were not statistically significant after stratifying by gender because of the much smaller sample size, but significant associations include males with CTD being enriched for PTVs in only one cilia-CHD gene (Fisher's exact OR=1.62, p=0.009497, Figure 8a), males with laterality defects being enriched for PTVs in two or more cilia-CHD genes (Fisher's exact OR=2.82, p= 0.008658, Figure 8b), and females being enriched for PTVs in two or more cilia-CHD genes when considering the entire CHD cohort (Fisher's exact OR=2.08, p=0.03467, Figure 8b). These results are driven by heterozygous genotypes, as only 2 cases and 4 controls have homozygous PTVs in cilia-CHD genes, none of whom carry PTVs in a second gene.

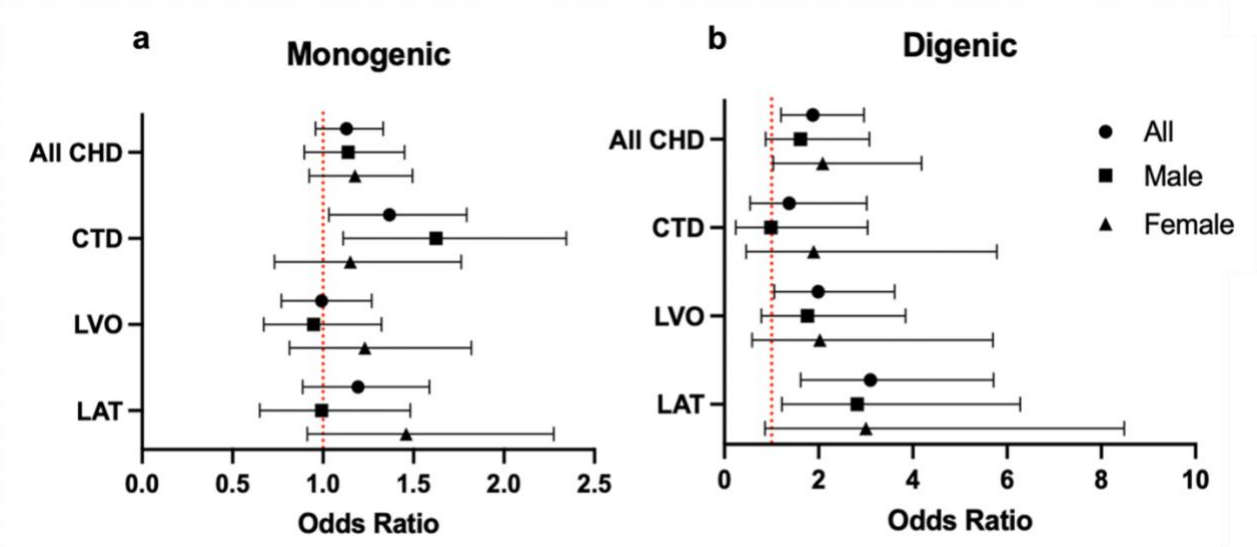

Figure 8. Enrichment of patients with one vs. two or more protein truncating variants in cilia-CHD genes

2.3.5 Network of genes with protein truncating variants co-occurring in CHD patients

We then sought to identify the specific gene pairs that co-occur in patients and contribute to this enrichment of digenic pairs in CHD cases. There are 51 patients with PTVs in two or more cilia-CHD genes, comprising an interaction network of 50 genes in total (Figure 9, Table 6). Although the sample size is too small to assess statistical significance of each gene pair individually, all of these co-occurring pairs are exclusive to CHD patients and not found in controls. Evaluation using StringDB shows this network of co-occurring genes is enriched for high confidence functional interactions ($p<1e-16$) and physical protein-protein interactions (PPIs, $p<1e-16$). While none of the control-only pairs share common neighbors in the PPI network, 11 of the case-pairs share common neighbors, or "linker" genes, in the PPI network (Figure 9). We also assessed whether genes with co-occurring PTVs are co-expressed in the fetal heart based on spatial transcriptomics data[105], and 21 out of 50 gene-gene pairs found only in CHD cases are co-

expressed within the same region and/or cell type of the developing heart, compared to 9 of 27 pairs found only in controls.

In addition, gene set enrichment analysis using Metascape shows that these co-occurring genes are enriched for genes in the smoothened signaling pathway (p=1.91e-21), which is the same as the cilia-transduced cell signaling sonic hedgehog (SHH) pathway, as well as overlapping sets of genes related to primary cilium development (p=2.57e-23), non-motile cilium development (p=2.40e-10), intraciliary transport (p=7.59e-7), and the centrosome cycle (p=0.0017, Figure 10). Further, six genes (*CNTRL*, *MKS1*, *C2CD3*, *CC2D2A*, *TCTN2*, and *TCTN3*) are direct protein-protein interactors in a complex involved in anchoring of the basal body to the cell membrane (Metascape p=1.58e-15). Interestingly, although there is an enrichment of genes intolerant to loss-of-function among those with PTVs in cases-only, there is a significant depletion of genes with pLI > 0.9 among genes in the case-only co-occurrence network (Fisher's exact OR=0.0438, p=2.92e-9). Conversely, there is a significant enrichment of genes that are predicted to be intolerant to gain of function (Fisher's exact OR=2.26, p=0.029).

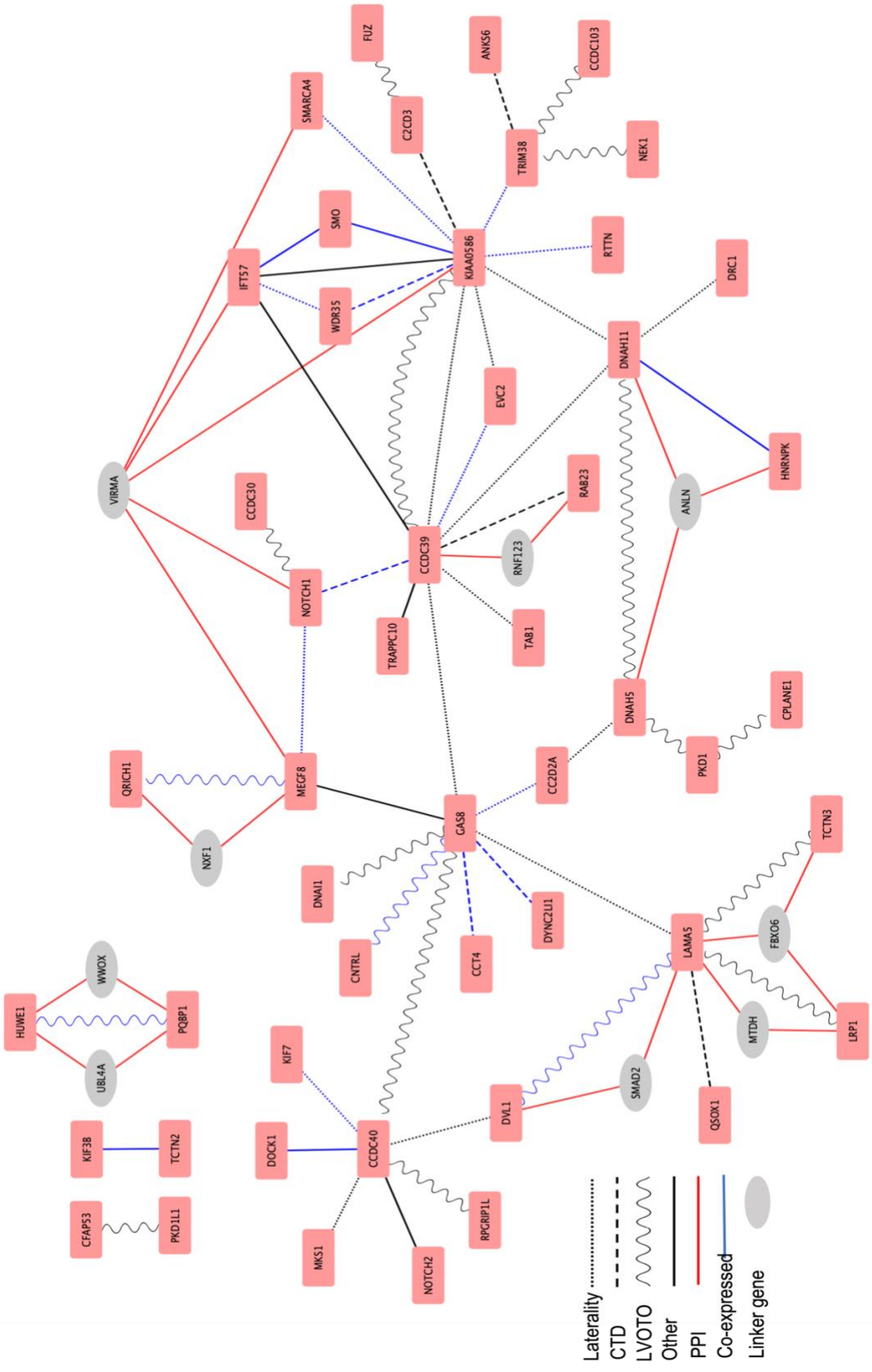
FUZ
ANKS6
CCDC103
SMARCA4
C2CD3
TRIM38
NEK1
SMO
KIAA0586
RTTN
IFT57
WDR35
DRC1
DNAH11
EVC2
VIRMA
HNRNPK
CCDC30
RAB23
ANLN
CCDC39
RNF123
NOTCH1
TRAPPC10
TA81
DNAH5
CPLANE1
QRICH1
MEGF8
CC2D2A
PKD1
CAS8
NXF1
TCTN3
DNAI1
FBXO6
DYNC2LI1
LAMA5
CNTRL
CCT4
LRP1
MTDH
WWOX
HUWE1
PQBP1
KIF7
UBL4A
SMAD2
DVL1
KIF38
TCTN2
DOCK1
QSOX1
CCDC40
CFAP53
PKD1L1
RPGRIP1L
MKS1
NOTCH2
Laterality
CTD
LVOTO
Other
PPI
Co-expressed
Linker gene

Figure 9. Digenic PTV network

Red nodes are cilia-CHD genes. Grey nodes are linker genes. Black lines indicate genes that co-occur in a patient. Blue lines indicate a co-expressed gene pair that co-occurs in a patient. Line type indicates phenotype group of the patient that the gene pair co-occurs in: Dotted – laterality defect, Dashed – conotruncal defect (CTD), Wavy – left ventricular outflow tract defect (LVOTO). Red lines indicate protein-protein interactions.

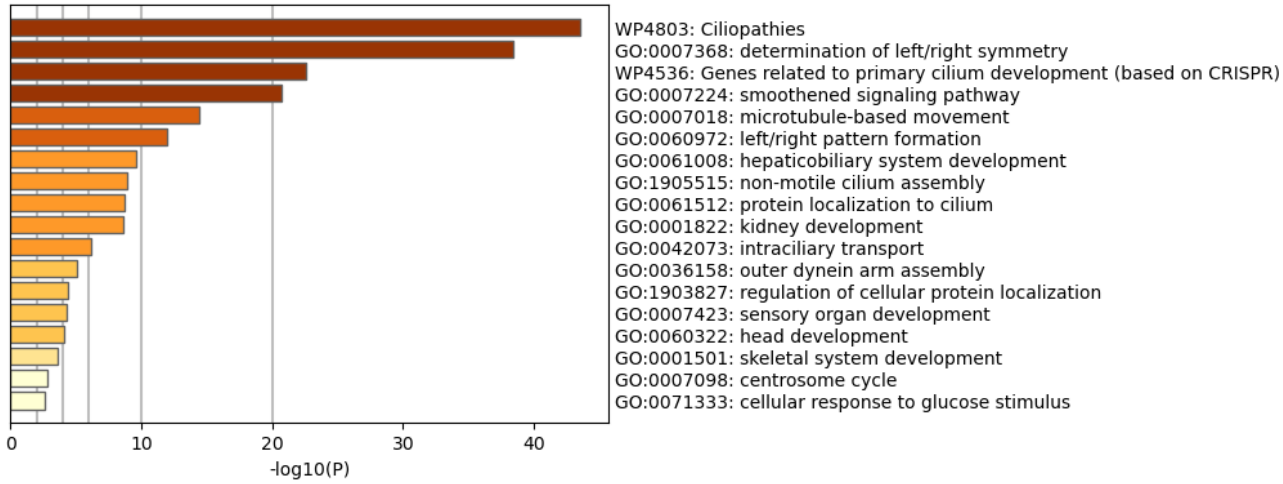

Figure 10. Digenic PTV network gene set enrichment

Gene set enrichment analysis and visualization performed in Metascape.

2.4 Discussion

Recent studies in mouse models and human cohorts have observed an association of cilia-related genes with congenital heart disease. However, previous studies of ciliary variants in CHD patients have focused on de novo variants or recessive genotypes. Here, we incorporate heterozygous genotypes into our association analysis of ciliary variants with CHD and investigate a complex model of disease, where variants in two ciliary genes are contributing to disease.

We show that in the context of protein truncating variants in ciliary genes, conotruncal defects follow a dominant monogenic model needing only a heterozygous variant in one gene to cause disease, laterality defects follow a recessive monogenic or digenic model needing two hits in either the same or different genes to cause disease, and left ventricular outflow tract obstructions follow a dominant digenic model of disease, needing heterozygous variants in at least two genes to cause disease. This is consistent with our previous findings in the *Ohia* mutant mouse line, where mutations in two genes, *Sap130* and *Pcdha9*, are needed to cause the severe LVOTO phenotype HLHS[43]. Overall, a significant proportion of CHD patients have PTVs in more than one cilia-CHD gene, 2.7% of CHD patients compared to 1.2% of controls. Although this digenic model of disease involving ciliary genes may explain only a small portion of cases, a monogenic model involving ciliary genes, such as shown here in CTD and laterality phenotypes, and a complex genetic model of disease involving non-cilia genes may explain a larger number of CHD cases.

When considering genes in which PTVs co-occur in CHD patients but not controls, these genes are associated with primary cilium development, giving context for the type of cilia genes that are pathogenic under a complex genetic model. Many of the genes in this network are either direct PPIs or part of gene pairs that share neighbors in the PPI network or are co-expressed within the fetal heart. This provides a broader biological context for their interaction leading to pathogenesis and suggests that the broader PPI network can help to identify novel candidate CHD genes that cause disease only in the context of other mutations.

There are two patients with PTVs in both *KIAA0586*, which is required for localization of the centrosome and cell polarity[113], and *CCDC39*, which is required for ciliary motility[110], one with HLHS and another with d-TGA and LVOTO. The d-TGA patient also has a PTV in *EVC2*, a positive regulator of the SHH pathway that localizes to the ciliary transition zone. This suggests

that in this case variants in two genes, *KIAA0586* and *CCDC39*, are required to cause CHD, while the variant in *EVC2* modifies phenotype. A separate patient with AVSD has PTVs in three genes, *KIAA0586*, *IFT57*, and *SMO*, all of which are members of the cilia-transduced SHH signaling pathway which is required for atrioventricular septation[40]. The six co-occurring gene pairs that are co-expressed in the fetal heart and share linker genes in the PPI network (*HUWE1-PQBP1*, *DVL1-LAMA5*, *MEGF8-NOTCH1*, *QRICH1-MEGF8*, *DNAH11-HNRNPK*, and *KIAA0586-SMARCA4*) also warrant further investigation. These genes that are close together in the PPI network are likely to lead to more severe perturbations when they are both disturbed, as with known pathogenic digenic interactions where 60% are also separated by just one gene in the PPI network[114].

