Causal Varian discovery in Familial Congenital Heart Disease - An Integrative -Omic Approach

Wendy Demos
Marquette University

CAUSAL VARIANT DISCOVERY IN FAMILIAL CONGENITAL HEART DISEASE – AN INTEGRATIVE –OMIC APPROACH

by

Wendy M. Demos

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

May 2012

ABSTRACT
CAUSAL VARIANT DISCOVERY IN FAMILIAL CONGENITAL HEART DISEASE – AN INTEGRATIVE –OMIC APPROACH

Wendy M. Demos

Marquette University, 2012

Background:

Hypoplastic left heart syndrome (HLHS) is a congenital heart defect that leads to neonatal death or compromised quality of life for those affected and their families. This syndrome requires extensive medical intervention for the affected to survive. It is characterized by significant underdevelopment or non-existence of the components of the left heart and the aorta, including the left ventricular cavity and mass. There are many factors ranging from genetics to environmental relationships hypothesized to lead to the development of the syndrome, including recent studies suggesting a link between hearing impairment and congenital heart defects (CHD). Although broadly characterized those factors remain poorly understood. The goal of this project is to systematically utilize bioinformatics tools to determine the relationships of novel mutations found in exome sequencing to a familial congenital heart defect.

Methods

A systematic genomic and proteomic approach involving exome sequencing, pathway analysis, and protein modeling was implemented to examine exome sequencing data of a patient with HLHS. Subsequent findings were examined in immediate family members and relatives to investigate inheritance and validate the relationship of novel variants to CHD in the family.

Conclusions

A rare, novel mutation in the LAMA4 gene carried by the proband and other family members may contribute to the development of HLHS in this family.

ACKNOWLEDGMENTS

Wendy M. Demos

I would like to thank the family for their willingness to be put under the microscope in order to better understand HLHS. My committee members: Dr. Craig Struble, for his support and willingness to see this through to the end with me, Dr. Aoy Tomita-Mitchell for the opportunity to join the study and for continually inspiring me with her positive encouragement, and Dr. Francis Peterson for taking me on late in the study and showing me the ropes of protein modeling with great patience.

I would also like to thank the institutions and organizations the made this work possible: Marquette University, The Medical College of Wisconsin, the MCW Department of Surgery, the Congenital Heart Disease Tissue Bank, the Wisconsin Pediatric Cardiac Registry, Families of the Herma Heart Center, Children's Research Institute, the Human and Molecular Genetics Center, and the Biotechnology and Bioengineering Center.

I would especially like to thank my husband, Chris, for his unlimited patience, encouragement, support, and humor through this process. This would not have been possible without you. My son, Evan, for reminding me how fragile, precious, and wonderful life is. My parents, Jim and Mary Lou my brother Wade, sister-in law Sara, mother-in-law Vivian, sister-in-law Michelle, brother-in-law Dominic, and all my nephews and nieces for their encouragement, graciously understanding of abbreviated family visits, babysitting, and support.

Members of the Mitchell lab, particularly: Mary Goetsch for help with PCR and sequencing reactions and her scientific expertise, Huan-Ling Liang for her assistance with PCR and sequencing reactions, Lisa Armitage for resourcefulness and assistance with scheduling, and Karl Stamm for providing bioinformatics insight and conducting haplotype and EVS analysis.

I would like to thank my friends who have provided much understanding and encouragement over the last few years especially Dr. Andrea Ferrante for additional help with protein modeling and scientific advice, Praful Aggarwal for much bioinformatics feedback and support, and Ejan and Rob Tyler for protein modeling insight and nights of keeping the balance.

It is impossible to mention everyone that has supported me through this endeavor and there are not enough words to convey my gratitude. I can only very simply say "Thank you".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 DNA Sequencing

The ability of researchers to comprehensively examine disease etiology at the genomic level has been revolutionized by the development of high throughput DNA sequencing technologies such as 454 pyrosequencing (Roche Diagnostics), HighSeq (Illuminia, Inc.), SOLiD (Applied Biosystems), and Ion semiconductor sequencing (Ion Torrent Systems Inc.) These technologies have drastically reduced DNA sequencing time and costs by generating massive quantities of data in a fraction of the time of traditional Sanger sequencing. Researchers and clinicians have capitalized on these technologies to further our understanding of the human genome and causes of disease. The field of 'genomic medicine', which utilizes the sequence of an individual's exome (coding regions of genes) has transformed the capability of modern medicine to diagnose and prescribe personalized treatment methods.

The chain terminator method of DNA sequencing (commonly referred to as the Sanger sequencing method) utilizes a DNA polymerase that adds nucleotides to generate a complementary strand of DNA to a single stranded DNA template. Applied Biosystems automated this type of DNA sequencing by creating machines that could run individual reactions in a single column rather than four simultaneous reactions on a polyacrylamide gel. These machines provided a visual readout of the DNA sequence called a chromatogram. At the onset of sequence automation a run of 24 samples would yield 12,000 bases of DNA per day (Adams 2008). Currently the Applied Biosystems 3130*xl* can generate more than 121,600 bases per day. (Applied Biosystems).

In comparison, pyrosequencing can yield 1,000,000,000 bases or 400,000 sequencing reads per run (Roche Diagnostics). Pyrosequencing is based on a method that measures the release of inorganic pyrophosphate, which is converted into visible light (Metzker 2005). A series of dNTPs are added to the reaction and a camera records emitted light after each round. The pattern of captured light is translated into the DNA sequence.

Unlike Sanger sequencing, 454 sequencing avoids the labor and time extensive step of cloning the DNA template into a vector. Roche 454 sequencing employs an emulsion PCR in which a single molecule of DNA binds to a single magnetic bead. Many beads are processed in a water and oil mixture which in turn isolates each bead into its own water droplet. The amplification process runs all beads in parallel and results in about 10 million clonal copies of a single DNA fragment, each bead holding a different fragment (Roche Diagnostics). This step alone vastly increases the starting quantity of DNA template over traditional Sanger sequencing.

Filtering through the resulting 454 sequence data to find allelic variations that may be related to a specific phenotype is a daunting task. The DNA Sequences returned from the exome sequencing reaction are aligned to a reference genome. This comparison aids in finding nucleotide differences from the patient sequence to the human genome reference sequence. These differences are termed *variant*. Some variants are not tolerated and presumably cause or contribute to the development of disease. A recent investigation of the case of a 15 month old male with a seemingly undiagnosable severe health condition supports the relationship of sequence variation to disease. Genomic DNA from the patient was subjected to whole exome sequencing. The sequencing reaction returned over 16,000 variants. Due to the uncommon presentation of symptoms and condition bioinformaticians focused on variants not found in the general population (novel) to determine which changes in the patient DNA sequence may have been linked to the condition under investigation. This approach ultimately led to a definitive diagnosis the patient carries a rare mutation in an apoptotic suppressor protein (XIAP) resulting in immunodeficiency (Worthey et. al. 2011).

Establishing prospective gene candidates and conducting the subsequent investigations is an extremely time consuming task as a result of the number of novel variants yielded by the sequencing method, future diagnosis, preventative measures, genetic testing, or advancements in treatment depend on an efficient approach to determine or rule out causal or contributory genes related to disease.

1.2 Congenital Heart Defects

A congenital heart defect (CHD) is described as a structural condition present at birth that can affect the heart and the way it works. CHD can affect the interior walls and valves of the heart and the arteries and veins that carry blood to the heart and body (Knapp et. al. 2010). There are many types of congenital heart defects which can be asymptomatic or complex with life-threatening symptoms (NHLBI). CHD affects 1 in 100 live births and is considered one of the most common types of birth defects. The incidence of moderate to severe forms is approximately 1% of the newborn population (Hoffman 1995, Hoffman et. al. 2002, Sander et. al. 2006). From 1999 to 2006, there were 41,494 CHD-related deaths and 27,960 deaths resulting from CHD in the United States (Gilboa et. al. 2010).

A sub category of CHD is critical congenital heart defect (CCHD). CCHD is a group of ductal dependent defects that cause severe and life-threatening symptoms that require intervention within the first year of life (Knapp et. al. 2010, Chang et. al. 2008, CDC).Failure to intervene within the first few days of life results in a fatal outcome (Barron et. al. 2009). The seven defects that are categorized as CCHD (Table 1) represent about 17-31% of all congenital heart defects (Knapp et. al. 2010).

Seven Defects Categorized as CCHD	
Hypoplastic left heart syndrome	Total anomalous pulmonary venous return
Pulmonary atresia (with intact septum)	Transposition of the great arteries
Tetralogy of Fallot	Tricuspid atresia
Truncus arteriosus	

Table 1. Types of CCHD. The seven types of CCHD according to the CDC in the context of newborn screening using pulse oximetry (Data retrieved from: http://www.cdc.gov/ncbddd/pediatricgenetics/pulse.html)

1.3Hypoplastic Left Heart Syndrome

Hypoplastic left heart syndrome (HLHS) is a classification of CCHD that refers to cardiac malformations characterized by significant underdevelopment or non-existence of the components of the left heart and the aorta, including the left ventricular cavity and mass

(Tchervenkov et. al. 2006). The left heart is responsible for moving oxygenated blood to the aorta which then carries and distributes it to the rest of the body (Fogoros 2003). Defects in this system can be severely debilitating and / or fatal to affected individuals.

Until recently, survival of the first five years of life in a person diagnosed with HLHS ranged between 50% and 69%. With improved diagnostics and management it is now expected that 70% of HLHS patients will survive into adulthood (Feinstein et. al. 2012). Motor or cognitive delays characterized by impaired social skills, hampered core communications skills, impulsive behavior, and among other related symptoms are associated to HLHS. Patients, parents, and clinicians affected by HLHS have participated in surveys to measure their and the affected individual's perceived quality of life. Findings have been inconsistent due to methodology of studies and the stress of dealing with a chronic illness (Feinstein et. al. 2012).

HLHS is described as a genetically inherited syndrome although a definitive method of inheritance has not yet been defined. The prevalence of HLHS is not linked to any race or gender; however, males are twice as commonly affected as females (Stumper 2010, CDC 2007). It is known there is an increased frequency in relatives of patients with HLHS to also be affected (Loffredo 2000, Lewin et. al. 2004). 55% of families had more than 1 affected family member and the recurrence rate for HLHS among siblings is 8% (Grossfeld 2007)

The etiology of HLHS has broadly been studied and the consensus is that the syndrome has multifactorial causes (Sander et. al. 2006). Linkage analysis suggesting genetic heterogeneity, mutations at multiple genetic loci, seasonal effects, chromosomal disorders, and environmental conditions also have been described as potential modes of the development of HLHS (Eghtesady et. al. 2011, Grossfeld 2007, Grossfeld 1999, McBride et. al. 2008, Aberdam et. al.2002). The hypothesis of multiple genetic loci is further supported by the strong association of HLHS with conditions such as Trisomy 13, Trisomy 18, Turner syndrome, and Jacobson syndrome (Grossfeld et. al. 2009).

1.4 Objective

At this time there is no gold standard or pipeline to manage these data and rapidly determine specific variations in relation to the disease under investigation. The goal of this thesis is to execute a systematic and efficient approach that incorporates genomic and proteomic analysis to investigate genomic variants found in exome sequencing of a patient (proband) with HLHS. This approach will incorporate the use of bioinformatic tools to: design PCR primers, align DNA sequences, determine mutation effects, conduct pathway analysis, and construct protein models with laboratory methods such as PCR to narrow the study set of genes and ultimately determine the path of inheritance and potential relationship of the novel variant or variants to CHD.

A mutation was found in the LAMA4 gene of the proband. The application of the described process provides evidence suggesting a potential causal or contributory relationship in regards to the development of HLHS in this family. Based on previous research that HLHS is a mutligeneic syndrome and the family history, the relationship of LAMA4 and GJB2 will also be investigated to determine if together they are linked to the development of HLHS in this family.

1.5 Outline of Thesis

Chapter 2 provides in depth background of the genes and proteins that were the focus of the study. Chapter 3 covers the applied approach and methods leading to the selection candidate genes for investigation. Chapter 4 presents the results and discussion. Chapter 5 summarizes the thesis and suggests future work.

CHAPTER 2. BACKGROUND

After the application of the research method, the LAMA4 gene was selected as the main gene of interest to study. LAMA4 is a member of the laminin family. Section 2.1 will provide a summary of laminin function, structure, and disease associations. Section 2.2 will specifically give a description of the LAMA4 gene, the main focus of the study. Section 2.3 will give a brief synopsis of the relationship of genetic disorders that present with hearing impairment and cardiovascular abnormalities. Section 2.4 will introduce gap junctions and the known functions and disease associations. Section 2.5describes the connection of GJB2 to inherited deafness.

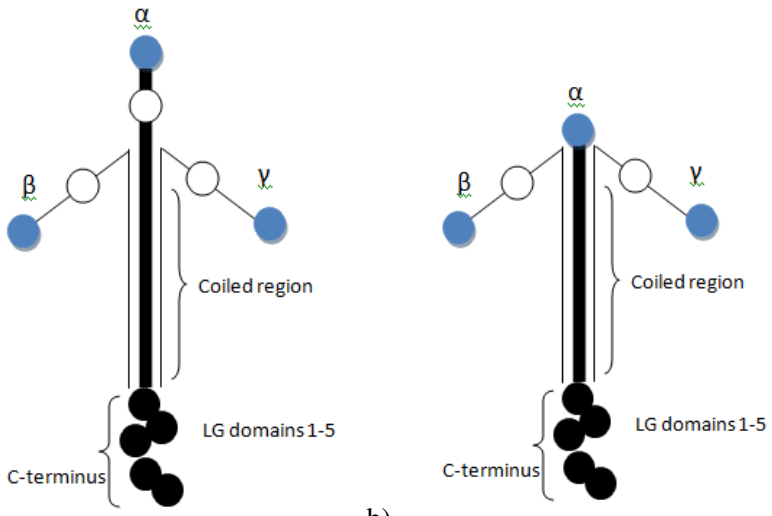
2.1 Laminin Family

Laminins are a family of extracellular matrix glycoproteins. They are essential for early embryonic development and organogenesis and have crucial functions in several tissues including muscle, nerve, skin, kidney, lung, and the vasculature (Durbeej 2010). They are the most abundant structural non-collagenous glycoprotein present in basement membranes and vital to cell recognition, adhesion, migration, and proliferation (Aberdam et. al. 2000).

Basement membranes are essential to the basic structure and mechanics of cells and tissues, forming barriers between cell types, and regulating interactions with other cells (Suzuki et. al. 2005). They are thin sheets of extracellular matrix that underlie epithelial and endothelial cells and surround muscle cells, Schwann cells, and fat cells. Expression of laminins is tissue specific and heavily determined by variations in the expression of the alpha chains (Durbeej 2010).

Laminins are composed of three chains: α, β, and γ. Each chain is encoded by its own gene: LAMA (α), LAMB (β), or LAMC (γ). To date, five α (named α 1-5), three β (named β 1-3), and three γ (named γ 1-3) chains have been identified. These chains assemble to create a cross like laminin structure. Thus far 18alpha, beta, and gamma combinations (named laminin-1 to 18) have been described (Durbeej 2010). The basic structures of the isoforms are relatively similar (Figure 1). An alpha, beta, and gamma chain each make up a unique short arm of the protein. The

chains coil together to form a long rigid rod-like structure creating a cruciform structure (Iivanainen et. al.1995). N-terminal domains, which are important for laminin network formation (Van Agtmael et. al. 2010) are present at the end of each short arm. The C-terminal domain of the alpha chain is preceded by the coiled region. The alpha chain C-terminal domain consists of five globular domains known as the laminin G (LG) domain. These globular domains are responsible for much of the binding interactions with cellular receptors. Eight laminin isoforms that contain α3A, α4, β3, and / or γ2 have a truncated arm due to a shorted N-terminus (Figure 1).



a) b)

Figure 1. Basic structure of a laminin. Blue circles represent the N-terminal domains of each chain. White circles are other globular domains. The black circles at the base of the figure represent the LG domain. G domains are numbered sequentially (LG1-5) from the base of the coiled region. a) General laminin structure. b) Laminin composed with a truncated alpha chain (Model adapted from Belkin et. al. 2000, Tunggal et. al. 2000, Van Agtmael et. al. 2010, Durbeej 2010)

Interactions with laminins are detrimental for basement membrane formation, tissue integrity, and biological processes. The LG domains bind to many proteins: heparin and sulfatides presumably bind the LG domains of all α chains (Durbeej 2010). The three main cellular receptors for laminins are integrins, dystroglycan, and syndecans. (Suzuki et. al. 2005, Durbeej 2010). Laminin - integrin binding regulates cell motility, survival, and proliferation. An inability of laminins to interact with syndecans compromises basement membrane formation in development.

