ANALYSIS OF GENETIC MUTATIONS CONTRIBUTING TO CONGENITAL HEART DEFECTS IN AN ENU INDUCED MOUSE MODEL

by

Laura Carolyn Smith

BS in Biochemistry, Eastern Nazarene College, 2012

Submitted to the Graduate Faculty of

Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Laura Carolyn Smith

It was defended on

December 5, 2013

and approved by

Thesis Advisor:
Cecilia Lo, Ph.D.
Professor and Chair
Department of Developmental Biology
School of Medicine, University of Pittsburgh

Committee Member:
M. Ilyas Kamboh, Ph.D.
Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Candace Kammerer, Ph.D.
Associate Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

ANALYSIS OF GENETIC MUTATIONS CONTRIBUTING TO CONGENITAL HEART DEFECTS IN AN ENU INDUCED MOUSE MODEL

Laura Carolyn Smith, M.S.

University of Pittsburgh, 2013

ABSTRACT

Congenital heart defects (CHD) are the most common birth malformation affecting about 8 infants per 1,000 live births and is responsible for the majority of prenatal deaths. Currently, the fundamental cause of most CHD cases remains unknown. Congenital heart defects pose a serious public health concern due to the high incidence and mortality among the population. By improving the understanding of causation of CHD, we may better understand the pathological basis of the disease and find disease risk for patients and their families.

We conducted a forward genetic screen using ENU-induced mice in order to find the majority of disease-causing genes for congenital heart defects. We were able to recover a total of 88 mutations within 62 disease-causing genes that contribute to both cardiac and non-cardiac mutant phenotypes from 146 mouse lines. Of these variants, 59 were missense, 16 were splicing, and 13 were nonsense.

There are 21 mutations from our screen that were found to be in the same domain as published human mutations. The mutations are similar in the sense that they occur in the same protein domain, but not necessarily the same exon. These mutations also present similar phenotypes in both mice and humans. Three of these mutations are incredibly similar to published human mutations due to the fact that they are within one or two amino acids or nucleotides from each other.

All in all, we can conclude that the 88 disease-causing mutations recovered from our genetic screen are mostly damaging. Three-fourths (75%) of the mutations were completely damaging. About 17% were possibly damaging, 2% were possibly tolerated, and 1% was tolerated. The remaining 5% are splicing mutations that were more than two base pairs from the exon and therefore did not have a definitive conservation call. Overall, there is only one mutation in our screen that is tolerated. This information proves that mutations which are disease causing occur in highly conserved locations. Mutations from our screen, especially those that are similar to published human mutations, are important in determining the genotype-phenotype relationship in congenital heart defects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank Dr. Cecilia Lo for giving me this opportunity to work in her lab. I have learned many techniques that will help me in my future career. I would also like to thank You Li for his guidance and extensive contribution to the ENU screen. Many thanks to Nikolai Klena for his hard work in validating the disease-causing genes and for his help in answering questions and teaching me about the screen. Thanks to Cheng Cui for his patience in teaching me the necessary lab skills. Thank you Shazina for your guidance, support, and encouragement. Also thank you to the rest of the Lo lab for your tremendous support.

I would also like to thank my friends and family for their continuous support throughout my time in graduate school. Even though I was far from home, I could still count on you all for your love, prayers, and support.

1.0 BACKGROUND AND SIGNIFICANCE

1.1 CONGENITAL HEART DEFECTS

Congenital heart defects (CHD) are the most common birth malformation affecting about 8 infants per 1,000 live births and is responsible for the majority of prenatal deaths (Mitchell 2007, Pierpont 2007). The genetic etiology and molecular mechanisms of CHD remain unknown despite many research studies, which suggest strong evidence for a genetic component. The prevalence of CHD in 2000 was estimated to be 623,000 people: 320,000 with simple lesions, 165,000 with moderately complex disease, and 138,000 with highly complex disease (Pierpont 2007). By improving the understanding of causation of CHD, we may better understand the pathological basis of the disease, find disease risk for patients and their families, and identify potential therapies.

It has been proposed that CHD is a multifactorial disease consisting of genetic, environmental, and teratogenic factors acting as disease modifiers (Botto 2003). Some teratogenic factors that may contribute to CHD include maternal folic acid intake, therapeutic drug exposure, PKU, and pregestational diabetes to name a few (Jenkins 2007). Most cases however are due to sporadic, de novo mutations as opposed to heritable, familial mutations making it difficult to deduce the precise cause of the phenotype. Additionally, there is a wide

range of phenotypes and associative diseases in CHD, which makes finding a genotype-phenotype relationship even more difficult.

The mechanistic complexity of CHD could account for the wide variation in phenotypes among humans. The genetic etiology has been hard to determine due to the limited correlation between the genetic mutation and the specific phenotype. Some reasons for the complexity of CHD include: genetic heterogeneity in which similar cardiac defects are caused by more than one genetic mutation (Benson 2000); variable expressivity in which one gene may cause multiple phenotypes (Benson 2000); reduced penetrance where a person may be a carrier for the disease-causing mutation but not express the CHD phenotype (Benson 1998); one or more causative genes for CHD associated with chromosomal alterations (Momma 1995); single gene defects with or without the expression of the CHD phenotype (McElhinney 2002).

1.2 ENU FORWARD GENETIC SCREEN

The Lo lab is conducting a large-scale forward genetic screen in mice using ethylnitrosourea in order to better understand the genetic etiology of congenital heart defects. Ethylnitrosourea, also known as ENU, is an efficient mutagen, which induces a point mutation every 1 to 2 megabases throughout the genome (Kile 2005). This mutation rate is approximately 100-fold higher than the spontaneous mutation rate and only affects a single locus, making it ideal for genetic screens. With the introduction of the ENU mutagen, it is now possible to accelerate the rate of mutagenesis in the mouse model instead of having to wait for spontaneous mutations.

1.3 WHOLE EXOME SEQUENCING

Whole exome sequencing (WES) is a technique used to identify and sequence the coding regions of all annotated protein coding genes (Goh 2012). It can be divided into three parts: preparation of genomic DNA libraries, hybridization of these libraries, and sequencing of the hybridized fragments.

There are some downfalls to using whole exome sequencing instead of whole genome sequencing. Since whole exome sequencing only identifies the sequence of the coding regions many structural and other functional parts are not being included that may also contribute to disease. WES does not cover untranslated regions, enhancers, and long-noncoding RNAs, which may be potentially functional (Goh 2012). Structural variations such as copy number variants, translocations, and inversions are also not included in whole exome sequencing (Goh 2012). Despite these downfalls, whole exome sequencing is still widely used among researchers due to its cost effectiveness, robustness, and timeliness.

When compared to whole genome sequencing, WES costs five times less, has a higher coverage depth of 100x versus 30x, and requires less processing time since only a sixth of the genome is being analyzed (Goh 2012). Additionally, most allelic variants that have been found to be causal of Mendelian disorders disrupt protein-coding sequences, which can easily be found through exome sequencing (Ng 2010). Variants in the exome also have a larger effect size than those within non-coding sequences even in highly conserved regions (Ng 2010). Given all these reasons our lab has chosen to perform whole exome sequencing on the mutant mice.

1.4 EVOLUTIONARY CONSERVATION

Evolutionary conservation is important in providing support for genes that may contribute to disease. Conservation means that the amino acid at a certain position is the same throughout a set of species. Mutations in regions that are highly conserved are more likely to cause disease than those that are less conserved or not conserved at all. It has been described as a basic criterion for examining the disease causing potential of an amino acid or nucleotide sequence change (Pierpont 2007). Research done by Pierpont et al. describes three criteria used to establish the disease causing potential of a mutation. These criteria include: a sequence change that alters the gene coding sense, gene splice site, or regulatory region of the protein; segregates with disease in a kindred; and is not found in unrelated, unaffected control chromosomes. Some mutations may also be considered conserved if the changed amino acid residue has similar properties of the original residue. Comparisons of the sequences of the same protein in different species help to pinpoint those residues that are structurally or functionally vital as these will not change (Bolsover 2011).

Mutations may be tolerated, as long as the change does not destroy or seriously alter the protein's function. Mutations are damaging, or possibly damaging, when changes in side chains that are involved in binding substrates or cofactors, that interact with other proteins, or that participate in catalytic mechanisms alter or destroy the protein's function (Betts 2003).

1.5 SPECIFIC AIMS

The objective of this study was to examine the disease-causing genes recovered from our ENU forward genetic screen looking at evolutionary conservation and specific protein domains. The objectives were accomplished by fulfilling the following aims:

Aim 1:

Determine if the disease-causing mutations recovered in the ENU screen are evolutionarily conserved throughout a set of species and the damaging status of the mutation.

Aim 2:

Determine the protein domain, if any, in which the mutation is located and explain how a mutation in this domain may cause the observed phenotype.

Aim 3:

Compare recovered disease-causing ENU mutations with published human mutations to determine if any of the mouse mutations are the same as those in humans or located in the same protein domain and associated with a similar CHD phenotype.

2.0 MATERIALS AND METHODS

2.1 ENU MOUSE BREEDING

G0 males are injected with the alkylating agent N-ethyl-N-nitrosourea (ENU). These males are crossed with a wild type C57BL6 female to generate G2 progeny. ENU mutagenized G0 males are backcrossed to their G2 daughters to generate G3 progeny. Pregnant G2 mothers are checked for G3 fetuses with congenital heart defects by ultrasound. Heart defects among other phenotypes should be found in G3 mice with homozygous mutations caused by the ENU mutagenesis. See *Figure 1* for the full breeding scheme.

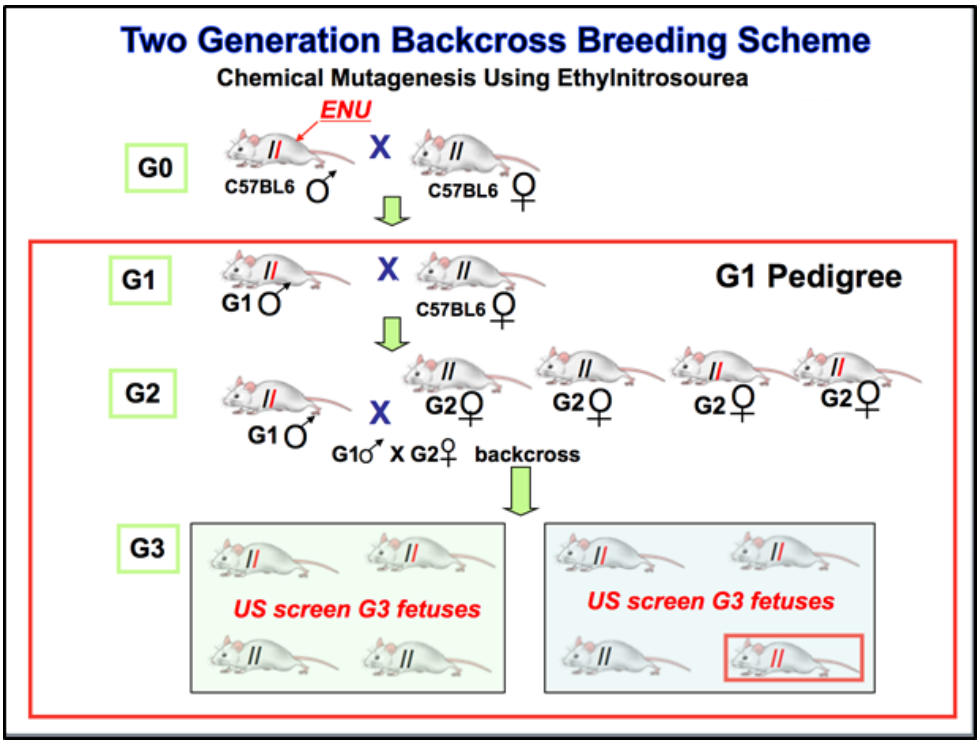

Figure 1. ENU breeding scheme.
Please note that this figure was adapted from a figure presented by Dr. You Li in the Lo lab.

2.2 MUTATION RECOVERY BY EXOME SEQUENCING

Mice that are phenotypically abnormal based on ultrasound screening, necropsy, and EFIC are sent out for full exome sequencing. Once the sequencing reads are retrieved by You Li, they must be filtered and sorted. Reads must be mapped to the genome and variants are detected by their deviation from the mouse genome. Reads that have low coverage across the exome are generally filtered out of the data due to the lack of accuracy. Next, the reads are annotated with their chromosomal location and nucleotide or amino acid change. Annotated mutations are then sorted based on the type: nonsynonymous (i.e. missense), frameshift (i.e. nonsense), and splicing. These reads are filtered against the dbSNP129 database and the Lo lab in house

database. We may then determine candidate disease causing genes based on their zygosity and are confirmed through next generation sequencing (NGS). Please refer to Figure 2 for the whole exome sequencing pipeline, which describes the technique in full.

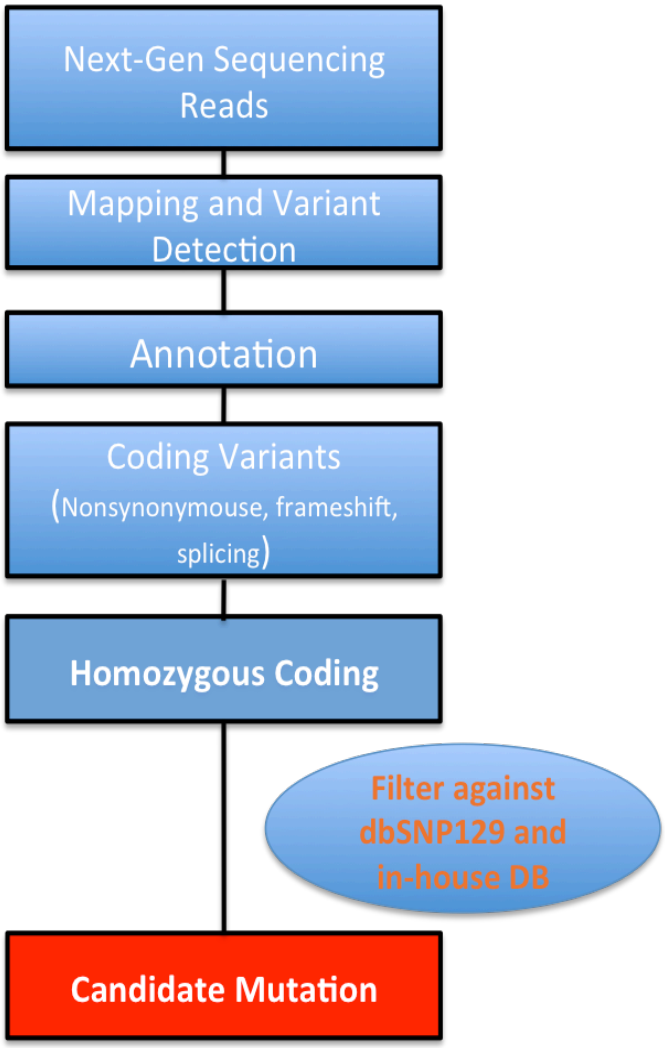

Figure 2. Whole exome sequencing pipeline
Please note that this figure was adapted from a figure presented by Dr. You Li in the Lo lab.

2.3 DISEASE-CAUSING GENE VALIDATION

Genes were validation for disease-causing status by genotyping of the mutant pup and its littermates. Please note that Nikolai Klena, a lab tech in the Lo lab, validated disease-causing genes. The recovered disease-causing genes are homozygous in the exome sequencing and this must be validated through genotyping. Once the mutant pup has been genotyped for the specific gene and homozygosity has been established we must genotype the littermates. The purpose of genotyping the littermates are to clarify that this mutation is actually causing the observed phenotype. If the mutation is present as homozygous in any of the littermates that do not have the observed mutant phenotype, then this mutation must not be causing the phenotype.

2.4 PROTEIN DOMAINS

Protein domains for each gene are determined using SMART, a Simple Modular Architecture Research Tool. This is a biological database that allows researchers to view the domain architecture of evolutionarily conserved protein domains. By inserting the entire amino acid sequence of the gene of interest, SMART comes up with a set of domains that are within that gene as well as the number and position of amino acids that are within that domain. The protein domain for the variants we recovered is found by looking at the location of the amino acid change in the mutation and comparing it with the location of the SMART domain for that gene.

2.5 CONSERVATION CALLS

There are four different conservation calls that could be made: damaging, possibly damaging, tolerated, and possibly tolerated. Conservation calls were made based on four criteria. A variant is considered damaging and thus highly conserved if the amino acid is the same across all species. If the amino acid is different among any of the species, we examine what the wild type amino acid for that particular species is compared to our recovered variant. If the wild type amino acid in a species is the same as the amino acid change in our variant, then it is not considered conserved and the call would be tolerated. If the wild type amino acid is different than the amino acid change in our variant, then we move on to the next rule. Next, we look at the type of amino acid change in our variant. If the amino acid change affects certain residue properties, then the conservation call would be possibly damaging. If the amino acid change does not affect these properties, then the conservation call would be possibly tolerated. Residue properties include: charge, size, structure, hydrophobicity, and polarity.

Splicing mutation calls are a bit different than the conservation calls for missense and nonsense mutations. Any splicing mutations that are within two base pairs of the exon are considered damaging, as these sites are canonical splice sites and the nucleotides at these positions are (generally) the same in every species. Any splicing mutation that is more than two base pairs from the exon does not have a conservation call. However, we still need to look at these locations to determine if the nucleotide change is present as wild type in any of the other species.

3.0 RESULTS

3.1 MUTATION RECOVERY

We have been able to recover over 9,200 mutation variants from nearly 150 mutant mouse lines through our ENU mutagenesis screen. Variants were recovered using whole mouse exome sequencing analysis to find coding mutations. This strategy has proven highly effective in providing a non-gene biased view of the genetic etiology of congenital heart disease. Out of the 9,235 mutations we recovered, only 1,338 (15%) of them were homozygous. We focused the rest of our research on analyzing the homozygous mutations.

Out of the 1,338 homozygous variants we recovered 96 (7.2%) disease-causing variants from sixty-two genes. Of these variants, 59 were missense, 16 were splicing, and 13 were nonsense. There were 8 variants that are the same; either occurring in the same mouse line but different mice or in different mouse lines. This gives us a final count of 88 disease-causing variants in 62 genes. Genes that had missense repeat variants include Adamts6, Armc4, Bicc1, and Smarca4. Genes that had splicing repeat variants are Cyp2c44 and Kif15. There was a nonsense mutation repeat in the Dnah11 gene. Please refer to *Table 1* to see all 88 variants that were recovered from the screen.

Table 1. Disease-causing mutations recovered from ENU mouse screen.