The variants considered in this analysis are all predicted to be PTVs, and depending on the location they can lead to loss of function (LoF), gain of function (GoF), or no effect on protein function if the variant is near the end of the protein's coding region. Interestingly, the network of co-occurring genes is depleted of heterozygous LoF intolerant genes, indicates that these LoF intolerant genes cause disease on their own and do not need hits in a secondary gene to cause disease. Conversely, for LoF tolerant genes in this network, haploinsufficiency of one gene may be tolerated, but haploinsufficiency of two genes causes enough perturbation in the biological network to lead to disease. The enrichment of GoF intolerant genes in this network suggests that a buildup of protein causing errant functions and interactions may be tolerated within heart development but is not well tolerated when perturbations occur at multiple places in the biological network.

3.0 CHD-associated gene-gene interactions in the ciliary interactome

3.1 Background

Studies in mouse and human have shown that many cases of CHD are likely driven by a more complex genetic model of disease[44,45,50]. Computational modeling in a cohort of LVOTO patients showed that a digenic model, where variants in two genes contribute to disease, is the most likely[46]. We showed that in the context of 150 known cilia genes, a digenic model is associated with CHD patients, particularly LVOTO and laterality defect phenotypes. There are likely many other genes, both cilia and non-cilia that contribute to the complex genetic model of CHD. However, combinatorial genome-wide analysis for association with CHD, even considering only two genes at a time, is infeasible due to the high demands in terms of computational resources and sample size of both CHD cases and controls.

Focusing on genes that are connected in the protein-protein interaction (PPI) network is one way to make this type of analysis possible. Most genes that have been associated with CHD converge into functional biological networks[53]. When these networks are perturbed by genetic variation, this can disrupt biological processes and lead to disease[115]. These perturbations can also be propagated through the network, having effects outside of their immediate function and interactions. The disruption of multiple genes is more likely to cause a major perturbation in tightly connected biological networks, even if the genes are not direct interactors, particularly if the genes serve redundant functions or are co-members of a multi-protein complex[116]. The multigenic etiology of congenital heart disease may arise in part from scenarios such as this where damage to multiple genes must propagate through the network to cause disease.

The importance of PPIs in CHD has been strongly supported by findings from network analysis of the genes recovered from our large-scale mouse forward genetic screen[11]. The interactome network revealed close connections of not only CHD-associated cilia genes but also direct PPIs between proteins encoded by CHD-associated genes and additional cilia genes. Many of the proteins encoded by the genes harboring CHD mutations recovered from the screen in fact showed direct PPIs that were interconnected within a tight network comprising the Ciliogenesis and PLANar cell polarity Effector (CPLANE) interactome[60]. This PPI network includes many genes already known to play a role in a broad range of ciliopathies such as Joubert syndrome, Meckel syndrome, Jeune and Sensenbrenner syndrome, and others. These findings suggest PPIs play an important role in CHD pathogenesis, with the cilia-CHD interactome providing the framework for investigating the role of cilia in mediating the complex genetics of CHD. Although protein-protein interaction network analysis has been used to characterize and prioritize results of association analyses, only one recent study has used this network as the starting point to identify novel genetic factors associated with CHD[117], and this was limited to interactors of the cardiac transcription factors GATA4 and TBX5. Here, we leveraged computational analysis of the ciliary PPI network to identify novel candidate genes underlying CHD pathogenesis.

3.2 Methods

3.2.1 Whole exome sequencing

All data access requests and human studies were approved by Institutional Review Board of the University of Pittsburgh School of Medicine and the Children's Hospital of Pittsburgh

(CHP). We obtained written informed consent from all participants and/or parents for children. Personal identities of the study participants were encrypted and secured in accordance with approved guidelines and regulations. This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided.

3.2.1.1 Germline variant calling

For 656 subjects recruited at the Children's Hospital of Pittsburgh UPMC (CHP), whole exome sequencing (WES) sequencing was carried out on Illumina HiSeq2000 with 100 paired-end reads at 80-100X coverage using Agilent V4 or V5 exome capture kit. For 2425 samples obtained from the Pediatric Cardiac Genomics Consortium (PCGC, dbGaP phs001194.v2.p2) and 5140 healthy control samples obtained from the Alzheimer's Disease Sequencing Project (ADSP) (with no personal or family history of dementia-related disease), SRA files were downloaded from the NCBI SRA database and converted to FASTQ files using sra-toolkit. BWA-MEM was used to align reads in fastq files to human reference genome GRCh38. Processing and quality control filtering was applied using GATK4 Best Practices workflows as described in section 2.2.1 above.

3.2.1.2 Annotation and filtering

Variants were annotated using Ensembl VEP v102 and CADD v1.6. Protein truncating variants (PTV) were used for this analysis, as PTVs are expected to have the most severe impact on biological activity. That includes variants with a consequence of start loss, stop loss, frameshift insertion-deletion, splice acceptor site, and splice donor site. Samples of European ancestry that pass quality control filtering as described in section 2.2.1.3 above were retained. 481 Pitt cases, 1451 PCGC cases, and 2602 ADSP controls passed sample-level filtering.

3.2.2 Network analysis

3.2.2.1 Gene list curation

Genes were identified that are known to be associated with CHD from our forward genetic mouse screen[11], the PCGC[22,70], MGI[92], CHDGene[93,94], and the literature[12]. Of these, cilia-genes were identified from known ciliopathy genes[63], the CiliaCarta database[95], the PCGC[70], the CPLANE network[60], and the literature[60,96–101]. In total, 150 genes were considered to be cilia-CHD related.

3.2.2.2 Genome-wide gene level association analysis

We performed single-gene association analysis of case-control sequencing data using Multi-marker Analysis of GenoMic Annotation (MAGMA)[102] as described in section 2.2.3.2 and gene set enrichment analysis using Metascape[103].

3.2.2.3 Protein-protein interaction network

We curated first-degree protein-protein interactions (PPI) with the cilia-CHD gene list using known physical interactions in human curated in BioGRID[118], which includes both direct biophysical protein interactions and protein complex co-membership. Genes were retained that had at least one variant in the WES data. Spatial transcriptomics data from fetal heart tissue of embryos at 4.5-5, 6.5, and 9 post-conception weeks were used to determine whether genes were expressed in the same cell types and/or heart regions[105]. Genes were considered co-expressed if they were expressed in the same ‘spot’ at least once.

3.2.2.4 Conditional correlation analysis

For cases and controls separately, using binary data indicating the presence or absence of a protein truncating variant in each gene in the cilia-CHD protein-protein interactome, the conditional correlation of all gene pairs was calculated under the truncated Poisson graphical model using the R package ModPGMInference[119]. This calculates the correlation of each pair of genes independent of the effects of other genes. This method is computationally fast, statistically efficient under the weakest possible sparsity assumption, robust to small *n*, large *p* scenarios, and provides a confidence interval, z-score, and p-value for the effect estimate of each interaction. We selected for those gene pairs that have statistically significant correlations in the CHD cases but not in the controls. The Benjamini-Hochberg correction was used to limit the false discovery rate to 0.05. This network of genes with PTVs that are correlated in patients is referred to as the gene-gene interaction (GGI) network.

3.2.2.5 Gene pair similarity measures

For each digenic pair in the GGI network that is statistically significant based on the conditional correlation analysis, we used protein-protein interaction network features and co-expression to prioritize the pairs. Spatial transcriptomic data was used to determine whether the digenic pair is co-expressed in the developing heart as described above.

3.2.2.6 Community Clustering

We used the Girvan-Newman community clustering algorithm[120] implemented in NetworkX v2.6.2 to identify densely connected regions of the GGI network. This method iteratively removes edges with the highest betweenness to optimize the modularity of the network, where there are dense connections within clusters and sparse connections between clusters. When

calculating modularity, each edge was weighted by the z-score of the correlation. Clusters were ordered by number of nodes and edges from largest to smallest for consistent naming. We used the Fisher's exact test to test for enrichment of known CHD genes or genes significantly associated with our case cohort in each cluster.

3.3 Results

3.3.1 Protein-protein interaction network creation

First, we created a protein-protein interaction (PPI) network which could be searched for novel CHD risk genes. We began with the 150 genes that are already known to cause CHD and are associated with the cilium, which comprise the cilia-CHD gene set. First degree interactors of these cilia-CHD were identified in the BioGRID database, which includes both direct biophysical protein interactions and protein complex co-membership. We filtered these interactions using spatial transcriptomic data from embryonic heart tissue for those in which both genes are expressed in the same region of the heart during development. The complete protein-protein interactome was comprised of 78,775 interactions between 4,566 genes.

3.3.2 Conditional correlation analysis to identify gene-gene interactions in CHD patients

We analyzed whole exome sequencing (WES) data from 481 CHD patients from the Children's Hospital of Pittsburgh UPMC (Pitt), 1451 CHD patients from the Pediatric Cardiac Genomics Consortium (PCGC), and 2602 controls from the Alzheimer's Disease Sequencing

Project (ADSP) with European ancestry. To reduce noise introduced by variants that are less likely to be damaging, variants were filtered for protein-truncating variants (PTVs). This includes variants with a consequence of start loss, stop loss, frameshift insertion-deletion, splice acceptor site, and splice donor site. Multi-marker Analysis of GenoMic Annotation (MAGMA)[102] was used to calculate the significance of PTVs in individual genes genome-wide in cases compared to controls.

3.3.3 Prioritizing genes for involvement in complex genetic model based on GGI network

We then performed conditional correlation analysis of CHD cases using ModPGMInference[119], resulting in a gene-gene interaction (GGI) network of 1285 genes participating in 2962 interactions, 142 of which are known CHD genes (Fisher's exact OR=1.71, p=2.26e-11) and 415 of which are known cilia-related genes (Fisher's exact OR=1.24, p=4.66e-8). Because we filtered specifically for PTVs, we also looked to see whether genes in the interactome are intolerant to the haploinsufficiency that can be caused by these types of variants. There are 177 genes in the GGI network that are intolerant to heterozygous loss of function (pLI > 0.9), which is a significant enrichment (OR=1.68, p=2.06e-13). Interestingly, this number of interactions is much greater than the 1925 interactions across 1052 genes found to be statistically significant in the control cohort, despite the control cohort having a larger sample size.

We prioritized these GGIs based on fetal heart co-expression and the PPI network. Of these interactions, 2951 were not also statistically significant in controls, 1334 of those were co-expressed in fetal heart involving 977 individual genes. 23 significant and co-expressed interactions are between known CHD genes (Figure 11). Many of the 1334 gene-gene interactions

are also close together in the PPI network, with 474 GGIs sharing common neighbors in the PPI network and 25 participating directly in PPIs (Figure 12).

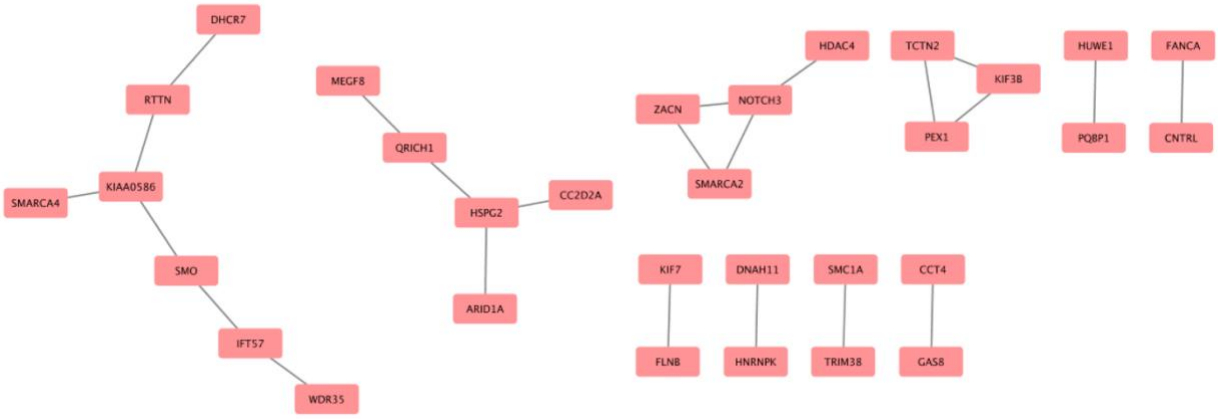

Figure 11. Significant interactions between known CHD genes

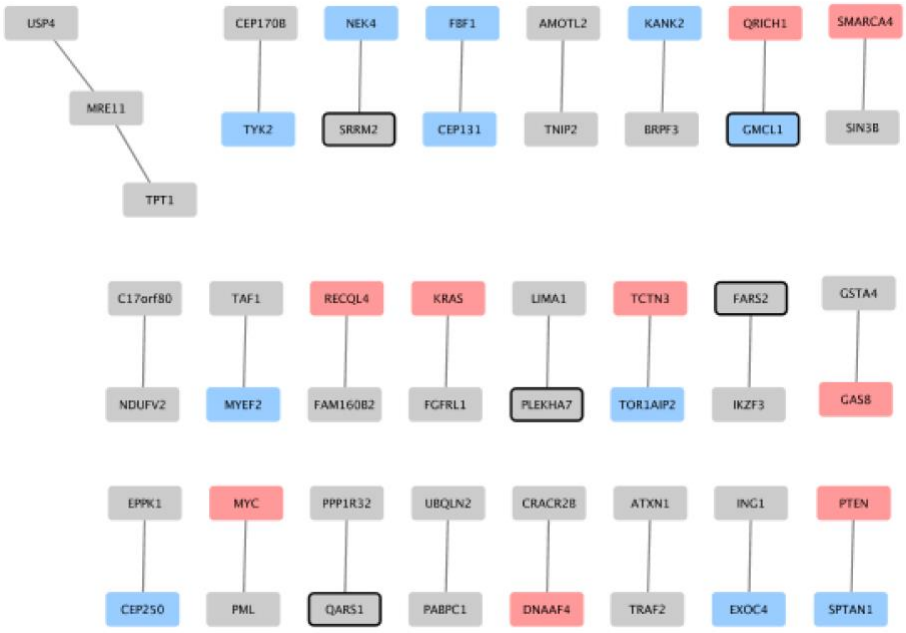

Figure 12. Interactions that are both significant GGIs and direct PPIs

Red nodes indicate known CHD genes; Blue nodes indicate known cilia genes that are not CHD associated; Grey nodes are neither CHD nor cilia related; Black outlines indicate MAGMA $p < 0.05$.

