

A Multi-omic Approach to Genetic Hearing Loss in the Newfoundland Founder Population

By:

© Justin Pater

May, 2019

A thesis submitted to the School of Graduate Studies (SGS) in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Medicine, Discipline of Genetics, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada

Abstract

Hearing loss is the most common sensory disorder and despite the identification of over 143 genes, many loci remain unsolved. Therefore, the identification of novel genes are of great importance to provide insight into disease pathways and improve the diagnosis and management of hearing loss. Due to a limited gene pool, genetic isolates, such as the island of Newfoundland, provide unprecedented opportunities for gene discovery. The objective of this thesis was to identify the genetic basis of hearing loss in several large Newfoundland families with either autosomal recessive or dominant hearing loss. Firstly, we identified a pathognomonic *deafness, autosomal recessive 29 (DFNB29; OMIM: 614035)* phenotype that was caused by a novel pathogenic *CLDN14* missense variant, which resided on a 1.4 Mb ancestral haplotype across four families. Even though *DFNB29* is associated with a highly variable, congenital phenotype, we observe cases of prelingual hearing loss that progresses to a distinct audioprofile. Subsequently, we identified a linked region (13q34; LOD: 4.77) within a large autosomal dominant hearing loss family that led to the discovery of a pathogenic splicing variant in a nascent *ATP11A* exon, which activates a cryptic splice site 153 bp downstream of the canonical splice site. This linked region overlaps with the *DFNA33* locus, and hearing loss due to *ATP11A* exhibits significant variable expressivity, which is consistent with the family from Germany used to map this locus. Unexpectedly, three families were found to have Usher syndrome, caused by homozygous or compound heterozygous *USH2A* splicing variants that co-segregated in two families that were initially ascertained as non-syndromic retinitis pigmentosa, while the remaining non-syndromic hearing loss family was positive for a novel pathogenic *ADGRV1* nonsense variant. Given that these families were reassigned to an Usher syndrome

diagnosis based on genetic testing, this highlights the importance of employing next-generation sequencing in the clinical setting. In summary, this thesis identified a novel autosomal dominant hearing loss gene, and reclassified four variants of unknown significance to pathogenic variants, ascertaining the genetic etiology of deafness within eight families in the Newfoundland genetic isolate. These discoveries accelerate the diagnosis and surveillance of those at-risk of developing hearing loss, as well as enrolment into promising clinical trials.

Acknowledgements

Where to start? So many people have helped me along the way in my PhD journey. This chapter of my life been quite exciting. There have been ups and downs along the way, but I wouldn't have it any other way. Most notably, I'd like to thank:

Dr. Terry-Lynn Young: Words can't describe my gratitude for my experience in the Young laboratory. Although I came from a functional genetic background in a completely different field of study, you took a chance on me. I will be forever grateful for the lessons I've learnt under your mentorship - am I 100% geneticist now?

Thanks to my supervisory committee, Drs. Darren O'Rielly, Bridget Fernandez, and Curtis French for their guidance, helpful advice and opportunities to grow as a scientist.

Thanks to the Young Laboratory for being supportive and incredible colleagues! A special thanks is in order to Jim Houston and Dante Galutira for your advice, patience and assistance you've provided over the last 6 years, as well as the use of your garbage can. Thank you Tammy Benteau, Cindy Penney, Elspeth Drinkell, Catherine Street, Geoff Woodland and Salem Werdyani for all your help and support – you've all made my PhD experience fun and exciting!

Anne Griffin and Sue Stanton, the field of audiology is massive! You've made my learning about clinical audiology so enjoyable and fun – thank you!

Dr. Kathy Hodgkinson, thanks for the many chats over the years, both from a personal and scientific perspective. You're insight into human genetics and the epidemiology of disease is invaluable!

Last, but not the least, my family: Megan, Mom, Dad, Vanessa, Blayne Josh, Jacqui, Mike Mitch, Trynda, Tyson, Marvin, Linda, Matt and Mojo. From celebrating milestones to listening to me when times are tough, each one of you has played a key role in my success and I wouldn't be where I am with you.

Dedication

“People are weird. When we find someone with weirdness that is compatible with ours, we team up and call it love”

- Dr. Seuss

This thesis is dedicated to my wife, Megan. Without your unconditional love, guidance and patience, none of this would be possible.

Table of Contents

Abstract	i
Acknowledgements	iii
Dedication	v
Table of Contents	vi
List of Tables	xi
List of Figures	xii
List of Abbreviations and Symbols	xvi
List of Appendices	xxii
Chapter 1: General Introduction	1
1.1 The Human Genome and Genomic Variation	1
1.1.1 Features of the Human Genome	1
1.1.2 An Overview of RNA Splicing	2
1.1.3 Cryptic and Alternative Splicing and Disease.....	9
1.2 Hereditary Hearing Loss and Its Syndromes	16
1.2.1 The Auditory System and Hearing Loss	16
1.2.2 The Genetics of Hearing Loss	24
1.2.3 The Newfoundland Founder Population and Hearing Loss Research.....	30
1.2.4 Next-Generation Sequencing Technologies	40
1.3 Summary and Study Goals	42
1.4 Bibliography	43
Chapter 2: A common variant in <i>CLDN14</i> causes precipitous, prelingual sensorineural hearing loss in multiple families due to founder effect	63

2.1	Co-author Statement	64
2.2	Copyright and License information:.....	64
2.3	Abstract.....	65
2.4	Introduction	66
2.5	Materials and Methods.....	68
2.5.1	DNA Preparation, Targeted Sequencing and Audioprofiling	71
2.5.2	Whole-Exome Sequencing and Variant Filtration.....	72
2.5.3	Cascade Sequencing and Haplotype Analysis.....	73
2.6	Results	74
2.6.1	Clinical Evaluation	74
2.6.2	Targeted Sequencing and Audioprofiling	74
2.6.3	Whole-Exome Sequencing.....	75
2.6.4	Cascade Sequencing and Haplotype Analysis.....	84
2.6.5	Genealogical Analysis	89
2.7	Discussion.....	91
2.8	Summary.....	96
2.9	Limitations	96
2.10	Acknowledgements	97
2.11	Bibliography	98

Chapter 3: A Pathogenic Splicing Variant In a Nascent <i>ATP11A</i> Exon Maps To <i>DFNA33</i> and Documents Its First Association With a Penetrant Mendelian Phenotype	105
3.1 Co-author Statement	106
3.2 Abstract.....	107
3.3 Introduction	108
3.4 Materials and Methods.....	109
3.4.1 Study participants and clinical evaluations.....	109
3.4.2 Targeted screening and lineage analysis	112
3.4.3 Whole exome sequencing (WES) and variant filtration.....	112
3.4.4 Comprehensive audioprofiling.....	113
3.4.5 Whole genome sequencing and variant filtration.....	114
3.4.6 Cascade sequencing, segregation and haplotype analysis	114
3.4.7 Experimental validation of splicing variants.....	115
3.5 Results.....	116
3.5.1 Hearing loss maps to 13q34 overlapping with the <i>DFNA33</i> locus	116
3.5.2 Whole exome sequencing fails to identify the genetic basis of hearing loss	116
3.5.3 Two candidate genetic variants co-segregate with dominant hearing loss.....	117
3.5.4 RNA analysis reveals multiple <i>ATP11A</i> transcripts.....	122
3.5.5 A pathogenic <i>ATP11A</i> splicing variant activates a cryptic donor splice site.....	125
3.6 Discussion.....	130
3.7 Acknowledgements	136

3.8	Bibliography	137
Chapter 4: Novel Usher syndrome pathogenic variants identified in cases with hearing and vision loss.....143		
4.1	Co-authorship Statement	144
4.2	Abstract.....	145
4.3	Introduction	146
4.4	Materials and Methods.....	147
4.4.1	Study Participants and Clinical Evaluations	147
4.4.2	Gene Panels	150
4.4.3	Linkage Analysis and Whole Exome Sequencing in Hearing Loss Family R2100.....	154
4.4.4	Splice variant <i>in silico</i> analysis	155
4.4.5	RNA-cDNA analysis	155
4.5	Results.....	155
4.5.1	<i>ADGRV1</i> c.17062C>T Genotype/Phenotype Analyses.....	155
4.5.2	<i>USH2A</i> c.5777-1G>A and c.10388-2A>G Genotype/Phenotype Analyses.....	161
4.5.3	<i>USH2A</i> c.5777-1G>A and c.10388-2A>G Experimental Validation of splicing effects...	168
4.6	Discussion.....	172
4.7	Conclusions	174
4.8	Acknowledgements	175

4.9	Bibliography	176
	Chapter 5: General Discussion and Conclusions	181
5.1	Future Directions.....	191
5.2	Bibliography	193
	Appendix.....	202

List of Tables

Table 2.1. Thirty-one Heterozygous variants identified in PID V-9, VI-4 and V-17 (Family R2010; Figure 2).....	77
Table 2.2. Four Homozygous variants identified in PID V-9, VI-4 and V-17 (Family R2010; Figure 2).....	78
Table 3.1. Fifty-one rare variants (<1% MAF; minimum of 30X coverage) that were present in four affected family R2070 members and absent in two unaffected members.....	118
Table 3.2. <i>In silico</i> predictions for <i>ATP11A</i> , chr13:113534962G>A.....	124
Table 4.1 R2100 Multipoint Linkage Analysis.....	156
Table 4.2. Whole exome sequencing variants that are shared between PID V-2 and V-3.....	158

List of Figures

Figure 1.1. Simplified Mechanism of RNA Splicing	4
Figure 1.2. Three protein isoforms produced by alternative RNA splicing.....	5
Figure 1.3. Possible alternative splicing patterns and mechanisms	6
Figure 1.4. The Regulation and Mechanisms of Alternative Splicing	8
Figure 1.5. The Mechanisms of Cryptic Splicing	10
Figure 1.6. Mechanisms of miRNA Biogenesis.....	14
Figure 1.7. Schematic of the auditory pathway	17
Figure 1.8. Anatomy and physiology of the cochlea.....	19
Figure 1.9. Tonotopic map of the cochlea	20
Figure 1.10. Schematic of hair cell bundles and mechanotransduction	22
Figure 1.11. Audiogram demonstrating hearing loss severity	25
Figure 1.12. Various audiograms illustrating the types of sensorineural hearing loss	26
Figure 1.13. Prevalence of genetic hearing loss according to mode of inheritance.....	29
Figure 1.14. Colonization of the Newfoundland genetic isolate.....	33
Figure 2.1. Rare, precipitous audiologic phenotype caused by <i>CLDN14</i> (c.488C>T; p.Ala163Val) in an Irish clan.....	69
Figure 2.2. Combined pedigrees of 3 families (R2033, R2075 and R2010) with rare, precipitous audiologic phenotype that connect to a common ancestor.....	70
Figure 2.3. <i>TMCI</i> c.421C>T Segregation Analysis in Family R2010	76
Figure 2.4. Sequence electropherogram of <i>CLDN14</i> (c.488C>T; p.Ala163Val). Red box highlights variant site	79
Figure 2.5. Location of pathogenic variants in Claudin-14	80

Figure 2.6. Conservation of the Claudin-14 protein using Clustal Omega and WebLogo display	81
Figure 2.7. Schematic diagram demonstrating the molecular structure of tight junctions.....	83
Figure 2.8. Cross-sectional diagram illustrating the anatomical location of the cochlear canals and their respective ionic composition.....	85
Figure 2.9. Combined pedigrees of 3 families (R2033, R2075 and R2010) with rare, precipitous audiologic phenotype connect to a founding couple and share an ancestral DFNB29-associated haplotype	86
Figure 2.10. Wild-type/normal <i>CLDN14</i> c.488C.....	87
Figure 2.11. Pedigree of family R2072, identified in screening of the NL deafness cohort, with the rare, precipitous audiologic phenotype who also share the <i>CLDN14</i> (c.488C>T; p.Ala163Val) allele and ancestral <i>DFNB29</i> -associated haplotype	88
Figure 2.12. Clan members lacking the recessive <i>CLDN14</i> [c.488C>T; p.(Ala163Val)] variant do not present with the characteristic steeply sloping hearing phenotype, exhibiting a different age of onset and hearing threshold progression.....	90
Figure 3.1. Family R2070 pedigree with a variable <i>DFNA33</i> hearing loss phenotype.....	110
Figure 3.2. Family R2070 Phenotype.....	111
Figure 3.3. Sequence electropherogram illustrating the <i>ATP11A</i> c.*11G>A substitution.....	120
Figure 3.4. Sequence electropherogram <i>COL4A1</i> c.3326-7dupT variant.....	121
Figure 3.5. <i>COL4A1</i> c.3326-7dupT RNA Analysis.....	123
Figure 3.6. <i>ATP11A</i> RT-PCR. Amplification in affected R2070 family members (PID III-1, III- 5, and III-7) and wild-types controls	123

Figure 3.7 Colony PCR amplification of TA-cloned RT-PCR amplification in affected R2070 family members (PID III-1, III-5, and III-7) and wild-types controls	123
Figure 3.8. Wild-type and mutant <i>ATP11A-201</i> sequence traces.....	123
Figure 3.9. Wild-type and mutant <i>ATP11A-202/212</i> sequence traces.....	123
Figure 3.10. Proposed <i>ATP11A</i> disease mechanism.....	123
Figure 4.1. Family R2100 Pedigree	148
Figure 4.2. Family R2100 Audiological Data.....	149
Figure 4.3. Family R4110 Pedigree.....	151
Figure 4.4. Family R4110 and R0723 Audiological Data.	152
Figure 4.5. Family R0723 Pedigree	153
Figure 4.6. Sequence electropherogram of <i>ADGRV1</i> c.17062C>T (p.Arg5688Ter.....	159
Figure 4.7. Family R2100 Pedigree with <i>ADGRV1</i> c.17063C>T.....	160
Figure 4.8. Retinal photograph of PID V-3 at age 27 (Family R2100).....	162
Figure 4.9. Central 24-2 visual threshold test of PID V-3 at age 27 (R2100) illustrating a deterioration of peripheral visual acuity.....	163
Figure 4.10. Sequence electropherogram of <i>USH2A</i> c.5777-1G>A, a pathogenic splicing variant that resides in intron 29.....	164
Figure 4.11. Sequence electropherogram of <i>USH2A</i> c.10388-2A>G, a pathogenic splicing variant that resides in intron 52.....	165
Figure 4.12. Family R4110 Pedigree.....	166
Figure 4.13. Family R0723 Pedigree.....	167
Figure 4.14. Family R0723 Ophthalmology Imaging.	169
Figure 4.15. <i>USH2A</i> c.5777-1G>A RNA analysis.	170

Figure 4.16. *USH2A* c.10388-2A>G RNA analysis.171

List of Abbreviations and Symbols

A	adenosine
A site	aminoacyl site
ACMG	American College of Medical Genetics and Genomics
ADGRV1	adhesion g-protein-coupled receptor-v1
ADPRHL1	ADP-Ribosylhydrolase Like 1
AGO2	Argonaut-2
ARVD5	arrhythmogenic right ventricular dysplasia type 5
ATP11A	ATPase Phospholipid Transporting 11A
ATP11C	ATPase Phospholipid Transporting 11C
ATP8A2	ATPase Phospholipid Transporting 8 A2
ATP8B1	ATPase Phospholipid Transporting 8 B2
BBS	Bardet-Biedl syndrome
bp	base pair
BRCA1	BRCA1, DNA repair associated
BRCA2	BRCA2, DNA repair associated
C	cytosine
Cas9	CRISPR associated protein 9
CDH23	cadherin-23
cDNA	Complementary DNA
chr	chromosome
<i>CLDN14</i> (gene)	Cluadin-14

CLDN14	Cluadin-14
cM	CentiMorgan
CNV	Copy number variation
COCH	cochlin
COL11A1	collagen type XI alpha 1 chain
COL11A2	collagen type XI alpha 2 chain
COL4A1	collagen type IV alpha 1 chain
CRISPR	Clustered regularly interspaced short palindromic repeats
dB	decibel
DFN	deafness
DFNA	autosomal dominant deafness
DFNB	autosomal recessive deafness
DFNX	X-linked deafness
DGCR8	DiGeorge syndrome critical region 8
DIAPH1	protein diaphanous homolog 1
DNA	deoxyribonucleic acid
E site	exit site
EDTA	ethylenediaminetetraacetic acid
ENCODE	encyclopedia of DNA elements
ENT	ear, nose and throat
ERG	electroretinogram
ESE	exonic splicing enhancers

ESS	exonic splicing silencers
F10	coagulation Factor X
G	guanine
GATK	genome analysis toolkit
GJB2	connexin-26
GJB3	connexin-31
GJB6	connexin-30
hnRNP	heterogenous nuclear ribonuclear proteins
Hz	hertz
INDEL	insertion/deletion variant
IRS2	insulin receptor substrate 2
ISE	intronic splicing enhancers
ISS	intronic splicing silencers
KCNQ4	potassium voltage-gated channel subfamily Q member 4
kHz	kilohertz
LOD	logarithm of the odds
MAF	minor allele frequency
MARVELD2	MARVEL Domain Containing 2
Mb	megabase
MIM	OMIM identifier
miRNA/miR	micro RNA
mL	milli litre

mRNA	messenger RNA
MT-RNR1	mitochondrially encoded 12S RNA
MYH9	myosin heavy chain 9
MYO6	myosin-6
MYO7A	myosin-7A
n	number
NGS	next-generation sequencing
NL	Newfoundland (island portion)
OMIM	online Mendelian inheritance in man
OTOA	otoancorin
P	peptidyl site
P4-ATPase	phospholipid flippase
PCDH15	protocadherin-15
PCR	polymerase chain reaction
PID	pedigree identification number
pol	polymerase
POU4F3	POU class 4 homeobox 3
pre-miRNA	premature-miRNA
pri-miRNA	primary miRNA
PS	phosphatidylserine
R	family
RISC	RNA-induce silencing complex

RNA	ribonucleic acid
RP	retinitis pigmentosa
RT-PCR	reverse-transcription PCR
SIFT	sorting Intolerant From Tolerant
SLC26A4	solute carrier family, member 4
SMPX	small muscle protein X-Linked
SNP	single nucleotide polymorphism
snRNP/U	small nuclear ribonucleoprotein
SNV	single nucleotide variant
SOX1	sry-box 1
SR	Ser/Arg-rich protein
STRC	stereocilin
Sxl	sex lethal
T	thymine
TECTA	alpha-tectorin
TMC1	transmembrane channel-like protein 1
TMC2	transmembrane channel-like protein 2
TMEM43	transmembrane protein 43
TMPRSS3	transmembrane serine protease 3
Tra	transformer
tRNA	transfer RNA
TTC8	tetratricopeptide Repeat Domain 8

U (nucleotide)	uracil
U2AF	U2 auxiliary factor
UCSC	University of California, Santa Cruz
uL	micro litre
USH	Usher syndrome
USH1B	Usher syndrome type 1 B
USH2A	usherin
USH2A (disease)	Usher syndrome type 2 A
USH2C	Usher syndrome type 2 C
UTR	untranslated region
v	version
VLGR1	very large G protein coupled receptor 1
VUS	variant of unknown significance
WES	whole exome sequencing
WFS1	wolframin
WHRN	Whirlin
WGS	whole genome sequence
YAC	yeast artificial chromosome

List of Appendices

Appendix A – ACMG Standards and guidelines for the interpretation of sequence variants ..	202
Appendix B – Loci and Genes of Hereditary Hearing Loss and its Syndromes	206
Appendix C – Recurrent Hearing Loss Variants in Newfoundland before this dissertation	214
Appendix D – Optimized TA-cloning procedure for the Young laboratory	215
Appendix E – Family R2070 Linkage Analysis Report Produced by SickKids.....	217
Appendix F – Whole exome sequencing identified 10 rare variants (<1% MAF) identified in family R2070	227
Appendix G – Whole exome sequencing identified 7 common variants (<10% MAF) identified in family R2070	228
Appendix H – Whole exome sequencing variant segregation testing	229
Appendix I – Family R2070 Audioprofiling.	241
Appendix J –Audioprofiling Report	289
Appendix K – Family R100 Linkage Analysis Report Produced by SickKids.....	298

Chapter 1: General Introduction

1.1 The Human Genome and Genomic Variation

Our understanding of genetics and the human genome has undergone many transformations in recent history. In a very short period of time, the field has discovered the structure of DNA, described the Central Dogma of Molecular Biology, mapped the entire human genome, and identified millions of variants across the genome. This rapid development of genomics has led to unprecedented improvements in the diagnosis, surveillance, and treatment of many human conditions. Moreover, the advent of advanced next-generation sequencing (NGS) technologies has broadened our understanding of molecular biology at both the DNA and RNA levels.

1.1.1 Features of the Human Genome

The human genome is approximately 3 billion nucleotides in length; however, mapping the human genome is not entirely complete (Green *et al.*, 2015). Several multi-megabase (Mb) sized gaps within the telomeres and centromeres of most chromosomes are challenging to sequence and assemble due to the repetitive nature of these regions and extreme sequence homogeneity (Miga, 2015). The initial draft of the Human Genome Project suggested that there were 30,000 – 40,000 protein coding genes (Lander *et al.*, 2001). However, more recent investigations suggest as little as 19,000 genes, representing 1 – 2% of the entire genome (Ezkurdia *et al.*, 2014). The other 99% of the genome is non-coding, and despite being initially referred to as “junk DNA” (S. Ohno, 1972), massive efforts such as the Encyclopedia of DNA Elements (ENCODE) project have described hundreds of thousands of functional regions that have significant roles in regulating gene expression (ENCODE, 2012). For

example, gene expression can be regulated through the epigenetic modification of non-coding regulatory regions, such as the methylation of distal gene enhancers and repressors, as well as CpG islands within promoter regions(Blattler & Farnham, 2013). Moreover, DNA methylation(Lev Maor *et al.*, 2015), non-coding RNA genes(Romero-Barrios *et al.*, 2018) and retrotransposons(Belancio *et al.*, 2006) have also been shown to regulate key processes, such as alternative RNA splicing.

1.1.2 An Overview of RNA Splicing

Even though the number of coding genes is much lower than once thought, more recent RNA studies have identified approximately 205,000 protein-coding transcripts and this number continues to grow(Hu *et al.*, 2015). Our understanding of the human genome is relatively complete; however, the same cannot be said at the RNA or whole transcriptome level. Alternative splicing is a tightly regulated process that increases protein biodiversity in an estimated 95% of all multiexon genes(Pan *et al.*, 2008). This crucial mechanism produces many different mRNA transcripts with different combinations of exons, enabling cell potency and the terminal differentiation of stem cells that are committed to a particular function in the body. With the exception of the testis and brain, isoform-specific gene expression explains 45% and 85% of the variance between individuals and tissues, respectively, and fewer than 200 genes are tissue-specific(Mele *et al.*, 2015).

Immediately after transcription, immature mRNA is spliced through a highly regulated and complex process that removes intronic sequences and ligates exons together, producing a mature mRNA (reviewed in Tom Strachan, 2010). This process is mediated by the spliceosome, a multimeric complex comprised of small nuclear ribonucleoprotein (snRNP; U) subunits. Each

snRNP contains a small nuclear RNA (snRNA) and proteins. Consensus sequences are required at the exon-intron boundaries for an intron to be removed by the spliceosome. Among these requirements are a G at the 5' end of the exon and GU at the 3' end of the intron, which denotes a donor site (also referred to as the 5' splice site). The consensus sequence marking an acceptor site (or 3' splice site) is comprised of an AG at the 3' end of the intron, as well as a G at the 5' end of the exon. A branch site that contains an adenosine toward the middle of the intron is also required for the removal of introns (Figure 1.1a). The first step of RNA splicing is the binding of the U1 snRNP to the donor splice site, followed by binding of the U2 snRNP to the branch site (Figure 1.1b). Next, the trimer of U4, U5 and U6 snRNPs binds in the intron region, completing spliceosome assembly (Figure 1.1c). Subsequently, the donor site is cut and the 5' end of the intron is ligated to the adenosine in the branch site to form a lariat structure (Figure 1.1d). At this time, the U1 and U4 snRNPs disassociate (Figure 1.1d) from the spliceosome and the U6 and U5 snRNPs shift positions within the intronic sequence (Figure 1.1e). Finally, the U5 snRNP cuts the acceptor splice site and ligates the exons together – the lariat intron and remaining components of the spliceosome are then released (Figure 1.1f).

The post-transcriptional regulation of immature mRNAs is quite complex. In addition to 3' polyadenylation and 5'-capping, immature mRNA molecules can be spliced in a variety of ways through alternative splicing, producing different mature mRNA that encode for different protein isoforms (Grabowski & Black, 2001; Kornblihtt *et al.*, 2013; Figure 1.2). Most exons in genes can be included or excluded in the mature mRNA, or they can become shortened or lengthened by altering the position of constitutive exons 5' or 3' to splice sites (Figure 1.3; Kornblihtt *et al.*, 2013; Lev-Maor *et al.*, 2007). Additionally, the entire length a transcript

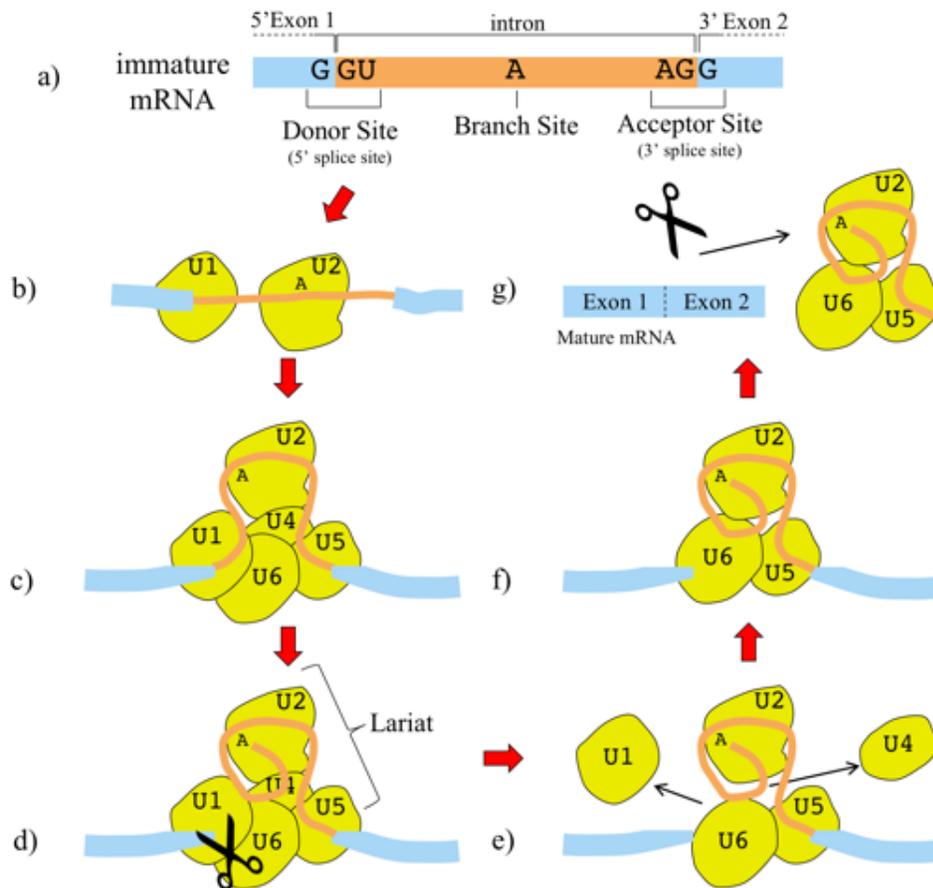


Figure 1.1. Simplified Mechanism of RNA Splicing. a) immature single-stranded mRNA that contains the highly conserved 5' GU and 3' AG motifs required for RNA splicing and producing a mature mRNA molecule. b) The U1 and U2 snRNPs bind to the 5' splice site and branch site, respectively, c) The U5-6 snRNPs are recruited to the spliceosome complex, d) the 5' intron is cleaved, producing a lariat structure, e) the U1 and U4 snRNPs dissociate, f) U2, U5 and U6 reposition to cleave the 3' end of the intron g). ssRNA: single-stranded RNA, G: Guanine, U: Uracil, A: Adenosine, U1-U6: snSRP 1-6. This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license, adapted from(Contributors11)

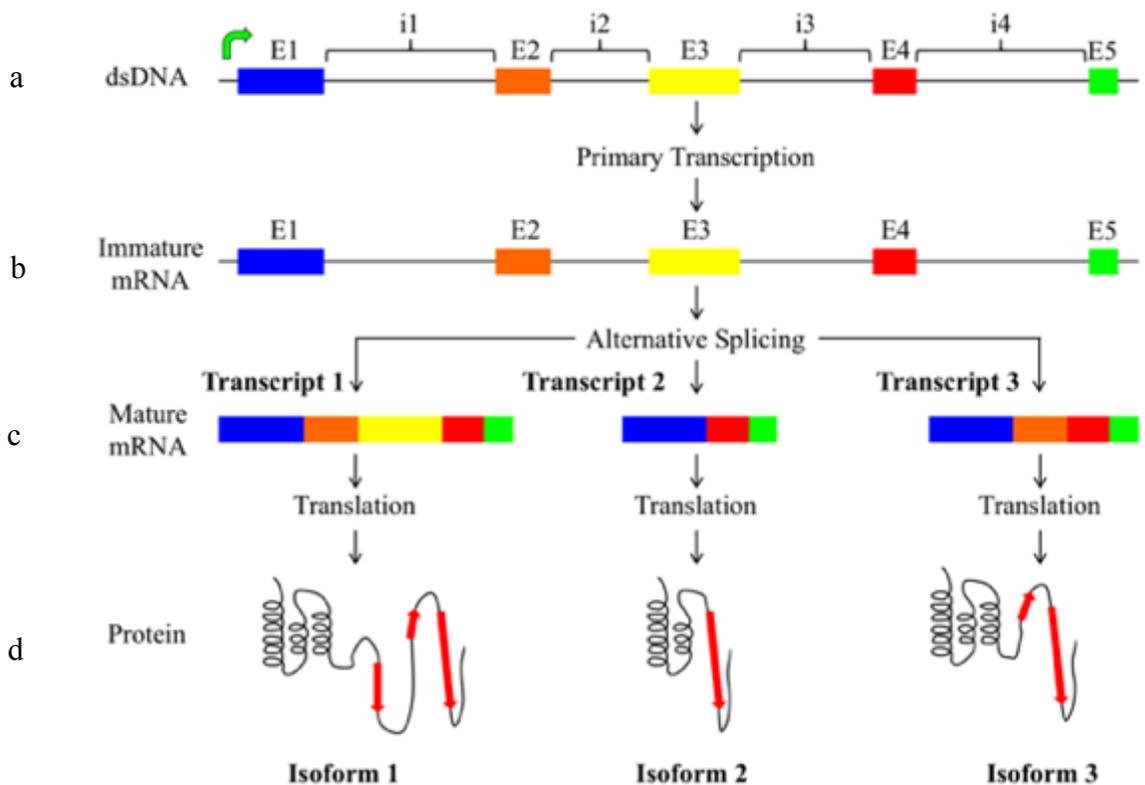


Figure 1.2. Three protein isoforms produced by alternative RNA splicing. a) Full-length dsDNA gene containing all possible coding exons, b) Immature mRNA molecule, c) Three different alternatively spliced transcripts that d) translate into three different protein isoforms. E: exon, i: intron, dsDNA: double-stranded DNA, ssRNA: single-stranded RNA, mRNA: messenger RNA. This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license, adapted from(Contributors1)

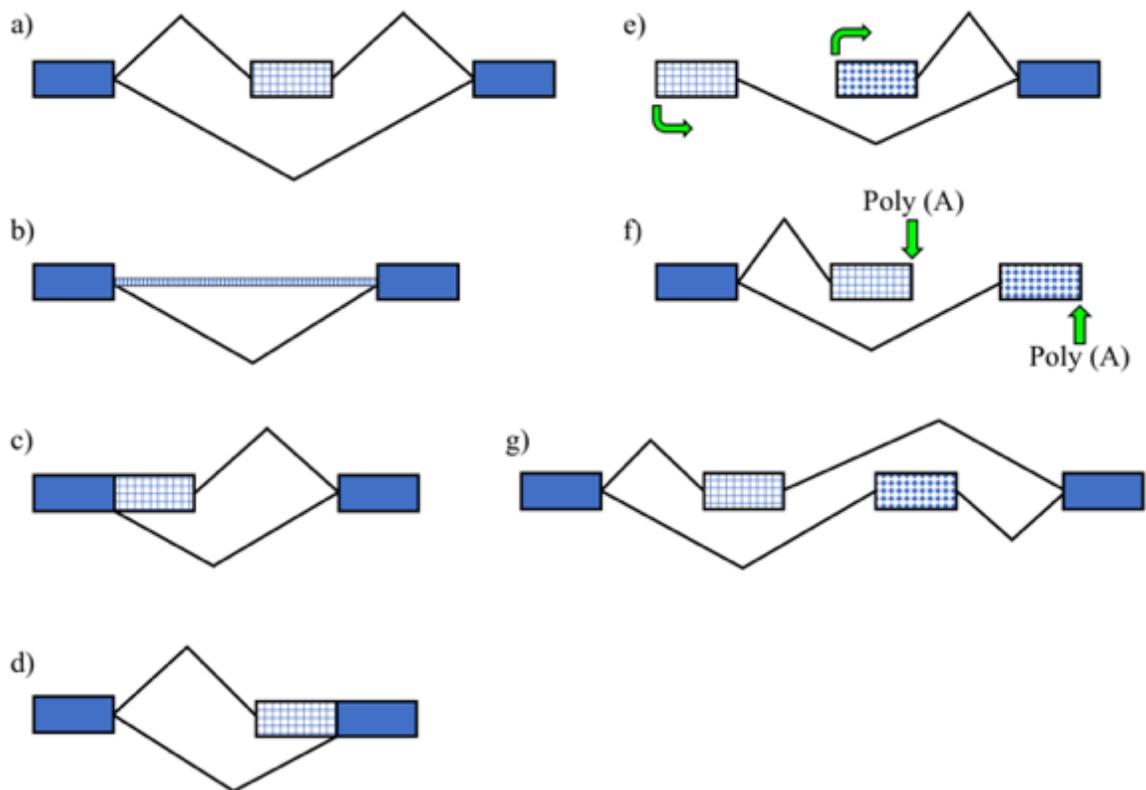


Figure 1.3. Possible alternative splicing patterns and mechanisms. Blue boxes represent constitutive exons present in all final mRNAs. Patterned boxes represent cassette exons that may or may not be included in final transcript, depending on the cell or tissue type in question. a) An exon can be either included or skipped, b) Introns can be retained or excluded from the mature mRNA, c/d) Alternative 5' donor or 3' acceptor splice sites and shorten or lengthen any given exon, e) Alternative promoters and f) Alternative poly(A) sites can lengthen or shorten the size of a transcript, g) Adjacent exons can be mutually exclusive such that only one exon can be included in the final transcript. This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license, adapted from(Contributors12)

can be changed through the use of alternative 3' terminal exons that alter polyadenylation sites, or alternative 5' exons that switch to a different internal promoter within a given gene(Kornblihtt *et al.*, 2013). Finally, some isoforms retain their introns and are considered as processed non-coding transcripts that are degraded by nonsense mediated decay, which do not produce a functional protein(Kim *et al.*, 2007; Sugnet *et al.*, 2004). These mechanisms serve a specific purpose for specific cells and tissues, and changing splice sites can drastically affect an encoded protein(Burke *et al.*, 1992). For example, more minute changes can significantly impact protein localization, enzymatic activity, or increase or decrease ligand affinity for a receptor. Conversely, alternative splicing that excludes several exons can result in the lack of an entire functional domain in a protein(Burke *et al.*, 1992; C. W. Smith & Nadal-Ginard, 1989).

Mechanisms that regulate alternative splicing are largely controlled by specific *cis*-regulatory sequence motifs that are divided into four main categories: intronic splicing silencers (ISSs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs) and exonic splicing enhancers (ESEs). These *cis*-regulatory sequence motifs regulate alternative splicing through their interaction with two large families of *trans*-acting proteins: Ser/Arg-rich proteins (SRs) and heterogenous nuclear ribonuclear proteins (hnRNPs; Kornblihtt *et al.*, 2013). This interaction targets critical components of the spliceosome complex by influencing the affinity of snRNPs at exon-intron boundaries; specifically, hnRNPs bind to ISEs and ISS and SRs interact with ESEs and ISEs to inhibit and promote RNA splicing, respectively(Kornblihtt *et al.*, 2013; Figure 1.4). Together, these mechanisms highlight processes that drive diversity of terminally differentiated cells in living systems, given that gene expression profiles, including genes that encode *trans*-acting SRs and hnRNPs are vastly different from cell to cell

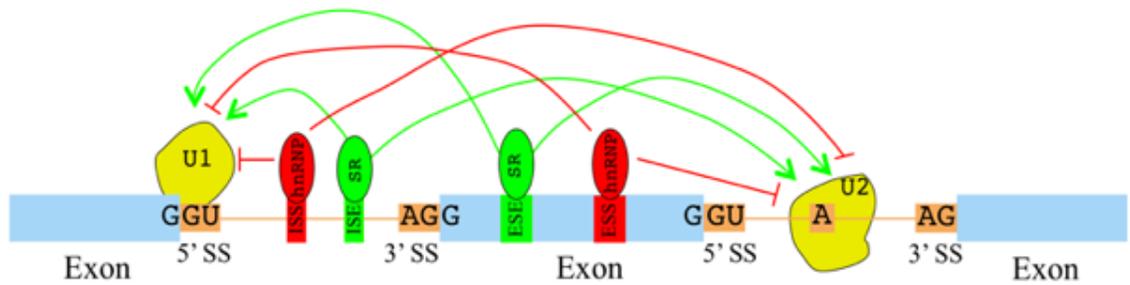


Figure 1.4. The Regulation and Mechanisms of Alternative Splicing. Many DNA elements that have proteins binding partners greatly influence which transcript is expressed in a cell. ESEs and ISEs interact with SRs to promote splicing. In contrast, hnRNPs bind to ISSs and ESSs to inhibit splicing. These *cis*-regulatory motifs and proteins regulate alternative splicing, promote protein biodiversity in a tissues, and permit terminal differentiation of cells from potent stem cells. Green arrow indicates a positive effect. Red, blunt ended lines indicate an inhibitory effect, ISS: intronic splicing silencers, ESS: exonic splicing silencers, ISE: intronic splicing enhancers ESE: exonic splicing enhancers. b-e) Black lines indicate splicing locations, SS: splice site, G: Guanine, U: Uracil, A: Adenosine, U1-U2: snSRP 1-2, SR: Ser/Arg-rich protein; shnRNPs: heterogenous nuclear ribonuclear proteins. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Z. Wang & Burge, 2008)

(Zhang *et al.*, 2016). These basic principles apply to all eukaryotic systems – as cells commit to a specific lineage, gene expression and the number of protein isoforms increasingly become more diverse(Black, 2000).

1.1.3 Cryptic and Alternative Splicing and Disease

Genomic variants that result in aberrant RNA splicing can have significant impacts on human health. The removal of intronic sequences from an immature RNA molecule is a tightly regulated process that is coordinated by highly conserved nucleotides at the exon-intron boundaries, as well as by recognition of these sequences by the spliceosome and additional auxiliary splicing factor proteins(Ward & Cooper, 2010). Therefore, variation within exon-intron boundaries is generally not tolerated and is for the most part, pathogenic. The first two and last two nucleotides of an intron, which are GU and AG, respectively, are of particular importance for spliceosome recognition (Figure 1.1; Chen & Manley, 2009; K. Ohno *et al.*, 2018). Even though this “GU-AG rule of splicing” is important, variation in the last 2 nucleotides of an exon and the first 6 nucleotides of an intron have a high likelihood of causing cryptic alternative splicing(Chen & Manley, 2009; K. Ohno *et al.*, 2018). In addition to exon skipping and intron inclusion, such variants that reside within exon-intron boundaries can also activate a cryptic donor or acceptor splice site (Figure 1.5).

Many state-of-the-art *in silico* algorithms, such as MaxEnt(Yeo & Burge, 2004), NNSPPLICE(Reese *et al.*, 1997), SpliceSiteFinder(Desmet *et al.*, 2009), and GeneSplicer(Pertea *et al.*, 2001) help predict the pathogenicity of candidate splicing variants. While these *in silico* predictions have many advantages in that they are sensitive for detecting the creation or destruction of splice sites(Leong *et al.*, 2015), they lack the ability to

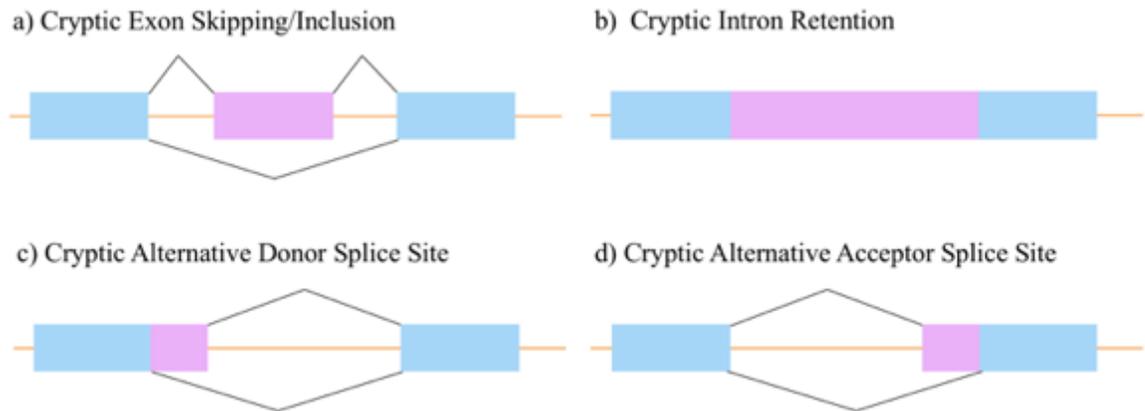


Figure 1.5. The Mechanisms of Cryptic Splicing. Variants that reside within exon-intron boundaries are likely to cause either: a) Cryptic exon skipping or inclusion, b) Cryptic intron retention, or they can activate a cryptic alternative donor splice site (c), or acceptor splice site (d). Blue boxes represent constitutive exons and pink boxes indicate cryptically spliced regions that are included in processed, mature mRNA. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors2)

consistently and accurately predict the specific splicing abnormality(Ernst *et al.*, 2018). For example, a recent study(Baert *et al.*, 2018) evaluated the performance of many *in silico* prediction tools and found that computational approaches were sensitive in detecting cryptic splicing effects in 21 *BRCA1* and *BRCA2* variants, but were unreliable in determining the specific splicing effect. Given these limitations, experimental validation of *in silico* predictions are essential to the interpretation of candidate splicing variants, and this is reflected in clinical variant interpretation guidelines, such as those published by the American College of Medical Genetics and Genomics (ACMG; Appendix A; Richards *et al.*, 2015) and EuroGentest(Matthijs *et al.*, 2016) . The pathogenicity of any given splicing variant is highly dependent on the sequence that is proximal to the exon-intron boundary. In the context of activating a cryptic donor or acceptor splice site, there is a 1/3 probability that this will create an in-frame insertion or deletion (INDEL). In light of the lack of specificity of *in silico* analyses, it is conceivable that candidate splicing variants that are predicted to cause exon skipping could experimentally validated as small, tolerated, in-frame insertions and deletions. This highlights the importance of studying candidate splicing variants at the RNA level.

Not only does alternative splicing facilitate the terminal differentiation of tissue-specific somatic cells, this process is also implicated in pleiotropy, or the ability of a single gene to influence two or more unrelated phenotypic traits(Drivas *et al.*, 2015) A recent survey of the human transcriptome identified that less than 200 genes are tissue specific(Mele *et al.*, 2015). Given that there are in excess of 200,000 protein coding transcripts(Hu *et al.*, 2015), genetic variants that reside in alternatively spliced, tissue-specific transcripts, could explain why different pathogenic variants in a single gene can cause both non-syndromic and syndromic Mendelian phenotypes(Dominguez *et al.*, 2016). An excellent example that portrays the effects

of alternative splicing on pleiotropy is *Tetratricopeptide Repeat Domain 8* (*TTC8*; MIM: 608132) that is within the *Bardet-Biedl syndrome 8 locus* (*BBS8*; MIM: 615985). Pathogenic biallelic variants in *TTC8* (NM_198309) cause Bardet-Biedl syndrome, an autosomal recessive condition characterized by overt clinical features, including developmental delay, obesity, postaxial polydactyly, hypogonadism and retinitis pigmentosa (RP; Ansley *et al.*, 2003). However, Riazuddin *et al* (2010) identified a pathogenic splicing variant in a retinal-specific *TTC8* transcript (NM_144596) that causes non-syndromic *retinitis pigmentosa 51* (MIM: 613464). Another such example of pleiotropy is the case of hereditary hearing loss and its syndromes. Until recently, pathogenic variants in *collagen type XI alpha 1 chain* (*COL11A1*), a gene that encodes an extracellular matrix protein, have been linked to connective tissue disorders that have hearing loss as a feature, such as Marshall syndrome (MIM: 154780; Annunen *et al.*, 1999), Stickler syndrome type II (MIM: 604841; Majava *et al.*, 2007; Rose *et al.*, 2005), and fibrochondrogenesis (MIM: 228520; Tompson *et al.*, 2010). Then Booth *et al* (2018) identified a pathogenic *COL11A1* variant that causes exon skipping in an inner ear-specific splicing transcript that maps to the *deafness, autosomal dominant 37 hearing loss locus* (*DFNA37*). In addition to having significant roles in the regulation of alternative splicing (Stower, 2013), non-coding regions of the genome also influence gene expression through microRNA (miRNA/miR)-mediated mechanisms (Gebert & MacRae, 2018).

Even though every cell type in an individual human body contains a nearly identical copy of every gene across the genome, the regulation of gene expression at specific developmental time points permits terminal differentiation from multipotent to tissue-specific somatic cells. While gene expression can be regulated by many mechanisms, such as the

epigenetic modification of histones and CpG methylation, miRNA-mediated gene silencing exhibits a profound role during normal development (Gebert & MacRae, 2018). miRNAs are promiscuous molecules that can inactivate hundreds of mRNAs, regulating the spatiotemporal expression of genes (Gebert & MacRae, 2018). Like coding genes, miRNAs are transcribed by RNA polymerase II (Figure 1.6a); however, miRNAs can be derived from gene introns or from long non-coding RNAs (Berezikov, 2011; Lau *et al.*, 2001). Upon transcription, miRNA biogenesis begins with the formation of primary miRNAs (pri-miRNAs), that contain a hairpin loop (Figure 1.6b; Gebert & MacRae, 2018). Before export to the cytoplasm, the RNase III enzyme, Drosha, and the DiGeorge syndrome critical region 8 (DGCR8) proteins complex with and cleave the pri-miRNA into a premature-miRNA (pre-miRNA) that is approximately 70 nucleotides in length (Figure 1.6c; Nguyen *et al.*, 2015; Nicholson, 2014). Subsequently, the pre-miRNA is carried out of the nucleus through a nuclear pore by the nuclear transporter, exportin-5 (Figure 1.6d; Okada *et al.*, 2009). In the cytoplasm the pre-miRNA is recognized by a large RNase III protein called Dicer, which cleaves the hairpin loop, yielding a shorter (20-22 nucleotide) double stranded mature miRNA (Figure 1.6e; Nicholson, 2014). Next, Argonaut-2 (AGO2) interacts with Dicer to bind the pre-miRNA (Figure 1.6f). After unwinding and releasing the “passenger strand” (Figure 1.6g), the “guide strand” and AGO2 recruit several proteins to form the RNA-induced silencing complex (RISC; Figure 1.6h; Gebert & MacRae, 2018). Thereafter, the RISC complex hybridizes and cleaves binding sites within the 3' UTR of genes that are complementary to the seed sequence of the “guide strand” (Figure 1.6i; Bartel, 2009), promoting translation repression and target mRNA degradation (Figure 1.6j). Thus, silencing the production of a functional protein (Jonas & Izaurralde, 2015). Like any other gene, the expression of miRNAs is a tightly regulated process (Gebert & MacRae, 2019). For

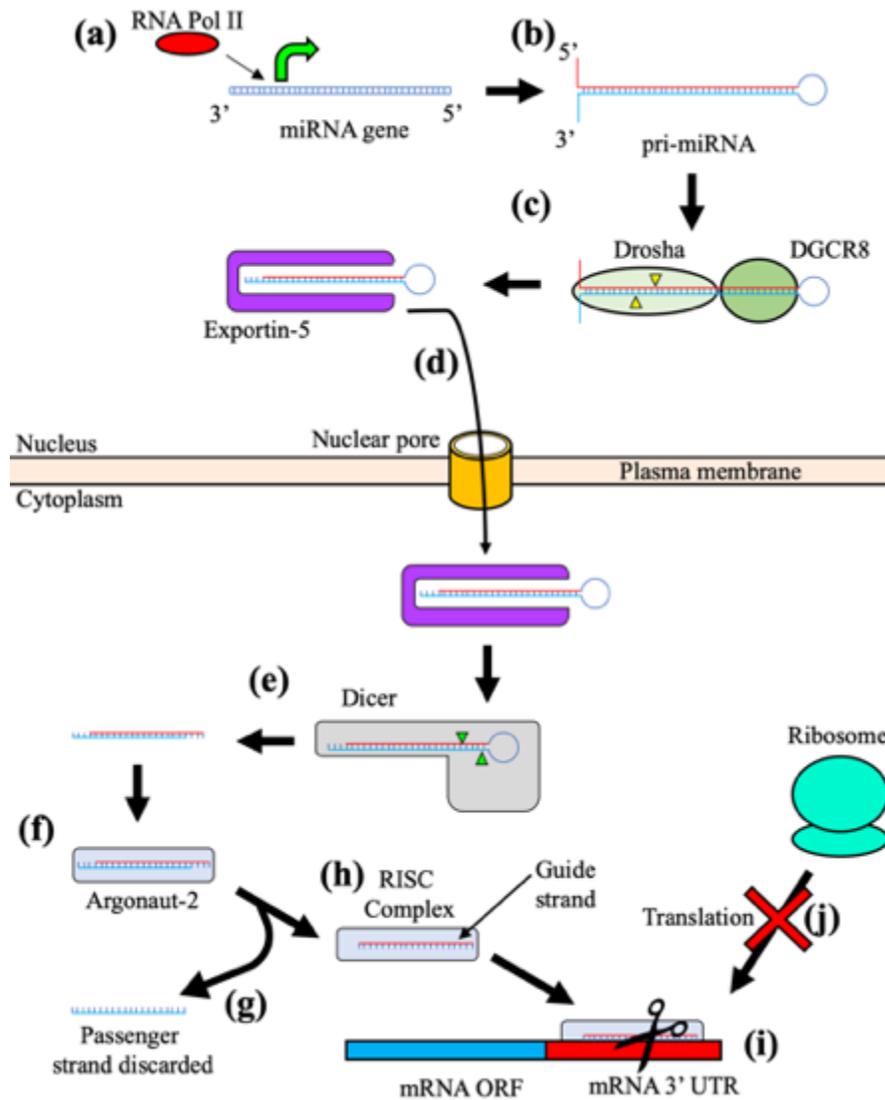


Figure 1.6. Mechanisms of miRNA Biogenesis. a) miRNA gene is transcribed by RNA Pol II into a pri-miRNA (b) that is cleaved by Drosha and DGCR8 into a ~70 nucleotide pre-miRNA (c). The pre-miRNA is carried out the nucleus through a nuclear pore by exportin-5 (d), and is cleaved by Dicer, yielding a shorter (20-22 nucleotide) double stranded mature miRNA (e). The mature miRNA is loaded onto Argonaute-2 (f), which releases the passenger strand (g), and recruits accessory proteins to form the RISC complex (h). The RISC complex hybridizes and cleaves binding sites within the 3' UTR of mRNAs that are complementary to the seed sequence of the guide strand (i), promoting translation repression and target mRNA degradation (j). miRNA: MicroRNA, Pol: Polymerase, DGCR8: DiGeorge syndrome critical region 8, pri-miRNA: primary miRNA, pre-miRNA: premature miRNA, mRNA: messenger RNA, RISC: RNA-induce silencing complex, UTR: untranslated region, ORF: open-reading frame. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors3)

example, most miRNA genes harbour highly polymorphic regions in their 5' and 3' ends, resulting in alternative splicing by Drosha and/or Dicer. Moreover, gene silencing is greatly dependant on the post-translational modification of proteins involved in miRNA biogenesis, such as the phosphorylation, PARylation and sumoylation of AGO2(Gebert & MacRae, 2019).

In addition to normal development, miRNAs also have established roles in disease pathogenesis. For example, pathogenic variants in *miR-96* (MIM: 611606) have been shown to cause autosomal dominant hearing loss (*DFNA50*; MIM: 613074). Given that these pathogenic variants reside within the seed sequence of *miR-96*, the processed guide strand within the RISC complex fails to hybridize to mRNA targets, leading to aberrant gene expression profiles within sensory hair cells that are required for normal auditory function(Mencia *et al.*, 2009). This remarkable finding prompted a paradigm shift in hearing loss gene discovery efforts, revealing that many other miRNAs, including *miR-182* and *miR-183*, are involved in the normal development of the auditory system, as well as hearing loss(Ushakov *et al.*, 2013). However, identifying miRNA-mediated hearing loss variants is not trivial, given that this mechanism involves many small cumulative events, rather than a single pathogenic variant(Rudnicki & Avraham, 2012). Conversely, this complexity is further confounded by variants that reside within miRNA binding sites(Ushakov *et al.*, 2013). Nonetheless, with respect to a ubiquitously expressed gene that is associated with a syndromic disease, a variant that introduces a tissue-specific miRNA binding site into the 3' UTR of that gene could theoretically result in a related but non-syndromic disorder(Ushakov *et al.*, 2013). Even though variants that reside in miRNA genes themselves are not a common cause of hearing loss(Hildebrand *et al.*, 2010), those variants within 3' UTR miRNA binding sites

represents a promising area of research in human disease. Collectively, variation at the RNA level, including our understanding of alternative splicing, cryptic splicing, and the miRNA-mediated spatiotemporal regulation of gene expression, represents a new, yet challenging area in human genetics that may uncover novel genes that have previously evaded discovery.

1.2 Hereditary Hearing Loss and Its Syndromes

1.2.1 The Auditory System and Hearing Loss

The auditory system has a single, yet complex task – to relay external acoustic stimuli to the auditory cortex, where it is perceived as sound (J. Anthony Seikel, 2015). Auditory perception begins with sound waves traveling into the outer ear and through the auditory canal where they meet and vibrate the tympanic membrane (or ear drum). The tympanic membrane is a cone-shaped structure that articulates with a chain of three small bones in the middle ear, consisting of the malleus, incus and stapes, which are collectively called the “auditory ossicles”. Auditory information of frequency and amplitude is passed on to the auditory system through vibration transfer from the footplate of the stapes to the oval window of the “bony labyrinth”. This labyrinth is filled with a sodium-rich fluid called perilymph. The round window is another flexible membrane, which permits the displacements of perilymph by the piston-like action of the stapes. The round window is found on the spiral portion of the bony labyrinth, known as the cochlea, which is the main sensory organ of the auditory system. (Figure 1.7).

Vibrations produced by the stapes first enter and ascend to the apex of the cochlea through the scala vestibuli, which in turn, descend back down the cochlea through the scala tympani, until they reach the round window. The cochlear duct (scala media), is a structure that

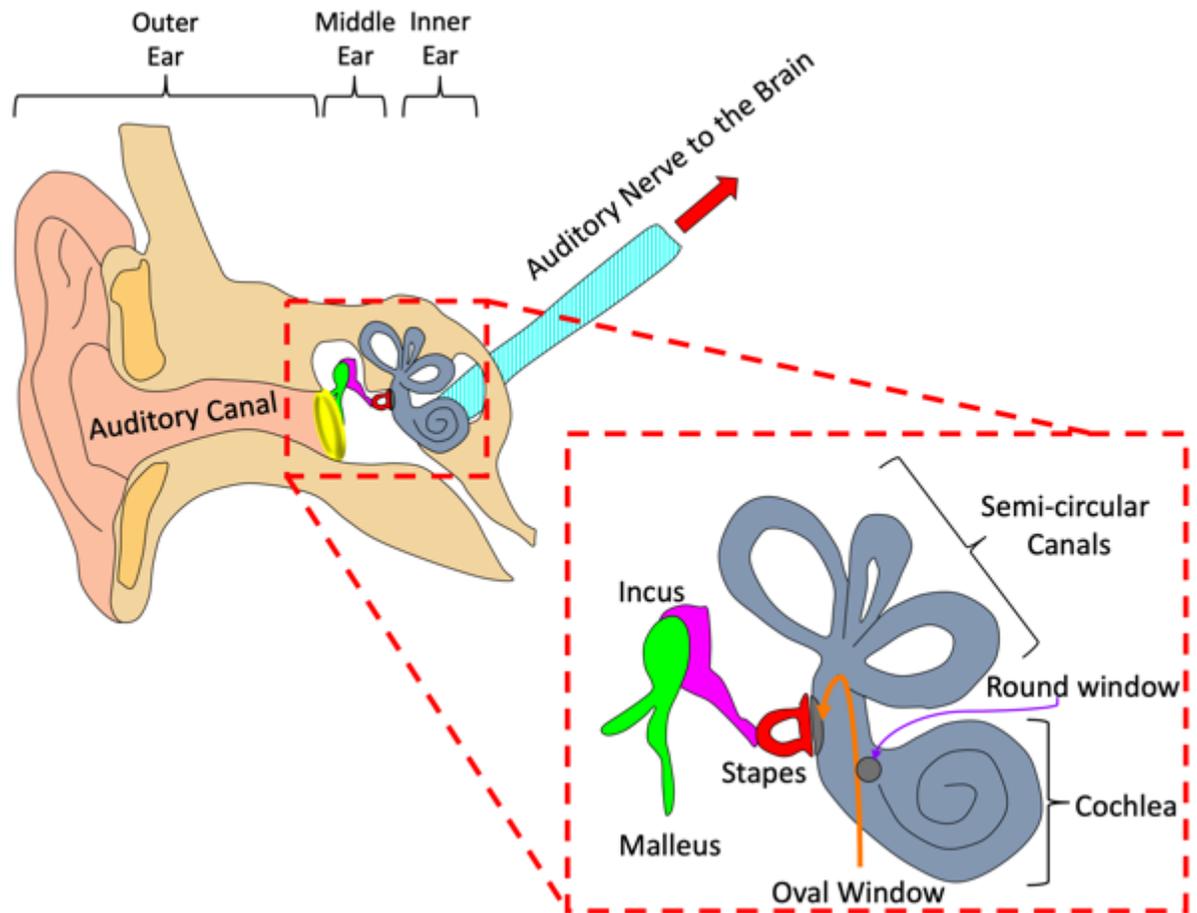


Figure 1.7. Schematic of the auditory pathway. Acoustic stimuli enter the auditory system in the outer ear, where sound waves with travel through the auditory canal until this energy is absorbed by the tympanic membrane (or ear drum). This energy is then transferred to the auditory ossicles of the middle ear, the malleus, incus and stapes. The cochlea is a snail-shaped, fluid-dilled sensory organ within the auditory system. The cochlear fluid (or perilymph) is displaced by piston-like action of the stapes on the oval window. This mechanism is possible due to an additional flexible membrane called the round window. Once incoming sound vibrations enter the cochlea, they elicit action potentials within sensory hair cells in the Organ of Corti, which in turn release neurotransmitters onto the auditory nerve that send electrical impulses to the brain. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors4)

contains potassium-rich endolymph and is situated between the scala vestibuli and scala tympani. Together, differential electrochemical gradients between the three cochlear compartments are maintained and separated by two flexible membranes that move in response to the vibrations traveling up the scala vestibuli; namely, the Reissner's membrane and the basilar membrane. The Organ of Corti is a highly specialized structure that is comprised of three rows of outer hair cells, a single row of inner hair cells, and additional supporting cells that collectively rest on the basilar membrane of the cochlear duct. This structure is filled with cortilymph, a perilymph-like medium that is sodium-rich and potassium-devoid. Despite the stark ionic composition differences between endolymph and cortilymph, the reticular lamina, which is comprised of tight junctions, creates a barrier between these two mediums and maintains these electrochemical gradients (Figure 1.8).