Gene	Chr	Exon	Nucleotide change	Amino acid change	Strand	Domain	Phenotype	Conservation Calls
2410089 E03Rik	15	Exon 8	T757C	S253P	forward	*unknown*	Joubert Syndrome	damaging
6030429 G01Rik	7	Exon 7	G710A	G237D	reverse	-	Laterality	Possibly damaging
Acan*	7	Exon 6	T1032A	Y344X	forward	LINK	Biventricular hypertrophy/ cleft palate/ crachyposidm	damaging
Adamts6*	13	Exon 3	C447G	S149R	forward	Pep_M12B_ propep	OFT/AVSD septal defects/ hypertrophy/ hydronephrosis	damaging
Anks6	4	Exon 2	T560A	M187K	reverse	ANK	OFT/Laterality	damaging
Ap1b1*	11	Exon 9	T1094C	V365A	forward	Adaptin_N	Laterality	damaging
Armc4	18	Exon 20	T2978A	M993K	reverse	ARM	Laterality	Possibly damaging
Bicc1	10	-	606+2T>C	-	reverse	intronic	Laterality	damaging
Bicc1	10	Exon 6	T586C	S196P	reverse	KH	Laterality	Possibly damaging
Cc2d2a	5	Exon 23	C2845T	R949X	forward	*unknown*	Mild laterality/ cystic kidney and lung	damaging
Ccdc151	9	Exon 7	828+2T>C	-	reverse	intronic	Laterality	damaging
Ccdc39	3	Exon 4	G445T	E149X	reverse	*unknown*	AVSD/Laterality	damaging
Ccdc39	3	Exon 4	357+2T>A	-	reverse	intronic/ CCD	Laterality	damaging
Ccdc39	3	Exon 1	T2A	M1K	reverse	*unknown*	SIT	damaging
Cep110*	2	Exon 13	T1886C	I629T	forward	*unknown*	OFT/cranio	Possibly damaging
Cep290	10	Exon 34	T4670A	L1557X	forward	CCD	Mild laterality/ cystic kidney	damaging
Cep290	10	Exon 12	1189+ 2T>C	-	forward	intronic/ CCD	Laterality	damaging
Cfc1	1	Exon 2	T68A	V23E	forward	*unknown*	Laterality	Possibly damaging
Cml5*	6	Exon 3	A289T	I97F	reverse	FR47/Acetyl transf_1	Laterality	damaging

Table 1 Continued

Cxcr4*	1	Exon 2	G295A	D99N	reverse	7tm_1	Left heart obstructive defects	Possibly damaging
D630037 F22Rik*	10	Exon 10	1071+ 2T>A	-	reverse	intronic	Laterality/OFT/ polydactyly	damaging
Dctn5*	7	Exon 4	348+6T>C	-	forward	intronic	DORV/VSD/ overriding aorta	N/A
Dnah11	12	Exon 16	C3184T	Q1062X	reverse	Spc7	Laterality	damaging
Dnah11	12	Exon 63	C10243T	Q3415X	reverse	MYSc	PCD/Laterality	damaging
Dnah11	12	Exon 60	T9837A	Y3279X	reverse	MYSc	Laterality/OFT	damaging
Dnah11	12	Exon 7	1137+T>C	-	reverse	intronic/ DHC_N1	Laterality	damaging
Dnah11	12	Exon 40	6489+ 2T>C	-	reverse	intronic/ AAA_5	Laterality	damaging
Dnah11	12	Exon 64	A10369T	I3457F	reverse	MYSc	PCD/Laterality	damaging
Dnah11	12	Exon 33	A5632G	T1878A	reverse	AAA	Laterality	damaging
Dnah11	12	Exon 41	T6641G	L2214R	reverse	AAA_5	Laterality	Possibly damaging
Dnah5	15	Exon 4	438+2T>A	-	forward	intronic	PCD/Laterality	damaging
Dnah5	15	Exon 44	7398+ 2T>A	-	forward	intronic	Laterality	damaging
Dnah5	15	Exon 76	T13169C	L4390P	forward	dynein heavy	PCD/Laterality	damaging
Dnah5	15	Exon 59	T10048C	S3350P	forward	DHC_N1	DORV/AVSD/ Laterality	damaging
Dnah5	15	Exon 34	T5503C	W1835R	forward	BRLZ/FYVE	PCD/Laterality	damaging
Dnah5	15	Exon 48	T7971G	N2657K	forward	AAA	PCD/Laterality	damaging
Dnah5	15	Exon 76	T13210A	S4404T	forward	dynein heavy	PCD/Laterality	damaging
Dnah5	15	Exon 77	13329= 10T>A	-	forward	intronic/ dynein heavy	PCD/Laterality	N/A
Dnah5	15	Exon 19	C2755A	H919N	forward	MBT/L27	PCD/Laterality	tolerated
Dnaic1	4	-	240+1G>A	-	forward	intronic	AVSD/DORV/ Laterality/ dextrocardia	damaging
Dnaic1	4	Exon 16	T1565C	I522T	forward	*unknown*	Laterality	damaging

Table 1 Continued

Dync2h1	9	Exon 5	T701A	V234E	reverse	DHC_N1	Mild Laterality	Possibly damaging
Dyx1c1	9	Exon 1	T2A	M1K	forward	*unknown*	Laterality	Possibly damaging
Foxj1*	11	Exon 2	T425A	I142N	reverse	FH	Laterality	damaging
Frem2	3	Exon 11	A6875G	Y2292C	reverse	Calx-beta	Kidney agenesis/ syn-polydactyly	damaging
Fuz	7	Exon 4	387+2T>A	-	forward	intronic	Pulmonary atresia/ polydactyly	damaging
Gm1060*	5	Exon 15	A2021G	Y674C	forward	*unknown*	Laterality	damaging
Gm1060*	5	Exon 15	T2033A	L678Q	forward	*unknown*	SIT/cranio/ hypertrophy	damaging
Gm572*	4	Intron 1	26+1G>A	-	forward	intronic	Laterality	damaging
Hace1*	10	Exon 19	A2060T	Y687F	forward	HECTc	RAA/DORV/ VSD	damaging
Hectd1*	12	Exon 22	T3264A	Y1088X	reverse	*unknown*	VSD/aortic atresia/ hypoplastic aortic arch	damaging
Hr*	14	Exon 13	G2732A	W911X	forward	*unknown*	SIT	damaging
Ift140	17	Exon 9	A1138G	N380D	forward	*unknown*	Laterality	damaging
Ift74*	4	Exon 19	T1645C	W549R	forward	*unknown*	Laterality	damaging
Kif7	7	Exon 3	T557A	V186E	reverse	KISc	OFT/VSD	damaging
Lox*	18	Exon 3	G854T	C285F	reverse	Lysyl_ oxidase	Diaphragmatic hernia	damaging
Lrp1	10	Exon 82	T12694C	C4232R	reverse	EGF	Left heart obstructive defects/ valvular anomalies	damaging
Lrp2	2	Exon 46	8456-3A>G	-	reverse	intronic/ LDLa	OFT/omphaloceol	N/A
Ltbp1*	17	Exon 7	T1134A	C378X	forward	*unknown*	OFT/cranio/ cystic kidney	damaging
Megf8	7	Exon 21	A3641T	N1214I	forward	EGF	Laterality	damaging
Megf8	7	Exon 3	A442G	N148D	forward	EGF_ Lam	SIT/AVSD/ DORV/ dextrocardia	Possibly damaging
Mical1*	10	Exon 22	T2786C	M929T	forward	DUF 3585	Laterality	Possibly damaging

Table 1 Continued

Mmp21*	7	Exon 2	G530T	W177L	reverse	ZnMc	Laterality	Possibly tolerated
Nek8	11	Exon 3	T371C	I124T	reverse	S_TKc	Laterality	damaging
Pcsk5*	19	Exon 4	A521G	D174G	reverse	Peptidase_S8	Laterality/OFT/ cystic kidney	damaging
Pcsk5*	19	Exon 35	T4795A	C1599S	reverse	FU	DORV/VSD/ aortic stenosis	damaging
Pde2a*	7	Exon 23	T2045G	M682R	forward	HDc	OFT/ valvular anomalies	damaging
Pkd1	17	Exon 38	T11084A	I3695N	forward	*unknown*	PKD	damaging
Plxnd1*	6	Exon 27	A4727G	D1576G	reverse	Plexin_ cytopl	AVSD/ASD/VSD/ RAA/PTA	Possibly damaging
Prickle1	15	Exon 5	G482T	C161F	reverse	LIM	Velocardiofacial syndrome	damaging
Pskh1*	8	Exon 2	T656C	L219P	forward	S_TKc	Laterality	damaging
Psme4*	11	Exon 28	A3247G	I1083V	forward	*unknown*	Laterality	damaging
Robo1*	16	Exon 6	T809C	I270T	forward	IGc2	Bilateral duplex cystic kidney/ anophthalmia	damaging
Sema6a*	18	Exon 10	T823C	S275P	reverse	Sema	Laterality	damaging
Smad6*	9	Exon 4	A982T	K328X	reverse	*unknown*	OFT	damaging
Smarca4	9	Exon 16	C2381T	T794I	forward	DEXDc & SNF2_N	OFT	damaging
Smarca4	9	Exon 8	C1249T	R417C	forward	Low complexity	OFT	damaging
Snx17*	5	Exon 5	A431G	E144G	forward	B41	OFT	Possibly damaging
Spata5*	3	Exon 9	C1493T	T498I	forward	AAA	Laterality	damaging
Sufu*	19	Exon 4	T530A	M177K	forward	SUFU	Polydactyly/cleft lip/domed head	damaging
Tab1*	15	Exon 9	T935A	M312K	forward	PP2Cc	DORV	damaging
Tmem67	4	Exon 11	T1120A	Y374N	reverse	Meckelin	Laterality/ cystic kidney	Possibly damaging
Tmem67	4	Exon 23	2322+6T>C	-	reverse	intronic/ Meckelin	Laterality	N/A
Tpbpa*	13	Exon 3	T162A	S54R	reverse	-	AVSD/DORV/ IAA/ noncompaction	Possibly damaging

Table 1 Continued

Wdr69*	1	Exon 2	T57A	Y19X	rorward	*unknown*	Laterality/kidney tubule cysts	damaging
Wdr69*	1	Exon 3	T197C	L66P	forward	*unknown*	Laterality	Possibly damaging
Xpnpep1*	19	Exon 15	T1246C	S416P	reverse	Peptidase_M 24	Unbalanced AVSD	damaging
Zfp161*	17	Exon 3	C419A	S140X	forward	*unknown*	DORV/ septal defects	damaging

Please see Appendix A for full domain information and phenotype acronyms.
Pink highlighted genes are known to contribute to CHD.
*novel CHD genes not associated with any human disease.

3.2 CONSERVATION CALLS FOR MISSENSE AND NONSENSE MUTATIONS

The tables in this section encompass the conservation for the region surrounding the mutation. Based on the conservation at the position of the mutation in addition to the conservation of the surrounding region, we can identify whether the mutation is damaging and how this mutation may affect the reading frame, particularly in nonsense mutations. DNA is read from the 5’ to 3’ end of the strand. The first line of each table shows which way the amino acid sequence is read. A sequence that is read 5’ to 3’ is on the forward strand and a sequence that is read 3’ to 5’ is on the reverse strand. The yellow highlighted region of each of the tables is where the mutation occurs. Any amino acids highlighted in red are deviations from the wild type amino acid in mice. Nonsense mutations are always considered damaging because they introduce a stop codon, which may truncate the coded protein and cause the observed phenotype. It is important to look at the location of the nonsense mutation to see where the stop codon is introduced in the exon. The following figure *(Figure 3*) shows the distribution of conservation across all 88 mutations.

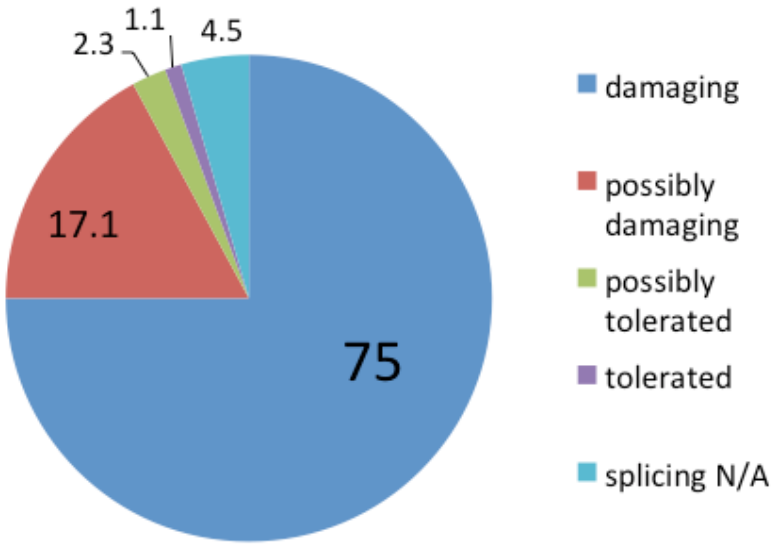

Figure 3. Pie graph of the conservation calls.

3.2.1 2410089E03Rik missense mutation

The 2410089E03Rik missense mutation is located in exon 8 at position 253 (S253P) of this gene, which changes a serine into a proline. Based on the conservation in Table 2, this missense mutation is highly conserved across all species and therefore is considered a damaging mutation.

Table 2. 2410089E03Rik missense mutation in exon 8 is damaging.

	5′----------→3′
Mouse	ECPLCSLIPRCASVKSRGALISAFSRDGLAL
Rat	ECRLCNLIPKCASVKSRGALISAFSRDGLAL
Human	DCHLCSLIPKCESVKSRGALISAFSRDGLTL
Orangutan	DCHLCSLIPKCESVKSRGALI==========
Dog	NCHLCNLIPKCASVKSRGALISAFSRDGVTL
Horse	DCLLSSLIPKCESVKSRGALISAFSRDGLTL
Opossum	ECFLCTLVPKCESIKSRGALIIAFSRDGLTL
Chicken	TCSLEKLVLTCVSIKSRGALL==========
Stickleback	TYPMSRLTPPCQPVKSRGALLPAFSPDGCLL

3.2.2 6030429G01Rik missense mutation

The missense mutation in the 6030429G01Rik gene is located in exon 7 at position 237 (G237D). The mutation is conserved across all species except for the opossum. The wild type amino acid for mouse is a glycine. The mutation recovered in this gene changes the small, hydrophobic, nonpolar glycine into a negatively charged, acidic, polar aspartic acid. This changes several properties of the original residue and is thus possibly damaging.

Table 3. 6030429G01Rik missense mutation in exon 7 is possibly damaging.

	3′←----------5′
Mouse	NPVHYASLDRLEFAVGTDRWRRFEQIHIVQA
Rat	NPVHYASLDRLEFAVGTDRWRRFEHFHIVQA
Human	NPVHYASSDRLEFAVGTDRWRRFEQPHIVQA
Orangutan	NPVHYASSDRLEFAVGTDRWRRFEQPHIVQA
Dog	NPVHYASSDRLEFAVGTDRWRRFERTHIVRA
Horse	NPVHYASSDRLEFAVGTDRWRRFERTHIVRA
Opossum	NPVHYASADRIEFAVSTERWRRFERSNIVRA
Chicken	
Stickleback	==========

3.2.3 Acan nonsense mutation

The mutation in the Acan gene is considered damaging since it is a stopgain mutation which inserts an early stop codon at the end of exon 6 at position 344 (Y344X).

Table 4. Acan damaging nonsense mutation is located at the end of exon 6 in a fairly conserved region.

	5′----------→3′
Mouse	VYLHANQTGYPDPSSRYDAICYTG…intron
Rat	VYLHANQTGYPDPSSRYDAICYTG…intron
Human	VYVHANQTGYPDPSSRYDAICYTG…intron
Orangutan	VYVHANQTGYPDPSSRYDAICYTG…intron
Dog	VYLHANQTGYPDPSSRYDAICYTG…intron
Horse	VYLHANQTGYPDPSSRYDAICYTG…intron
Opossum	VYLHANQTGYPDPLSRYDAICYTG…intron
Chicken	VYLPANQTGYPHPSSRYDAICYSG…intron
Stickleback	VYLPPNQTGYPNPDSRYEAVCFGG…intron

3.2.4 Adamts6 missense mutation

The Adamts6 mutation is present in three different mouse lines. The wild type amino acid in mice is a serine, which is conserved throughout the entire set of species. The mutation occurs at the end of exon 3 at position 149 (S149R) and is considered a damaging mutation.

Table 5. Adamts6 missense mutation in exon 3 is damaging.

	5′----------→3′
Mouse	TGYLQDQHSTTKVALSNCVGL…intron
Rat	TGYLQDQHSTTKVALSNCVGL…intron
Human	TGYLQDQRSTTKVALSNCVGL…intron
Orangutan	TGYLQDQRSTTKVALSNCVGL…intron
Dog	TGYLQDQRSTTKVALSNCIGL…intron
Horse	TGYLQDQRSTTKVALSNCVGL…intron
Opossum	TGYLQDQHSTTKVAISNCIGL…intron
Chicken	TGYLQDQHTTTKVAVSNCNGL…intron
Stickleback	VGSLQNQRGATRVALSNCKGL…intron

3.2.5 Anks6 missense mutation

A missense mutation was recovered in the Anks6 gene that is located on chromosome 4. The mutation is in exon 2 of the gene at position 187 (M187K). The wild type amino acid is a methionine, which is conserved throughout all species; making this a damaging mutation.

Table 6. Anks6 missense mutation in exon 2 is damaging.

	3′←----------5′
Mouse	MLLRVVAEHGHQVAAMLATIGLLEDGSGGTA
Rat	MLLRVVAEHGHQVAAMLATIGLLEDRSGGSA
Human	MLLRVVAEHGHQIAAMLATIDLPEDRSGGLG
Orangutan	MLLRVVAEHGHQIAAMLATIDLPEDRSGGLG
Dog	MLLRVVAEHGHQIAAMLATIDLLEDRSDGAG
Horse	MLLRVVAEHGHQIAAMLATIDLREDRSDGAG
Opossum	VLLHVVAEHGHQAAIMLATIDPLDEKNGD-G
Chicken	LLLHVVAEHGHQAAAMLPTIDPLDDRSSNVN
Stickleback	LLLRVAAEHGHQSAVMLATIDMF—-RGG===

3.2.6 Ap1b1 missense mutation

The Ap1b1 mutation is conserved throughout the species and is also in a highly conserved region; therefore it is a damaging mutation. It is located at the beginning of exon 9 at position 365 (V365A).

Table 7. Ap1b1 missense mutation in exon 9 is damaging and in a highly conserved region.

	5′----------===----------→3′
Mouse	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Rat	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Human	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Orangutan	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Dog	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Horse	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Opossum	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Chicken	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC
Stickleback	Intron…VLAELKEYATEVDVDFVRKAVRAIGRC

3.2.7 Armc4 missense mutation

We recovered one missense mutation in the Armc4 gene in two different mouse lines. This mutation occurs in the eighth amino acid from the end of exon 20 at position 993 (M993K). The wild type amino acid, methionine, is conserved throughout all species except the dog. When we look at the amino acid change in our mutation the nonpolar, hydrophobic methionine is converted into a positive, polar lysine. Since there is a change in the charge and polarity, this mutation is possibly damaging.

Table 8. Armc4 missense mutation in exon 20 is possibly damaging.

	3′←-------------------------5′
Mouse	Intron…KVAGNEHMTICNDADESLQYLAQ
Rat	Intron…KVAGNEHMTICNDADESLQYLAQ
Human	Intron…KVAGNEHMTICNDADESLQYLAQ
Orangutan	Intron…KVASNEHMTICNDADESLQYLAQ
Dog	Intron…KVAGNEHITICNDADESLQYLAQ
Horse	Intron…KVAGNEHMTICNDADESLQYLAQ
Opossum	Intron…KVAGNEHMTICNNADESLQYLAQ
Chicken	Intron…KVVGNEHMTICNSPDESLQYLAQ
Stickleback	Intron…KVVGRDHMTICNNPD========

3.2.8 Bicc1 missense mutation

The Bicc1 mutation is located at the twelfth amino acid of exon 6 at position 196 (S196P). The wild type amino acid is a serine, which is conserved in all species except the stickleback. The mutation is, however, located in a fairly conserved region considering it is towards the end of the exon. Our recovered mutation changes from a polar, uncharged, hydrophilic serine to a nonpolar, uncharged proline. Since there is a change in the polarity, this mutation is possibly damaging.

Table 9. Bicc1 missense mutation in exon 6 is possibly damaging.