Several diseases of the skin, muscle, and nerve are associated with laminins, either attributed to a defect in the laminin itself or the cellular receptors. Congenital muscular dystrophy is caused by the lack of laminin and alpha-dystroglycan-glycoprotein complex binding due to complete or near-complete deficiencies in laminin alpha2 (Suzuki et. al. 2005, McGowan et. al. 2000). Junctional epidermolysis bullosa (JEB) causes blistering of the skin in response to trauma or friction and severe cases can be fatal in first year of life. It is caused by mutations in the three genes LAMA3, LAMB3, and LAMC2 that are responsible for encoding the protein chains of laminin-5 (McGowan et. al. 2000). Mutations in LAMB2 leading to a deficiency of the laminin beta 2 chain contribute to the development of Pierson syndrome which causes ocular anomalies contributing to renal failure and loss of vision (VanDeVoorde et. al. 2006, Van Agtmael et. al. 2010). Relationships of LAMA1 and LAMA4 with rheumatoid arthritis and osteoarthritis have been examined (Salmon 1984, Fuerst et. al. 2010). Additionally, the relationship with laminin and Alzheimer's disease is currently under investigation (McGowan et. al. 2000).

In summary, the laminin family is not only important to the structure and organization of muscle and tissue. Defects in the genes chains can produce debilitating or even fatal functional problems.

2.2 LAMA4

The human LAMA4 gene encodes the protein laminin alpha 4. It is located on cytogenetic band 6q21. Although countless studies have shed light on the functions of laminin alpha 4, its exact function has eluded researchers (Stelzer et. al. 2012). It is predominantly located in the endothelial basement membranes of blood vessels. The main expression sites include: smooth muscle, fat, peripheral nerve neuromuscular junction, retina, and the central nervous system (Durbeej 2010). Research has indicated that LAMA4 is also strongly expressed in the adult heart, lung, ovary, small and large intestines, placenta, and liver (Iivanainen et. al. 1995). Mouse studies have been conducted to determine expression and function. These studies show it is also highly expressed in the aortic endothelium, endocardium and endothelium of blood vessels

(Freiser 1997), and cardiac muscle cells (Liu et. al. 1996) as in human, suggesting outcomes of LAMA4 studies with mouse models are relevant to human.

The structure of laminin alpha 4 is similar to other laminins. It has the closest resemblance to the alpha 3 chain however it is also homologous to regions in the alpha 1 and 2 chains. The short arm is closest in homology to laminin alpha 1 (49.3% sequence identity) although it is truncated at the N-terminal domain (Stelzer et. al. 2012, Iivanainen et. al. 1995). LAMA2 and LAMA4 also demonstrate homologies in the long arm and LG domain in that they are coded by a similar number of exons with few rearrangements. Six of the exons that code for these regions are identical in size and exhibit conserved protein motifs (Richards et. al. 1997).

Integrins α3β1, α6β1, αvβ3 (Suzuki 2005, Gonzales 2002, Lian et. al. 2006) and the syndecans act as cellular receptors for laminin alpha 4. Integrin interactions regulate cell motility, survival, and proliferation. Interactions with syndecans are involved in cell signaling, adhesion and migration in tumors and wound edges (Suzuki 2005).

Laminin alpha 4 is the alpha chain subunit that combines with beta and gamma chains to form laminin isoforms laminin-8(α4, β1, γ1),-9(α4, β2, γ1) and -14(α4, β2, γ3). Research has described its involvement in regulating endothelial cell survival through growth and proliferation and LAMA4 plays an integral part in proper development of vasculature and structural integrity of newly formed capillary basement membranes. Deletion of the alpha 4 protein in mice is not lethal but leads to impaired microvessel maturation and cardiomyopathy (DeHahn et. al. 2004, Lian et. al. 2006, Thyboll et. al.2001, and Knöll et. al. 2007). It has been shown to compensate in the absence or reduced expression of other laminins such as laminin alpha2 (Ringelmann et. al. 1999).

To date, no specific human disease has been associated with the laminins containing the alpha 4 chain, although mutations in the alpha 4 chain have been characterized in human cardiomyopathy (Knöll et. al. 2007). Over-expression of laminin-8 in human dermal microvascular endothelial cells has been related to promotion of angiogenesis related functions

(Lian et. al. 2006). The function of Laminin-14 expressed in the developing retina could be involved in the formation or stabilization of synapses in the CNS, in particular during the differentiation of the nervous system in development (Libby et. al. 2000).

2.3 Hearing Loss and Cardiac Abnormalities

Several genetic disorders present with phenotypes that include both hearing impairment and cardiovascular abnormalities due to mutual developmental controls (Table 2). More than 100 genes associated in genetic hearing impairment are also associated to congenital and progressive cardiac abnormalities. These genetic disorders present many modes of inheritance including matrilineal, de novo dominant mutations, genomic disorders, chromosomal disorders, and familial dominant; autosomal-recessive means. The types of genes that are associated with both phenotypes are involved in transcription factors, chromatin remodeling, signal transduction pathways, ion channels, extracellular matrix proteins, among others (Belmont et. al. 2011).

Genomic Disorder	Hearing/Ear association	Cardiovascular Abnormality
DiGeorge Syndrome & 22q11 Deletion	Otitis media, sinorhinitis, hearing loss	Tetralogy of Fallot, interrupted aortic arch type B, double outlet right ventricle, perimembranous
1p36 Deletion Syndrome	Deafness	Dialated cardiomyopathy, left ventricular noncompaction, ventricular septal defects
6p24 Deletion Syndrome	Deafness, auditory hypersensitivity	Aortic valve abnormalities
Williams Syndrome	Otitis media, conductive hearling loss	Supravalvar aortic stenosis, branch peripheral pulmonary arterial stenosis
Long QT syndrome	Congenital deafness	Heart arrhythmia
Leopard Syndrome	Sensorineural deafness	Hypertrophic cardiomyopathy
Alstrom Syndrome	Sensorineural hearing loss	Dialated cardiomyopathy
Osteogenesis Imperfecta	Sensorineural hearing loss	When present pathology of left-sided cardiac valves, aortic root and ascending aorta.

Table 2. Genomic disorders with both hearing impairment and cardiovascular defects (Belmont et. al. 2011).

2.4 Gap Junctions

The gap junction gene family encodes Connexin proteins that form permeable channels known as *gap junctions* that connect the cytoplasm of adjacent cells. They regulate cell to cell communication by allowing ions and small molecules to pass through. The type of molecules or ions that pass through the gap channel facilitates the electrical and biochemical signals sent

between cells and excitable tissues. Distribution of the twenty Connexin isoforms is tissue specific, however many cell types express multiple isoforms (Mammano 2000).

Mutations in gap junctions have been associated with several human pathologies. GJA1 is integral in the synchronized contraction of the heart and in embryonic development (Stelzer et. al. 2012). Mutations in this gene have been associated with severe heart malformations (Dasgupta et. al. 2001). Mutations in gap junction genes, GJB3 and GJB2 have also been associated with skin disorders such as Erythrokeratodermia variabilis (Gottfried et. al. 2002) and Vohwinkel syndrome (Mammano 2000). Charcot–Marie–Tooth disease, linked to mutations inGJA1, causes the degeneration of the peripheral nerve which leads to muscle atrophy, sensory disturbance, hyporeflexia, and foot deformities (Lin P.et. al. 2010).

2.5GJB2

Connexin 26 is encoded by the GJB2 gene. GJB2 is located at cytogenetic band 13q11. It is comprised of two exons, one of which contains the coding sequence (Kemperman et. al. 2002). Variants in the GJB2 genes are responsible for up to 50% of autosomal recessive nonsyndromic deafness in many world populations (Snoeckx et. al. 2005). Specifically, the deletion of a gaunonsine nucleotide at position 30 or 35 is the most frequently occurring of the mutations in GJB2 responsible for a deaf phenotype to be presented (Kemperman et. al. 2002).

CHAPTER 3. MATERIALS AND METHODS

3.1 Approach

A systematic approach was designed to efficiently and effectively search for genetic variants that may be causative or contributory to HLHS in the proband. This method begins with high throughput sequencing and the subsequent integration of various bioinformatics tools to analyze and investigate the relationship of these finding to the syndrome in the proband (Figure 2).

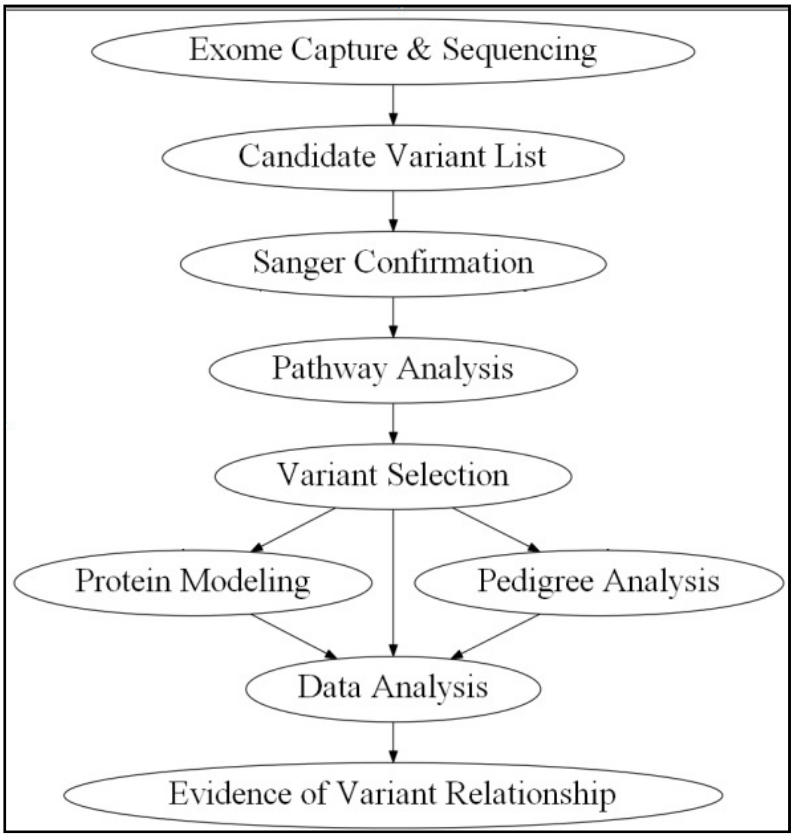


Figure 2: Novel variant detection and disease association approach.

3.2 Justification of Methodology

Although 454 pyrosequencing is a very reliable method, errors do occur. The dominant error type for the 454 platform is insertion-deletion, usually due to consecutive instances of the same base. Another error type that is possible is the miscalling of a nucleotide (Shendure&Ji2008). Sanger sequencing is a reliable method to validate pyrosequencing results

and typically used as the secondary analysis method to verify variant identification. Primers can easily be designed with publically available software from NCBI to flank the specific sequence variants. Turnaround time is also rather quick and sample processing of small sample sets is relatively inexpensive.

The mode of inheritance of HLHS is proposed to be autosomal recessive (Hinton et. al. 2007). Based on the rate of occurrence it is likely variants found in more than 1% of the general population are not the cause of this syndrome, therefore, the novelty of the variant needs to be verified before further investigation. Mutation prediction tools such as SIFT and PolyPhen are publically available and include web-interfaces for small queries. These tools not only provide that a specific variant is novel but also calculates the likelihood that the change from the reference sequence is non-synonymous, damaging, or tolerated. The algorithms use structural and sequence information to predict the impact of substitution on the structure and function of a protein. Determining the conservation of the amino acid provides a measure its functional importance at that position in the protein. Highly conserved amino acids are inferred as being more important to the function of the protein.

Pathway analysis can provide insight and potential avenues to explore and aid in the identification of genes and proteins associated with the etiology of a disease. Pathway-based analysis can mark the genes associated with a disease or phenotype and separates them from false positive hits and can also be useful in identifying biological pathways affected by the genes (Wu 2010). Pathway analysis has been implemented in many areas including cancer research (Wu 2010. Xiao et. al. 2011) and successfully identified significant group effects on several gene-gene pathway interactions relevant to IBD. Ingenuity Pathway Analysis (IPA, Ingenuity® Systems) is a commercially available tool and GeneMANIA (Warde-Farley et. al.2010) is free web-based software. Both tools were used as confirmation of results. GeneMANIA was also implemented as a quick scan for interactions.

Homology modeling is also an effective method used to infer the effect of a single variant in a gene. Mossé et. al. (2008) conducted a linkage study of twenty families with varying degrees of confidence of inheritance of neuroblastoma predisposition genes. Probands were sequenced for novel variants found in the anaplastic lymphoma kinase (ALK) gene. Homology modeling of the wild type and mutant forms of the gene made it apparent that the structure was dramatically altered due to the location of the mutations in an activation segment and in a glycine looping region. This data along with other experimental data in the study made it possible for researchers to determine that ALK is a critical neuroblastoma oncogene.

Pedigree studies have been an invaluable tool for the identification of causative and contributive genes in genetic disease. Perhaps the most famous pedigree is one that illustrates the inheritance of hemophilia in the family of Queen Victoria of England. Due to the well-mapped pedigree, researchers have recently identified a mutation in the F9 gene on the X chromosome as the causative gene responsible for hemophilia B in the Romanov branch of the royal family (Rogaev et. al. 2009). Familial studies have also been integral to the discovery of the Huntington's disease gene HTT (Chial 2008), mutations causative to development of cystic fibrosis (Danes et. al. 1977), and sickle cell anemia (Frenette et. al.2007). These types of studies are useful in determining whether there is a genetic relationship involved in a disease even if it does not tease out the causative genes. Hinton et. al. (2007) used family pedigree analysis to determine hypoplastic left heart syndrome is heritable. In a study of 38 families affected by this syndrome pedigree analysis supported that HLHS in almost entirely caused by genetic effects.

Extended pedigrees increase statistical power for gene discovery and represent a more homogeneous and limited set of causative genes and pathways especially in complex traits such as heart disease. Such complex traits tend to be common and cluster in families but do not always follow clean Mendelian segregation patterns. They may be caused by many genes and affected by multiple pathways in which multiple defects or mutations lead to the phenotypic variation. Familial studies are advantageous in revealing underlying causes as they benefit from the shared

genetic background and similar environmental exposures in comparison to studies with non-related cohorts (Borecki 2008).

3.3 Exome Capture

21µg of input DNA was used with the NimbleGen Sequence Capture Human Exome 2.1M array (Roche NimbleGen, Madison, WI). The Human Exome Array targets ~180,000 human protein coding exons and ~700 miRNA exons using 2.1 million long oligonucleotide probes (>60mer) on a single microarray. The captured fragments were sequenced with Roche protocol for Titanium chemistry on the 454 GS FLX instrument (Roche 454, Branford, CT). The sequence reads were aligned to human genome reference sequence NCBI 36.3 and variants were identified with Roche 454 gsMapper software.

Variants were annotated with a clinical in-house variant analysis platform, *Carpe Novo*, to determine candidate mutations to investigate (Worthey et. al.2011). *Carpe Novo* incorporates four stages of analysis. Primary analysis includes sequencing analysis. Secondary analysis is the sequence alignment, variant calling, and the consideration of data from the sequencing run such as depth of coverage and number of reads. Tertiary analysis consists of the annotation of variants. Quaternary analysis involves running queries on the annotated data to further narrow the potential candidate list (Worthey et. al.2011, Davies 2011).

3.4 Confirmation of Exome Sequencing Results.

Confirmation of the exome sequencing was a two part process. The first step in the process was to design primers that would detect the variant found in exome sequencing. This step involved submitting the nucleotide coordinates to validate the variant in build 36.3 with the NCBI Map Viewer tool found at the following web address: http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606&build=previous (Wolfsburg 2011). The second step was to verify the variants retained their novel and damaging status from the initial date of analysis to the date of primer design.

3.4.1 Primer Design

NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to generate forward and reverse primers flanking each variant. To narrow in on the variant nucleotide positions plus and minus 500 base pairs were entered in the Range Section. If initial testing resulted in several bands, this length allowed for easier nested primer design. Primer Pair specificity Checking parameters were as follows:

- Specificity Check: enable search for primer pairs specific to the intended PCR product = enabled.
- Organism: 9609 (Human)
- Database: Genome (reference assembly from selected organisms)
- Primer specificity stringency: 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 base pairs at the 3' end. Ignore targets that have 7 or more mismatches to the primer (default)

Primer selection criteria required 2 mismatches to be present within 5 bases of the 3' end of the reverse primer. The more mismatches present for other products in the primer pair, the more specific the primer is expected to be for the intended product. As several primers may cover the area of interest, primers with the greatest number of mismatches with unintended targets were selected. These mismatches on the end are to further ensure specific to the intended target sequence. Single and nested primer pairs were designed to confirm the variants found in the exome sequencing (Table 3).