This GGI network allows us to identify genes that may be novel candidate CHD genes, contributing to disease in the context of other variants. The GGI network is considered scale-free, such that the degree distribution follows the power law (r^2=0.911), indicating that there are many genes with few interactions and a small number of "hub" genes that participate in many interactions. 98 genes with the top 10% of degree centrality scores were considered hub genes (Table 7). We performed gene set enrichment analysis of these hub genes using Metascape, and surprisingly, they are enriched for genes related to neural stem cell population maintenance (4.47e-5) and DNA repair (p=0.00017, Figure 13). They are also enriched for genes causing disease phenotypes such as Joubert syndrome, (p=0.00015), microcephaly (p=5.01e-9), micrognathism (p=6.31e-6), and frontal lobe hypoplasia (p=3.16e-6, Figure 14).

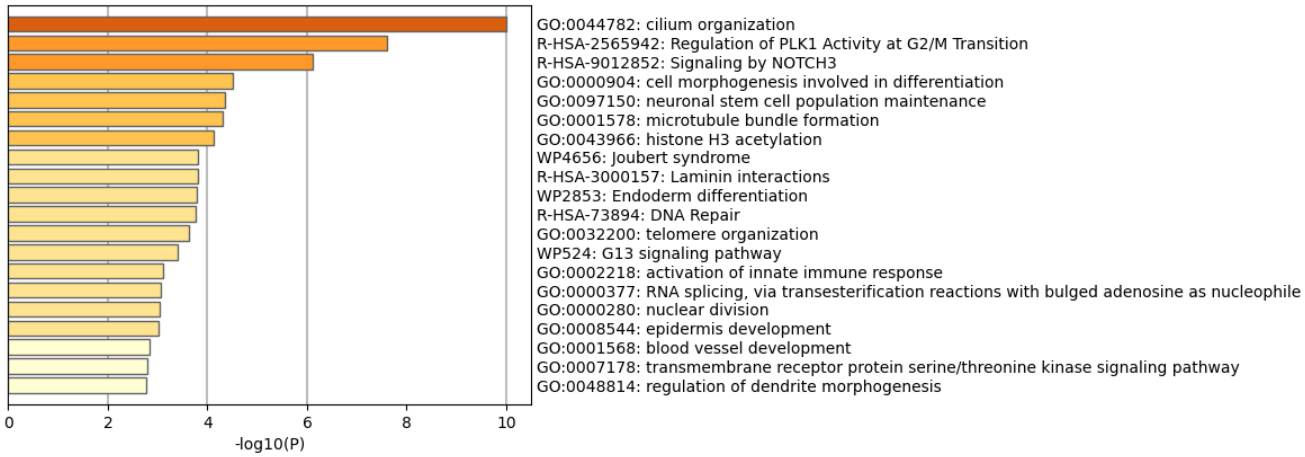

Figure 13. Gene set enrichment analysis of hub genes for pathways and ontology terms

Analysis and visualization performed in Metascape[103].

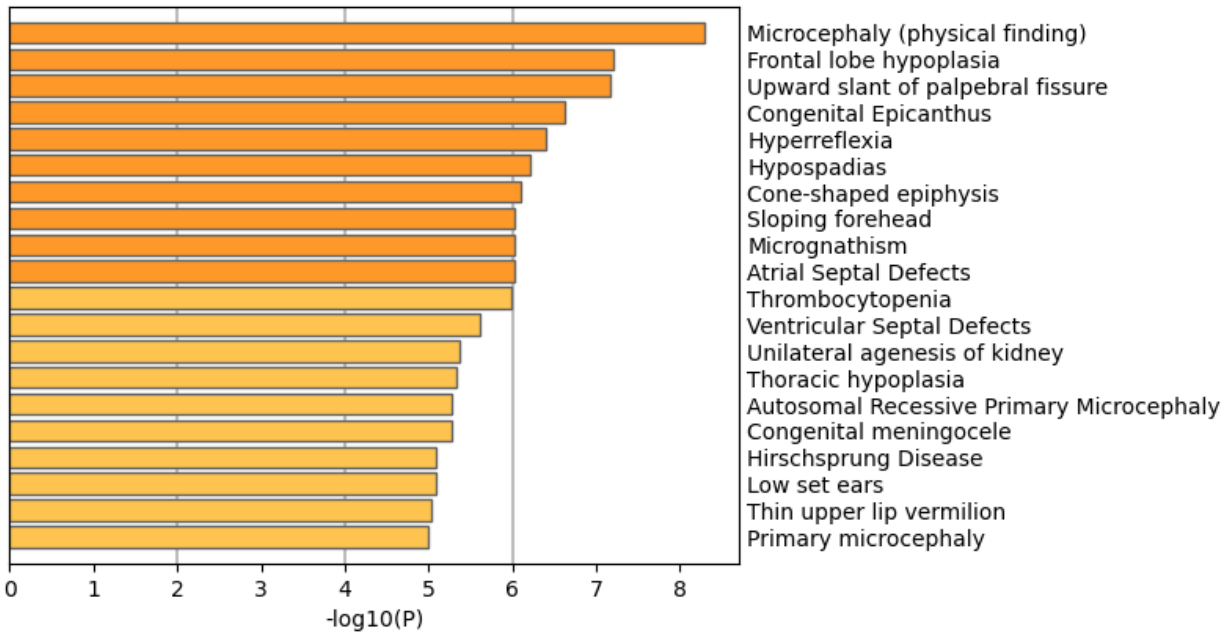

Figure 14. Gene set enrichment analysis of hub genes for human disease phenotypes

Analysis and visualization performed in Metascape[103].

We used the Girvan-Newman community clustering algorithm to identify densely connected regions using the network of 1334 gene-gene interactions/independent correlations that are statistically significant in cases but not controls and in which the genes are co-expressed in the fetal heart. The clustering cutoff was chosen to optimize the modularity score, with the edges weighted using the z-score of the correlation coefficient. This community clustering resulted in 78 clusters with an average of 12.5 nodes per cluster (Table 8). We prioritized the 30 clusters containing over 5 genes. Based on gene set enrichment analysis in Metascape, the largest cluster, cluster 1 with 65 genes, has greatest enrichment for genes associated with intrinsic apoptotic signaling pathway (p=5.4e-5), DNA repair (p=0.00012), and cell cycle (p=0.00019, Figure 15). Although no cluster is significantly enriched for known cilia-CHD genes after multiple testing correction, cluster 10 contains 6 cilia-CHD genes (8 total known CHD genes), more than any other cluster (cilia-CHD OR=2.94, p=0.030, Figure 16). Of the genes in the cluster not currently curated

to cause CHD, 5 have some evidence for involvement with heart development or have been associated with CHD in at least one human study (*CUBN*[121], *EXOC4*[122,123], *MAP3K11*[124], *MATR3*[125], and *PRPF38B*[126]). Cluster 12 is enriched for PPIs between any two genes in the cluster (OR=2.86, p=0.000968, Figure 17). Metascape gene set enrichment analysis of cluster 12 shows that it is also enriched for genes that are involved in cell division (p=4.8e-5).

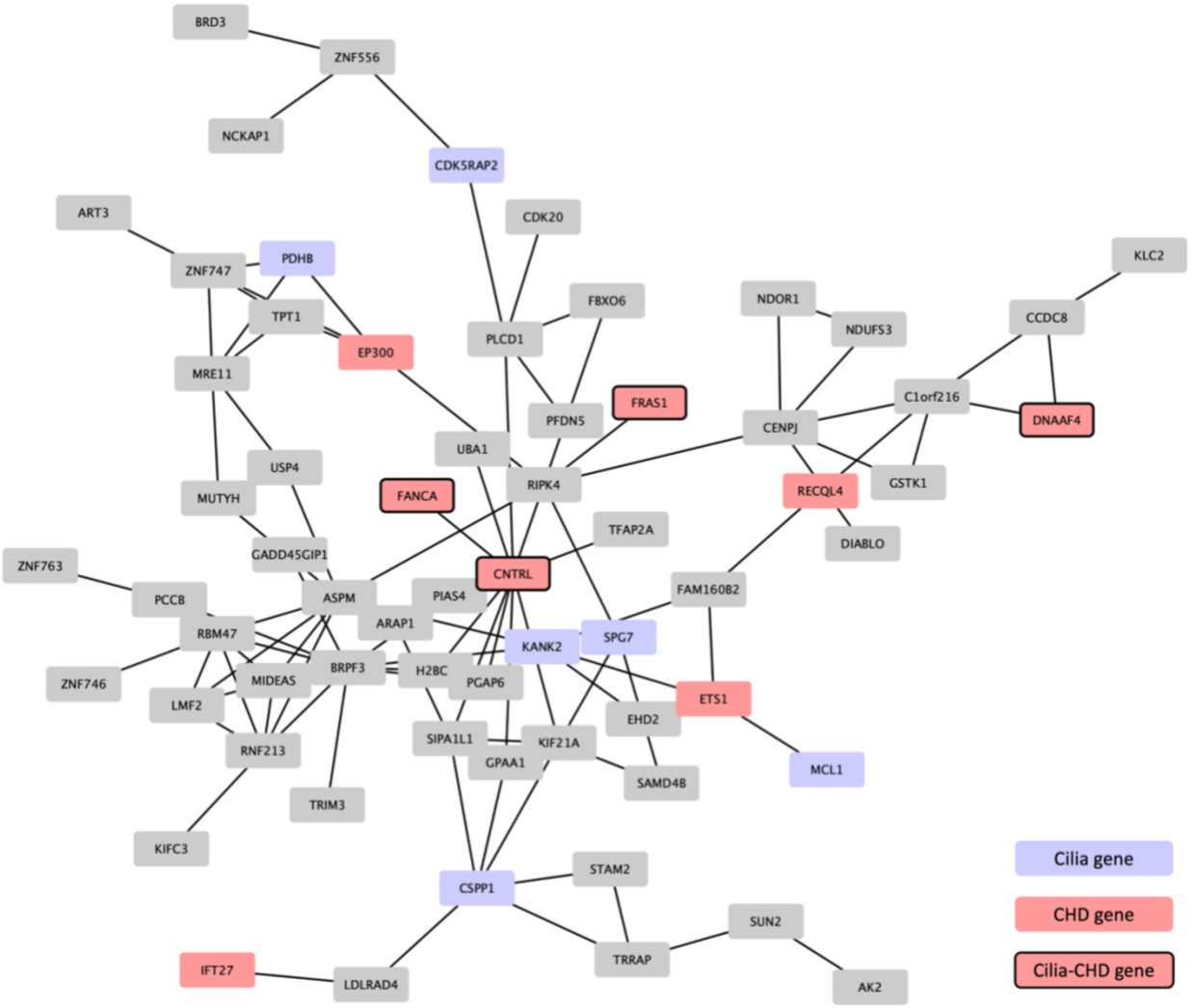

Figure 15. Cluster 1 is the largest cluster

Black lines indicate gene-gene interactions. Red nodes indicate known CHD genes. Blue nodes indicate known cilia genes that are not CHD associated. Red nodes with black outlines indicate cilia-CHD genes. Grey nodes are neither CHD nor cilia related.

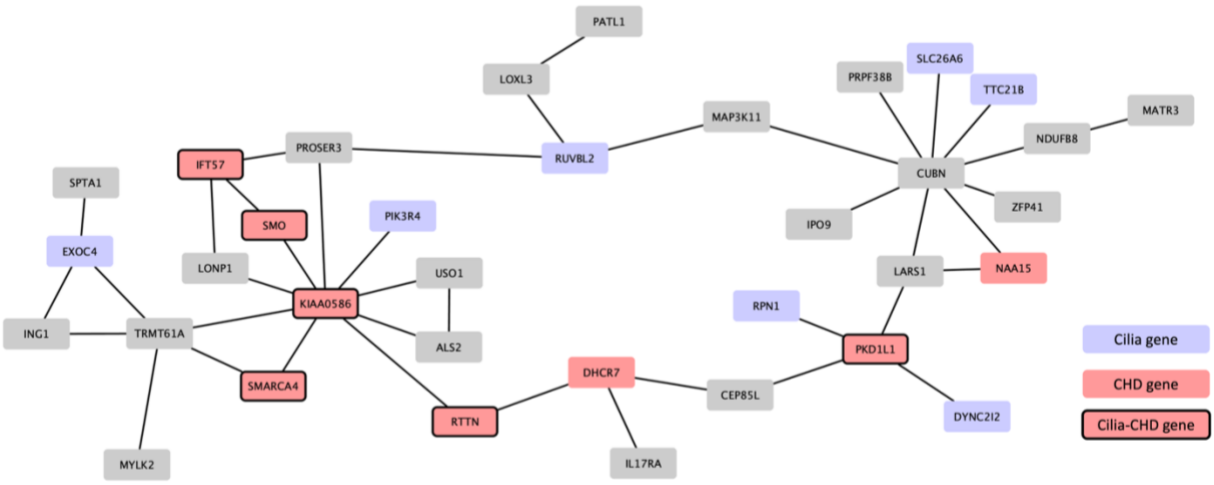

Figure 16. Cluster 10 is significantly enriched for known cilia-CHD genes

Black lines indicate gene-gene interactions. Red nodes indicate known CHD genes. Blue nodes indicate known cilia genes that are not CHD associated. Red nodes with black outlines indicate cilia-CHD genes. Grey nodes are neither CHD nor cilia related.

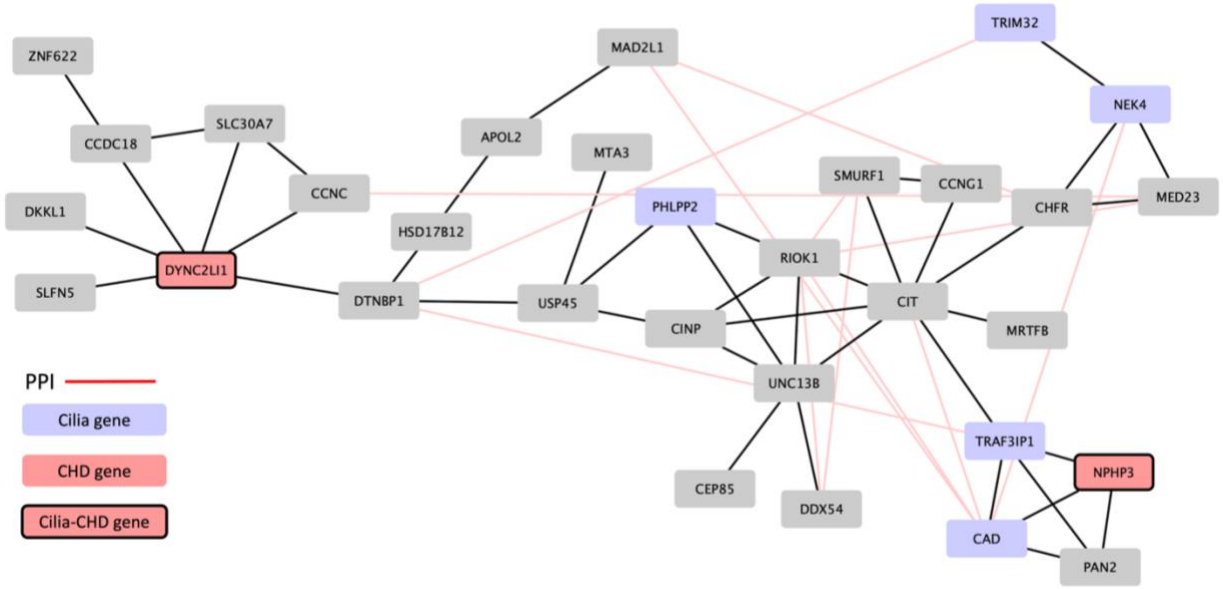

Figure 17. Cluster 12 is enriched for PPIs among genes

Black lines indicate gene-gene interactions. Red nodes indicate known CHD genes. Blue nodes indicate known cilia genes that are not CHD associated. Red nodes with black outlines indicate cilia-CHD genes. Grey nodes are neither CHD nor cilia related.