	3′←----------5′
Mouse	Intron…RIRARASEVGAPQGAISV…intron
Rat	Intron…RIRARASEVGAPQGAISV…intron
Human	Intron…RIRVRASEVGAPQGAISV…intron
Orangutan	Intron…RIRVRASEVGAPQGAISV…intron
Dog	Intron…RIRVRASEVGAPQGAISV…intron
Horse	Intron…RIRVRASEVGAPQGAISV…intron
Opossum	Intron…RIRVRASEVGAPQGAISV…intron
Chicken	Intron…RIRVRASEVGAPQGAISV…intron
Stickleback	Intron…RIKVRAAEVGGPQGAISV…intron

3.2.9 Cc2d2a nonsense mutation

The Cc2d2a nonsense mutation is located 5 amino acids into exon 23 at position 949 (R949X). This mutation is automatically considered damaging because it inserts a stop codon fairly early on in the exon. If you also look at the conservation for this mutation, it is conserved across all species.

Table 10. Cc2d2a nonsense mutation in exon 23 is damaging and in a fairly conserved region.

	5′----------→3′
Mouse	Intron…DYEKRLRDRNVIETKDHLD
Rat	Intron…DYEKRLRDRNVIETKDHLD
Human	Intron…DYEKRLRDRNVIETKEHID
Orangutan	Intron…DYEKRLRDRNVIETKEHID
Dog	Intron…DYEKRLQDRNVIETKDHID
Horse	Intron…DYEKRLRDRNVIETKDHID
Opossum	Intron…DYEKRIRDRGIIETKDHLD
Chicken	Intron…EYEKRLHERDVIDSKHYLD
Stickleback	Intron…DYEKRLKEGETVDTEEHLD

3.2.10 Ccdc39 missense and nonsense mutations

The missense mutation in the Ccdc39 gene is located at the first amino acid in exon 1 of this gene. The wild type amino acid is a methionine, which is conserved throughout all species; therefore the mutation is damaging.

Table 11. Ccdc39 missense mutation in exon 1 is damaging.

	3′←----------5′
Mouse	AFGDEWHLESLFESCM…intron
Rat	AFGDEWHLESLFESSM…intron
Human	AFGDEWHLEALFESSM…intron
Orangutan	AFGDEWHLEALFESSM…intron
Dog	AFGDDWHLEELLESSM…intron
Horse	AFGDEWHLEALFESSM…intron
Opossum	AFGNEWHLESLFDQSM…intron
Chicken	AFGDDWQLEALV---…intron
Stickleback	==========…intron

The nonsense mutation in Ccdc39 is located in exon 4 at position 149 (E149X). It introduces a stop codon in exon 4 and is considered damaging. This

Table 12. Ccdc39 nonsense mutation in exon 4 is conserved in all species except stickleback and is located in a highly conserved region.

	3′←----------5′
Mouse	SYKQLTLSDSDKHASEELWAELAQQDWNMQC
Rat	SYKQLTLSDSDKHASEELWAELAQQDWNMQC
Human	AYKQLTLADSDKHASEELWAELAQQDWNMQC
Orangutan	AYKQLTLADSDKHASEELWAELAQQDWNMQC
Dog	AYKQLTLADSDKHASEELWAELAQQDWNIQC
Horse	AYKQLTLADSDKHASEELWAELAQQDWNMQR
Opossum	TYKQLAFADSDKHASEELWAELAQQDWNMQC
Chicken	AYKQITIIDNDKRESEKIWSELVEEDCNMQQ
Stickleback	AYKIIAMTDEDKCNSDQLFADMTQQDWNMQN

3.2.11 Cep110 missense mutation

Cep110 has one missense mutation located in exon 13 at position 629 (I629T). The wild type amino acid is an isoleucine in mouse, rat, human, and orangutan. The wild type amino acid is a

valine in dog, horse, opossum, and chicken. The recovered mutation converts the aliphatic, nonpolar isoleucine to a polar, uncharged, hydrophilic threonine. Since there is a change in polarity, this mutation is possibly damaging.

Table 13. Cep110 missense mutation in exon 13 is possibly damaging.

	5′----------→3′
Mouse	ALKKDLEGVSGLQEYLGTIKGQATQAQNECRKLQ
Rat	ALKKDLEGVSGLQEYLGTIKGQAAQAQNECRKLQ
Human	ALKKDLEGVSGLQEYLGTIKGQATQAQNECRKLR
Orangutan	ALKKDLEGVSGLQEYLGTIKGQATQAQNECRKLQ
Dog	ALKKDLESVSGLQEYLGTVKGQAPQAQNECRKLQ
Horse	ALKKKLEGVSGLQEYLGTVKGQAAQAQNECRKLQ
Opossum	TLKKNLESISGLQDYLENVKGQVTKVNDECRVLQ
Chicken	ALKRDLESIIGLQEYLESVKHQAKQAQDECKELQ
Stickleback	

3.2.12 Cep290 nonsense mutation

The Cep290 nonsense mutation introduces a stop codon at the end of exon 34 at position 1557 (L1557X). The mutation is not completely conserved across all species; there is a different wild type amino acid in the stickleback. It is still considered damaging because it introduces an early stop codon, which will truncate the protein that is coded for by this gene.

Table 14. Cep290 nonsense mutation is located at the end of exon 34 in a fairly conserved region and is conserved throughout all species except the stickleback.

	5′----------→3′
Mouse	TIANMQARLNHKEEVLKKYQHLLEKARE…intron
Rat	TIANMQARLNHKEEVLKKYQHLLEKARE…intron
Human	TIANMQARLNQKEEVLKKYQRLLEKARE…intron
Orangutan	TIANMQARLNQKEEVLKKYQRLLEKARE…intron
Dog	TIANMQARLNQKEEVLKKYQHLLEKARE…intron
Horse	TIANMQARLNHKEEVLKKYQRLLEKARE…intron
Opossum	TIANMQARLDQKEEVLKKYQHLLEKARE…intron
Chicken	TIANMQARLSQKEEMLKKYQDLLAKARE…intron
Stickleback	TIRDLQARLNKKEDVIKKCHNQLAGARQ…intron

3.2.13 Cfc1 missense mutation

The Cfc1 missense mutation is not conserved across the species. It occurs at the end of exon 2 at position 23 (V23E). This gene is only present in three species: mouse, orangutan, and horse. There is a different wild type amino acid than the valine in the horse. If we look at our recovered mutation, the nonpolar valine is converted into a polar, negatively charged aspartic acid. Since there is a change in charge and polarity, this mutation is considered possibly damaging.

Table 15. Cfc1 missense mutation in exon 2 is possibly damaging.

	5′----------→3′
Mouse	Intron…RPLFLVTVALQLIGLGYS…intron
Rat	Intron…==========…intron
Human	
Orangutan	Intron…RLLFMVSLALQIISLGNS…intron
Dog	
Horse	Intron…RLLFMITLALQVIHLGNS…intron
Opossum	
Chicken	
Stickleback	

3.2.14 Cml5 missense mutation

Cml5 has a missense mutation in exon 3 at position 97 (I97F). This gene is only present in four out of the nine species in the set: mouse, rat, opossum and chicken. The recovered mutation is highly conserved across these four species and is therefore considered damaging.

Table 16. Cml5 missense mutation in exon 3 is damaging.

	3′←----------5′
Mouse	EAVWFSGHANLYSRTIDAMDTQLCQAVYDKF
Rat	EAVWFSGHANLYSKTIDAMDTQLCMAVYEKF
Human	
Orangutan	
Dog	
Horse	
Opossum	EAVWFSGKDSLYSKRIDRMDTHLAHDVYDSY
Chicken	EVVWFSDPARMYFSDIDRLD--SCREFST--
Stickleback	

3.2.15 Cxcr4 missense mutation

The Cxcr4 gene has a missense mutation in exon 2 at position 99 (D99N). It is conserved in all species except for the dog. In our recovered mutation, the polar, negatively charged aspartic acid is mutated into a polar, uncharged asparagine. This mutation is considered possibly damaging because there is a change in charge between the wild type amino acid and the mutated amino acid.

Table 17. Cxcr4 missense mutation in exon 2 is possibly damaging.

	3′←----------5′
Mouse	VAKCLFKGFYWDAMADVAWFPLTIVFLLDAV
Rat	VAKCLFKGFYWDAMADVAWFPLTIVFLLDAV
Human	VAKCLFNGFYWNAVADVAWFPLTIVFLLDAV
Orangutan	VAKCLFNGFYWNAVADVAWFPLTIVFLLDAV
Dog	VAKCLFNGFYWNAVAEVAWFPLTLVFLLDAV
Horse	VAKCLFKGFYWNAVADVAWFPLTLVFLLDAV
Opossum	VAKCLFNGFYWNAAADVAWFPLTLVFLLDAV
Chicken	VAKCLVNGFYWSIAADVSWFPLTIVFLLDAV
Stickleback	VSVCLFSGFYWSKAADVAWFPLTLVFLLDAV

3.2.16 Dnah11 has 3 missense and 3 nonsense mutations

The first missense mutation in Dnah11 occurs in exon 33 at position 1878 (T1878A). The mutation is conserved across all species and is located in a highly conserved region so it is a damaging mutation.

Table 18. Dnah11 missense mutation in exon 33 is damaging and located in a highly conserved region.

	3′←----------5′
Mouse	MGLARGLDKTTETKGTGAPGAPAGSMTLHLS
Rat	MGLARGLDKTTETKGTGAPGAPAGSMTLHLS
Human	MGLARGLDKTTETKGTGAPGAPAGSMTLHLS
Orangutan	MGLARGLDKTTETKGTGAPGAPAGSMTLHLS
Dog	MGLARGLDKTTETKGTGAPGAPAGSMTLHLS
Horse	MGLARGLDKTTETKGTGAPGAPAGSMTLHLS
Opossum	MGLAHGLDKTTETKGTGAPGAPAGSMTLHLS
Chicken	IGLARGLDKTTETKGTGAPGAPAGSMTLHLS
Stickleback	VGLARGLDKTTETKGTGAPGAPAGSMTLHLA

The next missense mutation in Dnah11 occurs at the seventh amino acid in exon 41 at position 2214 (L2214R). It is conserved across all species except for the stickleback. If we look at the recovered mutation, the uncharged leucine is changed into a positively charged arginine. Since there is a change in charge, this mutation is considered possibly damaging.

Table 19. Dnah11 missense mutation in exon 41 is possibly damaging.

	3′←----------**5′**
Mouse	VIWTPGDHTLNAQERLISSFLG…intron
Rat	VIWTPGDHTLNAQERLISSLLG…intron
Human	VIWKPGDHKLNAQERLISSFLG…intron
Orangutan	VIWKPGDHKLNAQERLISSFLG…intron
Dog	IIWKPGDHKLNAQERLISSFLG…intron
Horse	VIWKPGDHMLNAQERLVSSFLG…intron
Opossum	VIWKPGSHLLNAQERLLSSFLG…intron
Chicken	VLWKPGEHMAKTQERLLL----…intron
Stickleback	VIWKPGVHSIAAQERMLSSLLG…intron

The third missense mutation in Dnah11 is located in exon 64 at position 3457 (I3457F). The mutation is conserved across all species and is therefore a damaging mutation.

Table 20. Dnah11 missense mutation in exon 64 is damaging.

	3′←----------**5′**
Mouse	QQPDIMLPWRECHTLITANETSMRDSPLGEN
Rat	QQPDIMLPWRKCHTLITANETSMRDSPLGQN
Human	QQPDIVLPWRECHTLIAANETSMRDSPLGEN
Orangutan	QQPDIVLPWRECHTLIAANETSMRDSPLGEN
Dog	QQPDIMLPWRECHTLITANETSMRDNPLGEN
Horse	QQPDIMLPWRECHTLITANETSMRDNPLGEN
Opossum	QQPDIMLPWREAHTLITANETSMRDSPLGEN
Chicken	K*PDIMLPCHECHTLITAREASMQDGLLGRN
Stickleback	QQPDILLPWRESTTLISANETSMRDNPLGQN

The first nonsense mutation in Dnah11 is located at the fifth to last amino acid of exon 16 at position 1062 (Q1062X). The mutation is conserved across all species, except for stickleback. This mutation inserts a stop codon very early in the gene, which will cause the coded protein to be dysfunctional.

Table 21. Dnah11 nonsense mutation is located at the end of exon 16, which is not a highly conserved region.

	3′←----------**5′**
Mouse	Intron…QEKFQELTPPQLIDESAPAD
Rat	Intron…QEKFQELTPSQLIDEGAPAD
Human	Intron…QEKFQELTPPQPIEENAHAD
Orangutan	Intron…QEKFQELTPPQPIEENAHA-
Dog	Intron…QEKFQELTPPQPVEEYVHAD
Horse	Intron…QEKFQELTPPQPIEKNAHAD
Opossum	Intron…QEKFQEITPPQPLGENAQID
Chicken	Intron…QEKCQRTTPPWS--KDTFTD
Stickleback	Intron…QEKFNNITPPSLLADAGCAE

The next Dnah11 nonsense mutation is located in the second to last amino acid of exon 60 at position 3279 (Y3279X). It is conserved in all species except for the chicken. The

Table 22. Dnah11 nonsense mutation is located at the end of exon 60 in a fairly conserved region.

	3′←----------**5′**
Mouse	Intron…EYFRIINIVWACLGAAA
Rat	Intron…EYFKIINIVWACLGAAA
Human	Intron…EYFKIINIVWACLGAAA
Orangutan	Intron…EYFKIINIVWACLGAAA
Dog	Intron…QYFKVINIVWACLGAAA
Horse	Intron…EYFKIINIVWACLGAAA
Opossum	Intron…EYFKVINIVWACLGAAA
Chicken	Intron…EHFRVINVVWSCLGAAA
Stickleback	Intron…EYYRVINITWACLGAAA

The final nonsense mutation in Dnah11 is located in the second to last amino acid of exon 63 at position 3415 (Q3415X). This mutation is not highly conserved, as there are different wild

type amino acids in both the dog and the horse. However, since it is a nonsense mutation that introduces an early stop codon it is considered a damaging mutation.

Table 23. Dnah11 nonsense mutation is located at the end of exon 63 in a poorly conserved region.

	3′←----------5′
Mouse	Intron…KQQLFPIWKCDVLEQRY
Rat	Intron…KQQLFPIWKCDVLEQRY
Human	Intron…KQQLFPVWKCHVLEQRY
Orangutan	Intron…KQQLFPVWKCHVLEQRY
Dog	Intron…KHQLFPVWKCDVLEQRY
Horse	Intron…KRQLFPVWKCDVLEQRY
Opossum	Intron…KQKLFPIWKCQVLEERY
Chicken	Intron…KQKLFPIWFR*IQEQ--
Stickleback	Intron…KQRLFPIWENNLLQARY

3.2.17 Dnah5 has 6 missense mutations

Dnah5 has six different missense mutations. The first mutation is located at the sixth amino acid in exon 19 at position 919 (H919N). It is not conserved at all as the wild type amino acid in mice is only present in that species. Our mutation changes the wild type amino acid of a positively charged histidine to an uncharged asparagine. Asparagine is the wild type amino acid for the human and orangutan, therefore this mutation is considered tolerated because the mutated amino acid is present as the wild type amino acid for two species.

Table 24. Dnah5 missense mutation is located at the beginning of exon 19 and is tolerated.

	5′----------→3′
Mouse	Intron…GRREGHSEALASFNAGASSL
Rat	Intron…GGGEGCAEALASFNAGTSSL
Human	Intron…EREEGNFDTLTSINARANAL
Orangutan	Intron…E--EGNFDTLTSINARADAL
Dog	Intron…E=-EGKSDTLTPLRAPAGLL
Horse	Intron…E=-EGISDTLTSLNAGAGLL
Opossum	Intron…==========
Chicken	Intron…---ERRSKSLTSNNSWAGLS
Stickleback	Intron…---EGRLTTL-PCQCRPFL-

The next missense mutation is located at the tenth amino acid in exon 34 at position 1835 (W1835R). The amino acid is conserved across all species and is thus a damaging mutation.

Table 25. Dnah5 missense mutation is located at the beginning of exon 34 and is damaging.

	5′----------→3′
Mouse	Intron…VGLLGIQMLWTRDSEEALRNAKFDK
Rat	Intron…VGLLGIQMLWTRDSEEALQNAKFDK
Human	Intron…VGLLGIQMIWTRDSEEALRNAKFDK
Orangutan	Intron…VGLLGIQMIWTRDSEEALRNAKFDK
Dog	Intron…VGLLGIQMIWTRDSEEALRNAKYDK
Horse	Intron…VGLLGIQMIWTRDSEEALRNAKFDK
Opossum	Intron…VGLLGIQLIWTRDSEEALTNAKFDK
Chicken	Intron…VGLLGIQMIWTRDSEEALTNARYDK
Stickleback	Intron…VGLLGIQMIWTRDSEEALTNARYDR

The following Dnah5 mutation is located at the end of exon 48 at position 2657 (N2657K). It is conserved across all species and thus is a damaging mutation.

Table 26. Dnah5 missense mutation is located at the end of exon 48 and is damaging.

	5′----------→3′
Mouse	GPPAGKKMAVFIDDLNMPVINEWGDQ…intron
Rat	GPPAGKKMAVFIDDLNMPVINEWGDQ…intron
Human	GPPAGKKMTVFIDDVNMPIINEWGDQ…intron
Orangutan	GPPAGKKMTVFIDDVNMPIINEWGDQ…intron
Dog	GPPAGKKMTIFIDDINMPIINEWGDQ…intron
Horse	GPPAGKKMTVFIDDVNMPIINEWGDQ…intron
Opossum	GPPAGKKMTVFIDDVNMPVINEWGDQ…intron
Chicken	GPPAGKKMTVFIDDVNMPIINEWGDQ…intron
Stickleback	GPPAGKKMSVFIDDINMPVINEWGDQ…intron

The next missense mutation is also considered damaging because it is conserved across all species. It is located at the end of exon 59 at position 3350 (S3350P).

Table 27. Dnah5 missense mutation is located at the end of exon 59 and is damaging.

	5′----------→3′
Mouse	KIDVDKGCTMPSWQESLKLMTAGNFLQNLQ…intron
Rat	KIDLDKSCTVPSWQESLKLMTAGNFLQNLQ…intron
Human	KIDLEKSCTMPSWQESLKLMTAGNFLQNLQ…intron
Orangutan	KIDLEKSCTMPSWQESLKLMTAGNFLQNLQ…intron
Dog	KIDLEKSCTIPSWQESLKLMTAGNFLQNLQ…intron
Horse	KIDLEKSCTIPSWQESLKLMTAGNFLQNLQ…intron
Opossum	KIDLEKSCVIPSWQESLKLMTAGNFLQNLQ…intron
Chicken	KIDQEKSCTTPSWQESLKLMTAGNFLQNLQ…intron
Stickleback	KIDPEKNCNTPSWQESLKLMTAGNFMGSLQ…intron

The next Dnah5 mutation is located in exon 76 at position 4390(L4390P). It is conserved across all species and is located in a fairly conserved region. This mutation is also considered damaging.

Table 28. Dnah5 missense mutation in exon 76 is damaging.

	5′----------→3′
Mouse	ERLQKMGPFQPMNIFLRQEIDRMQRVLSLVR
Rat	ERLQKMGPFQPMNIFLRQEIDRMQRVLSLVR
Human	ERLQKMGPFQPMNIFLRQEIDRMQRVLSLVR
Orangutan	ERLQKMGPFQPMNIFLRQEIDRMQRVLSLVR
Dog	ERLQKMGPFQPMNIFLRQEIDRMQKLLSLVR
Horse	ERLQKMGPFQPMNIFLRQEIDRMQRVLTLVR
Opossum	DRLQKMGLFQPMNIFLRQEIDRMQRVISLVR
Chicken	EKLKNMGPFQPMNIFLRQEIQQMQRVISLVR
Stickleback	ERLQKMGPFQPMNIFLRQEIDRMQRVIVLVR

The final mutation in Dnah5 is located in exon 76 at position 4404 (S4404T) and is conserved throughout all species except for the stickleback. The wild type amino acid is an uncharged serine, which is then mutated into an uncharged threonine. Since there is no change in charge this mutation is possibly tolerated.