PCRs were conducted for each variant primer set. 5ng of proband DNA sample was denatured at 94°C for 15 minutes and amplified for 40 cycles, each consisting of 30 seconds at 95°C, 30 seconds at 53°C, and 45 seconds at 72°C, followed by a 10-minute extension at 72°C. 4µl of PCR product was visualized on a 2% agarose gel stained with ethidium bromide. PCR product was cleaned by and ExoSAP-IT reaction and sent to Retrogen, Inc. (San Diego, CA) for sequencing reactions. Sequencing results were analyzed with Sequencher DNA analysis software.

Gene	Size (bp)	Sequence
Taldo1 forward	782	GTGGGGAGCGGCCATTCGTC
Taldo1 reverse		GTGGGACGCACACGTGGGAG
KCNJ12 forward	630	TACGGGCTGCGCTGTGTGAC
KCNJ12 reverse		GGGGCGTAGAGGGCACCTCA
ARSD forward	332	TGGGTTGCTCCGTGACGTCA
ARSD reverse		GCGGGGATCACTGCCACCAC
GOLGB1 forward	271	CCACCCCTACCTGTGCAGGTTGT
GOLGB1 reverse		ACTGATGGGAGCAAGAAAAACCAGC
OR7C1 forward	322	AGGCCACCAGACTTGTCGCC
OR7C1 reverse		TCGAGCAGGGAACCCATGACACT
CDC27 forward	260	AGTACCAGCACCATCAATACGACTTTG
CDC27 reverse		GTTCCCTGCAGCACACCAGGC
Taldo1 forward	163	TCATAGTCAGGTGAACTGAG
Taldo1 reverse		TCACCCCCATTGCCTTTCTT
KCNJ12 forward	232	TAACCTGCGCAAGAGCCACA
KCNJ12 reverse		AAAGTCGTCCGTCTCCAGGT
ARSD forward	120	AACTGGGTGTAACCCCAGAG
ARSD reverse		GAACCACGGATTTGACTATT
GOLGB	123	AGGTTGTTCTGCCTGTGCCT
GOLGB		TGTAGAACTCTTGTCATGCTGA
OR7C1	121	ACCTAGAAGATGGTGTGGCT
OR7C1		TTGTGCTTTCTCCCAGCTGA
CDC27	168	AATGTCTAGGATTGACTCTG
CDC27		CAATATCATGTTCCCGTTGCA
MAGEC1 forward	719	TCCCCAGTCTCTTCTCCAGATTCCT
MAGEC1 reverse		CCTCCCCCTGAGGAGGGCTC
OR4C3 - 1 forward	605	ACACCTGCAGCATGCCTAGAACA
OR4C3 - 1 reverse		TGTTGTCTCCTGCACCTAGCACT
OR4C3 - 2 forward	608	ACACCTGCAGCATGCCTAGAAC
OR4C3 - 2 reverse		CATGTTGTCTCCTGCACCTAGCAC
OR4C3 - 3 forward	606	ACACCTGCAGCATGCCTAGAACA
OR4C3 - 3 reverse		TGTTGTCTCCTGCACCTAGCACT
AKAP13 forward	476	GGGATACCCAGGAACGTGCGGA
AKAP13 reverse		CTGGGATTGCTGGAGGGGCG
CD2BP2 forward	618	CCCTCCAAGAGGGCTTGGGC
CD2BP2 reverse		ACCCTGTGGCTGGGTCAGGG
LAMA4 forward	924	TTCCAGCCTCCCTGTGGCCA
LAMA4 reverse		AGCAAGTTGTTTAGTTCACCCAAACCA
OR1J2 forward	668	ACCAGAGCAGCGTGTCCGAGT
OR1J2 reverse		CCCTCAGGATGGTGGCCCCA
OR52N2 forward	375	CCTGCCCTATTGCCGGGGGA
OR52N2 reverse		TGGGTTCATGGTAGGAGGCAGTAGC

Table 3. Primer pairs for PCR confirmation of variants.

3.4.2. Confirmation of Novel Status

Amino acid conservation among species was verified with mouse, rat, chick, and zebrafish. UniProtKB identifiers for each species were retrieved from the UniProt Consortium Protein Knowledgebase (Jain 2009, Magrane 2011). BLAST alignment with the clustalo algorithm was conducted.

The UCSC liftOver tool (Haussler 2006) enables researchers to translate nucleotide positions aligned to previous builds in the NCBI reference genome to other or more current releases. A BED file with the minimum of the following three fields: chromosome (chrom), starting position of the chromosome (chromStart), and the ending position of the chromosome (chromEnd) is required for submission to the tool (Hinrichs et. al. 2006). The start and stop position of the chromosome(s) being queried are returned. The position of the variant of interest is subtracted from the end position of the chromosome in the original assembly in order to translate the updated coordinates in the queried assembly.

The NCBI37/hg19 nucleotide coordinate was used as search criteria to verify no SNP data is associated with the variant. A SNP identification number is necessary for determining if a SNP exists within dbSNP. Since the variants examined in this study were novel, there was no SNP identification value. The coordinates of the variant and sequence were very useful in confirming with NCBI whether a SNP had been identified with this variant and either missed by the initial carpe novo analysis or updated since that original analysis.

SIFT analysis (Ng & Henikoff 2003) was repeated with the updated nucleotide coordinates to ensure the novel and damaging status remained current. The chromosome number, variant position, orientation and alleles were entered into the web based submission form. OMIM Disease and Allele Frequencies (All HapMap Populations) output options were selected (Ng & Henikoff 2003).

Polymorphism Phenotyping (PolyPhen) analysis was repeated using the amino acid position and substitution. This pipeline calculates the potential impact of an amino acid substitution on the structure and function of a protein. A numerical score ranging from 0.00-1.00 is returned from the input queries. The higher the score, the more likely the amino acid change is damaging to the structure and function of the queried protein (Adzhubei et. al. 2010).

3.5 Protein Model Development

Homology models of wild type and mutated genes were generated with MODELLER v9.10 (Šali 1993) and PyMOL v0.99 software (The PyMOL Molecular Graphics System). Genes with resolved crystal structures were aligned with the sequences of the mutated genes with BLAST. The most homologous crystal structure found in Protein Data Bank (pdb - Bernstein 1977) and the Protein Families Database (Pfam - Punta et. al.2012) was used as the backbone for the mutated gene to be threaded upon. The .pdb file was loaded into PyMOL with the PDB Plug-in Loader Service. All hydrogen, water, and sugar molecules were deleted as they are not necessary for the MODELLER software. Blastp alignment was conducted to verify the sequence segment to utilize for the wild type and mutated models.

The MODELLER align2d.py script (Altschul 1990) was used to generate the final alignments to the known structures. The alignment generated by this step is input to the Model-single.py script for calculation of the 3D models in .pdb format. These models are viewed in the PyMOL software and analyzed. The model that retained the most structural homologies within the model set was selected as the final representative model.

3.6 Pathway Analysis

Two methods of pathway analysis were conducted with the IPA tool (Ingenuity® Systems). The first method generated a general sense of the gene neighborhoods to explore if any of the candidate genes are closely related in a pathway. Each gene was queried under the Genes and Chemicals search. The interaction network option was selected to view the figure representing the relations and the build tool was used to grow pathways from the query gene. Individual queries were run based on indirect/direct interactions and relationship types. Examples of interactions used as ‘grow’ criteria were: direct, protein-protein interaction, indirect protein-protein interaction, direct regulation of binding, etcetera. All relationship query types were examined in order to get a full picture of all possible relationships.

The interaction neighborhoods were exported to a .csv file which was then modified to include the gene queried, direct or indirect interaction, and relationship type. The .csv files were merged with a python script. The combined file was used to determine if any of the neighborhood genes overlapped with other gene queries. The overlapping genes were then compared to the exome date to determine if variants were found in these genes but then not considered for evaluation due to a non-damaging or non-novel status. An additional list of genes reported to be involved with CHD, HLHS, and Turner's syndrome were also examined in this manner (Table 4).

Gene	Association	Reference
ATRX	Turner's syndrome	Tomita-Mitchell 2012
BCOR	Turner's syndrome	Tomita-Mitchell 2012
EHMT1	CHD	Kleefstra 2006
ERBB4	CHD	McBride 2011
FKBP6	Turner's syndrome	Tomita-Mitchell 2012
FLNA	Turner's syndrome	Tomita-Mitchell 2012
FOXC1	HLHS	Grossfeld 2009
GATA4	CHD	Draus 2009, Rajagopal 2007, Tomita-Mitchell 2007, Tomita-Mitchell 2012
GJA1	HLHS	Grossfeld 2007, Grossfeld 2009, Hinton 2009
GPC3	Turner's syndrome	Tomita-Mitchell 2012
HAND1	HLHS	Grossfeld 2009, Reamon-Buettner 2008
MID1	Turner's syndrome	Tomita-Mitchell 2012
MYH11	CHD	Zhu 2006
NKX2.5	HLHS	Elliott 2003, Grossfield 2007, Grossfeld 2009, McBride 2008, Stallmeyer 2010
NOTCH1	HLHS	Grossfeld 2009, Grossfeld 2007, McBride 2008
SOX7	CHD	Wat 2009
ZIC3	Turner's syndrome	Tomita-Mitchell 2012

Table 4. Genes with cardiac or Turner's syndrome associations. These genes were used in IPA pathway analysis. See bibliography for references.

The second method of IPA network analysis was to generate a graphical representation of the molecular relationships between the candidate genes and HLHS. Cardiac disease and functions such as: Hypertrophy of cardiac muscle, HLHS, hypertrophy of heart, congestive heart failure, heart disease, coronary disease, and left ventricular dysfunction were overlaid to visually determine if any the candidate genes have a cardiac related function or were closely related to genes associated with HLHS.

Additional pathway analysis was conducted with GeneMANIA (Warde-Farley et. al.2010) to examine potential interactions of candidate genes in the study, known cardiac genes, and gap junction genes. GeneMANIA utilizes publicly available genomic and proteomic data sets such as GEO, BioGRID, Pathway Commons, and I2D for prediction of functional relationships or

shared properties of genes. It supports data for Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Mus musculus, Homo sapiens, and Saccharomyces cerevisiae (Warde-Farley et. al. 2010). GeneMANIA suggests annotations for genes based on Gene Ontology term enrichment of highly interacting genes with the query gene and is a gene recommendation system (The Gene Ontology Consortium 2000). The queries were weighted according to the Gene Ontology (GO) biological function weighting method. This weight reflects the data source relevance for predicting the function of interest (Warde-Farley et. al. 2010). GeneMANIA output was used in conjunction with IPA analysis to verify and generate hypothesis of gene relationships.

3.7 Family Pedigree Analysis

To examine the potential relationship of the CHD and deaf phenotype in the family, primers designed and published by Snoeckx et. al. for detecting the connexin26 deletion: Cx26F (5'-TCTTTTCCAGAGCAAACCGC-3') /Cx26R (5'- GGGCAATGCGTTAAACTGGC-3') were used in Sanger sequencing reactions with all available family DNA samples. DNA sample was denatured at 94°C for 15 minutes and amplified for 40 cycles, each consisting of 30 seconds at 95°C, 30 seconds at 63°C, and 45 seconds at 72°C, followed by a 10-minute extension at 72°C. 4µl of PCR product was visualized on a 2% agarose gel stained with ethidium bromide. PCR product was cleaned by and ExoSAP-IT reaction and sent to Retrogen, Inc. (San Diego, CA) for sequencing reactions. Sequencing results were analysis with Sequencher DNA analysis software.

CHAPTER 4. RESULTS AND DISCUSSION

4.1 Exome Sequencing and Sanger Sequencing Analysis

Exome sequencing resulted in an average depth of 36X coverage. Enrichment was approximately 75-fold. 60% of reads mapped to targeted regions and more than 99% of the intended targeted region was covered. Carpe novo data analysis identified more than 13,000 coding differences compared to the human genome reference (NCBI36/hg18) were found. More than 1400 of these coding differences (variants) were novel (Worthey et. al.2011).

PolyPhen (Adzhubei et. al 2010) and SIFT (Ng &Henikoff 2003) analyses were then utilized to further prioritize variants. The more conservative PolyPhen analysis was conducted first due to the ability to automate the process in house. Non-synonymous sequence variants in approximately 140 genes were predicted to be "damaging". This data set was then entered into SIFT (www.http://sift.jcvi.org/) and further reduced to 13 genes when sorted by status according to SIFT. The variant positions and nucleotide results were confirmed against NCBI map viewer human build 36.3 before confirmation with Sanger sequencing.

PCR Primer sets for variants found in genes TALDO1, KCNJ12, ARSD, LAMA4, and AKAP13 generated clean results when visualized on the 2% agarose gel, reducing the number of variants of interest to five (Figure 3). The DNA samples of immediate family members: father, mother, unaffected sibling, and fetal tissue were screened. Sanger sequencing results were aligned and analyzed with Sequencher DNA Sequencing Software, further reducing the number of variants to potentially investigate (Table 5).

Figure 3. Gel Electrophoresis results of TALDO1, KCNJ12, ARSD, LAMA4, and AKAP13. These products were sent for sequencing confirmation. Lane 1: 100bp DNA Ladder (Invitrogen) Lanes 2-5: Negative controls (TALDO1, KCNJ12, ARSD, AKAP13, LAMA4). Lanes6-10: Positive Negative controls (TALDO1, KCNJ12, ARSD, AKAP13, LAMA4). Lanes 11-15: TALDO1. Lanes16-20: KCNJ12. Lanes21-24: ARSD. 100bp DNA Ladder (Invitrogen).Lanes26-29:AKAP13. Lanes30-33: LAMA4.100bp DNA Ladder (Invitrogen) Sample order is: Proband, Father, Mother, Unaffected Sibling, Fetal tissue.

			Sanger Sequencing results				
Gene	Expected	454	Proband	Father	Mother	Sibling	Fetus
AKAP13	C	T	Variant*	Variant*	Expected	Variant*	Variant*
LAMA4	C	G	Variant*	Expected	Variant*	Expected	Variant*
ARSD	A	T	Expected	Expected	Expected	Expected	Expected
Taldo1	C	A	Expected	Expected	Expected	Expected	Expected
KCNJ12	T	G	Expected	Expected	Expected	Expected	Expected

Table 5. Sanger sequencing confirmation. Nucleotides listed in the 'Expected' column are expected according to the NCBI36/hg18 assembly. Nucleotides listed in the '454' column indicate nucleotide variants found in the exome sequencing reaction. Primer sets for variants found in genes ARSD, Taldo1, and KCNJ12 did not confirm the variants found in the exome sequencing reaction.

The AKAP13 PCR reaction confirmed the exome sequencing finding of a nucleic acid change from cytosine to thymine at position 83,924,813 (NCBI36/hg18 assembly). The initial decision was to not pursue the variant in this gene as the mode of inheritance is not straightforward since all family members except the mother carry the variant. A point of interest for this variant is that there is a second peak at this position in all members that carry the mutation. A cytosine peak is present that is approximately 10% the height of the thymine peak (Figure 4).

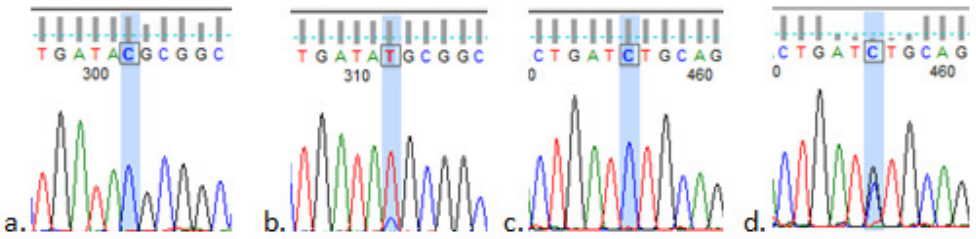


Figure 4. Chromatograms of AKAP13 and LAMA4 Sanger sequencing results: a.AKAP13 cytosine present in DNA sequence of the mother. b. Thymine variant found in other family members – note presence of slight cytosine peak. c. LAMA4 cytosine present in father and unaffected sibling. d. Heterozygous variant in proband, mother, and fetus.