3.4 Discussion

There is evidence that CHD has a complex genetic etiology, with a digenic model of disease being the most likely. However, performing genome-wide combinatorial analysis to identify associated gene pairs is computationally intractable. As most CHD associated genes exist within a tight functional network, using the protein-protein interactome is one way to limit the search to genes likely to be disease associated. Here, we identified GGIs significantly associated with CHD patients within the cilia-CHD interactome and identified candidate genes that are present in statistically significant digenic pairs with known CHD causing genes. We used community clustering to identify clusters of genes between which there are dense GGIs and that are also enriched for statistically significant genes or known PPIs.

We identified 98 hub genes that participate in GGIs with many other genes. These high degree centralities may be due to either individual samples with many PTVs or the presence of relatively common variants in these genes which may require a secondary hit in another gene to cause CHD. Interestingly, the hub genes are associated with phenotypes including brain abnormalities and craniofacial dysmorphism. This is interesting given the well-described occurrence of these defects, as well as neurodevelopmental defects, in CHD patients[127–129]. Notably, many HLHS patients also have microcephaly[130,131]. This suggests that these hub genes causing disease under a complex genetic model are also contributing to extracardiac phenotypes and that the co-occurrence of these phenotypes has genetic underpinnings.

The community clustering analysis identified groups of genes that are densely correlated in CHD patients. The enrichment of DNA repair, cell cycle, and apoptosis genes among the largest and most densely connected GGI cluster suggests that the ciliary contribution to the pathogenesis of CHD involves the disturbance of these cellular processes. The cluster that also densely

connected by PPIs further shows that the PPI network provides validation that investigation of GGIs within a PPI network is a promising way to identify novel disease genes.

The cluster with many known cilia-CHD genes also provides high interest candidate genes, as they are densely connected with many other known CHD genes, and this approach is validated by the knowledge that several genes in the cluster already have some known association with heart development or CHD. An intronic variant in the vitamin B12 receptor *CUBN* is associated for decreased risk of CHD[121]. Mutations in *EXOC4* and other genes encoding proteins involved in the exocyst, which is necessary for ciliogenesis and ciliary function, are associated with BAV[122]. A rare de novo CNV encompassing *MAP3K11* was identified in a CHD cohort[124]. *MATR3* is a gene in the CPLANE network that is highly expressed in the developing heart, and a homozygous mouse model shows incompletely penetrant bicuspid aortic valve, coarctation of the aorta, and patent ductus arteriosus, which are similar phenotypes to those seen in an LVOTO patient with a translocation of the gene[125]. Lastly, PRPF38B is a pre-mRNA processing factor that is found significantly more abundantly in patients with Ebstein's anomaly, a rare tricuspid valve defect, than in control patients[126].

This network analysis and clustering can help identify other candidate CHD genes that may cause disease in non-Mendelian or multigenic fashions. While these candidate genes and interactions must be further explored in model systems, the gene-gene interactions in addition to interactions in the protein-protein interaction network provide a broad biological context for pathogenicity.

4.0 General Discussion

4.1 Conclusions

Appreciation for the genetic complexity of congenital heart disease has grown recently. However, more specific knowledge of genes underlying pathogenic interactions has largely remained a mystery. Here, we show that variants in ciliary genes contribute to a subtype-specific complex genetic model of CHD and that the PPI network provides the biological context for these complex genetic underpinnings. The results of these studies align with what has previously been shown regarding the association of ciliary genes with CHD and the observations of CHD patients with variants in multiple CHD causing genes.

Specifically, we show that conotruncal defects follow a dominant model of disease, laterality defects follow a recessive or digenic model of disease, and left ventricular outflow tract obstructions follow a digenic model of disease. The genes involved in these digenic interactions are likely to be tolerant to heterozygous loss of function on their own but cause phenotype in combination with other variants.

We then utilized the protein-protein interaction network to make a broader analysis of gene-gene interactions more computationally feasible. Hub genes that participate in many interactions also likely contribute to the occurrence of extracardiac phenotypes among CHD patients. Many non-cilia genes in the cilia-CHD interactome also participate in significant interactions in CHD patients, particularly those involved in DNA repair, cell cycle, and apoptosis, providing another group of genes to investigate for involvement in the complex genetics of CHD.

4.2 Limitations

There are several limitations to these studies. The sample size of our case and control cohorts is relatively small, providing low power to detect associations. This adds particular difficulty when grouping patients by phenotype, as there is only a small number of patients with each specific defect. Therefore, there is limited power when performing subtype analysis, and we were unable to perform association analyses with more fine-grained phenotyping. In addition, combinatorial analyses pose an even greater hurdle due to the exponentially greater number of tests being performed as the number of genes included increases. Because of this, even subtype analyses using large groupings are infeasible for gene-gene interaction analysis.

Second, we only considered protein truncating variants in this study. Although this group of variants is most likely to lead to a damaged protein, some PTVs may have no effect, and other types of variants, such as missense variants and even noncoding variants, may have large effects that are not captured here. In addition, there is stringent quality control filtering due to the fact that samples were sequenced in separate batches, so some true PTVs may have been filtered out due to having low quality in at least one cohort or being outside the intersection of sequencing intervals.

Last, protein-protein interaction networks are incomplete, and proteins that are well-studied are likely to have more interactions. There may be interactions between understudied genes that are unknown or not curated in BioGRID and related PPI databases. In addition, there is limited expression data of the developing heart. Thus, there may be other co-expressed genes within the cilia-CHD interactome that are involved in complex genetic interactions that are not represented here.

4.3 Future Directions

Gene-gene interactions involving *KIAA0586* (formerly known as *TALPID3*) are of particular interest for future study. The protein encoded by this gene is a centrosomal protein involved in ciliogenesis that responds to signaling by the Shh pathway. It is associated with Joubert syndrome and other ciliopathies, and in MGI it is curated to cause abnormal heart looping morphology. Although PTVs in this gene are observed in controls, it significantly co-occurs with many other genes in cases, and these gene pairs are not significant in controls, suggesting that a heterozygous variant in this gene along with another gene is necessary to cause CHD. PTVs in the pair of cilia-CHD genes *KIAA0586* and *CCDC39* were identified in two CHD patients and no controls, and the presence of a PTV in a third cilia-CHD gene in one patient suggests that these two genes cause CHD, while the third modifies phenotype. *KIAA0586* is also a hub gene in the broader gene-gene interaction network, significantly co-occurring with 11 other co-expressed genes, two of which, *RTTN* and *PIK3R4*, are also known to be cilia-related. It would be interesting to observe whether a variant in *KIAA0586* is in fact necessary to cause CHD in the context of variants in these other genes and whether the secondary variant modifies CHD phenotype.

There are both statistical and experimental opportunities for future work to expand on these studies. While this study focuses exclusively on protein truncating variants, missense and even synonymous and noncoding variation can affect phenotype, and these types of variants should be explored for association with disease. The current methods can be used to test for significant co-occurrence of missense and other types of variants within cases and compare these correlations with controls. To adjust for any noise that may be introduced with a less stringent variant consequence, additional filters can be used, including limiting to variants that are rare in the gnomAD database with an allele frequency less than 5%, 1%, 0.1%, etc. and/or limiting to variants

that are predicted to be pathogenic based on variant deleteriousness and pathogenicity scores such as CADD, SIFT, and PolyPhen. A more fine-grained approach may include testing for correlations between specific variants in genes of interest. For specific GGIs of interest, it would also be important to test for significance of, or at least observe, co-occurrence of the genes within an independent cohort of CHD patients. It may also be interesting to test for the association of extracardiac defects with gene pairs in CHD patients, given that hub genes are significantly enriched for genes associated with brain and craniofacial abnormalities.

Gene pairs of interest can be experimentally validated through CRISPR mouse modeling. For variants that are predicted to cause loss-of function, knockout mouse lines can be generated using CRISPR gene editing for each gene separately and intercrossed to generate double heterozygous mice. Using imaging techniques such as prenatal ultrasound, color flow, and postmortem episcopic confocal microscopy, the phenotypes of the single mutants can be compared with that of the double mutant to see if the presence of multiple mutations causes a different CHD phenotype, a more severe phenotype or has a protective effect on phenotype. For variants that are likely to cause embryonic lethality, CRISPR can be used for heterozygous knockout, and imaging techniques can be used to assess for CHD phenotype of F0 embryos and mice. Successful validation of a statistically significant gene-gene interaction in a mouse model would further demonstrate the utility of analysis of variants for conditional correlation in humans and the digenic etiology of CHD.

Appendix

Table 5. 150 cilia-CHD genes and sources

	CHD						Cilia				
Symbol	Lo	Watkins	Sifrim	MGI	CHDGene	Williams	Reiter	Watkins	CiliaCarta	Lo	CPLANE
ACTB			x								x
ALDH1A1		x							x		
ANKS6	x			x			x		x	x	
AP1B1	x			x						x	
AP2B1	x									x	
ARL2BP				x			x	x	x		
ARMC4	x			x			x	x	x	x	
B3GALT6			x								x
B9D2				x			x	x	x		
BICC1	x	x								x	
C2CD3				x			x	x	x		
C9orf116				x					x		
CC2D2A	x			x			x	x	x	x	
CCDC103			x				x	x	x		
CCDC114			x				x	x	x		
CCDC151	x		x	x			x	x	x	x	
CCDC39	x		x	x			x	x	x	x	
CCDC40			x	x			x	x	x		
CCT4		x									x
CEP120				x			x				
CEP290	x		x	x			x	x	x	x	
CFAP53				x			x				
CLUAP1				x			x	x	x		
CNTRL	x									x	
CPLANE1	x						x			x	x
CPLANE2	x			x				x	x	x	x
CRX		x					x			x	
DAW1	x			x				x	x	x	
DCTN5	x			x						x	
DDX59			x				x				
DLL1				x				x			

Table 5 (continued)

DNAAF2				x			x	x	x		
DNAAF3	x		x	x			x	x	x	x	
DNAAF4	x		x	x			x	x	x	x	
DNAH11	x			x	x		x	x	x	x	
DNAH5	x			x			x	x	x	x	
DNAI1	x			x			x	x	x	x	
DNAI2	x						x	x	x	x	
DNM2	x									x	
DOCK1	x			x						x	
DPCD	x			x				x	x	x	
DPM1		x									x
DRC1	x			x			x	x	x	x	
DVL1						x		x	x		
DVL2				x				x			
DVL3	x									x	
DYNC2H1	x			x			x	x	x	x	
DYNC2I1			x				x	x	x		
DYNC2LI1				x			x	x	x		
DYNLL1				x					x		
DZIP1		x							x		
EFCAB1				x					x		
EFTUD2			x		x				x		
ENKUR				x				x	x		
EPHB4	x									x	
ERMP1	x									x	
EVC			x		x		x	x	x		
EVC2			x		x		x	x	x		
FGFR2			x		x			x			
FLNA			x		x			x	x		
FOXC1		x	x		x	x					x
FOXJ1	x			x		x		x	x	x	
FRAS1	x									x	
FUZ	x							x	x	x	x
GAS8				x			x	x	x		
GLI3			x		x		x	x	x		
HECTD1	x									x	
HHIP				x					x		

Table 5 (continued)

HNRNPK					x						x
HSPB11				x				x	x		
HUWE1		x							x		x
IFT122				x			x	x	x		
IFT140	x			x			x	x	x	x	
IFT172				x			x	x	x		
IFT27				x			x	x	x		
IFT46				x				x	x		
IFT57				x			x	x	x		
IFT74	x			x				x	x	x	
IFT88				x				x	x		
INVS				x	x		x	x	x		
KIAA0586				x			x	x	x		
KIF3A				x				x	x		
KIF3B				x				x	x		
KIF7	x		x				x	x	x	x	
LAMA5	x			x				x	x	x	
LOX	x									x	
LRP1	x									x	
LRP2	x	x	x							x	
LRRC6				x			x	x	x		
MAP2K1			x		x			x			
MAPK7	x									x	
MEGF8	x		x	x				x			
MGRN1				x							x
MKS1	x		x	x			x	x	x	x	
MNS1				x				x	x		
MYH10	x	x							x	x	
NEK1			x				x	x	x		
NEK7	x									x	
NEK8	x			x			x	x	x	x	
NME7				x			x	x	x		
NOTCH1			x	x	x	x		x			
NOTCH2			x	x	x	x		x			
NPHP3			x		x		x	x	x		
NPHP4					x		x	x	x		
OFD1				x			x	x	x		

Table 5 (continued)

PCMTD2		x									x
PDE2A	x									x	
PDGFRB	x									x	
PEX6		x									x
PHC1		x		x					x		
PHGDH			x								x
PKD1	x		x				x	x	x	x	
PKD1L1	x			x	x		x	x	x	x	
PKD2			x	x			x	x	x		
PQBP1		x	x						x		
PRICKLE1	x									x	
PRKAR1A		x							x		
PSKH1	x			x						x	
PSME4	x									x	
PTK7	x									x	
QRICH1	x									x	x
QSOX1	x									x	
RAB23			x		x		x	x	x		
RAF1		x			x						x
RBFOX2					x						x
RFX2		x							x		
RFX3	x			x		x		x	x	x	
RPGRIP1L				x			x	x	x		
RTTN				x				x	x		
SMARCA4	x		x		x					x	
SMC3					x						x
SMO				x			x	x	x		
SUFU	x			x				x	x	x	
TAB1	x									x	
TAZ	x									x	
TBC1D32	x		x	x			x	x	x	x	
TCTN2				x			x	x	x		
TCTN3				x			x	x	x		
TMEM67	x			x			x	x	x	x	
TRAPPC10	x							x	x	x	
TRIM38	x									x	
TTBK2				x			x	x	x		

Table 5 (continued)

TTC25				x			x				
TWF2	x									x	
VANGL2				x				x	x		
WDPCP	x						x	x	x	x	x
WDR35				x			x	x	x		
WNT5A	x		x							x	
ZIC2				x			x				
ZMYND10				x			x	x	x		