Table 29. Dnah5 missense mutation in exon 76 is damaging.

	5′----------→3′
Mouse	FLRQEIDRMQRVLSLVRSTLTE
Rat	FLRQEIDRMQRVLSLVRSTLTE
Human	FLRQEIDRMQRVLSLVRSTLTE
Orangutan	FLRQEIDRMQRVLSLVRSTLTE
Dog	FLRQEIDRMQKLLSLVRSTLTE
Horse	FLRQEIDRMQRVLTLVRSTLTE
Opossum	FLRQEIDRMQRVISLVRSTLTE
Chicken	FLRQEIQQMQRVISLVRSTLTE
Stickleback	FLRQEIDRMQRVIVLVRNTLTE

3.2.18 Dnaic1 missense mutation

Dnaic1 contains one missense mutation, which is located at the third amino acid of the end of exon 16 at position 522 (I522T). The amino acid is conserved throughout all nine species.

Table 30. Dnaic1 missense mutation is damaging and located at the end of exon 16 in a fairly conserved region.

	5′----------→3′
Mouse	KEIDYMFLVGTEEGKIYK…intron
Rat	KEIDYLFLVGTEEGKIYK…intron
Human	KEIDYMFLVGTEEGKIYK…intron
Orangutan	KEIDYMFLVGTEEGKIYK…intron
Dog	KEIDYMFLVGTEEGKIYK…intron
Horse	KEIDYMFLVGTEEGKIYK…intron
Opossum	KQIDYLFLVGTEEGKIYK…intron
Chicken	KKIDYLFLVGTEEGKIYK…intron
Stickleback	KQIDYLFLVGTGEGKIHK…intron

3.2.19 Dync2h1 missense mutation

The missense mutation in Dync2h1 is located in exon 5 at position 234 (V234E). It is conserved in all species except for the stickleback. The recovered mutation changes the wild type, uncharged, nonpolar valine into a negatively charged, polar glutamic acid. Since the mutation changes the charge and polarity, it is considered possibly damaging.

Table 31. Dync2h1 missense mutation in exon 5 is possibly damaging.

	3′←----------5′
Mouse	LMRSEPYHDHETQRWVDDVVDRTTEVLDVVE
Rat	LMRSEPYHDHETQRWVDDVVDRTTEVLDVVE
Human	LMRSEPYHDHETQRWVDDVVDQTTEVLDVVE
Orangutan	LMRSEPYHDHETQRWVDDVVDQTTEVLDVVE
Dog	LMRSEPYHDHETQRWVDDVVDRTTEVLDVVE
Horse	LMRSEPYHDYETQRWVDDVVDRTTEVLDVVE
Opossum	LMRAEPYHEHETQRWVDDVADRTTEVLDVVE
Chicken	RMRPQPFPDHETQRWVNDVTDRTTEVLDVVE
Stickleback	VMRTEPYPEYDTQRWTDDLADRSQEVLDMVD

3.2.20 Dyx1c1 missense mutation

The Dyx1c1 gene has one missense mutation located at the first amino acid in exon 1. It is conserved in almost all species except the opossum. The recovered mutation changes the uncharged, nonpolar methionine into a positively charged, polar lysine. This mutation is thus considered possibly damaging because of the change in charge and polarity.

Table 32. Dyx1c1 missense mutation is located in the first amino acid of exon 1 and is possibly damaging.

	5′----------→3′
Mouse	Intron…MPVRVSEFSWQQTPAT
Rat	Intron…MPVRVSEFSWQQTPAA
Human	Intron…MPLQVSDYSWQQTKTA
Orangutan	Intron…MPLQVSDYSWQQTKTA
Dog	Intron…MPLLVSDYNWQQTKTA
Horse	Intron…MPLQVSDYSWQQTKVA
Opossum	Intron…TPLQIRDYSWQQTEST
Chicken	Intron…MPLWLREYSWRQSGAA
Stickleback	Intron…==========

3.2.21 Foxj1 missense mutation

Foxj1 has one missense mutation, which is located in exon 2 at position 142 (I142N). The wild type amino acid is conserved across all species and so the recovered mutation is damaging.

Table 33. Foxj1 missense mutation in exon 2 is damaging.

	3′←----------5′
Mouse	FYCFNDTIWKYIASLTIKTAKSAQMAMCILTA
Rat	
Human	FYCFNDTIWKYIASLTIKTAKSAQMAMCILTA
Orangutan	FYCFNDTIWKYIASLTIKTAKSAQMAMCILTA
Dog	FYCFNDTIWKYIASLTIKTAKSAQMAMCILTA
Horse	
Opossum	FYCFNDTIWKYIASLTIKTAKSAQMAMCILTA
Chicken	FYCFNDTIWKYIASLTIKTKKSAEMAMCILTA
Stickleback	YYCFNETIWKYICALTIKTKKSAQMAMCILTA

3.2.22 Frem2 missense mutation

The Frem2 mutation is located at the seventh to last amino acid in exon 11 at position 2292 (Y2292C). It is conserved across all nine species and is thus a damaging mutation. It is also located in a fairly conserved region.

Table 34. Frem2 missense mutation in exon 11 is damaging and located in a conserved region at the end of the exon.

	3′←----------5′
Mouse	Intron…EESVPHYDEGSTASGDKTHVRV
Rat	Intron…EESVPHYDEGSTASGDKTHVRV
Human	Intron…EESVPHYDEGSTASGDKTHVRV
Orangutan	Intron…EESVPHYDEGSTASGDKTHVRV
Dog	Intron…EESVPHYDEGSTASGDKTHVRV
Horse	Intron…EESVPHYDEGSTASGDKTHVRV
Opossum	Intron…EESVPLYDEGSVASGDKTHVRV
Chicken	Intron…EESIPHYDEGSIASGDKTHVRV
Stickleback	Intron…EESIPHYDEGSTASGDKTHVRV

3.2.23 Gm1060 has 2 missense mutations

The Gm1060 missense mutation is conserved across all eight species even though the surrounding region is not highly conserved as you can see in *Table 35*. The mutation is located in exon 15 at position 674 (Y674C).

Table 35. Gm1060 missense mutation in exon 15 is damaging.

	5′------------------------→3′
Mouse	LKLKKETRDNSKDTEYWESLAAVIPFFKQNL
Rat	QRLKKDTRNNLKDTEYWESLAMVIPFSKQNL
Human	LRVQKNVRDNSKDSEYWQALTTVIPSSKQNL
Orangutan	LRVQKNVRDNSKDSEYWQALTTVIPSSKQNL
Dog	V--LKHVRDDSKDSEYWEALTTVISDTTQNL
Horse	VKLMKDVRDDSEDSEYWKALATVIPDATQNL
Opossum	VKVVREIRDTSKDSEYWTALCYIIPLQRQKL
Chicken	AKEVLQVRDSSKDGEYWEALAHVIPEPTLKL
Stickleback	

The second missense mutation in Gm1060 is located only 12 nucleotides, or 4 amino acids, after the first missense mutation. It is also located in exon 15 at position 678 (L678Q). Similarly to the first missense mutation, the wild type amino acid is conserved across all eight species and is located in a poorly conserved region.

Table 36. Gm1060 missense mutation in exon 15 is damaging.

	5′------------------------→3′
Mouse	KETRDNSKDTEYWESLAAVIPFFKQNLWDAL
Rat	KDTRNNLKDTEYWESLAMVIPFSKQNLWDAL
Human	KNVRDNSKDSEYWQALTTVIPSSKQNLWDAL
Orangutan	KNVRDNSKDSEYWQALTTVIPSSKQNLWDAL
Dog	KHVRDDSKDSEYWEALTTVISDTTQNLWDAL
Horse	KDVRDDSEDSEYWKALATVIPDATQNLWDAL
Opossum	REIRDTSKDSEYWTALCYIIPLQRQKLWDAL
Chicken	LQVRDSSKDGEYWEALAHVIPEPTLKLWDAL
Stickleback	

3.2.24 Hace1 missense mutation

The Hace1 missense mutation is conserved across all nine species and is also located in a highly conserved region and the surrounding region is also highly conserved. The mutation occurs in the sixteenth amino acid of exon 19 at position 687 (Y687F).

Table 37. Hace1 missense mutation in exon 19 is damaging and located at the beginning of the exon in a highly conserved region.

	5′----------→3′
Mouse	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Rat	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Human	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Orangutan	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Dog	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Horse	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Opossum	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Chicken	Intron…GIPVNYQDVASIDPEYAKNLQWILDNISDLG
Stickleback	Intron…GIPVNYQDVSSIDPEYAKNLQWILDNISDLG

3.2.25 Hectd1 nonsense mutation

Hectd1 nonsense mutation is located in the fourth amino acid of exon 22 at position 100 (Y1088X). The mutation is conserved throughout all nine species and is also located in a highly conserved region.

Table 38. Hectd1 nonsense mutation is located at the beginning of exon 22 in a highly conserved region.

	3′←----------5′
Mouse	STVVVLGYAAPNVWEYATK…intron
Rat	STVVVLGYAAPNVWEYATK…intron
Human	STVVVLGYAAPNVWEYATK…intron
Orangutan	STVVVLGYAAPNVWEYATK…intron
Dog	STVVVLGYAAPNVWEYATK…intron
Horse	STVVVLGYAAPNVWEYATK…intron
Opossum	STVVVLGYAAPNVWEYATK…intron
Chicken	STVVVLGYAAPNVWEYATK…intron
Stickleback	STVVVFGYAAPNVWEYATK…intron

3.2.26 Hr nonsense mutation

The Hr nonsense mutation is located at the end of exon 13 at position 911 (W911X). The mutation occurs in the ninth to last amino acid and introduces an early stop codon towards the end of the exon. It is conserved across all seven species, but is located in a poorly conserved region.

Table 39. Hr nonsense mutation is located at the end of exon 13 in a poorly conserved region.

	5′----------→3′
Mouse	ALGPPQPTNLDSTAFWEGFSHPET…intron
Rat	ALGPPQPTSLDSTAFWKGFSHPEA…intron
Human	PLGPPQPSSLGSTTFWEGFSWPEL…intron
Orangutan	PLGPPQPTSLDSTTFWEGFSWPEL…intron
Dog	PLGPPQPTSLRSATFWEGFSRPEI…intron
Horse	PLGPPQPTSLGSTAFWEGFSRPEV…intron
Opossum	LLGPPRPTDLSSTAFWKGFSRPEA…intron
Chicken	
Stickleback	

3.2.27 Ift140 missense mutation

Ift140 has a missense mutation that occurs at the sixth to last amino acid of exon 9 at position 380 (N380D). The mutation is conserved across all species and is thus considered damaging.

Table 40. Ift140 missense mutation in exon 9 is damaging.

	5′----------→3′
Mouse	GKDMWALQTPTELEGNITQIK…intron
Rat	GKDMWALQTPTELEGNITQIK…intron
Human	GKDRWALQTPTELQGNITQIQ…intron
Orangutan	GKDRWALQTPTELQGNITQIQ…intron
Dog	==========
Horse	GKDRWTLQTPTELEGNITQIK…intron
Opossum	GKDKWTLQTPTELEGNITQIK…intron
Chicken	GKEKWKLQASTELEGNITQIK…intron
Stickleback	AKAQWQLQTPTEVDGNVIQLQ…intron

3.2.28 Ift74 missense mutation

Ift74 missense mutation is located in the eight amino acid of exon 19 at position 549 (W549R), which is about half way into the exon. The mutation is conserved across all nine species and is also in a fairly conserved region therefore the mutation is damaging.

Table 41. Ift74 missense mutation in exon19 is damaging.

	5′----------→3′
Mouse	Intron…LTNLERKWQHLEQNNFVMKEF…intron
Rat	Intron…LTNLERKWQHLEQNNFVMKEF…intron
Human	Intron…LTNLERKWQHLEQNNFAMKEF…intron
Orangutan	Intron…LTNLERKWQHLEQNNFAMKEF…intron
Dog	Intron…LTNLERKWQHHEQNNFVMKEF…intron
Horse	Intron…LTNLERKWQHHEQNNFVMKEF…intron
Opossum	Intron…LTNLERKWQHHEQNNFVMKEF…intron
Chicken	Intron…LTNLERKWQNLEQNNFMMKEF…intron
Stickleback	

3.2.29 Kif7 missense mutation

The Kif7 missense mutation is located in the tenth amino acid of exon 3 at position 186 (V186E). The mutation is located in a fairly conserved region within the exon that is also conserved across all species and considered damaging.

Table 42. Kif7 missense mutation in exon 3 is damaging.

	3′←----------5′
Mouse	ANGMELLSLVEDLGEVDVEKVGCLV…intron
Rat	ANGMELLSLVEDLGEVDVEKVGCLV…intron
Human	ANGMELLSLVEDLGEVDVEKVGCLV…intron
Orangutan	ANGMELLSLVEDLGEVDVEKVGCLV…intron
Dog	ANGMELLSLVEDLGEVDVEKVGCLV…intron
Horse	ANGMELLSLVEDLGEVDVEKVGCLV…intron
Opossum	ANGVELLSLVEDLGEVEVEKVGCLV…intron
Chicken	TNGMELLSLVEDLGEVESEKVGCLV…intron
Stickleback	TNGSELLSLVEDLGEVECEKVGFLV…intron

3.2.30 Lox missense mutation

The missense mutation in the Lox gene is located in the third to last amino acid at the end of exon 3 at position 285 (C285F). The amino acid is conserved across all species and is located in a highly conserved region.

Table 43. Lox missense mutation at the end of exon 3 is in a highly conserved region and is damaging.

	3′←----------5′
Mouse	Intron…QHCSHWEWSYRPRSPLFD
Rat	Intron…QHCSHWEWSYRPRSPLFD
Human	Intron…QHCSHWEWSYRPRSPLFD
Orangutan	Intron…QHCSHWEWSYRPRSPLFD
Dog	Intron…QHCSHWEWSYRPRSPLFD
Horse	Intron…QHCSHWEWSYRPRSPLFD
Opossum	Intron…QHCSHWEWSYRPRSPLFD
Chicken	Intron…==========PLFD
Stickleback	Intron…QHCSHWEWSYRPRSPLFD

3.2.31 Lrp1 missense mutation

Lrp1 has one missense mutation, which is located in exon 82 at position 4232 (C4232R). The wild type amino acid is conserved across all species and is in a highly conserved region, with the exception of the stickleback species. This gene is not present in the chicken and there is not genetic information for the opossum.

Table 44. Lrp1 missense mutation in exon 82 is damaging and located in a highly conserved region.

	3′←----------5′
Mouse	CTGGNHCYEWCQDLECKDGTYRPQCRCKPQR
Rat	CTGGNHCYEWCQDLECKDGTYRPQCRCKPQR
Human	CTGGNRCHEWCQDLECKDGTYRPQCRCKPQR
Orangutan	CTGGNRCHEWCQDLECKDGTYRPQCRCKPQR
Dog	CTGGNRCYEWCQDLECKDGTYRPQCRCKPQR
Horse	CTGGNRCHEWCQDLECKDGTYRPQCRCKPQR
Opossum	==========\==========\==========
Chicken	
Stickleback	CTGGNKCYDRCLDIQCRDGGYNPQCHCKPQK

3.2.32 Ltbp1 nonsense mutation

Ltbp1 has a nonsense mutation located in exon 7 at position 378 (C378X). The amino acid at this position is conserved across all nine species, so the mutation is considered damaging. It is also located in a fairly conserved region of the exon.

Table 45. Ltbp1 missense mutation in exon 7 is damaging.

	5′----------→3′
Mouse	CMHPLSVHLTKQICCCSVGKAWGPHCEKCPL
Rat	CMHPLSVHLTKQICCCSVGKAWGPQCEKCPL
Human	CMHPLSVHLTKQLCCCSVGKAWGPHCEKCPL
Orangutan	CMHPLSVHLTKQLCCCSVGKAWGPHCEKCPL
Dog	CMHPLSVHLTKQLCCCSVGKAWGPHCEKCPL
Horse	CMHPLSVHLTKQLCCCSVGKAWGPHCEKCPL
Opossum	CMHPLSVQLSKQLCCCSVGKAWGPHCEKCPL
Chicken	CMHPLSVQLSKQLCCCSVGKAWGPNCEKCPL
Stickleback	CLHPVSTMLSKQLCCCSVGKAWGPRCDKCPP

3.2.33 Megf8 has 2 missense mutations

The first Megf8 missense mutation is located in exon 3 at position 148 (N148D). The amino acid at this position is not conserved across all species. The wild type amino acid that is present in mice is only present in rats also. The wild type amino acid is an uncharged, polar asparagine, which is changed into a negatively charged, polar aspartic acid. Since there is a change in charge, this mutation is considered possibly damaging.

Table 46. Megf8 missense mutation in exon 3 is possibly damaging.

	5′----------→3′
Mouse	FNASFRFSLCPGGCQNHGQCKSPGVCVCEPG
Rat	FNASFRFSLCPGGCQNHGQCKSPGVCVCEPG
Human	FNASFRFSLCPGGCQSHGQCQPPGVCACEPG
Orangutan	FNASFRFSLCPGGCRSHGQCQPPGVCACEPG
Dog	FNASYRFTLCPGGCSGHGQCQAPGVCTCEPG
Horse	FNASFRFSLCPGGCRSHGQCRPPGVCACEPG
Opossum	FNASFSFSLCPGGCSGRGLCEPSGLCTCDPG
Chicken	
Stickleback	FNATYTFSACPGACGGHGRCDPSTSCLCHQG

The second missense mutation in Megf8 is located in exon 21 at position 1214 (N1214I). This wild type amino acid is conserved across all species and is therefore considered damaging. It is also located in a fairly conserved region in the exon.

Table 47. Megf8 missense mutation in exon 21 is damaging.

	5′----------→3′
Mouse	FGNATGSGGCRPCQCNGHGDPRRGHCDNLTG
Rat	FGNATGSGGCRPCQCNGHGDPRRGHCDNLSG
Human	FGNATGSRGCRPCQCNGHGDPRRGHCDNLTG
Orangutan	FGNATGSRGCRPCQCNGHGDPRRGHCDNLTG
Dog	FGNATGSGGCRPCQCNGHGDPRRGHCDNLSG
Horse	FGNATGSGGCRPCQCNGHGDTRRGHCDNLSG
Opossum	FGNATGPGGCQPCQCNGHGDPQQGHCDGHTG
Chicken	
Stickleback	FGSALGGGGCVQCECNGHGDPARGYCHNQTG

3.2.34 Mical1 missense mutation

The Mical1 missense mutation is located at the end of exon 22 at position 929 (M929T). The amino acid is not conserved across all species. The wild type amino acid is a nonpolar, hydrophobic methionine, which is then changed into a polar, uncharged threonine. Since there is a change in polarity, this mutation is considered possibly damaging.

Table 48. Mical1 missense mutation is located at the end of exon 22 and is possibly damaging.

	5′----------→3′
Mouse	LEEKQRQLDHELRGYMNRE…intron
Rat	LEEKQRQLDHEFRG-INRE…intron
Human	LEEKQWQLDQELRGYMNRE…intron
Orangutan	LEEKQWQLDQELRGYMNQE…intron
Dog	LEEKQWHLDQELRGYLNQE…intron
Horse	LEEKQWQLDQELRGYMNRE…intron
Opossum	LEEKQWQLDQELRSYMDRK…intron
Chicken	LEEQQWQLDQELRWYIETE…intron
Stickleback	LEDKQSMLELELRKYMELN…intron

3.2.35 Mmp21 missense mutation

The missense mutation in Mmp21 is located in exon 2 at position 177 (W177L) and the wild type amino acid is not conserved across all species. At this position, the wild type amino acid is an uncharged, nonpolar tryptophan, which changes into a similarly uncharged, nonpolar leucine. Due to the lack of change in charge and polarity in this mutation it is considered possibly tolerated.