The list of variants of interest ultimately was reduced to the variant found in LAMA4 at position 112570046(NCBI36/hg18 assembly). The chromatograms of the proband, mother, and fetus indicate they all are heterozygous for the LAMA4 variant (guanine) and expected nucleotide (cytosine). The father and unaffected sibling both have a single cytosine peak at this position (Figure 2c, 2d). PCR sequence analysis of extended family members including a non-HLHS sibling, maternal grandparents, maternal half great uncle and aunt, maternal aunt, & maternal cousin showed that the non-HLHS sibling and maternal grandmother are also heterozygous for the LAMA4 variant.

4.2 Variant Pathway Analysis

The 13 genes of interest were analyzed through the use of IPA (Ingenuity® Systems, Inc. Redwood City, CA). Output files were combined and compared to find genes that overlapped among the twelve candidate's gene neighborhoods. Some neighborhood overlaps did occur, however further pathway analysis did not indicate a clear relationship to cardiac defects (Table 6). Variants in IL2 (interleukin 2 - cytokine) and SKIL (ski-like oncogene) were found in the exome sequencing results but were not novel or predicted as damaging. GeneMANIA queries of these genes, with LAMA4 did not result in physical or pathway relationships.

FOXM1 (cell proliferation), TNF (tumor necrosis factor), and ZHX1 (zinc finger and homeobox) were shared in LAMA4 neighborhoods but no mutations in these genes were found in the exome sequencing. These genes do not have implied cardiac functions, however to

potentially sift out potential relationships, a GeneMANIA query was run to screen for interactions with HLHS genes GJA1 and PDE5A to potentially view pathway connections that could be related to HLHS and no significant relationships were returned.

Symbol	Family	From Neighborhood	Relationship Type	Interaction
FOXM1	transcription regulator	CDC27	PP	D
FOXM1	transcription regulator	LAMA4	PD	D
FYN	kinase	CD2BP2	RB	I
FYN	kinase	CDC27	PP	D
HNF4A	transcription regulator	AKAP13	MB	D
HNF4A	transcription regulator	KCNJ12	PD	D
IL2	cytokine	AKAP13	LO	I
IL2	cytokine	CD2BP2	LO	I
SKIL	transcription regulator	CDC27	PP	D
SKIL	transcription regulator	GOLGB1	PP	D
TNF	cytokine	LAMA4	E	I
TNF	cytokine	TALDO1	E	I
ZHX1	transcription regulator	LAMA4	PP	D
ZHX1	transcription regulator	TALDO1	PP	D

Table 6. Overlapping genes found in query neighborhoods. IL2 and SKIL Genes in column one are also found to have non-novel mutations in the exome sequencing results. Key: PP: protein-protein binding, PD: protein-DNA binding, RB: regulation of biding, MB: Group/complex membership, LO: localization, E: expression (includes metabolism/synthesis for chemicals). D: direct interaction, I: indirect interaction.

The outputs from the variant neighborhood analysis were also compared to the gene list in Table 3. Overlapping genes were compared to the exome sequencing data to determine if variants were found but not considered for evaluation due to a non-damaging or non-novel status. The purpose of this comparison is to ascertain if the novel mutation alone is not responsible for the development of HLHS but coupled with other mutations (possibly not novel) HLHS could develop. PRKCA was found in these results, the 13 genes of interest neighborhood output, and 454 exome results (Table 7).

Output Gene	Variants of Interest	Associated Genes	Exome variant
CDKN1A	CDC27	NOTCH1	no
CTNNB1	AKAP13	GJA1	no
ESR1	AKAP13	ERBB4	no
IL5	AKAP13	GATA4	no
ITGB1	LAMA4	FLNA	no
ITGB3	LAMA4	FLNA	no
PRKCA	AKAP13	GJA1	YES
SMAD3	CDC27	NOTCH1	no
SRF	AKAP13	MYH11	no
TP53	LAMA4	NOTCH1	no

Table 7. Common genes in exome data and additional gene list. PRKCA was found in both the variant gene list and the additional gene list neighborhoods (Table 4). PRKCA was found to have a variant in the exome sequencing data. The PRKCA variant found in the exome sequencing results in the nonsynonymous coding change of amino acid valine (nonpolar neutral) to isoleucine (nonpolar neutral). The SIFT score is 1 deeming the change tolerated. The mutation is identified with dbSNP ID rs6504459 and therefore not novel.

PRKCA is a kinase that has been reported to play roles in many different cellular processes, such as cell adhesion, cell transformation, cell cycle checkpoint, and cell volume control (Stelzer et. al. 2012). In epithelial studies it has been indicated that the failure to down-regulate PRKCA increases leakiness of tight junctions (Mullin et. al. 2000). It also promotes angiogenic activity of human endothelial cells (Xu et. al. 2008) implying a potential relationship with GJA1 and LAMA4. It also has been studied in regards to its role in cardiac function. PRKCA controls myocardial contractility in mice and rats with cardiac disease. Knockout studies have shown null PRKCA mice have enhanced cardiac contractility and reduced susceptibility to heart failure whereas over-expression showed decreased cardiac contractility, ventricular dilation, and secondary hypertrophy (Ladage et. al. 2011)

The GeneMANIA query of PRKCA, known HLHS gene GJA1, and candidate genes LAMA4, and AKAP13 with Gene Ontology (GO) biological function based weighting returned one direct genetic interaction between LAMA4 and AKAP13 (Figure 5). This interaction returned the lowest weight of all interactions between query genes (4.55E-04) and was supported by one publication (Lin A. et. al. 2010). A similar pathway query did not show connectivity among the genes except in the case of LAMA4 and PRKCA, and again was the weakest interaction in the

network query (0.005128629). This pathway relationship was only supported by one publication. Physical Interactions accounted for 82.62% of the network interactions in this query meaning any interaction with these genes is more likely to be physical rather than pathway or genetic. Although the weights of the predicted interactions among the query genes were not exceptionally high, interactions between PRKCA and AKAP13 and PRKCA and GJA1 were supported by iRef-BIND, Bio-Grid-small scale studies, IREF-OPHID, and iRef -small scale studies. These genes may be found to indirectly work together in the formation of HLHS when more data is published.

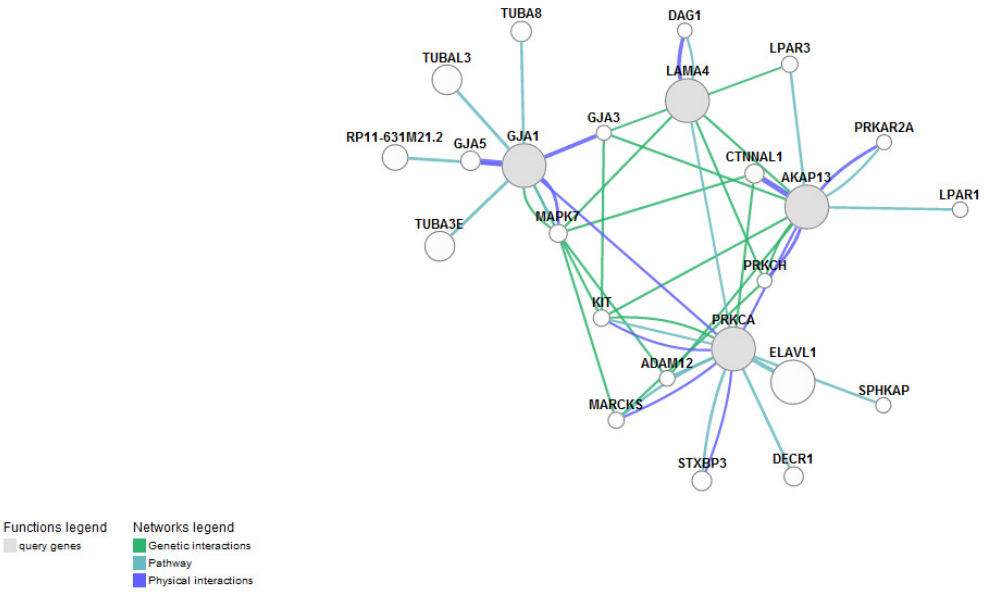


Figure 5. Genetic, pathway, and physical interactions. LAMA4, GJA1, PRKCA, and AKAP13(as predicted by GeneMANIA)share physical, pathway, and genetic interactions.

The second method of IPA network analysis generated a graphical representation of the molecular relationships between the twelve genes of interest and HLHS or cardiac disease or function. The input for this representation were the original thirteen genes, GJA1, and PDE5A. GJA1 and PDE5A were included as they were returned from an IPA Functions and Diseases query of HLHS. Cardiac associations shown in Figure 6 were overlaid to visually determine if any the genes have a cardiac related function. This visual representation indicates LAMA4 and AKAP13 are involved in several cardiac related diseases functions and therefore could potentially be involved in the development of HLHS in this family.

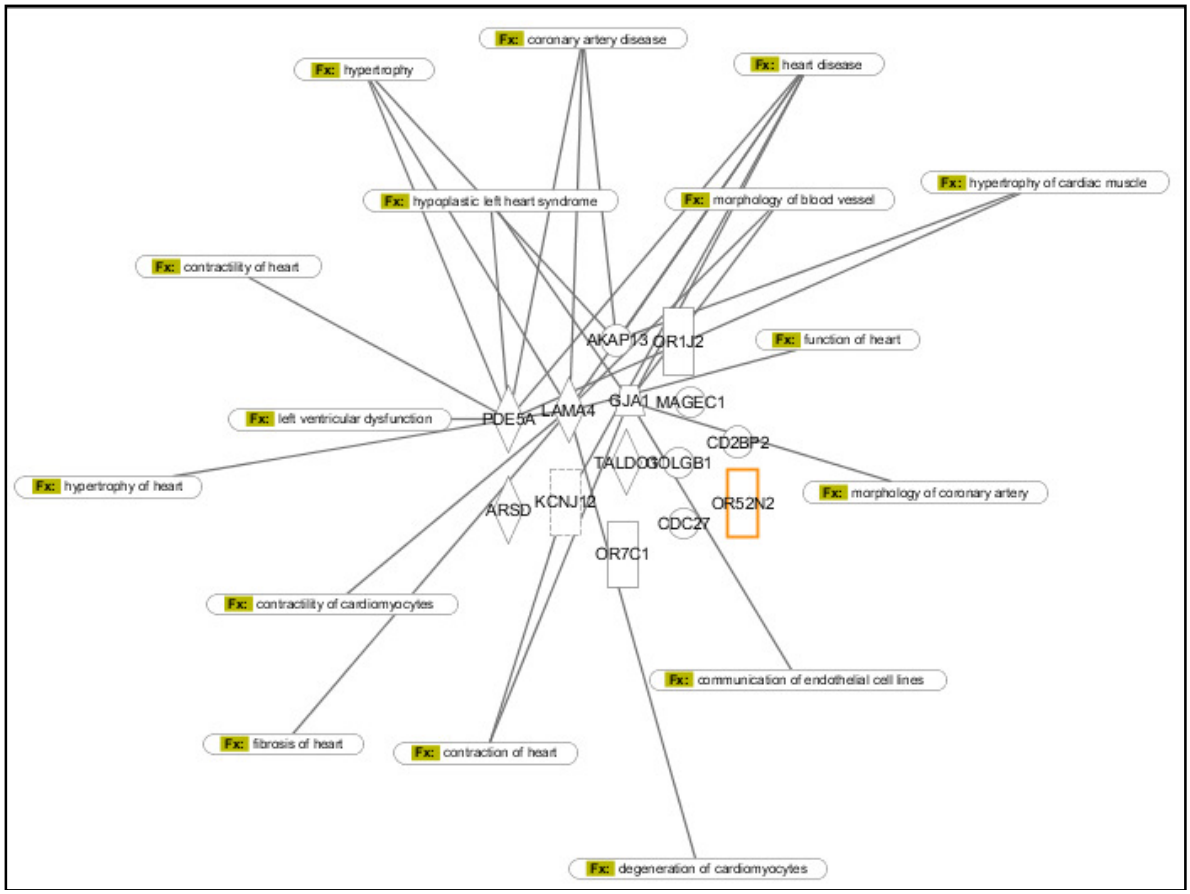


Figure 6. Variant and Cardiac Relationships. Disease Molecules and disease are represented as nodes and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one reference from the literature, from a textbook, or from canonical information stored in the Ingenuity Knowledge Base. Molecule nodes are displayed using various shapes that represent the functional class of the gene product. Function / disease nodes are oval (© 2000-2010 Ingenuity Systems).

4.3. Novel Variant Selection

Pathway analysis and Sanger sequencing confirmation of the first five family members that provided DNA samples indicated that LAMA4 was a potential candidate to play a role in the development in HLHS. Literature searches supported that there may be involvement of LAMA4 with HLHS due to the integral role of LAMA4 in cell communication, scaffolding, development, and the circulatory system (refer to Chapter 2.2).

4.3.1 Justification

Several hypotheses regarding the development of HLHS suggest the mutation found in the LAMA4 gene may be causative or contributive to the development of CHD in this family. One accepted hypothesis, supported by research in embryonic chick, states that HLHS develops late in embryogenesis as a result of embryonic alterations in blood flow, such as premature narrowing of the foramen ovale and aortic stenosis (Hinton et. al. 2007, Dasgupta et. al. 2001). Given the alpha-4 subunit is involved in angiogenesis and endothelial branching morphogenesis (a component of blood vessel development) it is feasible to consider the likelihood that LAMA4 plays a role in the development in HLHS. This is further supported by a second hypothesis that suggests that the extracellular matrix protein signaling molecules could serve as HLHS causing genes (Hinton et. al. 2007).

LAMA4 is located in close proximity to another region also related to cardiac defects on chromosome 6. A family-based genome-wide linkage analysis performed on families with an HLHS or bicuspid aortic valve (BAV) proband linked HLHS to chromosome 6q23 (Hinton et. al. 2009). The 6q22-23 region is also linked to familial dilated cardiomyopathy and conduction disease (Messina et. al. 1997). Char syndrome is caused by mutations in TFAP2B located in region 6p12-21.1. Cardiovascular abnormalities such as patent ductus arteriosus (the connection between the aorta and the pulmonary artery does not close after birth) are commonly present in patients with this syndrome (Vaughan & Basson 2000). Perhaps the proximity of LAMA4 to these regions can be considered as a risk factor.

Several studies have linked LAMA4 to cardiovascular development, function, and dysfunction. Knöll et. al. (2007) discovered two novel amino acid mutations in the LAMA4 gene that lead to a loss of integrin binding capacity. They discovered that a function of integrin –linked kinases is to maintain morphology of the ventricle during embryonic development and to maintain cardiomyocyte shape. This function is dependent on the LAMA4 gene suggesting that mutations in this gene causing conformational changes will interfere with the interaction of

endothelial cell and cardiomyocytes with the extracellular matrix. This interference may cause cardiomyopathy and ultimately lead to death (Knöll et. al.2007)

Studies with model organisms such as chick and mouse have been conducted to better understand human cardiac development. Nath et. al. (2009) determined that endothelial cell differentiation is a critical factor in heart development. These studies have also indicated that there are overlapping developmental pathways in vasculature and heart. Additional studies with laminin alpha-4 deficient mice have demonstrated an ischemic cardiac phenotype caused by a mutation in the laminin alpha 4 chain. Although mice survive development, 20% of the null mice died in the perinatal stage. Surviving litter mates did not express a notable phenotype but had increased frequency of sudden death. The null laminin alpha-4 mice possessed altered extracellular matrix structures that created an inefficient flow of blood to the heart due to cardiac hypertrophy and cardiac dysfunction (Wang et. al. 2009). Additional studies show null laminin alpha-4 mice leads to impaired blood vessel maturation. Capillary basement membranes are weak causing fatal hemorrhaging in 50% of the null laminin apha-4 mice (Thyboll et. al. 2001). This discovery could be pertinent if HLHS is acquired due to restriction of left heart blood flow in embryogenesis as suggested by Dasgupta et. al. (2001).

4.3.2. Supporting Pathway Analysis

Pathway analysis with IPA (Ingenuity® Systems) supported the pursuit of LAMA4 in regards to HLHS. Figure 6 is a graphical representation of the thirteen genes of interest and their relationships to cardiovascular disease and system development and function. Each relationship has a p-value calculated using the right-tailed Fisher Exact Test (Table 8). The p-value is a measure of the likelihood that the association between the molecules and a given function is due to random chance. The smaller the p-value, the less likely the association is random, and the more significant the association. In general p-values less than 0.05 indicate a statistically significant, non-random association (Ingenuity® Systems, www.ingenuity.com). The calculated p-values for the data set of the twelve genes of interest including GJA1 and PDE5A statistically support the

importance of the role of LAMA4 in the cardiovascular system. It was not indicated to have a relationship with HLHS; however, the other relationships within the cardiovascular system may lead to causal or contributory factors.