The cochlea exhibits tonotopic organization, meaning that high and low frequency sounds are received at the base and apex of the cochlea, respectively, providing a frequency-based map of the cochlea (Figure 1.9). In light of tonotopic organization, different frequencies peak at different positions along the basilar membrane, exciting distinctive subsets of hair cells. The manner in which the basilar membrane vibrates in response to sound is essential in our understanding of cochlear function. Sensory hair cells within the Organ of Corti are located between the basilar and tectorial membranes and are stimulated by incoming sound vibrations (Figure 1.8). While the stereocilia that protrude from the apical surfaces of inner hair cells detect and transmit this auditory information to the brain through the cochlear nerve in response to fluid motion, those of the outer hair cells are directly coupled to the tectorial membrane and amplify incoming sound vibrations. In response to incoming vibrations into the cochlea, outer hair cells strategically amplify these sounds, whereas inner hair cells detect and transmit this

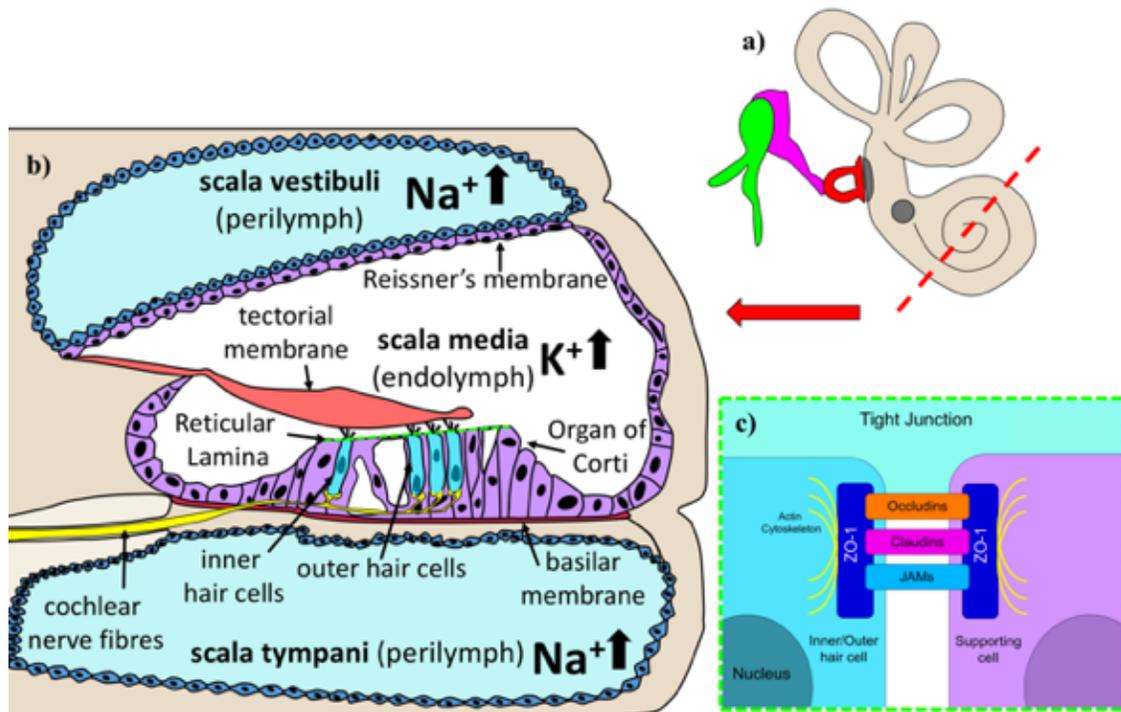


Figure 1.8. Anatomy and physiology of the cochlea.) Schematic of the cochlea, indicating the cross sectional plane (dashed red line) that is depicted in (b). The three main compartments of the cochlea are the scala vestibuli and scala tympani (filled with sodium-rich perilymph), as well as the scala median (filled with potassium-rich endolymph). These compartments are separated by the Reissner's and basilar membranes. The Organ of Corti contains three rows of outer hair cells and a single row of inner hairs, and resides on the basilar membrane. The Organ of Corti is also bathed in a perilymph-like fluid that is sodium rich. The endolymphatic and cortilymphatic electrochemical gradients are maintained by the reticular lamina (green dashed line). Tight junctions (c) at the apical surface of the Organ of Corti provide a barrier that prevents the mixing of endolymph and cortilymph. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors5; Contributors6)

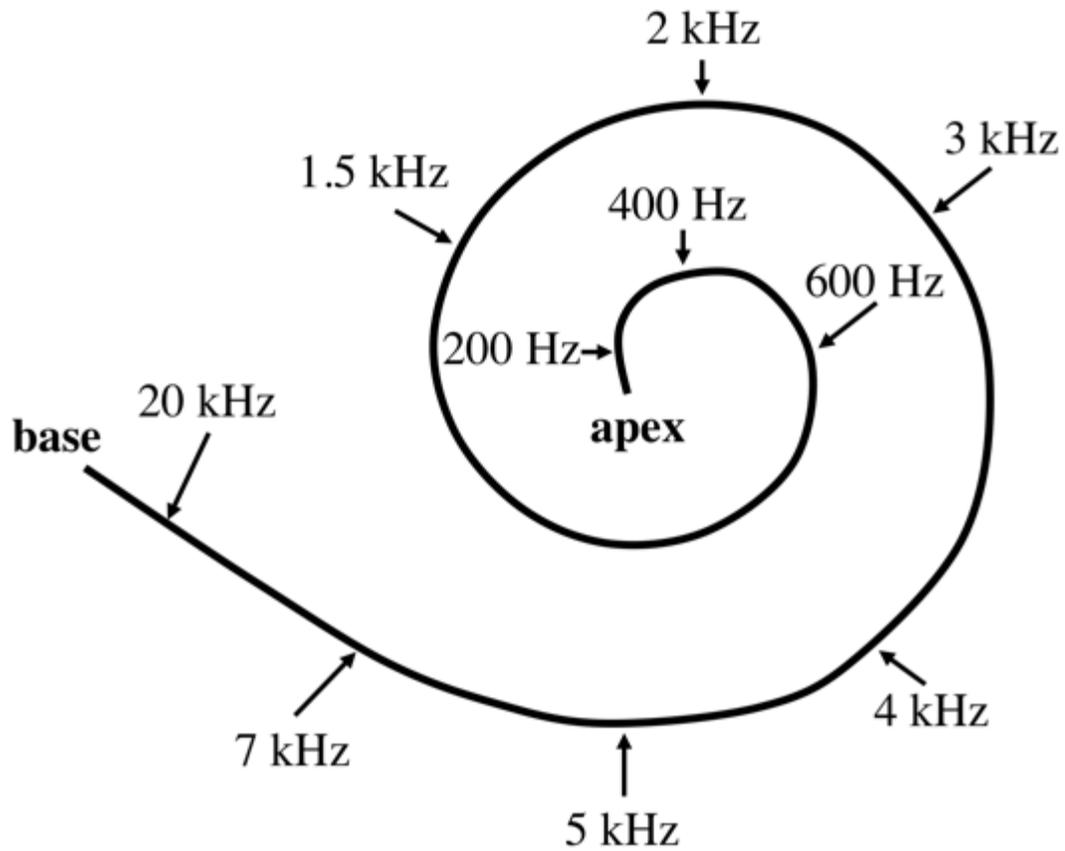


Figure 1.9. Tonotopic map of the cochlea. This topological schematic of the cochlea demonstrates that the sensory hair cells at the base of the cochlea are responsible higher frequency sound perception, which decreases in frequency as you ascend to the apex of the cochlea. Human hearing frequencies range from 20kHz to 200 Hz. Hz: hertz, kHz: kilohertz (1 kHz = 1,000 Hz). This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors7)

auditory information to the brain through the cochlear nerve. These processes are mediated through mechanotransduction of sensory hair cells in the cochlea. A defining feature of auditory hair cells is that they occur in bundles that are mechanically linked to one another at the tip (tip links), and at the base (ankle links). While the major molecular constituents of ankle links include adhesion g-protein-coupled receptor-v1 (ADGRV1) and usherin (USH2A), tip links are comprised of protocadherin-15 (PCDH15) and cadherin-23 (CDH23). These adhesion proteins complex with scaffolding molecules, including myosins, which are bound to actin filament networks of the stereocilia (Figure 1.10a). Moreover, tip links complex with non-specific, mechanically-gated ion channels, such as transmembrane channel-like protein 1 (TMC1) and transmembrane channel-like protein 2 (TMC2) at the apical tips of stereocilia. Given that stereocilia are tethered together, stereocilia bundles deflect as a single unit in response to sound stimuli. This deflection creates tension between the tips links, which opens the mechanically gated ion channels (Figure 1.10b). Subsequently, sensory hair cells are subject to a massive influx of potassium and calcium, which depolarizes the cell (Figure 1.10c). Once the membrane potential of the sensory hair cell reaches threshold, an action potential will be produced, resulting the release of neurotransmitters onto the auditory nerve. The genes that encode for such ion channels, and adhesion and scaffolding proteins that comprise the tip and ankle links, are essential for maintaining normal hearing and many of them have been implicated in hereditary hearing loss and it syndromes (Appendix B).

Hearing loss is one of the most common sensory disorders that affects an estimated 2-3 in 500 newborns annually(Helga V Toriello 2016; Richard JH Smith, 1999). According to a hearing loss survey in 2006(Statistics Canada, 2010), 5% of Canadians aged 15 and over exhibit some degree of hearing impairment. Approximately 466 million people exhibit some

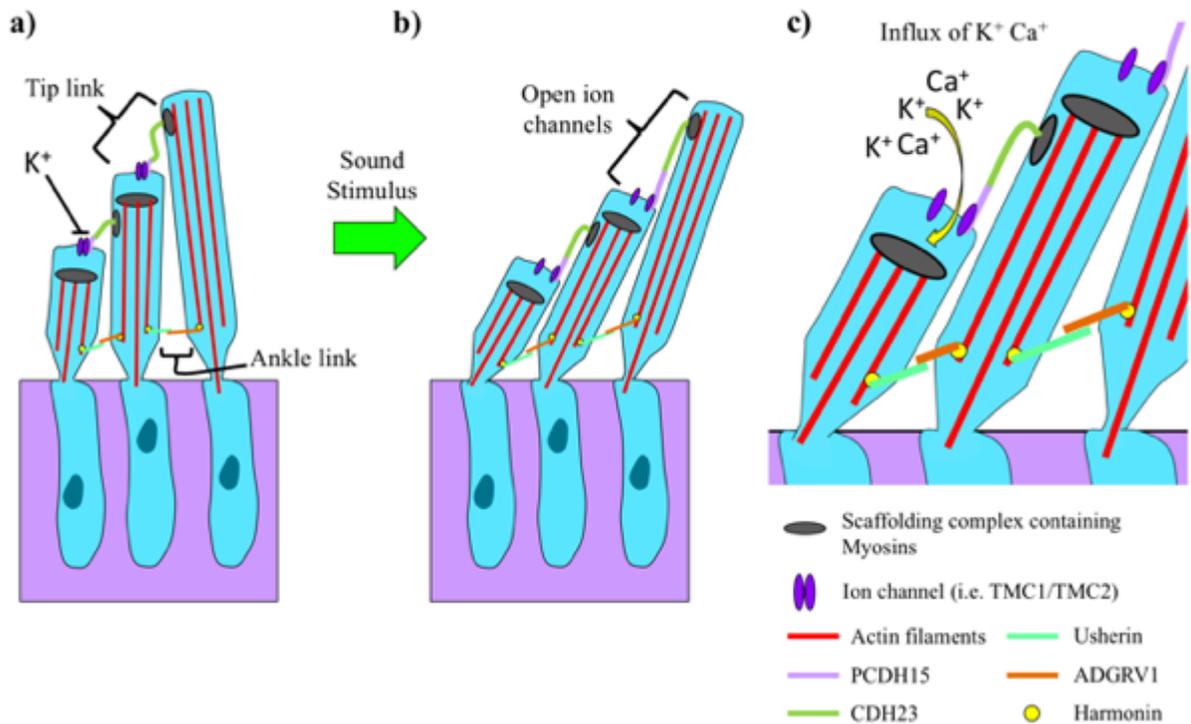


Figure 1.10. Schematic of hair cell bundles and mechanotransduction. A) Hair cell bundle illustrating the presence of tip links and ankle links. The main molecular constituents of tip links include protocadherin-15 (PCDH15) and cadherin-23 (CDH23), which tether the apical surface of one stereocilia to the side of the adjacent stereocilia. Ankle links are comprised of other adhesion proteins; namely, adhesion g-protein-coupled receptor-v1 (ADGRV1) and usherin (USH2A). Together, these adhesion proteins bind with the stereocilia actin filament networks, through their interaction with myosin scaffolding complexes. B) Upon the arrival of a sound stimulus, hair cell stereocilia will deflect and create tension on the tip and ankle links, c) This tension permits the opening of mechanically gated ion channels, which will allow for ion such as potassium and calcium to enter the cell body of sensory hair cells. In turn, this depolarizes the cell, leading to potentials and the release of neurotransmitters to the auditory nerve This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors8)

form of hearing loss(World Health Organization, 2018). Moreover, this overwhelming figure is projected to reach nearly 1 billion hearing impaired people by 2020(Van Eyken *et al.*, 2007). There are four main types of hearing loss: conductive, sensorineural, mixed and central auditory dysfunction(Richard JH Smith, 1999). While conductive hearing loss involves obstruction of sound waves in reaching the inner ear through abnormalities of the outer or middle ear, sensorineural hearing loss is caused by deficits in the auditory nerve or inner ear. The combination of both sensorineural and conductive hearing loss is referred to as mixed hearing loss. Central auditory dysfunction is different from sensorineural and conductive hearing loss in that it involves defects in the central nervous system, which prevent the transmission and processing of auditory information (Richard JH Smith, 1999 ; Van Camp G, 2015).The identification, diagnosis, and treatment of hearing and balance disorders falls within the field of clinical audiology(Gelfand, 2011). In order to assess a patients level of hearing, balance and speech perception, audiologists rely on several behavioural exams, from a sound and psychological perspective.

The key to understanding sound transmission is vibration (reviewed in Martin & Clark, 2014). When somebody speaks, their vocal chords vibrate, producing pressure waves that are propagated through the air. Collectively, sound waves consist of two main components: compressions and rarefactions. While nitrogen and oxygen molecules in the air are pushed together by high pressure during the compression phase of a sound wave, rarefactions pull them apart in areas of low pressure. Sound frequency, measured in hertz (Hz), is defined as the number of sound waves that pass a certain point in a given time. High frequency sounds are produced by shorter waves that move more quickly through the air. In contrast, low frequency sounds result from fewer, slower fluctuating waves. The “loudness” of a sound is dependent on

amplitude, or the difference between the compression and rarefaction phases of a sound wave, which is measured in a relative, logarithmic unit called decibels (dB).

Pure-tone audiometry measures hearing sensitivity by determining the lowest detectable sound amplitude at different frequencies. Specifically, this method measures air conduction hearing thresholds that range from 125 Hz to 8 kilohertz (kHz), which are the frequencies that are relevant to human speech. Pure tones are played through head phones to a maximal intensity (120 dB), until the subject indicates that they have heard the sound. Hearing thresholds are measured in both ears and are plotted on a behavioural audiogram, with the different tested frequencies on the X-axis and the sound amplitude (or intensity) on the Y-axis. On a standard audiogram, the left and right ears are annotated as X's and O's, respectively, and normal hearing is defined as having less than 20 dB hearing threshold across all frequencies. Mild hearing loss has hearing thresholds between 20-40dB. Moderate hearing loss has hearing thresholds between 41-70 dB. Severe hearing loss has hearing thresholds between 71-95 dB. Profound hearing loss has hearing thresholds in the excess of 95 dB (Figure 1.11). Audiogram configuration is typically classified into low, mid, and high frequencies. In addition, these configurations can be further characterized according to their slope as flat, gentle or steep, which can be present in one ear (unilateral), or more commonly, both ears (Mazzoli M, 2003; bilateral; Figure 1.12).

1.2.2 The Genetics of Hearing Loss

Given the wide-range hearing loss aetiologies that include aging, ototoxic drugs, noise exposure, and genetics, this disorder exhibits extreme heterogeneity in the clinical setting (Richard JH Smith, 1999; Van Camp G, 2015). Despite this diversity, 50-60% of

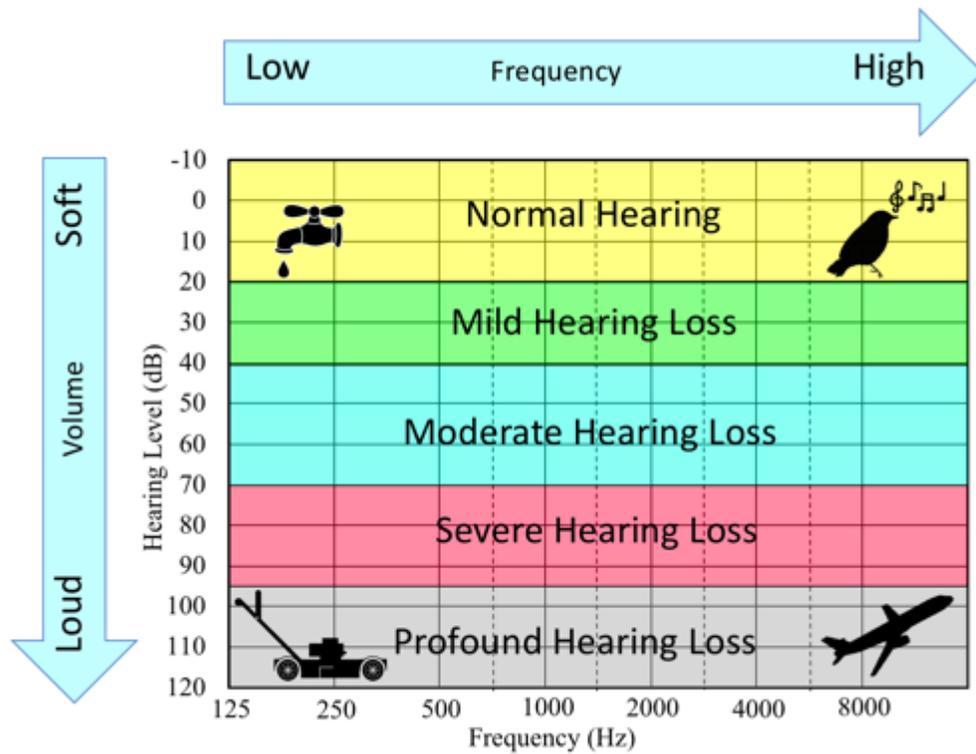


Figure 1.11. Audiogram demonstrating hearing loss severity. Normal hearing thresholds are characterized by the detection of sound above 20 dB at any frequency. Mild and moderate hearing loss thresholds range from 20-40 dB and 41-70 dB, respectively. While severe hearing loss demonstrates thresholds between 41-95 dB, a profound hearing loss is described as having threshold beyond 95 dB(Mazzoli M, 2003).

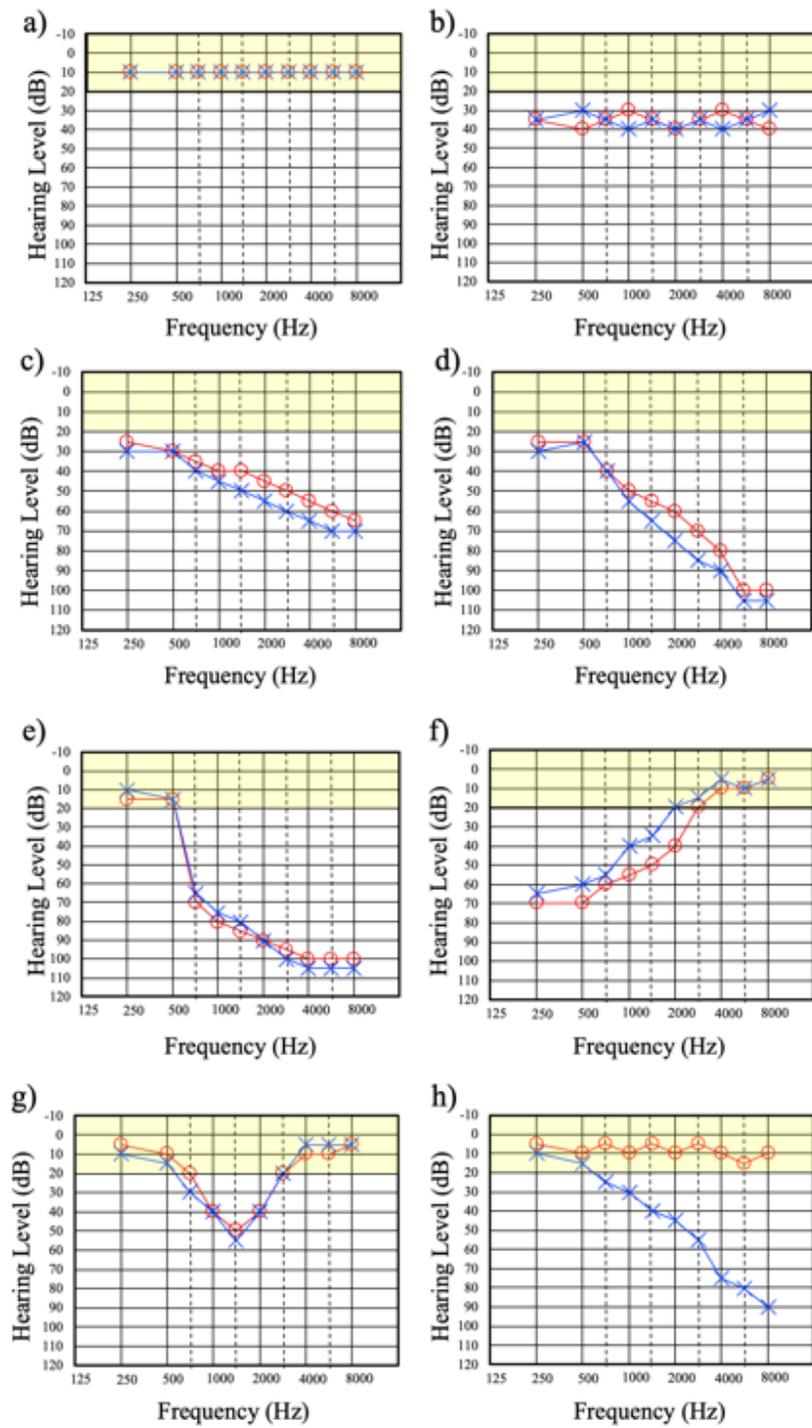


Figure 1.12. Various audiograms illustrating the types of sensorineural hearing loss. A) normal hearing, b) Flat hearing loss, c) Gently sloping hearing loss, d) Steeply sloping, e) Precipitous hearing loss, f) Low-frequency ascending (reverse slope), g) Mid frequency “cookie-bite” and unilateral hearing loss (h). Red and blue lines denote hearing thresholds of the right ear and left ear, respectively.

hearing loss is due to genetic factors(Morton & Nance, 2006), representing an unprecedented opportunity to identify and describe its molecular pathophysiology. However, phenocopies, or individuals that develop a given trait due environmental factors, are quite common in hearing loss(Arnett *et al.*, 2011). For this reason, phenocopies may resemble other study subjects within larger families members with genetically-defined hearing loss. Thus, it cannot be understated how important it is to have a complete and thorough understanding of the hearing phenotype when performing genetic hearing loss studies. For the most part, hearing loss is a monogenic disorder caused by a single gene pathogenic variant(Richard JH Smith, 1999); however, multifactorial forms exist where both gene-gene and gene-environment interactions significantly influence the development and progression of hearing loss(Cui *et al.*, 2017; Haider *et al.*, 2002; Sherry *et al.*, 2001; Speicher *et al.*, 2009)

In excess of 400 genetic syndromes involving deficits of the eye, neurological, kidney and musculoskeletal systems exhibit a hearing component(Helga V Toriello 2016). While non-syndromic hearing loss accounts for 70% of hereditary hearing loss, the remaining 30% is recognized as syndromic and is the most common feature of multi-organ disorders(Richard JH Smith, 1999; Toriello & Smith, 2013). Syndromic hearing loss is classically characterized by overt clinical features such as craniofacial and connective tissue abnormalities as in Treacher Collins syndrome. However, features of other hearing loss syndromes, such as visual deficits in Usher syndrome (USH), may manifest later in life, leading to delays or misdiagnosis(Keats & Corey, 1999; Kimberling *et al.*, 2010; Mathur & Yang, 2015). This presents a significant challenge in the clinical management, diagnosis and surveillance of those at-risk of developing syndromic forms of hearing loss, given the high demand of prioritized genetic tests for disorders of known aetiologies. However, the development of synergistic relationships between

clinical and research teams is an effective approach of mitigating such diagnostic challenges. In addition, this approach accelerates the pace of gene discovery, as well as improving health care policy, ethics, and routine genetic testing(K. Hodgkinson *et al.*, 2009).

Tremendous progress has been made since the first description of a genetic form of hearing loss in 1995(de Kok *et al.*, 1995). Specifically, a global effort in understanding the genetics of hereditary hearing loss and its syndromes has given rise to the identification of approximately 223 loci and 143 genes; thus, this disorder displays profound genetic heterogeneity(Van Camp G, 2015). Genetic loci that are associated with nonsyndromic hearing loss are designated as *DFN* (for *DeaFN*ess), and these are further annotated according to their inheritance patterns: *DFNA*, *DFNB*, and *DFNX* loci categorically denote autosomal dominant, autosomal recessive, and X-linked inheritance, respectively(Richard JH Smith, 1999). Although an estimated 75-80% of non-syndromic hearing loss is transmitted as an autosomal recessive trait, pathogenic variants that exhibit a dominant (20%), X-linked (2-5%) and mitochondrial (1%) inheritance pattern account for the remaining cases (Figure 1.13; R. J. Smith *et al.*, 2005). Importantly, a significant number of these genes exhibit pleiotropy, where they are associated with both syndromic and nonsyndromic hearing loss – the latter typically belonging to the *DFNA* and/or *DFNB* grouping. For example, *MYO7A* (MIM: 276903) can cause *DFNA11* (MIM: 601317) and *DFNB2* (MIM: 600060), as well as *USH1B* (MIM: 276900)(X. Z. Liu, Walsh, Mburu, *et al.*, 1997; X. Z. Liu, Walsh, Tamagawa, *et al.*, 1997; Weil *et al.*, 1995).

Even though hearing loss is predominately expressed as a autosomal recessive trait, this may not reflect the true epidemiology of genetic hearing loss, due to factors such as ascertainment biases and difficulties of identifying dominant genes(Speicher *et al.*, 2009).

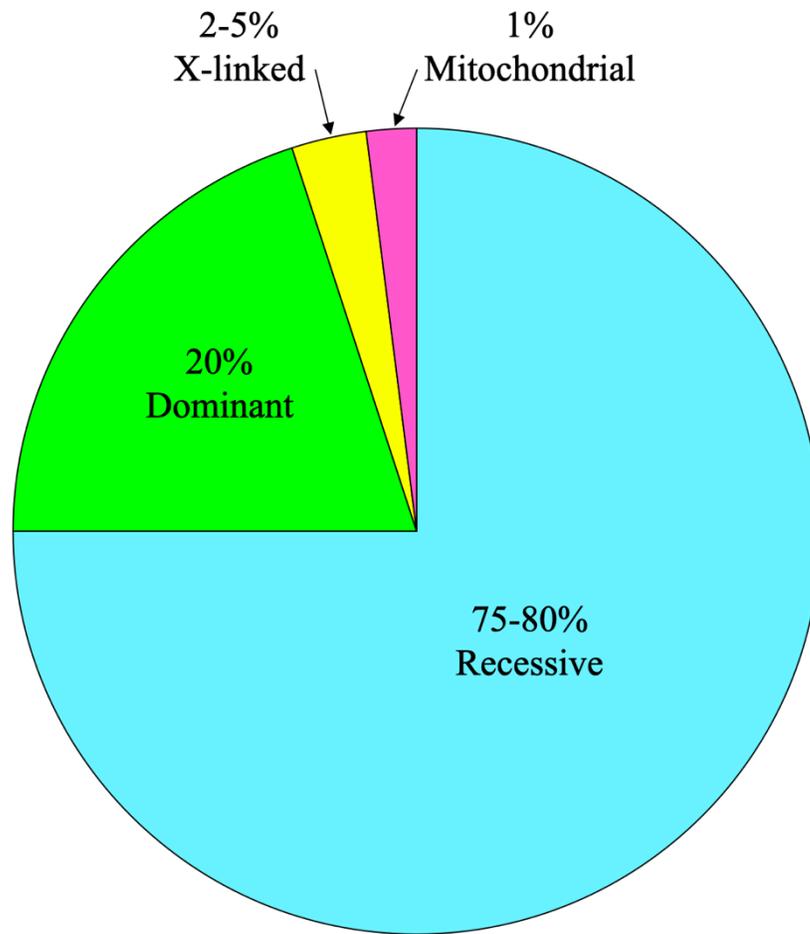


Figure 1.13. Prevalence of genetic hearing loss according to mode of inheritance. Data retrieved from Smith *et al*(R. J. Smith *et al.*, 2005).

Despite how far the hearing loss research field has come, many loci remain unsolved (Richard JH Smith, 1999; Van Camp G, 2015). Among these unsolved loci, *DFNA* genes are particularly challenging to identify, given broad critical regions coupled with the rarity of large extended families with well-described clinical data. This problem is further exacerbated by reduced penetrance and variable expressivity of hearing phenotypes (Richard JH Smith, 1999) as well as over 10 million heterozygous genetic variants in human genome (Sherry *et al.*, 2001). Variable expressivity refers to differing phenotypic features among individuals with the same genotype. For example, the recent discovery of *PTPRQ* (MIM: 603317) as the cause of *DFNA73* (Eisenberger *et al.*, 2018) found extensive variable expressivity within a 4-generation German family. Specifically, a pathogenic nonsense variant co-segregated in this family; however, hearing loss onset ranged from early childhood to the third decade with a variable audiogram configuration (Eisenberger *et al.*, 2018). Penetrance refers to the proportion of individuals carrying a particular genetic variant that express an associated phenotype. In cases of reduced penetrance, not all individuals with pathogenic alleles will display a given phenotype. For instance, a heterozygous, autosomal dominant, variant in *GJB2*, p.Val37Ile, is sufficient to cause *DFNA3* (MIM: 601544) with an estimated 17% penetrance (Chai *et al.*, 2015). These challenges highlight the importance of thorough clinical ascertainment and even though these factors present significant challenges in the discovery of dominant hearing loss genes, they can be overcome by studying large multiplex families, which are frequent within well-defined homogenous populations, such as the genetic isolate of Newfoundland.

1.2.3 The Newfoundland Founder Population and Hearing Loss Research

In recent history, the Newfoundland and Labrador (NL; island portion) population has provided an unprecedented opportunity to advance our knowledge of many monogenic disorders (Abdelfatah, McComiskey, *et al.*, 2013; Abdelfatah, Merner, *et al.*, 2013; Ahmed *et al.*, 2004; Doucette *et al.*, 2009; Merner *et al.*, 2008; Moore *et al.*, 2005; Young *et al.*, 2001). Compared to more ethnically diverse populations that possess a more heterogeneous gene pool, homogeneous founder populations present many advantages in gene discovery research endeavours, including both Mendelian and complex disorders. Relative to other founder populations, NL is comparatively young (<20 generations), providing an ideal research environment to identify the genetic basis that underlies many diseases (Rahman *et al.*, 2003).

According to archeological records, NL has been populated longer than any other location in North America. Approximately 18,000 years ago during the Last Glacial Maximum, the remnants of the Laurentide ice sheet disappeared, giving rise to vast landmasses positioned at the Northeastern shelf of Canada. While the landmass of Island portion of Newfoundland is approximately 100,000 km², Labrador is almost three times its size that covers an estimated 295,000 km². The availability of inhabitable land provided an ideal location for the first indigenous people, including the Maritime Archaic, Palaeoeskimo, and Beothuk, to establish Labrador and the island of Newfoundland roughly 10,000 and 6,000 years ago, respectively. Recent mtDNA studies suggest that the relationships between these culturally distinct peoples predate their arrival by land across the continent of North America (Duggan *et al.*, 2017).

In the turn of the 16th century, the arrival of John Cabot to Cape Bonavista marked the discovery of NL by European explorers. The earliest European colonies date back to the early 1600's, which were founded by English and Irish fisherfolk that came to the island to exploit the summer cod fishery. The fishing industry continued to grow, drawing in settlers that

immigrated to the island, reaching a population of roughly 12,000 people in 1775 (Figure 1.14a). By the late 19th century, the fisheries became so lucrative that hundreds of outport fishing communities were established along a rugged 10,000 kilometer of coastline (Figure 1.14b). Although an estimated 200,000 individuals resided in NL by 1890, more than 95% of residents were native born and the descendants of approximately 20,000 English (Protestant) and Irish (Roman Catholic) fisherfolk (Mannion & Memorial University of Newfoundland. Institute of Social and Economic Research., 1977). Geographic and religious segregation led to the establishment of this genetic isolate that exhibits high inbreeding coefficients and genetic homogeneity (Bear *et al.*, 1987, 1988). Today, NL has stratified into 3 distinct population clusters of Protestant, Roman Catholic and North American Indigenous peoples, the latter being relatively small (Zhai *et al.*, 2016). These properties make the NL population an extraordinary resource for studying monogenic disorders. An excellent example of this is *arrhythmogenic right ventricular dysplasia type 5* (*ARVD5*; MIM: 604400), an autosomal dominant condition that is characterized by the fibro-fatty replacement of the heart ventricles, leading to sudden death in young people. This lethal condition is caused by a heterozygous missense variant in *TMEM43* (p.Ser358Leu; Merner *et al.*, 2008), which was identified in >25 NL families. This pathogenic allele was found to be imported from Europe and dates back to early medieval ages (400 – 700 AD), predating today's European nations (Milting *et al.*, 2015). Compared to a global incidence of 1 in 5,000, *ARVD5* is enriched in the NL population (1 in 800 – 1,000) due to founder effects of the *TMEM43* (p.Ser358Leu) allele (Kathy Hodgkinson, Personal communication: Discussion on the *ARVD5* epidemiology in Newfoundland -October 2018).

Genetic drift due to founder effect is defined as the reduction in genetic diversity

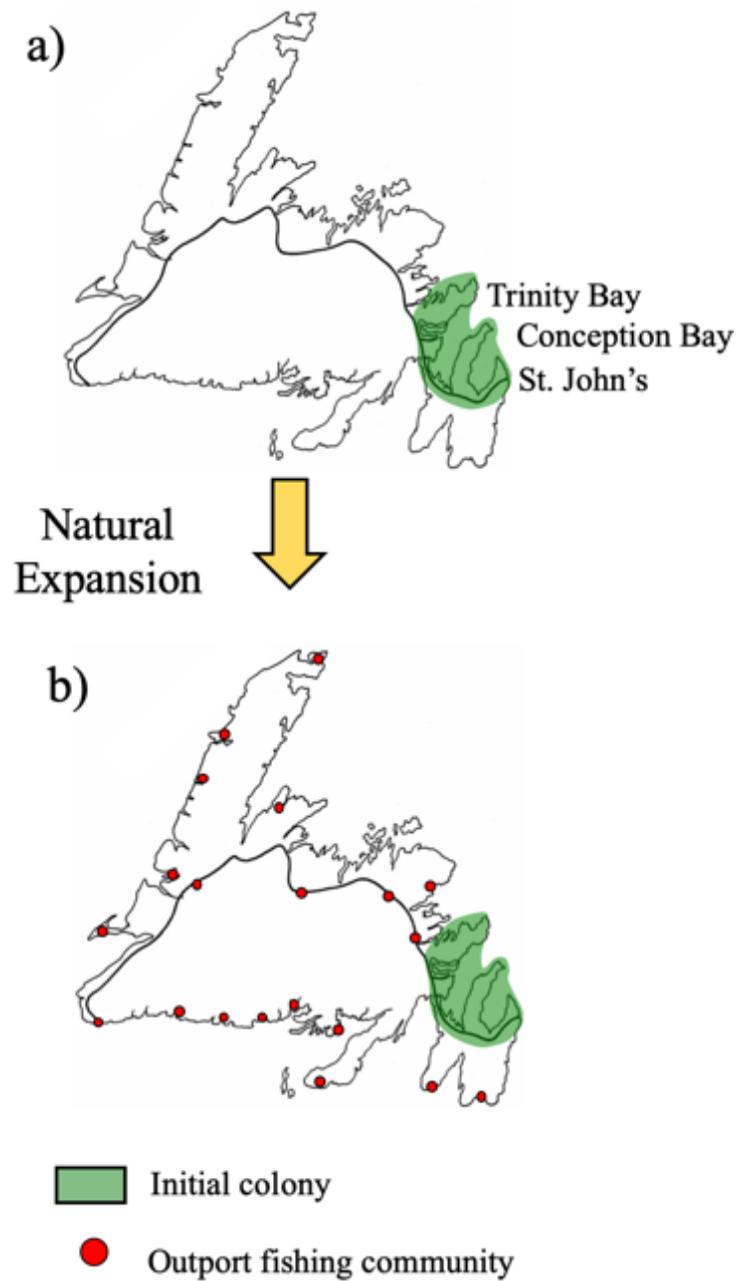


Figure 1.14. Colonization of the Newfoundland genetic isolate. a) The island population was initially colonized by English (Protestant) and Irish (Roman Catholic) fisherfolk which reached a population of ~12,000 in the Bonavista/Trinity/Conception Bay area, b) The natural expansion of the initial colony led to the development of hundreds of outport fishing communities by the late 19th century. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from (Contributors9)

through the establishment of a larger population, which has descended from a small number of colonizing ancestors(Tom Strachan, 2010). Founder populations possess many advantages in the mapping and identification of disease genes. Given the limited number of early settlers, disease alleles are often enriched in founder populations, which emerge from a common ancestor(Kristiansson *et al.*, 2008). Many well-known examples of founder effects include the Canadian Mennonites, French (Northeastern Québec), Hutterites, Icelandic, Dutch and the Finnish populations, which possess recurrent disease alleles that cause Fanconi anemia(de Vries *et al.*, 2012), myotonic dystrophy(Yotova *et al.*, 2005), limb girdle muscular dystrophy(Boycott *et al.*, 2008), early onset atrial fibrillation(Ebenesersdottir *et al.*, 2018), hearing loss(Fransen *et al.*, 2001), and Usher syndrome type 3a(Pakarinen *et al.*, 1995) respectively. In comparison to genetically diverse urban areas that rely on nuclear families to perform genetic studies, founder populations, such as NL, are typically characterized by a higher frequency of large families with deep genealogies. In addition, these families often descend from a common ancestor, forming a larger clan(Lupski *et al.*, 2011). This advantage permits cascade sequencing on candidate variants in many affected family members who share much more of the genome, compared to a single person in a nuclear family. Although this approach allows geneticists to more readily identify causative pathogenic variants, there are challenges related to using NL families in gene identification studies. Many large family members reside in remote areas that are isolated from tertiary hospitals, making clinical ascertainment more arduous. Without a complete family history and phenotype description, members could erroneously be labeled unaffected when they are in fact affected. Additionally, detecting phenocopies within larger

families is difficult with limited clinical data, which adds another level of complexity when interpreting co-segregation results from cascade sequencing.

The NL population has proven to be an exceptional resource for the study of the genetic basis of hereditary hearing loss. The inaugural Director of the Provincial Medical Genetics Program at Eastern Health, Dr. Elizabeth Ives, was consulted by the Department of Education (Dr. Claire Neville-Smith) to look at the increased incidence of deafness along the South shore of province. Subsequently, the genetic basis of this hearing loss for the first NL family was discovered in 2001 (Young *et al.*, 2001). To do this, Young *et al.* (Young *et al.*, 2001) performed genome-wide SNP genotyping and linkage analysis on a 6-generation family (Family C) recruited by Dr. Claire Neville-Smith and Dr. Ives. A logarithm of the odds (LOD) score of 11.58 was achieved at 4p16, which overlapped with, *DFNA6*, *DFNA14* and *DFNA38* (MIM: 600965). Full sequencing of *Wolframin* (*WFS1*; MIM: 606201) revealed 5 benign polymorphisms and 6 novel heterozygous variants. Importantly, pathogenic biallelic variants in *WFS1* also cause Wolfram syndrome (MIM: 222300), a rare and severe autosomal recessive neurodegenerative condition that is characterized by diabetes mellitus, optic atrophy, and hearing loss. Extensive genotyping in the extended pedigree and haplotype analysis identified a disease haplotype co-segregating with isolated hearing loss across multiple generations in Family C, demonstrating that *WFS1* c.2146G>A (p.Ala716Thr) was causing *DFNA38*. Cascade sequencing revealed that all affected *WFS1* c.2146G>A heterozygotes display nonsyndromic low-frequency hearing loss. Crucially, one Family C homozygote was identified who developed juvenile diabetes mellitus and a hearing profile that was not consistent with *WFS1* c.2146G>A heterozygotes. This finding was a seminal aspect of the *WFS1* discovery, indicating that some homozygous mutant alleles may manifest as a mild or intermediate Wolfram syndrome

phenotype. In addition, this discovery demonstrated that variants in a gene known to cause a recessive, syndromic form of hearing loss can also cause a dominant disorder, in this case isolated hearing loss (Young *et al.*, 2001).

Since the inaugural *WFSI* study, additional pathogenic hearing loss variants have been identified in the NL population (Abdelfatah, McComiskey, *et al.*, 2013; Abdelfatah, Merner, *et al.*, 2013; Ahmed *et al.*, 2004; Doucette *et al.*, 2009; McComiskey, 2010; Squires, 2015), which has led to the development of a screening protocol that aimed to ascertain families who were positive for previously described deafness alleles. This effort included the screening of all family probands for *WFSI* c.2146G>A (Young *et al.*, 2001), *TMPRSS3* (MIM: 605511) c.207delC (Ahmed *et al.*, 2004) and c.782+3delGAG, *PCDH15* (MIM: 605514) c.1583T>A (Doucette *et al.*, 2009), and *SMPX* (MIM: 300226) c.99delC (Abdelfatah, Merner, *et al.*, 2013), two large deletions in *GJB6* (MIM: 604418, delD1351830 and delD1351854) and the most common hearing loss variant, *GJB2* (MIM: 121011) c.35delG. Subsequently, families whose deafness could not be explained by known variants were prioritized for downstream genomic analyses, based the configuration of their audiogram and were categorically grouped as having low, mid or high frequency hearing loss audioprofiles. In order to identify candidate hearing loss genes, phenotypic data were submitted to Audiogene (Hildebrand *et al.*, 2009), a web-based software program that compares audioprofiles to known average audiograms of 34 deafness loci. Consequently, sequence variants in candidate genes were submitted to *in silico* programs (SIFT, Polyphen-2 and Mutation Taster) and tested for co-segregation within hearing loss families. Local MAF's were determined by screening each candidate variant in ethnically-matched population controls. Through these efforts, the Young laboratory identified *MT-RNR1*

m.1555A>G(Squires, 2015), *COCH* c.151 C>T(McComiskey, 2010), and *KCNQ4* c.806_808delCCT(Abdelfatah, McComiskey, *et al.*, 2013) throughout the NL population.

The NL Hereditary Hearing Loss Project has enrolled over 200 families into the study who have been screened for 23 pathogenic variants (Appendix C), which has led to establishing the Centre for Genomics-Based Research and Development in Hearing Science in Grand-Falls Windsor, NL. This facility is equipped with a comprehensive suite of clinical audiology tools designed to accelerate our understanding of genotype-phenotype correlations. Upon reviewing retrospective medical histories, extensive hearing loss clinical ascertainment, as well as in-depth study participant interviews, several family pedigrees were extended with well-described phenotype data. Several of these families were negative for known NL hearing loss variants; therefore, these kindreds were promising candidates for downstream genome-wide SNP genotyping and linkage analysis. Even though linkage analysis is a traditional “classical genetics” tool that is routinely used in forward genetics, it is quite powerful when used in combination with more modern next-generation sequencing (NGS) technologies, such as whole exome (WES) or whole genome (WGS) sequencing. Even though WES alone has the ability to identify several predicted pathogenic variants in known deafness-associated genes(Lewis *et al.*, 2018), the genetic etiology of hearing loss in many families remains unknown. Moreover, Lewis *et al* (2018) highlighted a significant pitfall of using WES in the clinical setting, noting that segregation analysis of candidate variants within larger families is an asset to hearing loss gene discovery and diagnosis. In addition to discerning pathogenic hearing loss variants, larger multiplex families permit linkage analysis, which allows geneticists to reduce a list of tens of thousands of variants to a smaller subset of more promising candidates. Given that linkage analysis is based on well-established statistical models, this approach provides sufficient

evidence to disregard variants residing outside of identified critical regions(Ott *et al.*, 2015). In light of the millions of variants across the human genome(Sherry *et al.*, 2001), combing NGS with linkage analysis in the post-genome area expedites the discovery of disease genes, especially with autosomal dominant conditions(Ott *et al.*, 2015).

Some of the earliest molecular characterization of hearing loss and its syndromes heavily relied on whole-genome linkage analysis within clinically well-described families, which maps informative genetic markers to defined critical chromosomal intervals that encompass a disease-causing variants(Botstein & Risch, 2003). However, due to a limited number of meioses, critical regions were broad, spanning several megabases that included hundreds of genes. These genes were prioritized for a variety of downstream positional cloning strategies, including artificial chromosome (BAC)-mediated cloning(Mburu *et al.*, 2003; A. Wang *et al.*, 1998), the sequencing of genes based on their functional relatedness or expression in the auditory system(X. Liu *et al.*, 2010), or candidate gene selection guided by hearing loss mouse models(Kurima *et al.*, 2002; Naz *et al.*, 2004). Using a linkage analysis approach, the first hearing loss locus was mapped to chromosome Xq in 1988(Wallis *et al.*, 1988), which led to the identifying *POU3F4* (MIM: 300039) as the first deafness gene several years later(de Kok *et al.*, 1995). Subsequently, Leon *et al* (1992) revealed the first autosomal deafness locus, *DFNA1*, where *DIAPH1* was found to be the causative hearing loss gene shortly thereafter(Lynch *et al.*, 1997). In the 1997, Kelsell *et al* (1997) first described *GJB2* as the *DFNB1* and *DFNA3* gene, a monumental discovery that led to great interest in hearing loss genetics. Within that same year, the most common *GJB2* variant (c.35delG) was reported(Carrasquillo *et al.*, 1997). Importantly, the *GJB2* c.35delG variant was identified with Belgian, the UK, and the American population, and was revealed that this variant resided on an

ancient founder haplotpye that arose in European many centuries ago(Van Laer *et al.*, 2001). These findings led to an international, multicenter collaboration that screened *GJB2* in ~1,500 affected subjects(Snoeckx *et al.*, 2005). This study found that *GJB2* c.35delG accounted for 65% of mutant alleles; although, 90% of participants were of European descent, with a carrier rate of 2-4%(Snoeckx *et al.*, 2005). Given that the mapping of these early loci and genes largely predated the human genome project, this painstaking process was of great value to the global hearing loss genetics field. Due to the limitations of molecular genetics tools of the time, many physical genetic maps contained significant errors that sometimes produced incorrect linkage mapping(Snoeckx *et al.*, 2004).

Other early pioneering hearing loss gene discoveries heavily relied on reverse genetic approaches using the mouse (*Mus musculus*) model organism. The identification of *MYO7A* (MIM: 276903) as the USH type 1B (MIM: 276900) gene exemplifies the important role of reverse genetics in human hereditary hearing loss. By investigating the *shaker-1* mouse model that was characterized by circling and head-tossing due to vestibular dysfunction, as well as progressive hearing loss, Weil *et al* (1995) identified a YAC subclone that hybridized to a 2-kb region on chromosome 7. This subclone was comprised of several *Myo7a* exons, which was homologous to the human *USH1B* locus at 11q13.5, prompting the sequencing of the *MYO7A* gene in several affected families. These sequencing efforts revealed five different pathogenic *MYO7A* variants in five unrelated families, identifying the genetic basis of USH type 1B. This discovery provided significant evidence of the role of *MYO7A* in hereditary hearing loss, which established its role in non-syndromic *DFNB2* (MIM:600060; Guilford *et al.*, 1994). In a similar approach, Avraham *et al* (1995), investigated the *Snell's waltzer* deafness mouse model. This mouse model harbored a radiation-induced inversion of chromosome 9 that was 2 cM in size.

Sequencing of inversion breaks identified a transcribed region that shared 89% sequence homology with porcine *Myo6*. Like the *shaker-1*, the *Snell's waltzer* model also displayed features, which were consistent with hearing loss and vestibular dysfunction. These data generated much interest in human *MYO6* (MIM: 600970), which lead to solving the *DFNB37* (MIM: 607821) and *DFNA22* (MIM: 606346) hearing loss loci.

1.2.4 Next-Generation Sequencing Technologies

In the early 2010s, many high throughput NGS technologies were commercially available; namely, Pacific Bioscience, Roche 454, Ion Torrent and Illumina(Loman *et al.*, 2012). However, the Ion Torrent and Illumina platforms are predominantly used today (reviewed in Levy & Myers, 2016). Ion Torrent NGS offers both targeted and WES platforms. Importantly, Ion Torrent library capture is PCR-based by multiplexing hundreds of thousands of primers that are designed against according to the desired target regions. Once amplified, targeted or whole exome libraries are then ligated on to an individual ion sphere particle, which are then loaded onto a semiconductor chip. Millions of wells are present within the semiconductor chip, each of which contains an individual ion sphere particle, and a PCR template stand for downstream sequencing. During the sequencing process, each of the four nucleotides are flowed into each well that allows for the synthesis of a complementary strand of DNA. Crucially, as each nucleotide is incorporated, protons (H⁺) are released from the growing complementary DNA strand. In turn, this process changes the pH within the well, which is measured and recorded as a specific base pair that was incorporated in the complementary DNA template stand. In contrast, Illumina offers targeted, WES and WGS solutions for gene discovery, and exploits a different approach to sequencing. Firstly, DNA sample preparation is

drastically different, relative to Ion Torrent. This process involves the fragmentation of genomic DNA through sonication or enzymatic digestion. Subsequently, universal adaptor sequences are ligated to the ends of the fragmented DNA molecules, followed by clustering on a glass flow cell slide. Each of the flow cell lanes are coated with lawns of oligonucleotides that are complementary to the universal adaptor sequences. Clustering involves the hybridization of the DNA library to the oligonucleotide lawns, which are then amplified by a polymerase into clonal “clusters”. After clustering, sequencing begins with the extension of the universal primer to produce the first read. With each cycle, fluorescently labeled nucleotides compete for addition to the growing sequence. Only one is incorporated based on the sequence of the template strand. After nucleotide incorporation, the cluster are excited by a light source and specific fluorescent signal is emitted, corresponding to a specific nucleotide – a method referred to as “Sequencing-by-synthesis”.

While WES and targeted sequencing strategies have drastically improved the molecular diagnosis of disease in the clinical setting, many cases go unresolved. From a genomics lens, this problem arises due to a significant limitation and biases in WES and targeted sequence capture platforms. For example, The Ion Torrent AmpliSeq RDY Exome (Thermo-Fisher-Scientific, 2019) and the Illumina TruSeq DNA Exome (Illumina, 2019) kits are designed to capture +96% of genes contained within the Consensus CDS (CCDS) project (Damiani *et al.*, 2016; Du *et al.*, 2016; O'Leary *et al.*, 2016; Q. Wang *et al.*, 2017). Even though this is a significant improvement over traditional Sanger sequencing methods, approximately +3% of genes within the CCDS project, and the over 8 million expressed sequence tags (ESTs) are not sequenced using this approach. However, one approach to overcoming these challenges is to perform WGS, which covers >98% of the entire genome with no biases towards specific

databases, and is more sensitive in detecting pathogenic variants in gene discovery projects (Belkadi *et al.*, 2015).

1.3 Summary and Study Goals

Hearing loss is one of the most common hereditary conditions, affecting millions of people around the world. Despite the identification of over 143 hearing loss genes, many loci remain unsolved. Although an estimated 1/3 of genes within autosomal dominant loci have evaded discovery (Van Camp G, 2015), the availability of large extended families with well-described phenotypes can reduce the burden of broad critical regions and a diverse human genome. In addition to relying on several genomic technologies, such as NGS and linkage analysis, incorporating functional RNA analyses will aid in deciphering the complexity of the transcriptome. This research project is a culmination of two decades of work to solve the genetic basis of several types of hearing loss. Throughout this thesis, we aim to identify and discover the genetic basis of autosomal dominant and recessive sensorineural hearing loss, and its syndromes, by exploiting a comprehensive multi-omic approach, discerning the functional consequences of genomic variation at the RNA-level and a thorough examination of hearing phenotypes.

1.4 Bibliography

- Abdelfatah, N., McComiskey, D. A., Doucette, L., Griffin, A., Moore, S. J., Negrijn, C., . . . Young, T. L. (2013). Identification of a novel in-frame deletion in KCNQ4 (DFNA2A) and evidence of multiple phenocopies of unknown origin in a family with ADSNHL. *Eur J Hum Genet*, *21*(10), 1112-1119. doi:10.1038/ejhg.2013.5
- Abdelfatah, N., Merner, N., Houston, J., Benteau, T., Griffin, A., Doucette, L., . . . Young, T. L. (2013). A novel deletion in SMPX causes a rare form of X-linked progressive hearing loss in two families due to a founder effect. *Hum Mutat*, *34*(1), 66-69. doi:10.1002/humu.22205
- Ahmed, Z. M., Li, X. C., Powell, S. D., Riazuddin, S., Young, T. L., Ramzan, K., . . . Wilcox, E. R. (2004). Characterization of a new full length TMPRSS3 isoform and identification of mutant alleles responsible for nonsyndromic recessive deafness in Newfoundland and Pakistan. *BMC Med Genet*, *5*, 24. doi:10.1186/1471-2350-5-24
- Annunen, S., Korkko, J., Czarny, M., Warman, M. L., Brunner, H. G., Kaariainen, H., . . . Alakokko, L. (1999). Splicing mutations of 54-bp exons in the COL11A1 gene cause Marshall syndrome, but other mutations cause overlapping Marshall/Stickler phenotypes. *Am J Hum Genet*, *65*(4), 974-983. doi:10.1086/302585
- Ansley, S. J., Badano, J. L., Blacque, O. E., Hill, J., Hoskins, B. E., Leitch, C. C., . . . Katsanis, N. (2003). Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. *Nature*, *425*(6958), 628-633. doi:10.1038/nature02030
- Arnett, J., Emery, S. B., Kim, T. B., Boerst, A. K., Lee, K., Leal, S. M., & Lesperance, M. M. (2011). Autosomal dominant progressive sensorineural hearing loss due to a novel

mutation in the KCNQ4 gene. *Arch Otolaryngol Head Neck Surg*, 137(1), 54-59.

doi:10.1001/archoto.2010.234

Avraham, K. B., Hasson, T., Steel, K. P., Kingsley, D. M., Russell, L. B., Mooseker, M. S., . . .

Jenkins, N. A. (1995). The mouse Snell's waltzer deafness gene encodes an unconventional myosin required for structural integrity of inner ear hair cells. *Nat Genet*, 11(4), 369-375. doi:10.1038/ng1295-369

Baert, A., Machackova, E., Coene, I., Cremin, C., Turner, K., Portigal-Todd, C., . . . Claes, K.

B. M. (2018). Thorough in silico and in vitro cDNA analysis of 21 putative BRCA1 and BRCA2 splice variants and a complex tandem duplication in BRCA2 allowing the identification of activated cryptic splice donor sites in BRCA2 exon 11. *Hum Mutat*, 39(4), 515-526. doi:10.1002/humu.23390

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2), 215-233. doi:10.1016/j.cell.2009.01.002

Bear, J. C., Nemeč, T. F., Kennedy, J. C., Marshall, W. H., Power, A. A., Kolonel, V. M., &

Burke, G. B. (1987). Persistent genetic isolation in outport Newfoundland. *Am J Med Genet*, 27(4), 807-830. doi:10.1002/ajmg.1320270410

Bear, J. C., Nemeč, T. F., Kennedy, J. C., Marshall, W. H., Power, A. A., Kolonel, V. M., &

Burke, G. B. (1988). Inbreeding in outport Newfoundland. *Am J Med Genet*, 29(3), 649-660. doi:10.1002/ajmg.1320290324

Belancio, V. P., Hedges, D. J., & Deininger, P. (2006). LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res*, 34(5), 1512-1521.

doi:10.1093/nar/gkl027

- Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., . . . Abel, L. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*, *112*(17), 5473-5478.
doi:10.1073/pnas.1418631112
- Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet*, *12*(12), 846-860. doi:10.1038/nrg3079
- Black, D. L. (2000). Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell*, *103*(3), 367-370.
- Blattler, A., & Farnham, P. J. (2013). Cross-talk between site-specific transcription factors and DNA methylation states. *J Biol Chem*, *288*(48), 34287-34294.
doi:10.1074/jbc.R113.512517
- Booth, K. T., Askew, J. W., Talebizadeh, Z., Huygen, P. L. M., Eudy, J., Kenyon, J., . . . Smith, S. D. (2018). Splice-altering variant in COL11A1 as a cause of nonsyndromic hearing loss DFNA37. *Genet Med*. doi:10.1038/s41436-018-0285-0
- Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*, *33* Suppl, 228-237. doi:10.1038/ng1090
- Boycott, K. M., Parboosingh, J. S., Chodirker, B. N., Lowry, R. B., McLeod, D. R., Morris, J., . . . Innes, A. M. (2008). Clinical genetics and the Hutterite population: a review of Mendelian disorders. *Am J Med Genet A*, *146A*(8), 1088-1098.
doi:10.1002/ajmg.a.32245
- Burke, J. F., Bright, K. E., Kellett, E., Benjamin, P. R., & Saunders, S. E. (1992). Alternative mRNA splicing in the nervous system. *Prog Brain Res*, *92*, 115-125.

- Canada, S. (2010). *Participation and Activity Limitation Survey 2006 Facts on Hearing Limitations 2006*. Ottawa, Ontario, Canada: Statistics Canada.
- Carrasquillo, M. M., Zlotogora, J., Barges, S., & Chakravarti, A. (1997). Two different connexin 26 mutations in an inbred kindred segregating non-syndromic recessive deafness: implications for genetic studies in isolated populations. *Hum Mol Genet*, 6(12), 2163-2172.
- Chai, Y., Chen, D., Sun, L., Li, L., Chen, Y., Pang, X., . . . Yang, T. (2015). The homozygous p.V37I variant of GJB2 is associated with diverse hearing phenotypes. *Clin Genet*, 87(4), 350-355. doi:10.1111/cge.12387
- Chen, M., & Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*, 10(11), 741-754. doi:10.1038/nrm2777
- Contributors1, W. C. (26 April 2014 12:00 UTC). File:DNA alternative splicing.gif. Retrieved from https://commons.wikimedia.org/w/index.php?title=File:DNA_alternative_splicing.gif&oldid=122314560
- Contributors2, W. C. (20 October 2017 14:15 UTC). File:SplicingSlide.png. Retrieved from <https://commons.wikimedia.org/w/index.php?title=File:SplicingSlide.png&oldid=263638969>
- Contributors3, W. C. (23 January 2018 22:46 UTC). File:MiRNA-biogenesis.jpg. Retrieved from <https://commons.wikimedia.org/w/index.php?title=File:MiRNA-biogenesis.jpg&oldid=282255200>

Contributors4, W. C. (1 December 2018 12:28 UTC). File:Anatomy of the Human Ear en.svg.

Retrieved from

https://commons.wikimedia.org/w/index.php?title=File:Anatomy_of_the_Human_Ear_en.svg&oldid=329963223

Contributors5, W. C. (10 October 2018 16:37 UTC). File:Cochlea-crosssection.svg. Retrieved

from <https://commons.wikimedia.org/w/index.php?title=File:Cochlea-crosssection.svg&oldid=323507750>

Contributors6, W. C. (7 December 2018 05:32 UTC). File:Tight Junction Transmembrane

Proteins.jpg. Retrieved from

https://commons.wikimedia.org/w/index.php?title=File:Tight_Junction_Transmembrane_Proteins.jpg&oldid=330533182

Contributors7, W. C. (29 November 2017 11:02 UTC). File:1408 Frequency Coding in The

Cochlea.jpg. Retrieved from

https://commons.wikimedia.org/w/index.php?title=File:1408_Frequency_Coding_in_The_Cochlea.jpg&oldid=269850966

Contributors8, W. C. (3 September 2018 11:28 UTC). File:Hair cell action.png. Retrieved from

https://commons.wikimedia.org/w/index.php?title=File:Hair_cell_action.png&oldid=318317473

Contributors9, W. C. (3 April 2017 18:26 UTC). File:Newfoundland map.png.

Contributors11, W. C. (13 July 2009 12:15 UTC). File:RNA splicing diagram en.svg. Retrieved

from

https://commons.wikimedia.org/w/index.php?title=File:RNA_splicing_diagram_en.svg&oldid=23563777

Contributors12, W. C. (3 May 2015 20:07 UTC). File:Alternative splicing.jpg. Retrieved from https://commons.wikimedia.org/w/index.php?title=File:Alternative_splicing.jpg&oldid=159614700

- Cui, J. J., Wang, L. Y., Zhu, T., Gong, W. J., Zhou, H. H., Liu, Z. Q., & Yin, J. Y. (2017). Gene-gene and gene-environment interactions influence platinum-based chemotherapy response and toxicity in non-small cell lung cancer patients. *Sci Rep*, 7(1), 5082. doi:10.1038/s41598-017-05246-8
- Damiati, E., Borsani, G., & Giacomuzzi, E. (2016). Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies. *Hum Genet*, 135(5), 499-511. doi:10.1007/s00439-016-1656-8
- de Kok, Y. J., van der Maarel, S. M., Bitner-Glindzicz, M., Huber, I., Monaco, A. P., Malcolm, S., . . . Cremers, F. P. (1995). Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. *Science*, 267(5198), 685-688.
- de Vries, Y., Lwiwski, N., Levitus, M., Kuyt, B., Israels, S. J., Arwert, F., . . . Meijers-Heijboer, H. (2012). A Dutch Fanconi Anemia FANCC Founder Mutation in Canadian Manitoba Mennonites. *Anemia*, 2012, 865170. doi:10.1155/2012/865170
- Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., & Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*, 37(9), e67. doi:10.1093/nar/gkp215
- Dominguez, D., Tsai, Y. H., Weatheritt, R., Wang, Y., Blencowe, B. J., & Wang, Z. (2016). An extensive program of periodic alternative splicing linked to cell cycle progression. *Elife*, 5. doi:10.7554/eLife.10288

- Doucette, L., Merner, N. D., Cooke, S., Ives, E., Galutira, D., Walsh, V., . . . Young, T. L. (2009). Profound, prelingual nonsyndromic deafness maps to chromosome 10q21 and is caused by a novel missense mutation in the Usher syndrome type IF gene PCDH15. *Eur J Hum Genet*, *17*(5), 554-564. doi:10.1038/ejhg.2008.231
- Drivas, T. G., Wojno, A. P., Tucker, B. A., Stone, E. M., & Bennett, J. (2015). Basal exon skipping and genetic pleiotropy: A predictive model of disease pathogenesis. *Sci Transl Med*, *7*(291), 291ra297. doi:10.1126/scitranslmed.aaa5370
- Du, C., Pusey, B. N., Adams, C. J., Lau, C. C., Bone, W. P., Gahl, W. A., . . . Adams, D. R. (2016). Explorations to improve the completeness of exome sequencing. *BMC Med Genomics*, *9*(1), 56. doi:10.1186/s12920-016-0216-3
- Duggan, A. T., Harris, A. J. T., Marciniak, S., Marshall, I., Kuch, M., Kitchen, A., . . . Poinar, H. (2017). Genetic Discontinuity between the Maritime Archaic and Beothuk Populations in Newfoundland, Canada. *Curr Biol*, *27*(20), 3149-3156 e3111. doi:10.1016/j.cub.2017.08.053
- Ebenesersdottir, S. S., Sandoval-Velasco, M., Gunnarsdottir, E. D., Jagadeesan, A., Guethmundsdottir, V. B., Thordardottir, E. L., . . . Helgason, A. (2018). Ancient genomes from Iceland reveal the making of a human population. *Science*, *360*(6392), 1028-1032. doi:10.1126/science.aar2625
- Eisenberger, T., Di Donato, N., Decker, C., Delle Vedove, A., Neuhaus, C., Nurnberg, G., . . . Bolz, H. J. (2018). A C-terminal nonsense mutation links PTPRQ with autosomal-dominant hearing loss, DFNA73. *Genet Med*, *20*(6), 614-621. doi:10.1038/gim.2017.155

- ENCODE. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57-74. doi:10.1038/nature11247
- Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R. K., & Hauke, J. (2018). Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med Genomics*, 11(1), 35. doi:10.1186/s12920-018-0353-y
- Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., . . . Tress, M. L. (2014). Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet*, 23(22), 5866-5878. doi:10.1093/hmg/ddu309
- Fransen, E., Verstreken, M., Bom, S. J., Lemaire, F., Kemperman, M. H., De Kok, Y. J., . . . Van Camp, G. (2001). A common ancestor for COCH related cochleovestibular (DFNA9) patients in Belgium and The Netherlands bearing the P51S mutation. *J Med Genet*, 38(1), 61-65.
- Gebert, L. F. R., & MacRae, I. J. (2018). Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol*. doi:10.1038/s41580-018-0045-7
- Gebert, L. F. R., & MacRae, I. J. (2019). Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol*, 20(1), 21-37. doi:10.1038/s41580-018-0045-7
- Gelfand, S. A. (2011). *Essentials of Audiology*: Thieme.
- Grabowski, P. J., & Black, D. L. (2001). Alternative RNA splicing in the nervous system. *Prog Neurobiol*, 65(3), 289-308.
- Green, E. D., Watson, J. D., & Collins, F. S. (2015). Human Genome Project: Twenty-five years of big biology. *Nature*, 526(7571), 29-31. doi:10.1038/526029a

- Guilford, P., Ayadi, H., Blanchard, S., Chaib, H., Le Paslier, D., Weissenbach, J., . . . Petit, C. (1994). A human gene responsible for neurosensory, non-syndromic recessive deafness is a candidate homologue of the mouse sh-1 gene. *Hum Mol Genet*, 3(6), 989-993.
- Haider, N. B., Ikeda, A., Naggert, J. K., & Nishina, P. M. (2002). Genetic modifiers of vision and hearing. *Hum Mol Genet*, 11(10), 1195-1206.
- Helga V Toriello , S. D. S. (2016). *Hereditary Hearing Loss and Its Syndromes, 3rd edn.* London, UK: Oxford University Press / Macmillan Publishers Limited.
- Hildebrand, M. S., DeLuca, A. P., Taylor, K. R., Hoskinson, D. P., Hur, I. A., Tack, D., . . . Smith, R. J. (2009). A contemporary review of AudioGene audioprofiling: a machine-based candidate gene prediction tool for autosomal dominant nonsyndromic hearing loss. *Laryngoscope*, 119(11), 2211-2215. doi:10.1002/lary.20664
- Hildebrand, M. S., Witmer, P. D., Xu, S., Newton, S. S., Kahrizi, K., Najmabadi, H., . . . Smith, R. J. (2010). miRNA mutations are not a common cause of deafness. *Am J Med Genet A*, 152A(3), 646-652. doi:10.1002/ajmg.a.33299
- Hodgkinson, K. (Personal communication: Discussion on the ARVD5 epidemiology in Newfoundland -October 2018, October 4, 2018).
- Hodgkinson, K., Dicks, E., Connors, S., Young, T. L., Parfrey, P., & Pullman, D. (2009). Translation of research discoveries to clinical care in arrhythmogenic right ventricular cardiomyopathy in Newfoundland and Labrador: lessons for health policy in genetic disease. *Genet Med*, 11(12), 859-865. doi:10.1097/GIM.0b013e3181c20bb3
- Hu, Z., Scott, H. S., Qin, G., Zheng, G., Chu, X., Xie, L., . . . Wei, C. (2015). Revealing Missing Human Protein Isoforms Based on Ab Initio Prediction, RNA-seq and Proteomics. *Sci Rep*, 5, 10940. doi:10.1038/srep10940

- Illumina. (2019). TruSeq DNA Exome. Retrieved from <https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-exome.html>
- J. Anthony Seikel, D. G. D., Douglas W. King. (2015). *Anatomy & Physiology for Speech, Language, and Hearing* (B. M. Ed. 5 ed.). Clifton Park, NY, USA: Cengage Learning.
- Jonas, S., & Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet*, *16*(7), 421-433. doi:10.1038/nrg3965
- Keats, B. J., & Corey, D. P. (1999). The usher syndromes. *Am J Med Genet*, *89*(3), 158-166.
- Kelsell, D. P., Dunlop, J., Stevens, H. P., Lench, N. J., Liang, J. N., Parry, G., . . . Leigh, I. M. (1997). Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature*, *387*(6628), 80-83. doi:10.1038/387080a0
- Kim, E., Magen, A., & Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res*, *35*(1), 125-131. doi:10.1093/nar/gk1924
- Kimberling, W. J., Hildebrand, M. S., Shearer, A. E., Jensen, M. L., Halder, J. A., Trzuppek, K., . . . Smith, R. J. (2010). Frequency of Usher syndrome in two pediatric populations: Implications for genetic screening of deaf and hard of hearing children. *Genet Med*, *12*(8), 512-516. doi:10.1097/GIM.0b013e3181e5afb8
- Kornblihtt, A. R., Schor, I. E., Allo, M., Dujardin, G., Petrillo, E., & Munoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol*, *14*(3), 153-165. doi:10.1038/nrm3525
- Kristiansson, K., Naukkarinen, J., & Peltonen, L. (2008). Isolated populations and complex disease gene identification. *Genome Biol*, *9*(8), 109. doi:10.1186/gb-2008-9-8-109

- Kurima, K., Peters, L. M., Yang, Y., Riazuddin, S., Ahmed, Z. M., Naz, S., . . . Griffith, A. J. (2002). Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function. *Nat Genet*, *30*(3), 277-284. doi:10.1038/ng842
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921. doi:10.1038/35057062
- Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, *294*(5543), 858-862. doi:10.1126/science.1065062
- Leon, P. E., Raventos, H., Lynch, E., Morrow, J., & King, M. C. (1992). The gene for an inherited form of deafness maps to chromosome 5q31. *Proc Natl Acad Sci U S A*, *89*(11), 5181-5184.
- Leong, I. U., Stuckey, A., Lai, D., Skinner, J. R., & Love, D. R. (2015). Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med Genet*, *16*, 34. doi:10.1186/s12881-015-0176-z
- Lev Maor, G., Yearim, A., & Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends Genet*, *31*(5), 274-280. doi:10.1016/j.tig.2015.03.002
- Lev-Maor, G., Goren, A., Sela, N., Kim, E., Keren, H., Doron-Faigenboim, A., . . . Ast, G. (2007). The "alternative" choice of constitutive exons throughout evolution. *PLoS Genet*, *3*(11), e203. doi:10.1371/journal.pgen.0030203
- Levy, S. E., & Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet*, *17*, 95-115. doi:10.1146/annurev-genom-083115-022413

- Lewis, M. A., Nolan, L. S., Cadge, B. A., Matthews, L. J., Schulte, B. A., Dubno, J. R., . . . Dawson, S. J. (2018). Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes. *BMC Med Genomics*, *11*(1), 77. doi:10.1186/s12920-018-0395-1
- Liu, X., Han, D., Li, J., Han, B., Ouyang, X., Cheng, J., . . . Yuan, H. (2010). Loss-of-function mutations in the PRPS1 gene cause a type of nonsyndromic X-linked sensorineural deafness, DFN2. *Am J Hum Genet*, *86*(1), 65-71. doi:10.1016/j.ajhg.2009.11.015
- Liu, X. Z., Walsh, J., Mburu, P., Kendrick-Jones, J., Cope, M. J., Steel, K. P., & Brown, S. D. (1997). Mutations in the myosin VIIA gene cause non-syndromic recessive deafness. *Nat Genet*, *16*(2), 188-190. doi:10.1038/ng0697-188
- Liu, X. Z., Walsh, J., Tamagawa, Y., Kitamura, K., Nishizawa, M., Steel, K. P., & Brown, S. D. (1997). Autosomal dominant non-syndromic deafness caused by a mutation in the myosin VIIA gene. *Nat Genet*, *17*(3), 268-269. doi:10.1038/ng1197-268
- Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*, *30*(5), 434-439. doi:10.1038/nbt.2198
- Lupski, J. R., Belmont, J. W., Boerwinkle, E., & Gibbs, R. A. (2011). Clan genomics and the complex architecture of human disease. *Cell*, *147*(1), 32-43. doi:10.1016/j.cell.2011.09.008
- Lynch, E. D., Lee, M. K., Morrow, J. E., Welsh, P. L., Leon, P. E., & King, M. C. (1997). Nonsyndromic deafness DFNA1 associated with mutation of a human homolog of the *Drosophila* gene diaphanous. *Science*, *278*(5341), 1315-1318.