Table 6. Protein truncating variants in case-only digenic network

Sample ID	Variant	GT	Gene	Consequence	Exon	Intron	CADD Phred	gnomAD AF	Phenotype group
PCGC_1133	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	LAT
PCGC_1133	chr20:62310313:T:C	0/1	LAMA5	Splice acceptor	.	76/79	32	8.51E-06	LAT
PCGC_11954	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	LAT
PCGC_11954	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	LAT
PCGC_1230	chr1:1341779:C:A	0/1	DVL1	Stop gain	5/15	.	39	.	LVO
PCGC_1230	chr20:62322110:CG:C	0/1	LAMA5	Frameshift	48/80	.	25.5	.	LVO
PCGC_14612	chr17:80047356:C:T	0/1	CCDC40	Stop gain	2/5	.	3.238	5.98E-03	LAT
PCGC_14612	chr15:89632945:G:A	0/1	KIF7	Stop gain	14/19	.	43	1.61E-05	LAT
PCGC_14630	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	LAT
PCGC_14630	chr4:5584777:AC:A	0/1	EVC2	Frameshift	17/23	.	26.1	.	LAT
PCGC_14630	chr14:58432438:AG:A	0/1	KIAA0586	Frameshift	4/33	.	33	3.05E-03	LAT
PCGC_1523	chr2:61868697:G:A	0/1	CCT4	Stop gain	14/14	.	41	2.79E-05	CTD
PCGC_1523	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	CTD
PCGC_15502	chr3:108222302:GA:G	0/1	IFT57	Frameshift	1/7	.	23.3	.	OTH
PCGC_15502	chr14:58432438:AG:A	0/1	KIAA0586	Frameshift	4/33	.	33	3.05E-03	OTH
PCGC_15502	chr7:129210550:T:C	0/1	SMO	Splice donor	.	4/4	34	.	OTH
PCGC_15850	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	CTD
PCGC_15850	chr9:136502337:G:GA	0/1	NOTCH1	Frameshift	28/34	.	33	.	CTD
PCGC_16701	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	LVO
PCGC_16701	chr9:136500579:C:CA	0/1	NOTCH1	Frameshift	31/34	.	33	.	LVO
PCGC_17253	chr20:32310885:GA:G	0/1	KIF3B	Frameshift	2/9	.	32	.	OTH
PCGC_17253	chr12:123690629:C:T	0/1	TCTN2	Stop gain	8/18	.	35	1.59E-05	OTH
PCGC_18390	chrX:53559498:GA:G	0/1	HUWE1	Frameshift	57/84	.	32	.	LVO
PCGC_18390	chrX:48902401:AGC:A	0/1	PQBP1	Frameshift	4/6	.	25.2	.	LVO
PCGC_19805	chr17:80047356:C:T	0/1	CCDC40	Stop gain	2/5	.	3.238	5.98E-03	LVO
PCGC_19805	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	LVO
PCGC_2010	chr19:42362115:TC:T	0/1	MEGF8	Frameshift	33/42	.	33	.	LVO

Table 6 (continued)

PCGC_2010	chr3:49057138:CA:C	0/1	QRICH1	Frameshift	4/11	.	32	.	LVO
PCGC_20164	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	OTH
PCGC_20164	chr3:108222320:C:A	0/1	IFT57	Start lost	1/7	.	22.4	1.61E-03	OTH
PCGC_20576	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	LVO
PCGC_20576	chr14:58432438:AG:A	0/1	KIAA0586	Frameshift	4/33	.	33	3.05E-03	LVO
PCGC_21711	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	CTD
PCGC_21711	chr6:57194813:TCTTAA:T	0/1	RAB23	Frameshift	5/7	.	32	.	CTD
PCGC_24058	chr7:21620016:C:T	0/1	DNAH11	Stop gain	25/83	.	39	1.22E-05	LAT
PCGC_24058	chr14:58432438:AG:A	0/1	KIAA0586	Frameshift	4/33	.	33	3.05E-03	LAT
PCGC_25811	chr7:21588578:C:T	0/1	DNAH11	Stop gain	11/83	.	35	1.93E-04	OTH
PCGC_25811	chr9:83971671:C:T	0/1	HNRNPK	Splice donor	.	11/15	33	.	OTH
PCGC_3085	chr7:21873273:G:C	0/1	DNAH11	Splice acceptor	.	74/82	33	.	LVO
PCGC_3085	chr5:13917252:C:T	0/1	DNAH5	Stop gain	8/79	.	37	.	LVO
PCGC_3340	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	LAT
PCGC_3340	chr22:39421960:C:T	0/1	TAB1	Stop gain	8/11	.	40	.	LAT
PCGC_3530	chr5:13727531:C:T	0/1	DNAH5	Stop gain	70/79	.	49	.	LVO
PCGC_3530	chr16:2090531:G:T	0/1	PKD1	Stop gain	45/46	.	38	.	LVO
PCGC_3673	chr4:15563395:C:T	0/1	CC2D2A	Stop gain	21/34	.	40	1.02E-04	LAT
PCGC_3673	chr5:13917252:C:T	0/1	DNAH5	Stop gain	8/79	.	37	.	LAT
PCGC_3787	chr17:80047356:C:T	0/1	CCDC40	Stop gain	2/5	.	3.238	5.98E-03	OTH
PCGC_3787	chr1:119948567:C:A	0/1	NOTCH2	Splice acceptor	.	16/33	32	.	OTH
PCGC_4319	chr14:58428357:T:TC	0/1	KIAA0586	Frameshift	1/33	.	22.8	2.85E-04	CTD
PCGC_4319	chr2:19938338:G:A	0/1	WDR35	Stop gain	19/28	.	40	3.18E-05	CTD
PCGC_4878	chr11:74074547:G:A	0/1	C2CD3	Stop gain	4/13	.	39	1.20E-05	LVO
PCGC_4878	chr19:49808825:T:G	0/1	FUZ	Splice acceptor	.	7/10	33	.	LVO
PCGC_4903	chr3:180661860:C:G	0/1	CCDC39	Splice donor	.	3/19	29.6	1.55E-04	OTH
PCGC_4903	chr21:44059187:C:T	0/1	TRAPPC10	Stop gain	6/24	.	39	.	OTH
PCGC_5710	chr17:44902645:CAG:C	0/1	CCDC103	Frameshift	4/4	.	27.6	0.00E+00	LVO

Table 6 (continued)

PCGC_5710	chr6:25969403:C:T	0/1	TRIM38	Stop gain	4/8	.	35	1.62E-02	LVO
PCGC_6298	chr9:121168270:C:T	0/1	CNTRL	Stop gain	36/42	.	44	.	LVO
PCGC_6298	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	LVO
PCGC_6347	chr14:58444071:CAA:C	0/1	KIAA0586	Frameshift	6/33	.	23.9	2.82E-05	LAT
PCGC_6347	chr19:10987003:G:T	0/1	SMARCA4	Stop gain	5/35	.	55	.	LAT
PCGC_659	chr20:62315083:G:T	0/1	LAMA5	Stop gain	59/80	.	33	.	LVO
PCGC_659	chr12:57210717:G:T	0/1	LRP1	Splice acceptor	.	82/88	34	.	LVO
PCGC_6688	chr20:62322732:C:A	0/1	LAMA5	Stop gain	46/80	.	51	.	CTD
PCGC_6688	chr1:180196997:TAAGGGCCCTG:T	0/1	QSOX1	Frameshift	12/12	.	32	.	CTD
PCGC_6802	chr19:42343479:C:T	0/1	MEGF8	Stop gain	9/42	.	36	0.00E+00	LAT
PCGC_6802	chr9:136505413:GC:G	0/1	NOTCH1	Frameshift	25/34	.	27.5	.	LAT
PCGC_7335	chr4:169424668:G:C	0/1	NEK1	Stop gain	29/34	.	36	1.25E-04	LVO
PCGC_7335	chr6:25969403:C:T	0/1	TRIM38	Stop gain	4/8	.	35	1.62E-02	LVO
PCGC_7960	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	OTH
PCGC_7960	chr19:42376557:GA:G	0/1	MEGF8	Frameshift	42/42	.	33	.	OTH
PCGC_811	chr5:37142481:C:G	0/1	CPLANE1	Splice acceptor	.	28/36	21.1	3.85E-04	LVO
PCGC_811	chr16:2089949:G:GT	0/1	PKD1	Frameshift	46/46	.	35	.	LVO
PCGC_884	chr9:98790541:CG:C	0/1	ANKS6	Frameshift	2/15	.	26.1	.	CTD
PCGC_884	chr6:25969403:C:T	0/1	TRIM38	Stop gain	4/8	.	35	1.62E-02	CTD
PCGC_8942	chr18:50250889:G:A	0/1	CFAP53	Stop gain	5/8	.	35	8.01E-06	LVO
PCGC_8942	chr7:47800714:G:A	0/1	PKD1L1	Stop gain	54/57	.	36	3.98E-06	LVO
PCGC_9141	chr2:43796796:G:A	0/1	DYNC2LI1	Splice donor	.	8/12	33	3.99E-06	CTD
PCGC_9141	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	CTD
PCGC_917	chr11:74122987:C:T	0/1	C2CD3	Splice donor	.	8/30	33	4.00E-06	CTD
PCGC_917	chr14:58432438:AG:A	0/1	KIAA0586	Frameshift	4/33	.	33	3.05E-03	CTD
PCGC_9514	chr9:34459054:G:GT	0/1	DNAI1	Splice donor	.	1/19	31	4.02E-04	LVO
PCGC_9514	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	LVO

Table 6 (continued)

PITT_7005	chr14:58432438:AG:A	0/1	KIAA0586	Frameshift	4/33	.	33	3.05E-03	LAT
PITT_7005	chr18:70168956:C:T	0/1	RTTN	Stop gain	12/48	.	27.1	.	LAT
PITT_7016	chr3:180616289:T:TA	0/1	CCDC39	Frameshift	2/5	.	26.1	7.92E-03	LAT
PITT_7016	chr7:21868989:C:T	0/1	DNAH11	Stop gain	74/83	.	57	.	LAT
PITT_7060	chr17:80039965:GC:G	0/1	CCDC40	Frameshift	3/11	.	0.04	4.41E-04	OTH
PITT_7060	chr10:127418543:T:C	0/1	DOCK1	Splice donor	.	45/51	12.94	1.06E-03	OTH
PITT_7128	chr3:108222320:C:A	0/1	IFT57	Start lost	1/7	.	22.4	1.61E-03	LAT
PITT_7128	chr2:19941796:A:C	0/1	WDR35	Stop gain	18/28	.	40	1.55E-04	LAT
PITT_7133	chr20:62315068:G:GC	0/1	LAMA5	Frameshift	59/80	.	30	.	LVO
PITT_7133	chr10:95682763:TA:T	0/1	TCTN3	Frameshift	12/14	.	29	.	LVO
PITT_7153	chr17:80058885:C:T	0/1	CCDC40	Stop gain	9/11	.	34	.	LVO
PITT_7153	chr16:53645725:C:CA	0/1	RPGRIP1L	Frameshift	17/25	.	33	7.96E-06	LVO
PITT_7367	chr7:21588578:C:T	0/1	DNAH11	Stop gain	11/83	.	35	1.93E-04	LAT
PITT_7367	chr2:26450086:G:A	0/1	DRC1	Splice donor	.	11/15	33	8.11E-05	LAT
PITT_7629	chr17:80047356:C:T	0/1	CCDC40	Stop gain	2/5	.	3.238	5.98E-03	LAT
PITT_7629	chr1:1342090:GTGAC:G	0/1	DVL1	Frameshift	4/15	.	32	.	LAT
PITT_7632	chr4:15579967:G:GT	0/1	CC2D2A	Frameshift	.	.	36	5.23E-05	LAT
PITT_7632	chr16:90031264:C:A	0/1	GAS8	Stop gain	3/6	.	1.58	7.31E-03	LAT
PITT_7638	chr17:80047356:C:T	0/1	CCDC40	Stop gain	2/5	.	3.238	5.98E-03	LAT
PITT_7638	chr17:58216138:G:A	0/1	MKS1	Stop gain	4/17	.	36	1.60E-05	LAT
PITT_7751	chr6:25969403:C:T	0/1	TRIM38	Stop gain	4/8	.	35	1.62E-02	LAT
PITT_7751	chr14:58432438:AG:A	0/1	KIAA0586	Frameshift	4/33	.	33	3.05E-03	LAT
PITT_7751	chr14:58474727:C:T	0/1	KIAA0586	Stop gain	9/22	.	38	.	LAT

Table 7. Hub genes in gene-gene interaction network

Gene	Degree
CEP170B	15
HSPG2	14
SEC16A	14
SENP3	13
ABL1	13
UBQLN2	12
CUBN	12
LAMA5	12
POLR3C	12
BRPF3	12
UBXN10	12
COL7A1	11
KIAA0586	11
MRE11	11
MYEF2	11
TNIP2	11
PKN1	10
GMCL1	10
CIT	10
ASPM	10
CNTRL	10
ERCC1	10
DNAH11	10
ADAMTSL4	9
RNF213	9
LMO7	9
SCNN1D	9
SRRM2	9
THBS3	9
SIN3B	9
ZNF223	9
ATN1	9
CEP131	8
EMP1	8
PQBP1	8
INCA1	8
NOTCH3	8
GPAT3	8
OBSL1	8
DCBLD2	8
NUMA1	8
LMF2	8
S100A4	8
DHX8	8
SLC2A8	8
CLUH	7
MIB1	7
CEP128	7
ASCC3	7
AAAS	7
PABPC1	7
SCYL1	7
AMOTL2	7
ELL	7
PDIA2	7
MAD2L2	7
MAGED1	7
DIAPH3	7
RTN4	7
NOTCH1	7
JAG2	7
WDR90	7
SSNA1	7
CEP250	7
DOT1L	7
SCAF11	7
RFX1	7
EP300	7
TYK2	6
DNAH5	6
ATXN1	6
SRC	6
JMJD8	6
PPM1J	6
ANKRD54	6
SIPA1L1	6
UNC13B	6
CC2D2A	6
RAPGEF3	6
DHCR7	6
WDR86	6
COP1	6
VPS33A	6
CEP152	6
GIGYF1	6
AAR2	6
ZBTB48	6
RBM47	6
FARS2	6
IFT57	6
IGSF1	6
PALD1	6
KANK4	6
BRD1	6
PCM1	6
DYNC2LI1	6
PPP1CB	6
IFT52	6