Category	Function	Function Annotation	p-value	Molecules	# Molecules
Cardiovascular Disease	hypoplastic left heart syndrome	hypoplastic left heart syndrome	1.66E-06	GJA1, PDE5A	2
Cardiovascular System Development and Function	morphology	morphology of blood vessel	8.92E-05	GJA1, **LAMA4**	2
Cardiovascular Disease	hypertrophy	hypertrophy of cardiac muscle	1.36E-04	AKAP13, PDE5A	2
Cardiac Hypertrophy	hypertrophy	hypertrophy of cardiac muscle	1.36E-04	AKAP13, PDE5A	2
Cardiovascular System Development and Function	morphology	morphology of coronary artery	5.46E-04	GJA1	1
Cardiovascular System Development and Function	coupling	coupling of proepicardial cells	5.46E-04	GJA1	1
Cardiovascular System Development and Function	migration	migration of proepicardial cells	5.46E-04	GJA1	1
Cardiovascular System Development and Function	proliferation	proliferation of proepicardial cells	5.46E-04	GJA1	1
Cardiac Proliferation	proliferation	proliferation of proepicardial cells	5.46E-04	GJA1	1
Organ Development	development	development of primary ovarian follicle	5.46E-04	GJA1	1
Cardiovascular Disease	atherosclerosis	atherosclerosis of abdominal aorta	1.09E-03	GJA1	1
Cardiovascular Disease	endocardial cushion defect	endocardial cushion defect	1.09E-03	GJA1	1
Cardiovascular System Development and Function	communication	communication of endothelial cell lines	1.09E-03	GJA1	1
Organ Development	development	development of secondary ovarian follicle	1.09E-03	GJA1	1
Cardiovascular System Development and Function	contraction	contraction of heart	1.49E-03	GJA1, KCNJ12	2
Cardiovascular Disease	angina pectoris	angina pectoris	1.69E-03	KCNJ12, PDE5A	2
Cardiovascular System Development and Function	branching	branching of endothelial cell lines	2.18E-03	**LAMA4**	1
Cardiovascular Disease	congestive heart failure	congestive heart failure	3.03E-03	KCNJ12, PDE5A	2
Cardiovascular Disease	cerebral vasospasm	cerebral vasospasm	3.81E-03	PDE5A	1
Cardiac Degeneration	degeneration	degeneration of cardiomyocytes	4.90E-03	**LAMA4**	1
Cardiovascular Disease	heart disease	heart disease	6.17E-03	AKAP13, GJA1, KCNJ12, **LAMA4**, PDE5A	5
Cardiovascular Disease	atherosclerosis	atherosclerosis of aortic root	8.16E-03	GJA1	1
Cardiovascular Disease	coronary disease	coronary disease	8.49E-03	AKAP13, KCNJ12, **LAMA4**, PDE5A	4
Cardiovascular Disease	cardiomyopathy	cardiomyopathy	9.57E-03	GJA1, PDE5A	2
Cardiovascular Disease	atherosclerosis	atherosclerosis	1.11E-02	AKAP13, GJA1, **LAMA4**, PDE5A	4
Cardiovascular Disease	patent ductus arteriosus	patent ductus arteriosus	1.25E-02	PDE5A	1
Cardiovascular Disease	left ventricular dysfunction	left ventricular dysfunction	1.36E-02	PDE5A	1
Cardiovascular Disease	intermittent claudication	intermittent claudication	1.41E-02	PDE5A	1
Cardiovascular Disease	ventricular fibrillation	ventricular fibrillation	1.41E-02	KCNJ12	1
Cardiovascular Disease	stroke	stroke	1.44E-02	GJA1, PDE5A	2
Cardiovascular Disease	valvulopathy	valvulopathy	1.73E-02	PDE5A	1
Cardiovascular Disease	cardiac arrest	cardiac arrest	1.79E-02	PDE5A	1
Cardiovascular System Development and Function	looping morphogenesis	looping morphogenesis of heart	2.05E-02	GJA1	1
Organ Development	looping morphogenesis	looping morphogenesis of heart	2.05E-02	GJA1	1
Cardiovascular Disease	chronic stable angina	chronic stable angina	2.16E-02	KCNJ12	1
Organ Development	gonadogenesis	gonadogenesis	2.29E-02	GJA1, OR7C1	2
Cardiovascular System Development and Function	contractility	contractility of cardiomyocytes	2.32E-02	**LAMA4**	1
Cardiovascular System Development and Function	morphogenesis	morphogenesis of blood vessel	3.01E-02	GJA1	1
Cardiovascular Disease	aortic stenosis	aortic stenosis	3.07E-02	PDE5A	1
Cardiovascular Disease	systemic sclerosis	systemic sclerosis	3.07E-02	PDE5A	1
Cardiovascular Disease	endothelial dysfunction	endothelial dysfunction	3.12E-02	PDE5A	1
Cardiovascular Disease	ischemic cardiomyopathy	ischemic cardiomyopathy	3.44E-02	PDE5A	1
Cardiovascular Disease	pulmonary hypertensive arterial disease	pulmonary hypertensive arterial disease	3.54E-02	PDE5A	1
Cardiovascular Disease	ventricular tachycardia	ventricular tachycardia	3.86E-02	KCNJ12	1
Cardiovascular Disease	preeclampsia	preeclampsia	3.91E-02	PDE5A	1
Cardiovascular System Development and Function	contractility	contractility of heart	4.12E-02	PDE5A	1
Cardiovascular System Development and Function	function	function of heart	4.18E-02	**LAMA4**	1
Organ Development	function	function of heart	4.18E-02	**LAMA4**	1
Cardiovascular Disease	coronary artery disease	coronary artery disease	4.30E-02	AKAP13, **LAMA4**, PDE5A	3
Cardiac Arteriopathy	coronary artery disease	coronary artery disease	4.30E-02	AKAP13, **LAMA4**, PDE5A	3
Cardiovascular System Development and Function	vasoconstriction	vasoconstriction of blood vessel	4.54E-02	PDE5A	1
Cardiovascular Disease	hypertension	hypertension	4.55E-02	AKAP13, CDC27, PDE5A	3

Table 8. Calculated p-values of figure 6 data. Only statistically significant p>0.05 values are included.

4.3.3 Verification of LAMA4 Gene Variant

To verify amino acid conservation among species, protein identification for LAMA4 in mouse, rat, chick, and zebrafish was retrieved from the UniProt Consortium Protein Knowledgebase (Jain 2009, Magrane 2011). BLAST alignment with the clustalo algorithm was conducted with the UniProtKB identifiers for each species: Q16363 (Human), P97927 (Mouse), Q5SNU9 (Zebrafish), F1NSZ5 (Chick), and F1RZM4 (Pig). The alignment confirmed

conservation of the expected amino acid in this position further supporting the importance of the change in amino acid sequence at this position (Figure 7).

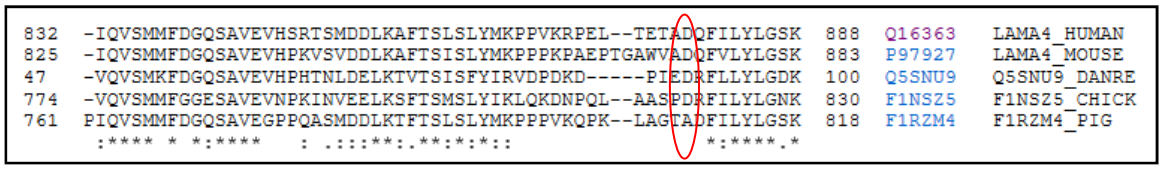


Figure 7. BLAST alignment of LAMA4 across species. Aspartic Acid (D) at amino acid 879 is conserved.

LAMA4 BED file: chr6 112,535,827 112,682,521 was submitted to UCSC liftOver (Hinrichs et. al. 2006) and the converted start and stop positions of the chromosome were returned. After translation to the latest build, the nucleotide coordinate 112,463,353 was used as search criteria to verify no SNP data is associated with this particular variant which would diminish the novel status. In 2221 NCBI SNPs for the LAMA4 gene there were no matches.

SIFT analysis (Ng 2003) was repeated with the updated coordinates to ensure the novel and damaging status remained current. Chromosome coordinates 6,112463353,1,C/G were entered into the submission form. Output options included OMIM Disease and Allele Frequencies (All HapMap Populations). The SIFT score (0) predicted the amino acid substitution is damaging. No dbSNP ID, allele frequencies, or OMIM disease associations were reported.

PolyPhen-2 (Polymorphism Phenotyping v2, Adzhubei et. al 2010) analysis was also repeated to ensure the amino acid change was still considered damaging. A score of 1.0 was returned confirming the damaging status. The multiple sequence alignment used in the PolyPhen-2 pipeline, further confirms sequence conservation at this location (Figure 8).

To further verify novelty of this variant, the Exome Variant Server (EVS) was queried and the variant was not found. The coordinates were expanded to look in the region flanking the variant of interest (4,875 alleles with data from this region) and remained novel (October 21, 2011). This analysis was repeated in February 2012, when the data set had been updated to 5400 exomes. There were no occurrences of the LAMA4 D879H variant in 7020 alleles of the

European cohort. One occurrence of this variant was identified in 1 of 3738 African American alleles that covered that region. No phenotype or clinical associations, known SNP IDs, or associated publications were assigned to this variant so it retains its novel and rare status (Exome Variant Server).

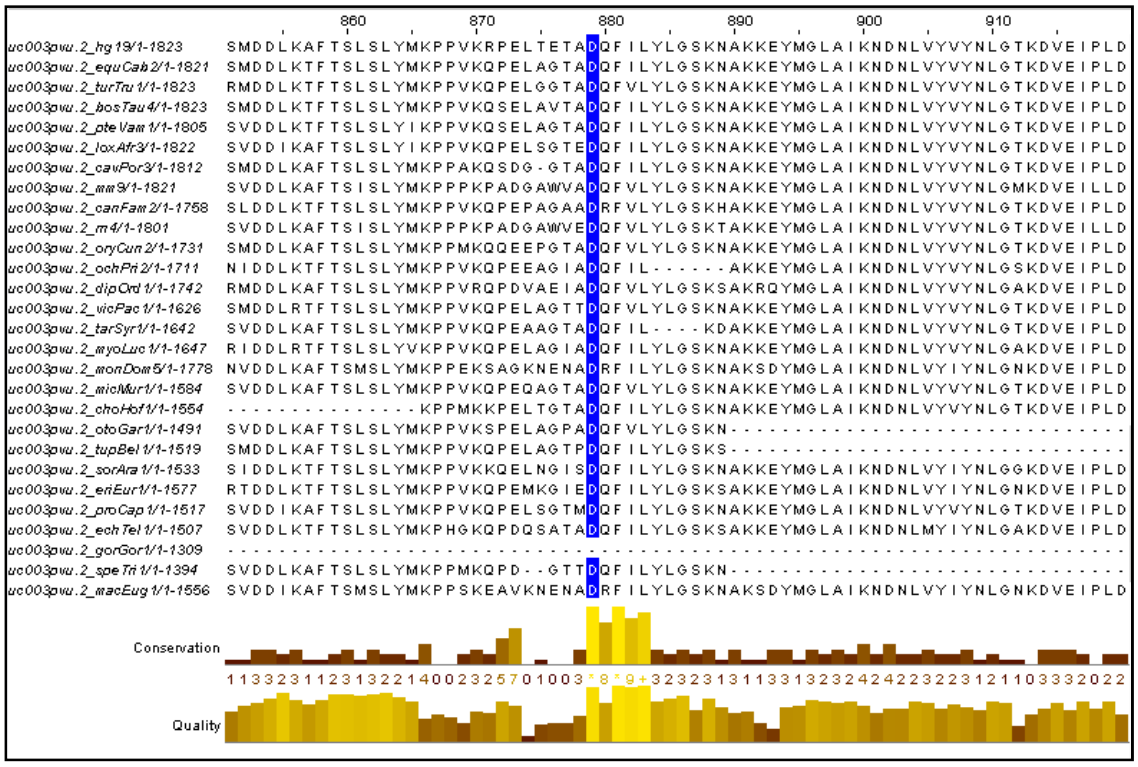


Figure 8. PolyPhen-2 alignment of LAMA4. Results illustrate the conservation of aspartic acid at position 879.

4.3.4 Laminin Alpha-4 Protein Modeling

The cytosine to guanine nucleotide change results in a change in the amino acid sequence. The expected triplet GAT encodes a negatively charged aspartic acid (D). The occurrence of a cytosine in place of the guanine (CAT) encodes for histidine (H), which is a positively charged amino acid at neutral pH. Designing a homology model would be extremely interesting to determine if the D to H mutation causes any conformational changes in the protein, especially since PolyPhen-2 and SIFT analysis predict the variant to be damaging.

No human model of laminin proteins is available. The G1 domain of mouse laminin alpha-2 is the most homologous to the G1 domain of human laminin alpha-4. The crystal

structure of the mouse LG1-3 region of the laminin alpha2 chain (PDB ID 2WJS) served as a backbone for the laminin-4 wild type and mutated models (Carafoli et. al. 2009).

Blastp alignments were conducted to verify the sequence segment to utilize for the model (Figure 9). The procedure was repeated for the mutated model; aspartic acid was replaced with a histidine at amino acid 879 (Figure 10). The laminin alpha-4 sequence is based on the Blastp alignment of amino acids 875-1013 which encode part of the laminin G1 domain. The amino acids in the pdb file after amino 2324 were deleted to eliminate the G2 and G3 domains in the 2WJS sequence, resulting in a more accurate depiction of the protein and simplifying the view of the end model. The MODELLER alignment algorithm (Altschul 1990) was used to generate the final sequence alignment of the human laminin alpha-4 and mouse laminin alpha-2 proteins.

```
 Score = 93.2 bits (230),  Expect = 2e-23, Method: Compositional matrix adjust.
 Identities = 118/546 (22%), Positives = 227/546 (42%), Gaps = 80/546 (15%)

Query  41    TAVADNLLFYLGSAKFI-DFLAIEMRKGKVSFLWDVGSGVGRVEYPDLTIDD-------S  92
             T  AD  + YLGS     +++ + ++    + +++++G+    VE P   +D        +
Sbjct  875   TETADQFILYLGSKNAKKEYMGLAIKNDNLVYVYNLGTK--DVEIP---LDSKPVSSWPA  929

Query  93    YWYRIEASRTGRNGSISVRALDGPKASMVPSTYHSVSPPGY-----TILDVD-ANAMLFV  146
             Y+  ++  R G++G + +         S+  +         G      ++LD+D  + + +V
Sbjct  930   YFSIVKIERVGKHGKVFLTV-----PSLSSTAEEKFIKKGEFSGDDSLLDLDPEDTVFYV  984

Query  147   GGLTGKIKKADAVRVITFTGCMGETYFDNKPIGLWNFREKEGDCKGCTVSPQVEDSEGTI  206
             GG+    K   ++ +  F GC+     +N  I L+NF+    +    T  P   D
Sbjct  985   GGVPSNFKLPTSLNLPGFVGCLELATLNNDVISLYNFKHIY-NMDPSTSVPCARDKLAFT  1043
```

Figure 9. Blastp alignment of human wild type laminin alpha-4 to mouse laminin alpha-2 protein.

```
 Score = 26.9 bits (58),  Expect = 2e-04, Method: Compositional matrix adjust.
 Identities = 28/149 (19%), Positives = 65/149 (44%), Gaps = 24/149 (16%)

Query  41   TAVADNLLFYLGSAKFI-DFLAIEMRKGKVSFLWDVGSGVGRVEYPDLTIDD-------S  92
            T  A   + YLGS     +++ + ++    + +++++G+    VE P   +D        +
Sbjct  2    TETAHQFILYLGSKNAKKEYMGLAIKNDNLVYVYNLGTK--DVEIP---LDSKPVSSWPA  56

Query  93   YWYRIEASRTGRNGSISVRALDGPKASMVPSTYHSVSPPGY-----TILDVD-ANAMLFV  146
            Y+  ++  R G++G + +         S+  +         G      ++LD+D  + + +V
Sbjct  57   YFSIVKIERVGKHGKVFLTV-----PSLSSTAEEKFIKKGEFSGDDSLLDLDPEDTVFYV  111

Query  147  GGLTGKIKKADAVRVITFTGCMGETYFDN  175
            GG+    K   ++ +  F GC+     +N
Sbjct  112  GGVPSNFKLPTSLNLPGFVGCLELATLNN  140
```

Figure 10. Blastp alignment of human mutated (D879H) laminin alpha-4 to mouse laminin alpha-2 protein.