Table 49. Mmp21 missense mutation in exon 2 is possibly tolerated.

	3′←----------5′
Mouse	EDGSLQSSYADGVLRWTLTKKSFARSAGTDS
Rat	EDGSLQSSYADGVLRWTLTKKSFVRSAGTDS
Human	DAVSLQSSLAEGLLRWSLTRKSFAQAAGGDP
Orangutan	EAASLQSSLAEGLLRWSLTKKSFAQAAGGDP
Dog	EAEALQGSAGERLLRWSLTRKAFARPAGTPA
Horse	EDLALQSSVGEGLLRRTLPRKSFARPGGAPP
Opossum	EDIPLQSSYGEGMLRWRLLKRSFAQPAG--E
Chicken	============================
Stickleback	EEVSLQSSYGEGILRWKLVRKTFAM-VPTG*

3.2.36 Nek8 missense mutation

The Nek8 missense mutation is located in exon 3 at position 124 (I124T). The wild type amino acid is conserved across all species and is also located in a highly conserved region. This mutation is considered damaging.

Table 50. Nek8 missense mutation in exon 3 is damaging.

	3′←----------5′
Mouse	HKDLLINQTKLDRHLILHTHVHHLALLIQVF
Rat	HKDLLINQTKLDRHLILHTHVHHLALLIQVF
Human	HKDLLINQTKLDRHLILHTHVHHLALLIQVF
Orangutan	HKDLLINQTKLDRHLILHTHVHHLALLIQVF
Dog	HKDLLINQTKLDRHLILHTHVHHLALLIQVF
Horse	HKDLLINQTKLDRHLILHTHVHHLALLIQVF
Opossum	QKDLLINQTKLDRHLILHSHVHHLALLIQVF
Chicken	HKDLLINQTKLDRHLIQKTHVHHLALLIQVF
Stickleback	HKDLLINQTKLDRHLILKNHVHYLALLIQVF

3.2.37 Pcsk5 has 2 missense mutations

The first mutation in Pcsk5 is located at the beginning of exon 4 at position 174 (D174G). The amino acid is conserved across al nine species and is therefore damaging.

Table 51. Pcsk5 missense mutation in exon 4 is damaging.

	3′←----------5′
Mouse	Intron…YNQMLDPHTREIGDDLITVVINKGTYGRK
Rat	Intron…YNQMLDPHTREIGDDLITVVINKGTYGRK
Human	Intron…YNQMLDPHTREIGDDLITVVINKGTYGRK
Orangutan	Intron…YNQMLDPHTREIGDDLITVVINKGTYGRK
Dog	Intron…YNQMLDPHTREIGDDLITVVINKGTYGRK
Horse	Intron…YNQMLDPHTREIGDDLITVVINKGTYGRK
Opossum	Intron…YNKVLDPHTREIGDDLITVVVDRGTYGRK
Chicken	Intron…YNQMLDPHNREIGDDLITVVVNKGTYGRK
Stickleback	Intron…YNQFLDPHNREIGDDLITVVVDKGTYGRK

The second missense mutation in Pcsk5 is located in exon 35 at position 1599 (C1599S). The mutation is conserved across all nine species and is therefore damaging.

Table 52. Pcsk5 missense mutation in exon 35 is damaging.

	3′←----------5′
Mouse	DCSQCDTPRPGRCSKCSKDCKECRGSTSEGY
Rat	DCSQCDTPQPGRCTKCSKDCKECRGSTSEGY
Human	DCSLCDTPRPGQCGKCSRNCRECRGTSNDAY
Orangutan	DCSLCDTPRPGQCGKCSRNCRECRGTSNDAY
Dog	DCSLCDAPGPGQCAKCSKSCRECRGTSIEAY
Horse	DCSLCDTPRPGRCAKCSRNCRECRGTANEAY
Opossum	HCSLCNTSHPGWCKRCDQNCKECLGSSNNAY
Chicken	DCFLCDTPQPGMCEKCGKHCHECSGTSNQAF
Stickleback	RCSLCNKNGPGWCRSCAPDCRECVMTNTNGF

3.2.38 Pde2a missense mutation

The Pde2a missense mutation is located in the middle of exon 23 at the thirteenth amino acid at position 682 (M682R). The mutation is located in a highly conserved region where it is conserved across all species and thus is a damaging mutation.

Table 53. Pde2a missense mutation in exon 23 is damaging.

	5′----------→3′
Mouse	Intron…EDIEIFALFISCMCHDLDHRGTNNSFQVAS…intron
Rat	Intron…EDIEIFALFISCMCHDLDHRGTNNSFQVAS…intron
Human	Intron…EDIEIFALFISCMCHDLDHRGTNNSFQVAS…intron
Orangutan	Intron…EDIEIFALFISCMCHDLDHRGTNNSFQVAS…intron
Dog	Intron…EDIEIFALFISCMCHDLDHRGTNNSFQVAS…intron
Horse	Intron…EDIEIFALFISCMCHDLDHRGTNNSFQVAS…intron
Opossum	Intron…DDIEIFALFVSCMCHDLDHRGTNNSFQVAS…intron
Chicken	
Stickleback	Intron…EEIEILALFVSCMCHDLDHRGTNNSFQVAS…intron

3.2.39 Pkd1 missense mutation

Pkd1 contains a missense mutation that is located at the end of exon 38 at position 3695 (I3695N). The mutation is conserved across all species and is thus a damaging mutation.

Table 54. Pkd1 missense mutation in exon 38 is damaging.

	5′----------→3′
Mouse	GDASCHGHAYRLQSAIKQELDSQAFLAITR…intron
Rat	GDASCHGHAYRLQSAIKQELDSQAFLAITR…intron
Human	GDASCHGHAYRLQSAIKQELHSRAFLAITR…intron
Orangutan	GDASCHGHAYRLQSAIKQELHSRAFLAITR…intron
Dog	GDASCHNHAYRLQSAIKQELGSQAFLAITR…intron
Horse	GDASCHSHAYRLQSAIKQELDSQAFLAITR…intron
Opossum	GDASSHTHAYRLQSSIKQELDSKAFLGITR…intron
Chicken	GDATRNSRAFLLQSSIKQQLGSSEFLHIKR…intron
Stickleback	==========I==========

3.2.40 Plxnd1 missense mutation

The Plxnd1 missense mutation is located in a highly conserved region of exon 27 at position 1576 (D1576G). The mutation is conserved across all species except for the stickleback species. The wild type amino acid at this position is a polar, negatively charged aspartic acid. The mutation changes this into a nonpolar, uncharged glycine. Since there is a change in charge and polarity, the mutation is considered possibly damaging.

Table 55. Plxnd1 missense mutation in exon 27 is possibly damaging.

	3′←----------5′
Mouse	FAELIKEKVQTLTDTDMARVSLSDMGCGQFS
Rat	FAELIKEKVQTLTDTDMARVSLSDMGCGQFS
Human	FAELIKEKVQTLTDTDMARVSLSDMGCGQFS
Orangutan	FAELIKEKVQTLTDTDMARVSLSDMGCGQFS
Dog	FAELIKEKVQTLTDTDMARVSLSDMGCGQFS
Horse	FAELIKEKVQTLTDTDMARVSLSDMGCGQFS
Opossum	FAELIKEKVQTLTDTDMARVSLSDMGCGQFS
Chicken	FAELIKEKVQSLTDTDMARVSLSDMGCGQFS
Stickleback	FAELIKEKVQCVTDTNMVRVPLSDMGCGQFS

3.2.41 Prickle1 missense mutation

The Prickle1 missense mutation is located in exon 5 at position 161 (C161F). The mutation is located in a highly conserved region and is also conserved across all nine species, making it a damaging mutation.

Table 56. Prickle1 missense mutation in exon 5 is damaging.

	3′←----------5′
Mouse	KGDQYFYILDVLLENCTFCVFCSPHWCVGPG
Rat	KGDQYFYILDVLLENCTFCVFCSPHWCVGPG
Human	KGDQYFYILDVLLENCTFCVFCSPHWCVGPG
Orangutan	KGDQYFYILDVLLENCTFCVFCSPHWCVGPG
Dog	KGDQYFYILDVLLENCTFCVFCSPHWCVGPG
Horse	KGDQYFYILDVLLENCTFCVFCSPHWCIGPG
Opossum	KGDQYFYILDVLLENCTFCVFCAPHWWIGPG
Chicken	KGDQYFYILDVLLENCTFCVFCSPHWCVGPG
Stickleback	KGNQYFYILDVLLEQCTTCVFCAPHWCPGLG

3.2.42 Pskh1 missense mutation

Pskh1 contains one missense mutation located in exon 2 at position 219 (L219P). The mutation is considered damaging because the wild type amino acid is conserved across all species and it is also in a highly conserved region.

Table 57. Pskh1 missense mutation in exon 2 is damaging.

	5′----------→3′
Mouse	DGVRYLHALGITHRDLKPENLLYYHPGTDSK
Rat	DGVRYLHALGITHRDLKPENLLYYHPGTDSK
Human	DGVRYLHALGITHRDLKPENLLYYHPGTDSK
Orangutan	DGVRYLHALGITHRDLKPENLLYYHPGTDSK
Dog	DGVRYLHALGITHRDLKPENLLYYHPGTDSK
Horse	DGVRYLHALGITHRDLKPENLLYYHPGTDSK
Opossum	DGVKYLHTLGITHRDLKPENLLYYHPGTDSK
Chicken	DGVRYLHTLGITHRDLKPENLLYYHPGTDSK
Stickleback	DGVKYLHTLGITHRDLKPENLLYYHPGSDSK

3.2.43 Psme4 missense mutation

The Psme4 mutation is located at the end of exon 28 at position 1083 (I1083V). The mutation is in a highly conserved region and is also conserved across all species so we may call it damaging.

Table 58. Psme4 missense mutation in exon 28 is damaging.

	5′----------→3′
Mouse	EKPSIVRLFDDLAEKIHRQYETIGLDFT…intron
Rat	EKPSIVRLFDDLAEKIHRQYETIGLDFT…intron
Human	EKPSIVRLFDDLAEKIHRQYETIGLDFT…intron
Orangutan	EKPSIVRLFDDLAEKIHRQYETIGLDFT…intron
Dog	EKPSIVRLFDDLAEKIHRQYETIGLDFA…intron
Horse	EKPSIVRLFDDLAEKIHRQYETIGLDFT…intron
Opossum	EKPSIVRLFDDLAEKIHRQYETIGLDFT…intron
Chicken	EKPSIVRLFDDLAEKIHRQYETIGLDFA…intron
Stickleback	EKPSIVRLFDDLADKIHRQYETIGVDFS…intron

3.2.44 Robo1 missense mutation

Robo1 missense mutation is located in the fourth amino acid of exon 6 at position 270 (I270T). The amino acid is conserved across all species so we can call it a damaging mutation.

Table 59. Robo1 missense mutation in exon 6 is damaging.

	5′--------------------→3′
Mouse	Intron…RYEIRDDHTLKIRKVTAGD
Rat	Intron…RYEIRDDHTLKIRKVTAGD
Human	Intron…RYEIRDDHTLKIRKVTAGD
Orangutan	Intron…RYEIRDDHTLKIRKVTAGD
Dog	Intron…RYEIRDDHTLKIRKVMAGD
Horse	Intron…RYEIRDDHTLKIRKVMAGD
Opossum	Intron…RYEIRDDHTLKIRKVMAGD
Chicken	Intron…RYEIRDDHTLKIRKVMAGD
Stickleback	Intron…RYEIREDHTLKIRRLTSAD

3.2.45 Sema6a missense mutation

The Sema6a missense mutation is located in a highly conserved region in exon 10 at position 275 (S275P). The wild type amino acid is conserved across all species and is considered damaging.

Table 60. Sema6a missense mutation in exon 10 is damaging.

	3′←-------------------------5′
Mouse	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Rat	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Human	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Orangutan	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Dog	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Horse	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Opossum	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Chicken	HSDGPVSCNLRAKLFSTWQKELVRQSGGMDN
Stickleback	HSDGPISCNLRSKLFSTWQKELVRPSGGRDN

3.2.46 Smad6 nonsense mutation

The Smad6 nonsense mutation is located in the tenth amino acid of exon 4 at position 328 (K328X). It is conserved across all nine species and is also located in a highly conserved region so the mutation is damaging.

Table 61. Smad6 nonsense mutation is located in a highly conserved region at the beginning of exon 4.

	3′←----------**5′**
Mouse	RTRHEWYAVSCWHSPKTADPSMSAD…intron
Rat	RTRHEWYAVSCWHSPKTADPSMSAD…intron
Human	RTRHEWYAVSCWHSPKTADPSMSAD…intron
Orangutan	RTRHEWYAVSCWHSPKTADPSMSAD…intron
Dog	RTRHEWYAVSCWHSPKTADPSMSAD…intron
Horse	RTRHEWYAVSCWHSPKTADPSMSAD…intron
Opossum	RTRHEWYAVNCWHSQKMTDPSMSAD…intron
Chicken	RTRHEWYAVNCWHSRKVADPSTSAD…intron
Stickleback	RTRHEWYAVNCWHSQKPADPSMSAD…intron

3.2.47 Smarca4 has 2 missense mutations

The first Smarca4 missense mutation is located in the second amino acid of exon 8 at position 417 (R417C). The amino acid at this position is conserved across all species and can therefore be called a damaging mutation.

Table 62. Smarca4 missense mutation at the beginning of exon 8 is damaging.

	5′----------**→3′**
Mouse	Intron…LRQEVVVCMRRDTALET
Rat	Intron…LRQEVVVCMRRDTALET
Human	Intron…LRQEVVVCMRRDTALET
Orangutan	Intron…LRQEVVVCMRRDTALET
Dog	Intron…LRQEVVVCMRRDTALET
Horse	Intron…LRQEVVVCMRRDTALET
Opossum	Intron…LRQEVVVCMRRDTALET
Chicken	Intron…LRQEVVACMRRDTTLET
Stickleback	Intron…LRQEVVVCMRRDTALET

The second missense mutation in Smarca4 is located in a highly conserved region within exon 16 at position 794 (T794I). The wild type amino acid is conserved across all species and is a damaging mutation.

Table 63. Smarca4 missense mutation in exon 16 is damaging.

	5′----------→3′
Mouse	DEMGLGKTIQTIALITYLMEHKRINGPFLII
Rat	DEMGLGKTIQTIALITYLMEHKRINGPFLII
Human	DEMGLGKTIQTIALITYLMEHKRINGPFLII
Orangutan	DEMGLGKTIQTIALITYLMEHKRINGPFLII
Dog	DEMGLGKTIQTIALITYLMEHKRINGPFLII
Horse	DEMGLGKTIQTIALITYLMEHKRINGPFLII
Opossum	DEMGLGKTIQTIALITYLMEHKRINGPFLII
Chicken	DEMGLGKTIQTIALITYLMEHKRLNGPYLII
Stickleback	DEMGLGKTIQTIALITYLMEYKRLNGPFLII

3.2.48 Snx17 missense mutation

The Snx17 missense mutation is located in the last amino acid of exon 5 at position 144 (E144G). The mutation is located in a fairly conserved region and is conserved across all species except for the stickleback. The wild type amino acid at this position is a negatively charged, polar glutamic acid. The mutated amino acid is an uncharged, nonpolar glycine. Since there is a change in charge and polarity, the mutation is possibly damaging.

Table 64. Snx17 missense mutation at the end of exon 5 is possibly damaging.

	5′----------→3′
Mouse	VLVTVLTSDQTEDVLE…intron
Rat	
Human	VLVNVLTSDQTEDVLE…intron
Orangutan	VLVNVLTSDQTEDVLE…intron
Dog	VLVNVLTSDQTEDVLE…intron
Horse	VLVSVLTSDQTEDVLE…intron
Opossum	VLVTILTSDQTEDVLE…intron
Chicken	VKVTILTSDQTEDVLE…intron
Stickleback	VTVNILTSDQTEDVLD…intron

3.2.49 Spata5 missense mutation

Spata5 has a missense mutation, which is located in the thirteenth amino acid of exon 9 at position 498 (T498I). The wild type amino acid is conserved across all species, so it is a damaging mutation.

Table 65. Spata5 missense mutation at the beginning of exon 9 is damaging.

	5′----------→3′
Mouse	Intron…EGSEGRVLVLGATNRPQALDAALRRPGR
Rat	Intron…EGSEGRVLVLGATNRPQALDAALRRPGR
Human	Intron…EVSEGQVLVLGATNRPHALDAALRRPGR
Orangutan	Intron…EVSEGQVLVLGATNRPHALDAALRRPGR
Dog	Intron…EGSEGQVLVLGATNRPHALDAALRRPGR
Horse	Intron…EGSEGQVLVLGATNRPHALDAALRRPGR
Opossum	Intron…EGSEGRVLVIGATNRLHSLDPALRRPGR
Chicken	Intron…EGSEGQLVVLGATNRPHALDAALRRPGR
Stickleback	Intron…EGMSGQLLVLGATNRPHALDPALRRPGR

3.2.50 Sufu missense mutation

Sufu missense mutation is located in a highly conserved region within exon 4 at position 177 (M177K). The amino acid is conserved across all species so it is considered damaging.

Table 66. Sufu missense mutation in exon 4 is damaging.

	5′----------→3′
Mouse	SWHSPLDNSESRIQHMLLTEDPQMQPVRTPF
Rat	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF
Human	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF
Orangutan	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF
Dog	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF
Horse	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF
Opossum	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF
Chicken	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF
Stickleback	SWHSPLDNSESRIQHMLLTEDPQMQPVQTPF

3.2.51 Tab1 missense mutation

The Tab1 missense mutation is located in a highly conserved region within exon 9. It is the fifth amino acid within this exon and is located at position 312 (M312K). The amino acid is conserved across all species and is thus considered damaging.

Table 67. Tab1 missense mutation at the beginning of exon 9 is damaging.

	5′----------→3′
Mouse	Intron…EIAAMIDTEFAKQTSLDAVA
Rat	Intron…EVAAMIDTEFAKQTSLDAVA
Human	Intron…EIAAMIDTEFAKQTSLDAVA
Orangutan	Intron…EIAAMIDTEFAKQTSLDAVA
Dog	Intron…EIAAMIDTEFAKQTSLDAVA
Horse	Intron…EIAAMIDTEFAKQTSLDSVA
Opossum	Intron…EIAAMIDTEFAKQTSLDAVA
Chicken	Intron…EIAAMIATEFAKQTSLDAVA
Stickleback	

3.2.52 Tmem67 missense mutation

The Tmem67 missense mutation is located in the thirteenth amino acid at the beginning of exon 11 at position 374 (Y374N). The amino acid at this position is conserved in all species except for the stickleback. The wild type amino acid is an uncharged, nonpolar tyrosine, which is changed into an uncharged, polar asparagine. Since there is no change in charge with this mutation it is considered possibly tolerated.

Table 68. Tmem67 missense mutation at the beginning of exon 11 is possibly tolerated.