- Majava, M., Hoornaert, K. P., Bartholdi, D., Bouma, M. C., Bouman, K., Carrera, M., . . . Mortier, G. R. (2007). A report on 10 new patients with heterozygous mutations in the COL11A1 gene and a review of genotype-phenotype correlations in type XI collagenopathies. *Am J Med Genet A*, *143A*(3), 258-264. doi:10.1002/ajmg.a.31586
- Mannion, J. J., & Memorial University of Newfoundland. Institute of Social and Economic Research. (1977). *The Peopling of Newfoundland : essays in historical geography*. St. John's, Nfld.: Institute of Social and Economic Research, Memorial University of Newfoundland.
- Martin, F. N., & Clark, J. G. (2014). *Introduction to Audiology*: Pearson Education Canada.
- Mathur, P., & Yang, J. (2015). Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities. *Biochim Biophys Acta*, *1852*(3), 406-420. doi:10.1016/j.bbadis.2014.11.020
- Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., . . . Bauer, P. (2016). Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet*, *24*(10), 1515. doi:10.1038/ejhg.2016.63
- Mazzoli M, G. V. C., Valerie Newton, Giarbini N, Declau F & Agnete Parving. (2003). Recommendations for the description of genetic and audiological data for families with nonsyndromic hereditary hearing impairment. *Audiological Medicine*, *1*(2), 148-150.
- Mburu, P., Mustapha, M., Varela, A., Weil, D., El-Amraoui, A., Holme, R. H., . . . Brown, S. D. (2003). Defects in whirlin, a PDZ domain molecule involved in stereocilia elongation, cause deafness in the whirler mouse and families with DFNB31. *Nat Genet*, *34*(4), 421-428. doi:10.1038/ng1208

- McComiskey, D. (2010). *Investigation of genetic cause of hearing loss in 28 autosomal dominant families within the Newfoundland founder population*. (Masters of Science (Medicine)), Memorial University, Retrieved from <https://research.library.mun.ca/9588/>
- Mele, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., . . . Guigo, R. (2015). Human genomics. The human transcriptome across tissues and individuals. *Science*, 348(6235), 660-665. doi:10.1126/science.aaa0355
- Mencia, A., Modamio-Hoybjor, S., Redshaw, N., Morin, M., Mayo-Merino, F., Olavarrieta, L., . . . Moreno-Pelayo, M. A. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet*, 41(5), 609-613. doi:10.1038/ng.355
- Merner, N. D., Hodgkinson, K. A., Haywood, A. F., Connors, S., French, V. M., Drenckhahn, J. D., . . . Young, T. L. (2008). Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *Am J Hum Genet*, 82(4), 809-821. doi:10.1016/j.ajhg.2008.01.010
- Miga, K. H. (2015). Completing the human genome: the progress and challenge of satellite DNA assembly. *Chromosome Res*, 23(3), 421-426. doi:10.1007/s10577-015-9488-2
- Milting, H., Klauke, B., Christensen, A. H., Musebeck, J., Walhorn, V., Grannemann, S., . . . Anselmetti, D. (2015). The TMEM43 Newfoundland mutation p.S358L causing ARVC-5 was imported from Europe and increases the stiffness of the cell nucleus. *Eur Heart J*, 36(14), 872-881. doi:10.1093/eurheartj/ehu077
- Moore, S. J., Green, J. S., Fan, Y., Bhogal, A. K., Dicks, E., Fernandez, B. A., . . . Parfrey, P. S. (2005). Clinical and genetic epidemiology of Bardet-Biedl syndrome in Newfoundland:

- a 22-year prospective, population-based, cohort study. *Am J Med Genet A*, 132A(4), 352-360. doi:10.1002/ajmg.a.30406
- Morton, C. C., & Nance, W. E. (2006). Newborn hearing screening--a silent revolution. *N Engl J Med*, 354(20), 2151-2164. doi:10.1056/NEJMra050700
- Naz, S., Griffith, A. J., Riazuddin, S., Hampton, L. L., Battey, J. F., Jr., Khan, S. N., . . . Friedman, T. B. (2004). Mutations of ESPN cause autosomal recessive deafness and vestibular dysfunction. *J Med Genet*, 41(8), 591-595. doi:10.1136/jmg.2004.018523
- Nguyen, T. A., Jo, M. H., Choi, Y. G., Park, J., Kwon, S. C., Hohng, S., . . . Woo, J. S. (2015). Functional Anatomy of the Human Microprocessor. *Cell*, 161(6), 1374-1387. doi:10.1016/j.cell.2015.05.010
- Nicholson, A. W. (2014). Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip Rev RNA*, 5(1), 31-48. doi:10.1002/wrna.1195
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44(D1), D733-745. doi:10.1093/nar/gkv1189
- Ohno, K., Takeda, J. I., & Masuda, A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip Rev RNA*, 9(1). doi:10.1002/wrna.1451
- Ohno, S. (1972). So much "junk" DNA in our genome. *Brookhaven Symp Biol*, 23, 366-370.
- Okada, C., Yamashita, E., Lee, S. J., Shibata, S., Katahira, J., Nakagawa, A., . . . Tsukihara, T. (2009). A high-resolution structure of the pre-microRNA nuclear export machinery. *Science*, 326(5957), 1275-1279. doi:10.1126/science.1178705

- Organization, W. H. (2018). Deafness and hearing loss. Retrieved from <http://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- Ott, J., Wang, J., & Leal, S. M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet*, *16*(5), 275-284. doi:10.1038/nrg3908
- Pakarinen, L., Karjalainen, S., Simola, K. O., Laippala, P., & Kaitalo, H. (1995). Usher's syndrome type 3 in Finland. *Laryngoscope*, *105*(6), 613-617. doi:10.1288/00005537-199506000-00010
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, *40*(12), 1413-1415. doi:10.1038/ng.259
- Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res*, *29*(5), 1185-1190.
- Rahman, P., Jones, A., Curtis, J., Bartlett, S., Peddle, L., Fernandez, B. A., & Freimer, N. B. (2003). The Newfoundland population: a unique resource for genetic investigation of complex diseases. *Hum Mol Genet*, *12 Spec No 2*, R167-172. doi:10.1093/hmg/ddg257
- Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. *J Comput Biol*, *4*(3), 311-323. doi:10.1089/cmb.1997.4.311
- Riazuddin, S. A., Iqbal, M., Wang, Y., Masuda, T., Chen, Y., Bowne, S., . . . Katsanis, N. (2010). A splice-site mutation in a retina-specific exon of BBS8 causes nonsyndromic retinitis pigmentosa. *Am J Hum Genet*, *86*(5), 805-812. doi:10.1016/j.ajhg.2010.04.001
- Richard JH Smith, E. S., Michael S Hildebrand, and Guy Van Camp. (1999 January 19, 2014). GeneReviews: Deafness and Hereditary Hearing Loss Overview. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK1434/>

- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, *17*(5), 405-424. doi:10.1038/gim.2015.30
- Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F., & Crespi, M. (2018). Splicing regulation by long noncoding RNAs. *Nucleic Acids Res*, *46*(5), 2169-2184. doi:10.1093/nar/gky095
- Rose, P. S., Levy, H. P., Liberfarb, R. M., Davis, J., Szymko-Bennett, Y., Rubin, B. I., . . . Francomano, C. A. (2005). Stickler syndrome: clinical characteristics and diagnostic criteria. *Am J Med Genet A*, *138A*(3), 199-207. doi:10.1002/ajmg.a.30955
- Rudnicki, A., & Avraham, K. B. (2012). microRNAs: the art of silencing in the ear. *EMBO Mol Med*, *4*(9), 849-859. doi:10.1002/emmm.201100922
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, *29*(1), 308-311.
- Smith, C. W., & Nadal-Ginard, B. (1989). Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell*, *56*(5), 749-758.
- Smith, R. J., Bale, J. F., Jr., & White, K. R. (2005). Sensorineural hearing loss in children. *Lancet*, *365*(9462), 879-890. doi:10.1016/S0140-6736(05)71047-3

- Snoeckx, R. L., Huygen, P. L., Feldmann, D., Marlin, S., Denoyelle, F., Waligora, J., . . . Van Camp, G. (2005). GJB2 mutations and degree of hearing loss: a multicenter study. *Am J Hum Genet*, 77(6), 945-957. doi:10.1086/497996
- Snoeckx, R. L., Kremer, H., Ensink, R. J., Flothmann, K., de Brouwer, A., Smith, R. J., . . . Van Camp, G. (2004). A novel locus for autosomal dominant non-syndromic hearing loss, DFNA31, maps to chromosome 6p21.3. *J Med Genet*, 41(1), 11-13.
- Speicher, M., Antonarakis, S. E., & Motulsky, A. G. (2009). *Vogel and Motulsky's Human Genetics: Problems and Approaches* (4 ed.). Germany: Springer Berlin Heidelberg.
- Squires, J. (2015). *The genetic etiology of early-onset hearing loss in Newfoundland and Labrador*. (Masters of Science (Medicine)), Memorial University, Retrieved from <https://research.library.mun.ca/8454/>
- Stower, H. (2013). Alternative splicing: Regulating Alu element 'exonization'. *Nat Rev Genet*, 14(3), 152-153. doi:10.1038/nrg3428
- Sugnet, C. W., Kent, W. J., Ares, M., Jr., & Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput*, 66-77.
- Thermo-Fisher-Scientific. (2019). Ion AmpliSeq™ Exome RDY Kit 1x8. Retrieved from <https://www.thermofisher.com/order/catalog/product/A38262?SID=srch-srp-A38262>
- Tom Strachan, A. R. (2010). *Human Molecular Genetics* (4th ed.). Italy: Garland Science.
- Tompson, S. W., Bacino, C. A., Safina, N. P., Bober, M. B., Proud, V. K., Funari, T., . . . Cohn, D. H. (2010). Fibrochondrogenesis results from mutations in the COL11A1 type XI collagen gene. *Am J Hum Genet*, 87(5), 708-712. doi:10.1016/j.ajhg.2010.10.009

- Toriello, H. V., & Smith, S. D. (2013). *Hereditary Hearing Loss and Its Syndromes*: Oxford University Press.
- Ushakov, K., Rudnicki, A., & Avraham, K. B. (2013). MicroRNAs in sensorineural diseases of the ear. *Front Mol Neurosci*, 6, 52. doi:10.3389/fnmol.2013.00052
- Van Camp G, S. R. (2015). Hereditary Hearing Loss Homepage. Retrieved from <http://hereditaryhearingloss.org>
- Van Eyken, E., Van Camp, G., & Van Laer, L. (2007). The complexity of age-related hearing impairment: contributing environmental and genetic factors. *Audiol Neurootol*, 12(6), 345-358. doi:10.1159/000106478
- Van Laer, L., Coucke, P., Mueller, R. F., Caethoven, G., Flothmann, K., Prasad, S. D., . . . Van Camp, G. (2001). A common founder for the 35delG GJB2 gene mutation in connexin 26 hearing impairment. *J Med Genet*, 38(8), 515-518.
- Wallis, C., Ballo, R., Wallis, G., Beighton, P., & Goldblatt, J. (1988). X-linked mixed deafness with stapes fixation in a Mauritian kindred: linkage to Xq probe pDP34. *Genomics*, 3(4), 299-301.
- Wang, A., Liang, Y., Fridell, R. A., Probst, F. J., Wilcox, E. R., Touchman, J. W., . . . Friedman, T. B. (1998). Association of unconventional myosin MYO15 mutations with human nonsyndromic deafness DFNB3. *Science*, 280(5368), 1447-1451.
- Wang, Q., Shashikant, C. S., Jensen, M., Altman, N. S., & Girirajan, S. (2017). Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep*, 7(1), 885. doi:10.1038/s41598-017-01005-x
- Wang, Z., & Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, 14(5), 802-813. doi:10.1261/rna.876308

- Ward, A. J., & Cooper, T. A. (2010). The pathobiology of splicing. *J Pathol*, *220*(2), 152-163.
doi:10.1002/path.2649
- Weil, D., Blanchard, S., Kaplan, J., Guilford, P., Gibson, F., Walsh, J., . . . *et al.* (1995).
Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature*,
374(6517), 60-61. doi:10.1038/374060a0
- Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with
applications to RNA splicing signals. *J Comput Biol*, *11*(2-3), 377-394.
doi:10.1089/1066527041410418
- Yotova, V., Labuda, D., Zietkiewicz, E., Gehl, D., Lovell, A., Lefebvre, J. F., . . . Laberge, C.
(2005). Anatomy of a founder effect: myotonic dystrophy in Northeastern Quebec. *Hum
Genet*, *117*(2-3), 177-187. doi:10.1007/s00439-005-1298-8
- Young, T. L., Ives, E., Lynch, E., Person, R., Snook, S., MacLaren, L., . . . King, M. C. (2001).
Non-syndromic progressive hearing loss DFNA38 is caused by heterozygous missense
mutation in the Wolfram syndrome gene WFS1. *Hum Mol Genet*, *10*(22), 2509-2514.
- Zhai, G., Zhou, J., Woods, M. O., Green, J. S., Parfrey, P., Rahman, P., & Green, R. C. (2016).
Genetic structure of the Newfoundland and Labrador population: founder effects
modulate variability. *Eur J Hum Genet*, *24*(7), 1063-1070. doi:10.1038/ejhg.2015.256
- Zhang, X., Chen, M. H., Wu, X., Kodani, A., Fan, J., Doan, R., . . . Walsh, C. A. (2016). Cell-
Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral
Cortex. *Cell*, *166*(5), 1147-1162 e1115. doi:10.1016/j.cell.2016.07.025

Chapter 2: A common variant in *CLDN14* causes precipitous, prelingual sensorineural hearing loss in multiple families due to founder effect

Justin A Pater¹, Tammy Benteau¹, Anne Griffin¹, Cindy Penney¹, Susan G. Stanton², Sarah Predham¹, Bernadine Kielley³, Jessica Squires¹, Jiayi Zhou¹, Quan Li¹, Nelly Abdelfatah¹,
Darren D. O'Rielly^{1,4}, Terry-Lynn Young^{1,2,4}

1. Craig L. Dobbin Genetics Research Centre, Faculty of Medicine, Memorial University, 300 Prince Phillip Drive, St. John's, Newfoundland and Labrador, Canada, A1B 3V6
2. Communication Sciences & Disorders, Western University, Elborn College, 1201 Western Road, London, Ontario, Canada, N6G 1H1
3. Department of Education and Early Childhood Development, Government of Newfoundland and Labrador, St. John's, Newfoundland and Labrador, Canada, A1B 4J6
4. Molecular Diagnostic Laboratory, Eastern Health, Craig L. Dobbin Genetics Research Centre, Faculty of Medicine, Memorial University, 300 Prince Phillip Drive, St. John's, Newfoundland and Labrador, Canada, A1B 3V6

A shorter version of this chapter is published in Human Genetics 2017; 136: 107-118.

2.1 Co-author Statement

The following authors were responsible for the following: AG, SGS, BK and SP were recruited family members and measured phenotypic data. AG, SGS and JAP interpreted family phenotype data. JAP Sanger sequenced all candidate variants that were identified by whole exome sequencing. JAP performed fragment analysis, genotyped all family members, phased and conducted haplotype analysis. JAP performed whole exome sequencing, JAP, JZ, QL, DDO performed bioinformatic data analysis. JAP, TB, DDO TLY wrote the manuscript.

2.2 Copyright and License information:

Open Access: This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

2.3 Abstract

Genetic isolates provide unprecedented opportunities to identify pathogenic variants and explore the full natural history of clinically heterogeneous phenotypes such as hearing loss. Our clinical audiologist noticed a unique audioprofile, characterized by prelingual and rapid deterioration of hearing thresholds at frequencies >0.5 kHz in several adults from unrelated families from the island population of Newfoundland. Previously performed targeted serial Sanger sequencing of probands for known deafness alleles in this founder population (n = 23) was negative. Whole exome sequencing in four members of the largest family (R2010) identified a *CLDN14* (*DFNB29*) variant [c.488C>T; p.(Ala163Val)] that causes autosomal recessive sensorineural hearing loss. Although not associated with deafness or disease, *CLDN14* p.(Ala163Val) has been previously reported as a variant of uncertain significance (VUS). Targeted sequencing of 169 deafness probands identified one homozygote and one heterozygous carrier. Genealogical studies, cascade sequencing and haplotype analysis across four unrelated families showed that all subjects with the unique audioprofile (n = 12) were also homozygous for p.(Ala163Val) and shared a 1.4 Mb *DFNB29*-associated haplotype on chromosome 21. Most significantly, sequencing 175 NL population controls revealed 1% of the population is heterozygous for *CLDN14* p.(Ala163Val), consistent with a major founder effect in Newfoundland. The youngest *CLDN14* [c.488C>T; p.(Ala163Val)] homozygote passed newborn screening and had normal hearing thresholds up to 3 years of age, which then deteriorated to a precipitous loss >1 kHz during the first decade. Our study suggests that genetic testing may be necessary to identify at-risk children in time to prevent speech, language and developmental delay.

2.4 Introduction

Hearing loss is one of the most common and genetic of all human phenotypes. Permanent bilateral sensorineural hearing loss affects 1/500 newborns, and almost twice as many adolescents (Morton & Nance, 2006; Richard JH Smith, 1999). Although approximately two-thirds of prelingual severe hearing loss cases are recessive, only a minority of hearing loss cases with a presumed recessive inheritance pattern can be conclusively diagnosed with a clear genetic etiology (Sloan-Heggen *et al.*, 2016). Therefore, many recessive cases may be due to genetic defects in genes yet to be identified. However, recent studies using new high-throughput technologies and broader application in multi-ethnic populations report *GJB2* yields of less than 25%, suggesting a larger role for other recessive genes in prelingual severe cases (Sloan-Heggen *et al.*, 2016; Yan *et al.*, 2016).

Sensorineural hearing loss is characterized by both degree (mild, moderate, severe or profound) and configuration (low, mid and/or high frequency) using the standard behavioral audiogram. Although clinically heterogeneous, rare pathognomonic audiograms may present with surprising regularity in clinics within genetically isolated populations and where patients often share a common ancestor due to founder effects. For example, the Finnish and Pakistani populations have been invaluable for discovery of deafness genes as population bottlenecks (genetic drift) and/or inbreeding increase the likelihood of inheriting recessive alleles that are identical by descent (Ahmed *et al.*, 2004; Z. E. Bashir *et al.*, 2013; Lee *et al.*, 2012; Nayak *et al.*, 2015). These populations are often characterized by large sibships, deep genealogies and higher consanguineous rates. The NL population was founded by ~20,000 Protestant English and Roman Catholic Irish settlers. Religious and geographic isolation within small coastal fishing (outport) communities (Manion, 1977) has resulted in a higher inbreeding coefficient in

the NL population(Bear *et al.*, 1987; Zhai *et al.*, 2015). We have previously identified several founder deafness alleles in the NL populations(Abdelfatah, McComiskey, *et al.*, 2013; Abdelfatah, Merner, *et al.*, 2013; Ahmed *et al.*, 2004; Doucette *et al.*, 2009; Young *et al.*, 2001).

A unique clinical audioprofile of steeply sloping sensorineural hearing loss was noted in several unrelated families. Herein, we report a founder missense variant in *CLDN14* causing precipitous prelingual sensorineural hearing loss in children born with normal hearing thresholds. The essential role of *CLDN14*, a component of tight junctions, was first discovered through studies in consanguineous families from the genetically isolated population of Pakistan(Wilcox *et al.*, 2001). Tight junctions have been shown to play a significant role in maintaining the structural integrity of cells within the inner ear. Other genes encoding tight junction proteins, such as *MARVEL Domain Containing 2* (*MARVELD2*; *DFNB49*; Nayak *et al.*, 2015; Riazuddin *et al.*, 2006), have also been implicated in recessive hearing loss. Claudin-14 is essential for the formation of tight junctions and is expressed in both hair cells and supporting cells of the organ of Corti; however, *CLDN14* exhibits preferential gene expression in sensory hair cells over supporting cells(Ben-Yosef *et al.*, 2003; Scheffer *et al.*, 2015; Wilcox *et al.*, 2001). Initially, *CLDN14* was considered the cause of congenital recessive and profound deafness(Wilcox *et al.*, 2001) and more recently of milder forms of hearing loss(Z. E. Bashir *et al.*, 2013). The *CLDN14* c.488C>T p.(Ala163Val) allele has previously been reported in multiple studies as a variant of uncertain significance (VUS; Purcell *et al.*, 2014; Thorleifsson *et al.*, 2009; Toka *et al.*, 2013) and recently identified by Sloan-Heggen *et al* (2016) as one of two VUS in a patient with congenital hearing loss. Our study shows children inheriting two copies of the *CLDN14* c.488C>T p.(Ala163Val) allele are born with normal hearing thresholds

and experience a rapid and progressive loss by 3–4 years of age. Extensive clinical recruitment and targeted screening suggest that *CLDN14* p.(Ala163Val) represents a major founder variant in the Newfoundland population.

2.5 Materials and Methods

This project is part of a large study of hereditary hearing loss in the Canadian province of NL. Informed consent, family history and permission to access medical records and audiograms were obtained from all participants by clinicians, as per approved institutional review board protocol #01.186 (Human Research Ethics Board, St. John's, NL, Canada). Sensorineural hearing loss was determined by our clinical team when hearing thresholds were abnormal, as defined by air and bone conduction results within 10 decibels (dB) of each other (i.e., air–bone gaps of 10 dB or less). In addition, the team obtained both retrospective and prospective audiograms and health records for all study participants.

In the course of ongoing clinical recruitment, a rare but consistent clinical audioprofile characterized as steeply sloping, sensorineural hearing loss above 0.5 kHz with mid- and high-frequency thresholds in the severe to profound range (Figure 2.1a – d) was noted by our clinical audiologist, Anne Griffin. When this project first began, we were unaware of common ancestry between Families R2033, R2075 and R2010 (Figure 2.2). In fact, during our outreach field trip to Burin Peninsula, we were only focused on the clinical ascertainment of Family R2010, as well as collecting DNA samples. When examining the pedigree structure and inheritance pattern of hearing loss in Family R2010, we were of the opinion that this was an autosomal dominant form of hearing loss. The reason we believed this is because PID IV-4 has hearing loss, which is vertically transmitted to all his descendants (Figure 2.2). Moreover, we

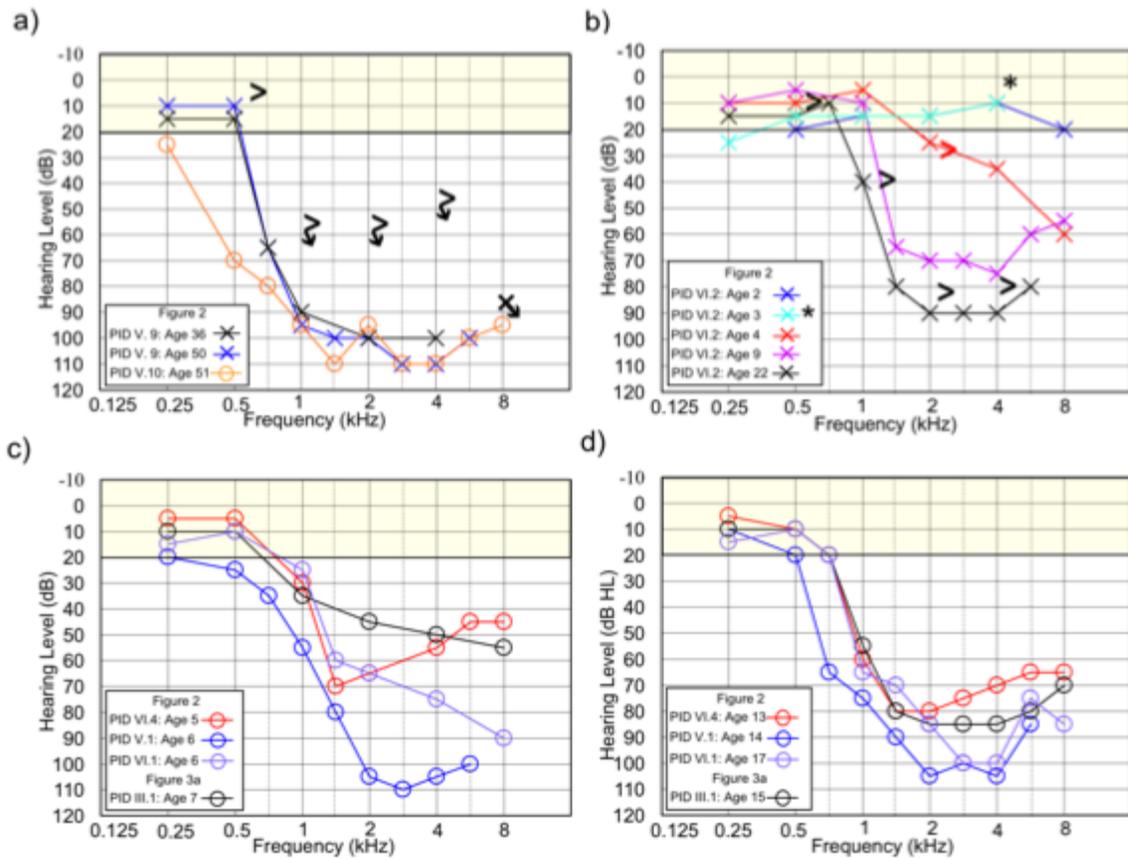


Figure 2.1. Rare, precipitous audiologic phenotype caused by *CLDN14* (c.488C>T; p.Ala163Val) in an Irish clan. a) Pure tone audiogram of Family R2010 proband (PID V-9) and sister (PID V-10), b) Pure tone audiogram series for PID VI-2 (Family R2075) showing normal hearing at age 2 years and a progressive hearing loss apparent by 4 years of age, c) and d) second decade pure tone audiogram of affected subjects. Yellow shaded area indicates range of normal hearing. Hearing thresholds are measured in decibels hearing level (dB HL), X = left ear (air conduction), O = right ear (air conduction), > = left ear (bone conduction), î = no response at the limits of the audiometer. * = 8 kHz was not measured

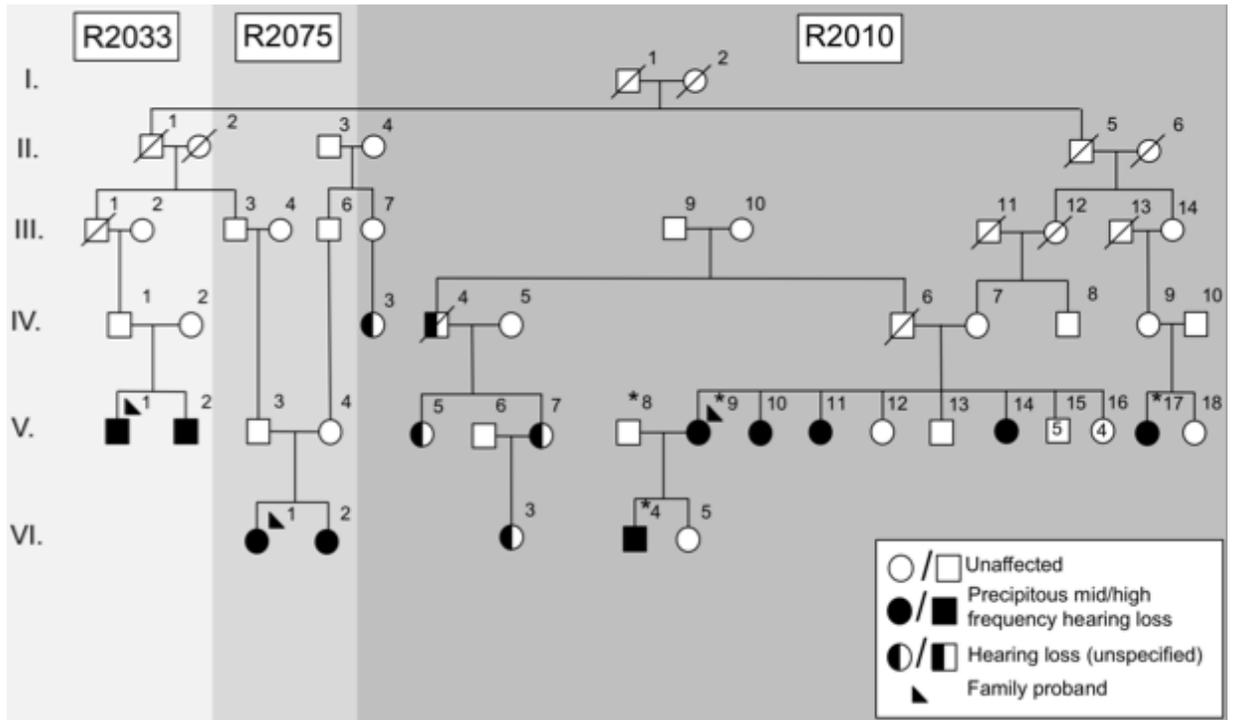


Figure 2.2. Combined pedigrees of 3 families (R2033, R2075 and R210) with rare, precipitous audiologic phenotype that connect to a common ancestor. Shaded symbols: precipitous sensorineural hearing loss. Half shaded symbols: unspecified hearing loss.

noted that a distant relative, PID VI-4, also inherited hearing loss from his mother, PID V-9, while his father, PID V-8, exhibited normal hearing thresholds across all frequencies (Figure 2.1 and Figure 2.2). In light of this strong family history, we considered that PID VI-4's deceased grandmother, PID IV-6, may have been clinically affected, which would support an autosomal dominant hearing loss inheritance pattern. Upon further inspection of other hearing loss families in our ongoing hearing loss study, our clinical audiologist noticed that this rare precipitous hearing loss phenotype was present in 2 additional families, R2033 and R2075 (Figure 2.1). Despite an apparent dominant inheritance pattern in family R2010, this precipitous hearing loss appeared to be recessive in families R2033 and R2075, which was rather puzzling during the early stages of this project. This prompted us to take a closer look at these families and consider the idea of shared ancestry between these three families. On the premise that subjects with this hearing loss pattern also shared a recent common ancestor, Anne Griffin and Sarah Predham (project genetic counsellor) used the distinct audioprofile to guide clinical recruitment.

2.5.1 DNA Preparation, Targeted Sequencing and Audioprofiling

Prior to this study, genomic DNA was extracted from peripheral blood using a simple salting out protocol (Miller *et al.*, 1988). All recruited probands were then screened for population-specific deafness alleles (Appendix C; Abdelfatah, McComiskey, *et al.*, 2013; Abdelfatah, Merner, *et al.*, 2013; Ahmed *et al.*, 2004; Doucette *et al.*, 2009; Young *et al.*, 2001). To identify other candidate genes to screen, audiograms were submitted to Audiogene (Hildebrand *et al.*, 2009) for computerized comparison with known average audiograms of 16 autosomal dominant loci, under the assumption that hearing loss was

segregating as an autosomal dominant trait in these NL families. Bidirectional Sanger sequencing (ABI PRISM 3500XL DNA Analyzer; Applied Biosystems, Foster City, CA, USA) with standard PCR assay using Primer3(Untergasser *et al.*, 2012) was used to screen candidate variants and genes(Merner *et al.*, 2008). DNA sequences were visualized using Mutation Surveyor Software (version 4.07, SoftGenetics LLC State College, PA 16803).

2.5.2 Whole-Exome Sequencing and Variant Filtration

Whole exome libraries were prepared for four members of Family R2010 using the Ion Torrent AmpliSeq RDY Exome Kit (Life Technologies, Cat. #A27193; Figure 2.2). Exome library purification, adapter ligation and barcoding were done using the Ion PI Hi-Q OT2 200 kit (Life Technologies, Cat. #A26434). Library purification was performed and quantified with the Ion Library Quantification Kit (Life Technologies, Cat. #4468802) and then loaded onto a PI v3 chip and sequenced with the Ion Torrent Proton Sequencer. Single-nucleotide variants (SNVs) and insertion/deletions (INDELS) were called (GATK, v3.5) and annotated by bioinformaticians using SnpEff (v4.1; <http://snpeff.sourceforge.net/>), and SNVs were filtered against publically available SNP databases (ExAC Browser, <http://exac.broadinstitute.org/>; dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>; 1000 genomes, <http://www.1000genomes.org>; ClinVar, <https://www.ncbi.nlm.nih.gov/clinvar/>). We assessed the impact of SNVs at the protein level with SIFT, PolyPhen-2, and MutationTaster. Filtered SNVs had a minimum of 20× coverage, a predicted moderate/high impact (nonsense, frameshift, missense, splice sites) and a minor allele frequency (MAF) of <1%. Two different variant filtration analyses were performed. In our first analysis, we filtered for all heterozygous variants that were present in affected (PID V-9, V-17 and VI-4) and absent in unaffected (PID V-9) family members (Figure

2.2), given that we initially thought that this was a dominant family. Since this rare precipitous hearing loss followed a recessive inheritance pattern in Families R2033 and R2075 (Figure 2.2), we filtered for all variants that were homozygous in affected (PID V-9, V-17 and VI-4) and heterozygous in unaffected (PID V-9) family members in our second analysis. The rationale behind conducting a recessive analysis was due the geographic location of these families on the Burin Peninsula. Carrier frequencies could be higher in this region due to founder effects resulting from consanguineous relationships, as well as religious and geographic barriers. This founder effect could certainly increase the likelihood of producing homozygous offspring. Even though pseudodominance was the more likely scenario, there wasn't enough evidence to exclude a dominant inheritance pattern. The apparent vertical transmission of hearing loss in the pedigree (Figure 2.2) could be due to either a dominant gene with reduced penetrance, or a recessive gene with a pseudodominant inheritance pattern, therefore we conducted both autosomal dominant and recessive analyses.

2.5.3 Cascade Sequencing and Haplotype Analysis

Potential pathogenic variants were subjected to cascade screening in all available relatives across three families observed to have the same rare audioprofile (Families R2010, R2033 and R2075), as well as 175 ethnically-matched controls. Microsatellites flanking candidate genes were genotyped according to standard procedures (Abdelfatah, McComiskey, *et al.*, 2013; Abdelfatah, Merner, *et al.*, 2013) and alleles were called using GeneMapper software v4.0. Haplotypes were reconstructed manually and compared across families. In addition, variants of interest were screened in 169 deafness probands with Newfoundland ancestry.

2.6 Results

2.6.1 Clinical Evaluation

Our clinical audiologist (AG) noted that probands (from Families R2010, R2075 and R2033) all shared a unique hearing loss pattern. The proband of Family R2010 (V-9; Figure 2.2) presented at 36 years of age with the characteristic pattern of normal low-frequency thresholds, steeply sloping to severe bilateral, symmetrical, sensorineural hearing loss throughout mid- and high-frequencies (Figure 2.1a). Age-appropriate audiologic tests of the proband's son (VI-4; Figure 2.2) at 1 month and 1 year of age were normal. Serial audiograms on PID VI-2 (Figure 2.1b; Family R2075) show normal hearing thresholds across frequencies up to 3 years of age, with subsequent rapid progression of hearing loss affecting high-frequencies first. Significant hearing loss of variable severity is already present in children aged 5–7 years (Figure 2.1c), which include probands of families R2033 and R2075. By the middle of the second decade of life, these children uniformly exhibit the distinctive steeply sloping audiogram (Figure 2.1d). The hearing loss progresses slowly during subsequent decades, primarily in the mid–high frequencies, with relatively well-preserved low-frequency thresholds. For adults, some variation in thresholds at 0.5 kHz is observed (PID V-10; Figure 2.1a) but otherwise the adult presentation is relatively uniform.

2.6.2 Targeted Sequencing and Audioprofiling

Prior to our study, targeted sequencing was carried out on probands for known deafness alleles (previously identified in this population; Appendix C) but none were found. Several gene candidates, as suggested by Audiogene(Hildebrand *et al.*, 2009), were also Sanger sequenced, including *COCH*, *KCNQ4* and *TMCI*. A rare variant in *TMCI* (c.421C>T; MAF of

0.01%) was identified and predicted to cause an arginine to tryptophan substitution at position 141 of the protein amino acid sequence. The variant was predicted to be deleterious by SIFT and probably damaging by PolyPhen-2 and Panther. Although identified in both the proband (PID V-9) of Family R2010 and her son (PID VI-4), the c.421C>T variant did not co-segregate with mid-high-frequency loss (Figure 2.3).

2.6.3 Whole-Exome Sequencing

Whole exome sequencing on Family R2010 using three affecteds (V-9, VI-4 and V-17) and one unaffected parent (V-8) (Figure 2.2) yielded >35,000 total variants. Under a dominant model, 34 heterozygous variants were filtered by our bioinformaticians (Table 2.1). However, none of these variants resided within known deafness genes/loci (<http://hereditaryhearing-loss.org/>). Under a recessive model, they filtered four homozygous variants (Table 2.2) One of these, *CLDN14* (*DFNB29*) is a known deafness gene expressed in the sensory epithelium of the organ of Corti of the inner ear (Scheffer *et al.*, 2015; Wilcox *et al.*, 2001). *CLDN14* consists of three exons and two isoforms, and encodes a protein containing four transmembrane domains. The *CLDN14* p.(Ala163Val) point variant (Figure 2.4) identified in Family R2010 predicts substitution of an alanine to a valine at the beginning of the fourth transmembrane domain (Figure 2.5) and is highly conserved (Figure 2.6). The *CLDN14* c.488C>T allele was first identified in the Icelandic population (Thorleifsson *et al.*, 2009). Globally, the *CLDN14* c.488C>T variant has an MAF of 0.02564% (ExAC Browser, <http://exac.broadinstitute.org/>) and has been reported in both European and African populations. The heterozygous *CLDN14* allele (human GRCh37/hg19: g. 37833506 G>A, NM_012130.3: c.488 C>T) is reported as a variant of uncertain significance in dbSNP (rs143797113) and ClinVar

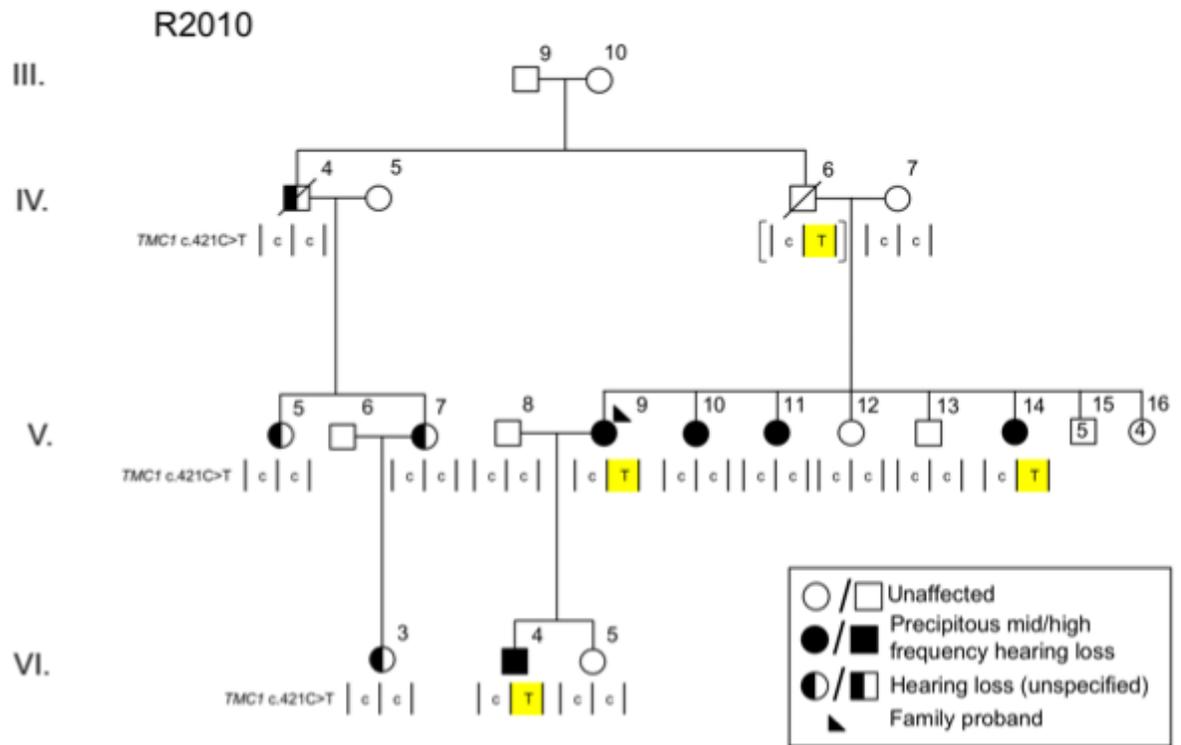


Figure 2.3. *TMC1* c.421C>T Segregation Analysis in Family R2010. This is a Sub-pedigree of the *CLDN14* Newfoundland illustrating that *TMC1* c.421C>T does not segregate with hearing loss. Shaded symbols: precipitous sensorineural hearing loss. Half shaded symbols: unspecified hearing loss.

Table 2.1. Thirty-one Heterozygous variants identified in PID V-9, VI-4 and V-17 (Family R2010; Figure 2). None of these variants resided within known deafness genes/loci or were functionally unrelated to hereditary hearing loss and was eliminated from downstream analyses.

Chromosome	Gene	Position	dbSNP	REF	ALT	ExAC MAF	Impact
1	<i>ZMYND12</i>	42902142	rs143810024	T	C	0.0032	MODERATE
1	<i>OR2L8</i>	248112762	-	TG	CA	-	MODERATE
1	<i>OR2T3</i>	248637416	-	GC	CT	-	MODERATE
2	<i>SLC4A3</i>	220494117	-	CA	C,CC	-	MODERATE
3	<i>GALNT15</i>	16260935	-	G	A	0.00004122	MODERATE
3	<i>LRIG1</i>	66433735	rs116809906	G	A	0.0099	MODERATE
3	<i>OR5K4</i>	98073241	rs143234586	C	G	0.0001	MODERATE
3	<i>NFKBIZ</i>	101571753	rs61730026	A	G	0.0014	MODERATE
4	<i>LEF1</i>	109086280	rs61752607	C	T	0.0084	MODERATE
4	<i>RFC1</i>	39306530	rs28903096	C	A	0.0004	MODERATE
4	<i>SHROOM3</i>	77677759	rs143947261	C	A	0.0007	MODERATE
4	<i>SEC24B</i>	110459736	rs199667887	A	G	0.0000911	MODERATE
5	<i>DMGDH</i>	78322282	rs113405129	G	A	0.00009885	MODERATE
5	<i>RASGRF2</i>	80390172	rs199615089	G	A	0.0002	MODERATE
6	<i>GUCA1B</i>	42153428	rs139923590	C	A	0.0087	MODERATE
10	<i>CHST15</i>	125804207	-	T	A	0.000008323	MODERATE
10	<i>DMBT1</i>	124380869	-	CAC	AGT	-	MODERATE
14	<i>ACOT4</i>	74060512	-	AGCTTA	TCAAAG	-	MODERATE
15	<i>NUSAP1</i>	41634587	-	AC	GA	-	MODERATE
16	<i>WDR90</i>	711429	rs199527716	C	T	0.0035	MODERATE
16	<i>IFT140</i>	1652418	rs146128830	C	T	0.0069	MODERATE
16	<i>TBL3</i>	2025080	rs147896871	G	A	0.001	MODERATE
16	<i>HYDIN</i>	71007809	rs783762	C	T	-	MODERATE
16	<i>HYDIN</i>	70975565	rs18157-7	T	C	-	MODERATE
17	<i>PFAS</i>	8167600	-	TG	CG,GC	-	MODERATE
17	<i>PFAS</i>	8167600	-	TG	CG,GC	-	MODERATE
19	<i>CLDND2</i>	51871194	rs61736500	C	T	0.0052	MODERATE
19	<i>PSG8</i>	43269704	-	TG	CA	-	MODERATE
19	<i>LILRB2</i>	54778617	rs143326661	T	C	0.0007	MODERATE
21	<i>KRTAP10-10</i>	46057417	-	AG	CC	-	MODERATE
22	<i>NDUFA6</i>	42482252	rs113437301	G	A	0.0028	MODERATE

Table 2.2. Four Homozygous variants identified in PID V-9, VI-4 and V-17 (Family R2010; Figure 2). No reports exist of *CUL7*, *PRKDC* and *ZNF404* being involved in hearing loss. Additionally, these three genes did not reside within known deafness genes/loci. *CLDN14* is a known hearing loss gene that causes DFNB29, which perfectly segregated within a Newfoundland clan and exhibited pseudodominance.

Chromosome	Gene	Position	dbSNP	REF	ALT	ExAC MAF	Impact
6	<i>CUL7</i>	43014298	-	TT	CC	-	MODERATE
8	<i>PRKDC</i>	48805816	rs370106149	A	AG,G	0.00003419	MODERATE
19	<i>ZNF404</i>	44377357	-	C	A,T	-	MODERATE
21	<i>CLDN14</i>	37833506	rs143797113	G	A	0.0003	MODERATE

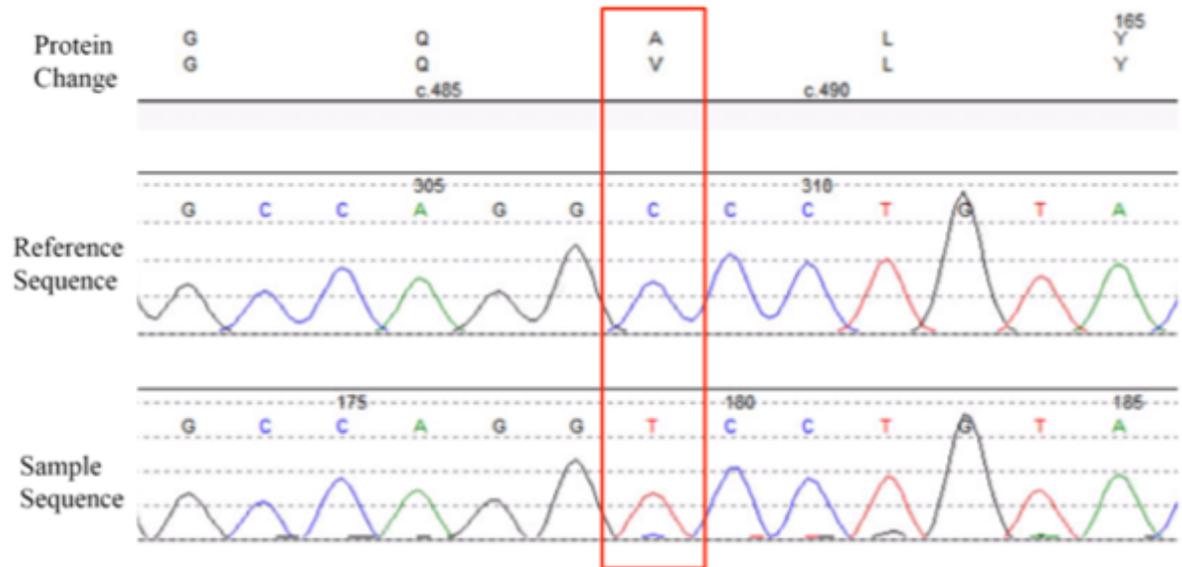


Figure 2.4. Sequence electropherogram of *CLDN14* (c.488C>T; p.Ala163Val). Red box highlights variant site.

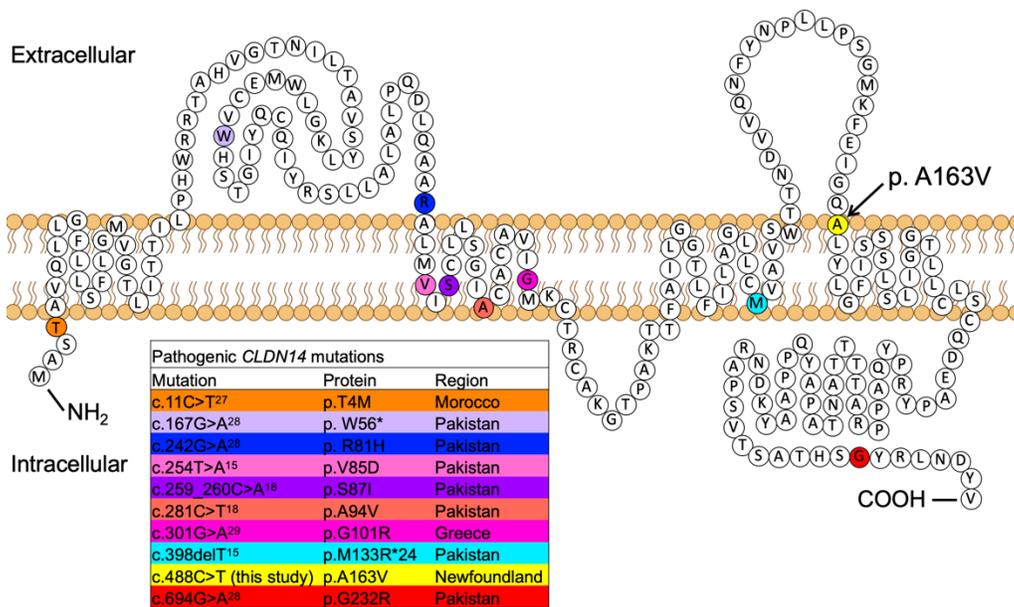


Figure 2.5. Location of pathogenic variants in Claudin-14. Colored amino acid residues indicate previously reported claudin-14 alleles. Arrow indicates position of *CLDN14* c.488C>T (p.Ala163Val). Adapted from: Bashir, Z.E., Latief, N., Belyantseva, I.A., Iqbal, F., Riazuddin, S.A., Khan, S.N. *et al.* Phenotypic variability of *CLDN14* variants causing *DFNB29* hearing loss in the Pakistani population. *J Hum Genet.* 58, 102-108 (2013). This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The article can be found at the following URL: <https://www.nature.com/articles/jhg2012143>

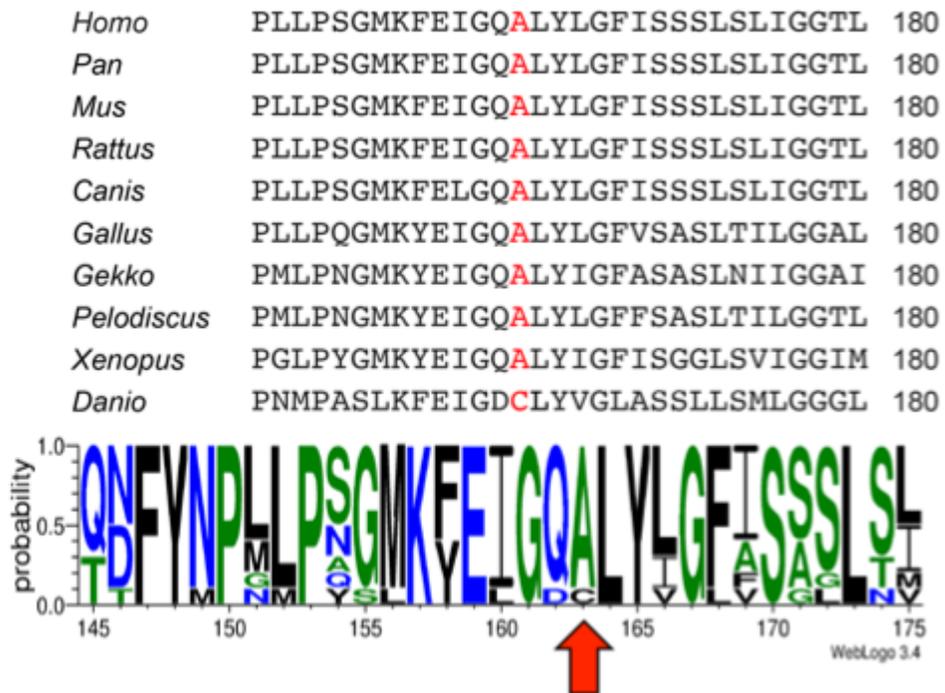


Figure 2.6. Conservation of the Claudin-14 protein using Clustal Omega and WebLogo display. *Homo sapiens* (NP_001139551.1), *Pan paniscus* (XP_008975916.1), *Mus musculus* (NP_001159398.1), *Rattus norvegicus* (NP_001013447.1), *Canis lupus familiaris* (XP_013965166.1), *Gallus gallus* (XP_015155717.1), *Gekko japonicas* (XP_015277878.1), *Pelodiscus sinensis* (XP_006126056.1), *Xenopus laevis* (NP_001086045.1), *Danio rerio* (NP_001004559.2). Red font and arrow indicates a highly conserved alanine residue at position 163

(Variation ID: 228519), and has been submitted to large scale sequencing projects, including ExAC browser (MAF: 0.02564%), 1000 genomes (MAF: 0.04%), and the Grand Opportunity Exome Sequencing Project (MAF: 0.05%). In addition, this allele has been reported in several control samples from other study cohorts within the USA (Toka *et al.*, 2013), Sweden (Purcell *et al.*, 2014), and Africa (ExAC browser). The majority of known pathogenic *CLDN14* variants reside within one of the transmembrane domains in Claudin-14 (Figure 2.5; R. Bashir *et al.*, 2010; Charif *et al.*, 2013; Lee *et al.*, 2012; Wattenhofer *et al.*, 2005; Wilcox *et al.*, 2001)). Functional studies of pathogenic *CLDN14* variants have demonstrated the importance of transmembrane domains with respect to protein topology and folding, as well as proper spatial localization within cells. For example, previous localization experiments have shown that the p.Val85Asp and p.Gly101Arg deafness variants within domain II (Figure 2.5) fail to form tight junctions due to the mislocalization of claudin-14 protein to the cytoplasm, in vitro (Wattenhofer *et al.*, 2005). Since p.(Ala163Val) is predicted to change a highly conserved amino acid within the fourth transmembrane domain (Figures 2.4 – 6), we suspect a similar impact regarding the spatial localization of claudin-14 to the plasma membrane, leading to the cells' inability to form tight junctions. While previous research demonstrated the importance of amino acid conservation within claudin-14 transmembrane domains, experimental functional studies are needed to prove *CLDN14* c.488C>T, p.(Ala163Val) pathogenicity. The *CLDN14* gene is essential in maintaining auditory function, as it has been identified as a critical component of tight junctions (Figure 2.7), which play a critical role in maintaining the electrochemical gradient observed in the organ of Corti. Briefly, hair cell stereocilia are bathed in potassium-rich endolymph, while the basolateral surface of the hair cell body is surrounded by an intercellular (or extracellular) fluid continuous with the perilymph. This electrochemical

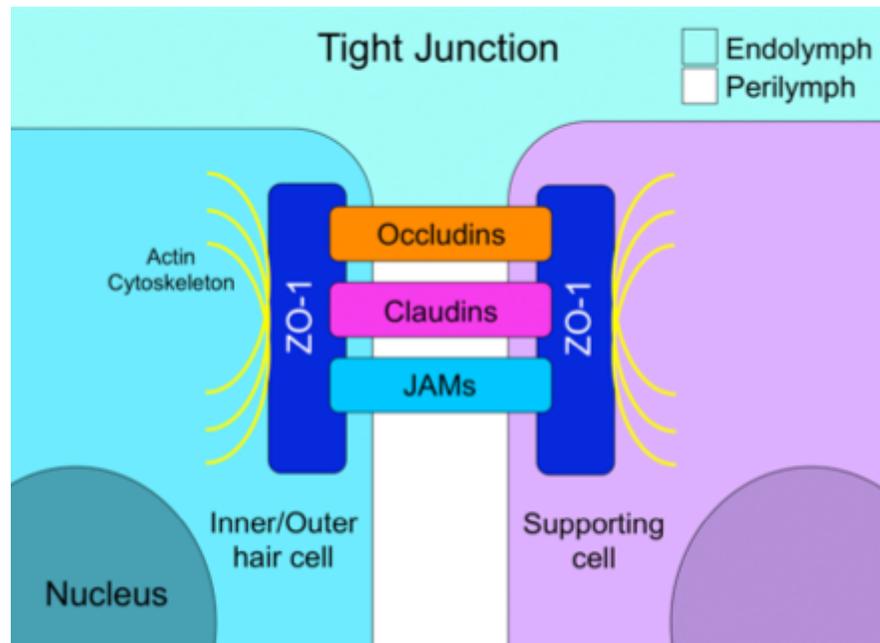


Figure 2.7. Schematic diagram demonstrating the molecular structure of tight junctions. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors6)

gradient is maintained by the reticular lamina, where *CLDN14* plays a key role in forming tight junctions (Figure 2.8).

2.6.4 Cascade Sequencing and Haplotype Analysis

Cascade sequencing in all available subjects from Families R2033 and R2075 showed that affecteds with the distinct precipitous mid–high-frequency hearing loss (Figure 2.9, filled symbols) were also homozygous for *CLDN14* p.(Ala163Val) (Figure 2.4). Subjects with a flat loss, such as PID IV-4 and his descendants (V-5, V-7 and VI-4) lacked the recessive *CLDN14* variant (Figure 2.10). This pattern is consistent with our hypothesis that *CLDN14* p.(Ala163Val) is a likely pathogenic, recessive allele where homozygosity results in a distinct precipitous mid–high-frequency hearing loss. Relatives inheriting a single copy (carriers) or wild type do not have this pattern. As per to the ACMG standards and guidelines (Richards *et al.*, 2015), *CLDN14* p.(Ala163Val) is classified as Pathogenic variant (PS4, PM1, PM2, PM3, PP1, PP2, PP3, PP4).

The cause of hearing loss in subjects with flat audioprofiles is not known, but is clearly not due to homozygosity for *CLDN14* c.488C>T. Future studies will explore the genetic etiology of their hearing loss. Furthermore, screening our cohort of 169 deafness probands identified an additional homozygous subject (Family R2072) and two heterozygous carriers. In Family R2072, the proband's mother (with the distinct audioprofile) was also found to be homozygous for *CLDN14* p.(Ala163Val) and her father a carrier (Figure 2.11). Screening population controls identified four carriers out of 175 subjects, estimating an MAF of 1.15% in the Newfoundland population and suggesting that this likely pathogenic variant is not rare.