Homology modeling did not demonstrate significant structural changes between the wild type laminin alpha-4 protein and the mutated version (Figure 11). The D879H coding change occurs in the G1 domain of the protein. In mouse, the laminin alpha-2 LG1 domain is hypothesized to offer a necessary protein surface for integrin binding. Large surface regions not blocked by carbohydrate modifications are present in the LG1 domain. The faces of the LG1 β – sandwich are free of carbohydrates in the alpha chain of mouse (Carafoli et. al. 2009) and the

virtual human model is expected to be free of carbohydrates as well, therefore capable of binding. Since this domain is suggested to act as an adhesive factor for endothelial cells (Lian et. al. 2006), it was proposed that this mutation could disrupt proper cell binding, leading to catastrophic effects on the base framework of the circulatory system and cardiogenesis. Although the general structure of the mutated model does not appear to be altered as compared to the wild type model, it is still formally possible that the cell binding process is disrupted.

The laminin alpha-4 mutation is 40 bases from the proposed binding site amino acids 919-1207 (Gonzales 2002) and points out into solution. Depending on the pH, the histidine will express a negative or neutral charge. The change from the negative aspartic acid to the neutral or positive histidine could affect binding or structure. The mutation could lead to the formation of salt bridges with lysine or histidine or arginine within that sequence, potentially altering the structure. A change in the charge could remotely affect binding interactions. Without the availability of a crystal structure it is difficult to ascertain if any of these rearrangements is actually taking place.

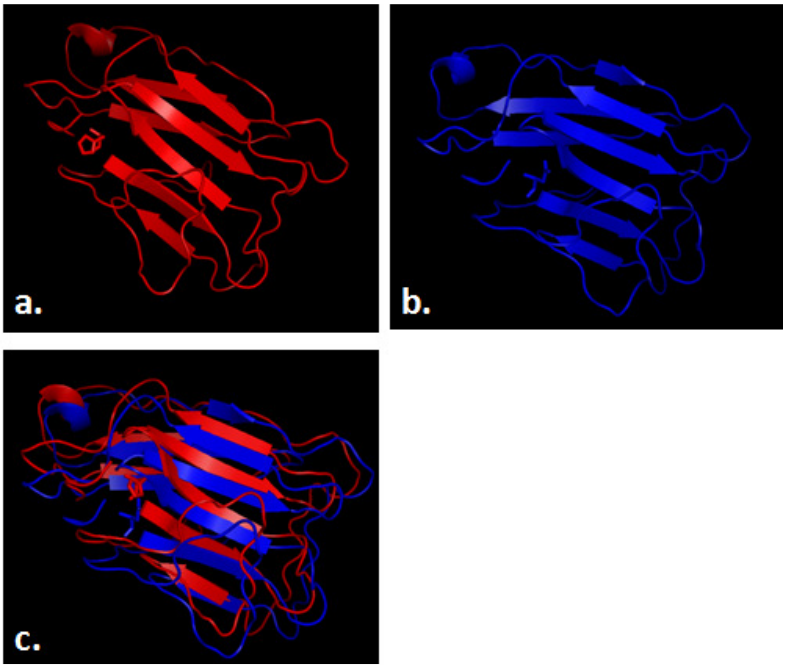


Figure 11. Homology models of laminin alpha-4 protein. (a) wild-type (b)D879H mutation mapped to the wild-type model (c) overlay of wild-type and mutated models.

This structural analysis, however, does not invalidate the potential connection that this variant may have with the HLHS pathway. It is difficult to analyze the impact of the LAMA4 mutation because the crystal structure of the human protein has not been solved. This homology model serves as a proposal of the protein and does not indicate this specific mutation causes structural differences between the two models.

4.4. LAMA4 and GJB2 Relationship - Pathway Analysis

GeneMANIA (Warde-Farley et. al. 2010) was employed to examine potential physical interactions of LAMA4 and GJB2. LAMA4 and GJB2 were queried with GO biological function based weighting. GeneMANIA provided evidence of physical interactions between the two genes mediated by DAG1 and CAV1 (Figure 12). DAG1 is a laminin binding component of the dystrophin-glycoprotein complex (Stelzer et. al. 2012). The interaction of LAMA2 and DAG1 has been well described. Caveolin1 (CAV1) interacts directly with G-protein alpha subunits and can functionally regulate their activity (Stelzer et. al. 2012). It is highly expressed in the vascular endothelium and to a lesser extent in cardiac fibroblasts and smooth muscle. CAV1 knockout mice express cardiac hypertrophy and left ventricular dilation and dysfunction. The heart of CAV1 knockout mice exhibit evidence for endothelial dysfunction as well (Krieger et. al. 2010).

The LAMA4-DAG1-CAV1-GJB2 physical interaction was heavily weighted based on the supporting reference data. The interaction between LAMA4 and DAG1 is supported by iRef-MINT and iRef - small scale studies sources. The DAG1 CAV1 interaction is supported by evidence from BioGRID small scale studies and iRef-small scale studies. BioGRID small scale studies, iRef -HPRD, iRef -OPHID, and iRef - small scale studies provide evidence of the interaction between CAV1 and GJB2. DAG, CAV1, and GJB2 are involved in cell-cell junction assembly. LAMA4 and DAG1 interactions are more closely related through ECM and basement membrane functions. DAG1 and CAV1 interactions involve protein maintenance and cell –cell junction assembly, as does GJB2.

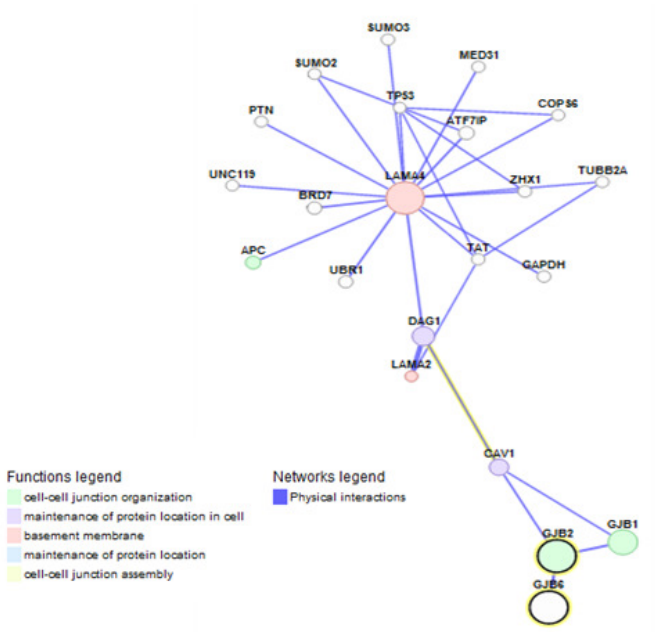


Figure 12: Physical interactions of Human LAMA4 and GJB2. GeneMANIA query results. The queried genes demonstrate direct and indirect physical interactions suggesting they may be involved in related biological processes. This is further supported by the common functions such as cell-cell junction organization, cell-cell junction assembly, and maintenance of protein locations shared by some genes in the network.

LAMA4, GJB2, GJA1, DAG1, and CAV1 were queried under similar conditions to determine if an obvious relationship was present that could be related associated to HLHS. Indirect physical interactions were present and well supported (Figure 13). A query for pathway relationships did not connect all genes involved unless the number of returned genes was expanded to 100 suggesting a distant relationship or a missing link between the two main groups and excluding GJB2 (Figure 14).

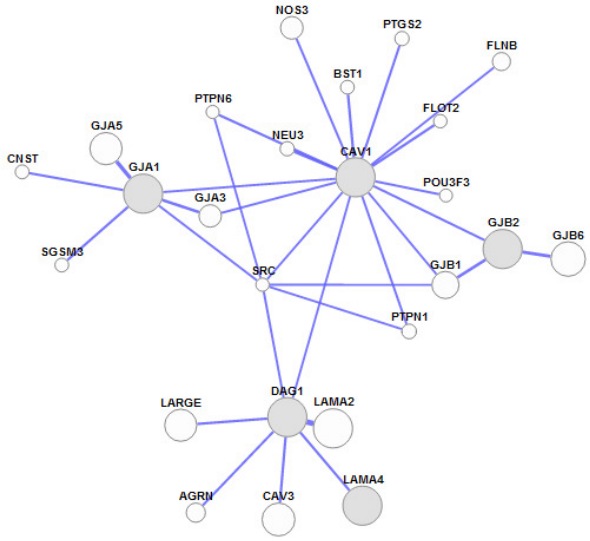


Figure 13. GeneMANIA analysis of LAMA4, GJB2, DAG1, CAV1, and GJA1 relationships. The queried genes demonstrate direct and indirect physical interactions suggesting they may be involved in related biological processes.

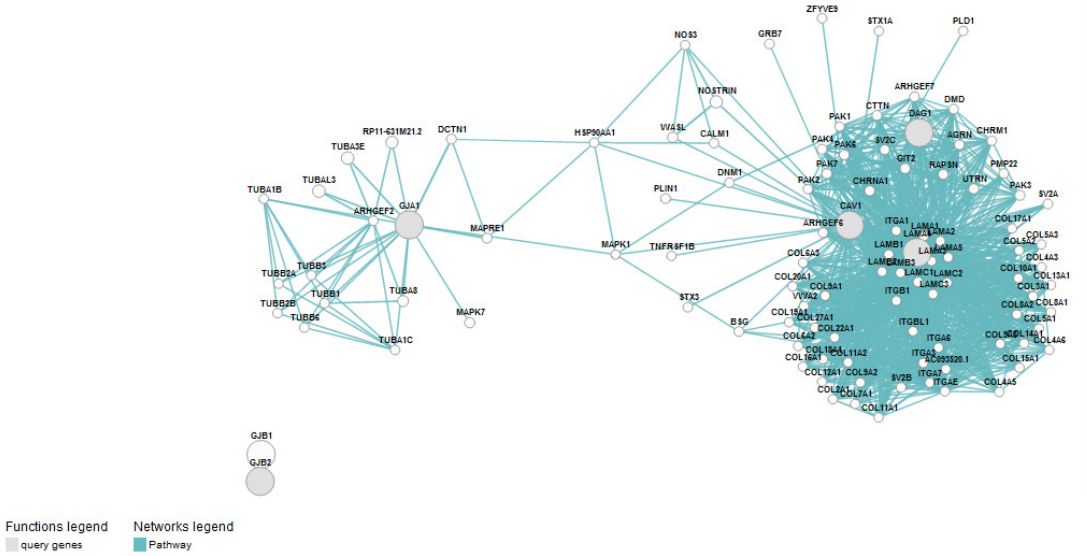


Figure 14. GeneMANIA analysis of GJB2, LAMA4, GJA1, DAG1, and CAV1. Network analysis shows extending the number of returned genes to 100 does not strongly connect all queried genes. Based on this analysis it is unlikely that together they are involved in the development of HLHS.

4.5 LAMA4 and GJB2 Relationship - Family Pedigree

The subject family is comprised of three sons. A history of cardiac anomalies on the maternal side has been documented. One of the sons does not carry any congenital heart defect (labeled as 'unaffected sibling'). One of the sons (labeled as 'proband') is affected by a serious

congenital heart defect known as hypoplastic left heart syndrome (HLHS). A daughter who did not survive due to a spontaneous abortion (fetal demise) also was affected by HLHS (labeled as 'fetus'). The DNA of the parents and these offspring were available at the onset of the study.

The DNA of another son became available later in the study. This child is affected with a longer than normal QT. A deaf phenotype caused by a deletion in the Connexin 26 (GJB2) gene is also present on the maternal side of the family. This child has a homozygous deletion in this gene and therefore expresses the deaf phenotype.

Sequencher DNA analysis indicated that both the father and mother are heterozygous for the deletion. The non-HLHS child is homozygous for the deletion and thereby expresses the deaf phenotype (Figure 15). The mother and non-HLHS child both are affected by non-HLHS cardiac abnormalities. It is unknown if the maternal grandmother also expresses a CHD phenotype. Perhaps some protection is created by the GJB2 deletion leading to less severe forms of CHD, however further experimentation will need to be conducted to verify this suggestion.

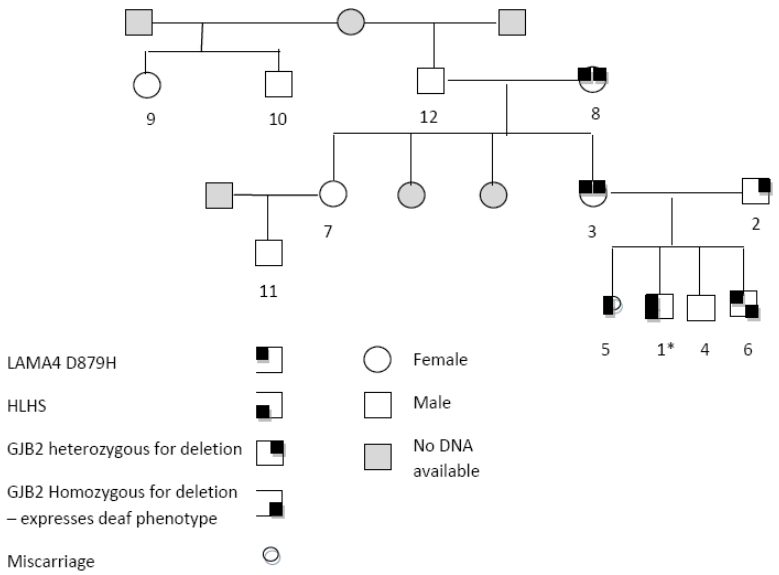


Figure 15. Family pedigree chart. Family members screened for LAMA4 and GJB2 by Sanger sequencing are indicated by the identification number under the nodes. Proband is indicated by *.

CHAPTER 5. CONCLUSION AND FUTURE DIRECTIONS

The evidence presented in this thesis suggests the D879H variant in LAMA4 is involved in the development of HLHS. Mutation predictors SIFT and PolyPhen-2 suggested that the amino acid change is damaging and novel. The expected amino acid is highly conserved across species implying that it is important to the function of the protein. Furthermore, all immediate family members that carry the mutated version of the allele exhibit a form of cardiac abnormalities.

Scientific literature indicates that the protein is expressed in the cardiovascular system during heart development and in the adult aorta. If HLHS in fact develops as a result of embryonic alterations in blood flow variations, this could indicate that LAMA4 is a culprit in the pathology of this syndrome. Losses of integrin binding during embryonic development affecting the shape of cardiomyocytes and subsequent changes in ventricle morphology would support this hypothesis due to the implications these changes would have on blood flow. Mouse knockout studies further support the importance of LAMA4 in the cardiovascular system and heart function by the appearance of leaky vasculature, premature death due to cardiac dysfunction, and the changes in cardiomyocytes and incidence of cardiomyopathy.

Homology modeling did not provide strong evidence that this particular mutation is causative. However, until the human structure is crystallized it is difficult to determine how accurate the model is. A model including the other LAMA4 non-novel variants found in the exome sequencing data may provide sufficient evidence to determine the inclusion of all the variations in this gene lead to an overall conformational change. The proposed structural effect could cause a disruption in protein binding leading to poor interaction of endothelial cell and cardiomyocytes with the extracellular matrix which could contribute to cardiomyopathy and compromised vasculature related to HLHS.

Pathway analysis indicates that LAMA4 interacts with genes such as PRKCA and AKAP13which have known cardiac associations. Perhaps examining only the variants with novel, nonsynonymous, and damaging characteristics is too limiting to determine the cause of this

syndrome. The indirect relationship of LAMA4 and GJA1 is interesting as GJA1 is described in the arrhythmogenic right ventricular cardiomyopathy pathway (ARVC). The suggested interaction with DAG1 is also promising as the KEGG database indicates it is involved with laminin in the Hypertrophic cardiomyopathy (HCM), Dilated cardiomyopathy (DCM), and Viral myocarditis disease pathways (Kanehisa et. al. 2000, 2012).

The relationship of AKAP13 and LAMA4 also may worthwhile of further exploration. Perhaps the coupling of mutations in these genes contributes to the development of HLHS. Three of the four children that carry both the LAMA4 and AKAP13 variants, two are diagnosed with HLHS, and the other has a longer than normal QT, possibly indicating dominant inheritance of the variant and reduced penetrance of the phenotype. IPA pathway analysis of the original variant list indicates that AKAP13 is related to cardiovascular disease and most specifically cardiac hypertrophy. AKAP13 is involved in a cellular response known as pathological cardiac hypertrophy. This syndrome includes increased cardiomyocyte size and greater organization of the sarcomere (Carnegie et. al. 2008). Since cardiomyocytes provide the foundation of sarcomeres which are responsible for heart contraction, mutations in these genes could affect blood flow. Interestingly DAG1 and CAV1 are members of subsarcolemmal complexes involved in contraction through supporting organ architecture, transmitting force between cells, and signal transduction pathways (Ahmad 2005). PRKCA is also involved in this pathway. GeneMANIA analysis indicated LAMA4 has a direct pathway connection to both PRKCA and AKAP13. PRKCA and AKAP13 have a direct physical interaction (Figure 4.) and their functional biological relationship has been well described (Carnegie et. al. 2004, Ahmad et. al. 2005). If HLHS is indeed due to an alteration in blood flow during embryogenesis, perhaps the combined effect of the mutations in these three genes is significant in this patient.