	3′←----------5′
Mouse	Intron…NQQYTTGFAYAADLRRETDPCL…intron
Rat	Intron…NQQYTTGFSYAADLRRETDPCL…intron
Human	Intron…NQQYTTGFSYAANLRTETDPCL…intron
Orangutan	Intron…NQQYTTGFSYAANLRTETDPCL…intron
Dog	Intron…NQQYTTGFSYAANLRTETDPCL…intron
Horse	Intron…NQQYTTGFSYAANLRTETDPCL…intron
Opossum	Intron…NQQYTTGFDYAADLRSVTDPCL…intron
Chicken	Intron…SQEYTTGFVYAANLKTQTDPCL…intron
Stickleback	Intron…SEQYATGFNFAAAQKSTTEPCL…intron

3.2.53 Tpbpa missense mutation

The Tpbpa gene is only present in a select few of the species in our set list. These species include the mouse, rat, human, orangutan, and horse. The wild type amino acid for the mouse is different than the wild type amino acid for all the other species. Our recovered mutation is located in exon 3 at position 54 (S54R). This region is poorly conserved, as is the wild type amino acid. The

recovered mutation shows us that the uncharged, polar serine is changed into a positively charged, polar arginine, meaning that the mutation is possibly damaging despite being in a poorly conserved region.

Table 69. Tpbpa missense mutation in exon 3 is possibly damaging.

	3′←----------5′
Mouse	SMEINSGETEQDNESSHLKNTKMFKSWVAKI
Rat	SMEIDLEEEEQDSKSNYLKVTKMFEDWVAKR
Human	NMAMTFSHKGERYEQNHQEIMKMNQ-WVARR
Orangutan	NMAMAFSHKGERYEQNHLEIMKMNKEWVARR
Dog	
Horse	NMAMTFGHKGQSYEQNHLEIMRMNKEWVARR
Opossum	
Chicken	
Stickleback	

3.2.54 Wdr69 has 1 missense and 1 nonsense mutation

The Wdr69 nonsense mutation is located at the beginning of exon 2 at position 19 (Y19X). The mutation is damaging because the amino acid is conserved across all species. This mutation introduces a stop codon very early in exon 2 and is therefore a damaging mutation.

Table 70. Wdr69 nonsense mutation is at the beginning of exon 2 in a fairly conserved region.

	5′----------→3′
Mouse	Intron…GIMLEYEKGGELKTKSIDLLE
Rat	Intron…GIMLEYEKGGELKTKSIDLLE
Human	Intron…GIMLEYEKHGELKTKSIDLLD
Orangutan	Intron…GIMLEYEKHGELKTKSIDLLD
Dog	Intron…GIMLEYEKSGELKTKSIDLLN
Horse	Intron…GIMLEYEKSGELKTKSIDLLD
Opossum	Intron…GIILEYEKSGELKTKSIDLLD
Chicken	Intron… IILEYVEGGELKTRSIDLLD
Stickleback	Intron… IILEYEKEGLLKTKSMDLLD

The missense mutation in Wdr69 is located in exon 3 at position 66 (L66P). The amino acid at this position is conserved in all species except for the orangutan. If we analyze this mutation further, the amino acid changes from a leucine to a proline both of which are uncharged. Since there is no change in charge, this mutation is possibly tolerated.

Table 71. Wdr69 missense mutation in exon 3 is possibly damaging.

	5′----------→3′
Mouse	AEPLITASRTKQVRLLVQRLQEKLRQHSDHN
Rat	VEPLITASRTKQVRLLVQRLQEKLRQHCDHN
Human	AEPLLTASRTEQVKLLIQRLQEKLGQNSNHT
Orangutan	AEPLLTASRTEQVKLFIQRLQEKLGQHSNHT
Dog	AEPLITASRSEQVKLLIQRLQDKLRQHSNHT
Horse	AEPLITASRKDQVKLLIQRLQDKLGQHSNHK
Opossum	AEPLITASRTDQVKHLILKLQDKLGQQGDHK
Chicken	EEPLITASCKEHVIRLVQRLQEKLGEKEDHK
Stickleback	SEPLVTESRADQVKQLILRLQQKQGQKDQFG

3.2.55 Xpnpep1 missense mutation

The Xpnpep1 missense mutation is located at the end of exon 15 at position 416 (S4196P). The mutation is located in a fairly conserved region and the wild type amino acid is conserved across all nine species making this a damaging mutation.

Table 72. Xpnpep1 missense mutation at the end of exon 15 is damaging.

	3′←----------5′
Mouse	Intron…KYQAGSDILYVEDLSLTRNTEPVPA…intron
Rat	Intron…KYQAGSDILYVEDLSLTRNTEPIPA…intron
Human	Intron…KYQAGSDILYVEDLSLTRNTEPVPA…intron
Orangutan	Intron…KYQAGSDILYVEDLSLTRNTEPVPA…intron
Dog	Intron…KYQAGSDILYVEDLSLTRNTEPVPA…intron
Horse	Intron…KYQAGSDILYVEDLSLTRNTEPVPA…intron
Opossum	Intron…KYQAGSDILYVENLSLMRNTEPVPS…intron
Chicken	Intron…KYQAGSDLLYIENVSLTRNTEPVPK…intron
Stickleback	Intron…VYQAGSDILYVENVSLTRNTEPLPR…intron

3.2.56 Zfp161 nonsense mutation

The Zfp161 nonsense mutation is located in a somewhat conserved region within exon 3 at position 140 (S140X). The mutation introduces an early stop codon in this exon, which is the reason why this mutation is damaging. It is damaging despite the fact that the wild type amino acid is not conserved in all species.

Table 73. Zfp161 nonsense mutation in exon 3 is damaging.

	5′-------------------------→3′
Mouse	SQKRDVSSPDESNGQSKSKYCLKLNRPIGDA
Rat	SQKRDVSSPDESNGQSKSKYCLKLNRPIGDA
Human	SQKRDVSSPDENNGQSKSKYCLKINRPIGDA
Orangutan	SQKRDVSSPDENNGQSKSKYCLKINRPIGDA
Dog	SQKRDVSSPDENNGQSKSKYCLKINRPIGDA
Horse	SQKRDVSSPDENNGQSKSKYCLKINRPIGDA
Opossum	SQKRDVSSPEENNTQAKNKYCLKINRPIGES
Chicken	SQKRDVSS--EENTQSKSKYCLKINRPIGEP
Stickleback	TQKRELTN--EDAGEP---------------

3.3 DISEASE-CAUSING GENES

There are a total of 62 disease-causing mutations that we were able to recover from the ENU screen. These genes all have varying functions that are involved in several different biological processes that may or may not contribute to cardiac development. *Figure 4* shows the chromosomal location of each of the 62 disease-causing genes that we recovered from our screen.

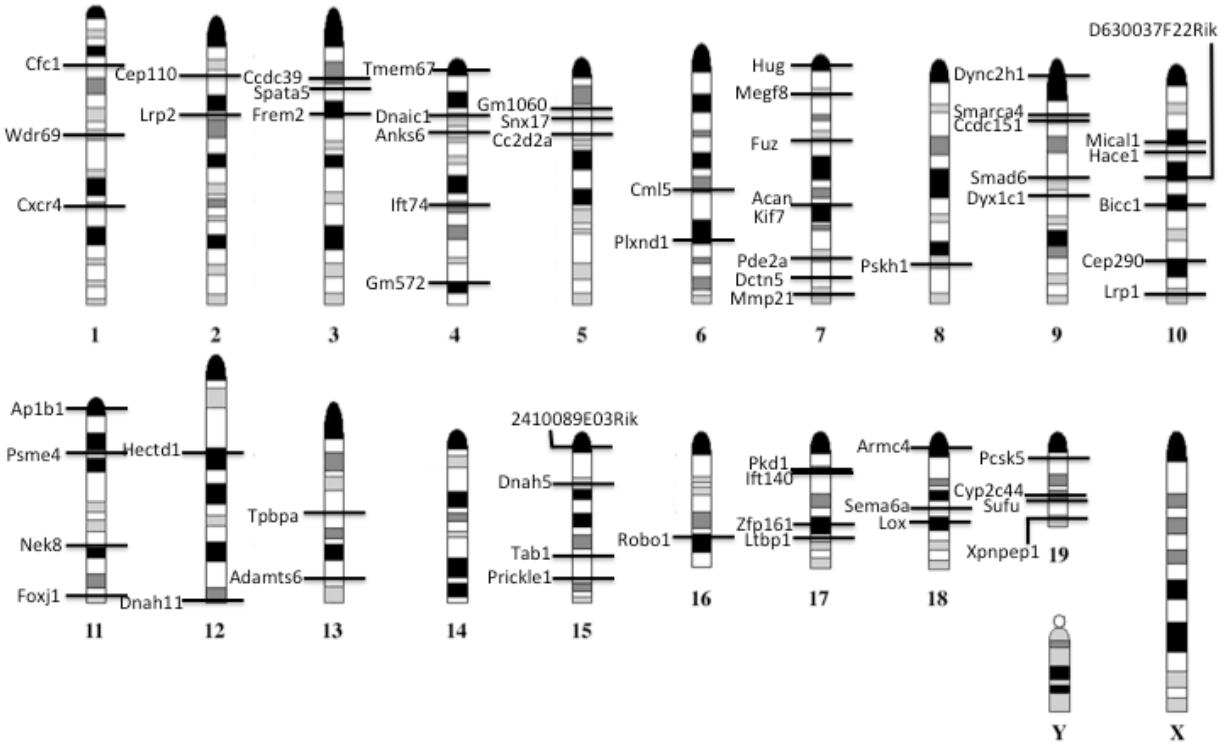

Figure 4. Chromosomal location of 65 disease-causing genes detected in ENU mice.

Figure 5 shows the number of disease causing homozygous mutations on each mouse chromosome. There are no mutations present on the sex chromosomes. This makes sense because any mutations on the X or Y chromosomes would result in a mutant phenotype in the father, and we did not see this in any of our mice. There are a varying number of mutations on each chromosome ranging from 1 to 11. However, there does not seem to be a saturation of genes on any of the chromosomes.

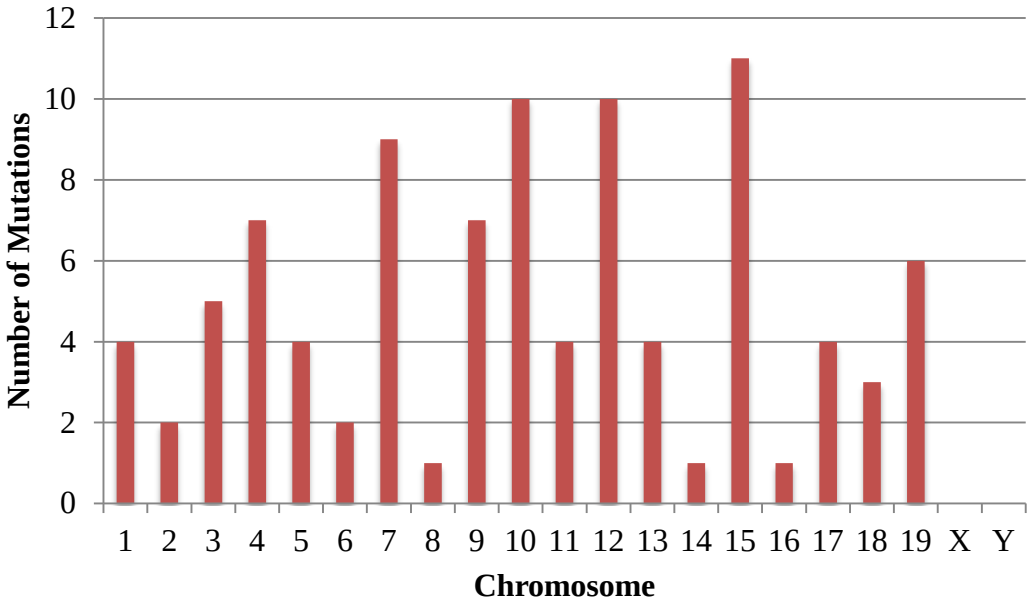

Figure 5. Graph of the number of disease-causing homozygous mutations on each mouse chromosome.

3.3.1 Multihit disease-causing genes

There are twelve disease-causing genes in which we have identified more than one mutation. *Table 74* shows these twelve genes as sorted by the gene size. The number of alleles in this table is the number of different mutations that were recovered.

Table 74. Disease-causing genes with multiple mutations as sorted by gene size.

Gene	Chr	Gene Size (CDS)	Human Ortholog	Number of Alleles
Wdr69	1	933	DAW1	2
Dnaic1	4	2106	DNA1	2
Gm1060/Drc1	5	2262	CCDC164	2
Ccdc39	3	2814	CCDC39	3
Bicc1	10	2934	BICC1	2
Tmem67	4	2988	TMEM67	2
Smarca4	9	4845	SMARCA4	2
Pcsk5	19	5634	PCSK5	3
Cep290	10	7440	CEP290	3
Megf8	7	8370	MEGF8	2
Dnah11	12	13467	DNAH11	8
Dnah5	15	13866	DNAH5	9

3.3.1.1 Bicc1—1 splicing and 1 missense mutation

The Bicc1 missense mutation in exon 6 is located in a KH domain, which binds RNA and can function in RNA recognition (Garcia-Mayoral 2007). The phenotype caused by this mutation is laterality.

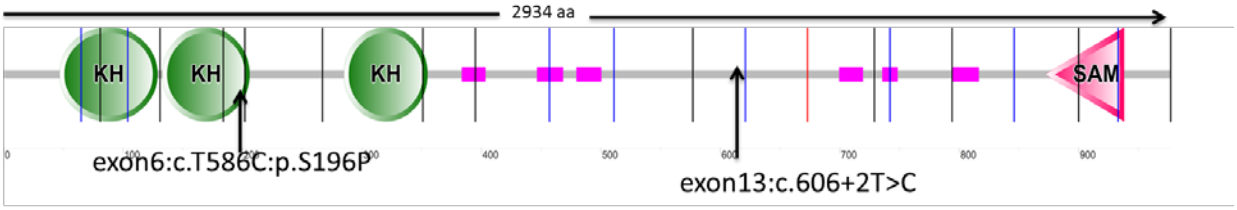

Figure 6. Bicc1 mutations are located in KH domain and in an unknown domain.

3.3.1.2 Ccdc39—1 splicing, 1 nonsense, and 1 missense mutation

The missense mutation in Ccdc39 occurs in the first amino acid, however it is not found in a functional domain. The phenotype caused by this mutation is SIT, or situs inversus totalis. The nonsense mutation in exon 4 is found just before a coiled coil domain. The phenotype involved with this mutation is AVSD (atrioventricular septal defect) and laterality.

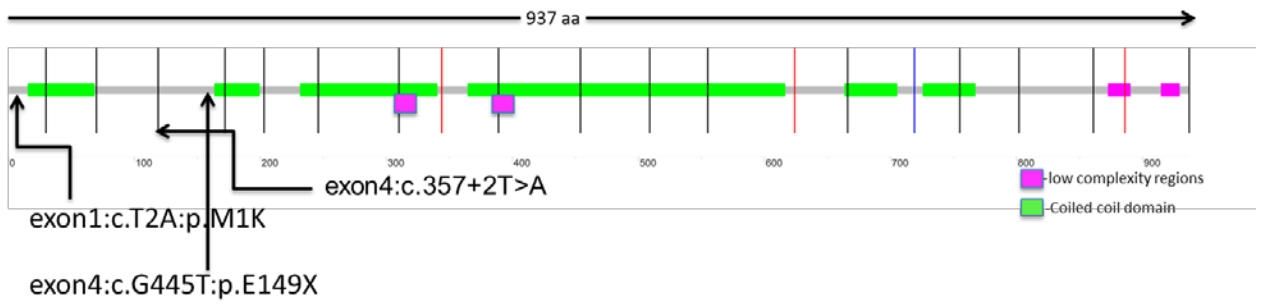

Figure 7. Ccdc39 mutations are all located in unknown domains.

3.3.1.3 Cep290—2 splicing and 1 nonsense mutation

The nonsense mutation in exon 34 is located in a coiled coil domain, which is involved in the regulation of gene expression. The phenotype associated with this mutation is mild laterality and cystic kidney. Both splicing mutations can also be found in coiled coil domains.

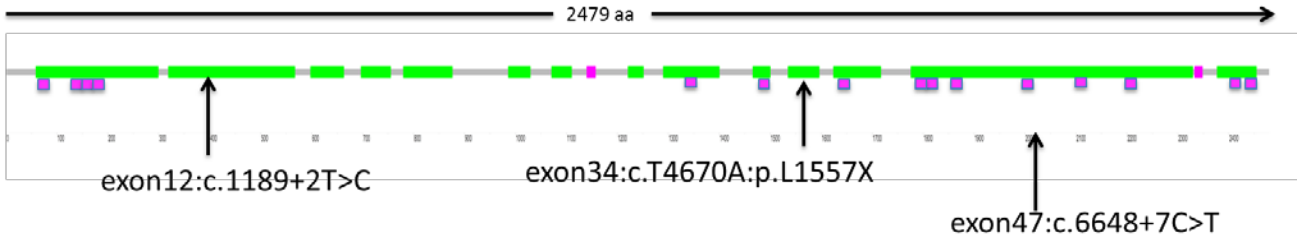

Figure 8. All three of the Cep290 mutations are in coiled coil domains.

3.3.1.4 Dnah5—3 splicing and 6 missense mutations

The first Dnah5 splicing mutation is in exon 4, which is located in an unknown domain. The phenotype related to this mutation is PCD (primary cystic kidney) and laterality. The second Dnah5 splicing mutation is located in exon 44 in an AAA domain also known as ATPases associated with a variety of cellular activities. The phenotype involved with this mutation is laterality defects.

The first missense mutation is in exon 19 which can be found in the MBT (SM005561) and L27 (SM00569) domains. The MBT domain, malignant brain tumor, encodes a protein involved in transcriptional regulation. The L27 domain is involved in protein binding. The next missense mutation is located in exon 34 which can be found in the FYVE (SM00064) domain. This encodes a zinc finger protein, which binds two zinc ions. The next missense mutation is located in exon 48 which can be found in the AAA (SM00382) domain that codes for ATPases involved in a variety of cellular processes such as membrane fusion, proteolysis, and DNA replication. The phenotype associated with this mutation is PCD and laterality. The next missense mutation is located in exon 59, which is found in both the DHC-N1 and BRLZ domains. The DHC-N1 (PF08393) domain stands for dynein heavy chain, N-terminal region 1 and is known to mediate interactions with other heavy chains and intermediate-light chain complexes. The BRLZ (SM00338) domain, or basic region leucine zipper, regulates the basic-leucine zipper (bZIP) transcription factors (Hurst 1995). The phenotype involved with this mutation is DORV (double outlet right ventricle), AVSD (atrioventricular septal defect), and laterality defects.

The following missense mutations and splicing mutation are located in the dynein heavy domain. The protein encoded by this domain is involved in ATPase activity, microtubule binding, and acts as a dynein motor for the movement of organelles and vesicles along microtubules (SMART). More importantly, dynein is involved in cilia and flagella movement, which have been proven to be involved in cardiac development (Hornef 2006).

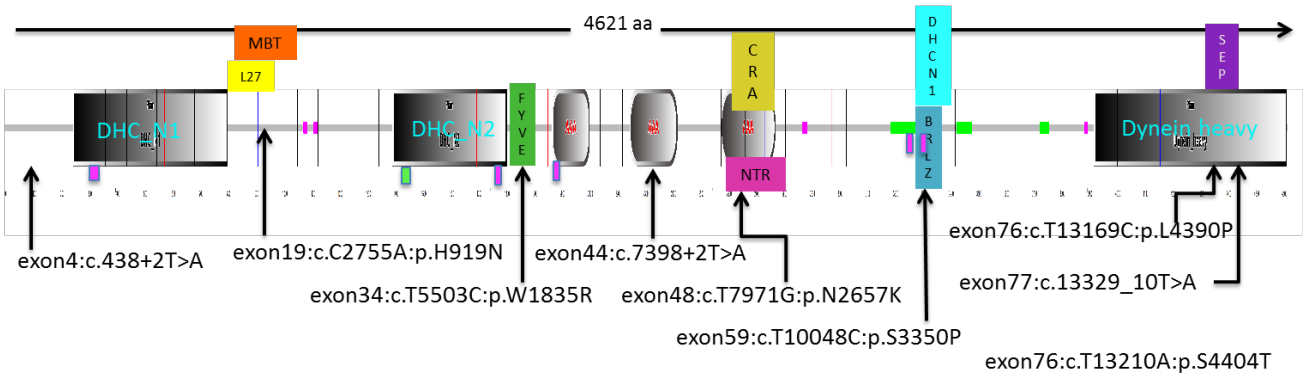

Figure 9. Dnah5 gene has 9 different mutations that are all in functional protein domains.