Table 8. Community clustering of gene-gene interaction network

Cluster #	Genes
1	RNF213, PFDN5, C1orf216, ETS1, NCKAP1, H2BC8, BRD3, TRIM3, PIAS4, CDK5RAP2, DIABLO, FRAS1, LDLRAD4, AK2, MRE11, FBXO6, MUTYH, SPG7, SAMD4B, UBA1, TRRAP, STAM2, FAM160B2, GSTK1, IFT27, PDHB, USP4, CSPP1, ART3, MCL1, ZNF763, EP300, TFAP2A, CCDC8, NDOR1, RECQL4, SUN2, FANCA, PGAP6, DNAAF4, MIDEAS, CENPJ, GPAA1, PCCB, KANK2, CNTRL, RIPK4, NDUFS3, CDK20, KIF21A, ZNF747, ARAP1, TPT1, LMF2, GADD45GIP1, ASPM, ZNF746, RBM47, EHD2, SIPA1L1, ZNF556, PLCD1, BRPF3, KIFC3, KLC2
2	SDHC, SPC24, FAM161B, FBXO11, NUDC, MPPE1, PEX12, CA14, CEP72, KIFBP, QSOX1, ADAMTS2, UBQLN2, POLR3C, TAF8, COP1, FBF1, ASS1, IKZF3, SENP3, RFX1, CC2D2A, CHEK2, AAR2, SCNM1, HSPG2, PABPC1, IGSF1, POMT1, RCC1, CEP131, NR0B2, PNMA1, OSBPL8, NR3C1, RRBP1, KCNJ8, TBC1D32, BNIP1, ATAD1, NIF3L1, FARS2, HCFC1, EEF1AKMT3, LIMCH1, GARRE1, EML2, ARID1A, RAB23, SLC25A25, BEGAIN, NPDC1
3	HLTF, CKAP2, ERBB2, RBM3, TICAM1, BRCA1, SMAD6, TTLL12, GRWD1, ZNF669, KANK4, ZNF764, RXRA, HMMR, ZNF575, KLHDC8A, WDR5B, TEAD2, ZNF644, TMEM30A, ANKRD26, CCDC114, BCKDHB, SPEN, LYPD6, SS18L1, RPS16, USP9X, SHBG, CRY1, CFAP94, WDR86, TNKS, ALKBH3, RREB1, ZBTB46, NPC2, YBX1, JAK3, BMPR1A, SMCHD1, GPATCH1, KIF3A, TRPM4, RAPGEF3, PALD1, RTN4, SELENOH, PLEKHG1
4	EMP1, FGFRL1, PSME3IP1, LIG4, UBE3A, LRRC49, TTC30B, PIP5K1C, MOV10, ACTR1B, PNKD, HINT1, PDCD11, ADAMTS4, MAPK7, MYH7B, DNAH5, FANCF, PTPN1, CNTN1, SMAP2, PLOD2, THAP10, TTC23L, POMK, ELL, PKD1, LOXL2, KRAS, WWP1, EIF5, JAG2, HEXIM2, PRKCSH, GPC1, AHCY, GEM, CDC6, CPNE2, C2CD3, PRKN, STK11, CRACR2B, GLUL, IFT74, RASAL3
5	SSR1, LLGL2, EP400, RAB11FIP2, GNAI2, MICAL1, MAP2K5, ITSN1, AHI1, CDK15, SLC2A8, BBS4, OBI1, TSSC4, SNX33, ACSL3, CDKN1A, ERCC1, ABCA3, NUF2, MCM4, ZUP1, COPE, PCNT, LMO7, DUSP18, GATA1, ERG, LAMC1, STX5, TRAPPC9, LOX, MNS1, TAFA3, PINK1, C9orf78, FGF21, ATP2B2, GPBP1L1, DIDO1, ARV1, NUMA1, RUSF1, BRD9
6	GOLGA2, BRD1, MYPOP, GPAT3, SPDL1, OFD1, MAGED1, CLNK, PLAT, P4HA2, CTCF, NEURL4, ZNRD2, CENPE, SLC25A19, ZNF223, ERO1A, ZBTB48, SCYL1, LZTS2, TYK2, LRRC36, GRAP2, SMYD1, AQP5, NDUFA5, FOXJ1, TCTN1, CEP170B, GATB, HAUS6, ISG15, HDAC4, TRAPPC5, AGRN, TMED9, DHX8, DALRD3, DISC1, DCDC2
7	PEX6, TNIP2, PIAS3, CACNA1A, MARK2, FSCN2, GLB1, ANKS6, NCLN, TMTC4, C18orf21, ARHGEF16, EHBP1, CASP6, SMC4, PCDH20, BCL11B, MED17, PKP1, GFPT2, SELENBP1, ABL1, MAPK8IP1, TM9SF2, PLEKHM2, CCN3, TERF2, MPP3, CTSH, MSLN, PLEKHA7, AMOTL2, PYHIN1, NINL, COL3A1, SEC16A, LIMA1, AR, UBN1
8	REEP5, DGAT1, SMC1A, SIN3B, RHPN1, UCK1, RIBC2, TRIM38, MME, IARS1, XRCC4, CCT4, RBL1, POMT2, GAS8, MYOT, EIF3I, SMARCD2, OBSL1, RBM42, CEP19, CCNE1, SUSD4, TLX3, GSTA4, PSMC2, CCDC40, NDUFS1, AP1M1, MDN1, DDB1, METTL14, CCDC88B, ANKRD54, MAD2L2, SSNA1, EMC1, XPO7
9	PASK, BYSL, TCHP, LCA5, FADS2, DVL1, LAMA5, LRP1, LSM14B, POR, CCDC136, TLCD1, TUBGCP2, LUZP1, KDM6A, TOR1AIP2, PEX7, TRAPPC2L, P4HA3, TRAPPC10, SIPA1L2, CLCC1, MFSD10, MOGS, GAS6, ECH1, TCTN3, VPS13B, TTC3, GAK, TTC21A, LDHA, SRP72, TRIM28, RFC4
10	IPO9, IFT57, LARS1, RPN1, PKD1L1, MAP3K11, USO1, TTC21B, ALS2, MYLK2, IL17RA, CEP85L, TRMT61A, PRPF38B, SPTA1, MATR3, PROSER3, LONP1, PATL1, DYNC2I2, EXOC4, NAA15, KIAA0586, ING1, CUBN, SMARCA4, SLC26A6, DHCR7, SMO, LOXL3, RTTN, NDUFB8, ZFP41, RUVBL2, PIK3R4

Table 8 (continued)

11	NUMB, DHX34, PPP6R2, SCNN1D, HK1, SLFN11, TSC2, RPL5, PIGQ ,SRRM2, ITGA4, DSP, RASSF6, KDELR2, PKN1, SCAF11, LYPLA2, RAVER1, RHOG, MEGF8, WDR35, IFT52, DRC1, C11orf49, NOP56, TMEM222, TK1, RAD50, GMCL1, MGAT1, NECAB2, QRICH1
12	TRIM32, MED23, DYNC2LI1, DTNBP1, MTA3, CCNC, RIOK1, TRAF3IP1, NPHP3, MAD2L1, CCDC18, CEP85, PHLPP2, DKKL1, ZNF622, CHFR, PAN2, SMURF1, SLFN5, CCNG1, DDX54, APOL2, USP45, CIT, UNC13B, CAD, MRTFB, NEK4, CINP, SLC30A7, HSD17B12
13	BAG3, COL7A1, PEX1, RFWD3, JAG1, UNC45A, MFAP1, TCTN2, DOCK1, SMC2, MPZL1, MARS1, LMAN1, CERS2, FAM120C, CEP128, SMAD9, KLK10, YIF1A, CALM2, AAAS, TACC2, IPO8, TRAF2, ARHGAP9, UBXN10, ATXN1, TUBAL3, PLD3, KIF3B, TNS2
14	QPCT, ELN, CIITA, TMEM209, TCOF1, JUP, KHSRP, STN1, EMILIN1, THRA, CEP70, CC2D1A, WWOX, TP53, ANKRA2, FSD1, TNFRSF10A, PPP1CB, RPS6KB2, ABCD3, AIMP1, AFF4, STK36, CASP8, GATD1, ICAM1, BBS2, MIB1, SLC25A30, DNAJC1
15	DIAPH3, INTU, NONO, NFATC2, WDTC1, RABEP2, CLCN7, EVI5L, PSMA1, RNF181, CTSB, KDM3A, LATS2, SUZ12, PDIA2, AHNAK2, TXLNB, NEK5, EVC, HOOK1, AXIN1, PLXNA2, MAPK12, RFT1, POLR3F, USP14, CYLD
16	ADAMTSL4, ATM, ELAC2, BIRC6, MRPL9, ZFP36L2, C19orf54, CYB5R3, PKP3, ASCC3, ARAF, PTEN, CCDC146, DGKE, SORBS1, DDRGK1, DNAJC16, SPTAN1, PHF19, TRAPPC8, DYNC2H1, LGALS9, LRRK2, FANCG, ALG9, DOCK6
17	ZNF385C, CPSF2, ATP9A, ACACB, HNRNPK, NDUFV2, EXOSC4, PPM1G, DCBLD2, BICD2, DYNC1LI1, USHBP1, INCA1, DNAH11, DTX1, SMAD4, IGFBP6, ZNF512B, PTPRS, SYNE1, NTRK1, C17orf80, LRCH3, POLD1, MED25, ITIH5
18	RALBP1, SPATA18, ZNF792, ERLIN1, PRMT2, CEP162, AMBRA1, TEKT5, NOTCH1, STX18, PRKCD, ESYT1, METTL3, PPP6C, ADD1, RNF180, LYPD3, PSMC5, TBK1, RNF40, SOAT1, ZNF212, MAPK11, DHX9, GIGYF1
19	ICE1, QARS1, ATP6V1C2, TRIM54, MAP3K1, CHD7, POSTN, ISOC1, TYMS, NTPCR, LAMP1, EHHADH, SNX21, RUBCN, PPP1R32, NTAQ1, ARL2BP, CCT5, RPL23, PLOD3, ACLY, LRRC6, PDGFD, CAPN7
20	PML, SPAG8, LRRIQ1, DCAF4L2, ZBTB39, FOXO1, METTL17, AFDN, NDUFA13, CRBN, FRYL, MYC, WDR90, CEP164, CABIN1, TBC1D2, CCDC33, BCL7C, CEP192, NEK1, FASN, ANO10
21	BCAM, ITPRIP, LEPROT, HNRNPU, ODAPH, BAG4, SMARCA2, SART1, NOTCH3, S100A4, DMAP1, RPP25L, ZACN, ATN1, EXOC7, ZNF224, ARHGEF18, ARHGEF2, ZBTB17, PHF10, EMG1, MAPK8IP2
22	HECTD3, PQBP1, SNRNP35, EGFL7, SRC, ITSN2, TRIM35, BRD4, HUWE1, SNRNP200, ZNF526, TMCO3, CEP250, DBN1, NAA10, ATP5F1C, CFAP53, BIRC3, SPAG5, CLUH, EPPK1
23	ALDH1B1, SDCBP, CARNMT1, UBE3C, RBM10, PPP6R3, CLPTM1, ROR2, PCM1, PTGES2, DPM1, MTFR1, HBS1L, SETDB1, JMJD8, ZBED6CL, ZBTB26, CENPH, MEOX2, USP50
24	GOLGB1, AP1B1, OLFM2, SYCE1, UCHL5, DTX4, CLASP1, KARS1, AKAP9, CNIH4, MYEF2, NF1, TTLL3, PARK7, SMARCC1, TAF1, CEP152, TEPSIN
25	DCPS, FAM167A, ECT2, UGGT1, SUFU, CWF19L2, SEC24C, CARD10, DOT1L, GOSR2, TMEM67, MAML2, GPD2, TSHR, STXBP4, HIP1, VPS33A
26	ZBTB33, SBF1, MEPCE, TRMT1L, SLC47A1, SH3RF1, SP100, SKIL, DHRS7, THBS3, TBXA2R, RPGRIP1L, NFIX, RSPH14, MYO18A, TEX101
27	TRIP10, RAB8A, PDHA1, EMC3, SLC30A5, HYAL2, PCK2, MLH3, MAGI1, EGFR, ZNF517, TMEM33, CCT6B
28	LRP2, DDX5, ASCC2, PLOD1, PPM1J, HMOX1, AIMP2, EPN1

Table 8 (continued)

29	CDKAL1, POC1A, FLNB, KIF7, AGTPBP1, MCPH1, KCTD18
30	PARD6A, NCAPD2, TCF12, SMAD7, PDIA3, INVS
31	CKAP4, ARMC6, DNAJB1, PLK1, BLOC1S2
32	PDS5A, URB1, WDFY3, SLC12A2
33	R3HCC1L, TMEM45A, AQP6, GRAMD1A
34	PIBF1, FAM83D, MANSC1, RAB3IL1
35	WRAP73, TP73, IPO5
36	SCFD1, GOLGA5, SSH3
37	LRP12, EPHA1, IFT172
38	LYN, DHX58, PTCH2
39	TRAF5, TRIM13, CEP89
40	HSDL2, INTS9, ANAPC5
41	PODN, CTNNAL1, NUDCD3
42	RUNX1, TASOR
43	CEP44, H2AC4
44	HGS, CRYL1
45	BRF2, FANCC
46	ANXA5, HCLS1
47	TMPRSS11B, DDIT4
48	RPUSD3, TMEM120B
49	ERCC6, CEP57L1
50	BICC1, LMO4
51	GRIP1, BCAR3
52	GPS1, FNDC11
53	ST8SIA4, PDHX
54	RFC1, TJP2

Table 8 (continued)

55	ARMC12, SHMT2
56	CCAR1, CCDC57
57	AAK1, NFE2L2
58	SETD7, PSMD8
59	ANO6, HOMEZ
60	CFTR, TTLL5
61	RUNDC3A, MPHOSPH9
62	CHEK1, CFAP20
63	MYL6B, MCC
64	DESI1, DIPK1B
65	SPG21, PKP2
66	RUVBL1, FBXL4
67	TRIP6, TMEM161A
68	RPS23, CPA4
69	SLC39A5, PLEKHA4
70	MYH10, VHL
71	DHPS, PPM1F
72	NOTCH2, MED31
73	PHB2, COMMD2
74	ETNK2, CLSTN3
75	CDC7, RIDA
76	POLL, PDE8A
77	ERGIC2, TMEM80
78	MAU2, LNX1

Bibliography

1. Kirby RS. The prevalence of selected major birth defects in the United States. *Seminars in Perinatology*. 2017;41(6):338–344.
2. Liu Y, Chen S, Zühlke L, Black GC, Choy M, Li N, Keavney BD. Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. *International Journal of Epidemiology*. 2019;48(2):455–463.
3. Van Der Linde D, Konings EEM, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJM, Roos-Hesselink JW. Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. *Journal of the American College of Cardiology*. 2011;58(21):2241–2247.
4. Mandalenakis Z, Giang KW, Eriksson P, Liden H, Synnergren M, Wåhlander H, Fedchenko M, Rosengren A, Dellborg M. Survival in Children With Congenital Heart Disease: Have We Reached a Peak at 97%? *Journal of the American Heart Association*. 2020;9(22):e017704.
5. Bouma BJ, Mulder BJM. Changing Landscape of Congenital Heart Disease. *Circulation Research*. 2017;120(6):908–922.
6. Brodwall K, Greve G, Leirgul E, Tell GS, Vollset SE, Øyen N. Recurrence of congenital heart defects among siblings — a nationwide study. *American Journal of Medical Genetics, Part A*. 2017;173(6):1575–1585.
7. Gill HK, Splitt M, Sharland GK, Simpson JM. Patterns of recurrence of congenital heart disease: An analysis of 6,640 consecutive pregnancies evaluated by detailed fetal echocardiography. *Journal of the American College of Cardiology*. 2003;42(5):923–929.