Extensive genotyping in the vicinity of *DFNB29* revealed that p.(Ala163Val) resides on

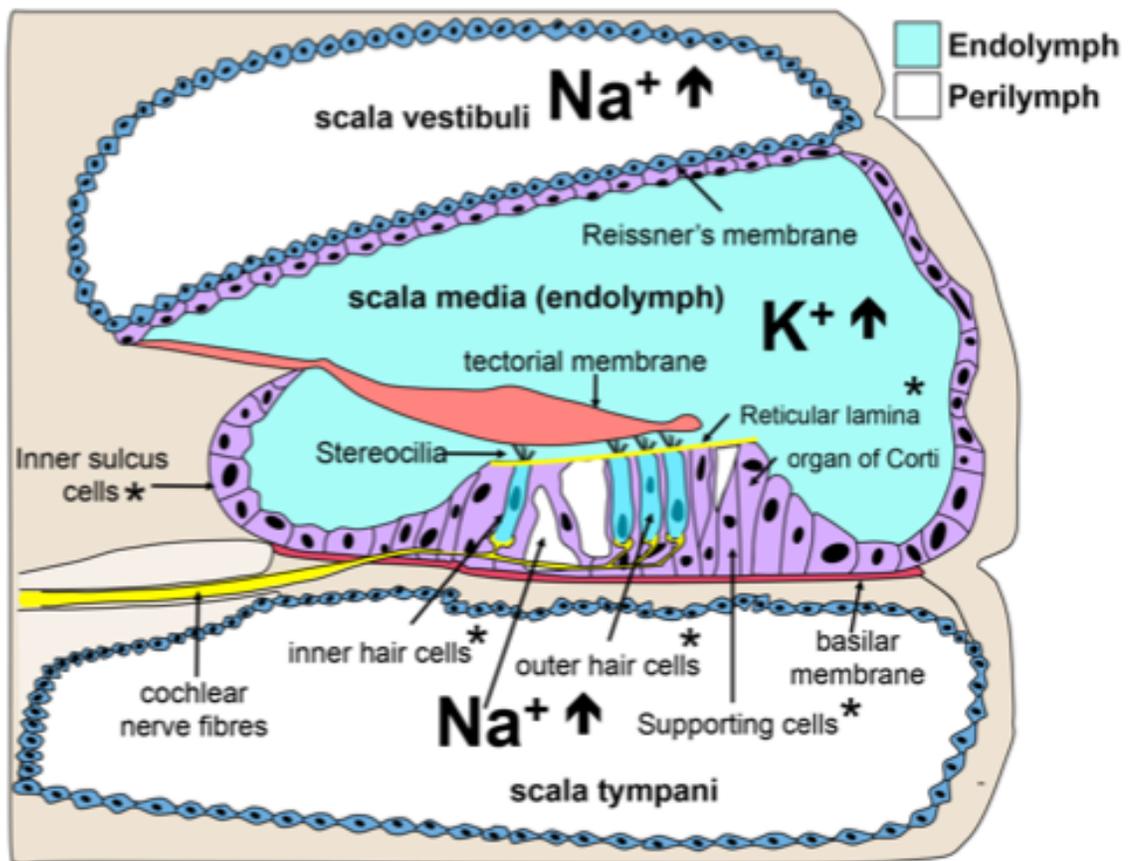


Figure 2.8. Cross-sectional diagram illustrating the anatomical location of the cochlear canals and their respective ionic composition. * denotes *CLDN14* expression. This image is licensed under Creative Commons Attribution 4.0 International License, adapted from(Contributors5; Contributors6)

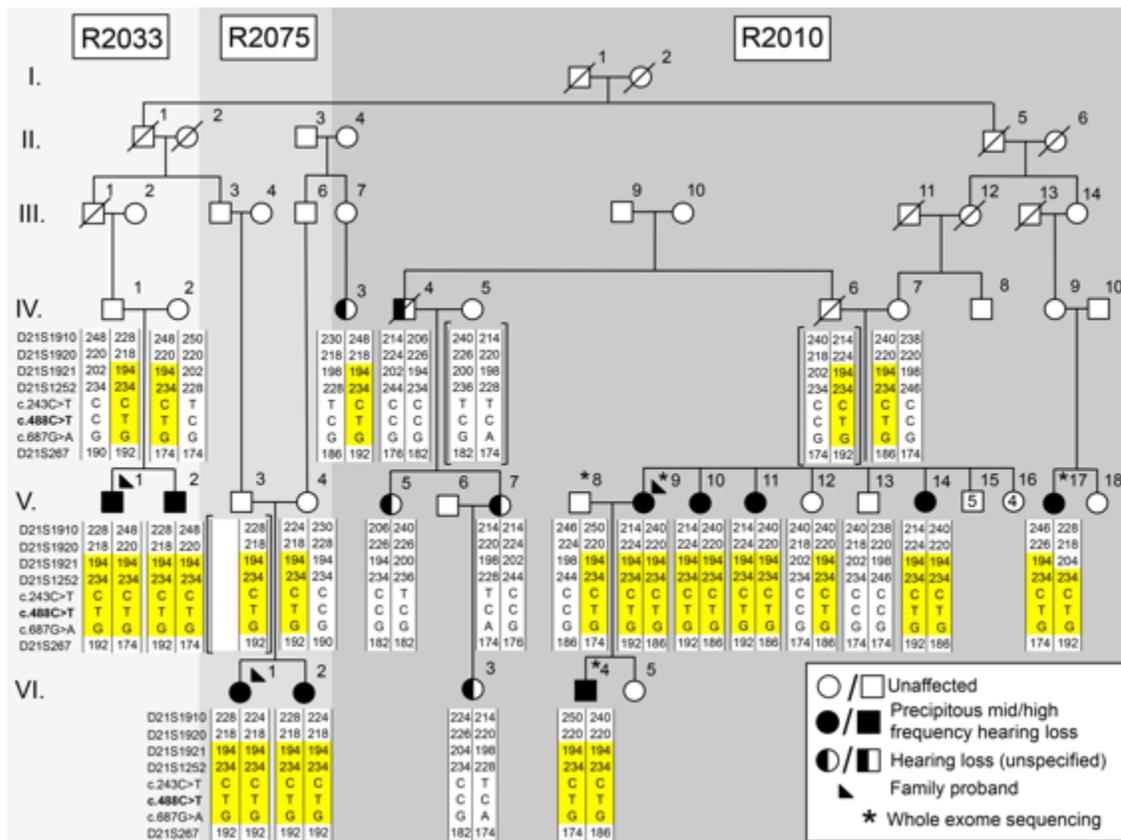


Figure 2.9. Combined pedigrees of 3 families (R2033, R2075 and R210) with rare, precipitous audiologic phenotype connect to a founding couple and share an ancestral DFNB29-associated haplotype. Shaded symbols: precipitous sensorineural hearing loss. Half shaded symbols: unspecified hearing loss.

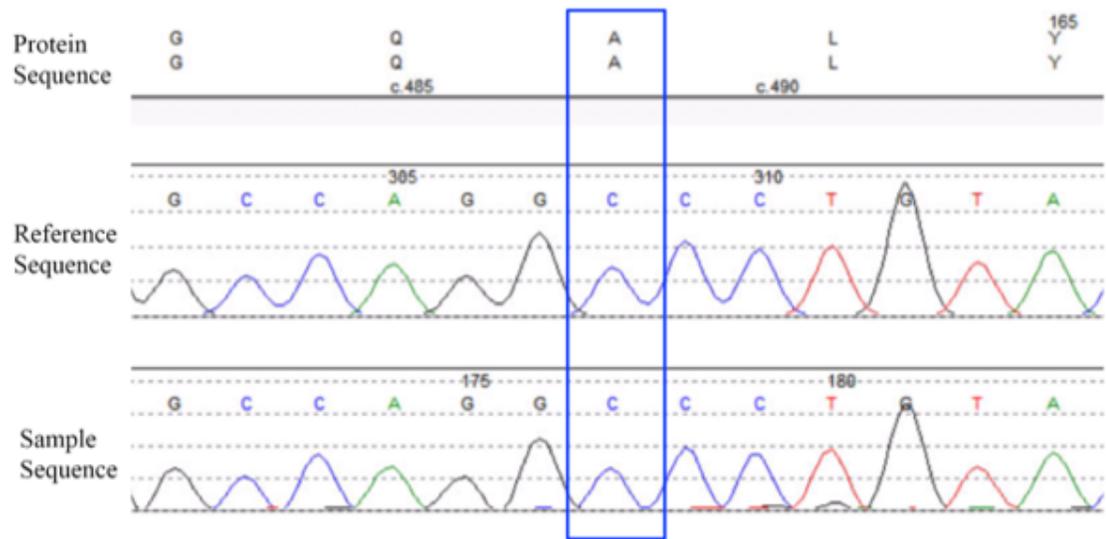
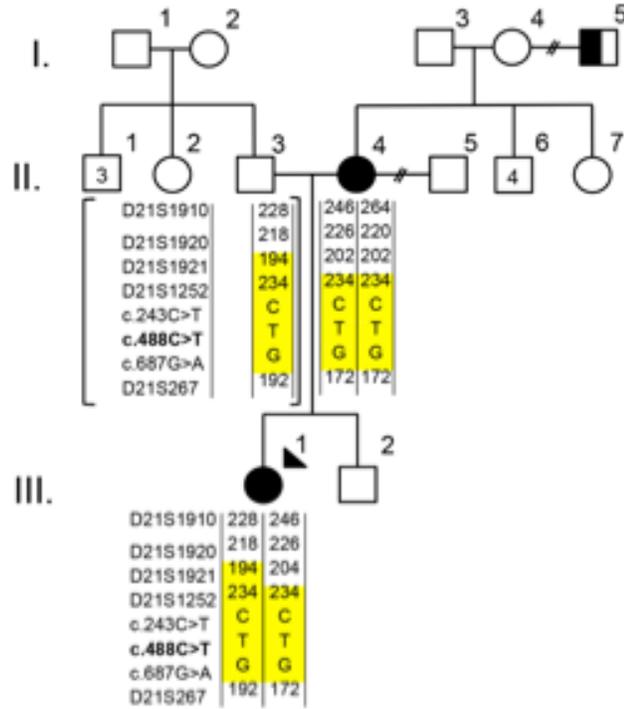


Figure 2.10. Wild-type/normal *CLDN14* c.488C. Blue box highlights wild-type sequence

Figure 2.11. Pedigree of family R2072, identified in screening of the NL deafness cohort, with the rare, precipitous audiologic phenotype who also share the *CLDN14* (c.488C>T; p.Ala163Val) allele and ancestral *DFNB29*-associated haplotype



a 1.4 Mb ancestral haplotype shared across all four families (Figure 2.7). Haplotype analysis shows affected individuals in the four families inherit an ancestral *DFNB29*-associated haplotype on chromosome 21q22.1, signifying clan membership, although biological connection for Family R2072 was not found (Figure 2.9 and Figure 2.11). Additionally, we sequenced all coding sequences of the *CLDN14* gene, including the exon/intron boundaries and 5' and 3' UTRs. We identified a common synonymous variant (c.243C>T; rs219799) within the clan, which was incorporated into our *DFNB29*-associated deafness haplotype.

2.6.5 Genealogical Analysis

Extension of the pedigrees and review of all audiograms by our audiologist identified 16 subjects with hearing loss; 10/16 subjects showed the distinct precipitous mid–high-frequency hearing loss (Figure 2.2). Subjects with hearing impairment not consistent with the distinct precipitous mid–high-frequency pattern include PID IV-3 (whom we have not connected to the founding couple) and all descendants of PID IV-4 (Figure 2.2). In these cases, the audiogram can be characterized as a flat loss across all frequencies; PID IV-4 had a profound flat loss and his descendants (V-5, V-7, VI-3) show a mild flat loss (Figure 2.12). Family interviews conducted by Anne Griffin and Sarah Predham determined surnames suggestive of Irish descent (Seary, 1977) and connected Families R2010, R2033 and R2075 to a single founding couple six generations ago. We noted that the inheritance pattern in the combined pedigree suggested either autosomal dominant (with reduced penetrance) or autosomal recessive (pseudodominant) inheritance (Figure 2.2). In summary, this population-based study using a targeted and whole exome sequencing approach identified a common *CLDN14* (*DFNB29*) variant (c.488C>T, p.(Ala163Val)) that causes recessive sensorineural hearing loss in several

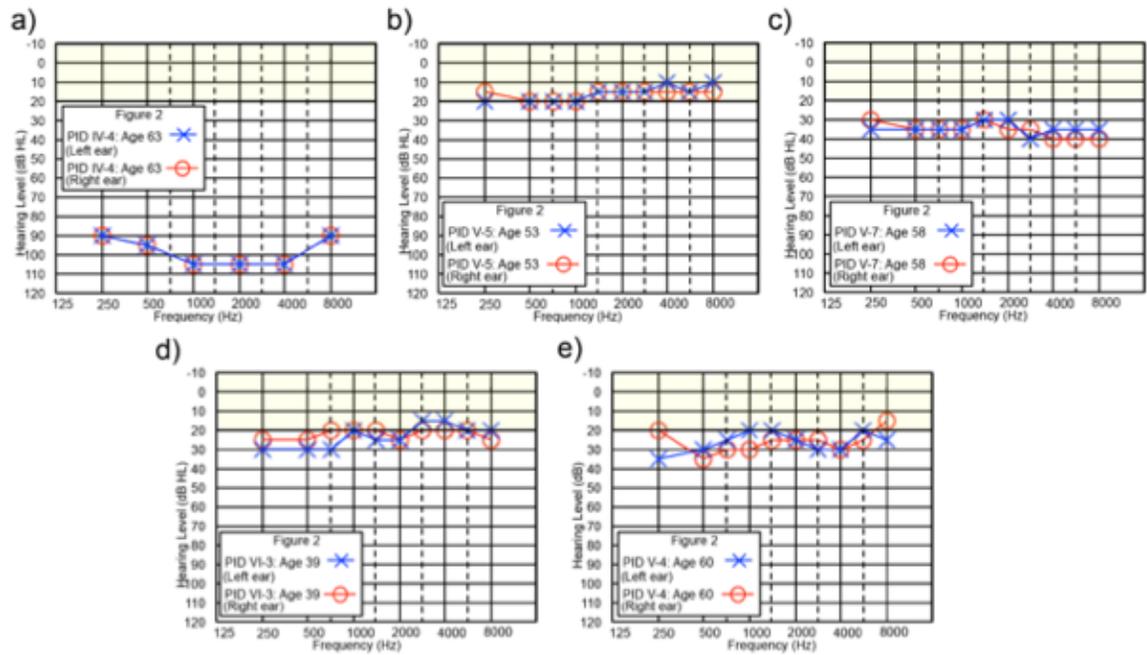


Figure 2.12. Clan members lacking the recessive *CLDN14* [c.488C>T; p.(Ala163Val)] variant do not present with the characteristic steeply sloping hearing phenotype, exhibiting a different age of onset and hearing threshold progression. a) Profound, flat sensorineural hearing loss with an unknown etiology at age 63 (PID IV-4), b) PID V-5 (age: 53) presents with borderline hearing thresholds, c) PID V-7 (age: 58) presents with mild hearing loss with a diagnosis of Meniere's disease, d) at age 39, PID VI-3 presents with mild hearing loss, and e) a heterozygous *CLDN14* [c.488C>T; p.(Ala163Val)] carrier (PID V-4) exhibits mild hearing loss at age 60. Yellow shaded area indicates range of normal hearing. Hearing thresholds are measured in decibels hearing level (dB HL), X = left ear (air conduction), O = right ear (air conduction)

Newfoundland families.

2.7 Discussion

We have determined that a known VUS (*CLDN14*, c.488C>T, p.(Ala163Val)) is likely pathogenic, and causes a precipitous, bilateral and rapid deterioration of hearing thresholds at frequencies >0.75 kHz in children, progressing gradually in adults. In addition, it has been determined that this Likely Pathogenic variant is enriched in the founder population of the island of Newfoundland and is present in ~1% of the population.

The role of *CLDN14* in nonsyndromic hearing loss was first described in two large consanguineous families from Pakistan with autosomal recessive profound congenital deafness (Wilcox *et al.*, 2001). Recessive *CLDN14* alleles manifest as nonsyndromic sensorineural hearing loss with considerable phenotypic variability and may present as a congenital, mild, moderate–severe or profound loss (R. Bashir *et al.*, 2010; Z. E. Bashir *et al.*, 2013). In this study, homozygous children had normal hearing thresholds up to 3 years of age and overall, a remarkably conserved hearing phenotype. Hearing loss onset is post-lingual during the first decade, which progresses from a normal gently sloping to moderate, to a normal sloping to profound by the second decade. A combination of pedigree extension and genotyping linked four families of Irish ancestry to a founding couple six generations back.

The claudin family of proteins consists of 24 members with tissue-specific expression. Claudin-14 plays a critical role in the formation of tight junction barriers that regulate paracellular ion transport (Mineta *et al.*, 2011) and is highly expressed in the kidney, liver and the inner ear (Ben-Yosef *et al.*, 2003; Wilcox *et al.*, 2001). Moreover, preferential gene expression has been observed in the inner ear, as *CLDN14* expression is lower in supporting

cells, relative to sensory hair cells(Scheffer *et al.*, 2015; Wilcox *et al.*, 2001). Normal hearing function and hair cell depolarization are dependent on tight junctions in the reticular lamina. In the organ of Corti, hair cell stereocilia are bathed in potassium-rich endolymph, while the basolateral surface of the hair cell body is surrounded by an intercellular (or extracellular) fluid continuous with the perilymph(Ferrary & Sterkers, 1998; Furuse & Tsukita, 2006). The reticular lamina, formed in part by tight junctions between the apical surfaces of hair cells and supporting cells of the sensory epithelium(Gulley & Reese, 1976), creates a barrier isolating the endolymphatic fluid from other cochlear compartments, which contain perilymph. Maintenance of this ionic gradient is essential for mechanotransduction, which depends on the modulation of potassium current flowing from the endolymph into the hair cells through the stereocilia, as they are displaced by sound-induced vibrations. The major molecular components of tight junctions include a broad group of genes of transmembrane occludins, claudins, and junctional adhesion molecules(Kitajiri *et al.*, 2004; Riazuddin *et al.*, 2006). In addition to *CLDN14*, other tight junction genes have been shown to cause hearing loss. For example, pathogenic variants in *MARVELD2* (MIM: 610572; Riazuddin *et al.*, 2006) and *ILDRI* (MIM: 609739; Borck *et al.*, 2011) are known to cause *DFNB49* (MIM: 610153) and *DFNB42* (MIM: 609646), respectively(Van Camp G, 2015). Disruption of this tight junction barrier alters the ionic gradient, increasing the potassium concentration around the hair cell body, compromising mechanotransduction, which causes hair cell toxicity, due to prolonged sensory cell depolarization, and eventual cell death(Ben-Yosef *et al.*, 2003).

The *CLDN14* p.(Ala163Val) variant reported in this study has been identified in previous studies, but not in association with disease. It was first reported as a VUS by Thorleifsson *et al*(Thorleifsson *et al.*, 2009) in a large Iceland/Netherlands GWAS cohort study

examining SNPs associated with kidney stones and bone mineral density, and more recently by Toka *et al* (2013), who detected the *CLDN14* p.(Ala163Val) allele in 3 of 1230 study participants for another kidney function study. The heterozygous p.(Ala163Val) allele was also found in a Swedish GWAS study examining the polygenic nature of schizophrenia (Purcell *et al.*, 2014). The heterozygous p.(Ala163Val) allele was submitted 31 times to ExAC browser, 29 alleles from European descent and 2 from the African population. In a recent American study including 1119 deafness probands, a cohort made up of 62.3% autosomal recessive cases (Sloan-Heggen *et al.*, 2016) used a targeted sequencing approach and the most commonly implicated genes were *GJB2*, *MYH9*, *OTOA*, *PCDH15*, *SLC26A4*, *STRC*, *TMCI*, *TMPRSS3* and *USH2A*. Interestingly, Sloan-Heggen *et al* (2016) reported the p.(Ala163Val) allele in a patient with congenital hearing loss; however, this was in a compound heterozygous state with a second *CLDN14* allele (p.Pro28Leu). In summary, these studies suggest that the likely pathogenic *CLDN14* p.(Ala163Val) allele is both rare and widely distributed around the globe.

Many reports claim that approximately 50% of autosomal recessive deafness is caused by either homozygous or compound heterozygous variants in the *DFNBI* locus (*GJB2*). Given that targeted NGS hearing loss panels have recently been implemented in the clinic with *GJB2*-negative cases, this represents a potential ascertainment bias, as children who been previously tested for *DFNBI* only typically were not offered any additional genetic testing. Recently, a large, ethnically diverse, cohort study demonstrated the importance of investigating *DFNBI*-negative deaf probands (Yan *et al.*, 2016). This study took a targeted panel approach in 342 probands (185 simplex and 157 multiplex families), sequenced 180 known hearing loss genes, and identified 151 variants in 119 families. Fifty-three families had pathogenic or likely pathogenic variants within 27 genes, while the remaining were variants of uncertain

significance. This study solved 25 and 7% of multiplex and simplex families, respectively, emphasizing the importance of large families and strong histories of disease in genetic studies (Yan *et al.*, 2016).

Pediatric hearing programs strive to identify and treat early in order to prevent delays in language, learning and social development. However, the detection of non-congenital and progressive forms of hearing loss remain a significant challenge. Children who are homozygous *CLDN14* p.(Ala163Val) pass newborn and early hearing tests. The proband's son (R2010, PID VI-4) was discharged after his normal hearing test at 1 year of age. Preschool testing 4 years later showed significant deterioration of both mid and high frequencies (Figure 2.1c). Delayed identification could result from incomplete testing of high frequencies in the preschool years, often complicated by limited testing tolerance in children. In this study, PID VI-2 had normal hearing at 8 kHz at 2 years of age. At 3 years, hearing was reported normal, although thresholds at 8 kHz were not performed. By 4 years, this child had developed a 55 dB HL threshold at 8 kHz and mild to moderate loss at all high frequencies required immediate hearing aid fitting. Retrospectively, if 8 kHz thresholds had been performed at 3 years, diagnosis and therapy could have been offered one year earlier. Conversely, genetic testing or prenatal/preconception parental carrier screening would have allowed appropriate hearing surveillance and minimized the risk of delays in language development and learning from hearing loss.

Adults who are homozygous for *CLDN14* p.(Ala163Val) also have a consistent phenotype, but there remain challenges in management. Hearing aids benefit affected children and young adults (up to the third decade), but most adults do not find them beneficial. For example, PIDs V-9 (age: 50), V-10 (age: 51) and V-17 (age: 57) reported some additional

sound with hearing aids, but no improvement of speech comprehension, consistent with the extreme erosion of mid and high frequencies. Older affected adults with well-preserved low-frequency sensitivity have limited ability to communicate by phone. PIDs V-10 and V-17, whose threshold at 0.5 kHz has deteriorated, can no longer communicate by phone. Adult members of this clan are highly skilled speech readers who can detect speech initiation and turn quickly to maximize the use of visual clues. Unfortunately, these skills can be mistaken for hearing and subjects have voiced concerns regarding safety in the workplace(Griffin, Personal communication: Discussion on hearing loss in NL. October 2018).

The development of the organ of Corti is unidirectional, and follows a base-to-apex hair cell degeneration in the *Cldn14*-null mouse cochlea. This may explain why we observe a sensorineural threshold loss progressing from the high to low frequencies in affected clan members. The cochlea discerns high- from low-frequency sound, based on a stiffness gradient along the basilar membrane(Ehret, 1978; Teudt & Richter, 2014). In *Cldn14*-null mice, the organ of Corti undergoes a base-to-apex deterioration beginning around postnatal day 10, with a more severe and rapid degeneration of outer hair cells compared to inner hair cells. By day 13, the three rows of outer hair cells are almost completely absent in the cochlear base, with partial loss and stereociliar disorganization in the middle and apical turns(Ben-Yosef *et al.*, 2003). The cochlear lesion then proceeds towards the cochlear apex, with a rapid deterioration of the outer hair cells accompanied by the onset of inner hair cell damage. By day 18, outer hair cell deterioration is severe with only a few remaining outer hair cells exhibiting damaged stereocilia in the most apical region; in contrast, only partial inner hair cell loss is reported throughout the cochlea by this age. Auditory brainstem responses measured in 4-week-old *Cldn14*-null mice indicate a significant hearing loss, in comparison to their wild-type and

heterozygous littermates (Ben-Yosef *et al.*, 2003). Since *CLDN14* p.(Ala163Val) causes a precipitous mid/high frequency hearing loss with normal hearing thresholds in the lower frequencies, these findings are consistent with hearing phenotypes observed *Cldn14*-null mice.

2.8 Summary

A population-based study of hearing loss in the NL population has clarified the role of *CLDN14* p.(Ala163Val), a VUS previously identified in the USA, Iceland and Sweden. *CLDN14* p.(Ala163Val) appears to be of Irish origin and causes a precipitous, prelingual autosomal recessive form of non-syndromic sensorineural hearing loss. This likely pathogenic variant is frequent in this island population of Northern European descent. *CLDN14* p.(Ala163Val) homozygotes have normal hearing thresholds at birth, and then experience rapid, progressive hearing loss in early childhood. Although missed by newborn hearing screening, genetic testing would ensure identification of at-risk children, allowing for appropriate monitoring and timely intervention, aural rehabilitation and counselling for families. We recommend the inclusion of *CLDN14* screening in NL newborn screening protocols for children with a family history of hearing loss.

2.9 Limitations

While exome sequencing is a powerful tool for elucidating disease-causing, coding variants, it does not explore non-coding regions. Additionally, there is no experimental proof of the predicted amino acid substitution, and without functional data, we cannot be certain that *CLDN14* p.(Ala163Val) impacts protein localization within tight junctions. For example, this variant could cause alternative splicing or alter gene expression. Although less likely, it is

plausible that a causal, non-coding variant at the *DFNB29* locus is in linkage disequilibrium with p.(Ala163Val). Even though our study presents several lines of evidence to suggest pathogenicity, experimental functional analysis of p.(Ala163Val) is required.

2.10 Acknowledgements

I would like to thank all the family members that I met during our clinical outreach trip to the Burin Peninsula. This experience gave me firsthand insight into the debilitating nature of deafness. I would also like to thank Elspeth Drinkell, Carol Negrijn, Jim Houston and Dante Galutira for technical assistance and Beverly King for key clinical insights. Sometimes the solutions to genetic problems are not as obvious as they may seem when beginning a project. A few words of advice for incoming students: be patient, be resilient and use your head.

2.11 Bibliography

- Abdelfatah, N., McComiskey, D. A., Doucette, L., Griffin, A., Moore, S. J., Negrijn, C., . . . Young, T. L. (2013). Identification of a novel in-frame deletion in KCNQ4 (DFNA2A) and evidence of multiple phenocopies of unknown origin in a family with ADSNHL. *Eur J Hum Genet*, *21*(10), 1112-1119. doi:10.1038/ejhg.2013.5
- Abdelfatah, N., Merner, N., Houston, J., Benteau, T., Griffin, A., Doucette, L., . . . Young, T. L. (2013). A novel deletion in SMPX causes a rare form of X-linked progressive hearing loss in two families due to a founder effect. *Hum Mutat*, *34*(1), 66-69. doi:10.1002/humu.22205
- Ahmed, Z. M., Li, X. C., Powell, S. D., Riazuddin, S., Young, T. L., Ramzan, K., . . . Wilcox, E. R. (2004). Characterization of a new full length TMPRSS3 isoform and identification of mutant alleles responsible for nonsyndromic recessive deafness in Newfoundland and Pakistan. *BMC Med Genet*, *5*, 24. doi:10.1186/1471-2350-5-24
- Bashir, R., Fatima, A., & Naz, S. (2010). Mutations in CLDN14 are associated with different hearing thresholds. *J Hum Genet*, *55*(11), 767-770. doi:10.1038/jhg.2010.104
- Bashir, Z. E., Latief, N., Belyantseva, I. A., Iqbal, F., Riazuddin, S. A., Khan, S. N., . . . Riazuddin, S. (2013). Phenotypic variability of CLDN14 mutations causing DFNB29 hearing loss in the Pakistani population. *J Hum Genet*, *58*(2), 102-108. doi:10.1038/jhg.2012.143
- Bear, J. C., Nemecek, T. F., Kennedy, J. C., Marshall, W. H., Power, A. A., Kolonel, V. M., & Burke, G. B. (1987). Persistent genetic isolation in outport Newfoundland. *Am J Med Genet*, *27*(4), 807-830. doi:10.1002/ajmg.1320270410

- Ben-Yosef, T., Belyantseva, I. A., Saunders, T. L., Hughes, E. D., Kawamoto, K., Van Itallie, C. M., . . . Friedman, T. B. (2003). Claudin 14 knockout mice, a model for autosomal recessive deafness DFNB29, are deaf due to cochlear hair cell degeneration. *Hum Mol Genet*, 12(16), 2049-2061.
- Borck, G., Ur Rehman, A., Lee, K., Pogoda, H. M., Kakar, N., von Ameln, S., . . . Kubisch, C. (2011). Loss-of-function mutations of ILDR1 cause autosomal-recessive hearing impairment DFNB42. *Am J Hum Genet*, 88(2), 127-137. doi:10.1016/j.ajhg.2010.12.011
- Charif, M., Bakhchane, A., Abidi, O., Boulouiz, R., Eloulid, A., Roky, R., . . . Barakat, A. (2013). Analysis of CLDN14 gene in deaf Moroccan patients with non-syndromic hearing loss. *Gene*, 523(1), 103-105. doi:10.1016/j.gene.2013.03.123
- Contributors5, W. C. (10 October 2018 16:37 UTC). File:Cochlea-crosssection.svg. Retrieved from <https://commons.wikimedia.org/w/index.php?title=File:Cochlea-crosssection.svg&oldid=323507750>
- Contributors6, W. C. (7 December 2018 05:32 UTC). File:Tight Junction Transmembrane Proteins.jpg. Retrieved from https://commons.wikimedia.org/w/index.php?title=File:Tight_Junction_Transmembrane_Proteins.jpg&oldid=330533182
- Doucette, L., Merner, N. D., Cooke, S., Ives, E., Galutira, D., Walsh, V., . . . Young, T. L. (2009). Profound, prelingual nonsyndromic deafness maps to chromosome 10q21 and is caused by a novel missense mutation in the Usher syndrome type IF gene PCDH15. *Eur J Hum Genet*, 17(5), 554-564. doi:10.1038/ejhg.2008.231
- Ehret, G. (1978). Stiffness gradient along the basilar membrane as a basis for spatial frequency analysis within the cochlea. *J Acoust Soc Am*, 64(6), 1723-1726.

- Ferrary, E., & Sterkers, O. (1998). Mechanisms of endolymph secretion. *Kidney Int Suppl*, 65, S98-103.
- Furuse, M., & Tsukita, S. (2006). Claudins in occluding junctions of humans and flies. *Trends Cell Biol*, 16(4), 181-188. doi:10.1016/j.tcb.2006.02.006
- Griffin, A. (Personal communication: Discussion on hearing loss in NL. October 2018, October 19, 2018).
- Gulley, R. L., & Reese, T. S. (1976). Intercellular junctions in the reticular lamina of the organ of Corti. *J Neurocytol*, 5(4), 479-507.
- Hildebrand, M. S., DeLuca, A. P., Taylor, K. R., Hoskinson, D. P., Hur, I. A., Tack, D., . . . Smith, R. J. (2009). A contemporary review of AudioGene audioprofiling: a machine-based candidate gene prediction tool for autosomal dominant nonsyndromic hearing loss. *Laryngoscope*, 119(11), 2211-2215. doi:10.1002/lary.20664
- Kitajiri, S. I., Furuse, M., Morita, K., Saishin-Kiuchi, Y., Kido, H., Ito, J., & Tsukita, S. (2004). Expression patterns of claudins, tight junction adhesion molecules, in the inner ear. *Hear Res*, 187(1-2), 25-34.
- Lee, K., Ansar, M., Andrade, P. B., Khan, B., Santos-Cortez, R. L., Ahmad, W., & Leal, S. M. (2012). Novel CLDN14 mutations in Pakistani families with autosomal recessive non-syndromic hearing loss. *Am J Med Genet A*, 158A(2), 315-321. doi:10.1002/ajmg.a.34407
- Manion, J. J. (1977). *The Peopleing of Newfoundland: Essays in Historical Geography*. Memorial University of Newfoundland, St. John's, Newfoundland, Canada: Institute of Social and Economic Research

- Merner, N. D., Hodgkinson, K. A., Haywood, A. F., Connors, S., French, V. M., Drenckhahn, J. D., . . . Young, T. L. (2008). Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *Am J Hum Genet*, 82(4), 809-821. doi:10.1016/j.ajhg.2008.01.010
- Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, 16(3), 1215.
- Mineta, K., Yamamoto, Y., Yamazaki, Y., Tanaka, H., Tada, Y., Saito, K., . . . Tsukita, S. (2011). Predicted expansion of the claudin multigene family. *FEBS Lett*, 585(4), 606-612. doi:10.1016/j.febslet.2011.01.028
- Morton, C. C., & Nance, W. E. (2006). Newborn hearing screening--a silent revolution. *N Engl J Med*, 354(20), 2151-2164. doi:10.1056/NEJMra050700
- Nayak, G., Varga, L., Trincot, C., Shahzad, M., Friedman, P. L., Klimes, I., . . . Riazuddin, S. (2015). Molecular genetics of MARVELD2 and clinical phenotype in Pakistani and Slovak families segregating DFNB49 hearing loss. *Hum Genet*, 134(4), 423-437. doi:10.1007/s00439-015-1532-y
- Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., . . . Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, 506(7487), 185-190. doi:10.1038/nature12975
- Riazuddin, S., Ahmed, Z. M., Fanning, A. S., Lagziel, A., Kitajiri, S., Ramzan, K., . . . Friedman, T. B. (2006). Tricellulin is a tight-junction protein necessary for hearing. *Am J Hum Genet*, 79(6), 1040-1051. doi:10.1086/510022

- Richard JH Smith, E. S., Michael S Hildebrand, and Guy Van Camp. (1999 January 19, 2014). GeneReviews: Deafness and Hereditary Hearing Loss Overview. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK1434/>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, *17*(5), 405-424. doi:10.1038/gim.2015.30
- Scheffer, D. I., Shen, J., Corey, D. P., & Chen, Z. Y. (2015). Gene Expression by Mouse Inner Ear Hair Cells during Development. *J Neurosci*, *35*(16), 6366-6380. doi:10.1523/JNEUROSCI.5126-14.2015
- Seary, E. R. (1977). *Family Names of the Island of Newfoundland*. St. John's, Newfoundland: McGill-Queen's University Press.
- Sloan-Heggen, C. M., Bierer, A. O., Shearer, A. E., Kolbe, D. L., Nishimura, C. J., Frees, K. L., . . . Smith, R. J. (2016). Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum Genet*, *135*(4), 441-450. doi:10.1007/s00439-016-1648-8
- Teudt, I. U., & Richter, C. P. (2014). Basilar membrane and tectorial membrane stiffness in the CBA/CaJ mouse. *J Assoc Res Otolaryngol*, *15*(5), 675-694. doi:10.1007/s10162-014-0463-y
- Thorleifsson, G., Holm, H., Edvardsson, V., Walters, G. B., Styrkarsdottir, U., Gudbjartsson, D. F., . . . Stefansson, K. (2009). Sequence variants in the CLDN14 gene associate with kidney stones and bone mineral density. *Nat Genet*, *41*(8), 926-930. doi:10.1038/ng.404

- Toka, H. R., Genovese, G., Mount, D. B., Pollak, M. R., & Curhan, G. C. (2013). Frequency of rare allelic variation in candidate genes among individuals with low and high urinary calcium excretion. *PLoS One*, *8*(8), e71885. doi:10.1371/journal.pone.0071885
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res*, *40*(15), e115. doi:10.1093/nar/gks596
- Van Camp G, S. R. (2015). Hereditary Hearing Loss Homepage. Retrieved from <http://hereditaryhearingloss.org>
- Wattenhofer, M., Reymond, A., Falciola, V., Charollais, A., Caille, D., Borel, C., . . . Antonarakis, S. E. (2005). Different mechanisms preclude mutant CLDN14 proteins from forming tight junctions in vitro. *Hum Mutat*, *25*(6), 543-549. doi:10.1002/humu.20172
- Wilcox, E. R., Burton, Q. L., Naz, S., Riazuddin, S., Smith, T. N., Ploplis, B., . . . Friedman, T. B. (2001). Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29. *Cell*, *104*(1), 165-172.
- Yan, D., Tekin, D., Bademci, G., Foster, J., 2nd, Cengiz, F. B., Kannan-Sundhari, A., . . . Tekin, M. (2016). Spectrum of DNA variants for non-syndromic deafness in a large cohort from multiple continents. *Hum Genet*, *135*(8), 953-961. doi:10.1007/s00439-016-1697-z
- Young, T. L., Ives, E., Lynch, E., Person, R., Snook, S., MacLaren, L., . . . King, M. C. (2001). Non-syndromic progressive hearing loss DFNA38 is caused by heterozygous missense mutation in the Wolfram syndrome gene WFS1. *Hum Mol Genet*, *10*(22), 2509-2514.

Zhai, G., Zhou, J., Woods, M. O., Green, J. S., Parfrey, P., Rahman, P., & Green, R. C. (2015).

Genetic structure of the Newfoundland and Labrador population: founder effects

modulate variability. *Eur J Hum Genet.* doi:10.1038/ejhg.2015.256

Chapter 3: A Pathogenic Splicing Variant In a Nascent *ATP11A* Exon Maps To *DFNA33* and Documents Its First Association With a Penetrant Mendelian Phenotype

Author list:

Justin A. Pater¹, Cindy Penney¹, Darren D. O’Rielly¹, Susan G. Stanton³, Anne Griffin¹, Nicole Roslin⁴, Daniel Vincent⁵, Pascale Marquis⁵, Geoffrey Woodland¹, Jim Houston¹, Taylor Burt¹,
Terry-Lynn Young^{1, 2, 3}

Affiliation:

¹Craig L. Dobbin Research Centre, Faculty of Medicine, Memorial University, 300 Prince Phillip Drive, St. John’s, Newfoundland and Labrador, Canada, A1B 3V6

²Molecular Diagnostic Laboratory, Eastern Health, Craig L. Dobbin Genetics Research Centre, Faculty of Medicine, Memorial University, 300 Prince Phillip Drive, St. John’s, Newfoundland and Labrador, Canada, A1B 3V6

³Communication Sciences and Disorders, Western University, Elborn College, 1201 Western Road, London, Ontario, Canada, N6G 1H1

⁴The Centre for Applied Genomics, The Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, 686 Bay Street, Toronto, Ontario, M5G 0A4

⁵Genome Quebec Innovation Centre, McGill University, 740 Dr. Penfield Avenue, Montréal, Québec, Canada H3A 0G1

⁶Canadian Centre for Computational Genomics, McGill University and Genome Quebec Innovation Center, 740 Dr. Penfield Avenue, Montréal, Québec, Canada H3A 0G1

A shorter version of this chapter has been submitted to the journal Genetics in Medicine

3.1 Co-author Statement

The following authors were responsible for the following: AG recruited/contributed patients. NR, DV: linkage analysis. JAP, AG, SGG, CP: collected/collated/evaluated clinical data for selecting whole exome sequencing family members. JAP whole exome sequencing. GW, JAP: whole exome sequencing bioinformatic data analysis. JAP, JH: sequenced/genotyped, JAP: performed a comprehensive audioprofiling analysis, JAP, DV, PM: whole genome sequencing. PM, JAP, DDO: whole genome sequencing bioinformatics analysis. JAP, TLY: haplotype analyses, JAP: RNA Extractions, JAP: TOPO-TA Cloning, JAP: Colony PCR, JAP, JH, DDO, TLY: wrote/edited the manuscript, TLY: directed the study.

3.2 Abstract

Hearing loss is the most genetic of all human phenotypes, and extremely genetically and clinically heterogeneous. In the current era, all Mendelian disorders are being solved at an accelerated pace, due to the routine use of whole exome sequencing. However, and particularly for hearing loss, new gene discoveries in dominant disorders lag behind. Herein, we report a six-generation family of Northern European ancestry with mid-high frequency (cookie bite) hearing loss progressing to a flat loss across all frequencies. In addition to cases of congenital hearing loss, some family members exhibit variable onset, progression and audiogram configuration. Linkage and haplotype defined the critical disease region to 769 Kb residing within the unsolved *DFNA33* locus on chromosome 13q34. Whole exome and segregation analysis failed to identify the causative gene. Whole genome sequencing yielded several potential candidate variants, including *ATP11A* c.*11G>A, a medium impact variant predicted to cause the activation of a cryptic donor splice site. So far, *ATP11A* has 17 reported isoforms. RNA analysis on control tissues yielded three distinct bands and on patients yielded three extra high molecular bands not seen in control tissue, for a total of six bands. The three higher molecular weight bands all contained the insertion of 153bp contiguous segment of intronic sequence, resulting in the extension of exon 29 (in isoform *ATP11A*-201) and exon 30 (in isoforms *ATP11A*-202/212). This is the first description of mutations in the *ATP11A* gene associated with human disorder.

3.3 Introduction

Hereditary hearing loss is a common sensory disorder that exhibits extensive genetic and clinical heterogeneity (Morton & Nance, 2006). So far, over 200 hearing loss genes have been identified (Van Camp G, 2015); however, approximately one-third of the 60 mapped dominant loci have evaded discovery. Dominant hearing loss is typically characterized by variable expressivity and reduced penetrance (Richard JH Smith, 1999), making disease genes particularly challenging to identify. In addition to the rarity of large extended families, there are thousands of loci heterozygous in individuals, making it almost impossible to solve without multiple, multigenerational families for study (Sherry *et al.*, 2001). Furthermore, using advanced bioinformatics that does not exclusively rely on the RefSeq database for sequence read alignment has been shown to be more robust at identifying causal variants, especially for the vast majority of genes represented by multiple isoforms (Belkadi *et al.*, 2015; Zhao & Zhang, 2015).

A key feature of eukaryotic membranes is the non-random distribution of phospholipids, an essential feature to maintaining the integrity of a cell (Segawa *et al.*, 2014). This non-random asymmetry is most evident at the plasma membrane (Zachowski, 1993) and is maintained by the action of three classes of proteins: scramblases, floppases and flippases. Phospholipid flippases (or P4-ATPases) specifically transport or “flip” of phospholipids from the outer to the inner leaflet of a phospholipid membrane (Paulusma & Elferink, 2010). Mouse studies have shown the importance of the P4-ATPase class of flippases in phospholipid metabolism and biology and in maintaining normal auditory function (Coleman *et al.*, 2014; Stapelbroek *et al.*, 2009). Furthermore, pathogenic variants in other P4-ATPases disease that encompass a hearing component. For example, *ATP8B1* (MIM: 605868), cause intrahepatic cholestasis type 1

(MIM: 211600), where patients sporadically develop hearing loss (Stapelbroek *et al.*, 2009). Moreover, phospholipids have been implicated in autoimmune conditions, such as antiphospholipid syndrome, a disorder characterized by the presence of antiphospholipid antibodies that invoke an autoimmune response, causing thrombosis, complications during pregnancy and hearing loss (Mouadeb & Ruckenstein, 2005; Wiles *et al.*, 2006).

Herein, we describe a pathogenic splicing variant in *ATP11A* (MIM: 605868) that causes autosomal dominant hearing loss in a multiplex family from a genetic isolate, documenting the first disease association for this gene.

3.4 Materials and Methods

3.4.1 Study participants and clinical evaluations

This project is a part of a large hereditary hearing loss study in NL, Canada. Informed consent, permission to access medical records and family history data were obtained from all research participants, as per approved institutional review board protocol #01.186 (Human Research Ethics Board, St. John's, NL, Canada). Family R2070 has a history of bilateral, sensorineural hearing loss spanning five generations (Figure 3.1).

Hearing loss was measured using air conduction thresholds and pure-tone audiometry methods noting severity progression and severity. The proband (PID IV-7; Figure 3.1) exhibited a progressive, sloping, bilateral sensorineural hearing loss (Figure 3.2A), and hearing severity was variable across 18/25 recruited members that were affected in this autosomal dominant pedigree (Figure 3.2B-D). Although the age of onset ranged between the first and second decade, an affected female in a recent generation failed newborn screening (PID V-5; Figure 3.1).

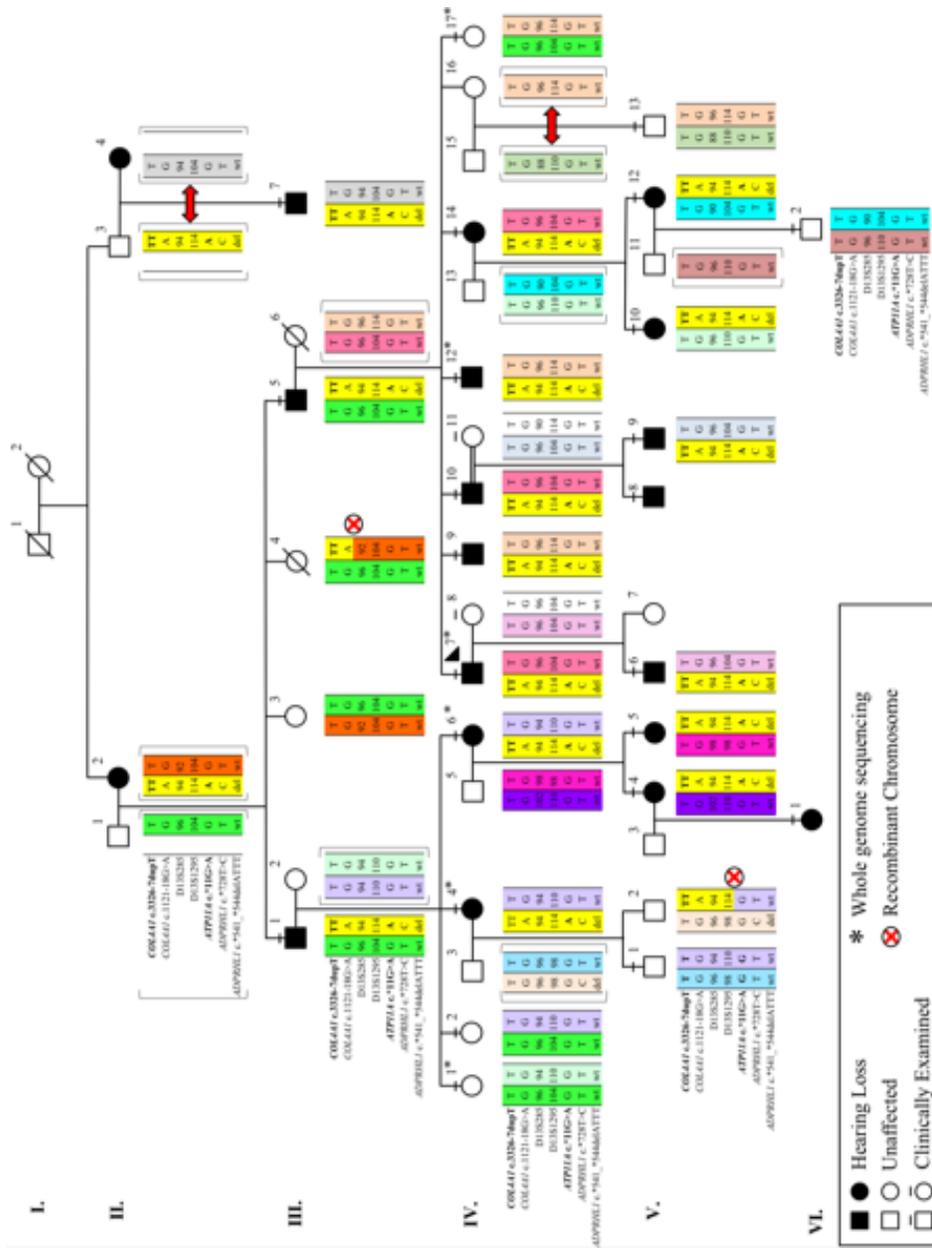


Figure 3.1. Family R2070 pedigree with a variable *DFNA33* hearing loss phenotype. This family is a core pedigree from an extended pedigree. Yellow chromosome indicates a shared ancestral *DFNA33*-associated haplotype. Filled symbols: *DFNA33* sensorineural hearing loss. Unfilled symbols: unaffected family member. Circle with X: recombinant chromosome. Asterisks indicate whole genome sequencing family members. Dashes above symbols indicate hearing testing by pure tone audiometry.

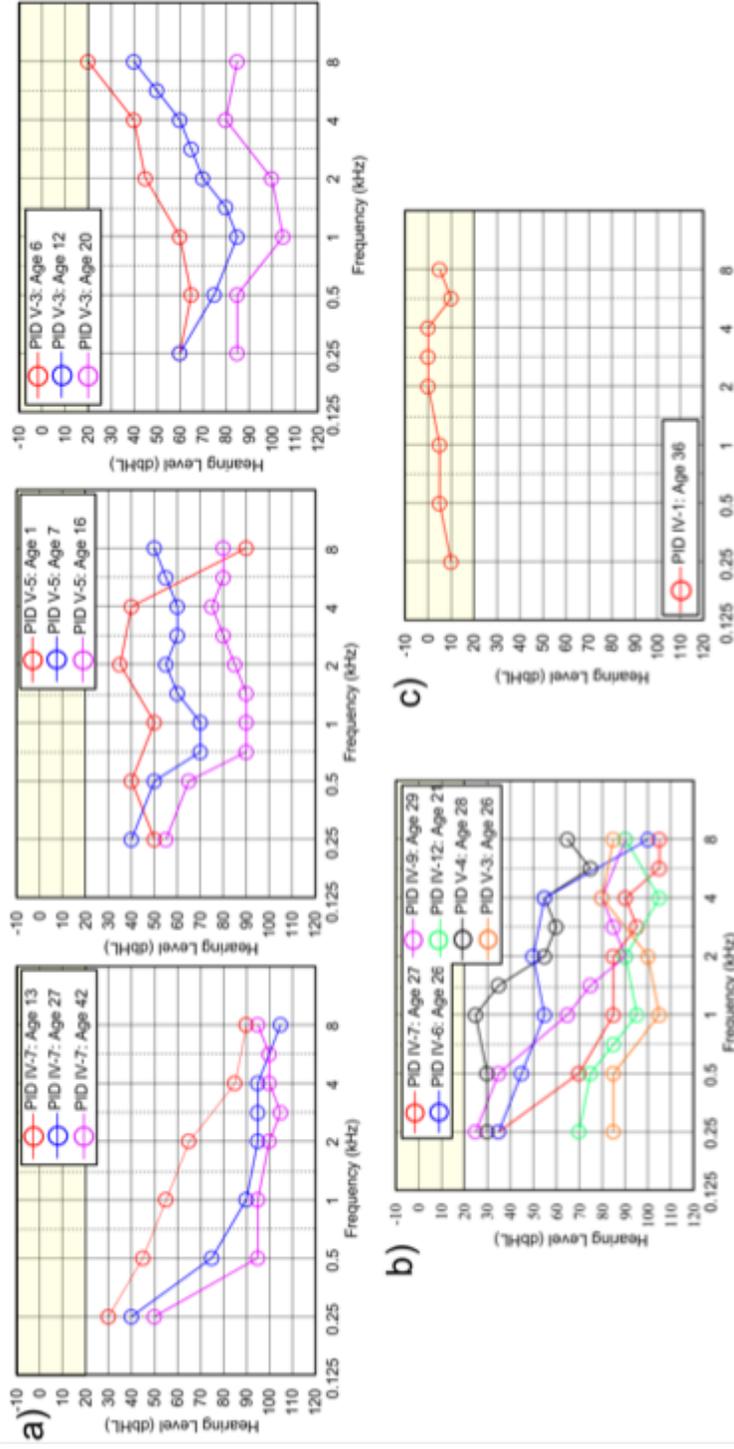


Figure 3.2. Family R2070 Phenotype. **a)** Sample audiograms measured longitudinally for 3 family members illustrate a decline in hearing sensitivity with age in each case, but differences in hearing loss severity. The audiograms also demonstrate individual differences in the frequency configuration of hearing sensitivity, specifically a sloping high frequency loss, a relatively flat configuration and a low frequency rising audiogram (top to bottom, respectively), which also change with increasing age. **b)** Audiograms for 6 affected family members in their 3rd decade reveal individual variability in both the severity and configuration of hearing loss, **c)** Unaffected family member with normal hearing thresholds across all pure tone frequencies (0.25-8 kHz).

3.4.2 Targeted screening and lineage analysis

Preceding to this study, genomic DNA was extracted from peripheral blood using a simple salting out protocol (Miller *et al.*, 1988). The Family R2070 proband PID IV-7, Figure 3.1) was screened for population-specific deafness alleles (Appendix C; Abdelfatah, McComiskey, *et al.*, 2013; Abdelfatah, Merner, *et al.*, 2013; Ahmed *et al.*, 2004; Doucette *et al.*, 2009; Young *et al.*, 2001). To identify candidate hearing loss genes, we submitted audiometric data from the proband to Audiogene (v4.0), a computational algorithm comparing sample audiograms to reference audioprofiles of 34 autosomal dominant deafness loci (Hildebrand *et al.*, 2009). Next, genome-wide single nucleotide polymorphism (SNP) genotyping was performed using the Illumina 610Quad genotyping chip (Illumina Inc., San Diego, CA, USA) on select family members. Starting with a set of >500,000 high quality SNP markers, a subset of informative markers (n=17,407) was used for parametric two-point linkage analysis using Superlink (v1.7) (Fishelson & Geiger, 2004). Linkage analysis was performed under an autosomal dominant model with 99% penetrance and a disease allele frequency of 0.0025. Genomic DNA on all available family members was extracted from peripheral blood using a modified salting out protocol (Miller *et al.*, 1988).

3.4.3 Whole exome sequencing (WES) and variant filtration

We prepared whole exome libraries for four affected (PID IV-4, IV-6, IV-7, and IV-12;) and two unaffected (PID IV-1 and IV-17; Figure 3.1) family members with the Ion Torrent AmpliSeq RDY Exome Kit (Life Technologies, Cat. #A27193), followed by purification, adapter ligation and barcoding using the Ion PI Hi-Q OT2 200 kit (Life Technologies, Cat. #A26434). Exome libraries were quantified using the Ion Library Quantification Kit (Life

Technologies, Cat. #4468802) and then loaded onto a PI v3 chip and sequenced with the Ion Torrent Proton Sequencer. Geoff Woodland called and annotated SNVs and IDELS using GATK (v3.5) and SnpEff (v4.1; <http://snpeff.sourceforge.net/>), respectively. Subsequently, variants were filtered against publically available SNP databases (ExAC Browser, <http://exac.broadinstitue.org/>; dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>; 1000 genomes, <http://www.1000-genomes.org>) and the impact of SNVs at the protein level was assessed using SIFT, PolyPhen-2, and MutationTaster. In addition, moderate and high impact variants (nonsense, frameshift, missense, splice sites) with a minimum of 30X coverage were included in our final list of candidate variants. Given that the family has a clear autosomal dominant inheritance pattern, we filtered for heterozygous variants that were shared across all affected family members and absent in unaffected members. Two separate analyses were conducted. First, we filtered for rare variants that had a MAF of <1%. Our second analysis included the filtration of common variants that have a MAF <10%, given that we have previously identified a common pseudodominant hearing loss allele (*CLDN14* c.488C>T) in the NL population. Sequencing coverage within our linked regions was assessed and any genes that did not meet our variant filtration criteria were manually sequenced.

3.4.4 Comprehensive audioprofiling

Given that no variants that were identified by WES, we considered that this may be due to a mutation that exhibits reduced penetrance or perhaps there are phenocopies in this family. Therefore, we (JAP and AG) decided to conduct an extensive audioprofile analysis on all available family members. This comprehensive exercise ensured that we were selecting family members with the most similar phenotypes for WGS.

3.4.5 Whole genome sequencing and variant filtration

Genomic DNA libraries were prepared on four affected (PID IV-4, IV-6, IV-7, and IV-12) and two unaffected relatives (PID IV-1 and IV-17) using the Lucigen Shotgun NxSeq AmpFREE Low DNA Library Kit (Cat. #14000-1, Lucigen Inc., Madison, WI, USA). Prepared libraries were loaded on an Illumina paired end 150 base pair (bp) sequencing lane, and sequenced on the HiSeqX Sequencer (Illumina Inc., San Diego, CA, USA). Sequence reads were aligned to GRch37 in both RefSeq and Ensembl reference genomes, and single nucleotide variants (SNV) and insertions and deletions (INDELs) were called using GATK (v4.0). Structural chromosomal variants were called using Lumpy (v0.2.13) and SVtyper (v0.5.2), while a Bioconductor package, cn.MOPS (v1.26.0), was used for copy number variation (CNV) analysis. Variants were functionally annotated with SNPeff (v4.3T). We filtered for rare variants using a MAF <1% with a minimum of 20X coverage in genes residing within the linked regions.

3.4.6 Cascade sequencing, segregation and haplotype analysis

Candidate variants were amplified using a standard touchdown PCR protocol and sequenced in all available family members and 326 ethnically-matched controls to determine the MAF. Haplotype analysis was performed using microsatellite markers and intergenic SNPs. Genotyping was performed using GeneMapper software (v4.0) and while SNVs were sequenced in all available family members (Pater *et al.*, 2017). Subsequently, genotypes were phased across the family pedigree. Microsatellites were genotyped according to standard

procedures(Abdelfatah, McComiskey, *et al.*, 2013) using GeneMapper software (v4.0) and haplotypes were reconstructed manually.

3.4.7 Experimental validation of splicing variants

Candidate splicing variants within exon-intron boundaries were analyzed *in silico* using MaxEnt, Human Splicing Finder (v3.1), and NNSPLICE (v0.9) to determine their effects on RNA splicing, which were experimentally validated using RNA from transformed patient-derived B-cell lymphocytes. RNA was extracted using TRIzol-based methods (Thermo-fisher, Cat. #15596026) and cDNA libraries were prepared with the Superscript VILO cDNA synthesis kit (Thermo-fisher, Cat. #11754050) followed by genomic DNA digestion using the Turbo DNA-free kit (Invitrogen, Cat. #1907). Reverse transcription PCR (RT-PCR) was performed using a standard touchdown PCR protocol and primers that flanked candidate splicing variants (*ATP11A*: 5' CCAGAGGGGTGTGAAGCA 3' and 5' CATCACACGAGCATTCCCAC 3'; *COL4A1*: 5' GTTCACCTGGCTTACCTGGA 3' and 5' AAACCCACCTCACCCCTTTG 3'). RT-PCR products were visualized using a 1.5% agarose Tris-Borate-EDTA gel stained with SYBR Safe (Invitrogen, Cat. #S33102). We employed TA-cloning technology using the TOPO TA-Cloning Kit for Sequencing with One Shot TOP10 Chemically Competent *E. coli* (Invitrogen, Cat. #K457540) according to the manufacturer's protocol, for candidate genes with multiple transcripts (Appendix D). Clones were amplified using colony PCR(Costa & Weiner, 2006), sequenced and visualized using Mutation Surveyor Software (v5.0, SoftGenetics LLC State College, PA, USA).

3.5 Results

3.5.1 Hearing loss maps to 13q34 overlapping with the *DFNA33* locus

The proband screened negative for all known pathogenic hearing loss variants in the Newfoundland (NL) population (Appendix C), prompting a more comprehensive genomic approach. SNP genotyping and linkage analysis identified a 3.6 Mb region at 13q34 (chr13: 110,708,368-114,312,000; Appendix E) with a LOD score of 4.77, overlapping with *DFNA33* (Bonsch *et al.*, 2009).

3.5.2 Whole exome sequencing fails to identify the genetic basis of hearing loss

In total, 40 variants were identified in four affected family members (PID IV-4, IV-6, IV-7, IV-12) and were absent two unaffected members (PID IV-1 and PID IV-17; Figure 3.1). After the removal of likely false positive and known benign calls, our first analysis identified 10 rare variants (<1% MAF) in four affected family members (Appendix F). Accounting for the possibility of reduced penetrance and pseudodominance, our second analysis identified an additional 12 common variants (1%-10% MAF; Appendix G). Between both analyses, 9 of these variants resided on chromosomes that were identified by our linkage analysis. One common variant, *CARS2* c.538A>T (rs72661692, MAF = 7.37%) was identified within our linked region on chromosome 13. However, it failed to segregate with hearing loss. All 10 rare variants were tested for segregation (Appendix H), but also did not segregate with hearing loss in the family. We decided that testing for segregation on common variants that did not reside in our linked region on chromosome 13 was a waste of resources, and therefore, these variants were not sequenced. All genes within our linked region met our coverage criteria for variant filtration, with the exception of two very small genes, *SOX1* (MIM: 602148) and *IRS2*

(MIM:600797). Upon reviewing these genes in UCSC Genome Browser (<https://genome.ucsc.edu/>), it was discovered that both of these genes contain >75% GC content, which explains why there had no coverage during WES. In order to overcome this challenge, we manually amplified *SOX1* and *IRS2* using 16 primer sets that covered all coding regions of these genes, followed by Sanger sequencing. While these data revealed several benign polymorphisms that are present in the global population in high numbers, no deleterious variants were identified (data not show). Consequently, comprehensive audioprofiling analysis was performed to determine if there were any potential phenocopies in the family (Appendix I). A detailed report was generated, which was crucial in deciding on what samples to use for whole genome sequencing and how to interpret genomic results in this family (Appendix J).