3.3.1.5 Dnah11—2 splicing mutations, 3 nonsense mutations, and 3 missense mutations

The first splicing mutation is located in exon 7 which can be found in the DHC-N1 domain. The second splicing mutation is in exon 40 which can be found in the AAA-5 domain. You can see a description of these two domains in the Dnah5 section above. The mice with these mutations have laterality defects.

The three nonsense mutations are all located in functional domains, two of which are in the same domain. The first nonsense mutation is in exon 16 which can be found in the Spc7 (SM00787) domain. This domain encodes a kinetochore protein and is required for kinetochore-spindle association (Kerres 2004). The mice with this mutation have laterality defects. The next two nonsense mutations are in exons 60 and 63, which can both be found in the MYSc (SM00242), myosin ATPases, domain. This domain encodes the myosin protein, which forms the thick filaments of the myofibril, which is involved in muscle contraction (Hayashida 1991). The phenotype for the first nonsense mutation is laterality and OFT (outflow tract). The phenotype for the second nonsense mutation is laterality and PCD.

There are three missense mutations, which are all found in different exons in different protein domains. The first missense mutation is located in exon 33 in the AAA domain. Please

see the explanation of the function of this domain in the Dnah5 section. The next missense mutation is found in exon 41 in the AAA-5 domain. This domain has the same function as the AAA domain. The third missense mutation can be found in exon 64, which is in the MYSc domain. This domain function is explained earlier in this section. All of these mutations are associated with laterality defects except the mutation in the MYSc domain also has PCD in addition to laterality.

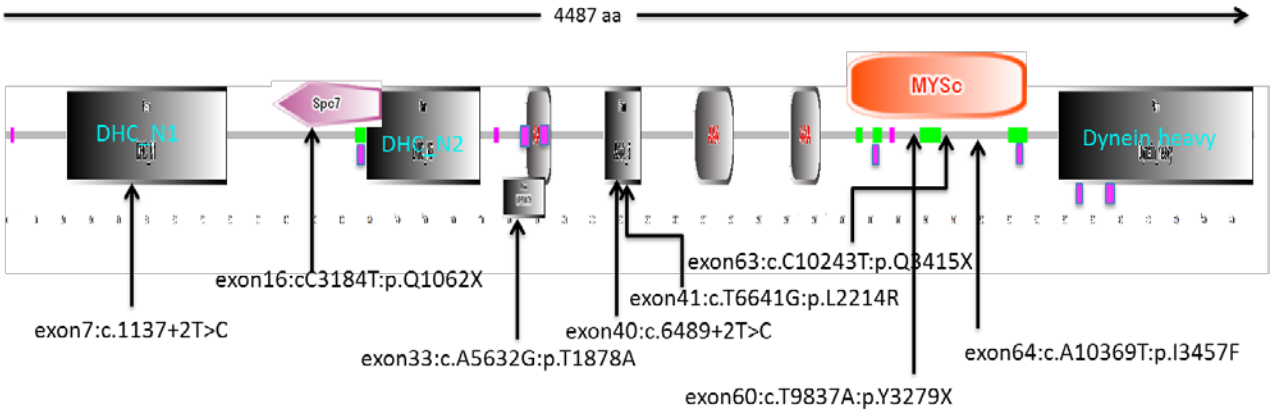

Figure 10. Dnah11 gene has 8 different mutations that are all located in functional domains. These domains include DHC_N1, Spc7, AAA, and MYSc.

3.3.1.6 Dnaic1—1 splicing mutation and 1 missense mutation

The Dnaic1 missense mutation is located in a WD40 (SM00320) domain, which is involved in protein binding. The phenotype for this mutation is laterality defects.

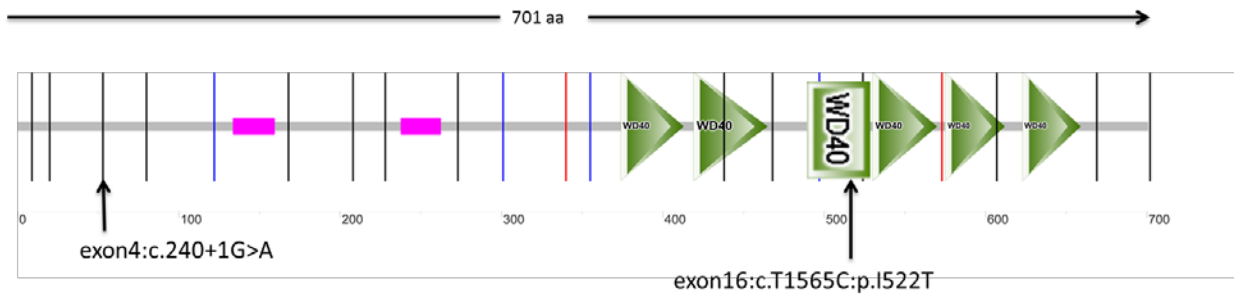

Figure 11. The Dnaic1 missense mutation is in a WD40 domain.

3.3.1.7 Gm1060—2 missense mutations

Both Gm1060 missense mutations are located in unknown protein domains. They are within four amino acids of one another in the same exon. This exon occurs just before a coiled coil domain so these mutations may affect that domain.

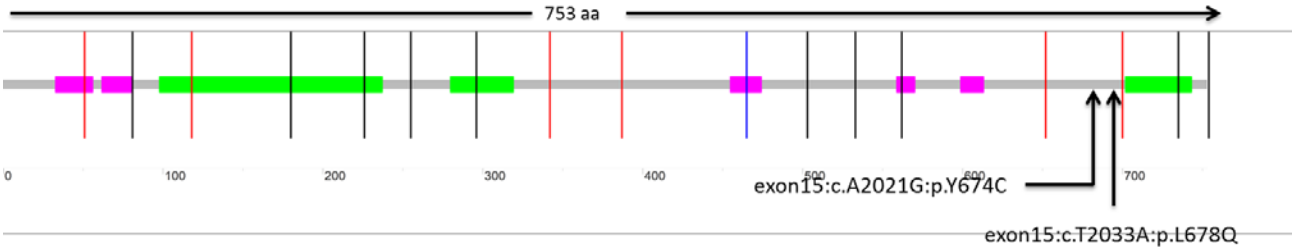

Figure 12. Neither of the Gm1060 missense mutations is in functional domains.

3.3.1.8 Megf8—2 missense mutations

Both of the Megf8 missense mutations are in EGF domains. The first mutation is in exon 3 which can be found in the first EGF (SM00181) domain for this gene. The EGF domain does not have a functional significance yet so it is difficult to determine how this causes the observed phenotype. The second mutation is in exon 21 which can be found EGF_Lam domain. This domain stands for laminin-type epidermal growth factor-like domain. Laminins are involved in mediating cell adhesion, growth migration, and differentiation (Beck 1990). These two mutations

however have different phenotypes. The mutation in exon 3 has several cardiac defects such as dextrocardia, AVSD, DORV, and SIT. The mutation in exon 21 only has laterality defects.

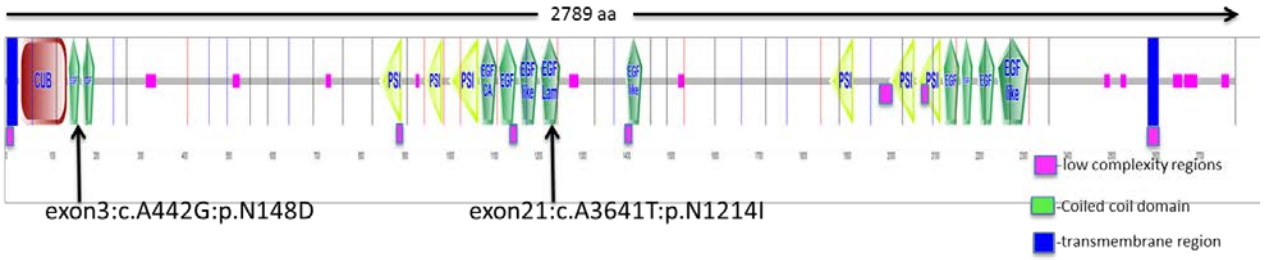

Figure 13. Both Megf8 missense mutations are in EGF domains.

3.3.1.9 Pcsk5—3 missense mutations

Two of the Pcsk5 missense mutations are in FU (SM00261) domains, although these domains are far apart across the gene. The FU domain stands for furin-like repeats which are involved in signal transduction by receptor tyrosine kinases (Raz 1991). The phenotype for the missense mutation in exon 22 is VSD (ventricular septal defect) and cystic kidney. The phenotype for the mutation in exon 35 is DORV, VSD, and aortic stenosis.

The other missense mutation occurs in exon 4 in the Peptidase_S8 domain. This domain is a part of the subtilase family, which is composed of serine proteases. The phenotype associated with this mutation is laterality defects, OFT, and cystic kidney.

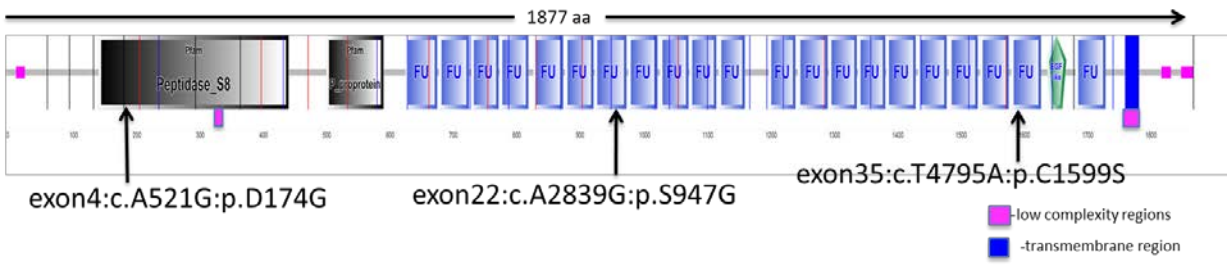

Figure 14. Two of the Pcsk5 missense mutations are in FU domains and the other is in a Peptidase_S8 domain.

3.3.1.10 Smarca4—2 missense mutations

The first missense mutation in Smarca4 at exon 8 is located in a low complexity region, which does not seem to have any function relevance to causing the phenotype. The second missense mutation at exon 16 is located in the DEXDc (SM00487) and SNF2_N domains. The DEXDc domain stands for DEAD-like helicases superfamily and encodes DNA and RNA helicases, which catalyze the separation of double-stranded nucleic acids (Caruthers 2002).

The SNF2_N domain is involved in several processes such as transcription regulation, DNA repair, DNA recombination, and chromatin unwinding among other minor processes (Pfam). The phenotype for this gene is outflow tract in both mutations.

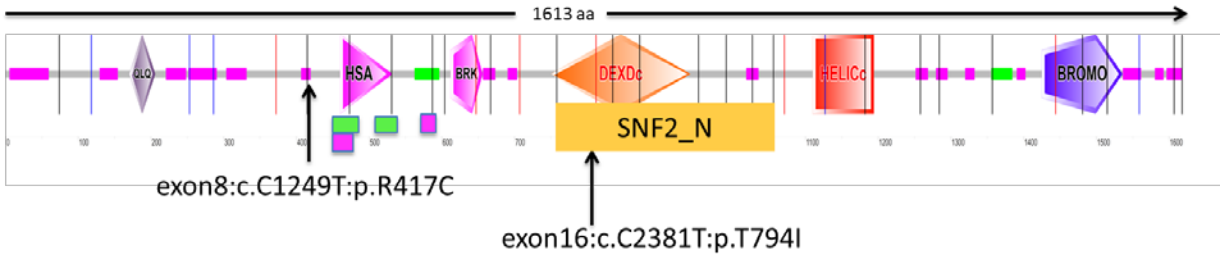

Figure 15. One Smarca4 mutation is located in two functional domains: SNF2_N and DEXDc.

3.3.1.11 Tmem67—1 splicing and 1 missense mutation

Both mutations in Tmem67 are located in the Meckelin (PF09773) domain; however the two mutations are fairly far apart because this domain is very large spanning almost the entire gene. The phenotype associated with the mutation in exon 11 is laterality and cystic kidney. The phenotype associated with the splicing mutation in exon 23 is laterality defects. The Meckelin domain is known to be associated with the basal ciliary body.

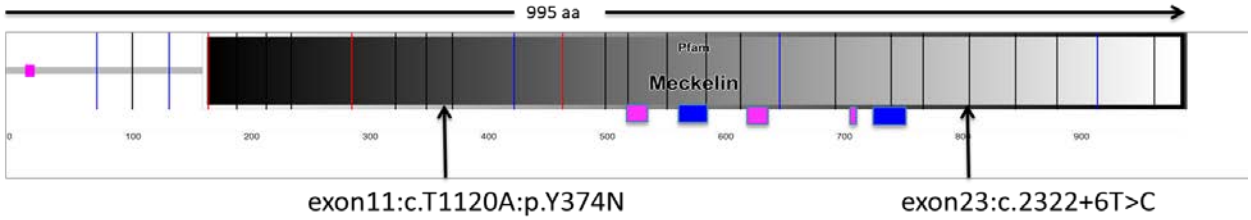

Figure 16. Both mutations in Tmem67 are in the Meckelin domain.

3.3.1.12 Wdr69—1 nonsense and 1 missense mutation

The nonsense mutation located in exon 2 is located in an unknown domain. The phenotype associated with this mutation is laterality defects and kidney tubule cysts. The missense mutation in exon 3 is found in a WD40 domain. The phenotype related to this mutation is laterality defects.

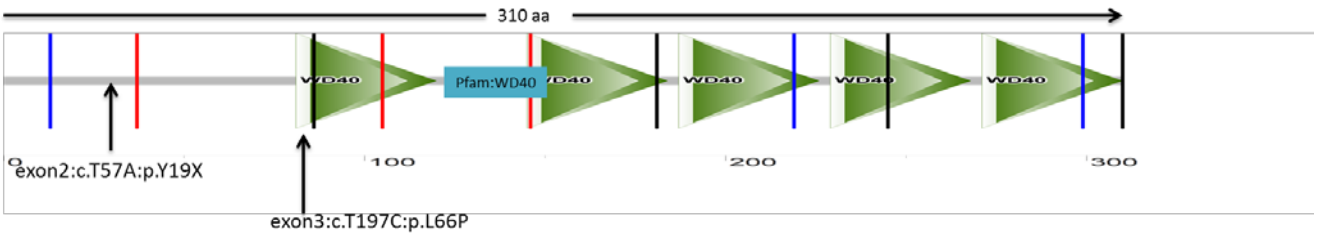

Figure 17. Wdr69 gene has a missense mutation in the WD40 domain and a nonsense mutation in an unknown domain.

3.3.2 Splicing Mutations

There are a total of 16 disease-causing splicing mutations that are described in *Table 75*. One quarter (25%) of these mutations are not within the two base pair canonical splicing sites. Almost all of the splicing mutations are at a conserved residue, with the exception of three mutations. The rest of the splicing mutations are in at least somewhat conserved regions as you can see in *Table 75*. There are four splicing mutations, which are not located in the canonical splice sites. The splicing mutations that are mostly conserved have a few changes in the wild type nucleotide at that position, but none that are the same as the mutation in our variant. Donor splicing mutations occur at the 5' end, or beginning, of the intron whereas acceptor splicing mutations occur at the 3' end, or the end, of the intron.

Table 75. Evolutionary conservation of splicing mutations in disease-causing genes.

Gene	Chr	Position	Nucleotide Change	Number of bases from exon	Splicing Site Affected	Conserved Residue
Dnaic1	4	41544973	c.240+1G>A	**1**	Donor	Conserved
Gm572	4	148017493	c.26+1G>A	**1**	Donor	Conserved
Bicc1	10	70420407	c.606+2T>C	**2**	Donor	Conserved
Ccdc39	3	33737998	c.357+2T>A	**2**	Donor	Conserved
Ccdc151	9	21799255	c.828+2T>C	**2**	Donor	Conserved
Cep290	10	99962906	c.1189+2T>C	**2**	Donor	Conserved
D630037F22Rik	10	55915160	c.1071+2T>A	**2**	Donor	Conserved
Dnah11	12	119425356	c.1137+2T>C	**2**	Donor	Conserved
Dnah11	12	119282224	c.6489+2T>C	**2**	Donor	Conserved
Dnah5	15	28159416	c.438+2T>A	**2**	Donor	Conserved
Dnah5	15	28275670	c.7398+2T>A	**2**	Donor	Conserved
Fuz	7	52152701	c.387+2T>A	**2**	Donor	Conserved
Lrp2	2	69316757	c.8456-3A>G	**3**	Acceptor	Mostly conserved
Tmem67	4	11974951	c.2322+6T>C	**6**	Donor	Mostly conserved
Dctn5	7	129281879	c.348+6T>C	**6**	Donor	Mostly conserved
Dnah5	15	28388293	c.13329-10T>A	**10**	Acceptor	Conserved

The following table shows three candidate genes that were recovered as splicing mutations in our screen. These three mutations are not conserved and would be considered tolerated under the conservation rules upon looking at the nucleotide change. These three mutations all have the nucleotide change as the wild type nucleotide in other species. In particular, the Kif15 mutation is present as wild type in all other species except the mouse. Upon further analysis, it was found that these mutations were not actually the cause for disease in the mice. This provides further

evidence that mutations, even in splicing regions, which are disease causing do not generally occur at a position where the change is present as the wild type.

Table 76. Nucleotides across the set of species for non-conserved splicing mutations.

Gene	Cypc2c44	Cep290	Kif15
Chromosome	19	10	9
Nucleotide change	T>**A**	C>**T**	C>**T**
Number of bases from exon	4	7	8
Mouse	T	C	C
Rat	T	**T**	-
Human	T	C	**T**
Orangutan	T	C	**T**
Dog	=	C	**T**
Horse	T	C	**T**
Opossum	**A**	A	**T**
Chicken	C	=	=
Stickleback		=	

Red lettering indicates a non-conserved residue
Bold means these nucleotides are the same as the nucleotide change observed.

The Cyp2c44 gene has a non-conserved mutation four base pairs from the exon located in an intronic region within the p450 protein domain. This mutation is present in two mice from line 386 who both have DORV, double outlet right ventricle. The nucleotide changes from a thymine to an adenosine in mice; however adenosine is the wild type nucleotide in the opossum. Since the opossum is lower on the species list, it is safe to assume that this mutation may still be disease causing.

The Cep290 gene has a mutation seven base pairs from the exon located in the intronic region of a coiled coil protein domain. This mutation is present in one mouse line with a

phenotype of hypertrophy, craniofacial defects, and laterality defects. The mutation changes the wild type cysteine nucleotide to a tyrosine. This mutation may be tolerated because the tyrosine is present as the wild type nucleotide in the rat.