8. Pierpont ME, Brueckner M, Chung WK, Garg V, Lacro RV, McGuire AL, Mital S, Priest JR, Pu WT, Roberts A, Ware SM, Gelb BD, Russell MW. Genetic Basis for Congenital Heart Disease: Revisited: A Scientific Statement From the American Heart Association. *Circulation*. 2018;138(21):653–711.
9. van Nisselrooij AEL, Lugthart MA, Clur S-A, Linskens IH, Pajkrt E, Rammeloo LA, Rozendaal L, Blom NA, van Lith JMM, Knegt AC, Hoffer MJV, Aten E, Santen GWE, Haak MC. The prevalence of genetic diagnoses in fetuses with severe congenital heart defects. *Genetics in Medicine*. 2020;22(7):1206–1214.
10. Zaidi S, Brueckner M. Genetics and Genomics of Congenital Heart Disease. *Circulation Research*. 2017;120:923–940.
11. Li Y, Klena NT, Gabriel GC, Liu X, Kim AJ, Lemke K, Chen Y, Chatterjee B, Devine W, Damerla RR, Chang C, Yagi H, San Agustin JT, Thahir M, Anderton S, et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature*. 2015;521(7553):520–524.
12. Williams K, Carson J, Lo C. Genetics of Congenital Heart Disease. *Biomolecules*. 2019;9(12):879.
13. Olson EN. Gene regulatory networks in the evolution and development of the heart. *Science (New York, N.Y.)*. 2006;313(5795):1922–7.
14. Andersen TA, Troelsen K de LL, Larsen LA. Of mice and men: molecular genetics of congenital heart disease. *Cellular and molecular life sciences*. 2014;71(8):1327–52.
15. Granados-Riveron JT, Pope M, Bu'Lock FA, Thornborough C, Eason J, Setchfield K, Ketley A, Kirk EP, Fatkin D, Feneley MP, Harvey RP, Brook JD. Combined mutation screening of

NKX2-5, GATA4, and TBX5 in congenital heart disease: Multiple heterozygosity and novel mutations. *Congenital Heart Disease*. 2012;7(2):151–159.

16. Sylva M, Van den Hoff MJB, Moorman AFM. Development of the human heart. *American Journal of Medical Genetics, Part A*. 2014;164(6):1347–1371.
17. Brade T, Pane LS, Moretti A, Chien KR, Laugwitz K-L. Embryonic Heart Progenitors and Cardiogenesis. *Cold Spring Harbor Perspectives in Medicine*. 2013;3(10):a013847–a013847.
18. Gittenberger-De Groot AC, Bartelings MM, Deruiter MC, Poelmann RE. Basics of cardiac development for the understanding of congenital heart malformations. *Pediatric Research*. 2005;57(2):169–176.
19. Niessen K, Karsan A. Notch Signaling in Cardiac Development. *Circulation Research*. 2008;102(10):1169–1181.
20. Lin CJ, Lin CY, Chen CH, Zhou B, Chang CP. Partitioning the heart: Mechanisms of cardiac septation and valve development. *Development (Cambridge)*. 2012;139(18):3277–3299.
21. Botto LD, Lin AE, Riehle-Colarusso T, Malik S, Correa A. Seeking causes: Classifying and evaluating congenital heart defects in etiologic studies. *Birth Defects Research Part A: Clinical and Molecular Teratology*. 2007;79(10):714–727.
22. Sifrim A, Hitz M-P, Wilsdon A, Breckpot J, Turki SHA, Thienpont B, McRae J, Fitzgerald TW, Singh T, Swaminathan GJ, Prigmore E, Rajan D, Abdul-Khaliq H, Banka S, Bauer UMM, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nature Genetics*. 2016;48(9):1–9.
23. Øyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PKA, Melbye M. Recurrence of congenital heart defects in families. *Circulation*. 2009;120(4):295–301.

24. Costain G, Silversides CK, Bassett AS. The importance of copy number variation in congenital heart disease. *npj Genomic Medicine*. 2016;1(1):16031.

25. Krishnan A, Samtani R, Dhanantwari P, Lee E, Yamada S, Shiota K, Donofrio MT, Leatherbury L, Lo CW. A detailed comparison of mouse and human cardiac development. *Pediatric research*. 2014;76(6):500–7.

26. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, Depalma SR, Zeng X, Qi H, Chang W, Sierant MC, Hung WC, Haider S, Zhang J, Knight J, Bjornson RD, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nature Genetics*. 2017;49(11):1593–1601.

27. Kodo K, Takahashi T, Nakanishi T, Yamagishi H, Makino S, Fukuda K, Oda M, Nishizawa T, Furutani M, Arai S, Matsuoka R, Ishihara K. Genetic analysis of essential cardiac transcription factors in 256 patients with non-syndromic congenital heart defects. *Circulation Journal*. 2012;76(7):1703–1711.

28. Montgomery RL, Davis CA, Potthoff MJ, Haberland M, Fielitz J, Qi X, Hill JA, Richardson JA, Olson EN. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes & Development*. 2007;21(14):1790–1802.

29. MacGrogan D, Münch J, de la Pompa JL. Notch and interacting signalling pathways in cardiac development, disease, and regeneration. *Nature Reviews Cardiology*. 2018;15(11):685–704.

30. Timmerman LA, Grego-Bessa J, Raya A, Bertrán E, Pérez-Pomares JM, Díez J, Aranda S, Palomo S, McCormick F, Izpisúa-Belmonte JC, De La Pompa JL. Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. *Genes and Development*. 2004;18(1):99–115.

31. Ackerman C, Locke AE, Feingold E, Reshey B, Espana K, Thusberg J, Mooney S, Bean LJH, Dooley KJ, Cua CL, Reeves RH, Sherman SL, Maslen CL. An excess of deleterious variants in VEGF-A pathway genes in down-syndrome-associated atrioventricular septal defects. *American Journal of Human Genetics*. 2012;91(4):646–659.
32. Rochais F, Mesbah K, Kelly RG. Signaling pathways controlling second heart field development. *Circulation Research*. 2009;104(8):933–942.
33. Bonachea EM, Zender G, White P, Corsmeier D, Newsom D, Fitzgerald-Butt S, Garg V, McBride KL. Use of a targeted, combinatorial next-generation sequencing approach for the study of bicuspid aortic valve. *BMC Medical Genomics*. 2014;7(1):56.
34. Chen H, Shi S, Acosta L, Li W, Lu J, Bao S, Chen Z, Yang Z, Schneider MD, Chien KR, Conway SJ, Yoder MC, Haneline LS, Franco D, Shou W. BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development*. 2004;131(9):2219–2231.
35. Dyer LA, Kirby ML. Sonic hedgehog maintains proliferation in secondary heart field progenitors and is required for normal arterial pole formation. *Developmental Biology*. 2009;330(2):305–317.
36. Washington Smoak I, Byrd NA, Abu-Issa R, Goddeeris MM, Anderson R, Morris J, Yamamura K, Klingensmith J, Meyers EN. Sonic hedgehog is required for cardiac outflow tract and neural crest cell development. *Developmental Biology*. 2005;283(2):357–372.
37. Liu J, Cheng H, Xiang M, Zhou L, Wu B, Moskowitz IP, Zhang K, Xie L. Gata4 regulates hedgehog signaling and Gata6 expression for outflow tract development. Conlon FL, ed. *PLoS Genetics*. 2019;15(5):e1007711.

38. Dyer LA, Makadia FA, Scott A, Pegram K, Hutson MR, Kirby ML. BMP signaling modulates hedgehog-induced secondary heart field proliferation. *Developmental Biology*. 2010;348(2):167–176.
39. Zhou Z, Wang J, Guo C, Chang W, Zhuang J, Zhu P, Li X. Temporally Distinct Six2-Positive Second Heart Field Progenitors Regulate Mammalian Heart Development and Disease. *Cell Reports*. 2017;18(4):1019–1032.
40. Goddeeris MM, Rho S, Petiet A, Davenport CL, Johnson GA, Meyers EN, Klingensmith J. Intracardiac septation requires hedgehog-dependent cellular contributions from outside the heart. *Development (Cambridge, England)*. 2008;135(10):1887–95.
41. Burnicka-Turek O, Steimle JD, Huang W, Felker L, Kamp A, Kweon J, Peterson M, Reeves RH, Maslen CL, Gruber PJ, Yang XH, Shendure J, Moskowitz IP. Cilia gene mutations cause atrioventricular septal defects by multiple mechanisms. *Human Molecular Genetics*. 2016;25(14):3011–3028.
42. Rexiati M, Sun M, Guo W. Muscle-Specific Mis-Splicing and Heart Disease Exemplified by RBM20. *Genes*. 2018;9(1):18.
43. Liu X, Yagi H, Saeed S, Bais AS, Gabriel GC, Chen Z, Peterson KA, Li Y, Schwartz MC, Reynolds WT, Saydmohammed M, Gibbs B, Wu Y, Devine W, Chatterjee B, et al. The complex genetics of hypoplastic left heart syndrome. *Nature Genetics*. 2017;49(7):1152–1159.
44. Winston JB, Erlich JM, Green CA, Aluko A, Kaiser KA, Takematsu M, Barlow RS, Sureka AO, Lapage MJ, Janss LL, Jay PY. Heterogeneity of genetic modifiers ensures normal cardiac development. *Circulation*. 2010;121(11):1313–1321.
45. Nora JJ. Multifactorial inheritance hypothesis for the etiology of congenital heart diseases. The genetic-environmental interaction. *Circulation*. 1968;38(3):604–617.

46. Mcbride KL, Pignatelli R, Lewin M, Ho T, Fernbach S, Menesses A, Lam W, Leal SM, Kaplan N, Schliekelman P, Towbin JA, Belmont JW. Inheritance Analysis of Congenital Left Ventricular Outflow Tract Obstruction Malformations: Segregation, Multiplex Relative Risk, and Heritability. *American Journal of Medical Genetics*. 2005;134:180–186.
47. Andrews T, Honti F, Pfundt R, Leeuw N de, Hehir-Kwa J, Silfhout AV, Vries B de, Webber C. The clustering of functionally related genes contributes to CNV-mediated disease. *Genome Research*. 2015;25(6):802–813.
48. Teekakirikul P, Zhu W, Gabriel GC, Young CB, Williams K, Martin LJ, Hill JC, Richards T, Billaud M, Phillippi JA, Wang J, Wu Y, Tan T, Devine W, Lin J, et al. Common deletion variants causing protocadherin-α deficiency contribute to the complex genetics of BAV and left-sided congenital heart disease. *Human Genetics and Genomics Advances*. 2021;2(3):100037.
49. De Luca A, Sarkozy A, Consoli F, Ferese R, Guida V, Dentici ML, Mingarelli R, Bellacchio E, Tuo G, Limongelli G, Digilio MC, Marino B, Dallapiccola B. Familial transposition of the great arteries caused by multiple mutations in laterality genes. *Heart*. 2010;96(9):673–677.
50. Gifford CA, Ranade SS, Samarakoon R, Salunga HT, Soysa TY de, Huang Y, Zhou P, Elfenbein A, Wyman SK, Bui YK, Metzler KRC, Ursell P, Ivey KN, Srivastava D. Oligogenic inheritance of congenital heart disease involving a NKX2-5 modifier. *Science*. 2019;364:865–870.
51. Priest JR, Osoegawa K, Mohammed N, Nanda V, Kundu R, Schultz K, Lammer EJ, Girirajan S, Scheetz T, Waggott D, Haddad F, Reddy S, Bernstein D, Burns T, Steimle JD, et al. De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. Jiao K, ed. *PLoS Genetics*. 2016;12(4):e1005963.

52. Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. *Nat Rev Genet*. 2014;15(11):722–733.

53. Lage K, Greenway SC, Rosenfeld JA, Wakimoto H, Gorham JM, Segre AV, Roberts AE, Smoot LB, Pu WT, C. Pereira A, Mesquita SM, Tommerup N, Brunak S, Ballif BC, Shaffer LG, et al. Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. *Proceedings of the National Academy of Sciences*. 2012;109(35):14035–14040.

54. Pediatric Cardiac Genomics Consortium. The Congenital Heart Disease Genetic Network Study. *Circulation Research*. 2013;112(4):698–706.

55. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK, Carriero NJ, Cheung YH, Deanfield J, Depalma S, Fakhro KA, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013;498(7453):220–223.

56. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, DePalma SR, McKean D, Wakimoto H, Gorham J, Jin SC, Deanfield J, Giardini A, Porter GA, Kim R, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*. 2015;350(6265):1262–6.

57. Lage K, Møllgård K, Greenway S, Wakimoto H, Gorham JM, Workman CT, Bendsen E, Hansen NT, Rigina O, Roque FS, Wiese C, Christoffels VM, Roberts AE, Smoot LB, Pu WT, et al. Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Molecular systems biology*. 2010;6(1):381.

58. Sevim Bayrak C, Zhang P, Tristani-Firouzi M, Gelb BD, Itan Y. De novo variants in exomes of congenital heart disease patients identify risk genes and pathways. *Genome Medicine*. 2020;12(1):9.

59. Gabriel GC, Young CB, Lo CW. Role of cilia in the pathogenesis of congenital heart disease. *Seminars in Cell & Developmental Biology*. 2021;110:2–10.

60. Toriyama M, Lee C, Taylor SP, Duran I, Cohn DH, Bruel A-L, Tabler JM, Drew K, Kelly MR, Kim S, Park TJ, Braun D, Pierquin G, Biver A, Wagner K, et al. The ciliopathy-associated CPLANE proteins direct basal body recruitment of intraflagellar transport machinery. *Nature Genetics*. 2016;48(April).

61. Nonaka S, Tanaka Y, Okada Y, Takeda S, Harada A, Kanai Y, Kido M, Hirokawa N. Randomization of Left–Right Asymmetry due to Loss of Nodal Cilia Generating Leftward Flow of Extraembryonic Fluid in Mice Lacking KIF3B Motor Protein. :9.

62. Shinohara K, Hamada H. Cilia in Left–Right Symmetry Breaking. *Cold Spring Harbor Perspectives in Biology*. 2017:a028282.

63. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. *Nature Reviews Molecular Cell Biology*. 2017;18(9):533–547.

64. Avasthi P, Marshall WF. Stages of Ciliogenesis and Regulation of Ciliary Length. *Differentiation; Research in Biological Diversity*. 2012;83(2):S30–S42.

65. Lu Q, Insinna C, Ott C, Stauffer J, Pintado PA, Rahajeng J, Baxa U, Walia V, Cuenca A, Hwang Y-S, Daar IO, Lopes S, Lippincott-Schwartz J, Jackson PK, Caplan S, et al. Early steps in primary cilium assembly require EHD1/EHD3-dependent ciliary vesicle formation. *Nature Cell Biology*. 2015;17(3):228–240.

66. Reiter JF, Mostov K. Vesicle transport, cilium formation, and membrane specialization: The origins of a sensory organelle. *Proceedings of the National Academy of Sciences*. 2006;103(49):18383–18384.