3.5.3 Two candidate genetic variants co-segregate with dominant hearing loss

Comprehensive audioprofiling confirmed that PID IV-4, IV-6, IV-7, IV-12 and PID IV-1 and PID IV-17 (Figure 3.1), were the most representative family members that were affected and unaffected, respectively (Appendix I and J). An average coverage of 44X was obtained with 94% of the genome covered at 25X. A total of 15,071 variants (minimum of 20X coverage) were identified in four affected family members and absent in two unaffected members, with 49 low impact and two medium impact rare variants (<1% MAF) residing within the linked 3.6 Mb region (Table 3.1). No structural chromosomal rearrangements or CNVs were identified (data not shown). Two candidate splicing variants in *ATP11A* (c.*11G>A; *ATP11A*-203; ENST00000415301.1; Figure 3.3) and *COL4A1* (c.3326-7dupT; NM_001845; rs532261610; Figure 3.4) co-segregated with the disease-associated haplotype (Figure 3.1). *ATP11A* c.*11G>A is a novel, unreported medium impact variant

Table 3.1. Fifty-one rare variants (<1% MAF; minimum of 30X coverage) that were present in four affected family R2070 members and absent in two unaffected members. We exclusively focused on all genes that resided within our 11.13 cM region at 13q34 (chr13: 110,708,368 - 114,312,000). In order to ensure that we didn't miss any variants and compensate for the possibility of reduced penetrance in our linkage analysis, we took a conservative approach by examining a slightly broader region on chromosome 13 (chr13: 95,000,000 - terminus). MAF: minor allele frequency, cM: centimorgan

Chromosome	Start	End	Gene	Variant	dbSNP	Impact	MAF (%)
13	98300574	98300575	<i>PSMA6P4-RP11-120E13.1</i>	n.98300575G>C	None	Low	-
13	98300580	98300581	<i>PSMA6P4-RP11-120E13.1</i>	n.98300581G>C	None	Low	-
13	98300584	98300585	<i>PSMA6P4-RP11-120E13.1</i>	n.98300585A>C	None	Low	-
13	108527191	108527192	<i>FAM155A-LIG4</i>	n.108527192C>T	rs547032144	Low	0.89
13	108542600	108542601	<i>FAM155A-LIG4</i>	n.108542601G>A	rs149383683	Low	0.29
13	108559560	108559561	<i>FAM155A-LIG4</i>	n.108559561G>A	rs541712696	Low	-
13	108857507	108857508	<i>LIG4</i>	c.*3372_*3373insTTT	rs146739883	Low	-
13	108915581	108915582	<i>TNFSF13B</i>	n.78-6866A>G	rs117976973	Low	0.50
13	109053517	109053518	<i>TNFSF13B-HCFC2P1</i>	n.109053519_109053524delACACAC	rs143572234	Low	-
13	109076930	109076931	<i>TNFSF13B-HCFC2P1</i>	n.109076931_109076932insTCTCTCTC	None	Low	-
13	109154792	109154793	<i>HCFC2P1-MYO16</i>	n.109154793G>A	rs187868150	Low	0.40
13	109325175	109325176	<i>MYO16</i>	c.226+6679T>C	rs72664976	Low	0.89
13	109453319	109453320	<i>MYO16</i>	c.676-5700dupA	rs377390365	Low	0.80
13	109595133	109595134	<i>MYO16</i>	c.1860-14902T>G	rs118168489	Low	0.89
13	109646663	109646664	<i>MYO16</i>	c.2376+1886_2376+1887dupCA	rs146565717	Low	-
13	109835672	109835673	<i>MYO16</i>	c.5349+3692A>C	rs187349074	Low	0.70
13	109863191	109863192	<i>MYO16</i>	c.*4008G>A	None	Low	-
13	110001293	110001294	<i>LINC01067-LINC00359</i>	n.110001294A>G	rs147705218	Low	0.40
13	110183896	110183898	<i>LINC00399-LINC00676</i>	n.110183898delT	rs544892286	Low	-
13	110320676	110320680	<i>LINC00399-LINC00676</i>	n.110320678_110320680delTTC	rs147586219	Low	-
13	110350350	110350351	<i>LINC00399-LINC00676</i>	n.110350351T>A	rs182466123	Low	0.80
13	110492036	110492047	<i>IRS2-RN7SKP10</i>	n.110492038_110492047delCATAATATAA	None	Low	-

Table 3.1. Continued

Chromosome	Start	End	Gene	Variant	dbSNP	Impact	MAF (%)
13	110826878	110826879	<i>COL4A1</i>	c.3326-7dupT	rs532261610	Medium	0.37
13	110942287	110942288	<i>COL4A1</i>	c.84+17003G>A	rs187157903	Low	0.60
13	111189067	111189068	<i>RAB20</i>	c.173-12524A>C	None	Low	-
13	111997038	111997039	<i>TEX29</i>	c.*569C>T	None	Low	-
13	112203312	112203313	<i>TEX29-RP11-65D24.2</i>	n.112203313A>G	None	Low	-
13	112313508	112313509	<i>RP11-65D24.2</i>	c.170-11236C>T	rs551876124	Low	0.15
13	112605700	112605701	<i>ALI36302.1-SWORD44</i>	n.112605701G>T	rs529136683	Low	0.14
13	112798120	112798121	<i>LINC00403-LINC01070</i>	n.112798121G>A	rs565231257	Low	0.41
13	112840752	112840753	<i>LINC00403-LINC01070</i>	n.112840753_112840754insT	rs562370447	Low	-
13	112883767	112883769	<i>LINC01070-LINC01043</i>	n.112883769delA	rs369608209	Low	-
13	113276405	113276406	<i>TUBGCP3-C13orf35</i>	n.113276406G>C	None	Low	-
13	113281868	113281869	<i>TUBGCP3-C13orf35</i>	n.113281869_113281870insTGTG	None	Low	-
13	113353553	113353554	<i>C13orf35</i>	c.259+1402G>A	None	Low	-
13	113534962	113534963	<i>ATP11A</i>	c.*11G>A	None	Medium	-
13	113699222	113699223	<i>MCF2L</i>	c.450-362T>G	rs145185774	Low	0.85
13	113756650	113756651	<i>ALI137002.1</i>	c.-43G>A	None	Low	-
13	113801835	113801836	<i>F10</i>	c.865+26C>T	rs183118165	Low	0.35
13	114073963	114073964	<i>ADPRHL1</i>	c.*3173C>T	rs147978889	Low	0.99
13	114075847	114075848	<i>ADPRHL1</i>	c.*1289T>C	rs112596428	Low	0.99
13	114075907	114075908	<i>ADPRHL1</i>	c.*1229C>A	rs111450368	Low	0.99
13	114075927	114075928	<i>ADPRHL1</i>	c.*1209A>G	rs186651486	Low	0.99
13	114076021	114076022	<i>ADPRHL1</i>	c.*1115T>A	rs112353951	Low	0.99
13	114076640	114076641	<i>ADPRHL1</i>	c.*496G>C	rs144952998	Low	0.99
13	114088197	114088198	<i>ADPRHL1</i>	c.380-16T>G	rs369286904	Low	0.70
13	114088198	114088199	<i>ADPRHL1</i>	c.380-17C>A	rs373227737	Low	0.71
13	114172890	114172891	<i>TMCO3</i>	c.1226-2040G>A	rs138992456	Low	0.50
13	114454632	114454633	<i>LINC00552-TMEM255B</i>	n.114454633C>T	rs142234274	Low	0.50
13	114481059	114481060	<i>TMEM255B</i>	c.252+892T>G>A	rs146119780	Low	0.30
13	114638627	114638628	<i>LINC00565-RASL3</i>	n.114638628C>T	rs113366916	Low	0.40

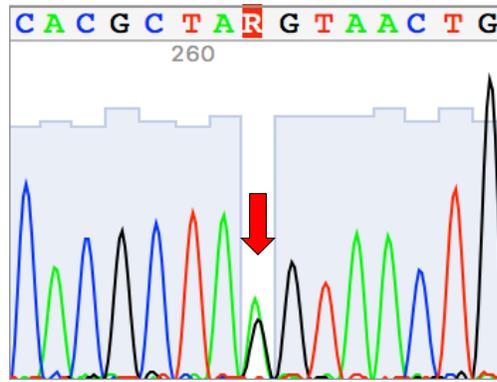


Figure 3.3. Sequence electropherogram illustrating the *ATP11A* c.*11G>A substitution. Red arrow indicates site of the heterozygous G to A variant.

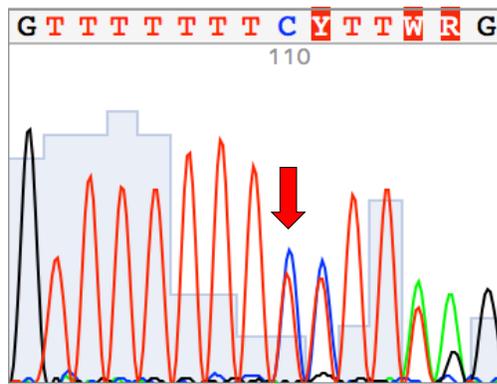


Figure 3.4. Sequence electropherogram *COL4A1* c.3326-7dupT variant. Red arrow indicates site of the heterozygous duplicated T.

(ENST00000415301.1; GERP: 5.160; CADD: 12.03), whereas *COL4A1* c.3326-7dupT medium impact variant (ENST00000375820.8; GERP: 2.860; ExAC: 0.002314; GnomAD: 0.00245815) has a 0.25% MAF. Both of these variants are absent in 326 ethnically-matched population controls. Importantly, two recombinant family members (PID III-4 and V-2; Figure 3.1) were recruited during the transcriptional analysis of the two splicing variants. Separate crossing over events occurred between *COL4A1* c.1221-18G>A and D13S285 (PID III-4), and D13S1295 and *ATP11A* c.*11G>A (PID V-2), which reduced the critical region to 769 Mb and excluded *COL4A1* from the disease haplotype. Several *in silico* tools predict that *COL4A1* c.3326-7dupT does not disrupt splicing, which was confirmed by RNA analysis (Figure 3.5). Consequently, this variant was not further investigated.

3.5.4 RNA analysis reveals multiple *ATP11A* transcripts

The *ATP11A* variant was identified as a medium impact variant in exon 2 of *ATP11A*-203. This transcript is poorly supported by a suspect EST and has an incomplete annotation of the 5' coding (CDS) sequence. In total, the *ATP11A* gene encodes 17 transcripts (Human GRCh38.p12 Ensembl 93 build(Yates *et al.*, 2016)), most of which are incompletely annotated and includes two RefSeq transcripts: 8,768 bp isoform b (*ATP11A-201* ENST00000375630.6; NM_032189) and 8,795 bp isoform a (*ATP11A-202* ENST00000375645.7; NM_015205). Given that *in silico* algorithms predicted that the *ATP11A* variant functionally disrupts a canonical donor splice site (Table 3.2), we performed RT-PCR with *ATP11A*-203-specific primers on EBV-transformed B-cell lines from peripheral blood of three controls (wild-type) and three affected *ATP11A*-203 c.*11G>A carriers (PID III-1, III-5 and III-7; Figure 3.1). Notably, multiple bands were observed in controls and *ATP11A* c.*11G>A carriers

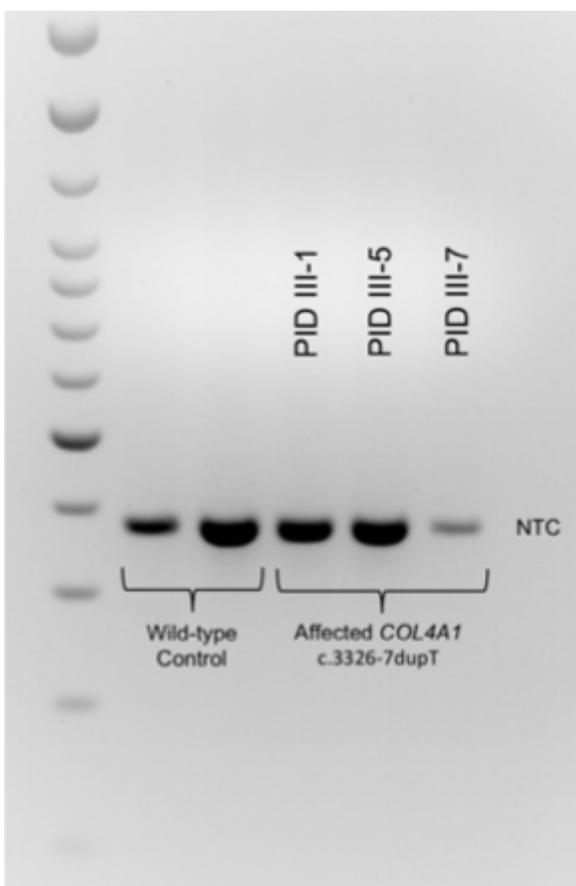


Figure 3.5. *COL4A1* c.3326-7dupT RNA Analysis. RT-PCR amplification in affected R2070 family members (PID III-1, III-5, and III-7) and wild-type controls. Despite our *in silico* analyses predicting that *COL4A1* c.3326-7dupT would not alter splicing, we felt it was necessary to experimentally validate this prediction in order to confidentially exclude this variant. As expected, *COL4A1* c.3326-7dupT RT-PCR experiments amplifies a single amplicon, shows that this variant does not affect splicing and concurs with our *in silico* analyses. NTC: non-templated control.

Table 3.2. *In silico* predictions for *ATP11A*, chr13:113534962G>A

Allele	Reference	Alternate	Difference
MaxENT	6.43	1.43	5
Human Splice Finder	70.8	41.85	28.95
Maximum Dependence Decomposition Model	10.08	3.98	6.1
First-order Markov Model	5.09	1.41	3.68
Weight Matrix Model	7.2	4.03	3.17

(Figure 3.6). Subsequent cloning revealed three bands in controls (1x~300 bp; 2x~370 bp) and three additional bands (1x~500 bp; 2x~550 bp) in the *ATP11A* c.*11G>A carrier (PID III-1; Figure 3.7). Sanger sequencing of the common lower molecular weight bands of *ATP11A*-203 transcript showed that the clones map to three alternatively spliced full-length transcripts: *ATP11A*-201 (Figure 3.8), *ATP11A*-202 (Figure 3.9) and *ATP11A*-212 (ENST00000487903.5; Figure 3.9). These data identify a missing exon from the open reading frame in the current *ATP11A*-201 and in the 3'UTR of *ATP11A*-202/212 Human (GRCh38.p12) Ensembl 93 build (Yates *et al.*, 2016). The cloning experiments indicate that the 104 bp exon (*ATP11A*-203) is likely the true previously uncharacterized exon in *ATP11A*-201 and *ATP11A*-202/212, preserving the 3' end of each transcript (Figure 3.8b & 3.9b). In one of the two RefSeq transcripts (*ATP11A*-201; Figure 3.8a), the addition of the missing exon (104 bp) is spliced to the 3' end of exon 28 revealing that the true stop codon (TAG) is located 159 bp upstream of the reported stop codon (Figure 3.8b). RefSeq transcripts (*ATP11A*-202/212; Figure 3.9a) have the missing exon (104 bp) spliced to the 3' end of exon 29, extending the 3'UTR (Figure 3.9b,d,e). Consequently, the c.*11 variant in exon 2 of *ATP11A*-203 is located at the same position in *ATP11A*-201 but positioned at c.*113 in *ATP11A*-202/212.

3.5.5 A pathogenic *ATP11A* splicing variant activates a cryptic donor splice site

According to splicing analyses, the *ATP11A* variant is predicted to activate a cryptic splice site 153 bp downstream of the canonical donor splice site. After correcting the annotation of the RefSeq transcripts, the composition of the three additional low molecular weight transcripts reveal that the variant of interest resides in the 3' boundary of exon 29 of *ATP11A*-201 (Figure 3.8c) and exon 30 of *ATP11A*-202/212 (Figure 3.9c). Sequencing of the three

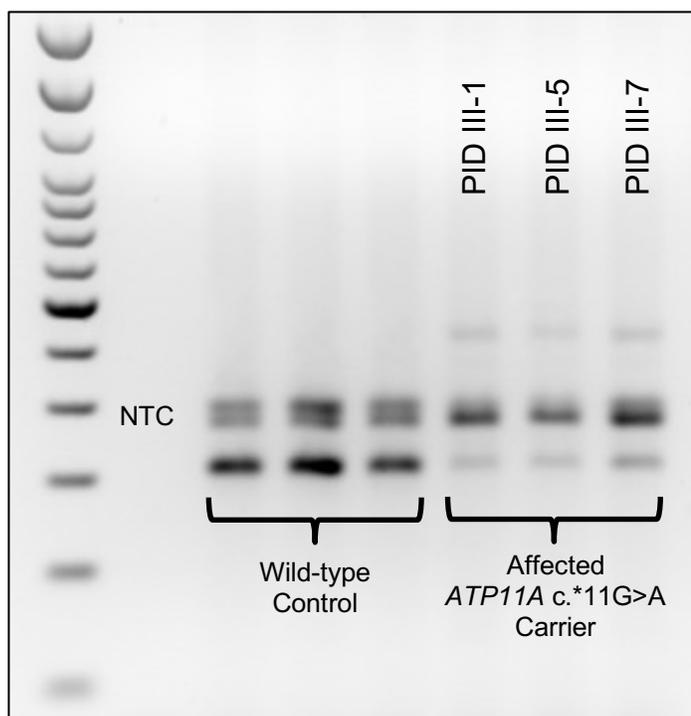


Figure 3.6. *ATP11A* RT-PCR. Amplification in affected R2070 family members (PID III-1, III-5, and III-7) and wild-types controls. NTC: Non-templated control

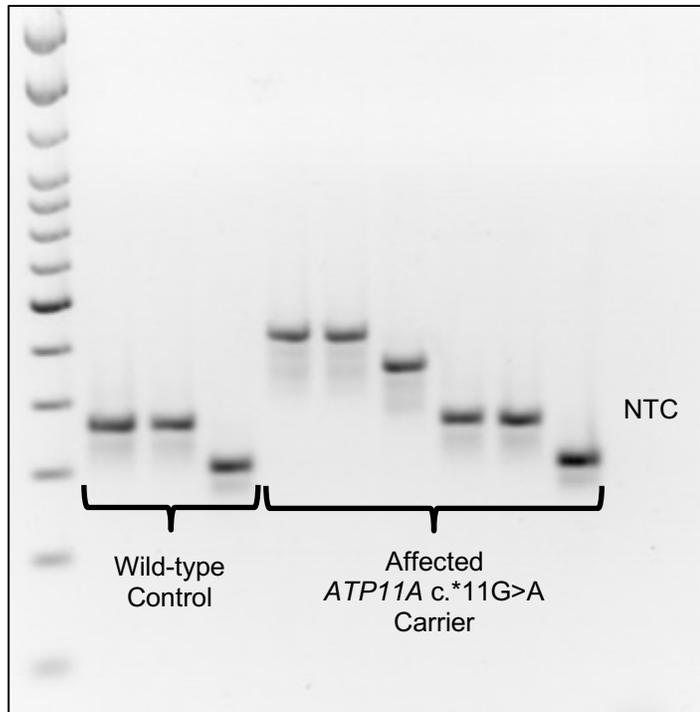


Figure 3.7 Colony PCR amplification of TA-cloned RT-PCR amplification in affected R2070 family members (PID III-1, III-5, and III-7) and wild-types controls. NTC: Non-templated control

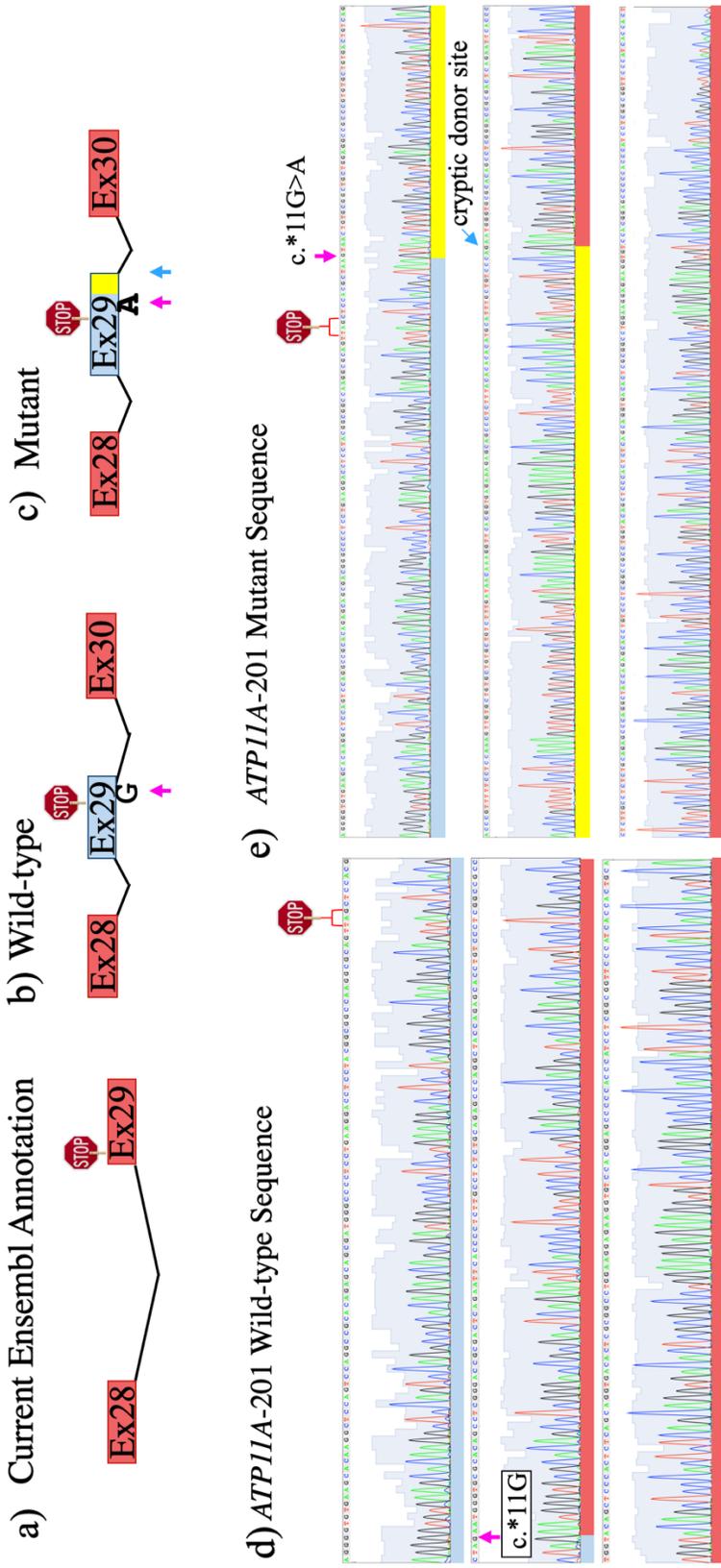


Figure 3.8. Wild-type and mutant *ATP11A-201* sequence traces. a) Current Ensembl annotation, lacking exon 29 (b), a newly identified exon. The addition of this exon introduces a new stop codon 156 base pairs upstream of the current Ensembl annotation. c) A pathogenic *ATP11A* variant activates a cryptic donor splice site. d) Wild-type sequence electropherogram illustrating c. *11G, e) Mutant sequence electropherogram demonstrating that c. *11G>A causes cryptic alternative splicing 153 base pairs into the intron (intrinsic sequence: yellow bar)

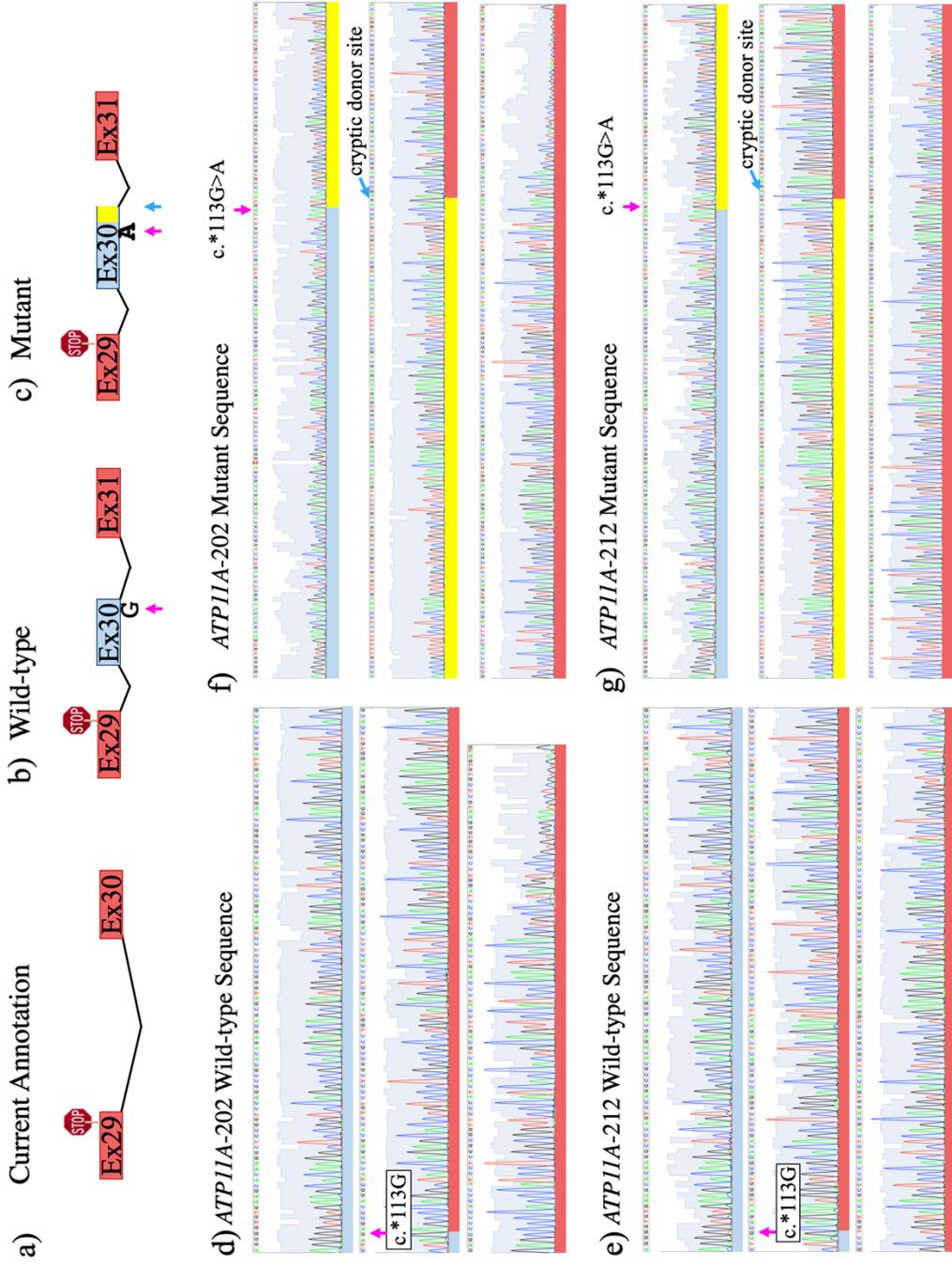


Figure 3.9. Wild-type and mutant *ATP11A-202* and *212* sequence traces. a) Current Ensembl annotation, lacking exon 30 (b), a newly identified exon. The addition of this exon adds 104 base pairs into the 3' UTR of these transcripts. c) A pathogenic *ATP11A-202/213* variant activates a cryptic donor splice site. d) Wild-type sequence electropherogram illustrating c. *113G, e) Mutant sequence electropherogram demonstrating that c. *113G>A causes cryptic alternative splicing 153 base pairs into the intron (intronic sequence: yellow bar)

aberrant higher molecular weight bands (Figure 3.7) in the *ATP11A* c.*11G>A/*113G>A carrier revealed that all contain the same 153 bp intronic sequence extending exon 29 (*ATP11A-201*; Figure 3.8e) and exon 30 (*ATP11A-202/212*; Figure 3.9f & g) experimentally confirming *in silico* prediction.

3.6 Discussion

We report a pathogenic splicing variant in the *ATP11A* gene that maps within the 6cM disease interval as the *DFNA33* locus (13q34), as first described by Bönsch *et al.* (Bönsch *et al.*, 2009). *ATP11A* c.*11G>A carriers in the family were diagnosed with low to mid frequency hearing loss during the first decade progressing to a sloping configuration. Two key critical recombination events on the disease-associated haplotype in unaffected relatives excluded all but a medium impact variant in *ATP11A*. Other P4-ATPase members are associated with syndromic forms of hearing loss and this study documents the first association of *ATP11A* with a highly penetrant Mendelian phenotype. The hearing loss in the 4 generation German family that map *DFNA33* exhibits similar audioprofiles, progressing to a flat hearing loss across all frequencies. Although we cannot be certain that *ATP11A* is *DFNA33*, it was noted by Bönsch *et al.* to be a functional candidate based on mouse studies. (Bönsch *et al.*, 2009) Although a decade has passed since *DFNA33* was mapped to chromosome 13q34-qter, no other families have been reportedly mapped to this locus.

This study highlights the importance of combining whole-genome sequencing, a comprehensive bioinformatics pipeline targeting all known transcripts and experimentally validating genomic findings in patient tissues. Using Ensembl, of the 17 annotated *ATP11A*

transcripts, we mapped the PCR amplicon containing the putative mutation to a short, poorly supported, isoform containing 3 exons total. However, downstream cDNA analysis in transformed white blood cells identified 3 long isoforms in control and patient samples. The *ATP11A* c.*11G>A carriers also had 3 additional high molecular weight isoforms. Cloning experiments suggest that the 104 bp exon (*ATP11A*-203) likely represents the true previously uncharacterized exon in *ATP11A*-201/202/212 that is unreported in Ensembl 93 build (Yates *et al.*, 2016). Our pathogenic *ATP11A* variant destroys the canonical donor splice site, activating a cryptic donor splice site 153 bp downstream. Though, the disease mechanism underlying the insertion of 153 bp into the 3' UTR of *ATP11A* is unclear. Given that the splicing variant is located in 3'UTR, it is unlikely that it affects protein structure; however, it might affect protein function through modulating *ATP11A* gene expression at the post-transcriptional level. The poor annotation of some of the *ATP11A* transcripts made it difficult to assess the effect of the variant on the mature RNA. The availability of tissue samples from affected and unaffected within the family and the multiplex pedigree structure were critical resources that helped to discern the altered RNA species expressed in the disease state.

A recent study demonstrated that there could be as little as 19,000 protein coding gene in the human genome (Ezkurdia *et al.*, 2014). While less than 200 genes are tissue specific (Mele *et al.*, 2015), it has been estimated that there is at least 205,000 protein-coding transcripts in the human genome (Hu *et al.*, 2015). One reason as to why we think this family remained unsolved until now is due to whole exome sequencing capture bias in primitive capture systems. The Ion Torrent AmpliSeq RDY Exome Kit systems is PCR-based and designed to capture 97.5% of genes contained within the Consensus CDS (CCDS) project (Damiati *et al.*, 2016) (O'Leary *et al.*, 2016; Wang *et al.*, 2017). Not only does this WES method create PCR amplification

bias(Aird *et al.*, 2011), it neglects transcripts that are not within the CCDS project(Damiati *et al.*, 2016), which misses an estimated 3% of coding variants and the majority of the over 8 million ESTs(Belkadi *et al.*, 2015; dbEST, 2019; Nagaraj *et al.*, 2007). In contrast, our WGS approach utilized a shotgun, PCR-free capture system that minimizes PCR-biases. This approach captures >98% of the entire genome and has been shown to be more robust in identifying causal variants during exploratory research(Belkadi *et al.*, 2015). In retrospect, we evaluated coverage and read-depth of *ATP11A*, and according to Ion Torrent metrics, all coding regions of this gene met our specified quality score values for variant filtration. Therefore, the Ion Torrent AmpliSeq RDY exome system must lack probes that amplify the coding regions of *ATP11A-203* and may explain why our *ATP11A* variant went undetected during our WGS analyses. Given that WGS identified a variant in an *ATP11A* exon that was previously unknown to the RefSeq database, this speaks to the complexity of the human transcriptome.

The *ATP11A* gene encodes for an integral membrane P4-ATPase, a phospholipid flippase that specifically catalyzes the energy dependant transport or “flip” of phospholipids from the outer to the inner leaflet of a phospholipid membrane(Paulusma & Elferink, 2010). While 14 P4-ATPase are annotated in the human genome, many are functionally related(Paulusma & Elferink, 2010; van der Mark *et al.*, 2013). For example, *ATP11A*, *ATP11C*, *ATP8A1*, and *ATP8A2* specifically utilize phosphatidylserine (PS) as a substrate(Lee *et al.*, 2015; Takatsu *et al.*, 2014). These P4-ATPases ensure that the outer leaflet of the plasma membrane is devoid of PS, which acts as an phagocytic “eat me” signal for in cells undergoing apoptosis when presented at the cell surface(Segawa *et al.*, 2014). Impressively, Segawa *et al.*(Segawa *et al.*, 2016; Segawa *et al.*, 2014) have demonstrated that *ATP11A* and *ATP11C* are the major flippases of mammalian cells and the loss of *ATP11A* results in PS presentation at the

cell surface, *in vitro*; however, other P4-ATPases, such as *ATP11C*, possess redundant compensatory roles. We propose that *ATP11A* haploinsufficiency leads to aberrant increase phagocytic signals due to increased PS levels at the surface of many cells of the auditory system (Figure 3.10). Additionally, the genetic background of compensatory epistatic factors that regulate the expression of other P4-ATPases must be considered. Given that our *ATP11A* variant is specific to 3 of 16 *ATP11A* transcripts, perhaps other transcripts are sufficient to maintain normal physiology across the body, as well as maintaining normal hearing during the first decade.

Mouse investigations that explore the role of P4-ATPases, such as *Atp8b1* and *Atp8a2*, demonstrate that these genes assimilate their human phenotype counterparts, including auditory deficits. Coleman *et al.* (Coleman *et al.*, 2014) found severe visual and auditory system defects in *Atp8a2*-deficient mice. Despite having intact sensory hair cells, *Atp8a2* knockout mice exhibit reduced auditory startle responses. Relative to wild-type mice, auditory brainstem responses and light microscopy in 2-month-old *Atp8a2*-deficient mice identified significantly higher hearing thresholds at 16 kHz and a marked reduction in the number of spiral ganglion cells in cross sections through the basal turns of cochlea, respectively. This finding is conceivable, as both the auditory and visual systems have been implicated in other human phenotypes, such as Usher syndrome (Keats & Corey, 1999). Another study has found that *ATP8B1* is essential for maintaining normal hearing (Stapelbroek *et al.*, 2009). This intrahepatic cholestasis mouse model harbored a homozygous missense variant, *Atp8b1*^{G308V/G308V}, which significantly decreased *Atp8b1* expression in mice (Pawlikowska *et al.*, 2004), and resulted in hair cell degeneration and abnormal auditory brainstem responses at 1, 3 and 6 months (Stapelbroek *et al.*, 2009).

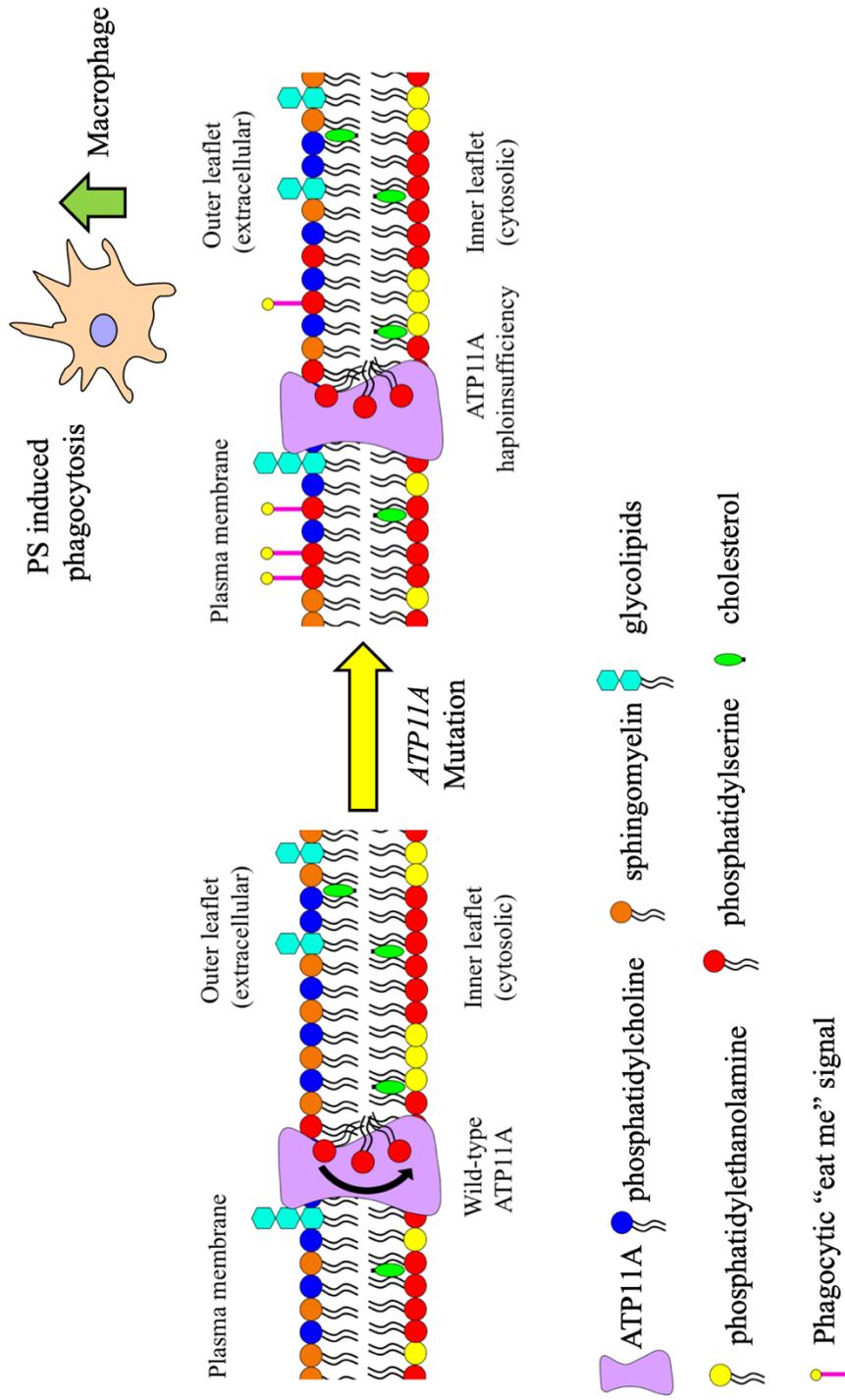


Figure 3.10. Proposed *ATP11A* disease mechanism. The main biological role of the ATP11A protein is to maintain phospholipid symmetry at the plasma membrane. Specifically, ATP11A ensures that the outer leaflet of the plasma membrane is devoid of phosphatidylserine. This is an important process, as the presentation of phosphatidylserine at the cell surface acts as a phagocytic “eat me” signal, marking cells that are undergoing apoptosis. Therefore, pathogenic *ATP11A* variants that were haploinsufficient would lead to aberrant phosphatidylserine induced phagocytosis in many cells of the auditory system.

Future directions include the quantitative analyses of *ATP11A* splicing variant, in addition to loss *in vitro* and *in vivo* functional characterization of the molecular mechanisms. Phagocytic signals, such as PS at the cell surface, are known pharmaceutical targets (Birge *et al.*, 2016), so there is potential that hearing loss due to *ATP11A* could be mitigated, especially since there is a brief therapeutic in those affected by non-congenital hearing loss. In order to functionally characterize our *ATP11A* splicing variant, several biological assays would be required. Firstly, a dual luciferase assay would demonstrate that defects would silence this gene at the post-transcriptional level. Briefly, two different luciferase assays would be performed – one group would have a wild-type *ATP11A* 3'UTR, while the other group would have our mutant *ATP11A* 3'UTR. Conceptually, the wild-type group would produce robust luciferase fluorescence, whereas the mutant group would exhibit a reduced signal. These findings would indicate that the cryptically spliced mutant *ATP11A* 3'UTR harbours miRNA sites that are not present in its wild-type counterpart. Subsequently, generating CRISPR-Cas9 *ATP11A* null HEI-OC1 cells (mammalian auditory sensory hair cell line) would enable researchers to determine the biological consequences of the loss of *ATP11A*. Since ATP11A is known to translocate PS to the inner leaflet of the plasma membrane, immunofluorescence would indicate the cytological location of PS in *ATP11A*-null HEI-OC1 cells. In addition, given that aberrant PS presentation at the cell surface induces apoptosis, a combination of immunoblotting for cleaved caspase-3 and TUNEL assays in wild-type and *ATP11A*-null HEI-OC1 cells would indicate whether programmed cell death was imminent. Assuming that these hypotheses held true, the final step would be to investigate the *atp11a*-null mouse, which is currently cryopreserved at The Jackson Laboratories in Bar Harbour, Maine, USA.

3.7 Acknowledgements

First and foremost, we thank the family members for participating. We would like to thank Dr. Curtis French for his expertise in TA-cloning. This work was supported by a grant from the Canadian Institutes for Health Research-Regional Partnership Program with the Research and Development Corporation of Newfoundland and Labrador, the Canadian Foundation for Innovation (New Investigator Award no. 9384 and Leaders Opportunity Fund no. 13120) and Genome Canada (Atlantic Medical Genetics and Genomics Initiative). The authors also gratefully acknowledge financial support from the Janeway Children's Hospital Foundation, Memorial University and the Government of Newfoundland and Labrador.

3.8 Bibliography

- Abdelfatah, N., McComiskey, D. A., Doucette, L., Griffin, A., Moore, S. J., Negrijn, C., . . . Young, T. L. (2013). Identification of a novel in-frame deletion in KCNQ4 (DFNA2A) and evidence of multiple phenocopies of unknown origin in a family with ADSNHL. *Eur J Hum Genet*, *21*(10), 1112-1119. doi:10.1038/ejhg.2013.5
- Abdelfatah, N., Merner, N., Houston, J., Benteau, T., Griffin, A., Doucette, L., . . . Young, T. L. (2013). A novel deletion in SMPX causes a rare form of X-linked progressive hearing loss in two families due to a founder effect. *Hum Mutat*, *34*(1), 66-69. doi:10.1002/humu.22205
- Ahmed, Z. M., Li, X. C., Powell, S. D., Riazuddin, S., Young, T. L., Ramzan, K., . . . Wilcox, E. R. (2004). Characterization of a new full length TMPRSS3 isoform and identification of mutant alleles responsible for nonsyndromic recessive deafness in Newfoundland and Pakistan. *BMC Med Genet*, *5*, 24. doi:10.1186/1471-2350-5-24
- Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., . . . Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*, *12*(2), R18. doi:10.1186/gb-2011-12-2-r18
- Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., . . . Abel, L. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*, *112*(17), 5473-5478. doi:10.1073/pnas.1418631112
- Birge, R. B., Boeltz, S., Kumar, S., Carlson, J., Wanderley, J., Calianese, D., . . . Herrmann, M. (2016). Phosphatidylserine is a global immunosuppressive signal in efferocytosis,

infectious disease, and cancer. *Cell Death Differ*, 23(6), 962-978.

doi:10.1038/cdd.2016.11

- Bonsch, D., Schmidt, C. M., Scheer, P., Bohlender, J., Neumann, C., Am Zehnhoff-Dinnesen, A., & Deufel, T. (2009). [A new gene locus for an autosomal-dominant non-syndromic hearing impairment (DFNA 33) is situated on chromosome 13q34-qter]. *HNO*, 57(4), 371-376. doi:10.1007/s00106-008-1832-9
- Coleman, J. A., Zhu, X., Djajadi, H. R., Molday, L. L., Smith, R. S., Libby, R. T., . . . Molday, R. S. (2014). Phospholipid flippase ATP8A2 is required for normal visual and auditory function and photoreceptor and spiral ganglion cell survival. *J Cell Sci*, 127(Pt 5), 1138-1149. doi:10.1242/jcs.145052
- Costa, G. L., & Weiner, M. P. (2006). Colony PCR. *CSH Protoc*, 2006(1). doi:10.1101/pdb.prot4141
- Damiati, E., Borsani, G., & Giacomuzzi, E. (2016). Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies. *Hum Genet*, 135(5), 499-511. doi:10.1007/s00439-016-1656-8
- dbEST. (2019). dbEST release 130101. Retrieved from https://www.ncbi.nlm.nih.gov/genbank/dbest/dbest_summary/
- Doucette, L., Merner, N. D., Cooke, S., Ives, E., Galutira, D., Walsh, V., . . . Young, T. L. (2009). Profound, prelingual nonsyndromic deafness maps to chromosome 10q21 and is caused by a novel missense mutation in the Usher syndrome type IF gene PCDH15. *Eur J Hum Genet*, 17(5), 554-564. doi:10.1038/ejhg.2008.231

- Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., . . . Tress, M. L. (2014). Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet*, *23*(22), 5866-5878. doi:10.1093/hmg/ddu309
- Fishelson, M., & Geiger, D. (2004). Optimizing exact genetic linkage computations. *J Comput Biol*, *11*(2-3), 263-275. doi:10.1089/1066527041410409
- Hildebrand, M. S., DeLuca, A. P., Taylor, K. R., Hoskinson, D. P., Hur, I. A., Tack, D., . . . Smith, R. J. (2009). A contemporary review of AudioGene audioprofiling: a machine-based candidate gene prediction tool for autosomal dominant nonsyndromic hearing loss. *Laryngoscope*, *119*(11), 2211-2215. doi:10.1002/lary.20664
- Hu, Z., Scott, H. S., Qin, G., Zheng, G., Chu, X., Xie, L., . . . Wei, C. (2015). Revealing Missing Human Protein Isoforms Based on Ab Initio Prediction, RNA-seq and Proteomics. *Sci Rep*, *5*, 10940. doi:10.1038/srep10940
- Keats, B. J., & Corey, D. P. (1999). The usher syndromes. *Am J Med Genet*, *89*(3), 158-166.
- Lee, S., Uchida, Y., Wang, J., Matsudaira, T., Nakagawa, T., Kishimoto, T., . . . Arai, H. (2015). Transport through recycling endosomes requires EHD1 recruitment by a phosphatidylserine translocase. *EMBO J*, *34*(5), 669-688. doi:10.15252/embj.201489703
- Mele, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., . . . Guigo, R. (2015). Human genomics. The human transcriptome across tissues and individuals. *Science*, *348*(6235), 660-665. doi:10.1126/science.aaa0355
- Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, *16*(3), 1215.

- Morton, C. C., & Nance, W. E. (2006). Newborn hearing screening--a silent revolution. *N Engl J Med*, *354*(20), 2151-2164. doi:10.1056/NEJMra050700
- Mouadeb, D. A., & Ruckenstein, M. J. (2005). Antiphospholipid inner ear syndrome. *Laryngoscope*, *115*(5), 879-883. doi:10.1097/01.MLG.0000158666.15447.37
- Nagaraj, S. H., Gasser, R. B., & Ranganathan, S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform*, *8*(1), 6-21. doi:10.1093/bib/bbl015
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, *44*(D1), D733-745. doi:10.1093/nar/gkv1189
- Pater, J. A., Benteau, T., Griffin, A., Penney, C., Stanton, S. G., Predham, S., . . . Young, T. L. (2017). A common variant in CLDN14 causes precipitous, prelingual sensorineural hearing loss in multiple families due to founder effect. *Hum Genet*, *136*(1), 107-118. doi:10.1007/s00439-016-1746-7
- Paulusma, C. C., & Elferink, R. P. (2010). P4 ATPases--the physiological relevance of lipid flipping transporters. *FEBS Lett*, *584*(13), 2708-2716. doi:10.1016/j.febslet.2010.04.071
- Pawlikowska, L., Groen, A., Eppens, E. F., Kunne, C., Ottenhoff, R., Looije, N., . . . Freimer, N. B. (2004). A mouse genetic model for familial cholestasis caused by ATP8B1 mutations reveals perturbed bile salt homeostasis but no impairment in bile secretion. *Hum Mol Genet*, *13*(8), 881-892. doi:10.1093/hmg/ddh100
- Richard JH Smith, E. S., Michael S Hildebrand, and Guy Van Camp. (1999 January 19, 2014). GeneReviews: Deafness and Hereditary Hearing Loss Overview. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK1434/>

- Segawa, K., Kurata, S., & Nagata, S. (2016). Human Type IV P-type ATPases That Work as Plasma Membrane Phospholipid Flippases and Their Regulation by Caspase and Calcium. *J Biol Chem*, *291*(2), 762-772. doi:10.1074/jbc.M115.690727
- Segawa, K., Kurata, S., Yanagihashi, Y., Brummelkamp, T. R., Matsuda, F., & Nagata, S. (2014). Caspase-mediated cleavage of phospholipid flippase for apoptotic phosphatidylserine exposure. *Science*, *344*(6188), 1164-1168. doi:10.1126/science.1252809
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, *29*(1), 308-311.
- Stapelbroek, J. M., Peters, T. A., van Beurden, D. H., Curfs, J. H., Joosten, A., Beynon, A. J., . . . Houwen, R. H. (2009). ATP8B1 is essential for maintaining normal hearing. *Proc Natl Acad Sci U S A*, *106*(24), 9709-9714. doi:10.1073/pnas.0807919106
- Takatsu, H., Tanaka, G., Segawa, K., Suzuki, J., Nagata, S., Nakayama, K., & Shin, H. W. (2014). Phospholipid flippase activities and substrate specificities of human type IV P-type ATPases localized to the plasma membrane. *J Biol Chem*, *289*(48), 33543-33556. doi:10.1074/jbc.M114.593012
- Van Camp G, S. R. (2015). Hereditary Hearing Loss Homepage. Retrieved from <http://hereditaryhearingloss.org>
- van der Mark, V. A., Elferink, R. P., & Paulusma, C. C. (2013). P4 ATPases: flippases in health and disease. *Int J Mol Sci*, *14*(4), 7897-7922. doi:10.3390/ijms14047897

- Wang, Q., Shashikant, C. S., Jensen, M., Altman, N. S., & Girirajan, S. (2017). Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep*, 7(1), 885. doi:10.1038/s41598-017-01005-x
- Wiles, N. M., Hunt, B. J., Callanan, V., & Chevretton, E. B. (2006). Sudden sensorineural hearing loss and antiphospholipid syndrome. *Haematologica*, 91(12 Suppl), ECR46.
- Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., . . . Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Res*, 44(D1), D710-716. doi:10.1093/nar/gkv1157
- Young, T. L., Ives, E., Lynch, E., Person, R., Snook, S., MacLaren, L., . . . King, M. C. (2001). Non-syndromic progressive hearing loss DFNA38 is caused by heterozygous missense mutation in the Wolfram syndrome gene WFS1. *Hum Mol Genet*, 10(22), 2509-2514.
- Zachowski, A. (1993). Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement. *Biochem J*, 294 (Pt 1), 1-14.
- Zhao, S., & Zhang, B. (2015). A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genomics*, 16, 97. doi:10.1186/s12864-015-1308-8

Chapter 4: Novel Usher syndrome pathogenic variants identified in cases with hearing and vision loss

Justin A. Pater¹, Jane Green¹, Anne Griffin¹, Sara Fernandez², Bridgette Fernandez^{1,2}, Darren D. O’Rielly³, Jim Houston¹, Susan G. Stanton⁴, Jiayi Zhou¹, Nicole Roslin⁵, Terry-Lynn Young^{1,3,4},

¹Craig L. Dobbin Research Centre, Faculty of Medicine, Memorial University, 300 Prince Phillip Drive, St. John’s, Newfoundland and Labrador, Canada, A1B 3V6

²Provincial Medical Genetics, Craig L. Dobbin Research Centre, Eastern Health, 300 Prince Phillip Drive, St. John’s, Newfoundland and Labrador, Canada, A1B 3V6

³Molecular Diagnostic Laboratory, Eastern Health, Craig L. Dobbin Genetics Research Centre, Faculty of Medicine, Memorial University, 300 Prince Phillip Drive, St. John’s, Newfoundland and Labrador, Canada, A1B 3V6

⁴Communication Sciences and Disorders, Western University, Elborn College, 1201 Western Road, London, Ontario, Canada, N6G 1H1

⁵The Centre for Applied Genomics, The Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, 686 Bay Street, Toronto, Ontario, M5G 0A4

A shorter version of this chapter has been accepted and in press at BMC Medical Genetics

(MGTC-D-18-00418R1)

4.1 Co-authorship Statement

The following authors were responsible for the following: BF, SF, JG, AG: contributed patients/clinical data, JAP, AG: evaluated audiology records, NR: performed linkage analysis, JAP, JH: sanger sequencing, JAP: whole exome sequencing, JZ, JAP, DDO: bioinformatic data analysis, JAP: RNA extractions, JAP: RT-PCR and TOPO-TA cloning, JAP: colony PCR, JAP, JH, DDO TLY: wrote/edited the manuscript, TLY, JG: directed the study.

4.2 Abstract

Background: Usher syndrome, the most common form of inherited deaf-blindness, is unlike many other forms of syndromic hereditary hearing loss in that the extra aural clinical manifestations are also detrimental to communication. Usher syndrome patients with early onset deafness also experience vision loss due to progressive retinitis pigmentosa that can lead to legal blindness in their third or fourth decade. **Methods:** Using a multi-omic approach, we identified three novel pathogenic variants in two Usher syndrome genes (*USH2A* and *ADGRV1*) in cases initially referred for isolated vision or hearing loss. **Results:** In a multiplex hearing loss family, two affected sisters, the product of a second cousin union, are homozygous for a novel nonsense pathogenic variant in *ADGRV1* (c.17062C>T, p.Arg5688*), predicted to create a premature stop codon near the N-terminus of *ADGRV1*. Ophthalmological examination of the sisters confirmed typical retinitis pigmentosa and prompted a corrected Usher syndrome diagnosis. In an unrelated clinical case, a child with hearing loss tested positive for two novel *USH2A* splicing variants (c.5777-1G>A, p. Glu1926_Ala1952del and c.10388-2A>G, p.Asp3463Alafs*6) and RNA studies confirmed that both pathogenic variants cause splicing errors. Interestingly, these same *USH2A* variants are also identified in another family with vision loss where subsequent clinical follow-up confirmed pre-existing hearing loss since early childhood, eventually resulting in a reassigned diagnosis of Usher syndrome. **Conclusion:** These findings provide empirical evidence to increase Usher syndrome surveillance of at-risk children. Given that novel antisense oligonucleotide therapies have been shown to rescue retinal degeneration caused by *USH2A* splicing pathogenic variants, these solved *USH2A* patients may now be eligible to be enrolled in therapeutic trials.

4.3 Introduction

Approximately 30% of inherited hearing loss is syndromic and is classically characterized by overt clinical features, such as distinctive craniofacial and eye abnormalities, and joint problems as in Stickler syndrome (Baker *et al.*, 2011; MIM: 108300). However, syndromic forms of hearing loss such as Usher syndrome (USH), present more insidiously, often resulting in delayed or misdiagnosis. USH is an autosomal recessive condition characterized by bilateral sensorineural hearing loss with or without vestibular dysfunction, and progressive retinitis pigmentosa (RP; Keats & Corey, 1999; Kimberling *et al.*, 2010; Mathur & Yang, 2015). Most children with USH are born with congenital hearing loss; however, progressive RP may present in the second decade, making diagnosis difficult due to the subtle changes in visual function over time (Yan & Liu, 2010). Historically, USH was considered an extremely rare disorder with a frequency of 1 in 25,000 (Boughman *et al.*, 1983); however, a recent study suggests a higher prevalence of 1 in 6,000 individuals in the European (Non-Finnish) population (Kimberling *et al.*, 2010).

USH is an extremely deleterious disorder and is the most common cause of inherited deaf-blindness (Kimberling *et al.*, 2010). So far, 13 USH genes have been identified which adversely affect the development of sensory hair cells within the inner ear and of photoreceptors in the eye (Yan & Liu, 2010). The most common subtype, USH type 2A (USH2A), accounts for two-thirds of all cases. Many *USH2A* pathogenic variants cause splicing defects such as exon skipping and the creation or destruction of canonical acceptor and donor splice sites (Yan & Liu, 2010). Novel therapies that target *USH2A* show great promise as retinal degeneration in USH2A patients can be rescued using antisense oligonucleotide-based therapy targeting cryptic splicing variants (Slijkerman *et al.*, 2016). Additionally, antioxidant-based

therapies have also shown great promise in preventing cone degeneration in USH1 mice, which is linked to oxidative stress (Trouillet *et al.*, 2018). Oxidative stress has well-established roles in many retinal dystrophies, where polymorphisms in *GLO1* may explain RP susceptibility and clinical heterogeneity (Donato *et al.*, 2018). Enrollment of patients in therapeutic trials requires a molecular diagnosis which can be challenging in the clinical setting. A comprehensive approach that includes linkage analysis, exome sequencing and functional analysis is often required, especially for novel splicing variants (Lewis *et al.*, 2018; Sakuma *et al.*, 2016). Herein, we report three novel USH pathogenic variants in *USH2A* or *ADVRG1* identified in cases of vision and hearing loss using a comprehensive multi-omic approach.

4.4 Materials and Methods

4.4.1 Study Participants and Clinical Evaluations

The study involved three families from the Newfoundland population, including two multiplex families. Clinical evaluations included air conduction thresholds using pure-tone audiometry, noting the audiogram configuration, severity, onset and progression. Vision was assessed with ocular examination, visual acuity and visual field testing, electroretinography (ERG) and fluorescein angiography of the retina.

For Family R2100, hearing loss (HL) is present in three sibships with varying audioprofiles, including two sisters who are the product of a consanguineous union (Figure 4.1). The proband (PID V-2) diagnosed with hearing loss at 3 years, presents by age 7 with a mild to moderate bilateral sensorineural HL, and her younger sister (PID V-3) was diagnosed at age 3 with a similar audioprofile (Figure 4.2).

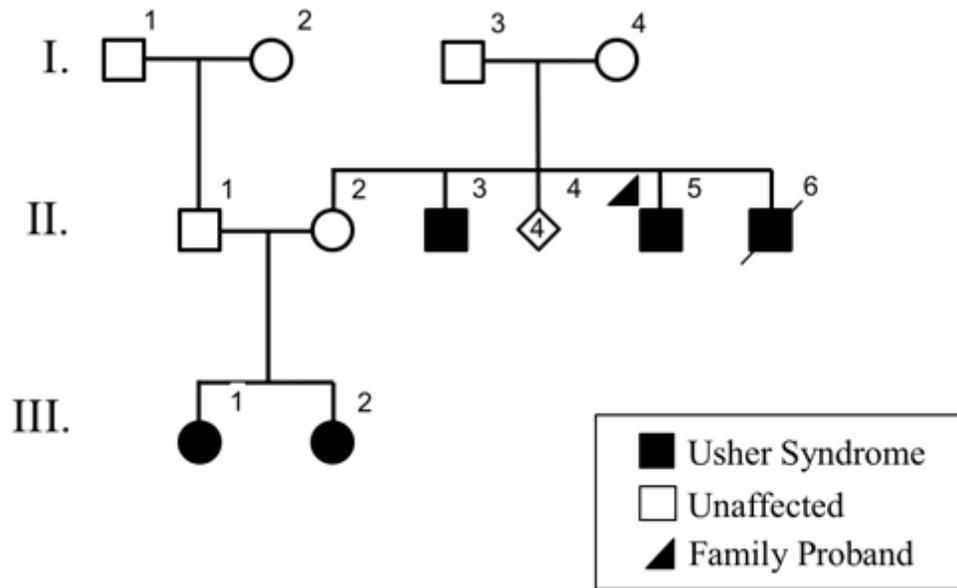


Figure 4.1. Family R2100 Pedigree. This family is a hereditary hearing loss pedigree with three affected sibships.

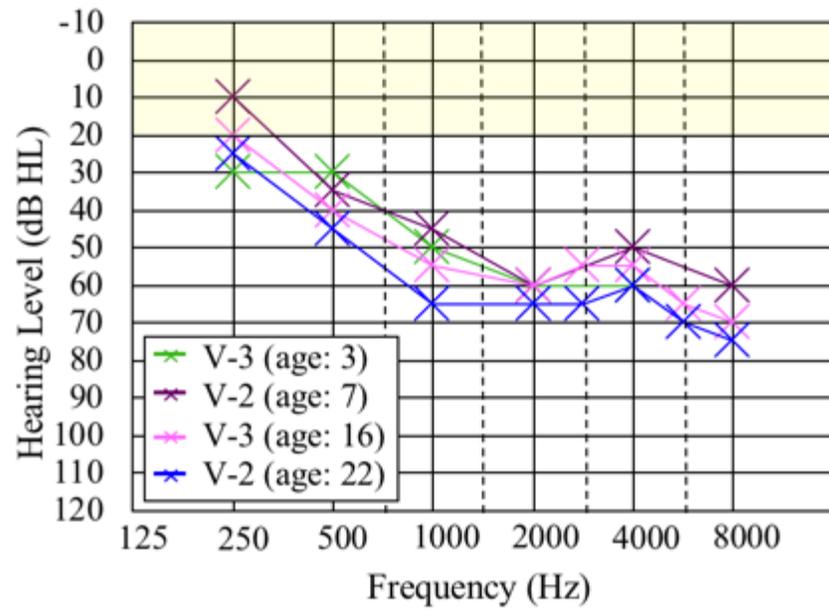


Figure 4.2. Family R2100 Audiological Data. Serial audiograms from PID V-2 and V3 demonstrates a stable hearing loss over two decades

We also recruited a case from our local medical genetics' clinic (Family R4110; Figure 4.3) of a child diagnosed at 3 months (following newborn hearing screening), who presents by age 3 with mild to moderate bilateral sensorineural HL (Figure 4.4). In another multiplex family (R0723), the proband and his brothers (PIDs II-5, II-3 and II-6, respectively) were first diagnosed with RP in mid 5th decade when their central vision decreased to the point that they met criteria for legally recognized blindness (Figure 4.5). They reported reduced night vision since the mid-second decade, and hearing loss since young childhood. The proband had been fitted for hearing aids for moderate to severe hearing loss. In R0273, the proband and his brothers (PIDs II-5, II-3 and II-6 respectively) had reported experiencing reduced night vision since their mid-third decade and were all diagnosed with RP in the mid-fifth decade when their central vision decreased. Throughout the course of ongoing clinical assessment, it was noted that the proband was fitted for hearing aids due to a moderate to severe HL and although the age of onset was unknown, he had HL at a young age. Two nieces with early hearing loss were also diagnosed with RP on follow-up, which prompted targeted genetic testing for known USH genes (Figure 4.4).

4.4.2 Gene Panels

In the case of family R0723, a targeted gene panel for 13 USH genes (CEI Molecular Diagnostic Laboratory, Portland, OR, USA) was offered through a research study on hereditary vision loss. In the clinical case of the child who failed newborn hearing screening (Family R4110), the family was offered targeted screening (158 syndromic and non-syndromic hearing loss genes, Blueprint Genetics, Comprehensive Hearing Loss and Deafness Panel, version 1, San Francisco, CA, USA). To validate variants of interest and check for co-segregation with RP

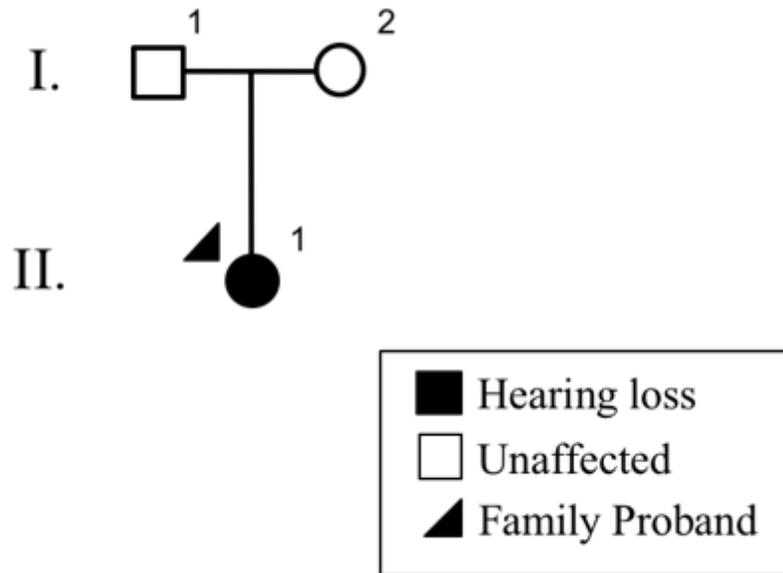


Figure 4.3. Family R4110 Pedigree. This family was recruited through the Newfoundland Provincial Genetics clinic. The proband of this family did not pass newborn hearing screens and was *GJB2*-negative

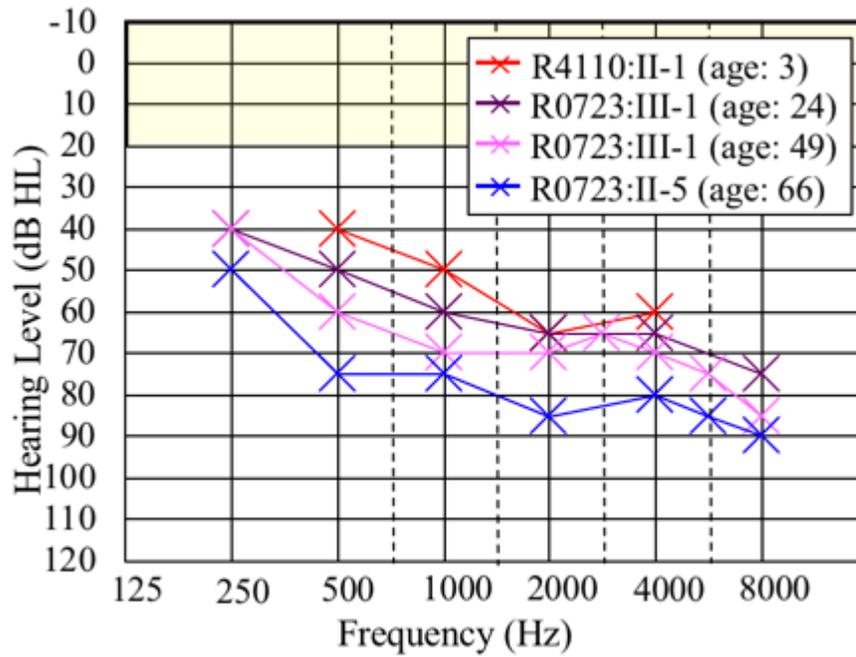


Figure 4.4. Family R4110 and R0723 Audiological Data. Audioprofile of PID II-1 (R4110) and PID III-1 and II-5 (R0723).

and/or HL trait in the families, genomic DNA was amplified using custom primers and sequenced in both directions using standard touchdown PCR protocols (ABI PRISM 3500XL DNA Analyzer; Applied Biosystems, Foster City, CA, USA). Sequence traces were analyzed using Mutation Surveyor Software (version 5.00, SoftGenetics LLC State College, PA 16803).