The Kif15 gene has a mutation eight base pairs from the exon located in the intronic region before the kinesin protein domain. This mutation is present in two different mouse lines; one mouse has laterality defects and the other mouse has PTA and pulmonary atresia with MAPCA. The base pair changes from a cytosine to a thymine in our mouse model. This particular mutation is definitely not conserved because the nucleotide change, thymine, is the wild type nucleotide in all other species, not including the rat and stickleback which do not have nucleotide information at that position. Due to the fact that this mutation is present in two mouse lines with different phenotypes and the conservation information, it is very unlikely that this mutation is disease causing.

3.4 MOUSE AND HUMAN MUTATION CORRELATION

The following table (***Error! Reference source not found.***) gives a comparison of the recovered ENU mutations with published human mutations. We were able to recover several mutations in our screen that are in the same domain as the human mutations. Within that subsection, we also recovered a few mutations that are a few amino acids apart and have similar phenotypes. The genes in which these mutations occur include: Cc2d2a, Ccdc39, Smarca4, and Tmem67.

The Cc2d2a gene contains a homozygous nonsense mutation at position 949 (R949X) in the mouse and at position 950 (R950X) in humans. There are similar phenotypes in mice and humans; mice have mild laterality, cystic kidney, and cystic lung while humans have retinal

dystrophy, renal disease, and molar tooth sign from MRI. Both mice and humans have some form of renal defect. Unfortunately, these mutations occur in unknown domains within this gene.

The Ccdc39 gene has a homozygous splicing mutation in mice two base pairs from intron 4 while humans have a similar mutation one base pair from exon 4 that can be either homozygous or heterozygous. The phenotype in mice is laterality defects and in humans is Kartagener syndrome, SIT, and Ivemark syndrome, all of which are caused by laterality defects.

The Smarca4 gene has a heterozygous missense mutation at position 766 (V766G) in mice whereas humans have a missense mutation one amino acid earlier at position 764 (W764R) with unknown zygosity. Unfortunately the phenotype was not explained by Medina, so we cannot compare it with the mouse phenotype. However, this mutation occurs in the DEXDc protein domain which, as explained earlier, encodes helicases which catalyze the separation of double stranded nucleic acids.

The Tmem67 gene was found to have mutations in the Meckelin protein domain in both humans and mice. The homozygous mouse mutation is found in exon 11 at position 374 (Y374N) whereas the heterozygous human mutation is found in the amino acid before that at position 372 (T372K). The human mutation has a vast range of phenotypic anomalies with this mutation including molar tooth sign, retinal abnormalities, cystic kidney, elevated liver enzymes, and hepatosplenomegaly. The mouse phenotype simply consists of a cystic kidney, similarly to the human, and laterality defects.

Table 77. Disease-causing homozygous mutations recovered from ENU screen and published human mutations with similar phenotypes that are in the same protein domains.

Gene	Zygosity	Location	Nucleotide Change	Amino acid change	Domain	Mouse/Human Phenotype	Citation
ARMC4	-	-	T2780G	L927W	ARM	**PCD**	Hjeij
Armc4	H	Exon 20	T2978A	M993K	ARM	**Laterality**	-
CC2D2A	H	-	-	R950X	*unknown*	Retinal dystrophy, **renal disease**, molar tooth sign from MRI	Gordon
Cc2d2a	H	Exon 23	C2845T	R949X	*unknown*	Mild laterality/**cystic kidney** & lung	-
CCDC39	H/R	Intron 3	357+1G>C	-	intronic	Kartagener, SIT, Ivemark—**laterality defects**	Blanchon
Ccdc39	H	Exon 4	357+2T>A	-	intronic	**Laterality**	-
CEP290	R	Exon 36	A4723T	K1575X	CCD	LCA, mental retardation, molar tooth sign, muscular weakness	Perrault
Cep290	H	Exon 34	T4670A	L1557X	CCD	Mild laterality/ Cystic kidney	-
Dnah11	H	Exon 64	-	A3474T	*unknown*	-	Bartoloni
DNAH11	-	Exon 64	G10284A	G3429R	*unknown*	-	Pifferi
Dnah11	H	Exon 64	A10369T	I3457F	*unknown*	PCD/Laterality	-
DNAH11	R	Exon 60	T9764C	L3255S	*unknown*	**Situs solitus**	Knowles
Dnah11	H	Exon 63	C10243T	Q3415X	*unknown*	**PCD/Laterality**	-
DNAH5	R	Exon 48	A7888T	R2630W	AAA	-	Failly
Dnah5	H	Exon 48	T7971G	N2657K	AAA	PCD/Laterality	-
DNAH5	R	Exon 73	G12614T	G4205V	Dynein heavy	-	Failly & Hornef
Dnah5	H	Exon 76	T13169C	L4390P	Dynein heavy	PCD/Laterality	-
Dnah5	H	Exon 76	T13210A	S4404T	Dynein heavy	PCD/Laterality	-
DNAH5	H/R	Intron 76	13338+ 5G>A	-	intronic/ dynein heavy	-	Failly & Hornef
Dnah5	H	Exon 77	T13329A	-	intronic/ dynein heavy	PCD/Laterality	-
DNAI1	R	Exon 16	G1543A	G515S	*unknown*	**SIT**, respiratory complications	Guichard & Zietkiewicz
DNAI1	R	Exon 16	T1538C	L513P	*unknown*	-	Zietkiewicz
Dnaic1	H	Exon 16	T1565C	I522T	*unknown*	**Laterality**	-

Table 77. Continued

DYNC2H1	R	Exon 57	A9044G	D3015G	*unknown*	ATD; scoliosis, syndactyly	Dagoneau & Schmidts
Dync2h1	R	Exon 57	G9031T	V3011L	*unknown*	Unbalanced AVSD	-
DYNC2H1	H	Exon 6	C988T	R330C	DHC_N1	Scoliosis, syndactyly	Schmidts & Hokayem
Dync2h1	H	Exon 5	T701A	V234E	DHC_N1	Mild Laterality	-
FREM2	-	Exon 5	C5914A	E1972K	Calx-beta	**Syndactyly**, cryptophthalmos, urinary tract/ear/nose/lung abnormalities	Van Haelst
Frem2	H	Exon 11	A6875G	Y2292C	Calx-beta	Kidney agenesis/**syndactly/ polydactyly**	-
IFT140	R	-	A932G	Y311C	*unknown*	Mainzer-Saldino syndrome	Perrault
Ift140	H	Exon 9	A1138G	N380D	*unknown*	Laterality	-
MEGF8	H	-	G595C	G199R	EGF	craniosynostosis, prominent forehead, 5th finger clinodactyly, preaxial polydactyly, **dextrocardia** w/ SIT, bilateral cryptorchidism, wide-spread nipples	Twigg
Megf8	H	Exon 3	A442G	N148D	EGF	SIT, AVSD, **dextrocardia**, DORV	-
PRICKLE1	-	Exon 3	G361A	V121I	LIM	*Predicted as benign & tolerant	Bosoi
PRICKLE1	-	Exon 3	G370A	A124T	LIM	*Predicted as benign & tolerant	Bosoi
Prickle1	H	Exon 5	G482T	C161F	LIM	Velocardiofacial Syndrome	-
SMARCA4	-	Exon 16	T2290A	W764R	DEXDc	-	Medina
Smarca4	R	Exon 16	T2297G	V766G	DEXDc	DORV, VSD, cystic kidney	-
Smarca4	H	Exon 16	C2381T	T794I	DEXDc & SNF2_N	DORV, VSD	-

Table 77 Continued

TMEM67	H	Exon 10	T1046C	L349S	Meckelin	encephalomeningocele, bile duct proliferation, cleft palate, bile duct proliferation, tetralogy of Fallot *Predicted as possibly damaging	Iannicelli
TMEM67	R	Exon 11	C1115A	T372K	Meckelin	Molar tooth sign, chorioretinal colobomas, chronic renal insufficiency, **cystic kidney**, elevated liver enzymes, hepatosplenomegaly	Iannicelli
Tmem67	H	Exon 11	T1120A	Y374N	Meckelin	Laterality/**cystic kidney**	-

Please see Appendix A for full domain and phenotype information.
H-homozygous, R-heterozygous.
Yellow highlighted mutations are within 2 amino acids of one another.
Phenotypes in **bold** are the same in both the mouse and human for that gene.

4.0 DISCUSSION

As you can see in *Figure 4* and *Figure 5*, there is no saturation of disease-causing genes on any specific chromosome. However, we did not recover any genes the sex chromosomes. This makes sense because any genes on only one of the sex chromosomes would have resulted in a lethal phenotype in males; this is not the case with our mutations.

As for the mutation that is considered tolerated, I have no hypothesis as to why this is able to cause disease despite the fact that the mutated amino acid is present as a wild type in other species. Since it is a missense mutation, we cannot argue that there is an introduction of a stop codon that will truncate the protein. The mutation is found in two lesser known protein domains: MBT and L27. The MBT, malignant brain tumor, domain is involved in transcriptional repression and chromatin modification. The L27 domain is found in receptor targeting proteins Lin-2 and Lin-7 as well as some protein kinases.

5.0 CONCLUSIONS

We conducted a forward genetic screen using ENU-induced mice in order to find the majority of disease-causing genes for congenital heart defects. We were able to recover a total of 88 different mutations within 62 disease-causing genes that contribute to both cardiac and non-cardiac mutant phenotypes from nearly 150 mouse lines.

Out of the 1,338 homozygous variants we recovered 96 (7.2%) disease-causing variants from sixty-five genes. Of these variants, 59 were missense, 16 were splicing, and 13 were nonsense. There were 8 variants that are the same, either occurring in the same mouse line but different mice or in different mouse lines. This gives us a final count of 88 different disease-causing variants in 62 genes.

There are 21 mutations from our screen that were found to be similar to previously published human mutations. The mutations are similar in the sense that they occur in the same protein domain, but not necessarily the same exon. These mutations also present similar phenotypes in both mice and humans. Three of these mutations are incredibly similar to published human mutations due to the fact that they are within one or two amino acids or nucleotides from each other.

All in all, we can conclude that the 88 disease-causing mutations recovered from our genetic screen are mostly damaging. Three-fourths (75%) of the mutations were completely damaging. About 17% were possibly damaging, 2% were possibly tolerated, and 1% was

tolerated. The remaining 5% are splicing mutations that were more than two base pairs from the exon and therefore did not have a definitive conservation call. Overall, there is only one mutation in our screen that was tolerated.

This information proves that mutations which are disease causing generally occur in highly conserved amino acids across a set of species. It has been previously theorized that there is biological significance in having mutations with

APPENDIX A: FULL DOMAIN NAMES FROM TABLE 1

7tm_1—7 transmembrane receptor (rhodopsin family)

AAA— ATPase family associated with various cellular activities

AAA_5—AAA family-5

Acetyltransf_1—acetyltransferase (GNAT) family

Adaptin_N—adaptin N terminal region

ANK—ankyrin repeat

ARM—armadillo/beta-catenin-like repeat

B41—band 4.1 homologues

Calx-beta—Calx-beta domain

CCD—coiled coil domain

DEXDc—DEAD-like helicases superfamily

DHC_N1—dynein heavy chain, N-terminal region 1

DUF3585—domain of unknown function 3585

Dynein heavy—dynein heavy chain and region D6 of dynein motor

EGF—epidermal growth factor-like

FH—forkhead

FR47—FR47-like protein

FU—furin-like repeats

HDc—metal dependent phosphohydrolases with conserved ‘HD’ motif

HECTc—domain homologous to E6-AP Carboxyl terminus

IGc2—immunoglobulin C-2 type

KH—K homology RNA-binding domain

KISc—kinesin motor, catalytic domain, ATPasw

LDLa—low-density lipoprotein receptor domain class A

LIM—zinc-ion binding domain present in Lin-11, Isl-1, Mec-3

LINK—LINK (hyaluronan binding)

Lysl_oxidase—

Meckelin—Meckelin (transmembrane protein 67)

P450—cytochrome p450

Pep_M12B_propep—reprolysin family propeptide

Peptidase_M24— metallopeptidase family M24

Peptidase_S8—subtilase family

Plexin_cytopl—plexin cytoplasmic RasGAP domain

PP2Cc—serine/threonine phosphatases, family 2C, catalystic domain

Sema—semaphorin domain

SNF2_N—SNF2 family N-terminal domain

S_TKc—serine/threonine protein kinases, catalytic domain

SUFU—suppressor of fused protein

Tryp_SPc—trypsin-like serine protease

ZnMc—zinc-dependent metalloprotease

APPENDIX B: PHENOTYPE ACRONYMS

ASD—atrial septal defect

ATD—asphyxiating thoracic dystrophy

AVSD—atrioventricular septal defect

DORV—double outlet right ventricle

D-TGA—dextro-transposition of the great arteries

HLHS—hypoplastic left heart syndrome

HRHS—hypoplastic right heart syndrome

IAA—interrupted aortic arch

LCA—Leber congenital amaurosis

MAPCA—major aortopulmonary collateral artery

OFT—outflow tract

PCD—primary ciliary dysfunction

PKD—primary cystic kidney disease

PTA—percutaneous transluminal angioplasty

RAA—right aortic arch

SIT—situs inversus totalis

TGA—transposition of the great arteries

VSD—ventricular septal defect

BIBLIOGRAPHY

Bartoloni, *et al.* Mutations in the DNAH11 (Axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. *PNAS* 99(16), 10282-10286 (2002).

Beck, Hunter, Engel. Structure and function of laminin: anatomy of a multidomain glycoprotein. *FASEB J* 4(2), 148-160 (1990).

Benson. Advances in cardiovascular genetics and embryology: role of transcription factors in congenital heart disease. *Curr Opin Pediatr* 12, 497-500 (2000).

Benson, Sharkey, Fatkin, *et al*. Reduced penetrance, variable expressivity, and genetic heterogeneity of familial atrial septal defects. *Circulation* 97, 2043-2048 (1998).

Betts, Russell, and Barnes. Amino acid properties and consequences of subsitutions. *In* Bioinformatics for Geneticists, I.C. Gray eds, Wiley (2003).

Blanchon, *et al.* Delineation of CCDC39/CCDC40 mutation spectrum and associated phenotypes in primary ciliary dyskinesia. *J Med Genet* 49, 410-416 (2012).

Bolsover *et al.* When a different amino acid will do: conservative mutations. *In* Cell Biology, A short course, 3rd ed. Wiley (2011).

Bosoi, Capra, Allache, *et al*. Identification and characterization of novel rare mutations in the planar cell polarity gene PRICKLE1 in human neural tube defects. *Hum Mutat* 32(12), 1371-1375 (2011).

Botto and Correa. Decreasing the burden of congenital heart anomalies: an epidemiologic evaluation of risk factors and survival. *Prog in Ped Cardio* 18, 111-121 (2003).

Caruthers and McKay. Helicase structure and mechanism. *Curr Opin Struct Biol* 12(1), 123-133 (2002).

Clancy. RNA splicing: introns, exons and spliceosome. *Nature Ed* 1(1), 31 (2002).

Dagoneau, Goulet, Genevieve, *et al.* DYNC2H1 mutations cause asphyxiating thoracic dystrophy and short rib-polydactyly syndrome, type III. *Am J Hum Genet* 84, 706-711 (2009).

Failly, *et al.* Mutations in DNAH5 account for only 15% of a non-preselected cohort of patients with primary ciliary dyskinesia. *J Med Genet* 46, 281-286 (2009).

Garcia-Mayoral, Hollingworth, Masino, *et al*. The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degredation. *Structure* 15(4), 485-498 (2007).

Goh and Choi. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. *Genomics Inform* 10(4), 214-219 (2012).

Gorden, *et al.* CC2D2A is mutated in Joubert syndrome and interacts with the ciliopathy-associated basal body protein CEP290. *Am J Hum Genet* 83, 559-571 (2008).

Guichard, *et al.* Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome). *Am J Hum Genet* 68, 1030-1035 (2001).

Hayashida, Maita, and Matsuda. The primary structure of skeletal muscle myosin heavy chain: Sequence of the amino-terminal 23 kDa fragment. *J Biochem* 110(1), 54-59 (1991).

Hjeij, *et al.* ARMC4 mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry. *Am J Hum Genet* 93, 357-367 (2013).

Hornef, *et al.* DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arms defects. *Am J Resp* 174, 120-126 (2006).

Hokayem, Huber, Couve, *et al.* NEK1 and DYNC2H1 are both involved in short rib polydactyly Majewski type but not in Beemer Langer cases. *J Med Genet* 49, 227-233 (2012).

Hurst. Transcription factors 1: bZIP proteins. *Protein Profile* 2(2), 101-168 (1995).

Iannicelli, Brancati, Mougou-Zerelli, *et al.* Novel TMEM67 mutations and genotype-phenotype correlates in Meckelin-related ciliopathies. *Hum Mutat Brief* 31, E1319-E1331 (2010).

Jenkins, Correa, Feinstein, *et al*. Noninherited risk factors and congenital cardiovascular defects: current knowledge—A scientific statement from the American Heart Association Council on cardiovascular disease in the young. *Circulation* 115, 2995-3014 (2007).

Kerres, Vietmeier-Decker, Ortiz, *et al*. The fission yeast kinetochore component Spc7 associates with the EB1 family member Mal2 and is required for kinetochore-spindle association. *Mol Biol Cell* 15(12), 5255-5267 (2004).

Knowles, *et al.* Mutations of DNAH11 in primary ciliary dyskinesia patients with normal ciliary ultrastructure. *Thorax* 67(5), 433-441 (2012).

McElhinney, Krantz, Bason, *et al*. Analysis of cardiovascular phenotype and genotype-phenotype correlation in individuals with a JAG1 mutation and/or Alagille syndrome. *Circulation* 160, 2567-2574 (2002).

Medina, Romero, Kohno, *et al.* Frequent BRG1/SMARCA4-inactivating mutations in human lung cancer cell lines. *Hum Mutat* 29(5), 617-622 (2008).

Mitchell, Sander, Klinker, and Tomita-Mitchell. The molecular basis of congenital heart disease. *Semin Thorac Cardiovasc Surg* 19, 228-237 (2007).

Momma, Kondo, Ando, *et al*. Tetralogy of Fallot associated with chromosome 22q11 deletion. *Am J Cardiol* 76, 618-621 (1995).

Ng, Buckingham, Lee, *et al*. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 42(1), 30-35 (2010).

Perrault, *et al.* Spectrum of NPHP6/CEP290 mutations in Leber congential amaurosis and delineation of the associated phenotype. *Hum Mutat* 956, 1-10 (2007).

Perrault, Saunier, Hanein, *et al.* Manizer-Saldino syndrome is a ciliopathy caused by IFT140 mutations. *Am J Hum Genet* 90, 864-870 (2012).

Pierpont, Basson, Benson, *et al.* Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* 115(23), 3015-38 (2007).

Pifferi, *et al.* New DNAH11 mutations in primary ciliary dyskinesia with normal Axonemal ultrastructure. *Euro Resp J* 35(6), 1413-1416 (2010).

Raz, Schejter, and Shilo. Interallelic complementation among DER/flb alleles: implications for the mechanism of signal transduction by receptor-tyrosine kinases. *Genetics* 129(1), 191-201 (1991).

Schmidts, Arts, Bongers, *et al.* Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome) without major polydactyly, renal or retinal involvement. *J Med Genet* 50, 309-323 (2013).

Twigg, Lloyd, Jenkins, *et al.* Mutations in multidomain protein MEGF8 identify a Carpenter syndrome subtype associated with defective lateralization. *Am J Hum Genet* 91, 897-905 (2012).

Van Haelst, Maiburg, Baujat, *et al*. Molecular study of 33 families with Fraser syndrome: New data and mutation review. *Am J Med Genet* 146A, 2252-2257 (2008).

Zietkiewicz, Nitka, *et al.* Population specificity of the DNAI1 gene mutation spectrum in primary ciliary dyskinesia (PCD). *Resp Research* 11(174), 1-11 (2010).