DNA sequencing combined with physical interaction analysis suggests that there may be an indirect relationship between hearing impairment caused by a deletion in GJB2, the D879H

mutation in the LAMA4 gene, and CHD. However, since neither of the HLHS children carry the GJB2 deletion, it is not possible to conclude this relationship is specific to HLHS. Additionally, the relationship is based on physical evidence that does not necessarily imply that the interactions are part of a biological pathway of the syndrome. The hypothesis cannot totally be discarded because even though the proteins are not known to interact at this time does not definitively mean the interaction is not possible. Research has suggested that there are links between hearing impairment and cardiac dysfunction, therefore the relationship of GJB2 to HLHS may not have been uncovered as of yet. Further investigation of these relationships may link these mutations to the development of CHD or reveal that the GJB2 deletion modulates the severity of CHD. This hypothesis is corroborated by evidence that the mother and one of the children do not show HLHS, and they are both positive for the LAMA4 variant and a deleted GJB2 allele.

Additional analysis such as binding and expression assays will be useful to further determine the involvement of the LAMA4 mutation in regards to HLHS. The inclusion of genes such as PRKCA, AKAP13, DAG1, GJB2, and CAV1 may result in a more in-depth understanding of the potential physical or pathway interactions. It also may prove worthwhile to compare pathway analysis of Osteogenesis Imperfecta (a connective tissue disorder related collagen, another integral basement membrane protein. Belmont et. al. 2011) with HLHS to further investigate the relationship between expression of the deaf phenotype and CHD in this family.

The systematic approach implemented in this study serves as an efficient and effective guideline to conduct unique variant detection in relationship to disease. Integrating genomic and proteomic methods along with the application of bioinformatic tools provides ample information to draw conclusions about the relationships of variants and the syndrome under investigation in an efficient and effective manner.

BIBLIOGRAPHY

Aberdam D, Virolle T, Simon-Assmann P. Transcriptional regulation of laminin gene expression. Microsc Res Tech. 2000 Nov 1;51(3):228-37. Review. PubMed PMID: 11054873.

Adams, J. DNA sequencing technologies. Nature Education (2008). 1(1)

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248-9. PubMed PMID: 20354512; PubMed Central

Ahmad F, Seidman JG, Seidman CE. The genetic basis for cardiac remodeling. Annu Rev Genomics Hum Genet. 2005;6:185-216. Review. PubMed PMID: 16124859.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10. PubMed PMID: 2231712.

Barron DJ, Kilby MD, Davies B, Wright JG, Jones TJ, Brawn WJ. Hypoplastic left heart syndrome. Lancet. 2009 Aug 15;374(9689):551-64. Review. PubMed PMID: 19683641.

Belkin AM, Stepp MA. Integrins as receptors for laminins. Microsc Res Tech. 2000 Nov 1;51(3):280-301. Review. PubMed PMID: 11054877.

Belmont JW, Craigen W, Martinez H, Jefferies JL. Genetic disorders with both hearing loss and cardiovascular abnormalities. AdvOtorhinolaryngol. 2011;70:66-74. Epub 2011 Feb 24. Review. PubMed PMID: 21358187.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol. 1977 May 25;112(3):535-42. PubMed PMID: 875032.

Borecki IB, Province MA. Genetic and genomic discovery using family studies. Circulation. 2008 Sep 2;118(10):1057-63. Review. PubMed PMID: 18765388.

PDB ID: 2WJS. Carafoli F, Clout NJ, Hohenester E. Crystal structure of the LG1-3 region of the laminin alpha2 chain. J Biol Chem. 2009 Aug 21;284(34):22786-92. Epub 2009 Jun 24. PubMed PMID: 19553699; PubMed Central PMCID: PMC2755686.

Carnegie GK, Soughayer J, Smith FD, Pedroja BS, Zhang F, Diviani D, Bristow MR, Kunkel MT, Newton AC, Langeberg LK, Scott JD. AKAP-Lbc mobilizes a cardiac hypertrophy signaling pathway. Mol Cell. 2008 Oct 24;32(2):169-79. PubMed PMID: 18951085; PubMed Central PMCID: PMC3169907.

Chang RK, Gurvitz M, Rodriguez S. Missed diagnosis of critical congenital heart disease. Arch Pediatr Adolesc Med. 2008 Oct;162(10):969-74. PubMed PMID: 18838650.

Centers for Disease Control and Prevention. Racial Differences by gestational Age in Neonatal Deaths Attributable to Congenital Heart Defects – United states 2003-2006. MMWR 2007. 59: 1208-1209.

National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention March 13, 2012 Pediatric Genetics, Screening for Critical Congenital Heart Defects Retrieved from:http://www.cdc.gov/ncbddd/pediatricgenetics/pulse.htmlLast Accessed March 18, 2012

Danes BS, Hodson ME, Batten J. Cystic fibrosis: Evidence for a genetic compound from a family study in cell culture. Clin Genet. 1977 Feb;11(2):83-90. PubMed PMID: 837566.

Dasgupta C, Martinez AM, Zuppan CW, Shah MM, Bailey LL, Fletcher WH. Identification of connexin43 (alpha1) gap junction gene mutations in patients with hypoplastic left heart syndrome by denaturing gradient gel electrophoresis (DGGE). Mutat Res. 2001 Aug 8;479(1-2):173-86. PubMed PMID: 11470490.

Davies Kevin. Seize the New: Laying the Clinical Genomics Pipeline in Wisconsin. Bio-IT World. (2011). 10(5):19-20, 22

DeHahn KC, Gonzales M, Gonzalez AM, Hopkinson SB, Chandel NS, Brunelle JK, Jones JC. The alpha4 laminin subunit regulates endothelial cell survival. Exp Cell Res. 2004 Mar 10;294(1):281-9. PubMed PMID: 14980521.

Draus JM Jr, Hauck MA, Goetsch M, Austin EH 3rd, Tomita-Mitchell A, Mitchell ME. Investigation of somatic NKX2-5 mutations in congenital heart disease. J Med Genet. 2009 Feb;46(2):115-22. PubMed PMID: 19181906; PubMed Central PMCID: PMC2628540.

Durbeej M. Laminins. Cell Tissue Res. 2010 Jan;339(1):259-68. Epub 2009 Aug 20. Review. PubMed PMID: 19693542.

Eghtesady P, Brar A, Hall M. Seasonality of hypoplastic left heart syndrome in the United States: a 10-year time-series analysis. J ThoracCardiovasc Surg. 2011 Feb;141(2):432-8. PubMed PMID: 20817208.

Elliott DA, Kirk EP, Yeoh T, Chandar S, McKenzie F, Taylor P, Grossfeld P, Fatkin D, Jones O, Hayes P, Feneley M, Harvey RP. Cardiac homeobox gene NKX2-5 mutations and congenital heart disease: associations with atrial septal defect and hypoplastic left heart syndrome. J Am CollCardiol. 2003 Jun 4;41(11):2072-6. PubMed PMID: 12798584.

Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/). Accessed February 16, 2012.

Feinstein JA, Benson DW, Dubin AM, Cohen MS, Maxey DM, Mahle WT, Pahl E, Villafañe J, Bhatt AB, Peng LF, Johnson BA, Marsden AL, Daniels CJ, Rudd NA, Caldarone CA, Mussatto KA, Morales DL, Ivy DD, Gaynor JW, Tweddell JS, Deal BJ, Furck AK, Rosenthal GL, Ohye RG, Ghanayem NS, Cheatham JP, Tworetzky W, Martin GR. Hypoplastic left heart syndrome: current considerations and expectations. J Am CollCardiol. 2012 Jan 3;59(1 Suppl):S1-42. Review. Erratum in: J Am CollCardiol. 2012 Jan 31;59(5):544. PubMed PMID: 22192720.

Fogoros M.D. Richard N. The Heart's Chambers and Valves. About.com Health's Disease and Condition Guide. Updated December 01, 2003. http://biology.about.com/od/anatomy/ss/ventricles.htm Date of last access: March 12, 2012.

Frenette PS, Atweh GF. Sickle cell disease: old discoveries, new concepts, and future promise. J Clin Invest. 2007 Apr;117(4):850-8. Review. PubMed PMID: 17404610; PubMed Central PMCID: PMC1838946.

Frieser M, Nöckel H, Pausch F, Röder C, Hahn A, Deutzmann R, Sorokin LM. Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of endothelium. Eur J Biochem. 1997 Jun 15;246(3):727-35. PubMed PMID: 9219532.

Fuerst FC, Kielhauser SM, Gruber G, Stradner MH, Angerer H, Kremser ML, Setzagel D, Graninger WB. Matrix Metalloproteinase 3 is regulated by the laminin LAMA4 in Human Osteoarthritis. Ann Rheum Dis. 2011. 70(Suppl 2):A1-A94

The Gene Ontology Consortium. "Gene ontology: tool for the unification of biology." Nat. Genet. May 2000;25(1):25-9.

Gilboa SM, Salemi JL, Nembhard WN, Fixler DE, Correa A. Mortality resulting from congenital heart disease among children and adults in the United States, 1999 to 2006. Circulation. 2010 Nov 30;122(22):2254-63. Epub 2010 Nov 22. PubMed PMID: 21098447.

Gonzalez AM, Gonzales M, Herron GS, Nagavarapu U, Hopkinson SB, Tsuruta D, Jones JC. Complex interactions between the laminin alpha 4 subunit and integrins regulate endothelial cell behavior in vitro and angiogenesis in vivo. ProcNatlAcadSci U S A. 2002 Dec 10;99(25):16075-80. Epub 2002 Nov 26. PubMed PMID: 12454288; PubMed Central PMCID:PMC138567.

Gottfried I, Landau M, Glaser F, Di WL, Ophir J, Mevorah B, Ben-Tal N, Kelsell DP, Avraham KB. A mutation in GJB3 is associated with recessive erythrokeratodermiavariabilis (EKV) and leads to defective trafficking of the connexin 31 protein. Hum Mol Genet. 2002 May 15;11(11):1311-6. PubMed PMID: 12019212.

Grossfeld P, Ye M, Harvey R. Hypoplastic left heart syndrome: new genetic insights. J Am Coll Cardiol. 2009 Mar 24;53(12):1072-4. PubMed PMID: 19298922.

Grossfeld PD. Hypoplastic left heart syndrome: it is all in the genes. J Am CollCardiol. 2007 Oct 16;50(16):1596-7. Epub 2007 Oct 1. PubMed PMID: 17936160.

Grossfeld PD. The genetics of hypoplastic left heart syndrome. Cardiol Young. 1999 Nov;9(6):627-32. Review. PubMed PMID: 10593278. Cited in PMCRelated citations

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D590-8. PubMed PMID: 16381938; PubMed Central PMCID: PMC1347506.

Hinton RB, Martin LJ, Rame-Gowda S, Tabangin ME, Cripe LH, Benson DW. Hypoplastic left heart syndrome links to chromosomes 10q and 6q and is genetically related to bicuspid aortic valve. J Am CollCardiol. 2009 Mar 24;53(12):1065-71. PubMed PMID: 19298921; PubMed Central PMCID: PMC2703749.

Hinton RB Jr, Martin LJ, Tabangin ME, Mazwi ML, Cripe LH, Benson DW. Hypoplastic left heart syndrome is heritable. J Am CollCardiol. 2007 Oct 16;50(16):1590-5. Epub 2007 Oct 1. PubMed PMID: 17936159.

Hoffman JI. Incidence of congenital heart disease: I. Postnatal incidence. PediatrCardiol. 1995 May-Jun;16(3):103-13. Review. PubMed PMID: 7617503.

Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am CollCardiol. 2002 Jun 19;39(12):1890-900. Review. PubMed PMID: 12084585.

Iivanainen A, Sainio K, Sariola H, Tryggvason K. Primary structure and expression of a novel human laminin alpha 4 chain.FEBS Lett.1995 May29;365(2-3):183-8.PubMedPMID:7781776.

Imbeault S, Gauvin LG, Toeg HD, Pettit A, Sorbara CD, Migahed L, DesRoches R, Menzies AS, Nishii K, Paul DL, Simon AM, Bennett SA. The extracellular matrix controls gap junction protein expression and function in postnatal hippocampal neural progenitor cells. BMC Neurosci. 2009 Feb 24;10:13. PubMed PMID: 19236721; PubMed Central PMCID: PMC2655299.

Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, Suzek BE, Martin MJ, McGarvey P, Gasteiger E. Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics. 2009 May 8;10:136. PubMed PMID: 19426475; PubMed Central PMCID: PMC2686714.

Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.; KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Res. 40, D109-D114 (2012).

Kemperman MH, Hoefsloot LH, Cremers CW. Hearing loss and connexin 26. J R Soc Med. 2002 Apr;95(4):171-7. Review. PubMed PMID: 11934905; PubMed Central PMCID:PMC1279509.

Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, Geneviève D, Cormier-Daire V, van Esch H, Fryns JP, Hamel BC, Sistermans EA, de Vries BB, van Bokhoven H. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am J Hum Genet. 2006 Aug;79(2):370-7.Epub 2006 Jun 13. PubMed PMID: 16826528; PubMed Central PMCID: PMC1559478.

Knapp, AA, Metterville, DR, Kemper, AR, Prosser, L, Perrin, JM. Evidence review: Critical congenital cyanotic heart disease, Final Draft, September 3, 2010. Prepared for the Maternal and Child Health Bureau, Health Resources and Services Administration.

Knöll R, Postel R, Wang J, Krätzner R, Hennecke G, Vacaru AM, Vakeel P, Schubert C, Murthy K, Rana BK, Kube D, Knöll G, Schäfer K, Hayashi T, Holm T, Kimura A, Schork N, Toliat MR, Nürnberg P, Schultheiss HP, Schaper W, Schaper J, Bos E, Den Hertog J, van Eeden FJ, Peters PJ, Hasenfuss G, Chien KR, Bakkers J. Laminin-alpha4 and integrin-linked kinase mutations cause human cardiomyopathy via simultaneous defects in cardiomyocytes and endothelial cells. Circulation. 2007 Jul 31;116(5):515-25. Epub 2007 Jul 23. PubMed PMID: 17646580.

Krieger MH, Di Lorenzo A, Teutsch C, Kauser K, Sessa WC. Telmisartan regresses left ventricular hypertrophy in caveolin-1-deficient mice. Lab Invest. 2010 Nov;90(11):1573-81. Epub 2010 Jun 28. PubMed PMID: 20585312; PubMed Central PMCID: PMC3248785.

Ladage D, Tilemann L, Ishikawa K, Correll RN, Kawase Y, Houser SR, Molkentin JD, Hajjar RJ. Inhibition of PKCα/β with ruboxistaurin antagonizes heart failure in pigs after myocardial infarction injury. Circ Res. 2011 Dec 9;109(12):1396-400. Epub 2011 Oct 13. PubMed PMID: 21998327; PubMed Central PMCID: PMC3237728.

Lewin MB, McBride KL, Pignatelli R, Fernbach S, Combes A, Menesses A, Lam W, Bezold LI, Kaplan N, Towbin JA, Belmont JW. Echocardiographic evaluation of asymptomatic parental and sibling cardiovascular anomalies associated with congenital left ventricular outflow tract lesions. Pediatrics. 2004 Sep;114(3):691-6. PubMed PMID: 15342840; PubMed Central PMCID: PMC1361301.

Lian J, Dai X, Li X, He F. Identification of an active site on the laminin alpha4 chain globular domain that binds to alphavbeta3 integrin and promotes angiogenesis. Biochem Biophys Res Commun. 2006 Aug 18;347(1):248-53. Epub 2006 Jun 21. PubMed PMID: 16824487.

Libby RT, Champliaud MF, Claudepierre T, Xu Y, Gibbons EP, Koch M, Burgeson RE, Hunter DD, Brunken WJ. Laminin expression in adult and developing retinae: evidence of two novel CNS laminins. J Neurosci. 2000 Sep 1;20(17):6517-28. PubMed PMID: 10964957; PubMed Central PMCID: PMC2924637.

Lin A, Wang RT, Ahn S, Park CC, Smith DJ. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. Genome Res. 2010 Aug;20(8):1122-32. Epub 2010 May 27. PubMed PMID: 20508145; PubMed Central PMCID: PMC2909575.