4.4.3 Linkage Analysis and Whole Exome Sequencing in Hearing Loss Family R2100

We initially screened the proband of Family 2100 for deafness alleles previously identified in the NL population (Appendix C) and submitted representative audiograms to Audiogene, a program comparing these to average audiograms of 34 deafness loci (Taylor *et al.*, 2013). As this targeted approach failed to solve Family 2100, traditional linkage study was done. For the linkage analysis, we selected three affected and two unaffected relatives (PID V-2, V-3, III-9, and IV-3, IV-4 respectively; Figure 4.1) and genotyped 17,407 polymorphic markers with the Illumina Human610-Quad chip. Multipoint linkage analysis (Merlin v1.1.2) (Abecasis *et al.*, 2002) was performed under a recessive model with a disease allele frequency of 0.07 and a penetrance of 99%. In order to screen candidate genes within linked regions, whole exome sequencing was carried out on 5 family members (two affected offspring and their parents: PID V-2, V-3, IV-3, IV-4 respectively) using the Ion Torrent AmpliSeq RDY Exome Kit (Life Technologies, Cat. #A27193). Purified libraries were loaded onto an Ion Proton PI v3 chip and sequenced with the Ion Torrent Proton. Only rare variants (MAF <1%) that mapped to linked regions, had a depth of coverage >20X and were of medium to high impact were validated by Sanger sequencing and selected for cascade screening. Population frequencies were determined using 124 ethnically-matched controls.

4.4.4 Splice variant *in silico* analysis

For variants of interest that reside within canonical ± 1 or 2 splice sites, we conducted *in silico* analyses using Alamut Visual (Interactive Biosoftware Inc., Rouen, France), a program that provides a splicing alteration report by linking to the MaxEnt, NNSPPLICE, SplicSiteFinder, and GeneSplicer algorithms.

4.4.5 RNA-cDNA analysis

In order to experimentally validate splicing predictions, we extracted total RNA from B-cell lymphocytes using standard TRIzol-based methods (Thermo-fisher, Cat. #15596026) and prepared cDNA libraries with the Superscript III First Stand Synthesis System (Thermo-fisher, Cat. #18080093). RT-PCR was carried out with primers that spanned candidate splicing regions, followed by TOPO TA-Cloning using One Shot TOP10 Chemically Competent E. coli cells (Invitrogen, #K457540) according to the manufacturer's protocol. RT-PCR products were Sanger sequenced and then analyzed using Mutation Surveyor Software (version 5.00, SoftGenetics LLC State College, PA 16803).

4.5 Results

4.5.1 *ADGRV1* c.17062C>T Genotype/Phenotype Analyses

In the step-wise analysis of hearing loss Family 2100, the proband (PID V-2; Figure 4.1) screened negative for all hearing loss variants previously identified in the NL population. Genome-wide linkage analysis (assuming autosomal recessive inheritance) yielded positive LOD scores suggestive of linkage for 8 genomic regions and the theoretical maximum LOD (1.68) for regions on chromosomes 3, 4, 5, 6 and 15 (Table 4.1; Appendix K). Subsequently,

Table 4.1 R2100 Multipoint Linkage Analysis Result

Chromosome	LOD	Start		End		Size (Mb)
		dbSNP	Position	dbSNP	Position	
1	0.58	rs3933251	64,095,165	rs591540	71,866,132	7.8
3	1.68	rs1400207	2,825,953	rs13084851	4,115,819	1.3
4	1.68	rs7671597	41,665,756	rs2035906	45,815,310	4.1
5	1.68	rs12110158	78,378,722	rs257239	97,997,738	19.6
6	1.67	rs1322633	125,082,133	rs1490388	126,514,509	1.4
11	1.64	rs1320211	15,301,410	rs10833818	22,795,026	7.5
15	1.68	rs937302	33,627,848	rs11070349	41,792,819	8.2
20	1.41	rs237417	5,770,142	rs4140471	7,226,740	1.5

exome sequencing identified 278 variants that were shared between the proband (PID V-2) and her affected sister (PID V-3). Of these, only eight variants remained after filtering for rare variants (MAF <1%) of medium to high impact that mapped to linked regions and had a depth of coverage >20X (Table 4.2). Seven of these variants were shown to be false positive INDELS (did not validate with Sanger sequencing) or did not reside within genes associated with syndromic or non-syndromic HL (Van Camp G, 2015). The remaining candidate (*ADGRV1* c.17062C>T; p.Arg5688*) is a nonsense variant associated with *USH2C* (Figure 4.6). Co-segregation analysis confirmed that the affected sisters were homozygous for *ADGRV1* c.17062C>T and their parents were unaffected carriers (Figure 4.7). The only other available affected relative for cascade sequencing was a maternal uncle (PID III-9) who was wild-type (two normal copies) and subsequently confirmed to have acquired his hearing loss after a serious diving accident. The nonsense *ADGRV1* variant is predicted to create a premature stop codon nearing the N-terminus of ADGRV1, preventing the translation of all 7 transmembrane domains. Furthermore, the *ADGRV1* c.17062C>T variant is absent in the population controls and has a single heterozygous entry in ExAC browser from the European (Non-Finnish) population. According to ACMG guidelines, the *ADGRV1* nonsense variant should be classified as pathogenic as it meets the following criteria: PVS1, PM2, PM3, PP1, PP3.

At the start of this study, we were aware of hearing loss (HL) in three sibships with varying audioprofiles, including two sisters who are the product of a consanguineous union. On the basis of the pending molecular diagnosis of Usher syndrome and the serious prognosis, the clinic contacted the sisters in order to request a visual examination (PID V-2 and V-3; Figure 4.1). The sisters are now in their late third and early fourth decade. Both women report impaired vision for some years. Ophthalmology reports on both sisters indicated definite

Table 4.2. Eight variants identified by whole exome sequencing

Chr	Gene	REF	ALT	SNP Position	dbSNP	ExAC MAF (%)	Protein Effect
6	<i>HEY2</i>	A	AC,C	126080841	None	None	INDEL
3	<i>WDR48</i>	GTTTTG	GTTT, GTTTG	39136139	None	None	INDEL
3	<i>RPL14</i>	A	ACTGCTG	40503520	None	None	INDEL
5	<i>MTX3</i>	G	C	79281458	None	None	Missense
5	<i>ADGRV1</i>	C	T	90144496	rs747622607	0.003235	Nonsense
5	<i>SPATA9</i>	A	G	95011189	rs55796768	0.958466	Missense
5	<i>ERAP1</i>	C	G	96139250	None	None	Missense
15	<i>GOLGA8A</i>	C	T	34673722	rs76522922	None	Missense

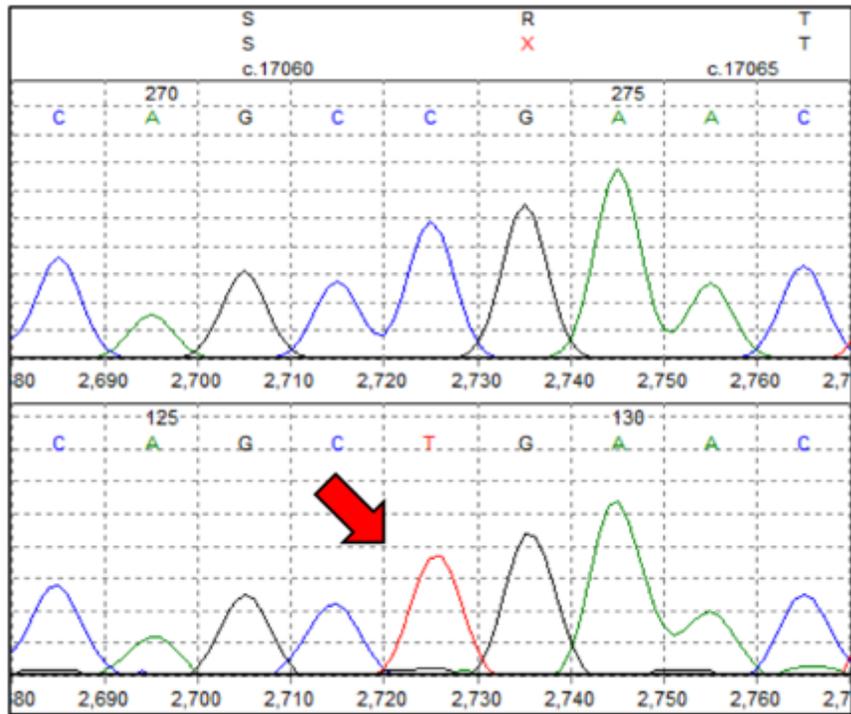


Figure 4.6. Sequence electropherogram of *ADGRV1* c.17062C>T (p.Arg5688Ter). Red arrow indicates this homozygous substitution

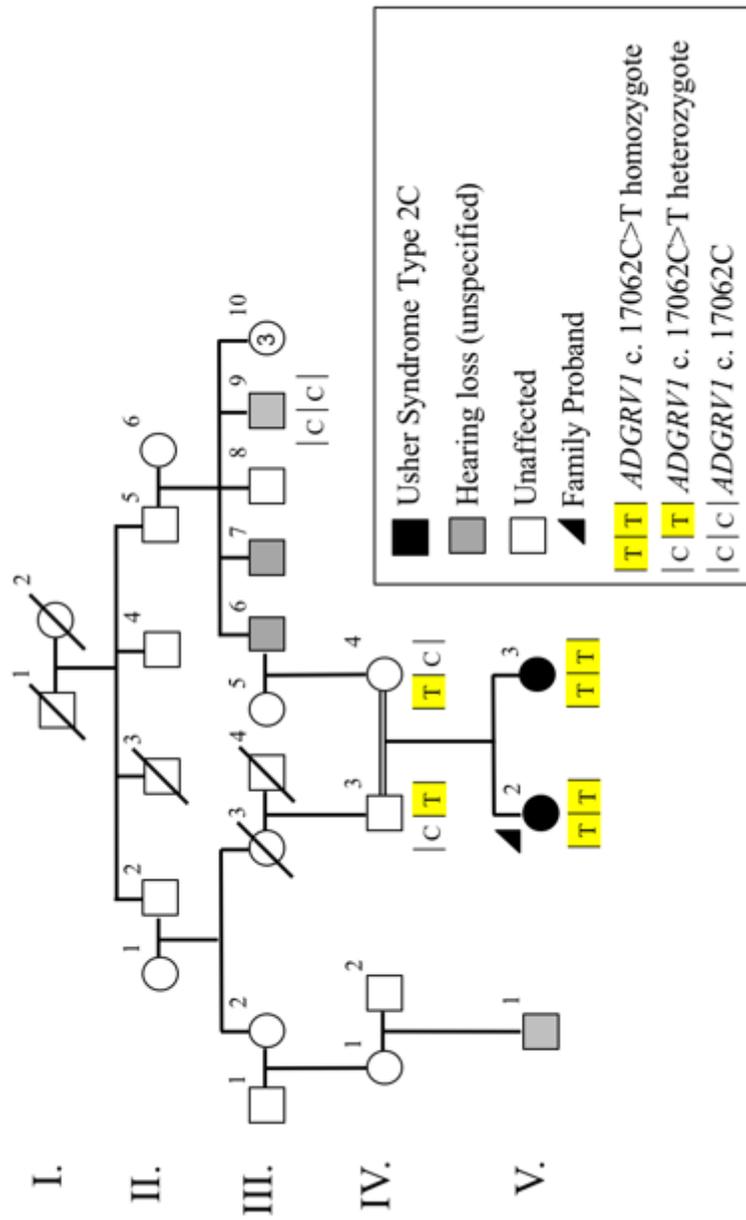


Figure 4.7. Family R2100 Pedigree with a pathogenic *ADGRV1* variant. This family is a hereditary hearing loss pedigree with three affected sibships. Whole exome sequencing identified a nonsense variant. *ADGRV1*c.17062C>T (p.Arg568Ter) that segregates in two sisters that have Usher Syndrome Type 2C. Prior to advent of genetic testing, vision impairment was not known.

features of RP. Further testing of PID V-3 identified bone spicule pigmentation of the retina (Figure 4.8) and a significant reduction in peripheral visual acuity (Figure 4.9), which are consistent with “typical RP”. These findings prompted the clinic to counsel the women regarding their new diagnosis of USH2C.

4.5.2 *USH2A* c.5777-1G>A and c.10388-2A>G Genotype/Phenotype Analyses

The comprehensive gene panel that was offered to the clinical case of the 3-year-old child diagnosed with isolated hearing loss at 3 months (Family 4110; Figure 4.3) identified two novel *USH2A* splicing variants: c.5777-1G>A (Figure 4.10) and c.10388-2A>G (Figure 4.11). Cascade sequencing confirmed the maternal contribution as c.5777-1G>A and the paternal contribution as c.10388-2A>G, and verified these novel variants reside *in trans* (Figure 4.12). However, given this child’s young age and the novelty (variants of unknown significance) of the *USH2A* variants, the genetic testing results are of limited value.

Fortuitously, the targeted USH gene panel offered to Family R0723 identified these same *USH2A* splicing variants. The proband (PID II-5) and his brother (PID II-3) are homozygous for *USH2A* c.5777-1G>A, their nieces (PIDs III-1 and III-2) are compound heterozygotes (c.5777-1G>A; c.10388-2A>G; Figure 4.13). Even though the deceased brother (PID II-6) was not available for genetic testing, he is likely a *USH2A* c.5777-1G>A homozygote, given the strong family history of RP. Retrospective audiological data on the proband’s niece, PID III-1, from mid-third decade to mid-fifth decade show a stable hearing loss according to GenDeaf guidelines (Figure 1d; Mazzoli M, 2003). The proband (PID II-5) has moderate to severe hearing loss in his seventh decade, not significantly worse than his younger niece (PID III-1) in her late fifth decade (Figure 4.4). His other niece, PID III-2,

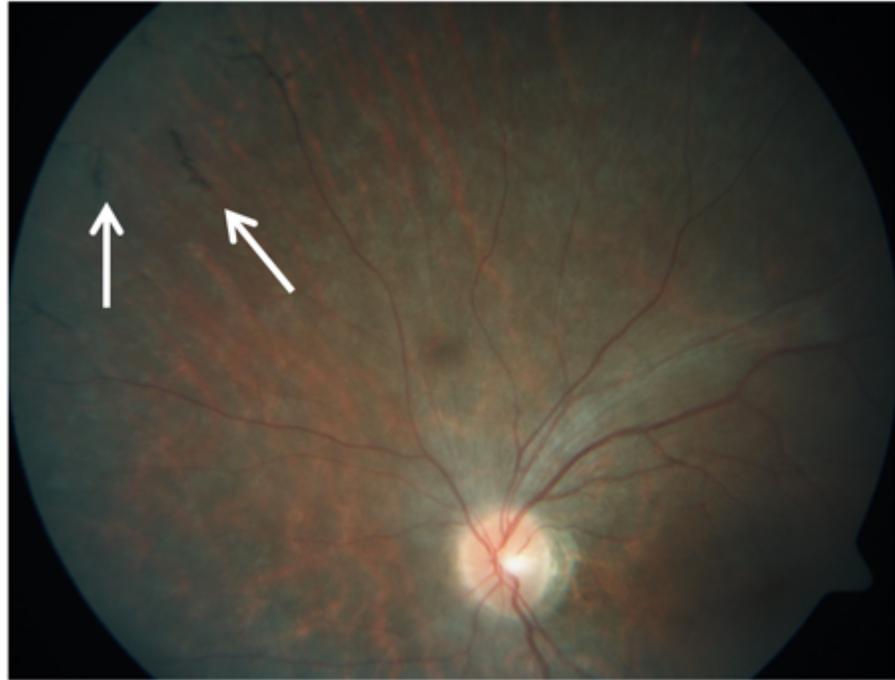


Figure 4.8. Retinal photograph of PID V-3 at age 27 (Family R2100). White arrows highlight the presence of bone spicules

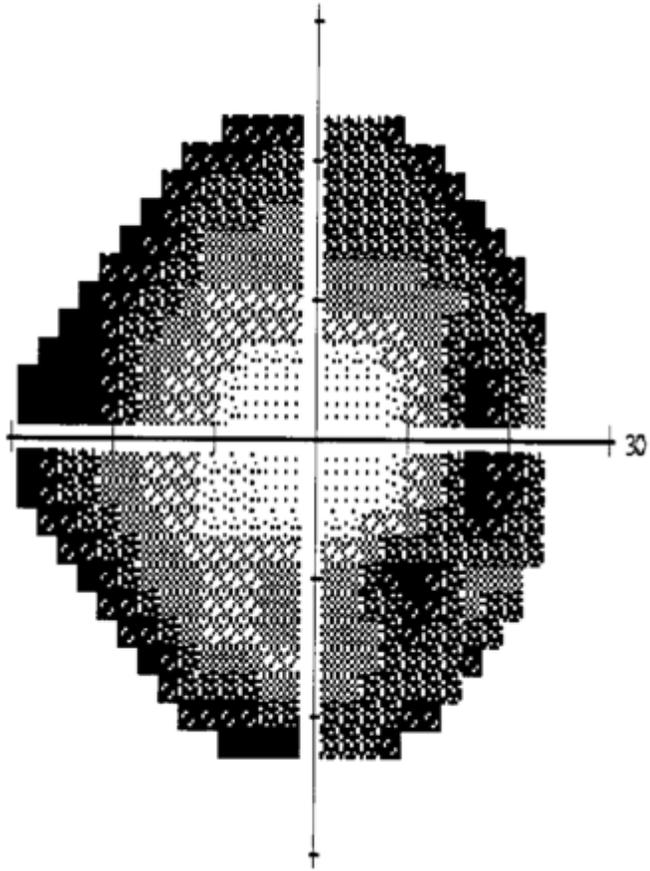


Figure 4.9. Central 24-2 visual threshold test of PID V-3 at age 27 (R2100) illustrating a deterioration of peripheral visual acuity. Darker shaded areas indicate loss of vision.

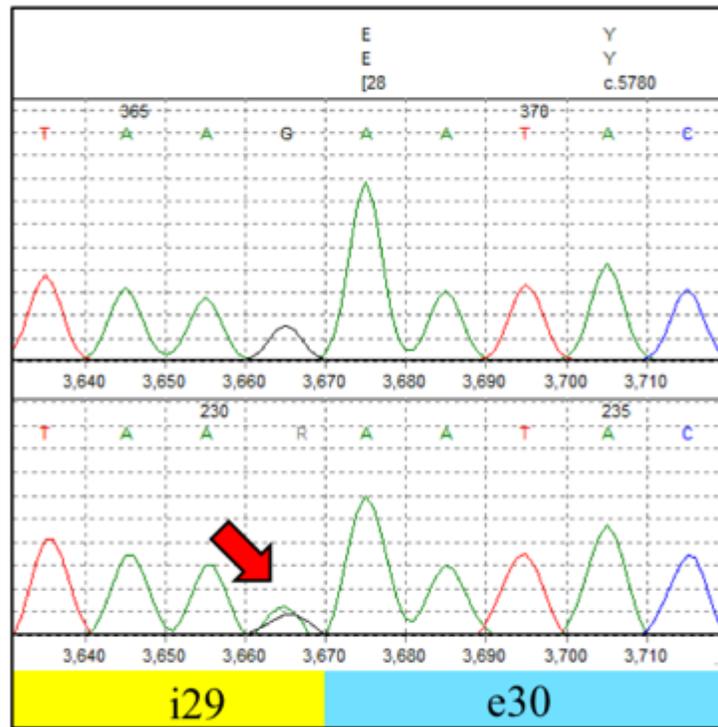


Figure 4.10. Sequence electropherogram of *USH2A* c.5777-1G>A, a pathogenic splicing variant that resides in intron 29. Red arrow indicates this homozygous substitution. i: intron, e: exon

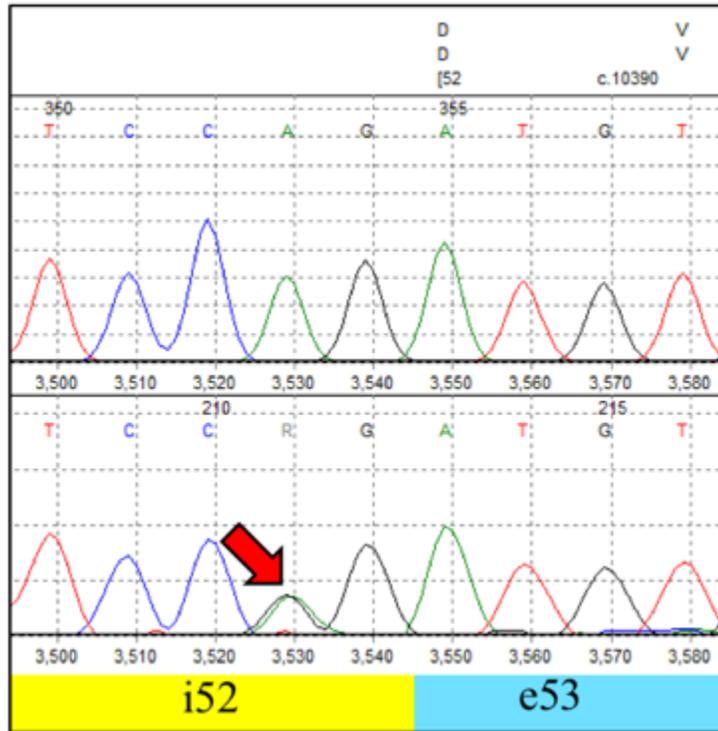


Figure 4.11. Sequence electropherogram of *USH2A* c.10388-2A>G, a pathogenic splicing variant that resides in intron 52. Red arrow indicates this homozygous substitution. i: intron, e: exon

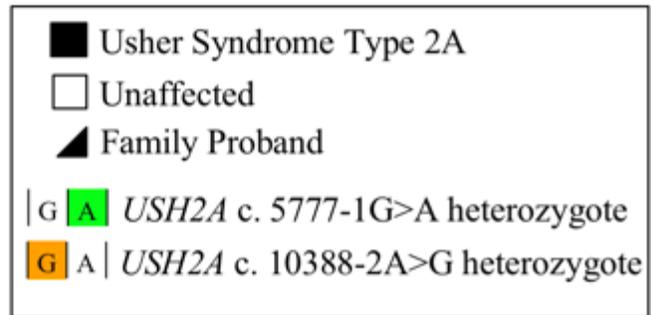
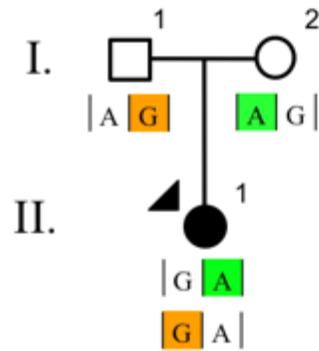


Figure 4.12. Family R4110 Pedigree. This proband was recruited through the clinic as a hereditary hearing loss family. A comprehensive gene panel identified two *USH2A* variants of unknown significance: c.5777-1G>A and c.10388-2A>G. PID II-1 is yet to present with retinitis pigmentosa, as she is too young to display this clinical feature of *USH2A*. Prior to advent of genetic testing, knowledge of potential vision impairment was not known.

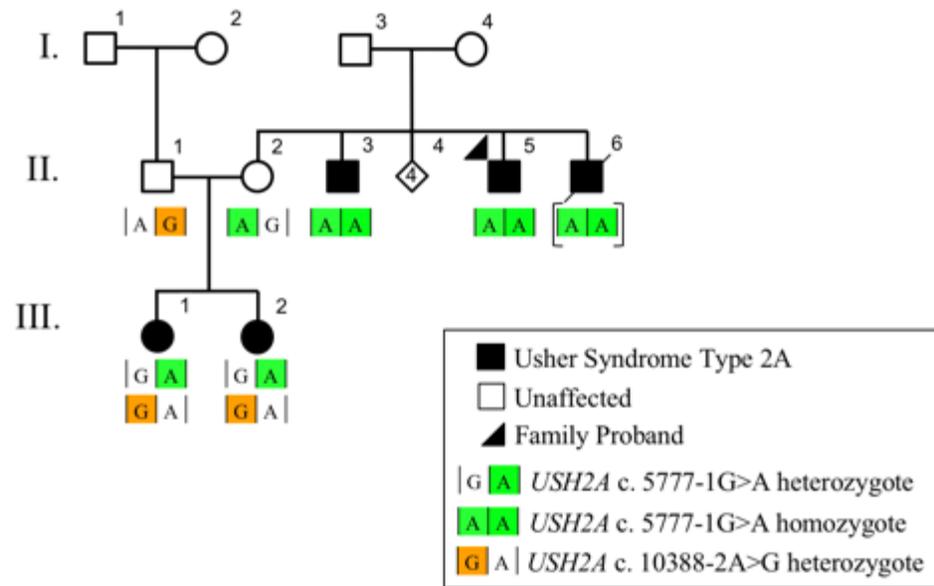


Figure 4.13. Family R0723 Pedigree. This proband was recruited through the clinic as an Usher syndrome family through a hereditary eye study. An Usher syndrome gene panel identified two *USH2A* variants of unknown significance: c.5777-1G>A and c.10388-2A>G. Three brothers, PID II-3, II-5 and II-6 are homozygous for c.5777-1G>A, while PID III-1 and III-2 are compound heterozygous for both c.5777-1G>A and c.10388-2A>G.

reveals a similar clinical phenotype (data not shown). With respect to RP, the proband (PID II-5) and his two brothers (PID II-3 and II-6) reported decreased night vision by their late 20s (Figure 1e); however, RP as seen in retinal photographs of the proband was not diagnosed in the brothers until their late 40's (Figure 4.14). Following the diagnosis of RP in the uncles, their nieces who had documented hearing loss were closely monitored, and reduced visual fields noted at age 14 in PID III-2, indicating the first symptoms of RP. Abnormal dark adaptation and ERG responses were recorded in both nieces in the third decade and retinal photographs of PID III-2 illustrate arterial attenuation, a characteristic sign of early RP (Figure 4.14).

4.3.3 *USH2A* c.5777-1G>A and c.10388-2A>G Experimental Validation of splicing effects

Cascade sequencing revealed that both *USH2A* c.5777-1G>A and *USH2A* c.10388-2A>G co-segregate with disease in families R0723 and R4110. *In silico* analyses using Alamut Visual suite of algorithms predicted that both variants cause exon skipping (MaxEnt: -100.0%, NNSPLICE: -100.0% and SSF: -100.0%). Using patient-derived cells, Sanger sequencing of cDNA confirmed that *USH2A* c.5777-1G>A causes the skipping of exon 29 leading to an in-frame deletion (p. Glu1926_Ala1952del) in an affected individual (PID III-2) compared with a control sample (Figure 4.15). The sequencing of patient cDNA also determined that *USH2A* c.10388-2A>G activates a cryptic acceptor site 14 bps downstream of the canonical splice site (Figure 4.16), resulting in a premature stop codon (p.Asp3463Alafs*6). Based on cascade sequencing within these families and subsequent RNA analysis, *USH2A* c.5777-1G>A and c.10388-2A>G can both be classified as pathogenic variants according to the ACMG guidelines (PVS1, PS3, PM2, PM3, PP1, PP3) (Richards *et al.*, 2015).

R0723: III-2 (age: 21)



R0723: II-5 (age: 45)

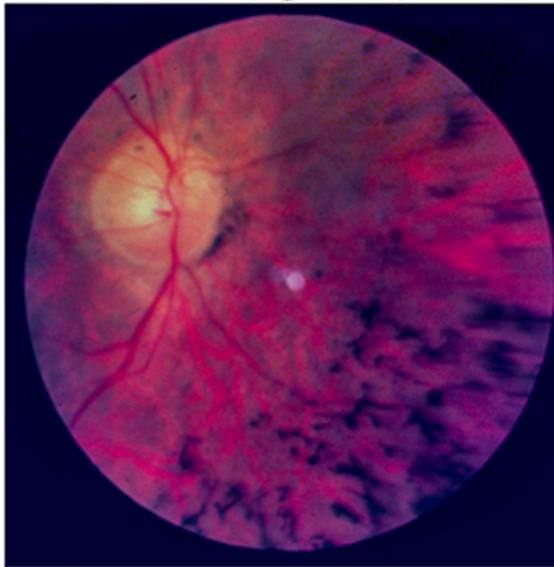
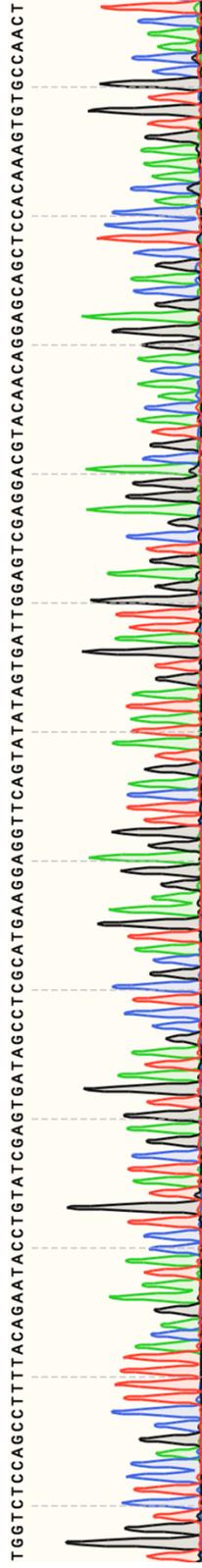


Figure 4.14. Family R0723 Ophthalmology Imaging. Retinal photograph of PID III-2 at age 21 demonstrates arterial attenuation in the retina, which further deteriorates by the fifth decade as seen in PID II-5 at age 45 (Family R0723; Figure 4.5).

USH2A c.5777-1G (wild-type allele, NM_206933.3)



USH2A c.5777-1G>A (mutant allele)

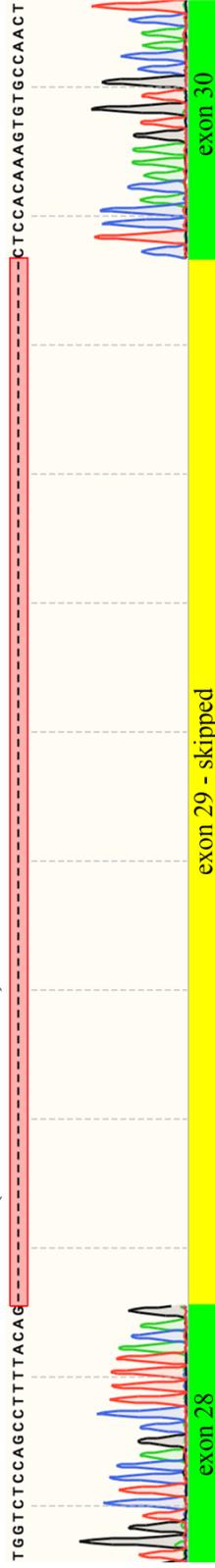


Figure 4.15. *USH2A* c.5777-1G>A RNA analysis. **a)** Sequence electropherogram of RT-PCR *USH2A* c.5777-1G>A. Top electropherogram show the reference sequence, while the middle and bottom electropherogram show the wild-type *USH2A* c.5777-1G and *USH2A* c.5777-1G>A alleles, respectively. Red horizontal line in *USH2A* c.5777-1G>A electropherogram indicates the skipping of exon 29, **b)** Schematic of cryptic alternative splicing, due to *USH2A* c.5777-1G>A. i: intron, e: exon

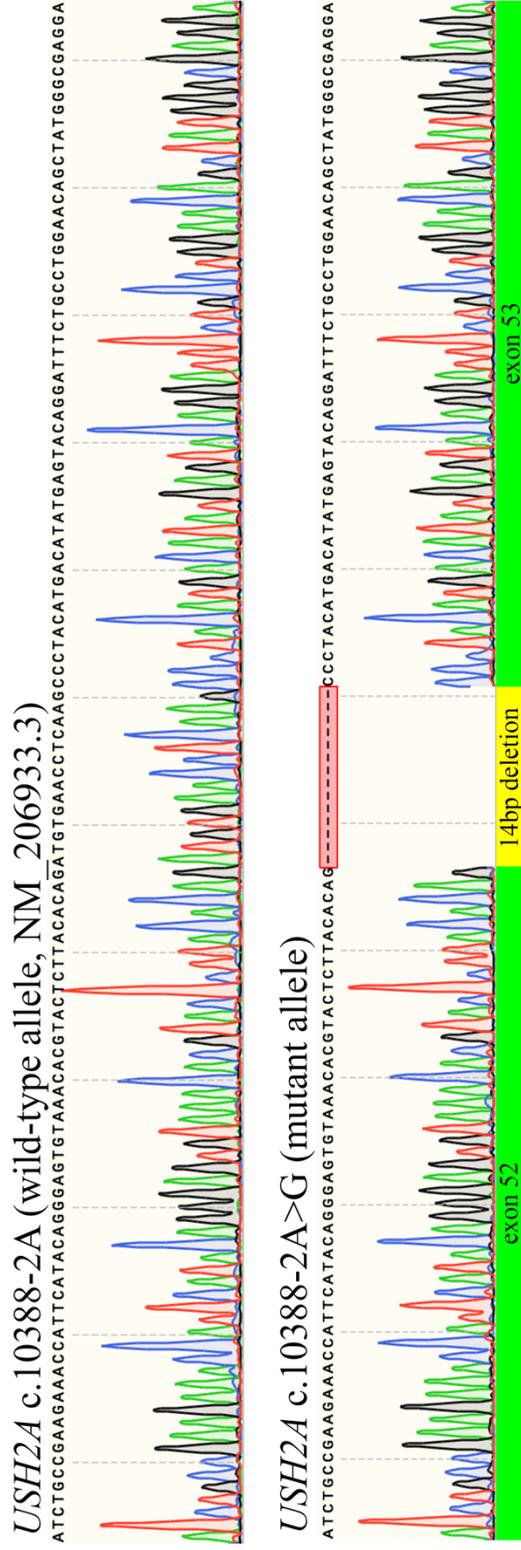


Figure 4.16. *USH2A* c.10388-2A>G RNA analysis. **a)** Sequence electropherogram of RT-PCR *USH2A* c.10388-2A>G. Top electropherogram show the reference sequence, while the middle and bottom electropherogram show the wild-type c.10388-2A and *USH2A* c.10388-2A>G alleles, respectively. Red horizontal line in *USH2A* c.10388-2A>G electropherogram indicates the activation of a cryptic acceptor site 14 bp downstream of the natural splice site, **b)** Schematic of cryptic alternative splicing, due to c.10388-2A>G. i: intron, e: exon

4.6 Discussion

Two novel pathogenic variants in *USH2A* account for cases recruited or referred as isolated hearing or vision loss in two families in this study. Clinical evidence suggests that the two novel *USH2A* pathogenic variants result in congenital moderate to severe HL, and RP in the pre/post-pubertal period; findings similar to that of previously reported pathogenic variants in *USH2A* (Lentz & Keats, 1993). Several affected family members present as compound heterozygotes, suggesting that both *USH2A* c.5777-1G>A and c.10388-2A>G pathogenic variants are sufficient to cause USH2A and therefore are USH2A-specific. This finding is consistent with the allelic hierarchy model of *USH2A* alleles, which suggests that certain alleles are USH2A-specific and others RP-specific, and the presence of at least one RP-specific allele causes isolated RP with normal hearing (Lenassi *et al.*, 2015). Given our clinical case with a young girl who also tested positive for both of these alleles, we are now increasing surveillance for visual symptoms, leading to improved management of USH.

Similarly, we find that two sisters with hereditary hearing loss, the product of a second cousin union, are homozygous for a nonsense pathogenic variant in *ADGRV1* (c.17062C>T, p.Arg5688*). Visual examination secondary to molecular analyses confirmed typical RP (late third and fourth decade) in addition to hearing loss (first decade) and prompted a corrected USH diagnosis. This is consistent with previous reports of *ADGRV1* pathogenic variants associated with early onset of hearing loss with delayed visual impairment (Abadie *et al.*, 2012; Fuster-Garcia *et al.*, 2018), most of which are located in the calx- β motif (Schwartz *et al.*, 2005), and the *ADGRV1* c.17062C>T lies downstream (3') to this calx- β motif. This variant is rare, and to our knowledge, has only been reported once before when it was identified in 1/31 French non-*USH2A* patients (Besnard *et al.*, 2012). In addition to causing USH2C, nonsense

ADGRV1 pathogenic variants have been shown to cause dominant audiogenic epilepsy (Nakayama *et al.*, 2002; Skradski *et al.*, 2001). However, the two affected sisters from R2100 whom are homozygous for *ADGRV1* c.17062C>T do not present with audiogenic epilepsy.

Clinically, USH2 should be suspected in patients with bilateral, congenital, sensorineural, mild to severe hearing loss, normal vestibular function, and post-pubertal RP, most often in the second decade (Lentz J, December 10, 1999; Mathur & Yang, 2015). Visual examinations revealed a ‘typical RP’ phenotype in patients diagnosed with USH2A or USH2C (Schwartz *et al.*, 2005). Likewise, from an audiological standpoint, our data is consistent with previous reports of a stable moderate to severe hearing loss (Abadie *et al.*, 2012; Besnard *et al.*, 2012; Eandi *et al.*, 2017; Garcia-Garcia *et al.*, 2013). These results indicate that USH2A and USH2C are not readily discerned phenotypically (Schwartz *et al.*, 2005). The USH2A, *ADGRV1* and *WHRN* proteins co-localize at the stereocilia base in developing cochlear hair cells and together form the Ankle-link complex at the base of sensory hair cells and at the periciliary membrane complex of photoreceptors (Liu *et al.*, 2007; Richardson & Petit, 2019; Yang *et al.*, 2010), so it is not surprising that the USH2A and USH2C phenotypes are indistinguishable.

To determine variant pathogenicity, clinical best-practice guidelines, such as ACMG (Richards *et al.*, 2015) and EuroGentest (Matthijs *et al.*, 2016) are important to follow. For the splicing variants, we used *in silico* prediction algorithms for preliminary assessment only, and experimentally confirmed the splicing effects using patient-derived B-cell lines. For the nonsense variant, we confirmed that the parents were unaffected carriers and that their affected children received one copy of the novel nonsense *ADGRV1* c.17062C>T variant from each of

them, establishing that we are detecting two disease alleles *in trans* and confirming the recessive pattern for Usher syndrome.

In a recent meta-analysis including all of the known genes causing usher syndrome, *USH2A* (50%) mutations are the most common with *ADGRVI* mutations being less frequent (5%) in patients with both visual and hearing impairments (Jouret *et al.*, 2019). In patients with seemingly isolated sensorineural deafness, 7.5% had disease-causing mutations in USH genes, and are therefore at high risk of developing RP. In isolated cases of ‘hearing loss’ or ‘vision loss’, it is important to screen both USH and RP genes, as an accurate diagnosis of Usher syndrome is essential for patient clinical follow-up, particularly the referral and access to the correct support systems.

4.7 Conclusions

Recognition of syndromic forms of both hearing and vision loss, especially Usher syndrome, is important given the major impact of these types of sensory losses on the acquisition of speech in children and quality of life for adults. In this report, USH was not considered in these cases until genetic testing was performed. Close collaboration between local clinics and molecular genetics researchers was necessary to fully categorized three novel USH variants as pathogenic using ACMG criteria. Accurate molecular diagnosis of patients is essential to provide new opportunities for patients and their families to enroll in therapeutic trials.

4.8 Acknowledgements

We are grateful to the families who participated in this study, to Cindy Penney and Amanda Morgan for technical assistance.

4.9 Bibliography

- Abadie, C., Blanchet, C., Baux, D., Larrieu, L., Besnard, T., Ravel, P., . . . Roux, A. F. (2012). Audiological findings in 100 USH2 patients. *Clin Genet*, *82*(5), 433-438.
doi:10.1111/j.1399-0004.2011.01772.x
- Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*, *30*(1), 97-101.
doi:10.1038/ng786
- Baker, S., Booth, C., Fillman, C., Shapiro, M., Blair, M. P., Hyland, J. C., & Ala-Kokko, L. (2011). A loss of function mutation in the COL9A2 gene causes autosomal recessive Stickler syndrome. *Am J Med Genet A*, *155A*(7), 1668-1672. doi:10.1002/ajmg.a.34071
- Besnard, T., Vache, C., Baux, D., Larrieu, L., Abadie, C., Blanchet, C., . . . Roux, A. F. (2012). Non-USH2A mutations in USH2 patients. *Hum Mutat*, *33*(3), 504-510.
doi:10.1002/humu.22004
- Boughman, J. A., Vernon, M., & Shaver, K. A. (1983). Usher syndrome: definition and estimate of prevalence from two high-risk populations. *J Chronic Dis*, *36*(8), 595-603.
- Donato, L., Scimone, C., Nicocia, G., Denaro, L., Robledo, R., Sidoti, A., & D'Angelo, R. (2018). GLO1 gene polymorphisms and their association with retinitis pigmentosa: a case-control study in a Sicilian population. *Mol Biol Rep*, *45*(5), 1349-1355.
doi:10.1007/s11033-018-4295-4
- Eandi, C. M., Dallorto, L., Spinetta, R., Micieli, M. P., Vanzetti, M., Mariottini, A., . . . Marchese, C. (2017). Targeted next generation sequencing in Italian patients with Usher syndrome: phenotype-genotype correlations. *Sci Rep*, *7*(1), 15681. doi:10.1038/s41598-017-16014-z

- Fuster-Garcia, C., Garcia-Garcia, G., Jaijo, T., Fornes, N., Ayuso, C., Fernandez-Burriel, M., . . . Millan, J. M. (2018). High-throughput sequencing for the molecular diagnosis of Usher syndrome reveals 42 novel mutations and consolidates CEP250 as Usher-like disease causative. *Sci Rep*, *8*(1), 17113. doi:10.1038/s41598-018-35085-0
- Garcia-Garcia, G., Besnard, T., Baux, D., Vache, C., Aller, E., Malcolm, S., . . . Roux, A. F. (2013). The contribution of GPR98 and DFNB31 genes to a Spanish Usher syndrome type 2 cohort. *Mol Vis*, *19*, 367-373.
- Jouret, G., Poirsier, C., Spodenkiewicz, M., Jaquin, C., Gouy, E., Arndt, C., . . . Lebre, A. S. (2019). Genetics of Usher Syndrome: New Insights From a Meta-analysis. *Otol Neurotol*, *40*(1), 121-129. doi:10.1097/MAO.0000000000002054
- Keats, B. J., & Corey, D. P. (1999). The usher syndromes. *Am J Med Genet*, *89*(3), 158-166.
- Kimberling, W. J., Hildebrand, M. S., Shearer, A. E., Jensen, M. L., Halder, J. A., Trzupek, K., . . . Smith, R. J. (2010). Frequency of Usher syndrome in two pediatric populations: Implications for genetic screening of deaf and hard of hearing children. *Genet Med*, *12*(8), 512-516. doi:10.1097/GIM.0b013e3181e5afb8
- Lenassi, E., Vincent, A., Li, Z., Saihan, Z., Coffey, A. J., Steele-Stallard, H. B., . . . Webster, A. R. (2015). A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants. *Eur J Hum Genet*, *23*(10), 1318-1327. doi:10.1038/ejhg.2014.283
- Lentz J, K. B. (December 10, 1999, 2016 Jul 21). GeneReviews: Usher Syndrome Type II. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK1341/>

- Lentz, J., & Keats, B. (1993). Usher Syndrome Type II. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews*((R)). Seattle (WA).
- Lewis, M. A., Nolan, L. S., Cadge, B. A., Matthews, L. J., Schulte, B. A., Dubno, J. R., . . . Dawson, S. J. (2018). Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes. *BMC Med Genomics*, *11*(1), 77. doi:10.1186/s12920-018-0395-1
- Liu, X., Bulgakov, O. V., Darrow, K. N., Pawlyk, B., Adamian, M., Liberman, M. C., & Li, T. (2007). Usherin is required for maintenance of retinal photoreceptors and normal development of cochlear hair cells. *Proc Natl Acad Sci U S A*, *104*(11), 4413-4418. doi:10.1073/pnas.0610950104
- Mathur, P., & Yang, J. (2015). Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities. *Biochim Biophys Acta*, *1852*(3), 406-420. doi:10.1016/j.bbadis.2014.11.020
- Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., . . . Bauer, P. (2016). Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet*, *24*(10), 1515. doi:10.1038/ejhg.2016.63
- Mazzoli M, G. V. C., Valerie Newton, Giarbini N, Declau F & Agnete Parving. (2003). Recommendations for the description of genetic and audiological data for families with nonsyndromic hereditary hearing impairment. *Audiological Medicine*, *1*(2), 148-150.

- Nakayama, J., Fu, Y. H., Clark, A. M., Nakahara, S., Hamano, K., Iwasaki, N., . . . Ptacek, L. J. (2002). A nonsense mutation of the MASS1 gene in a family with febrile and afebrile seizures. *Ann Neurol*, *52*(5), 654-657. doi:10.1002/ana.10347
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, *17*(5), 405-424. doi:10.1038/gim.2015.30
- Richardson, G. P., & Petit, C. (2019). Hair-Bundle Links: Genetics as the Gateway to Function. *Cold Spring Harb Perspect Med*. doi:10.1101/cshperspect.a033142
- Sakuma, N., Moteki, H., Takahashi, M., Nishio, S. Y., Arai, Y., Yamashita, Y., . . . Usami, S. (2016). An effective screening strategy for deafness in combination with a next-generation sequencing platform: a consecutive analysis. *J Hum Genet*, *61*(3), 253-261. doi:10.1038/jhg.2015.143
- Schwartz, S. B., Aleman, T. S., Cideciyan, A. V., Windsor, E. A., Sumaroka, A., Roman, A. J., . . . Jacobson, S. G. (2005). Disease expression in Usher syndrome caused by VLGR1 gene mutation (USH2C) and comparison with USH2A phenotype. *Invest Ophthalmol Vis Sci*, *46*(2), 734-743. doi:10.1167/iovs.04-1136
- Skradski, S. L., Clark, A. M., Jiang, H., White, H. S., Fu, Y. H., & Ptacek, L. J. (2001). A novel gene causing a mendelian audiogenic mouse epilepsy. *Neuron*, *31*(4), 537-544.
- Slijkerman, R. W., Vache, C., Dona, M., Garcia-Garcia, G., Claustres, M., Hetterschijt, L., . . . Van Wijk, E. (2016). Antisense Oligonucleotide-based Splice Correction for USH2A-

associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. *Mol Ther Nucleic Acids*, 5(10), e381. doi:10.1038/mtna.2016.89

Taylor, K. R., Deluca, A. P., Shearer, A. E., Hildebrand, M. S., Black-Ziegelbein, E. A., Anand, V. N., . . . Casavant, T. L. (2013). AudioGene: predicting hearing loss genotypes from phenotypes to guide genetic screening. *Hum Mutat*, 34(4), 539-545.

doi:10.1002/humu.22268

Trouillet, A., Dubus, E., Degardin, J., Estivalet, A., Ivkovic, I., Godefroy, D., . . . Picaud, S. (2018). Cone degeneration is triggered by the absence of USH1 proteins but prevented by antioxidant treatments. *Sci Rep*, 8(1), 1968. doi:10.1038/s41598-018-20171-0

Van Camp G, S. R. (2015). Hereditary Hearing Loss Homepage. Retrieved from <http://hereditaryhearingloss.org>

Yan, D., & Liu, X. Z. (2010). Genetics and pathological mechanisms of Usher syndrome. *J Hum Genet*, 55(6), 327-335. doi:10.1038/jhg.2010.29

Yang, J., Liu, X., Zhao, Y., Adamian, M., Pawlyk, B., Sun, X., . . . Li, T. (2010). Ablation of whirlin long isoform disrupts the USH2 protein complex and causes vision and hearing loss. *PLoS Genet*, 6(5), e1000955. doi:10.1371/journal.pgen.1000955

Chapter 5: General Discussion and Conclusions

The overarching goal of this dissertation was to define and characterize the genetic basis of hereditary hearing loss in the NL genetic isolate. This thesis employs a “multi-omic” approach to gene discovery by combining both traditional robust methods such as linkage analysis and haplotyping, when large families are available, with modern NGS technologies and molecular tools that discern the impact of genomic variation at the RNA-level. Genetic isolates and founder populations have provided unprecedented opportunities in uncovering the genetic factors that contribute to disease. In lieu of identifying many pathogenic variants that have been enriched in the NL population, due to founder effects (Abdelfatah *et al.*, 2013; Ahmed *et al.*, 2004; Kopciuk *et al.*, 2009; Merner *et al.*, 2008; Olufemi *et al.*, 1998; Spirio *et al.*, 1999), this genetic isolate further exemplifies the importance and power of performing genomic studies on homogenous populations.

Non-syndromic, autosomal recessive, hearing loss accounts for the vast majority of monogenic deafness and is most commonly caused by homozygous or compound heterozygous variants in *GJB2* (Richard JH Smith, 1999). Until recently, many clinical molecular genetics laboratories only tested *GJB2* during routine newborn hearing screening investigations, representing a significant ascertainment bias that may skew the true genetic epidemiology of recessive hearing loss (Snoeckx *et al.*, 2005). Despite these biases, targeted genomic enrichment and NGS has accelerated the molecular diagnosis of hearing loss, with a diagnostic rate ranging from 10% - 84% (Shearer & Smith, 2015). Many factors may be contributing to this variable diagnostic rate, including platform enrichment bias that captures a specific set of known hearing loss genes and study participant bias that examines specific ethnicities with a known family history of deafness (Sloan-Heggen *et al.*, 2016). As outlined in the general introduction, the

hereditary hearing loss project began with the targeted sequencing of known pathogenic variants in all recruited families in the study. All families who screened negative were subject to more comprehensive genomic investigations.

After screening negative for known hearing loss alleles, Family R2010 underwent WES, which identified a pathogenic missense variant, *CLDN14* c.488C>T (p.Ala163Val; *DFNB29*), in several NL families. Upon initial inspection of hearing loss inheritance in Family 2010, we believed that this was a dominant trait, due to the vertical transmission from the proband to her son. Throughout ongoing clinical recruitment during a rural outreach field trip to the Burin peninsula, our hearing loss reach team noted that that the audioprofile within this family was remarkably similar amongst the kinship, with intact hearing threshold in the lower frequencies that precipitously deteriorate after 500 Hz. Our clinical audiologist noted that this specific phenotype was present in two additional families, R2033 and R2075, who screened positive for *CLDN14*. Subsequently, the sequencing of hearing loss probands revealed a fourth *CLDN14* family, R2072, who also presented with this precipitous, pathognomonic, audioprofile. Although genealogical studies identified a common ancestor between families R2010, R2033, and R2075, haplotype analysis identified a 1.4 Mb disease haplotype across all four families, indicating shared ancestry between R2072 and the larger clan. This *CLDN14* allele is common to the NL population, which explains the pseudodominant inheritance pattern that is observed in two of the families.

Within ExAC browser, *CLDN14* c.488C>T has 31 heterozygous submissions, with a global MAF of 0.02564% (Lek *et al.*, 2016). After first being identified in Iceland(Thorleifsson *et al.*, 2009), carriers for this pathogenic variant was reported in the USA(Toka *et al.*, 2013), Sweden(Purcell *et al.*, 2014), and Africa(Lek *et al.*, 2016). Given the presence

CLDN14 c.488C>T in many populations, this allele should be of interest to molecular diagnostic laboratories around the world. Until our *DFNB29* study, pathogenic variants in *CLDN14* have been shown to cause congenital hearing loss with a phenotype that exhibits variable expressivity (R. Bashir *et al.*, 2010; Z. E. Bashir *et al.*, 2013; Wilcox *et al.*, 2001). However, we have shown that pathogenic *CLDN14* variants can cause non-congenital, prelingual hearing loss, with considerable conservation in audiogram configuration. Without genetic testing, children who are *CLDN14*-positive would not be detected at newborn hearing screening.

Since the publication of Chapter 2, *CLDN14* c.488C>T has been included in a targeted hearing loss gene panel that has identified several NL families who were *CLDN14*-positive, highlighting the urgent need for NGS technologies in the clinical setting. Not only can this approach expedite the molecular diagnosis of hearing loss, it can also provide appropriate hearing surveillance for children at-risk of developing significant delays that lead to language and learning deficits. Even though NGS is an effective diagnostic modality, the lack of extended pedigrees in the clinic presents a significant challenge during variant interpretation and it is recommended to use caution when the phasing of disease alleles and segregation analysis is unavailable (Lewis *et al.*, 2018). Given that there are the millions of heterozygous variants across the human genome (Sherry *et al.*, 2001), performing segregation analysis is especially relevant to autosomal dominant traits. Despite the identification of approximately 143 hearing loss genes, approximately 1/3 of *DFNA* loci have evaded discovery^{72,76}, due to the lack of large kinships, broad critical regions, well-described phenotype data, as well as reduced penetrance and the inter- and intrafamilial variability of dominant hearing phenotypes (Richard JH Smith, 1999).

Chapter 3 represents the most significant scientific discovery of this dissertation, the identification of a novel autosomal dominant gene that maps to the *DFNA33* locus. The purpose of this study was to discern the genetic basis of hearing loss in Family R2070, a large multiplex kinship from the Western region of NL with apparent autosomal dominant hearing loss. Several genomic technologies were employed throughout this project. Previous work including genome-wide SNP genotyping and multipoint linkage analysis mapped the hearing loss trait in family R2070 to chromosome 13q34. Even though WES failed to solve this family, the combination of WGS and a comprehensive bioinformatics approach identified two candidate variants within this region. The first variant, *COL4A1* c.3326-7dupT, was in close proximity to the exon-intron boundary; however, experimental RNA evidence demonstrates that this variant exerts no effect at the RNA-level. The second variant, *ATP11A* c.*11G>A, was predicted to destroy the canonical +/- 1 or 2 donor splice site and was first identified in *ATP11A*-203, a poorly supported suspect expressed sequence tag (EST). Haplotype analysis established a 3.3 Mb critical region, which was further reduced to 769 Kb by two separate recombination events to the disease-associated haplotype that excluded *COL4A1* as a candidate gene.

For the most part, our current understanding of the human genome is relatively complete; conversely, the same cannot be said about the human transcriptome and the non-coding DNA elements that regulate alternative splicing (Macaulay *et al.*, 2017). An excellent example of this is the identification of *COL11A*, a gene that causes *DFNA37* (Booth, Askew, *et al.*, 2018). However, this discovery was not trivial (Talebizadeh, 2018). In a recent commentary entitled *Lessons learned from the DFNA37 gene discovery odyssey* (Talebizadeh, 2018), Dr. Zohreh Talebizadeh eloquently outlined the many challenges associated with hearing loss gene discoveries. Upon the discovery of *DFNA37* on chromosome 1p21, Dr. Talebizadeh prioritized

the Sanger sequencing of the complete *COL11A1* cDNA sequence, due to its association with Marshall syndrome(Griffith *et al.*, 1998). At the RNA level, Talebizadeh *et al* found a messy electropherogram pattern that spanned exons 4 and 5 in the *DFNA37* proband, which was absent in unaffected family members. Yet, the sequencing of genomic DNA samples found an out-of-phase electropherogram within the intron 4 and exon 5 boundary in both affected and unaffected samples. Taken together, the cDNA sequencing results was perceived as a PCR artifact, resulting in the false-negative exclusion of *COL11A1* as the *DFNA37* gene. Twenty years later, a collaborative, multicentre study re-examined the *DFNA37* family using NGS and discovered a novel pathogenic splicing variant, c.652-2A>C, which caused the skipping of exon 5(Macaulay *et al.*, 2017). Furthermore, an out-of-frame deletion, c.652-6_-17del, was revealed in both the proband and unaffected family members, which masked the c.652-2A>C allele and explained the out-of-phase electropherogram when this study first began. Due to the lack of selection pressure, introns can vary in both in length and sequence, relative to exons(Tom Strachan, 2010). In addition to the polymorphic nature of intronic sequences, the difficulty of studying DNA variation at the RNA level is further confounded by over 8 million ESTs across the genome(dbEST, 2019). When performing RT-PCR, several amplicons can be produced that represent multiple gene transcripts that are either be under- or over-represented, depending on the tissue or cell type in question(Nagaraj *et al.*, 2007). In order to properly interpret these confusing results, more advanced molecular tools are required, such as TA-cloning. Alternatively, *in vitro* assays can also be employed in determining the role of candidate splicing variants, as was performed in a recent investigation of pathogenic *DFNA5* missense variants that cause exon skipping(Booth, Azaiez, *et al.*, 2018). These examples highlight the importance of knowing and understanding our assumptions during experimental design, as well as how

complicated gene discovery efforts can be when deciphering the intricate functional impact of variants at the RNA level.

While many studies routinely use commercially available cell line to investigate splicing variants, we were fortunate in that we could extract RNA from patient-derived B-cell lymphocytes. RNA analyses revealed that exon 2 from *ATP11A-203* was in fact marking a nascent exon that was missing from three alternatively spliced transcripts, *ATP11A-201*, *ATP11A-212* and *ATP11A-202*. Accounting for the addition of this exon, the c.*11 variant in exon 2 of *ATP11A-203* is located at the same position in *ATP11A-201* but positioned at c.*113 in *ATP11A-202/212*. As demonstrated by RT-PCR, cloning, and sequencing, this variant activates a cryptic splice site that inserts 153 bp into the 3' UTR of *ATP11A-201*, *ATP11A-212* and *ATP11A-202*. P4-ATPases play a critical role in maintaining phospholipid asymmetry in eukaryotic membranes (Paulusma & Elferink, 2010). Specifically, *ATP11A* and its paralog *ATP11C* are the most abundant P4-ATPases in mammalian cells, ensuring that PS is enriched and devoid in cytosolic and extracellular leaflets of plasma membrane, respectively (Segawa *et al.*, 2016; Segawa *et al.*, 2014). Under normal physiology, PS is enriched in the extracellular leaflet of damaged or weakened cells, marking them for phagocytosis and the completion of apoptosis (Paulusma & Elferink, 2010). Given that *ATP11A* is highly expressed in sensory hair cells and spiral ganglion neurons (Shen *et al.*, 2015), this dissertation proposes that this pathogenic *ATP11A* variant causes haploinsufficiency, leading to aberrant phagocytosis of healthy cells that are required normal auditory function. Together, we are confident that we have identified that genetic cause of hearing loss in this large NL family; although, functional luciferase assays are required to confirm this hypothesis.

To say the least, the discovery of *ATP11A* was quite challenging. Despite the many obstacles throughout this project, many important lessons were learnt along the way that emphasizes the importance of understanding the fundamental assumptions that are made during NGS variant filtration and the strengths and limitations of WES and WGS capture platforms. Upon the discovery of *ATP11A* using a PCR-free WGS capture platform, Geoff Woodland and I retrospectively reviewed how well this gene was covered by Ion Torrent WES. According to Ion Torrent WES metrics, the entire *ATP11A* gene was sufficiently covered to meet our variant filtration criteria. However, the Ion Torrent WES sequencing platform does not include probes that captures our *ATP11A* variant at chr13: 113,534,962. Provided that the human transcriptome is comprised of over of 200,000 protein coding transcripts(Hu *et al.*, 2015) and the Ion Torrent Hi-Q WES platform was designed to sequence 97.5% of all exons within the Consensus CDS (CCDS) project(Damiati *et al.*, 2016), it is unlikely that poorly supported transcripts (i.e. *ATP11A-203*) that are found within other reference databases are captured. Even though the RefSeq Gene, Ensembl, and UCSC reference databases display a considerable amount of discordance with respect to the number of genes, transcripts, and exons, it is important to note that no database is perfect, which may result in inaccurate variant annotations(Zhao & Zhang, 2015). According to Wu *et al*(Wu *et al.*, 2013), it is behaving to employ a more reproducible and robust reference genome when performing research at the clinical level, while more complex reference genomes are more appropriate for exploratory research, such as gene discovery.

The final research chapter of the thesis was extremely enlightening with respect to the clinical features of syndromic hearing loss (i.e. USH) and how they can present when least expected. USH is the leading cause of combined hearing and vision loss, which is most

commonly caused by pathogenic variants in the USH2 subtype; namely, *USH2A*, *ADGRV1*, and *WHRN* (MIM: 607928), (Lentz J, December 10, 1999). Family R2100 was first recruited to our study with apparent dominant non-syndromic hearing loss. However WES revealed a deleterious nonsense variant, *ADGRV1* c.17062C>T (p.Arg5688Ter) in two affected sisters. Upon review of the USH literature, we were concerned about visual impairment, even though no signs or symptoms were known at ascertainment. Subsequent follow-up with these siblings identified “typical RP” findings, prompting an USH2C diagnosis. Concurrently, two likely pathogenic *USH2A* splicing variants, c.5777-1G>A and c.10388-2A>G, were identified by the clinic in a 3-year old girl who was *GJB2*-negative. Through Dr. Green’s hereditary eye study, these same *USH2A* variants were identified in Family R0723, an apparent non-syndromic RP kinship; however, it was noted that some of family members were fitted with hearing aids, which was followed by an *USH2A* diagnosis. RNA analyses confirmed that these *USH2A* variants cause cryptic splicing. Despite exon skipping predictions, *in silico*, only c.5777-1G>A was experimentally validated, while c.10388-2A>G creates a frameshift by activating a cryptic acceptor site 14 bp downstream. While these variants can be reclassified as pathogenic variants, it’s important to appreciate that *in silico* predictions lack the ability to predict the precise splicing effect, and these variants could have easily caused a more tolerable in-frame INDEL(Ernst *et al.*, 2018). Throughout this study, I was perplexed as to why USH went undetected in these families until the identification of these variants. This story highlights one of the major limitations in studying genetic isolates, the lack of accessible health care in remote communities(Green, Personal communication: Discussion on health care access in NL - October 4 2018). Without a complete family history, the affection status of members may be erroneously annotated on pedigrees and misguide molecular studies. Collectively, this research

chapter revealed several variants that can be translated into the clinic to diagnose, treat and manage patients at-risk of developing USH.

For those patients with USH2, pathogenic variants in *USH2A* account for an estimated 75% – 90% of cases(Aller *et al.*, 2006; Baux *et al.*, 2007; Dreyer *et al.*, 2008), while *ADGRV1* is more rare at 3% – 6%(Besnard *et al.*, 2012; Ebermann *et al.*, 2009; Weston *et al.*, 2004). Previously, the prevalence of USH was once thought to be 1 in 26,000(Boughman *et al.*, 1983). However, more recent estimate of 1 in 6,000 was found in the European population(Kimberling *et al.*, 2010); although, the frequency of USH and the underlying causative genetic defect is dependent on ethnicity(Khalaileh *et al.*, 2018). For example, *USH2A* variant, c.2299delG, is the most common allele in the European population, while it is present in isolated cases within the South American, African and Asian populations(Dreyer *et al.*, 2001). Similarly, pathogenic variants in *ADGRV1* appear to be more frequent in the Europe(Besnard *et al.*, 2012), given that 2018 marks the first reports of USH2C in the Israeli and Palestinian populations(Khalaileh *et al.*, 2018). While we observe much higher allele frequencies within the European population, it is important to note that the vast majority of participants that are enrolled in large-scale sequencing projects are of European descent(Popejoy & Fullerton, 2016). This may skew our understanding of the prevalence of certain diseases, such as USH, and allele frequencies amongst minority groups.