67. Yang H, Huang K. Dissecting the Vesicular Trafficking Function of IFT Subunits. *Frontiers in Cell and Developmental Biology*. 2020;7:352.

68. Koefoed K, Veland IR, Bang L, Lars P, Larsen A, Christensen ST. Cilia and coordination of signaling networks during heart development. 2014;10(1):108–125.

69. Wheway G, Nazlamova L, Hancock JT. Signaling through the Primary Cilium. *Frontiers in Cell and Developmental Biology*. 2018;6:8.

70. Watkins WS, Hernandez EJ, Wesolowski S, Bisgrove BW, Sunderland RT, Lin E, Lemmon G, Demarest BL, Miller TA, Bernstein D, Brueckner M, Chung WK, Gelb BD, Goldmuntz E, Newburger JW, et al. De novo and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes. *Nature Communications*. 2019;10(1):1–12.

71. Klink BU, Gatsogiannis C, Hofnagel O, Wittinghofer A, Raunser S. Structure of the human BBSome core complex. *eLife*. 9:e53910.

72. Taschner M, Lorentzen E. The Intraflagellar Transport Machinery. *Cold Spring Harbor Perspectives in Biology*. 2016;8(10):a028092.

73. Bhogaraju S, Engel BD, Lorentzen E. Intraflagellar transport complex structure and cargo interactions. *Cilia*. 2013;2(1):10.

74. Kousi M, Söylemez O, Ozanturk A, Mourtzi N, Akle S, Jungreis I, Muller J, Cassa CA, Brand H, Mokry JA, Wolf MY, Sadeghpour A, McFadden K, Lewis RA, Talkowski ME, et al. Evidence for secondary-variant genetic burden and non-random distribution across biological modules in a recessive ciliopathy. *Nature Genetics*. 2020;52(11):1145–1150.

75. Akhirome E, Walton NA, Nogee JM, Jay PY. The Complex Genetic Basis of Congenital Heart Defects. *Circulation Journal*. 2017;81(5):629–634.

76. Gelb BD, Chung WK. Complex genetics and the etiology of human congenital heart disease. *Cold Spring Harbor Perspectives in Medicine*. 2014;4(7).

77. Beecham GW, Bis JC, Martin ER, Choi S-H, DeStefano AL, van Duijn CM, Fornage M, Gabriel SB, Koboldt DC, Larson DE, Naj AC, Psaty BM, Salerno W, Bush WS, Foroud TM, et al. The Alzheimer's Disease Sequencing Project: Study design and sample selection. *Neurology Genetics*. 2017;3(5):e194.

78. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.

79. Auwera GAV der, Carneiro MO, Hartl C, Poplin R, Angel G del, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*. 2013;43(1):11.10.1-11.10.33.

80. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biology*. 2016;17(1):122.

81. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*. 2007;35(Database issue):D61-5.

82. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill andrew J, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285–291.

83. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*. 2016;44(D1):D862–D868.

84. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 2009;4(7):1073–1082.

85. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics*. 2013;76(1):7.20.1-7.20.41.

86. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. 2014;46(3):310–315.

87. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology*. 2010;6(12).

88. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M, Kang HM. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *American journal of human genetics*. 2012;91(5):839–48.

89. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867–2873.

90. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007;81(3):559–75.

91. Raghavan NS, Brickman AM, Andrews H, Manly JJ, Schupf N, Lantigua R, Wolock CJ, Kamalakaran S, Petrovski S, Tosto G, Vardarajan BN, Goldstein DB, Mayeux R, The Alzheimer's Disease Sequencing Project. Whole-exome sequencing in 20,197 persons for rare variants in Alzheimer's disease. *Annals of Clinical and Translational Neurology*. 2018;5(7):832–842.
92. Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Research*. 2018;46(D1):D836–D842.
93. Szot Justin O., Cuny Hartmut, Blue Gillian M., Humphreys David T., Ip Eddie, Harrison Katrina, Sholler Gary F., Giannoulatou Eleni, Leo Paul, Duncan Emma L., Sparrow Duncan B., Ho Joshua W.K., Graham Robert M., Pachter Nicholas, Chapman Gavin, et al. A Screening Approach to Identify Clinically Actionable Variants Causing Congenital Heart Disease in Exome Data. *Circulation: Genomic and Precision Medicine*. 2018;11(3):e001978.
94. Alankarage D, Ip E, Szot JO, Munro J, Blue GM, Harrison K, Cuny H, Enriquez A, Troup M, Humphreys DT, Wilson M, Harvey RP, Sholler GF, Graham RM, Ho JWK, et al. Identification of clinically actionable variants from genome sequencing of families with congenital heart disease. *Genetics in Medicine*. 2019;21(5):1111–1120.
95. Dam TJP van, Kennedy J, Lee R van der, Vrieze E de, Wunderlich KA, Rix S, Dougherty GW, Lambacher NJ, Li C, Jensen VL, Leroux MR, Hjeij R, Horn N, Texier Y, Wissinger Y, et al. CiliaCarta: An integrated and validated compendium of ciliary genes. *PLOS ONE*. 2019;14(5):e0216705.
96. Ishikawa H, Thompson J, Yates JR, Marshall WF. Proteomic Analysis of Mammalian Primary Cilia. *Current Biology*. 2012;22(5):414–419.

97. Mick DU, Rodrigues RB, Leib RD, Adams CM, Chien AS, Gygi SP, Nachury MV. Proteomics of Primary Cilia by Proximity Labeling. *Developmental Cell*. 2015;35(4):497–512.

98. Sang L, Miller JJ, Corbit KC, Giles RH, Brauer MJ, Otto EA, Baye LM, Wen X, Scales SJ, Kwong M, Huntzicker EG, Sfakianos MK, Sandoval W, Bazan JF, Kulkarni P, et al. Mapping the NPHP-JBTS-MKS Protein Network Reveals Ciliopathy Disease Genes and Pathways. *Cell*. 2011;145(4):513–528.

99. Garcia-Gonzalo FR, Corbit KC, Sirerol-Piquer MS, Ramaswami G, Otto EA, Noriega TR, Seol AD, Robinson JF, Bennett CL, Josifova DJ, García-Verdugo JM, Katsanis N, Hildebrandt F, Reiter JF. A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nature Genetics*. 2011;43(8):776–784.

100. Wang L, Gu L, Meng D, Wu Q, Deng H, Pan J. Comparative Proteomics Reveals Timely Transport into Cilia of Regulators or Effectors as a Mechanism Underlying Ciliary Disassembly. *Journal of Proteome Research*. 2017;16(7):2410–2418.

101. Blackburn K, Bustamante-Marin X, Yin W, Goshe MB, Ostrowski LE. Quantitative Proteomic Analysis of Human Airway Cilia Identifies Previously Uncharacterized Proteins of High Abundance. *Journal of Proteome Research*. 2017;16(4):1579–1592.

102. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. Tang H, ed. *PLoS Computational Biology*. 2015;11(4):e1004219.

103. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*. 2019;10(1):1523.

104. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-

controlled protein-protein association networks, made broadly accessible. *Nucleic acids research*. 2017;45(D1):D362–D368.

105. Asp M, Giacomello S, Larsson L, Wu C, Fürth D, Qian X, Wärdell E, Custodio J, Reimegård J, Salmén F, Österholm C, Ståhl PL, Sundström E, Åkesson E, Bergmann O, et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell*. 2019;179(7):1647-1660.e19.

106. Helle E, Córdova-Palomera A, Ojala T, Saha P, Potiny P, Gustafsson S, Ingelsson E, Bamshad M, Nickerson D, Chong JX, Ashley E, Priest JR. Loss of function, missense, and intronic variants in NOTCH1 confer different risks for left ventricular outflow tract obstructive heart defects in two European cohorts. *Genetic Epidemiology*. 2019;43(2):215–226.

107. Girdauskas E, Geist L, Disha K, Kazakbaev I, Groß T, Schulz S, Ungelenk M, Kuntze T, Reichenspurner H, Kurth I. Genetic abnormalities in bicuspid aortic valve root phenotype: Preliminary results. *European Journal of Cardio-thoracic Surgery*. 2017;52(1):156–162.

108. Iascone M, Ciccone R, Galletti L, Marchetti D, Seddio F, Lincesso AR, Pezzoli L, Vetro A, Barachetti D, Boni L, Federici D, Soto AM, Comas JV, Ferrazzi P, Zuffardi O. Identification of de novo mutations and rare variants in hypoplastic left heart syndrome. *Clinical Genetics*. 2012;81(6):542–554.

109. Zahavich L, Bowdin S, Mital S. Use of Clinical Exome Sequencing in Isolated Congenital Heart Disease. *Circulation: Cardiovascular Genetics*. 2017;10(3).

110. Panizzi JR, Becker-Heck A, Castleman VH, Al-Mutairi DA, Liu Y, Loges NT, Pathak N, Austin-Tse C, Sheridan E, Schmidts M, Olbrich H, Werner C, Häffner K, Hellman N, Chodhari R, et al. CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. *Nature Genetics*. 2012;44(6):714–719.

111. Tarkar A, Loges NT, Slagle CE, Francis R, Dougherty GW, Tamayo JV, Shook B, Cantino M, Schwartz D, Jahnke C, Olbrich H, Werner C, Raidt J, Pennekamp P, Abouhamed M, et al. DYX1C1 is required for axonemal dynein assembly and ciliary motility. *Nature Genetics*. 2013;45(9):995–1003.

112. Coban-Akdemir Z, White JJ, Song X, Jhangiani SN, Fatih JM, Gambin T, Bayram Y, Chinn IK, Karaca E, Punetha J, Poli C, Boerwinkle E, Shaw CA, Orange JS, Gibbs RA, et al. Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. *The American Journal of Human Genetics*. 2018;103(2):171–187.

113. Stephen LA, Tawamie H, Davis GM, Tebbe L, Nürnberg P, Nürnberg G, Thiele H, Thoenes M, Boltshauser E, Uebe S, Rompel O, Reis A, Ekici AB, McTeir L, Fraser AM, et al. TALPID3 controls centrosome and cell polarity and the human ortholog KIAA0586 is mutated in Joubert syndrome (JBTS23). Dietz HC, ed. *eLife*. 2015;4:e08077.

114. Mukherjee S, Cogan JD, Newman JH, Phillips JA, Hamid R, Meiler J, Capra JA. Identifying digenic disease genes via machine learning in the Undiagnosed Diseases Network. *The American Journal of Human Genetics*. 2021;108(10):1946–1963.

115. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterology and Hepatology From Bed to Bench*. 2014;7(1):17–31.

116. Boucher B, Jenna S. Genetic interaction networks: Better understand to better predict. *Frontiers in Genetics*. 2013;4(DEC):290.

117. Gonzalez-Teran B, Pittman M, Felix F, Richmond-Buccola D, Thomas R, Choudhary K, Moroni E, Colombo G, Alexanian M, Cole B, Samse-Knapp K, McGregor M, Gifford CA,

Huttenhain R, Gelb BD, et al. Integration of Protein Interactome Networks with Congenital Heart Disease Variants Reveals Candidate Disease Genes. *bioRxiv*. 2021:2021.01.05.423837.

118. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*. 2006;34(Database issue):D535–D539.

119. Zhang R, Ren Z, Celedón JC, Chen W. Inference of Large Modified Poisson-type Graphical Models: Application to RNA-seq Data in Childhood Atopic Asthma Studies. *The Annals of Applied Statistics*. 2021 Jun;15(2):831-55.

120. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Physical Review E*. 2004;69(2):026113.

121. Wang J, Zhao J-Y, Wang F, Peng Q-Q, Hou J, Sun S-N, Gui Y-H, Duan W-Y, Qiao B, Wang H-Y. A genetic variant in vitamin B12 metabolic genes that reduces the risk of congenital heart disease in Han Chinese populations. *PloS One*. 2014;9(2):e88332.

122. Fulmer D, Toomer K, Guo L, Moore K, Glover J, Moore R, Stairley R, Lobo G, Zuo X, Dang Y, Su Y, Fogelgren B, Gerard P, Chung D, Heydarpour M, et al. Defects in the Exocyst-Cilia Machinery Cause Bicuspid Aortic Valve Disease and Aortic Stenosis. *Circulation*. 2019;140(16):1331–1341.

123. Mlynarski EE, Xie M, Taylor D, Sheridan MB, Guo T, Racedo SE, McDonald-McGinn DM, Chow EWC, Vorstman J, Swillen A, Devriendt K, Breckpot J, Digilio MC, Marino B, Dallapiccola B, et al. Rare copy number variants and congenital heart defects in the 22q11.2 deletion syndrome. *Human Genetics*. 2016;135(3):273–285.

124. Glessner JT, Bick AG, Ito K, Homsy JG, Rodriguez-Murillo L, Fromer M, Mazaika E, Vardarajan B, Italia M, Leipzig J, DePalma SR, Golhar R, Sanders SJ, Yamrom B, Ronemus

M, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circulation Research*. 2014;115(10):884–896.

125. Quintero-Rivera F, Xi QJ, Keppler-Noreuil KM, Lee JH, Higgins AW, Anchan RM, Roberts AE, Seong IS, Fan X, Lage K, Lu LY, Tao J, Hu X, Berezney R, Gelb BD, et al. MATR3 disruption in human and mouse associated with bicuspid aortic valve, aortic coarctation and patent ductus arteriosus. *Human Molecular Genetics*. 2015;24(8):2375–2389.
126. Abu-Halima M, Wagner V, Becker LS, Ayesh BM, Abd El-Rahman M, Fischer U, Meese E, Abdul-Khaliq H. Integrated microRNA and mRNA Expression Profiling Identifies Novel Targets and Networks Associated with Ebstein's Anomaly. *Cells*. 2021;10(5):1066.
127. Morton PD, Ishibashi N, Jonas RA. Neurodevelopmental Abnormalities and Congenital Heart Disease: Insights into Altered Brain Maturation. *Circulation Research*. 2017;120(6):960–977.
128. Kasatwar A, Borle R, Bhola N, K R, Prasad GSV, Jadhav A. Prevalence of congenital cardiac anomalies in patients with cleft lip and palate – Its implications in surgical management. *Journal of Oral Biology and Craniofacial Research*. 2018;8(3):241–244.
129. Egbe A, Uppu S, Lee S, Ho D, Srivastava S. Prevalence of Associated Extracardiac Malformations in the Congenital Heart Disease Population. *Pediatric Cardiology*. 2014;35(7):1239–1245.
130. Hinton RB, Andelfinger G, Sekar P, Hinton AC, Gendron RL, Michelfelder EC, Robitaille Y, Benson DW. Prenatal Head Growth and White Matter Injury in Hypoplastic Left Heart Syndrome. *Pediatric research*. 2008;64(4):364–369.

131. Hangge PT, Cnota JF, Woo JG, Hinton AC, Divanovic AA, Manning PB, Ittenbach RF, Hinton RB. Microcephaly is associated with early adverse neurologic outcomes in hypoplastic left heart syndrome. *Pediatric Research*. 2013;74(1):61–67.