Lin P, Mao F, Liu Q, Yang W, Shao C, Yan C, Gong Y. A novel deletion mutation in GJB1 causes X-linked Charcot-Marie-Tooth disease in a Han Chinese family. Muscle Nerve. 2010 Dec;42(6):922-6. PubMed PMID: 21104867.

Liu J, Mayne R. The complete cDNA coding sequence and tissue-specific expression of the mouse laminin alpha 4 chain. Matrix Biol. 1996 Dec;15(6):433-7. PubMed PMID: 9049981.

Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011 Mar 29;2011:bar009. Print 2011. PubMed PMID: 21447597; PubMed Central PMCID: PMC3070428.

Mammano F. Gap Junctions: Cell-Cell Channels in Animals. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-.Accessed March 16, 2012.

McBride KL, Zender GA, Fitzgerald-Butt SM, Seagraves NJ, Fernbach SD, Zapata G, Lewin M, Towbin JA, Belmont JW. Association of common variants in ERBB4 with congenital left ventricular outflow tract obstruction defects. Birth Defects Res A ClinMolTeratol. 2011 Mar;91(3):162-8. doi: 10.1002/bdra.20764. Epub 2011 Feb 2. PubMed PMID: 21290564.

McBride KL, Riley MF, Zender GA, Fitzgerald-Butt SM, Towbin JA, Belmont JW, Cole SE. NOTCH1 mutations in individuals with left ventricular outflow tract malformationsreduce ligand-induced signaling. Hum Mol Genet. 2008 Sep 15;17(18):2886-93. Epub 2008 Jun 30. PubMed PMID: 18593716; PubMed Central PMCID: PMC2722892.

McGowan KA, Marinkovich MP. Laminins and human disease. Microsc Res Tech. 2000 Nov 1;51(3):262-79. Review. PubMed PMID: 11054876.

Messina DN, Speer MC, Pericak-Vance MA, McNally EM. Linkage of familial dilated cardiomyopathy with conduction defect and muscular dystrophy to chromosome 6q23. Am J Hum Genet. 1997 Oct;61(4):909-17. PubMed PMID: 9382102; PubMed Central PMCID: PMC1715999.

Metzker ML. Emerging technologies in DNA sequencing. Genome Res. 2005 Dec;15(12):1767-76. Review. PubMed PMID: 16339375.

Mossé YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF, Laquaglia MJ, Sennett R, Lynch JE, Perri P, Laureys G, Speleman F, Kim C, Hou C, Hakonarson H, Torkamani A, Schork NJ, Brodeur GM, Tonini GP, Rappaport E, Devoto M, Maris JM. Identification of ALKas a major familial neuroblastoma predisposition gene. Nature. 2008 Oct 16;455(7215): 930-5. Epub 2008 Aug 24. PubMed PMID: 18724359; PubMed CentralPMCID:PMC2672043.

Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. 2008;9 Suppl 1:S4. Epub 2008 Jun 27. PubMed PMID: 18613948; PubMed Central PMCID: PMC2447538.

Mullin JM, Laughlin KV, Ginanni N, Marano CW, Clarke HM, Peralta Soler A. Increased tight junction permeability can result from protein kinase C activation/translocation and act as a tumor promotional event in epithelial cancers. Ann N Y Acad Sci. 2000;915:231-6. Review. PubMed PMID: 11193580.

Nath AK, Krauthammer M, Li P, Davidov E, Butler LC, Copel J, Katajamaa M, Oresic M, Buhimschi I, Buhimschi C, Snyder M, Madri JA. Proteomic-based detection of a protein cluster dysregulated during cardiovascular development identifies biomarkers of congenital heart defects. PLoS One. 2009;4(1):e4221. Epub 2009 Jan 19. PubMed PMID: 19156209; PubMed Central PMCID: PMC2626248.

National Institues of Health, Department of Health and Human Servies. July 1, 2011. National Heart Lung and Blood Institute (NHLBI). What are Congenetial Heart defects? Retrieved from. http://www.nhlbi.nih.gov/health/health-topics/topics/chd/. Most recently accessed March 22, 2012

Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003 Jul 1;31(13):3812-4. PubMed PMID: 12824425; PubMed Central PMCID: PMC168916.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. ThePfam protein families database. Nucleic Acids Res. 2012 Jan;40(Database issue):D290-301. Epub 2011 Nov 29. PubMed PMID: 22127870; PubMed Central PMCID: PMC3245129.

The PyMOL Molecular Graphics System, Version 0.99 Schrödinger, LLC.

Rajagopal SK, Ma Q, Obler D, Shen J, Manichaikul A, Tomita-Mitchell A, Boardman K, Briggs C, Garg V, Srivastava D, Goldmuntz E, Broman KW, Benson DW, Smoot LB, Pu WT. Spectrum of heart disease associated with murine and human GATA4 mutation. J Mol Cell Cardiol. 2007 Dec;43(6):677-85. Epub 2007 Jun 21. PubMed PMID: 17643447; PubMed Central PMCID: PMC2573470.

Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002 Sep 1;30(17):3894-900. PubMed PMID: 12202775; PubMed Central PMCID: PMC137415.

Reamon-Buettner SM, Ciribilli Y, Inga A, Borlak J. A loss-of-function mutation in the binding domain of HAND1 predicts hypoplasia of the human hearts. Hum Mol Genet. 2008 May 15;17(10):1397-405. Epub 2008 Feb 14. PubMed PMID: 18276607.

Richards A, Luccarini C, Pope FM. The structural organisation of LAMA4, the gene encoding laminin alpha4. Eur J Biochem. 1997 Aug 15;248(1):15-23. PubMed PMID: 9310354.

Ringelmann B, Röder C, Hallmann R, Maley M, Davies M, Grounds M, Sorokin L. Expression of laminin alpha1, alpha2, alpha4, and alpha5 chains, fibronectin, and tenascin-C in skeletal muscle of dystrophic 129ReJ dy/dy mice. Exp Cell Res. 1999 Jan 10;246(1):165-82. PubMed PMID: 9882526.

Roche Applied Science. 2007. Retrieved fromhttp://www.roche-applied-science.com/publications/multimedia/genome_sequencer/amplicon_07/wbt.htm. The genome Sequencer introduced by Roche Diagnostics and 454 Life Sciences.Last Accessed March 22, 2012

Rogaev EI, Grigorenko AP, Faskhutdinova G, Kittler EL, Moliaka YK. Genotype analysis identifies the cause of the "royal disease". Science. 2009 Nov 6;326(5954):817. Epub 2009 Oct 8. PubMed PMID: 19815722.

Šali A. and Blundell T. L.. Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 199

Sander TL, Klinkner DB, Tomita-Mitchell A, Mitchell ME. Molecular and cellular basis of congenital heart disease. PediatrClin North Am. 2006 Oct;53(5):989-1009, x. Review. Erratumin: PediatrClin North Am. 2009 Aug;56(4):xix. PubMed PMID: 17027620.

Scott DL, Salmon M, Morris CJ, Wainwright AC, Walton KW. Laminin and vascular proliferation in rheumatoid arthritis. Ann Rheum Dis. 1984 Aug;43(4):551-5. PubMed PMID: 6383233; PubMed Central PMCID: PMC1001406.

Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008 Oct;26(10):1135-45. PubMed PMID: 18846087.

Siffel C, Alverson CJ, Correa A. Analysis of seasonal variation of birth defects in Atlanta. Birth Defects Res A ClinMolTeratol. 2005 Oct;73(10):655-62. PubMed PMID: 16240376.

Snoeckx RL, Hassan DM, Kamal NM, Van Den Bogaert K, Van Camp G. Mutation analysis of the GJB2 (connexin 26) gene in Egypt. Hum Mutat. 2005 Jul;26(1):60-1. PubMed PMID: 15954104.

Snoeckx RL, Huygen PL, Feldmann D, Marlin S, Denoyelle F, Waligora J, Mueller-Malesinska M, Pollak A, Ploski R, Murgia A, Orzan E, Castorina P, Ambrosetti U, Nowakowska-Szyrwinska E, Bal J, Wiszniewski W, Janecke AR, Nekahm-Heis D, Seeman P, Bendova O, Kenna MA, Frangulov A, Rehm HL, Tekin M, Incesulu A, Dahl HH, du Sart D, Jenkins L, Lucas D, Bitner-Glindzicz M, Avraham KB, Brownstein Z, del Castillo I, Moreno F, Blin N, Pfister M, Sziklai I, Toth T, Kelley PM, Cohn ES, Van Maldergem L, Hilbert P, Roux AF, Mondain M, Hoefsloot LH, Cremers CW, Löppönen T, Löppönen H, Parving A, Gronskov K, Schrijver I, Roberson J, Gualandi F, Martini A, Lina-Granade G, Pallares-Ruiz N, Correia C, Fialho G, Cryns K, Hilgert N, Van de Heyning P, Nishimura CJ, Smith RJ, Van Camp G. GJB2 mutations and degree of hearing loss: a multicenter study. Am J Hum Genet. 2005 Dec;77(6):945-57. Epub 2005 Oct 19. PubMed PMID: 16380907; PubMed Central PMCID: PMC1285178.

Stallmeyer B, Fenge H, Nowak-Göttl U, Schulze-Bahr E. Mutational spectrum in the cardiac transcription factor gene NKX2.5 (CSX) associated with congenital heart disease. Clin Genet. 2010 Dec;78(6):533-40. doi: 10.1111/j.1399-0004.2010.01422.x. PubMed PMID: 20456451.

Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, Oz-Levi D, Olender T, Belinky F, Bahir I, Krug H, Perco P, Mayer B, Kolker E, Safran M, Lancet D. In-silico human genomics with GeneCards. Hum Genomics. 2011 Oct;5(6):709-17. PubMed PMID: 22155609. http://www.genecards.org/cgi-in/carddisp.pl?gene=LAMA4&search=LAMA4 Date Last Accessed: March 15, 2012.

Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, Oz-Levi D, Olender T, Belinky F, Bahir I, Krug H, Perco P, Mayer B, Kolker E, Safran M, Lancet D. In-silico human genomics with GeneCards. Hum Genomics. 2011 Oct;5(6):709-17. PubMed PMID: 22155609. http://www.genecards.org/cgi-bin/carddisp.pl?gene=GJA1 Date Last Accessed: March 15, 2012.

Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, Oz-Levi D, Olender T, Belinky F, Bahir I, Krug H, Perco P, Mayer B, Kolker E, Safran M, Lancet D. In-silico human genomics with GeneCards. Hum Genomics. 2011 Oct;5(6):709-17. PubMed PMID: 22155609. http://www.genecards.org/cgi-bin/carddisp.pl?gene=PRKCA&search=prkca Date Last Accessed: March 15, 2012.

Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, Oz-Levi D, Olender T, Belinky F, Bahir I, Krug H, Perco P, Mayer B, Kolker E, Safran M, Lancet D. In-silico human genomics with GeneCards. Hum Genomics. 2011 Oct;5(6):709-17. PubMed PMID: 22155609. http://www.genecards.org/cgi-bin/carddisp.pl?gene=DAG1 Date Last Accessed: March 15, 2012.

Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, Oz-Levi D, Olender T, Belinky F, Bahir I, Krug H, Perco P, Mayer B, Kolker E, Safran M, Lancet D. In-silico human genomics with GeneCards. Hum Genomics. 2011 Oct;5(6):709-17. PubMed PMID: 22155609. http://www.genecards.org/cgi-bin/carddisp.pl?gene=CAV1 Date Last Accessed: March 15, 2012.

Stumper O. Hypoplastic left heart syndrome. Postgrad Med J. 2010 Mar;86(1013):183-8. Review. PubMed PMID: 20237013.

Suzuki N, Yokoyama F, Nomizu M. Functional sites in the laminin alpha chains. Connect Tissue Res. 2005;46(3):142-52. Review. PubMed PMID: 16147852.

Tchervenkov CI, Jacobs JP, Weinberg PM, Aiello VD, Béland MJ, Colan SD, Elliott MJ, Franklin RC, Gaynor JW, Krogmann ON, Kurosawa H, Maruszewski B, Stellin G. The nomenclature, definition and classification of hypoplastic left heart syndrome. Cardiol Young. 2006 Aug;16(4):339-68. Review. PubMed PMID: 16839428.

Thyboll J, Kortesmaa J, Cao R, Soininen R, Wang L, Iivanainen A, Sorokin L, Risling M, Cao Y, Tryggvason K. Deletion of the laminin alpha4 chain leads to impaired microvessel maturation. Mol Cell Biol. 2002 Feb;22(4):1194-202. PubMed PMID: 11809810; PubMed Central PMCID: PMC134646.

Tomita-Mitchell A, Mahnke DK, Struble CA, Tuffnell ME, Stamm KD, Hidestrand M, Harris SE, Goetsch MA, Simpson PM, Bick DP, Broeckel U, Pelech AN, Tweddell JS, Mitchell ME. Human Gene Copy Number Spectra Analysis in Congenital Heart Malformations. Physiol Genomics. 2012 Feb 7. [Epub ahead of print] PubMed PMID: 22318994.

Tomita-Mitchell A, Maslen CL, Morris CD, Garg V, Goldmuntz E. GATA4 sequence variants in patients with congenital heart disease. J Med Genet. 2007 Dec;44(12):779-83. PubMed PMID: 18055909; PubMed Central PMCID: PMC2652815.

Tunggal P, Smyth N, Paulsson M, Ott MC. Laminins: structure and genetic regulation. Microsc Res Tech. 2000 Nov 1;51(3):214-27. Review. PubMed PMID: 11054872.

Van Agtmael T, Bruckner-Tuderman L. Basement membranes and human disease. Cell Tissue Res. 2010 Jan;339(1):167-88. Epub 2009 Sep 15. Review. PubMed PMID: 19756754.

VanDeVoorde R, Witte D, Kogan J, Goebel J. Pierson syndrome: a novel cause of congenital nephrotic syndrome. Pediatrics. 2006 Aug;118(2):e501-5. Epub 2006 Jul 24. PubMed PMID: 16864643.

Vaughan CJ, Basson CT. Molecular determinants of atrial and ventricular septal defects and patent ductus arteriosus. Am J Med Genet. 2000 Winter;97(4):304-9. Review. PubMed PMID: 11376442.

Wang J, Hoshijima M, Lam J, Zhou Z, Jokiel A, Dalton ND, Hultenby K, Ruiz-Lozano P, Ross J Jr, Tryggvason K, Chien KR. Cardiomyopathy associated with microcirculation dysfunction in laminin alpha4 chain-deficient mice. J Biol Chem. 2006 Jan 6;281(1):213- 20. Epub 2005 Oct 3. PubMed PMID: 16204254.

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010 Jul;38(Web Server Issue):W214-20. PubMed PMID: 20576703; PubMed Central PMCID: PMC2896186.

Wat MJ, Shchelochkov OA, Holder AM, Breman AM, Dagli A, Bacino C, Scaglia F, Zori RT, Cheung SW, Scott DA, Kang SH. Chromosome 8p23.1 deletions as a cause of complex congenital heart defects and diaphragmatic hernia. Am J Med Genet A. 2009 Aug;149A(8):1661-77. PubMed PMID: 19606479; PubMed Central PMCID: PMC2765374.

Wolfsberg, T. G. Using the NCBI Map Viewer to Browse Genomic Sequence Data. Current Protocols in Human Genetics. 2011. 69:18.5.1–18.5.25.

Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP Making a definitive diagnosis: successful clinical application of wholeexome sequencing in a child withintractable inflammatory bowel disease. Genet Med.2011 Mar;13(3):255-62. PubMed PMID: 21173700.

Xu H, Czerwinski P, Hortmann M, Sohn HY, Förstermann U, Li H. Protein kinase C alpha promotes angiogenic activity of human endothelial cells via induction of vascular endothelial growth factor. Cardiovasc Res. 2008 May 1;78(2):349-55. Epub 2007 Dec 4. PubMed PMID: 18056764.

Zhu L, Vranckx R, Khau Van Kien P, Lalande A, Boisset N, Mathieu F, Wegman M, Glancy L, Gasc JM, Brunotte F, Bruneval P, Wolf JE, Michel JB, Jeunemaitre X. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. Nat Genet. 2006 Mar;38(3):343-9. Epub 2006 Jan 29. PubMed PMID: 16444274.

Xiao Wu, Sharpe K., Tianyi Zhang, Hongyan Chen, Wei Zhu, Li, E., Taghavi S., Van Der Lelie, D. Comparative genetic pathway analysis using structural equation ModelingComputational Advances in Bio and Medical Sciences (ICCABS), 2011 IEEE 1st International Conference OnDigital Object Identifier: 10.1109/ICCABS.2011.5729878Publication Year: 2011 Page(s):190 195