The genes of the USH2 subtype, *USH2A*, *ADGRV1* and *WHRN* encode for the Usherin, VLGR1, and Whirlin proteins, respectively. Given that these proteins interact with one another to form a multimeric ankle-link complex at the base of sensory hair cells and at the periciliary

membrane complex of photoreceptors(Liu *et al.*, 2007; Richardson & Petit, 2019; Yang *et al.*, 2010), it is no surprise that they cause an indistinguishable phenotype. The Usherin and VLGR1 proteins are very similar in that they possess very large extracellular domains and relatively short C-terminal cytoplasmic regions that carry PDZ domain-binding motifs(Adato *et al.*, 2005; Michalski *et al.*, 2007). In the cytoplasm of sensory hair cells and photoreceptors, Whirlin and PDZD7, a protein encoded by *PDZD7*, which is a modifier of USH2C (MIM: 602851) (Ebermann *et al.*, 2010) and also causes *DFNB57* (MIM: 618003)(Booth *et al.*, 2015), form a heterodimer scaffolding complex; in turn, this heterodimer interacts with the PDZ domains of Usherin and VLGR1, recruiting them to the ankle link complex(Chen *et al.*, 2014). Together, this complex is believed to participate in linking the base of sensory hair cells and at the periciliary membrane complex of photoreceptors to various extracellular matrix proteins, as well as cell adhesion proteins in order to provide structural support(Yang *et al.*, 2012)

In conclusion, these projects emphasize the usefulness of applying a comprehensive “multi-omic” approach while investigating hereditary hearing loss in large extended families from genetic isolates, such as NL. In addition to combining NGS technologies, such as WES and WGS, with the power of linkage analysis, I also employed tools to study genetic variants at the RNA-level, which solved seven families and discovered a novel dominant hearing loss gene. Together, these studies have significantly contributed to the hereditary hearing loss field and are of great importance to the clinical diagnosis, management and treatment of these families, including the enrolment into gene therapy and small molecule trials. By understanding the genetic basis of hearing loss in affected families, innovative treatments can be offered that improve quality of life and contribute to the era of precision medicine.

5.1 Future Directions

While the *CLDN14*, *ADGRV1* and *USH2A* variants that have been identified in this thesis sufficiently fulfill the pathogenic ACMG criteria, our pathogenic *ATP11A* variant requires functional analyses to determine the molecular mechanism of disease. Given that our *ATP11A* variant resides within the 3' UTR, we hypothesize that aberrant miRNA silencing of the mRNA is the most likely molecular disease mechanism. In order to answer this question, a dual luciferase assay would be required to determine if *ATP11A* mRNA destabilization was occurring at post-transcriptional level. In short, both the wild-type and mutant *ATP11A* 3'UTRs would be cloned into the pmirGLO Dual-Luciferase miRNA Target Expression Vector (Promega Cat.# E1330), and transfected into a mammalian auditory sensory hair cell line (HEI-OC1 cells). Preceding this, quantitative PCR for known auditory system miRNAs (Rudnicki *et al.*, 2014) would be performed to determine the suitability of using this HEI-OC1 cell line. Assuming that our *ATP11A* variant causes dysregulated gene expression, our next steps would be to knockout *ATP11A* in HEI-OC1 cells using CRISPR-Cas9, followed by immunofluorescence and immunoblot assays. We would perform immunofluorescence on mutant and wild-type *ATP11A* in HEI-OC1 cells with Anti-Phosphatidylserine antibodies to determine the spatial localization of this phagocytic signal. Subsequently, immunoblotting for cleaved caspase-3 would indicate whether the loss of *ATP11A* induces apoptosis. In being certain of these results, TUNEL assays comparing wild-type and *ATP11A* null HEI-OC1 cells would confirm these findings. Finally, investigating *ATP11A*, *in vivo*, would be paramount in establishing this gene's role in hearing loss. Luckily, a *atp11a*-null mouse has already been established, which is cryopreserved at The Jackson Laboratories in Bar Harbour, Maine, USA. Assessing the role of *ATP11A* in hearing loss would involve measuring auditory brainstem

responses periodically throughout life; specifically, at birth, one month, three months, six months. Since hearing loss due to *ATP11A* can present as an adult, it is important to assess auditory brainstem responses outside the neonatal period. Thereafter, mice would be sacrificed, where components of the auditory system (including the cochlea and spiral ganglion tissues) would be resected, sectioned, and subject to H&E staining to visualize any gross morphology defects.

5.2 Bibliography

- Abdelfatah, N., Merner, N., Houston, J., Benteau, T., Griffin, A., Doucette, L., . . . Young, T. L. (2013). A novel deletion in SMPX causes a rare form of X-linked progressive hearing loss in two families due to a founder effect. *Hum Mutat*, *34*(1), 66-69. doi:10.1002/humu.22205
- Adato, A., Lefevre, G., Delprat, B., Michel, V., Michalski, N., Chardenoux, S., . . . Petit, C. (2005). Usherin, the defective protein in Usher syndrome type IIA, is likely to be a component of interstereocilia ankle links in the inner ear sensory cells. *Hum Mol Genet*, *14*(24), 3921-3932. doi:10.1093/hmg/ddi416
- Ahmed, Z. M., Li, X. C., Powell, S. D., Riazuddin, S., Young, T. L., Ramzan, K., . . . Wilcox, E. R. (2004). Characterization of a new full length TMPRSS3 isoform and identification of mutant alleles responsible for nonsyndromic recessive deafness in Newfoundland and Pakistan. *BMC Med Genet*, *5*, 24. doi:10.1186/1471-2350-5-24
- Aller, E., Jaijo, T., Beneyto, M., Najera, C., Oltra, S., Ayuso, C., . . . Millan, J. M. (2006). Identification of 14 novel mutations in the long isoform of USH2A in Spanish patients with Usher syndrome type II. *J Med Genet*, *43*(11), e55. doi:10.1136/jmg.2006.041764
- Bashir, R., Fatima, A., & Naz, S. (2010). Mutations in CLDN14 are associated with different hearing thresholds. *J Hum Genet*, *55*(11), 767-770. doi:10.1038/jhg.2010.104
- Bashir, Z. E., Latief, N., Belyantseva, I. A., Iqbal, F., Riazuddin, S. A., Khan, S. N., . . . Riazuddin, S. (2013). Phenotypic variability of CLDN14 mutations causing DFNB29 hearing loss in the Pakistani population. *J Hum Genet*, *58*(2), 102-108. doi:10.1038/jhg.2012.143

- Baux, D., Larrieu, L., Blanchet, C., Hamel, C., Ben Salah, S., Vielle, A., . . . Roux, A. F. (2007). Molecular and in silico analyses of the full-length isoform of usherin identify new pathogenic alleles in Usher type II patients. *Hum Mutat*, 28(8), 781-789. doi:10.1002/humu.20513
- Besnard, T., Vache, C., Baux, D., Larrieu, L., Abadie, C., Blanchet, C., . . . Roux, A. F. (2012). Non-USH2A mutations in USH2 patients. *Hum Mutat*, 33(3), 504-510. doi:10.1002/humu.22004
- Booth, K. T., Askew, J. W., Talebizadeh, Z., Huygen, P. L. M., Eudy, J., Kenyon, J., . . . Smith, S. D. (2018). Splice-altering variant in COL11A1 as a cause of nonsyndromic hearing loss DFNA37. *Genet Med*. doi:10.1038/s41436-018-0285-0
- Booth, K. T., Azaiez, H., Kahrizi, K., Simpson, A. C., Tollefson, W. T., Sloan, C. M., . . . Smith, R. J. (2015). PDZD7 and hearing loss: More than just a modifier. *Am J Med Genet A*, 167A(12), 2957-2965. doi:10.1002/ajmg.a.37274
- Booth, K. T., Azaiez, H., Kahrizi, K., Wang, D., Zhang, Y., Frees, K., . . . Smith, R. J. (2018). Exonic mutations and exon skipping: Lessons learned from DFNA5. *Hum Mutat*, 39(3), 433-440. doi:10.1002/humu.23384
- Boughman, J. A., Vernon, M., & Shaver, K. A. (1983). Usher syndrome: definition and estimate of prevalence from two high-risk populations. *J Chronic Dis*, 36(8), 595-603.
- Chen, Q., Zou, J., Shen, Z., Zhang, W., & Yang, J. (2014). Whirlin and PDZ domain-containing 7 (PDZD7) proteins are both required to form the quaternary protein complex associated with Usher syndrome type 2. *J Biol Chem*, 289(52), 36070-36088. doi:10.1074/jbc.M114.610535

- Damiati, E., Borsani, G., & Giacomuzzi, E. (2016). Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies. *Hum Genet*, *135*(5), 499-511. doi:10.1007/s00439-016-1656-8
- dbEST. (2019). dbEST release 130101. Retrieved from https://www.ncbi.nlm.nih.gov/genbank/dbest/dbest_summary/
- Dreyer, B., Brox, V., Tranebjaerg, L., Rosenberg, T., Sadeghi, A. M., Moller, C., & Nilssen, O. (2008). Spectrum of USH2A mutations in Scandinavian patients with Usher syndrome type II. *Hum Mutat*, *29*(3), 451. doi:10.1002/humu.9524
- Dreyer, B., Tranebjaerg, L., Brox, V., Rosenberg, T., Moller, C., Beneyto, M., . . . Nilssen, O. (2001). A common ancestral origin of the frequent and widespread 2299delG USH2A mutation. *Am J Hum Genet*, *69*(1), 228-234.
- Ebermann, I., Phillips, J. B., Liebau, M. C., Koenekoop, R. K., Schermer, B., Lopez, I., . . . Bolz, H. J. (2010). PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. *J Clin Invest*, *120*(6), 1812-1823. doi:10.1172/JCI39715
- Ebermann, I., Wiesen, M. H., Zrenner, E., Lopez, I., Pigeon, R., Kohl, S., . . . Bolz, H. J. (2009). GPR98 mutations cause Usher syndrome type 2 in males. *J Med Genet*, *46*(4), 277-280. doi:10.1136/jmg.2008.059626
- Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R. K., & Hauke, J. (2018). Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med Genomics*, *11*(1), 35. doi:10.1186/s12920-018-0353-y
- Green, J. (Personal communication: Discussion on health care access in NL - October 4 2018, October 1, 2018).

- Griffith, A. J., Sprunger, L. K., Sirko-Osadsa, D. A., Tiller, G. E., Meisler, M. H., & Warman, M. L. (1998). Marshall syndrome associated with a splicing defect at the COL11A1 locus. *Am J Hum Genet*, *62*(4), 816-823. doi:10.1086/301789
- Hu, Z., Scott, H. S., Qin, G., Zheng, G., Chu, X., Xie, L., . . . Wei, C. (2015). Revealing Missing Human Protein Isoforms Based on Ab Initio Prediction, RNA-seq and Proteomics. *Sci Rep*, *5*, 10940. doi:10.1038/srep10940
- Khalaileh, A., Abu-Diab, A., Ben-Yosef, T., Raas-Rothschild, A., Lerer, I., Alswaiti, Y., . . . Khateb, S. (2018). The Genetics of Usher Syndrome in the Israeli and Palestinian Populations. *Invest Ophthalmol Vis Sci*, *59*(2), 1095-1104. doi:10.1167/iovs.17-22817
- Kimberling, W. J., Hildebrand, M. S., Shearer, A. E., Jensen, M. L., Halder, J. A., Trzupsek, K., . . . Smith, R. J. (2010). Frequency of Usher syndrome in two pediatric populations: Implications for genetic screening of deaf and hard of hearing children. *Genet Med*, *12*(8), 512-516. doi:10.1097/GIM.0b013e3181e5afb8
- Kopciuk, K. A., Choi, Y. H., Parkhomenko, E., Parfrey, P., McLaughlin, J., Green, J., & Briollais, L. (2009). Penetrance of HNPCC-related cancers in a retrolective cohort of 12 large Newfoundland families carrying a MSH2 founder mutation: an evaluation using modified segregation models. *Hered Cancer Clin Pract*, *7*(1), 16. doi:10.1186/1897-4287-7-16
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285-291. doi:10.1038/nature19057
- Lentz J, K. B. (December 10, 1999, 2016 Jul 21). GeneReviews: Usher Syndrome Type II. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK1341/>

- Lewis, M. A., Nolan, L. S., Cadge, B. A., Matthews, L. J., Schulte, B. A., Dubno, J. R., . . . Dawson, S. J. (2018). Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes. *BMC Med Genomics*, *11*(1), 77. doi:10.1186/s12920-018-0395-1
- Liu, X., Bulgakov, O. V., Darrow, K. N., Pawlyk, B., Adamian, M., Liberman, M. C., & Li, T. (2007). Usherin is required for maintenance of retinal photoreceptors and normal development of cochlear hair cells. *Proc Natl Acad Sci U S A*, *104*(11), 4413-4418. doi:10.1073/pnas.0610950104
- Macaulay, I. C., Ponting, C. P., & Voet, T. (2017). Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet*, *33*(2), 155-168. doi:10.1016/j.tig.2016.12.003
- Merner, N. D., Hodgkinson, K. A., Haywood, A. F., Connors, S., French, V. M., Drenckhahn, J. D., . . . Young, T. L. (2008). Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *Am J Hum Genet*, *82*(4), 809-821. doi:10.1016/j.ajhg.2008.01.010
- Michalski, N., Michel, V., Bahloul, A., Lefevre, G., Barral, J., Yagi, H., . . . Petit, C. (2007). Molecular characterization of the ankle-link complex in cochlear hair cells and its role in the hair bundle functioning. *J Neurosci*, *27*(24), 6478-6488. doi:10.1523/JNEUROSCI.0342-07.2007
- Nagaraj, S. H., Gasser, R. B., & Ranganathan, S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform*, *8*(1), 6-21. doi:10.1093/bib/bbl015

- Olufemi, S. E., Green, J. S., Manickam, P., Guru, S. C., Agarwal, S. K., Kester, M. B., . . . Chandrasekharappa, S. C. (1998). Common ancestral mutation in the MEN1 gene is likely responsible for the prolactinoma variant of MEN1 (MEN1Burin) in four kindreds from Newfoundland. *Hum Mutat*, *11*(4), 264-269. doi:10.1002/(SICI)1098-1004(1998)11:4<264::AID-HUMU2>3.0.CO;2-V
- Paulusma, C. C., & Elferink, R. P. (2010). P4 ATPases--the physiological relevance of lipid flipping transporters. *FEBS Lett*, *584*(13), 2708-2716. doi:10.1016/j.febslet.2010.04.071
- Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, *538*(7624), 161-164. doi:10.1038/538161a
- Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., . . . Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, *506*(7487), 185-190. doi:10.1038/nature12975
- Richard JH Smith, E. S., Michael S Hildebrand, and Guy Van Camp. (1999 January 19, 2014). GeneReviews: Deafness and Hereditary Hearing Loss Overview. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK1434/>
- Richardson, G. P., & Petit, C. (2019). Hair-Bundle Links: Genetics as the Gateway to Function. *Cold Spring Harb Perspect Med*. doi:10.1101/cshperspect.a033142
- Rudnicki, A., Isakov, O., Ushakov, K., Shivatzki, S., Weiss, I., Friedman, L. M., . . . Avraham, K. B. (2014). Next-generation sequencing of small RNAs from inner ear sensory epithelium identifies microRNAs and defines regulatory pathways. *BMC Genomics*, *15*, 484. doi:10.1186/1471-2164-15-484

- Segawa, K., Kurata, S., & Nagata, S. (2016). Human Type IV P-type ATPases That Work as Plasma Membrane Phospholipid Flippases and Their Regulation by Caspase and Calcium. *J Biol Chem*, *291*(2), 762-772. doi:10.1074/jbc.M115.690727
- Segawa, K., Kurata, S., Yanagihashi, Y., Brummelkamp, T. R., Matsuda, F., & Nagata, S. (2014). Caspase-mediated cleavage of phospholipid flippase for apoptotic phosphatidylserine exposure. *Science*, *344*(6188), 1164-1168. doi:10.1126/science.1252809
- Shearer, A. E., & Smith, R. J. (2015). Massively Parallel Sequencing for Genetic Diagnosis of Hearing Loss: The New Standard of Care. *Otolaryngol Head Neck Surg*, *153*(2), 175-182. doi:10.1177/0194599815591156
- Shen, J., Scheffer, D. I., Kwan, K. Y., & Corey, D. P. (2015). SHIELD: an integrative gene expression database for inner ear research. *Database (Oxford)*, *2015*, bav071. doi:10.1093/database/bav071
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, *29*(1), 308-311.
- Sloan-Heggen, C. M., Bierer, A. O., Shearer, A. E., Kolbe, D. L., Nishimura, C. J., Frees, K. L., . . . Smith, R. J. (2016). Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum Genet*, *135*(4), 441-450. doi:10.1007/s00439-016-1648-8
- Snoeckx, R. L., Huygen, P. L., Feldmann, D., Marlin, S., Denoyelle, F., Waligora, J., . . . Van Camp, G. (2005). GJB2 mutations and degree of hearing loss: a multicenter study. *Am J Hum Genet*, *77*(6), 945-957. doi:10.1086/497996

- Spirio, L., Green, J., Robertson, J., Robertson, M., Otterud, B., Sheldon, J., . . . Leppert, M. (1999). The identical 5' splice-site acceptor mutation in five attenuated APC families from Newfoundland demonstrates a founder effect. *Hum Genet*, *105*(5), 388-398.
- Talebizadeh, Z. (2018). Lessons learned from the DFNA37 gene discovery odyssey. *Genet Med*. doi:10.1038/s41436-018-0395-8
- Thorleifsson, G., Holm, H., Edvardsson, V., Walters, G. B., Styrkarsdottir, U., Gudbjartsson, D. F., . . . Stefansson, K. (2009). Sequence variants in the CLDN14 gene associate with kidney stones and bone mineral density. *Nat Genet*, *41*(8), 926-930. doi:10.1038/ng.404
- Toka, H. R., Genovese, G., Mount, D. B., Pollak, M. R., & Curhan, G. C. (2013). Frequency of rare allelic variation in candidate genes among individuals with low and high urinary calcium excretion. *PLoS One*, *8*(8), e71885. doi:10.1371/journal.pone.0071885
- Tom Strachan, A. R. (2010). *Human Molecular Genetics* (4th ed.). Italy: Garland Science.
- Weston, M. D., Luijendijk, M. W., Humphrey, K. D., Moller, C., & Kimberling, W. J. (2004). Mutations in the VLGR1 gene implicate G-protein signaling in the pathogenesis of Usher syndrome type II. *Am J Hum Genet*, *74*(2), 357-366. doi:10.1086/381685
- Wilcox, E. R., Burton, Q. L., Naz, S., Riazuddin, S., Smith, T. N., Ploplis, B., . . . Friedman, T. B. (2001). Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29. *Cell*, *104*(1), 165-172.
- Wu, P. Y., Phan, J. H., & Wang, M. D. (2013). Assessing the impact of human genome annotation choice on RNA-seq expression estimates. *BMC Bioinformatics*, *14 Suppl 11*, S8. doi:10.1186/1471-2105-14-S11-S8

- Yang, J., Liu, X., Zhao, Y., Adamian, M., Pawlyk, B., Sun, X., . . . Li, T. (2010). Ablation of whirlin long isoform disrupts the USH2 protein complex and causes vision and hearing loss. *PLoS Genet*, *6*(5), e1000955. doi:10.1371/journal.pgen.1000955
- Yang, J., Wang, L., Song, H., & Sokolov, M. (2012). Current understanding of usher syndrome type II. *Front Biosci (Landmark Ed)*, *17*, 1165-1183.
- Zhao, S., & Zhang, B. (2015). A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genomics*, *16*, 97. doi:10.1186/s12864-015-1308-8

Appendix

Appendix A – ACMG Standards and guidelines for the interpretation of sequence variants (Richards *et al.*, 2015). Provided here are the 28 criteria for the annotation of sequence variants. All variants identified in this thesis are bolded and in brackets after the criteria statement.

Evidence for variant pathogenicity:

Very Strong

- 1) PVS1:** null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease (*ADGRV1*, *USH2A*)

Strong

- 2) PS1:** Same amino acid change as a previously established pathogenic variant regardless of nucleotide change
- 3) PS2:** De novo (both maternity and paternity confirmed) in a patient with the disease and no family history
- 4) PS3:** Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product (*USH2A*)
- 5) PS4:** The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls (*CLDN14*; *USH2A*; *ADGRV1*; *ATP11A*)

Moderate

- 6) PM1:** Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation (*CLDN14*)
- 7) PM2:** Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium (*CLDN14*; *ADGRV1*; *USH2A*; *ATP11A*)
- 8) PM3:** For recessive disorders, detected in *trans* with a pathogenic variant (*CLDN14* ; *ADGRV1*)
- 9) PM4:** Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants
- 10) PM5:** Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before
- 11) PM6:** Assumed de novo, but without confirmation of paternity and maternity

Supporting

- 12) PP1:** Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease (*CLDN14* ; *ADGRV1*; *USH2A*; *ATP11A*)
- 13) PP2:** Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease (*CLDN14*)
- 14) PP3:** Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) (*CLDN14* ; *ADGRV1*; *USH2A*; *ATP11A*)

- 15) PP4:** Patient's phenotype or family history is highly specific for a disease with a single genetic etiology (*CLDN14*; *ADGRV1*; *USH2A*; *ATP11A*)
- 16) PP5:** Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

Criteria for the annotation of benign variants:

Stand Alone

- 17) BA1:** Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium

Strong

- 18) BS1:** Allele frequency is greater than expected for disorder
- 19) BS2:** Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age
- 20) BS3:** Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing
- BS4:** Lack of segregation in affected members of a family
- 21) BS4:** Lack of segregation in affected members of a family

Supporting

- 22) BP1:** Missense variant in a gene for which primarily truncating variants are known to cause disease
- 23) BP2:** Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in *cis* with a pathogenic variant in any inheritance pattern
- 24) BP3:** In-frame deletions/insertions in a repetitive region without a known function
- 25) BP4:** Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)
- 26) BP5:** Variant found in a case with an alternate molecular basis for disease
- 27) BP6:** Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
- 28) BP7:** A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Evidence of pathogenicity	Category
Very strong	<p>PVS1 null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease</p> <p>Caveats:</p> <ul style="list-style-type: none"> • Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>) • Use caution interpreting LOF variants at the extreme 3' end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	<p>PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level</p> <p>PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity.</p> <p>PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.</p> <p>PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case-control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.</p>
Moderate	<p>PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation</p> <p>PM2 Absent from controls (or at extremely low frequency if recessive) (Table 6) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing.</p> <p>PM3 For recessive disorders, detected in <i>trans</i> with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase.</p> <p>PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants</p> <p>PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.</p> <p>PM6 Assumed de novo, but without confirmation of paternity and maternity</p>
Supporting	<p>PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data</p> <p>PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease</p> <p>PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.</p> <p>PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology</p> <p>PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation</p>

LOF, loss of function; OR, odds ratio.

Evidence Framework

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in <i>trans</i> with a dominant variant BP2 Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Appendix B – Loci and Genes of Hereditary Hearing Loss and its Syndromes

Non-syndromic Sensorineural Hearing Loss

<u>Autosomal Dominant Sensorineural Hearing Loss Loci and Genes</u>		
<u>Locus (OMIM)</u>	<u>Location</u>	<u>Gene (OMIM)</u>
<i>DFNA1</i>	5q31	<i>DIAPH1</i>
<i>DFNA2A</i>	1p34	<i>KCNQ4</i>
<i>DFNA2B</i>	1p35.1	<i>GJB3</i>
<i>DFNA2C</i>		<i>IFNLR1</i>
<i>DFNA3A</i>	13q11-q12	<i>GJB2</i>
<i>DFNA3B</i>	13q12	<i>GJB6</i>
<i>DFNA4A</i>	19q13	<i>MYH14</i>
<i>DFNA4B</i>	19q13.32	<i>CEACAM16</i>
<i>DFNA5</i>	7p15	<i>GSDME</i>
<i>DFNA6</i>	4p16.3	<i>WFS1</i>
<i>DFNA7</i>	1q21-q23	<i>LMX1A</i>
<i>DFNA8</i>	see DFNA12	
<i>DFNA9</i>	14q12-q13	<i>COCH</i>
<i>DFNA10</i>	6q22-q23	<i>EYA4</i>
<i>DFNA11</i>	11q12.3-q21	<i>MYO7A</i>
<i>DFNA12</i>	11q22-24	<i>TECTA</i>
<i>DFNA13</i>	6p21	<i>COL11A2</i>
<i>DFNA14</i>	see DFNA6	
<i>DFNA15</i>	5q31	<i>POU4F3</i>
<i>DFNA16</i>	2q24	<i>unknown</i>
<i>DFNA17</i>	22q	<i>MYH9</i>
<i>DFNA18</i>	3q22	<i>unknown</i>
<i>DFNA19</i>	10 (pericentric)	<i>unknown</i>
<i>DFNA20</i>	17q25	<i>ACTG1</i>
<i>DFNA21</i>	6p21	<i>unknown</i>
<i>DFNA22</i>	6q13	<i>MYO6</i>
<i>DFNA23</i>	14q21-q22	<i>SIX1</i>
<i>DFNA24</i>	4q	<i>unknown</i>
<i>DFNA25</i>	12q21-24	<i>SLC17A8</i>
<i>DFNA26</i>	see DFNA20	
<i>DFNA27</i>	4q12	<i>REST</i>
<i>DFNA28</i>	8q22	<i>GRHL2</i>
<i>DFNA30</i>	15q25-26	<i>unknown</i>

<i>DFNA31</i>	6p21.3	<i>unknown</i>
<i>DFNA32</i>	11p15	<i>unknown</i>
<i>DFNA33</i>	13q34-qter	<i>unknown</i>
<i>DFNA34</i>	1q44	<i>NLRP3</i>
<i>DFNA36</i>	9q13-q21	<i>TMC1</i>
<i>DFNA37</i>	1p21	<i>COL11A1</i>
<i>DFNA38</i>	see DFNA6	
<i>DFNA39</i>	4q21.3	<i>DSPP</i>
<i>DFNA40</i>	16p12.2	<i>CRYM</i>
<i>DFNA41</i>	12q24-qter	<i>P2RX2</i>
<i>DFNA42</i>	5q31.1-q32	<i>unknown</i>
<i>DFNA43</i>	2p12	<i>unknown</i>
<i>DFNA44</i>	3q28-29	<i>CCDC50</i>
<i>DFNA47</i>	9p21-22	<i>unknown</i>
<i>DFNA48</i>	12q13-q14	<i>MYO1A</i>
<i>DFNA49</i>	1q21-q23	<i>unknown</i>
<i>DFNA50</i>	7q32.2	<i>MIRN96</i>
<i>DFNA51</i>	9q21	<i>TJP2</i>
<i>DFNA52</i>	4q28	<i>unknown</i>
<i>DFNA53</i>	14q11.2-q12	<i>unknown</i>
<i>DFNA54</i>	5q31	<i>unknown</i>
<i>DFNA56</i>	9q31.3-q34.3	<i>TNC</i>
<i>DFNA57</i>	19p13.2	<i>unknown</i>
<i>DFNA58</i>	2p12-p21	<i>unknown</i>
<i>DFNA59</i>	11p14.2-q12.3	<i>unknown</i>
<i>DFNA60</i>	2q21.3-q24.1	<i>unknown</i>
<i>DFNA64</i>	12q24.31-q24.32	<i>SMAC/DIABLO</i>
<i>DFNA65</i>	16p13.3	<i>TBC1D24</i>
<i>DFNA66</i>	6q15-21	<i>CDI64</i>
<i>DFNA67</i>	20q13.33	<i>OSBPL2</i>
<i>DFNA68</i>	15q25.2	<i>HOMER2</i>
<i>DFNA69</i>	12q21.32-q23.1	<i>KITLG</i>
<i>DFNA70</i>	3q21.3	<i>MCM2</i>
<i>DFNA73</i>	12q21.31	<i>PTPRQ</i>

Autosomal Recessive Sensorineural Hearing Loss Loci and Genes

Locus (OMIM)	Location	Gene (OMIM)
<i>DFNB1A</i>	13q12	<i>GJB2</i>
<i>DFNB1B</i>	13q12	<i>GJB6</i>
<i>DFNB2</i>	11q13.5	<i>MYO7A</i>
<i>DFNB3</i>	17p11.2	<i>MYO15A</i>
<i>DFNB4</i>	7q31	<i>SLC26A4</i>
<i>DFNB5</i>	14q12	<i>unknown</i>
<i>DFNB6</i>	3p14-p21	<i>TMIE</i>
<i>DFNB7/11</i>	9q13-q21	<i>TMC1</i>
<i>DFNB8/10</i>	21q22	<i>TMPRSS3</i>
<i>DFNB9</i>	2p22-p23	<i>OTOF</i>
<i>DFNB10</i>	see <i>DFNB8</i>	
<i>DFNB11</i>	see <i>DFNB7</i>	
<i>DFNB12</i>	10q21-q22	<i>CDH23</i>
<i>DFNB13</i>	7q34-36	<i>unknown</i>
<i>DFNB14</i>	7q31	<i>unknown</i>
<i>DFNB15/72/95</i>	3q21-q25 19p13	<i>GIPC3</i>
<i>DFNB16</i>	15q21-q22	<i>STRC</i>
<i>DFNB17</i>	7q31	<i>unknown</i>
<i>DFNB18</i>	11p14-15.1	<i>USH1C</i>
<i>DFNB18B</i>	11p15.1	<i>OTOG</i>
<i>DFNB19</i>	18p11	<i>unknown</i>
<i>DFNB20</i>	11q25-qter	<i>unknown</i>
<i>DFNB21</i>	11q	<i>TECTA</i>
<i>DFNB22</i>	16p12.2	<i>OTOA</i>
<i>DFNB23</i>	10p11.2-q21	<i>PCDH15</i>
<i>DFNB24</i>	11q23	<i>RDX</i>
<i>DFNB25</i>	4p13	<i>GRXCRI</i>
<i>DFNB26</i>	4q31	<i>GAB1</i>
<i>DFNB27</i>	2q23-q31	<i>unknown</i>
<i>DFNB28</i>	22q13	<i>TRIOBP</i>
<i>DFNB29</i>	21q22	<i>CLDN14</i>
<i>DFNB30</i>	10p11.1	<i>MYO3A</i>
<i>DFNB31</i>	9q32-q34	<i>WHRN</i>
<i>DFNB32/105</i>	1p13.3-22.1	<i>CDC14A</i>
<i>DFNB33</i>	9q34.3	<i>unknown</i>

<i>DFNB35</i>	14q24.1-24.3	<i>ESRRB</i>
<i>DFNB36</i>	1p36.3	<i>ESPN</i>
<i>DFNB37</i>	6q13	<i>MYO6</i>
<i>DFNB38</i>	6q26-q27	<i>unknown</i>
<i>DFNB39</i>	7q21.1	<i>HGF</i>
<i>DFNB40</i>	22q	<i>unknown</i>
<i>DFNB42</i>	3q13.31-q22.3	<i>ILDR1</i>
<i>DFNB44</i>	7p14.1-q11.22	<i>ADCY1</i>
<i>DFNB45</i>	1q43-q44	<i>unknown</i>
<i>DFNB46</i>	18p11.32-p11.31	<i>unknown</i>
<i>DFNB47</i>	2p25.1-p24.3	<i>unknown</i>
<i>DFNB48</i>	15q23-q25.1	<i>CIB2</i>
<i>DFNB49</i>	5q12.3-q14.1.	<i>MARVELD2/BDP1</i>
<i>DFNB51</i>	11p13-p12	<i>unknown</i>
<i>DFNB53</i>	6p21.3	<i>COL11A2</i>
<i>DFNB55</i>	4q12-q13.2	<i>unknown</i>
<i>DFNB59</i>	2q31.1-q31.3	<i>PJKV</i>
<i>DFNB60</i>	5q23.2-q31.1	<i>SLC22A4</i>
<i>DFNB61</i>	7q22.1	<i>SLC26A5</i>
<i>DFNB62</i>	12p13.2-p11.23	<i>unknown</i>
<i>DFNB63</i>	11q13.2-q13.4	<i>LRTOMT / COMT2</i>
<i>DFNB65</i>	20q13.2-q13.32	<i>unknown</i>
<i>DFNB66</i>	6p21.2-22.3	<i>DCDC2</i>
<i>DFNB66/67</i>	6p21.31	<i>LHFPL5</i>
<i>DFNB68</i>	19p13.2	<i>SIPR2</i>
<i>DFNB71</i>	8p22-21.3	<i>unknown</i>
<i>DFNB72</i>	see DFNB15	
<i>DFNB73</i>	1p32.3	<i>BSND</i>
<i>DFNB74</i>	12q14.2-q15	<i>MSRB3</i>
<i>DFNB76</i>	19q13.12	<i>SYNE4</i>
<i>DFNB77</i>	18q12-q21	<i>LOXHD1</i>
<i>DFNB79</i>	9q34.3	<i>TPRN</i>
<i>DFNB80</i>	2p16.1-p21	<i>unknown</i>
<i>DFNB81</i>	19p	<i>unknown</i>
<i>DFNB82</i>	1p13.1	<i>(see note 4)</i>
<i>DFNB83</i>	see DFNA47	
<i>DFNB84</i>	12q21.2	<i>PTPRQ / OTOGL</i>
<i>DFNB85</i>	17p12-q11.2	<i>unknown</i>

<i>DFNB86</i>	16p13.3	<i>TBC1D24</i>
<i>DFNB88</i>	2p12-p11.2	<i>ELMOD3</i>
<i>DFNB89</i>	16q21-q23.2	<i>KARS</i>
<i>DFNB90</i>	7p22.1-p15.3	<i>unknown</i>
<i>DFNB91</i>	6p25	<i>SERPINB6</i>
<i>DFNB93</i>	11q12.3-11q13.2	<i>CABP2</i>
<i>DFNB94</i>		<i>NARS2</i>
<i>DFNB95</i>	see <i>DFNB15</i>	
<i>DFNB96</i>	1p36.31-p36.13	<i>unknown</i>
<i>DFNB97</i>	7q31.2-q31.31	<i>MET</i>
<i>DFNB98</i>	21q22.3-qter	<i>TSPEAR</i>
<i>DFNB99</i>	17q12	<i>TMEM132E</i>
<i>DFNB100</i>	5q13.2-q23.2	<i>PPIP5K2</i>
<i>DFNB101</i>	5q32	<i>GRXCR2</i>
<i>DFNB102</i>	12p12.3	<i>EPS8</i>
<i>DFNB103</i>	6p21.1	<i>CLIC5</i>
<i>DFNB104</i>	6p22.3	<i>FAM65B</i>
<i>DFNB105</i>	see <i>DFNB32</i>	
<i>DFNB106</i>	11p15.5	<i>EPS8L2</i>
<i>DFNB108</i>	1p31.3	<i>ROR1</i>

X-linked Sensorineural Hearing Loss Loci and Genes

<u>Locus (OMIM)</u>	<u>Location</u>	<u>Gene (OMIM)</u>
<i>DFNX1</i>	Xq22	<i>PRPS1</i>
<i>DFNX2</i>	Xq21.1	<i>POU3F4</i>
<i>DFNX3</i>	Xp21.2	<i>unknown</i>
<i>DFNX4</i>	Xp22	<i>SMPX</i>
<i>DFNX5</i>	Xq26.1	<i>AIFM1</i>
<i>DFNX6</i>	Xq22.3	<i>COL4A6</i>

Syndromic Hearing Loss

Alport Syndrome

Gene (OMIM)	Genomic Location	Inheritance
<i>COL4A3</i>	2q36.3	Autosomal Recessive
<i>COL4A4</i>	2q36.3	Autosomal Recessive
<i>COL4A5</i>	Xq22.3	X-linked Recessive

CHARGE Syndrome

Gene (OMIM)	Genomic Location	Inheritance
<i>SEMA3E</i>	7q21.11	Autosomal Dominant
<i>CHD7</i>	8q12.2	Autosomal Dominant

Pendred Syndrome

Gene (OMIM)	Genomic Location	Inheritance
<i>SLC26A4</i>	7q22.3	Autosomal Recessive
<i>FOXI1</i>	5q35.1	Autosomal Recessive
<i>KCNJ10</i>	1q23.2	Autosomal Recessive

Branchio-Oto-Renal Syndrome

Locus	Gene (OMIM)	Genomic Location	Inheritance
<i>BOR1</i>	<i>EYAI</i>	8q13.3	Autosomal Dominant
<i>BOR2</i>	<i>SIX5</i>	19q13.32	Autosomal Dominant
	<i>Unknown</i>	1q31	Autosomal Dominant
<i>BOR3</i>	<i>SIX1</i>	14q23.1	Autosomal Dominant

Jervell & Lange-Nielsen Syndrome

Locus	Gene (OMIM)	Genomic Location	Inheritance
<i>JLNS1</i>	<i>KNCQ1</i>	11p15.5-15.4	Autosomal Recessive
<i>JLNS2</i>	<i>KCNE1</i>	21q22.12	Autosomal Recessive

Norrie Disease

Locus	Gene (OMIM)	Genomic Location	Inheritance
<i>NDP1</i>	<i>NDP</i>	Xp11.3	X-linked Recessive

Perrault Syndrome

Locus	Gene (OMIM)	Genomic Location	Inheritance
<i>PRLTS1</i>	<i>HSD17B4</i>	5q23.1	Autosomal Recessive
<i>PRLTS2</i>	<i>HARS2</i>	5q31.3	Autosomal Recessive
<i>PRLTS3/DFNB81</i>	<i>CLPP*</i>	19p13.3	Autosomal Recessive
<i>PRLTS4</i>	<i>LARS2</i>	3p21.31	Autosomal Recessive
<i>PRLTS5</i>	<i>TWNK</i>	10q24.21	Autosomal Recessive
<i>PRLTS6</i>	<i>ERAL1</i>	17q11.2	Autosomal Recessive

Stickler Syndrome

Locus	Gene (OMIM)	Genomic Location	Inheritance
<i>STL1</i>	<i>COL2A1</i>	12q13.11	Autosomal Dominant
<i>STL2</i>	<i>COL11A1</i>	1p21	Autosomal Dominant
<i>STL3</i>	<i>COL11A2</i>	6p21.32	Autosomal Recessive
<i>STL4</i>	<i>COL9A1</i>	6q13	Autosomal Recessive
<i>STL5</i>	<i>COL9A2</i>	1p34.2	Autosomal Recessive

Treacher Collins Syndrome

Locus	Gene (OMIM)	Genomic Location	Inheritance
<i>TCOF1</i>	<i>TCOF1</i>	5q32-q33.1	Autosomal Dominant
<i>TCOF2</i>	<i>POLR1D</i>	13q12.2	Autosomal Dominant
<i>TCOF3</i>	<i>POLR1C</i>	6p21.1	Autosomal Recessive

Usher Syndrome

Type	Locus	Gene (OMIM)	Genomic Location	Inheritance
Usher 1	<i>USH1A</i>	-	14q32	Autosomal Recessive
Usher 1	<i>USH1B</i>	<i>MYO7A</i>	11q13.5	Autosomal Recessive
Usher 1	<i>USH1C</i>	<i>USH1C</i>	11p15.1	Autosomal Recessive
Usher 1	<i>USH1D</i>	<i>CDH23</i>	10q22.1	Autosomal Recessive
Usher 1	<i>USH1E</i>	-	21q21	Autosomal Recessive
Usher 1	<i>USH1F</i>	<i>PCDH15</i>	10q21.1	Autosomal Recessive
Usher 1	<i>USH1G</i>	<i>SANS/USH1G</i>	17q25.1	Autosomal Recessive
Usher 1	<i>USH1H</i>	-	15q22-23	Autosomal Recessive
Usher 1	<i>USH1J</i>	-	15q25.1	Autosomal Recessive
Usher 1	<i>USH1K</i>	-	10p11.21-q21.1	Autosomal Recessive
Usher 2	<i>USH2A</i>	<i>USH2A</i>	1q41	Autosomal Recessive
Usher 2	<i>USH2B</i>	-	3p23-24.2	Autosomal Recessive
Usher 2	<i>USH2C</i>	<i>ADGRV1</i>	5q14.3	Autosomal Recessive

Usher 2	<i>USH2D</i>	<i>WHRN</i>	9q32	Autosomal Recessive
Usher 3	<i>USH3A</i>	<i>CLRN1</i>	3q25.1	Autosomal Recessive
Usher 3	<i>USH3B</i>	<i>HARS</i>	5q31.1	

Waardenburg Syndrome

Type	Locus	Gene (OMIM)	Genomic Location	Inheritance
Type I	<i>WS1</i>	<i>PAX3</i>	2q36.1	Autosomal Dominant
Type II	<i>WS2A</i>	<i>MITF</i>	3p13	Autosomal Dominant
Type II	<i>WS2B</i>	-	1p21-p13.3	Autosomal Dominant
Type II	<i>WS2C</i>	-	8p23	Autosomal Dominant
Type II	<i>WS2D</i>	<i>SNAI2</i>	8q11	Autosomal Recessive
Type II	<i>WS2E</i>	<i>SOX10</i>	22q13.1	Autosomal Dominant
Type III	<i>WS3</i>	<i>PAX3</i>	2q36.1	Autosomal Recessive
Type IV	<i>WS4A</i>	<i>EDNRB</i>	13q22.3	Autosomal Recessive
Type IV	<i>WS4B</i>	<i>EDN3</i>	20q13.32	Autosomal Recessive
Type IV	<i>WS4C</i>	<i>SOX10</i>	22q13.1	Autosomal Dominant

Appendix C – Recurrent Hearing Loss Variants in Newfoundland before this dissertation. Screening for these known pathogenic hearing loss variants was performed by Mrs. Jessica Squires.

Gene	Accession No.	Variant	DFN Locus	Locus (Cyto)
<i>TMPRSS3</i>	NM_024022	c.207delC c.268 G>C c.782+3delGAG c.757 A>G c.612-2insTA	DFNB8/10	21q22
<i>WFS1</i>	NM_006005.3	c.2146 G>A c.1832 A>G	DFNA6/14/38	4p16.3
<i>PCDH15</i>	NM_033056	c.1583 T>A c.1590 + 20 A>G	DFNB23	10p11.2-q21
<i>KCNQ4</i>	NM_004700	c.806delCCT	DFNA2A	1p34
<i>GJB2</i>	NM_004004	c.-23+1G>A c.35delG c.101 T>C c.229 T>C c.249 C>G c.167delT	DFNA3A/B1	13q11-q12
<i>GJB6</i>	NM_006783	delD1351830 delD1351854	DFNA3B	13q12
<i>GJB3</i>	NM_024009.2	c.109 G>A	DFNA2B	1p35.1
<i>SMPX</i>	NM_014332.2	c.99delC	DFNX4	Xp22
<i>COCH</i>	NM_004086.2	c.151 C>T	DFNA9	14q12-q13
<i>TECTA</i>	NM_005422	c.26557 A>G	DFNB21	11q
<i>TMCI</i>	NM_138691	c.421C>T	DFNA36	9q13-q21

Appendix D – Optimized TA-cloning procedure for the Young laboratory

General Notes:

- Make sure that you use PCR products coming from a reaction a general *taq* polymerase. High fidelity *taq* polymerases generate PCR amplicons that have blunt ends. Since this is referred to as “TA cloning”, the TOPO reaction relies on the presence of “A overhangs” on the amplicon that are complementary to “T overhangs” on the TOPO vector.
- Before Starting the TOPO cloning procedure
 - o Make sure to take out the SOC cell media out of the fridge and allow it to come to room temperature.
 - o Turn on the water bath and set it to 42°C
- Use filter tips
- Start preheating your agar plates at 37°C
- Turn on the shaker and allow it to warm up to 37°C
- Acceptable Nanodrop values after purification:
 - o 260/280: >1.7
 - o 260/230: >1.7
- The amount of template you input into the TOPO reaction greatly influences the ligation efficiency of your PCR amplicon into the vector. The reaction requires a 10:1 ratio of TOPO:Purified PCR Product. The TOPO vector is 4kb in size. Therefore, if you’re working with a 400bp amplicon, you should be use 1 ng of template (4000bp:400bp = 10:1). If you’re using an 800bp PCR product, you would use 0.5ng of template (4000bp:800bp = 5:1). If you over template the TOPO reaction, it will actually inhibit the topoisomerase that is attached to the vector. Inhibition of the TOPO reaction occurs in cases of extreme over templating. The TOPO vector tolerates slight over templating and I usually use 2ng of template as a starting point. After quantifying my purified PCR product, I dilute my sample to 2ng/ul and input 1ul of this into the TOPO Master Mix. If using 2ng and the TOPO reaction fails, repeat the procedure using 1ng of DNA.

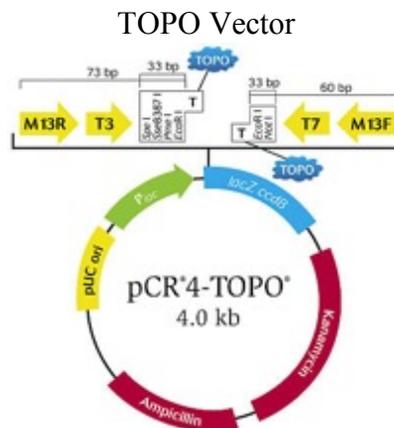
Procedure:

- 1) Combine the following reaction

TOPO Master Mix	
Reagent	Volume (uL)
Purified PCR product	1
Salt Solution	1
Water	3
TOPO Vector	1
Total	6

- 2) Incubate the reaction for 5-30 minutes at room temperature. The incubation period is at the discretion of the user. I find that I get the best results with a 15 minute incubation.

- 3) Add 2 uL of the TOPO reaction to One Shot TOP10 Chemically Competent Cells. Do not mix by pipetting.
 - Note: One Shot TOP10 Chemically Competent Cells are suspended in DMSO and are nutrient deprived, making them extremely unstable at this point. Make sure to minimize the time spent on ice before the transformation step. Avoid vigorous shaking or pipetting, until SOC media is added to the cells later on in the protocol. I usually allow them to sit on ice for 5 minutes. Therefore, after 10 minutes have gone by in the TOPO reaction, I bring the competent cells out of the -80 and put them on ice.
- 4) Incubate on ice for 20 minutes
- 5) Heat-shock cell for 30 second at 42°C without shaking.
 - Note: make sure the water bath is already at 42°C before beginning this procedure.
- 6) Immediately transfer the tubes to ice.
- 7) Add 250 uL of room temperature SOC media.
- 8) Shake the tube horizontally at 200 RPM at 37°C for 1 hour.
- 9) Spin the cells down by centrifuging at 3000 RPM for 3 minutes.
- 10) At this point there is approximately 300 uL in each tube. Aspirate 200 uL, leaving 100uL behind.
- 11) Resuspend cells by gently pipetting to mix. Make sure there are no cell clumps left in the cell suspension.
- 12) Spread 50 uL of cells to 2 separate onto pre-warmed selective agar plates (ampicillin or kanamycin) and incubate at 37°C overnight.
 - Note: Ideally, you should not allow the overnight incubation to exceed 16 hours.
- 13) Continue to Colony PCR the following morning



Linkage analysis in deafness pedigree 2070

Data providers: Terry Lynn Young (tlyoung@mun.ca), Nelly Abdelfatah (nellya@mun.ca), Lance Doucette (lanced@mun.ca), Tammy Benteau (tbenteau@mun.ca)

Report prepared by: Nicole Roslin (nroslin@sickkids.ca)

Last modified: 24 October 2013

Data summary

Chip: Human610-Quadv1_B.bpm

Number of markers passing QC: 573,708

Number of markers used in linkage analysis: 17,407

Family: 2070 (20 genotyped samples)

Introduction

Twenty individuals from family 2070, from Newfoundland and Labrador, were genotyped using the Human610-Quad chip from Illumina. The genotypes were of excellent quality, and were consistent with the pedigree structure and genders provided for this family; see the document "Analysis of SNP genotypes from the Illumina Human610-Quad panel in AMGGI Deafness pedigrees," dated 22 May 2013, for details.

Samples, pedigree and phenotype

Family 2070 is a multigenerational pedigree with many individuals with deafness (Figure 1). The hearing loss was early onset that progressed to profound deafness. Eleven individuals had deafness that was confirmed by audiogram analysis, and another two had unconfirmed deafness. The analysis was run twice, once including all 13 individuals known to have deafness, and once including just the 11 individuals with confirmed deafness.

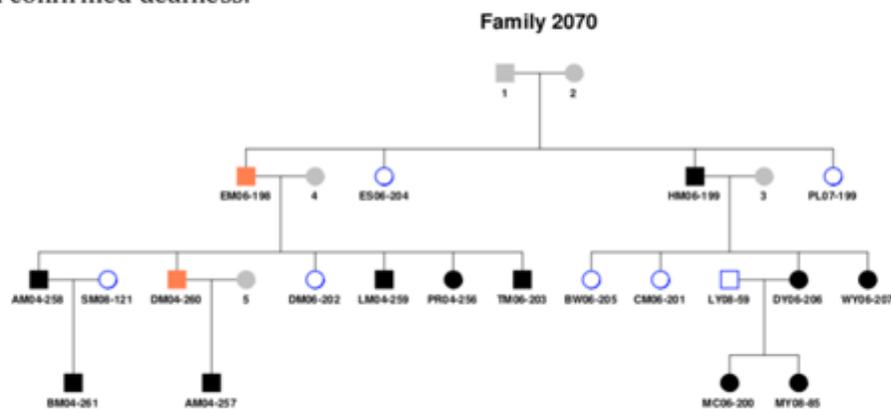


Figure 1. Family 2070. White symbols outlined in blue are unaffected for hearing loss, grey symbols have unknown phenotype, black symbols have deafness that has been confirmed by audiogram analysis, and orange symbols have unconfirmed deafness. Symbols with hyphenated labels have been genotyped.

The genetic model to be used in the linkage analysis requires the specification of a disease allele frequency. In linkage analysis, this frequency is used to estimate of the probability that a founder carries the disease allele. Generally, the frequency is derived from the prevalence of the disease in the population. However, statistics are not available regarding the prevalence of hearing loss in Newfoundland and Labrador. Statistics from the Canadian population indicates hearing loss is found in approximately 5% of the adult population. However, there are many different types of hearing loss, with different causes, so the prevalence of the particular type of hearing loss observed in this family will be much less frequent. Therefore, a range of disease allele frequencies was used, with an upper bound corresponding to 5% prevalence.

Reduced pedigree

The complete pedigree, including all genotyped members and individuals required to link the pedigree, was composed of 25 individuals. This pedigree is too large to be analyzed using Merlin, which performs exact multipoint analysis. Therefore, the pedigree was reduced to a size that could be handled by Merlin. Five individuals who would provide the least amount of linkage information were removed: the unaffected individuals PL07-199, ES06-204, DM06-202, BW06-205 and CM06-201. The resulting pedigree is shown in Figure 2, and will be referred to as the reduced pedigree in this report.

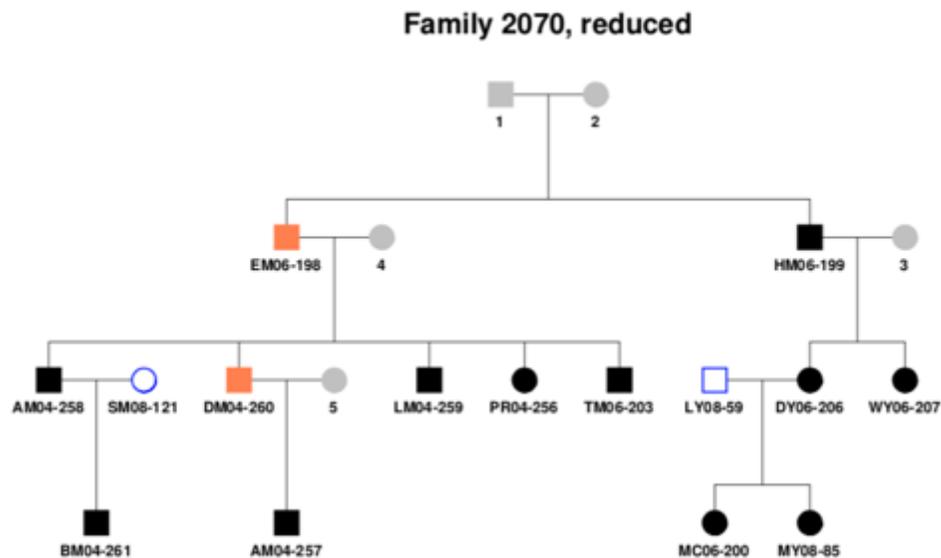


Figure 2. A reduced version of family 2070. Five unaffected individuals were removed so that exact multipoint linkage could be performed. The symbol legend is the same as in Figure 1.

Simulations

Simulations were performed under the alternate hypothesis of linkage to determine the maximum possible LOD score for the pedigree, under ideal conditions. These conditions included an autosomal disease-causing locus that was perfectly linked to the marker locus, a marker locus that provided perfect information regarding the

segregation of alleles in the pedigree, and an analysis model that was correctly specified. It was also assumed that the pedigree structure and affection status were correct. Genotypes for pedigrees identical to the one in Figure 2 were simulated under dominant and recessive models with a range of disease allele frequencies and penetrances using SLINK 3.02 (Schaffer, et al., 2011), and analyzed under the same model using Merlin 1.1.2 (Abecasis, et al., 2002). The phenocopy rate was fixed at 0.2% for all models. The maximum LOD score obtained from the analysis of 1000 simulated pedigrees was declared the maximum LOD score for any particular model.

For the dominant model, disease allele frequencies of 0.025, 0.005 and 0.0025 were used, corresponding to prevalences of 5%, 1% and 0.5%, respectively, under a fully penetrant model. For the recessive model, the disease allele frequencies used were 0.22, 0.1 and 0.07 (corresponding to the same prevalences). For both dominant and recessive models, simulations were performed using penetrances of 50, 60, 70, 80, 90, 95 and 99%.

For analysis 1, including all 13 individuals with deafness, the maximum simulated LOD score was 3.22, under a dominant model with a disease allele frequency of 0.0025 (Figure 3). Dominant models had consistently higher LOD scores than recessive models.

For analysis 2, including 11 individuals with deafness, the maximum simulated LOD score was also 3.22, under a dominant model with disease allele frequency of 0.0025. The trends were essentially identical to those in analysis 1 (results not shown).

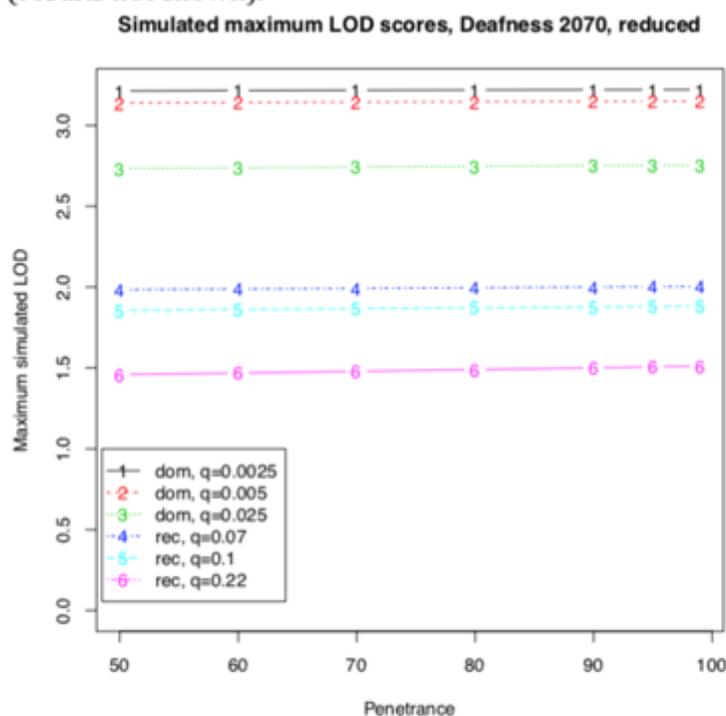


Figure 3. Simulated maximum LOD scores under the alternative hypothesis of linkage, under a range of genetic models, for analysis 1. Five unaffected individuals were removed from the pedigree to reduce the overall size, as in Figure 2. dom = dominant, rec = recessive, q = disease allele frequency.

Multipoint linkage analysis

SNP filtering

Starting with a set of 572,708 high quality SNPs from the Illumina 610Quad genotyping chip, markers were removed in order to end up with a smaller set suitable for linkage analysis. First, SNPs on chromosomes other than 1 to 22 and X were removed (mitochondrial, Y, XY and unplaced markers). Next, markers that had alleles ambiguous for strand information (A/T and G/C variants) were removed, in order to facilitate strand matching with HapMap data (Altshuler, et al., 2010). Markers that were monomorphic in all genotyped samples were also removed. At this point, approximately 550,000 markers remained. Markers that were present in both this set and the CEU and TSI samples (of western European origin) from HapMap were retained, matching on marker name. Out of this set of about 525,000 markers, SNPs with minor allele frequency >0.45 were kept, which should select for markers with the most linkage information. In order to avoid inconsistencies between the genotypes and the genetic map, only markers with unique locations on the genetic map were extracted. Arbitrarily, for a group of markers at the same genetic position, the marker with the lowest physical position was retained. Since linkage disequilibrium (LD) between the SNPs can arbitrarily inflate multipoint LOD scores (Huang, et al., 2004), the set of markers was pruned for LD between pairs of markers: only SNPs with pairwise $r^2 < 0.1$ were kept. These steps resulted in a set of 17,407 SNPs across the genome that was suitable for linkage analysis. For this set, the average intermarker distance was 0.26 cM, or 197 kb.

Multipoint linkage analysis

Linkage analysis was performed using Merlin 1.1.2 (Abecasis, et al., 2002) using the observed genotypes. Multipoint linkage analysis was performed under a dominant model with a disease allele frequency (q) of 0.0025 and penetrances of 0.2, 99 and 99% for 0, 1, 2 copies of the disease-causing allele, respectively. These parameters correspond to a disease prevalence of approximately 0.5%, and formed the model with the largest simulated maximum LOD score for this family (out of the models tested). Due to the large number of individuals, the pedigree in Figure 1 needed to be reduced in order to be analyzed. Five unaffected individuals (PL07-199, ES06-204, DM06-202, BW06-205 and CM06-201) were removed. Although these individuals contain information for linkage, they do not provide as much information as affected individuals. The reduced pedigree that was used is shown in Figure 2.

For analysis 1 (13 individuals affected with hearing loss, 5 unaffected individuals removed), the largest observed LOD score was 3.22, on chromosome 13 (Table 1, Figure 4, black line). Note that there are two peaks on chromosome 13 that are approximately 5 cM apart. There was also a peak on chromosome 4 with a LOD of 1.07. Using a threshold of $\text{LOD} > 1$, the linked regions spanned approximately 37 cM; this represents 1% of the genome, assuming a total genetic length of 36M. The average information content across the genome was 0.99.

For analysis 2 (11 individuals affected with hearing loss, 5 unaffected individuals removed), the largest observed LOD score was 3.22 on chromosome 13 (Table 2). The results were nearly identical between analysis 1 and analysis 2

(Figure 4, red line). The main difference was that there was a peak with LOD = 1.07 on chromosome 5 in analysis 2; this peak had a maximum LOD of 0.64 in analysis 1. However, in both analyses, the peaks on chromosome 13 dominated the results.

Based on the genetic model, autosomal dominant with nearly full penetrance, it should be expected that the results of analyses 1 and 2 are nearly identical. Given this model, and the fact that both individuals with unconfirmed deafness have at least one child with confirmed deafness, there is a high probability that both individuals are also affected.

Table 1. Regions with maximum observed LOD scores >1 for analysis 1 (5 unaffected individuals removed), under a dominant model with 99% penetrance and a disease allele frequency of 0.0025. A region was defined as the 1-LOD support interval around the maximum. An asterisk following the SNP name indicates that the marker was either the first or last analyzed marker on the chromosome, and so the actual boundary of the linked region is not known. cM = centiMorgans (provided by Illumina), bp = base pairs (GRCh37).

Chr	Max LOD	Start			End		
		SNP	cM	bp	SNP	cM	bp
4	1.07	rs4861512	183.73	183,391,730	rs11727845	189.64	185,287,351
13	3.21	rs2297319	92.48	99,361,959	rs7988924	94.56	100,588,210
13	3.22	rs2065391	100.18	104,373,191	rs9525300*	129.45	114,960,232

Table 2. Regions with maximum observed LOD scores >1 for analysis 2 (5 unaffected individuals removed), under a dominant model with 99% penetrance and a disease allele frequency of 0.0025. See the caption for Table 1 for definitions.

Chr	Max LOD	Start			End		
		SNP	cM	bp	SNP	cM	bp
4	1.06	rs4861512	183.73	183,39,1730	rs11727845	189.64	185,287,351
5	1.07	rs753279	140.10	140,023,818	rs4868254	188.51	172,732,331
13	3.21	rs2297319	92.48	99,361,959	rs7988924	94.56	100,588,210
13	3.22	rs2065391	100.18	104,373,191	rs9525300*	129.45	114,960,232

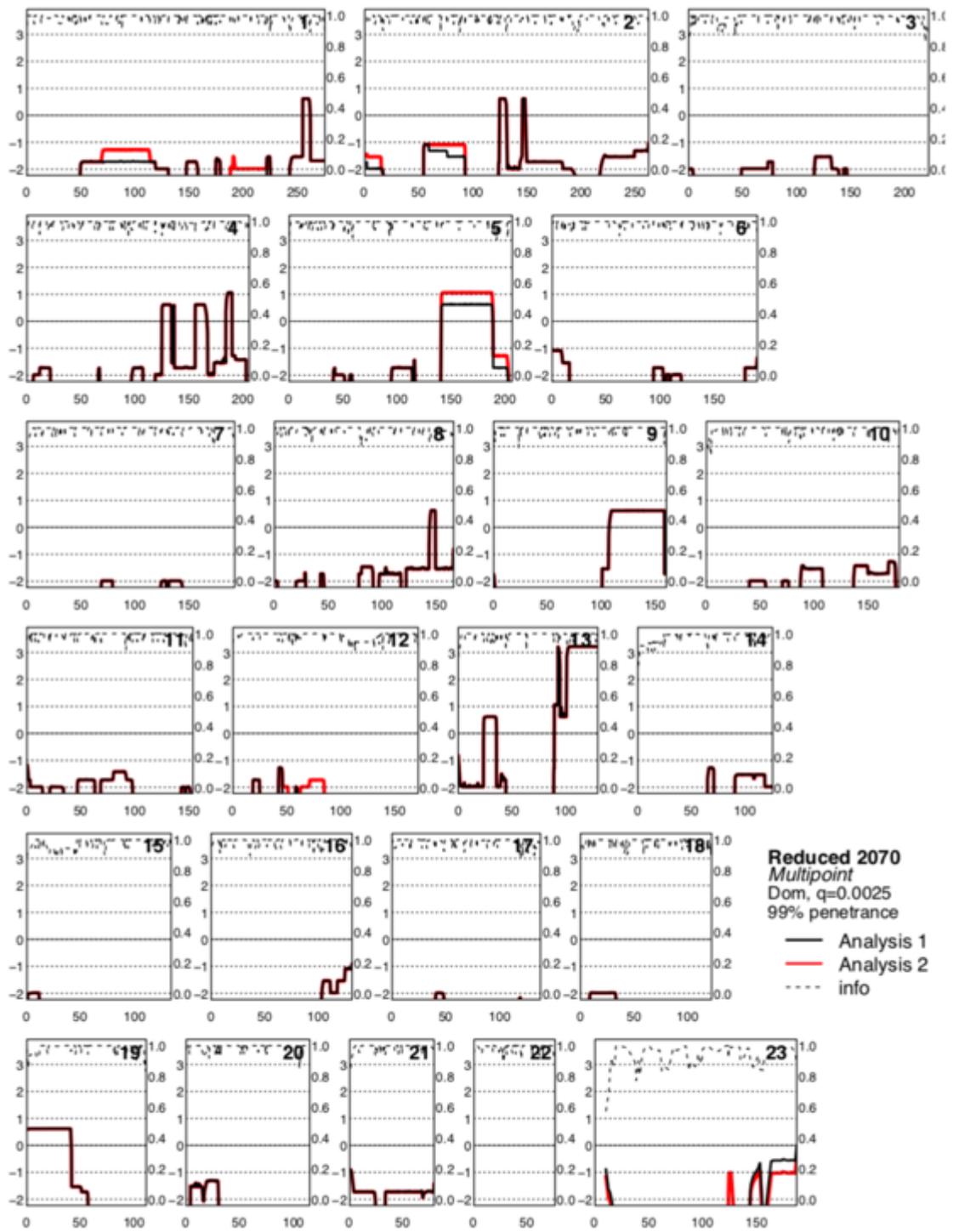


Figure 4. Multipoint LOD scores (solid lines) and information content (dashed line) for the family shown in Figure 2, under a dominant model with 99% penetrance and $q=0.0025$ for analysis 1 (black line) and analysis 2 (red line). LOD scores use the vertical axis on the left, and information content uses the vertical axis on the right. Only $\text{LOD} > -2$ are shown.

Full pedigree

It was not possible to perform exact multipoint linkage analysis on the pedigree in Figure 1, due to computational limitations. In order to extract information from the complete pedigree, two options are available: either perform approximate multipoint analysis, or perform exact singlepoint analysis. It was decided to perform singlepoint analysis, since the re-addition of the five samples back into the pedigree did not increase the complexity (ie, did not add inbreeding loops), it just increased the pedigree size. The resulting analysis will run quickly, but will suffer from a loss of information.

Simulations

Simulations were performed under the alternative hypothesis for the full pedigree shown in Figure 1. The same genetic models as in the reduced pedigree were used, and the same assumptions apply. Genotypes were simulated using SLINK, as before, however this time the simulations were analyzed using Superlink 1.7 (Fishelson M, 2004). For analysis 1 (13 affected individuals), the maximum simulated LOD score was 4.77, under a dominant model with 99% penetrance and $q=0.0025$. Similar results were seen for analysis 2 (11 affected individuals), where the maximum simulated LOD score was also 4.77, under the same model (results not shown).

Simulated maximum LOD scores, Deafness 2070

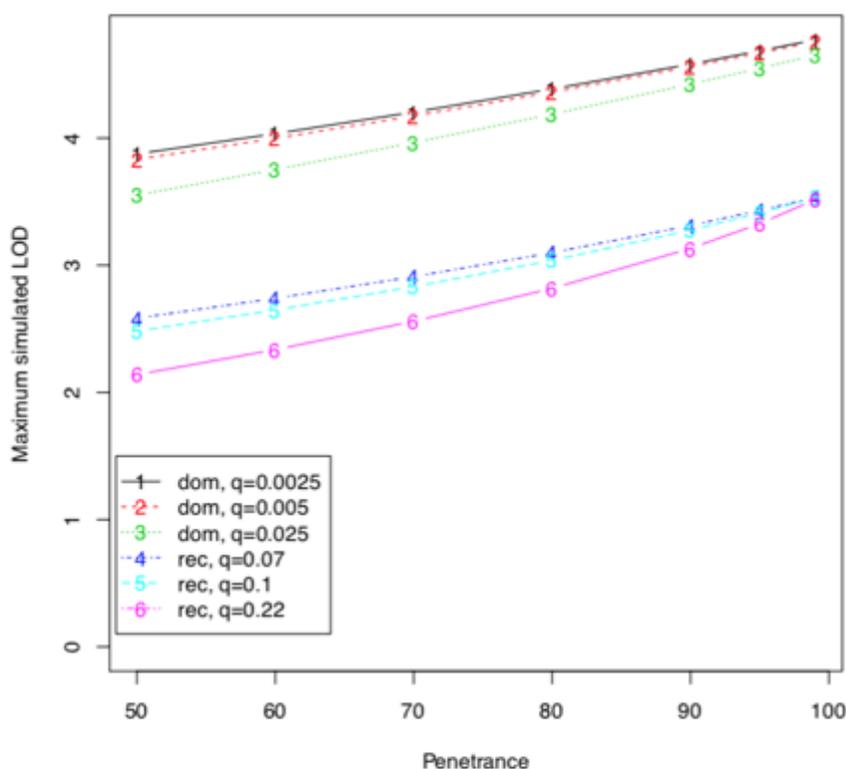


Figure 5. Simulated maximum LOD scores under the alternative hypothesis of linkage, under a range of genetic models, for analysis 1. The full pedigree shown in Figure 1 was used. dom = dominant, rec = recessive, q = disease allele frequency.

Singlepoint linkage analysis

In order to obtain as much information as possible, windows of five SNPs were examined. For each window, singlepoint linkage was performed at the location of the middle SNP, while the remaining four SNPs were used only for descent information. This approach allowed the genome to be analyzed in approximately 24h. The same genetic model as in the analysis of the reduced family was used: dominant, with $q=0.0026$ and 99% penetrance. The analysis was performed using Superlink 1.7 (Fishelson M, 2004); this program does not report the information content. Also, the same set of 17,407 SNPs were used in the analysis.

For analysis 1, the maximum LOD score was 4.77, on chromosome 13 (Table 3). This was equal to the maximum LOD score obtained by simulations, indicating that the approach used here was able to extract the full information out of the biallelic markers, at least in this location. Although the 1-LOD support interval may not be the most appropriate way to define a linked region, due to the nature of singlepoint analysis, the 1-LOD support interval extends from rs872484 (117.51 cM) to rs9324254 (128.64 cM) - near the end of the chromosome. This interval is narrower than the one from the multipoint analysis on the reduced pedigree. The next highest LOD score was 2.54, on chromosome 1. Two other chromosomes had LOD >1.5 (chromosomes 5 and 19). Note that the LOD score on chromosome 13 exceeds the threshold for significant linkage proposed by Lander and Kruglyak (1995). Chromosome 4, which had a maximum LOD score of 1.07 in the multipoint analysis, did not achieve LOD >1.5 in the singlepoint analysis. Also, a region on chromosome 1 had LOD >2 in the singlepoint analysis, but had negative LOD scores in the multipoint analysis.

For analysis 2, the results were largely similar. The maximum LOD score was 4.76, also on chromosome 13 (results not shown).

Table 3. Regions with maximum observed LOD >1.5 in the singlepoint analysis of the full pedigree, analysis 1, under a dominant model with 99% penetrance and a disease allele frequency of 0.0025.

Chr	Max LOD	Start			End		
		SNP	cM	bp	SNP	cM	bp
1	2.54	rs591979	85.54	61,368,955	rs9629017	87.62	62,083,960
1	2.14	rs6593523	101.44	76,486,908	rs1360878	101.82	76,749,088
1	1.87	rs1325278	109.17	85,400,182	rs817485	109.42	85,573,095
5	1.56	rs253604	161.03	155,960,089	rs6892282	163.77	159,360,485
5	1.55	rs11954477	167.38	163,374,345	rs253537	169.34	164,600,485
13	4.77	rs872484	117.51	110,708,368	rs9324254	128.64	114,312,000
19	1.83	rs4527136	25.72	8,186,519	rs2042300	26.79	8,580,602
19	1.83	rs2060260	38.32	15,704,783	rs1558139	38.55	15,997,564

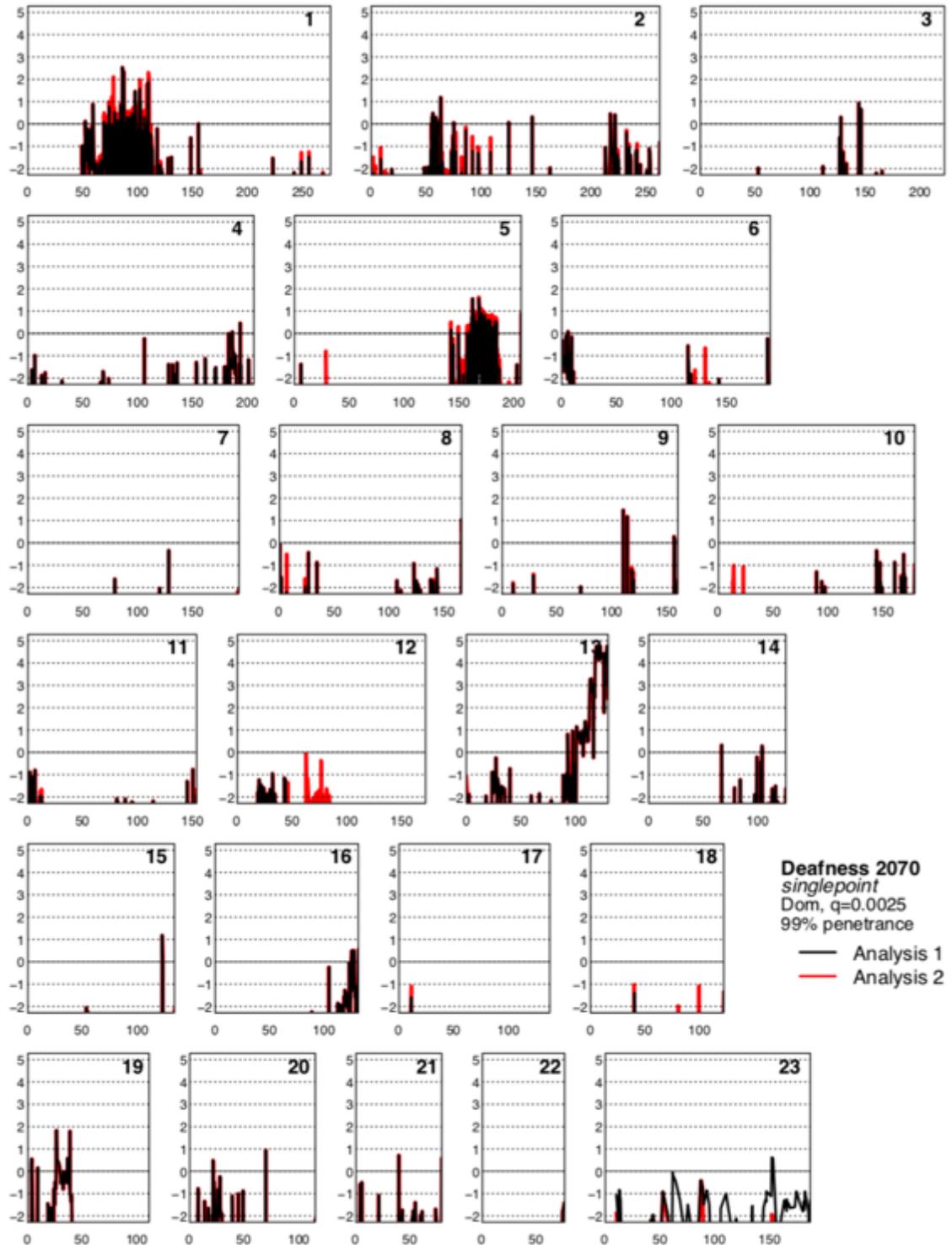


Figure 6. Singlepoint LOD scores for the full pedigree under a dominant model with 99% penetrance and $q=0.0025$. The black curve represents the score obtained for analysis 1 (13 affecteds), and the red curve represents the scores for analysis 2 (11 affecteds).

Summary

Multipoint linkage analysis was performed on family 2070, under a dominant model with a disease allele frequency of 0.0025 and a penetrance of 99%, corresponding to a prevalence of 0.5%. Due to the size of the pedigree, two approaches were used to perform the linkage analysis. Multipoint analysis was performed by excluding 5 unaffected individuals. In analysis 1, including all 13 individuals known to have deafness, three regions on two chromosomes attained a LOD score >1, spanning approximately 37 cM (13.7 Mb). Two adjacent regions on chromosome 13 had a LOD of 3.2, the maximum possible for this family and model. In analysis 2, including 11 individuals with deafness confirmed by audiogram analysis, the same three regions had LOD >1, plus one additional region on chromosome 5. Again, two adjacent regions on chromosome 13 had a LOD of 3.2. Singlepoint analysis was also performed on the complete pedigree, using 4 flanking markers for additional information. The largest LOD score was 4.77, observed on chromosome 13 (4.76 for analysis 2). Both singlepoint and multipoint approaches strongly indicate the presence of a linked region at the end of chromosome 13. The two individuals whose deafness was not confirmed by audiogram analysis (EM06-198 and DM04-260) did not appear to strongly influence the linkage results.

References

Abecasis, G.R., *et al.* (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees, *Nature Genetics*, **30**, 97-101.

Altshuler, D.M., *et al.* (2010) Integrating common and rare genetic variation in diverse human populations, *Nature*, **467**, 52-58.

Fishelson M, G.D. (2004) Optimizing exact genetic linkage computations, *Journal of Computational Biology*, **11**, 263-275.

Huang, Q., Shete, S. and Amos, C.I. (2004) Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis, *American Journal of Human Genetics*, **75**, 1106-1112.

Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results, *Nature Genetics*, **11**, 241-247.

Schaffer, A.A., *et al.* (2011) Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees, *Human Heredity*, **71**, 126-134.

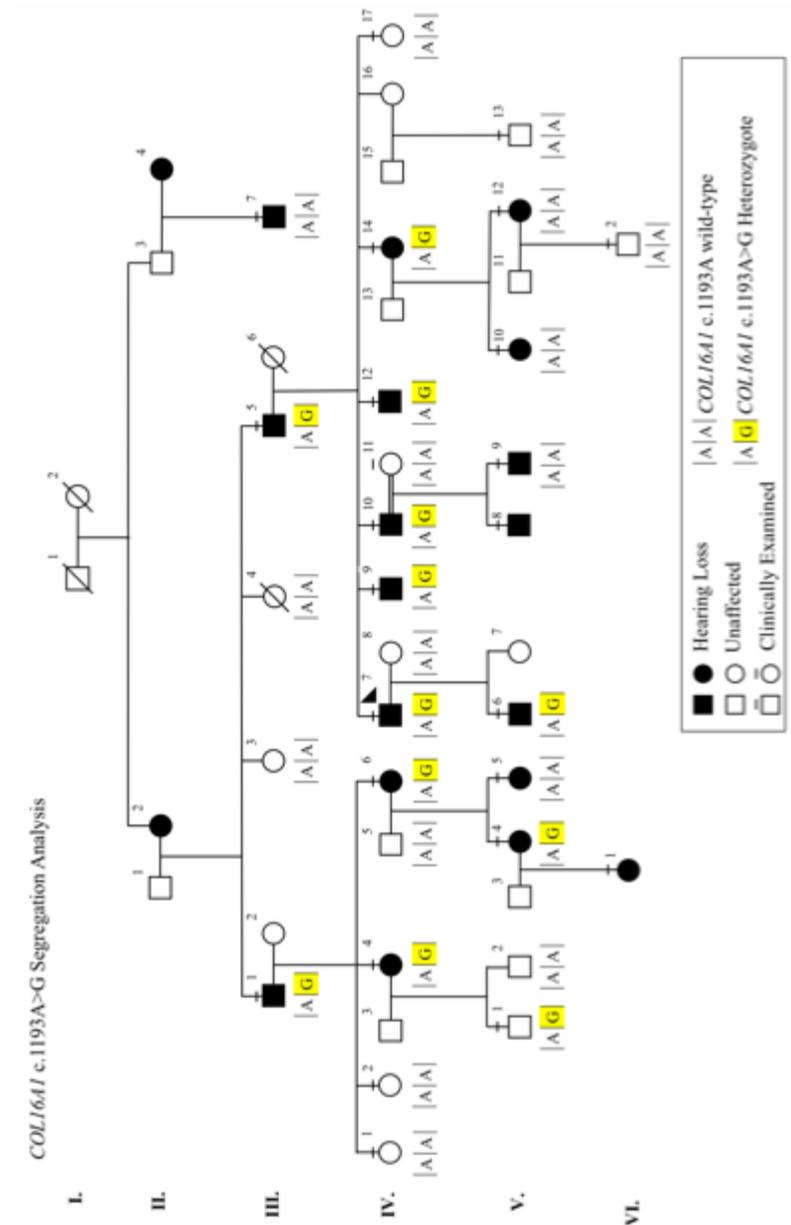
Appendix F – Whole exome sequencing identified 10 rare variants (<1% MAF) identified in family R2070

Damage	Chromosome	dbSNP	Gene	Mutation - Gene	Mutation - Protein	Sequenced?	ExAC (%)
14	1	rs115606583	<i>COL16A1</i>	c.1193A>G	p.K298R	Yes	0.51
5	1	rs146617729	<i>CSF3R</i>	c.2422G>A	p.E808K	Yes	0.62
8	2	rs76399854	<i>CAPN14</i>	c.829C>T	p.R277W	Yes	0.06
14	4	None	<i>TRAPPC11</i>	c.1436T>G	p.V479G	Yes	Novel
2	4	None	<i>WFS1</i>	c.1507G>A	p.V503I	Yes	0.007425
0	14	rs79957669	<i>GPHB5</i>	c.155G>C	p.G53A	Yes	Novel
999	19	None	<i>GTF2F1</i>	c.1111delAA	Splice	Yes	0.01
4	19	rs148952767	<i>RTBDN</i>	c.517G>A	p.V205M	Yes	0.59
6	20	rs45469293	<i>PDYN</i>	c.575A>T	p.E192V	Yes	0.86
1	20	rs201155774	<i>SIGLEC1</i>	c.3566C>T	p.A1189V	Yes	0.02

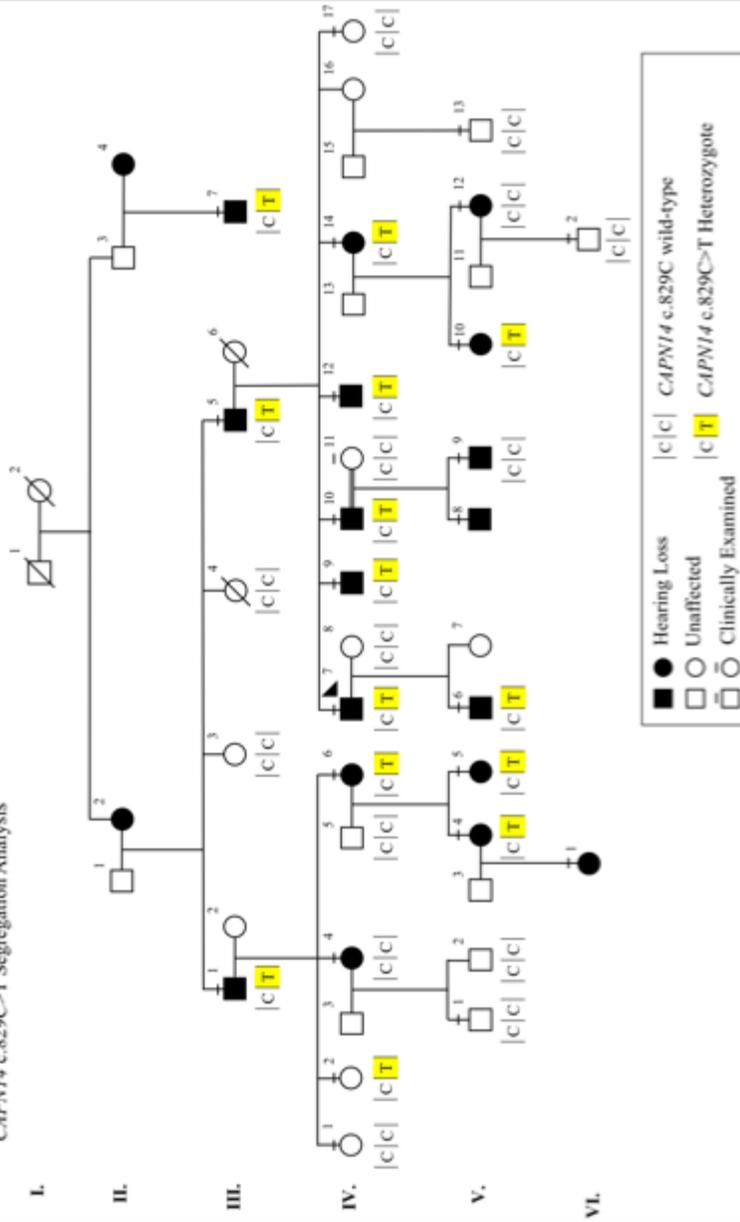
Appendix G – Whole exome sequencing identified 7 common variants (<10% MAF) identified in family R2070

Damage	Chromosome	dbSNP	Gene	Mutation - Gene	Mutation - Protein	Sequenced?	ExAC (%)
4	1	rs138630887	<i>TCTEX1D4</i>	c.140C>G	p.P47L	No	1.63
2	1	rs17107806	<i>PODN</i>	c.28G>C	p.A10P	No	4.88
10	1	rs41285372	<i>ALG6</i>	c.1357C>G	p.L455V	No	1.21
3	2	rs76758152	<i>ESPNL</i>	c.853C>T	p.R653W	No	2.26
0	2	rs34000641	<i>GPR75</i>	c.346G>A	p.A116T	No	2.49
4	4	rs112132883	<i>ZNF732</i>	c.181G>A	p.E61V	No	4.29
2	6	rs36019638	<i>TMC5</i>	c.325G>A	p.A355T	Yes	3.92
5	13	rs72661692	<i>CARS2</i>	c.538A>T	p.I180F	Yes	7.37
1	16	rs61741503	<i>CI6orf95</i>	c.636G>A	p.R151H	No	1.07
0	19	rs117420388	<i>SLC25A41</i>	76C>G	p.L26V	No	2.11
0	20	rs1135711	<i>TMX4</i>	c.644A>G	p.Y215C	No	3.95
15	22	rs8140287	<i>ISX</i>	c.248G>A	p.R83Q	No	4.3

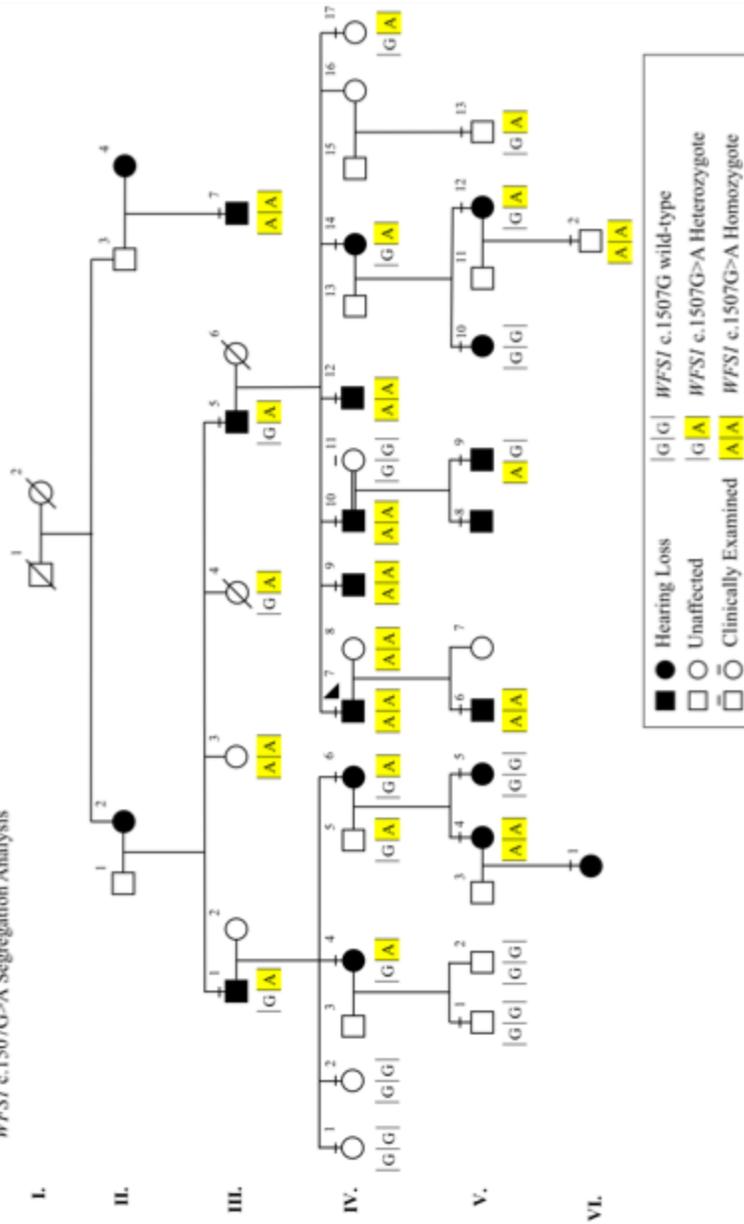
Appendix H – Whole exome sequencing variant segregation testing



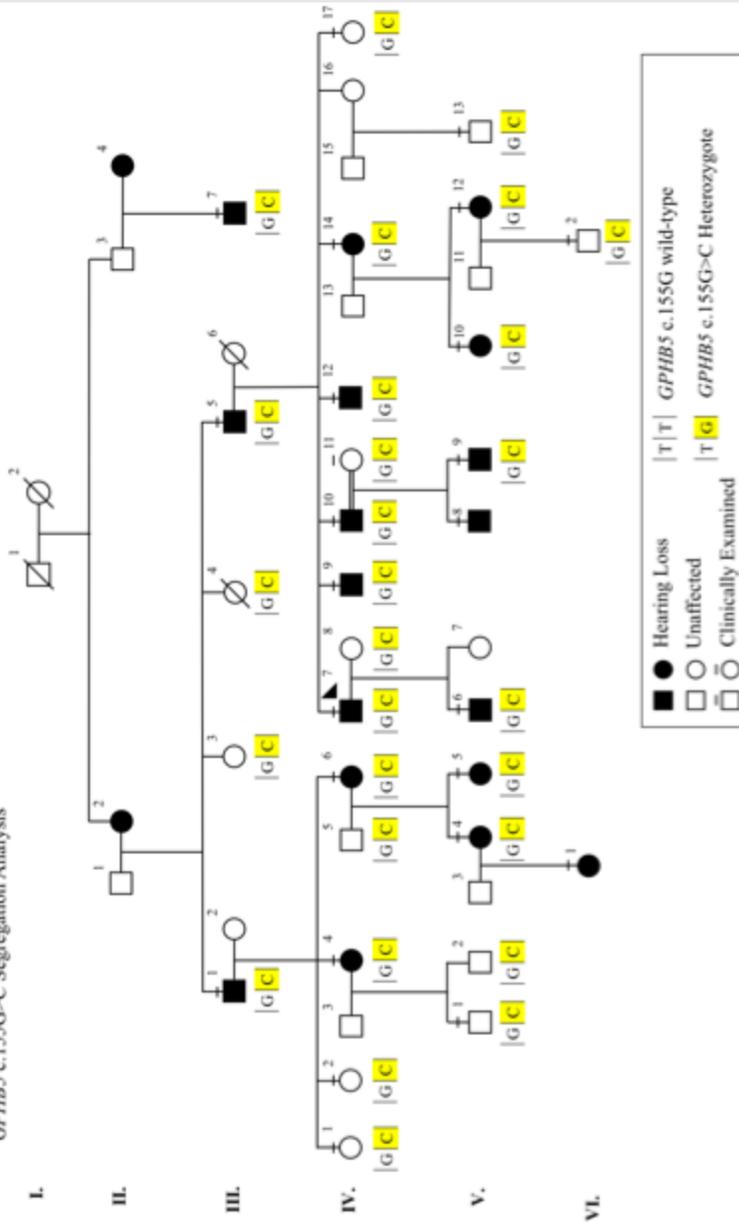
CAPN14 c.829C>T Segregation Analysis



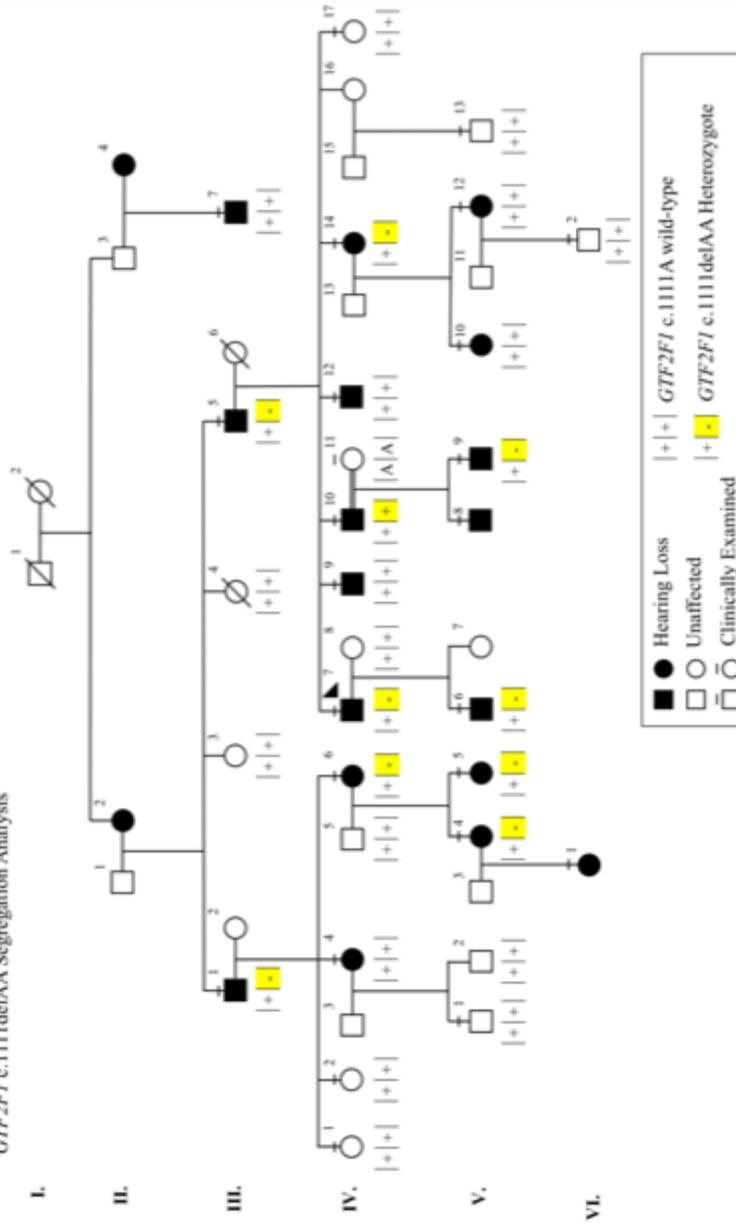
WFS1 c.1507G>A Segregation Analysis



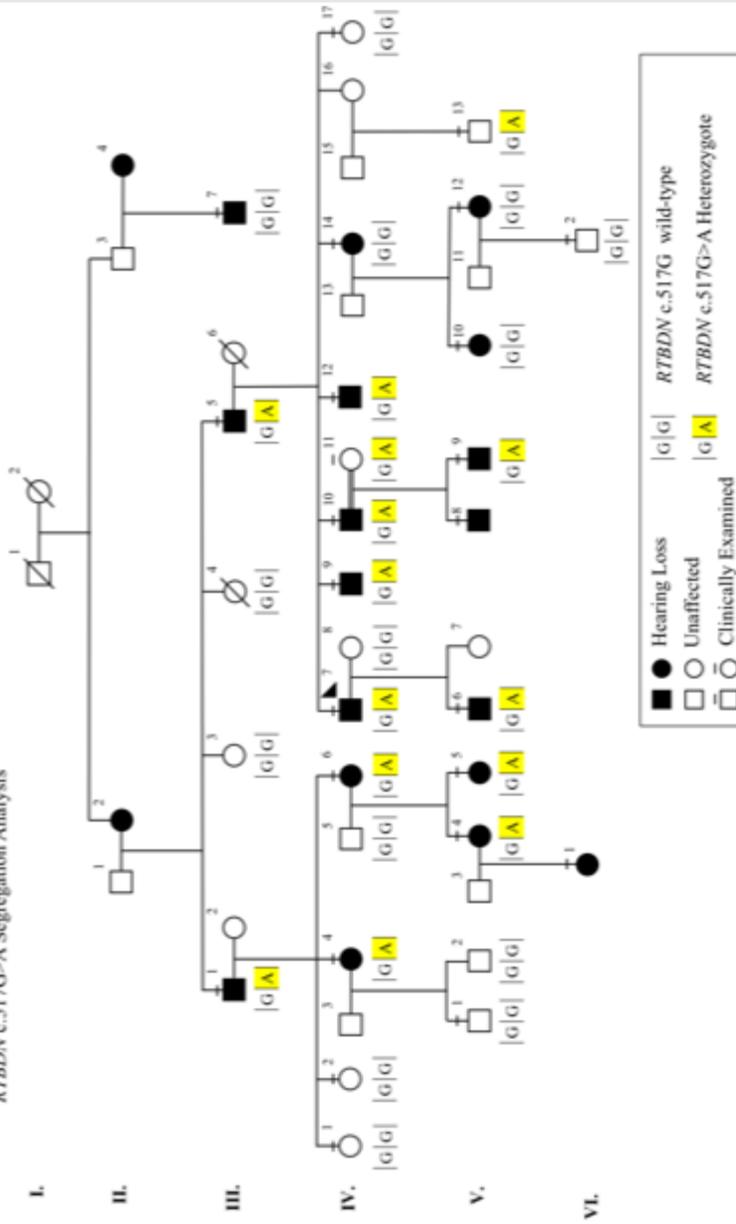
GPHB5 e.155G>C Segregation Analysis



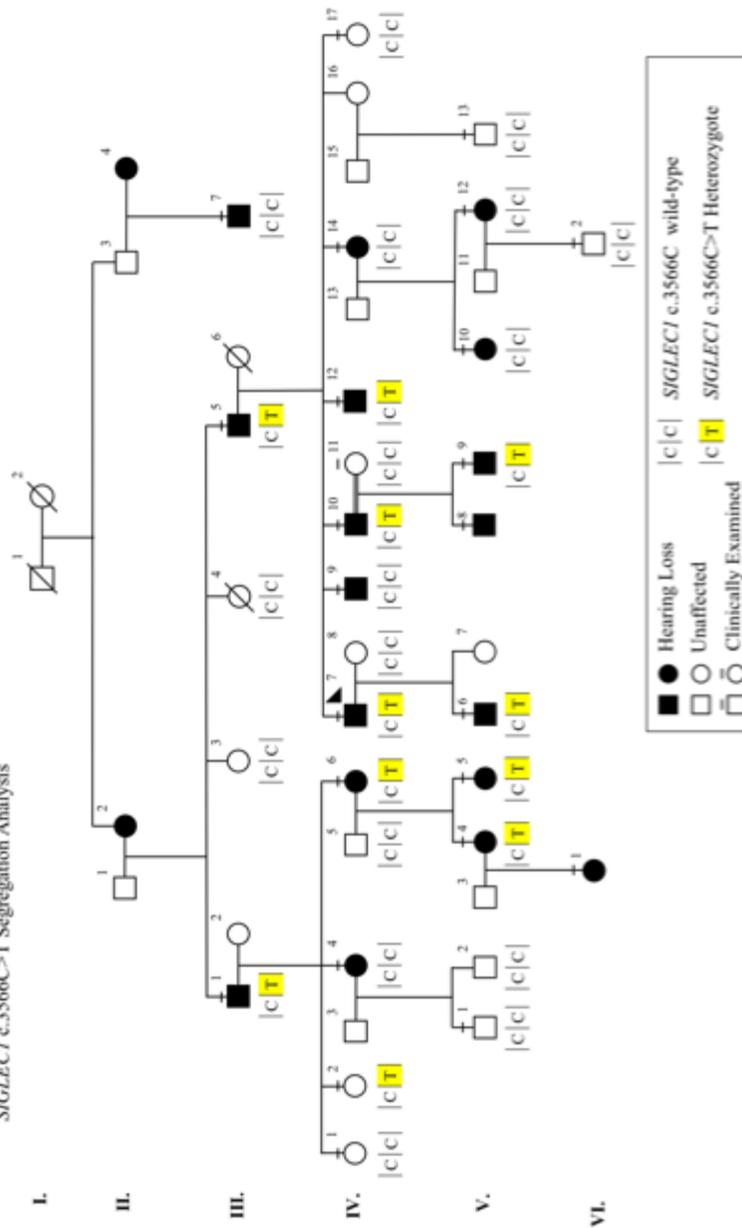
GTF2F1 c.1111delAA Segregation Analysis



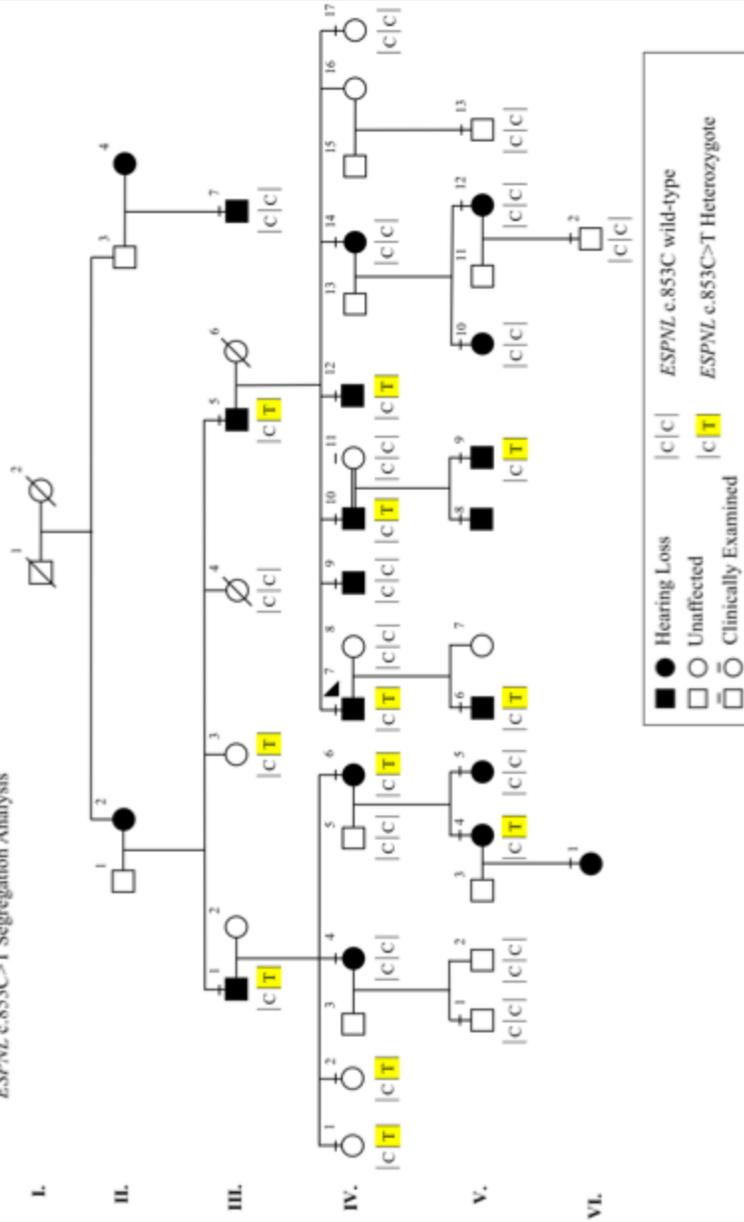
RTBDN c.517G>A Segregation Analysis



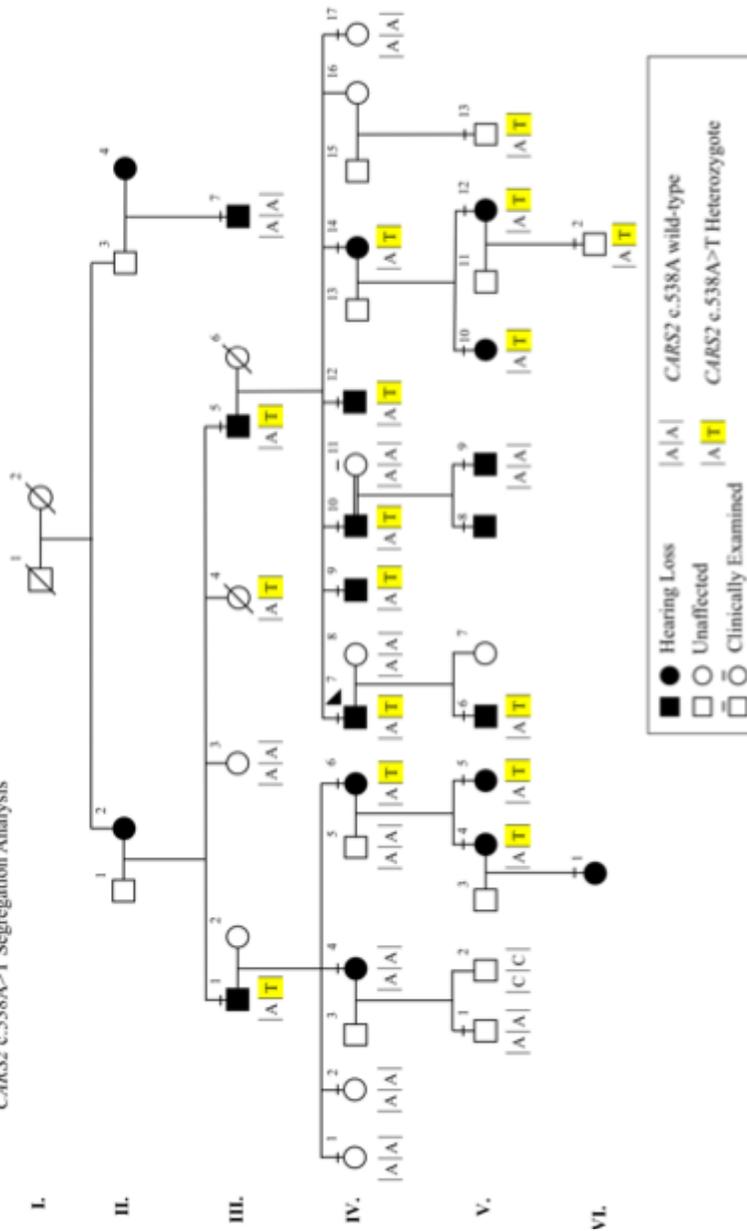
SIGLEC1 c.3566C>T Segregation Analysis



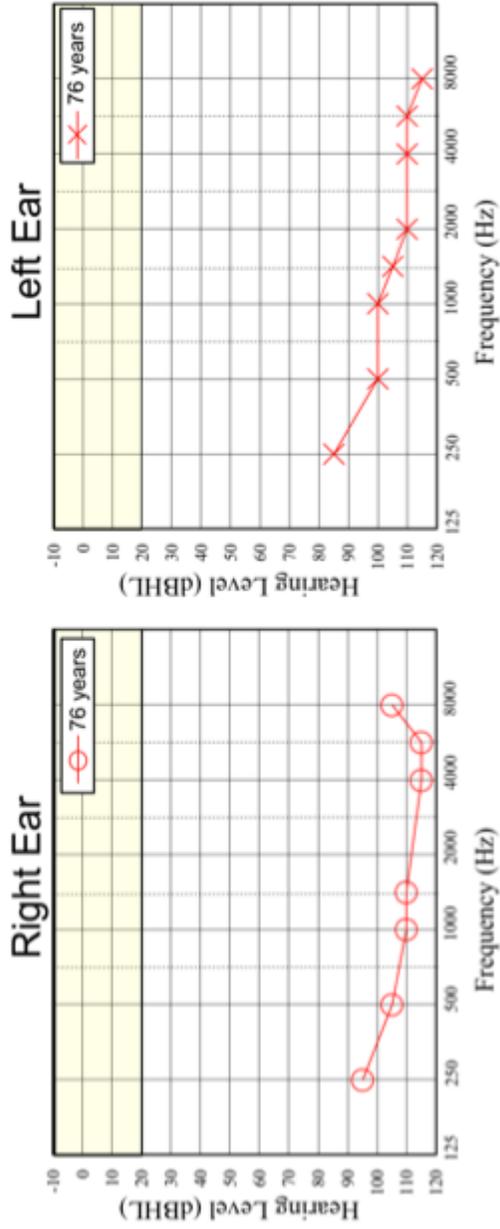
ESP/NL c.853C>T Segregation Analysis



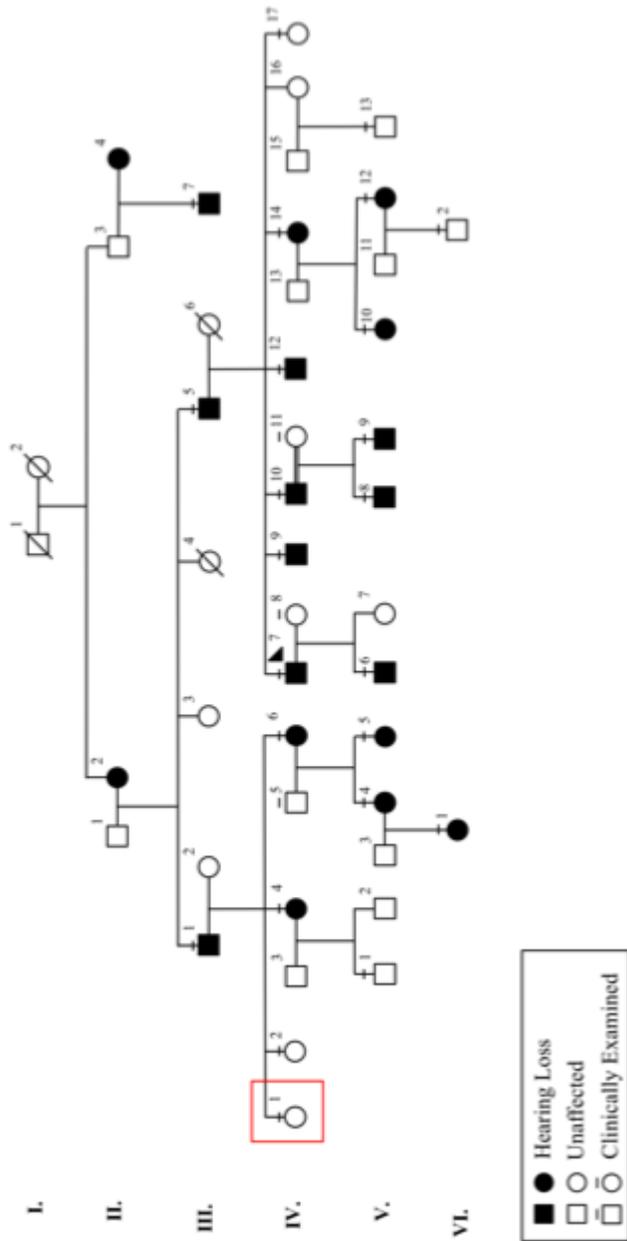
CARS2 c.538A>T Segregation Analysis



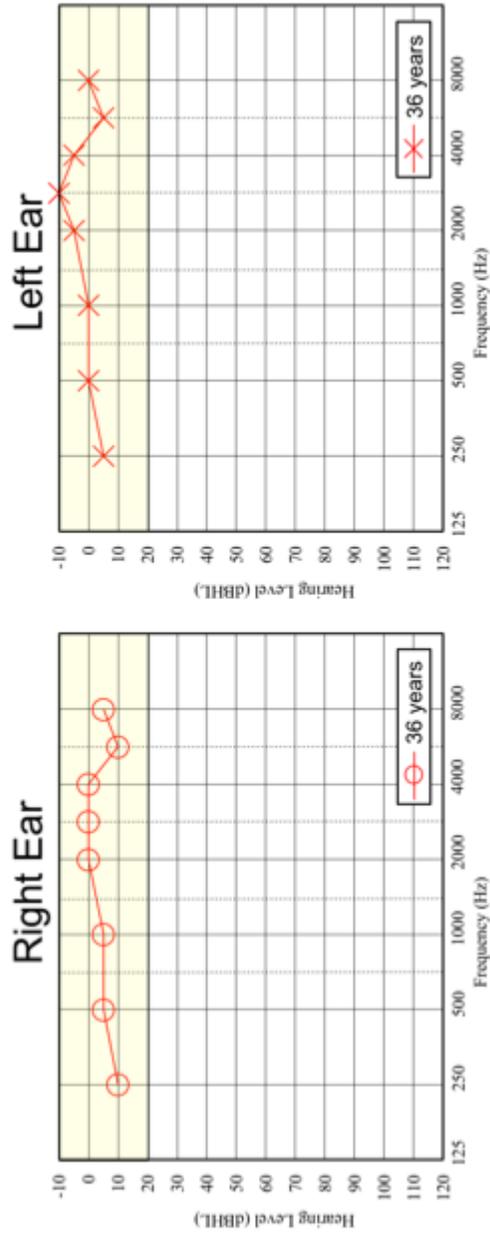
Audiological Data of PID III-1: Yellow area on audiogram indicates normal hearing levels



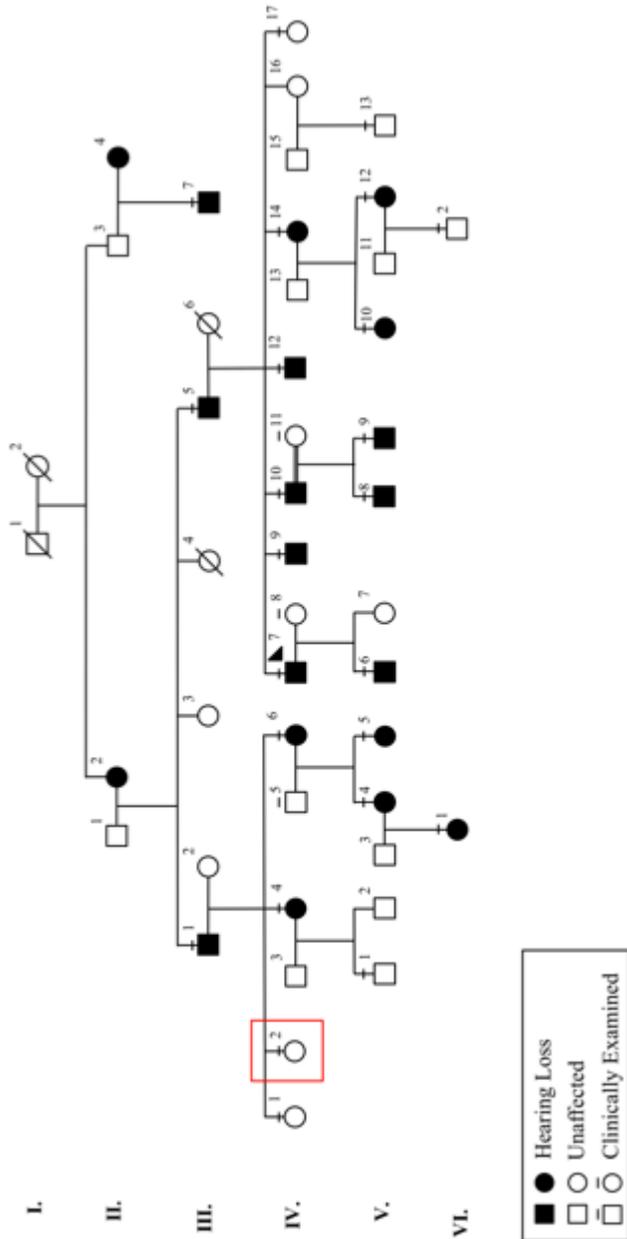
Family R2070 Pedigree indicating location of PID IV-1



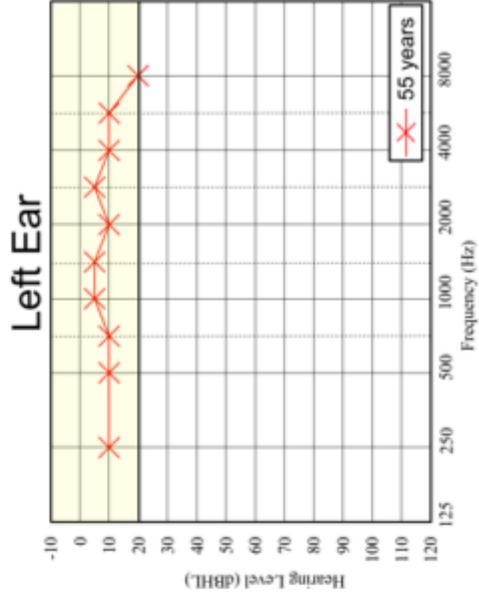
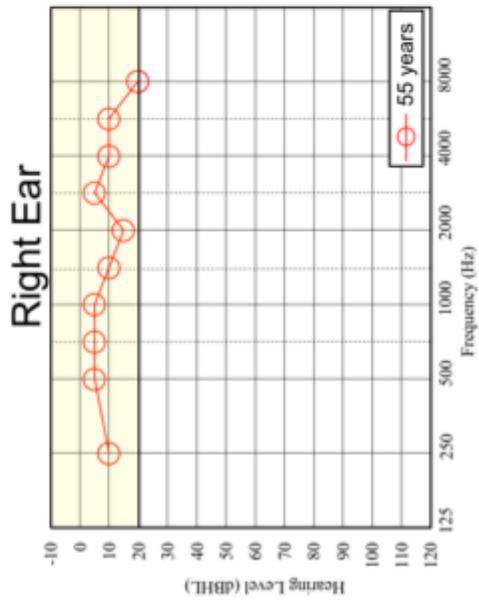
Audiological Data of **PID IV-1**: Yellow area on audiogram indicates normal hearing levels



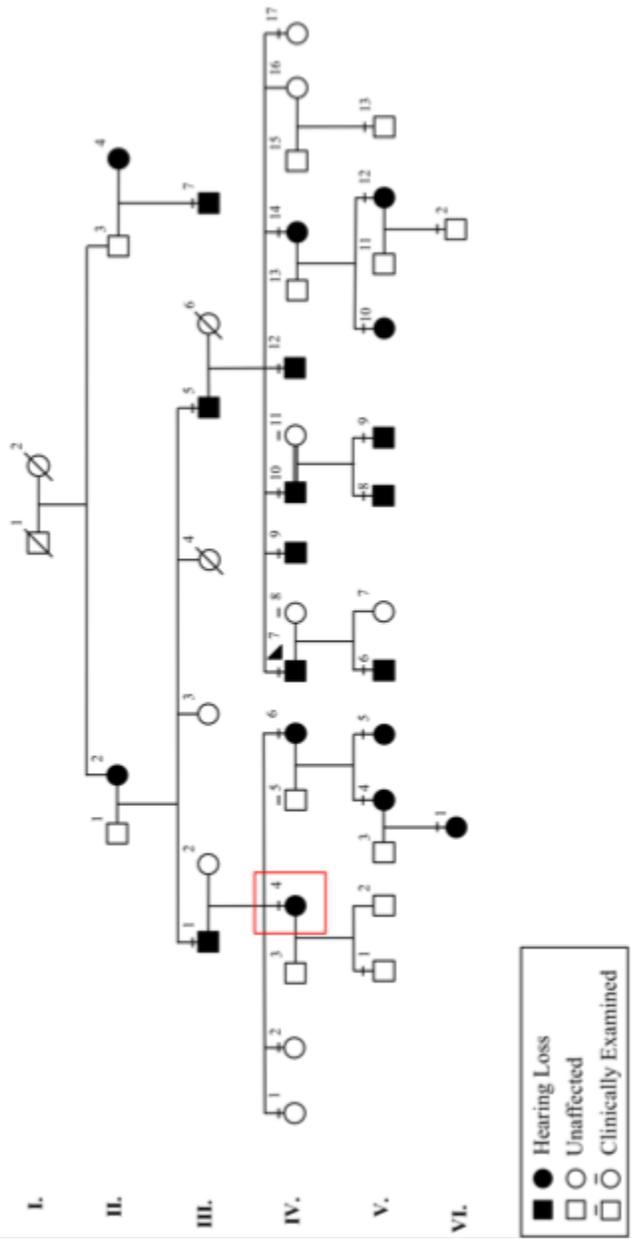
Family R2070 Pedigree indicating location of PID IV-2



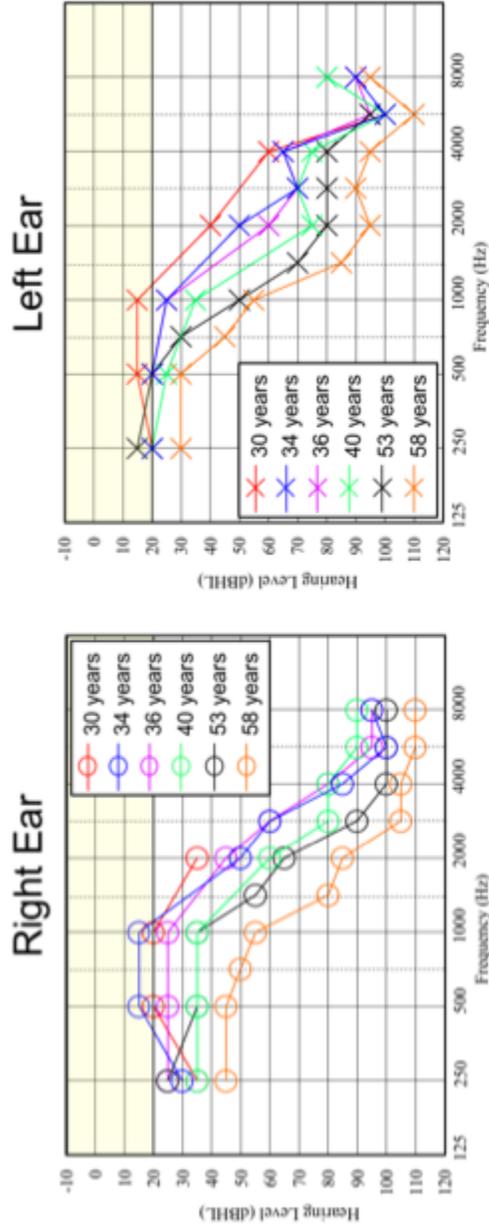
Audiological Data of PID IV-2: Yellow area on audiogram indicates normal hearing levels



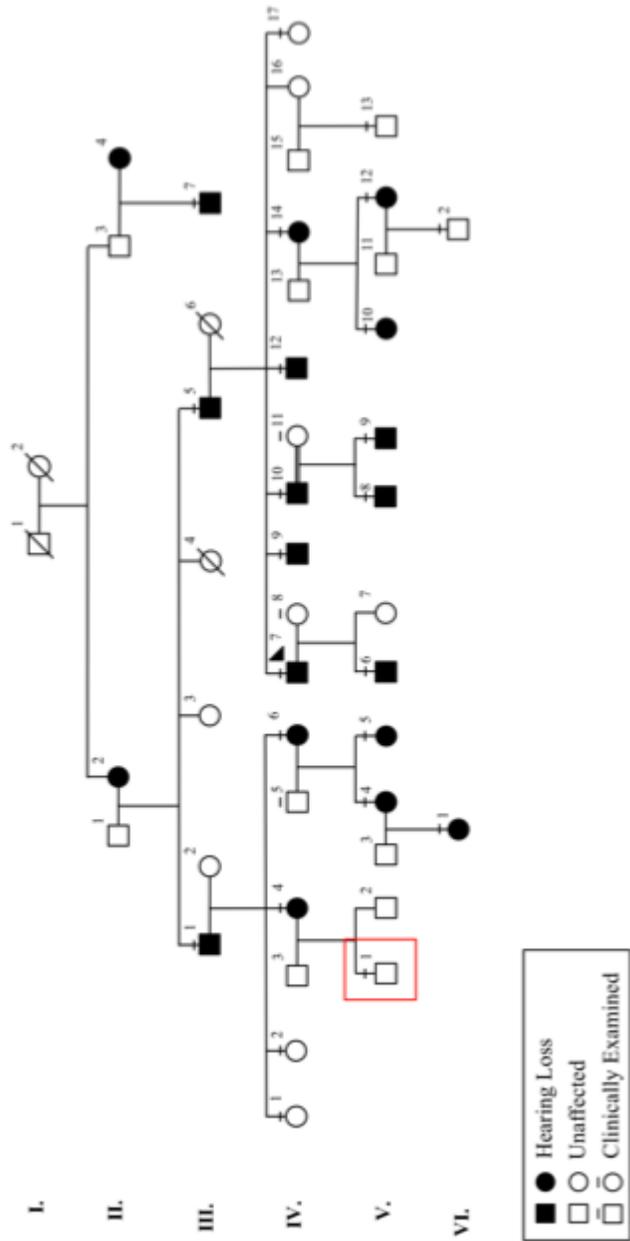
Family R2070 Pedigree indicating location of PID IV-4



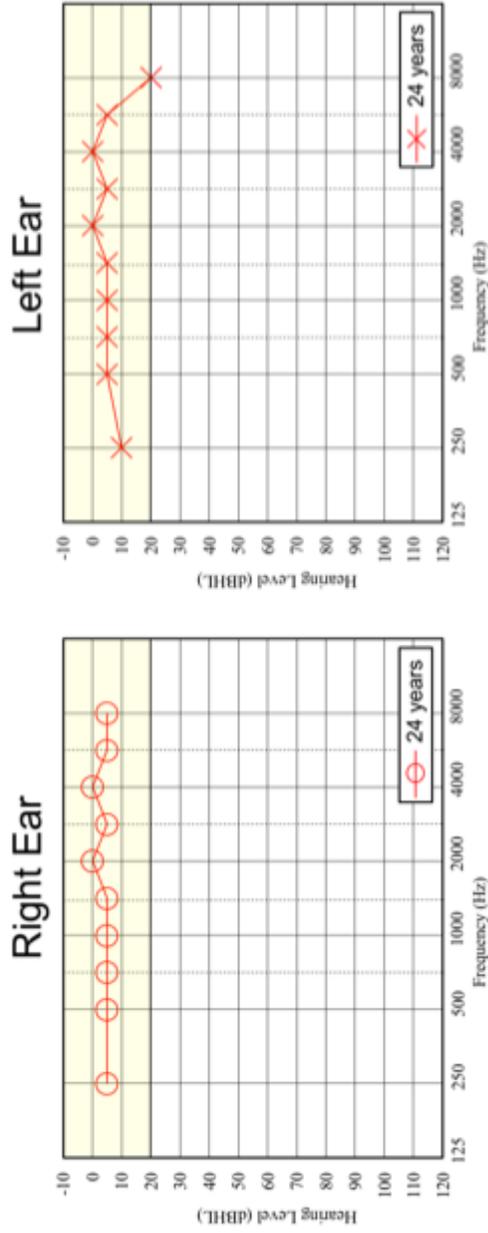
Audiological Data of PID IV-4: Yellow area on audiogram indicates normal hearing levels



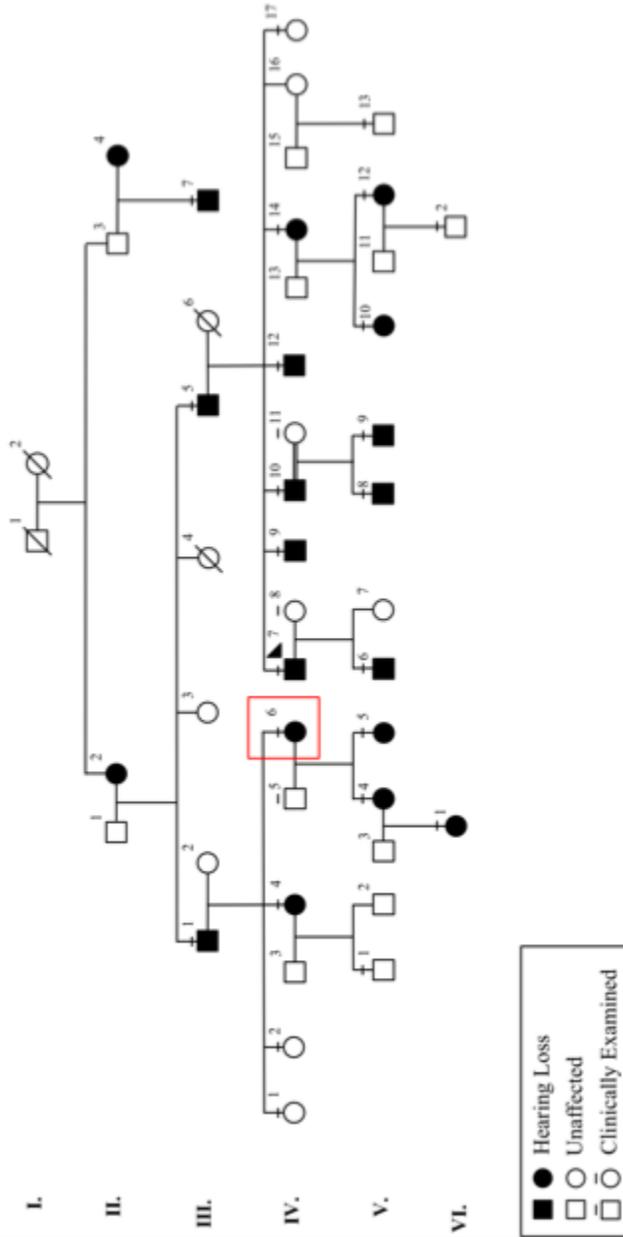
Family R2070 Pedigree indicating location of PID V-1



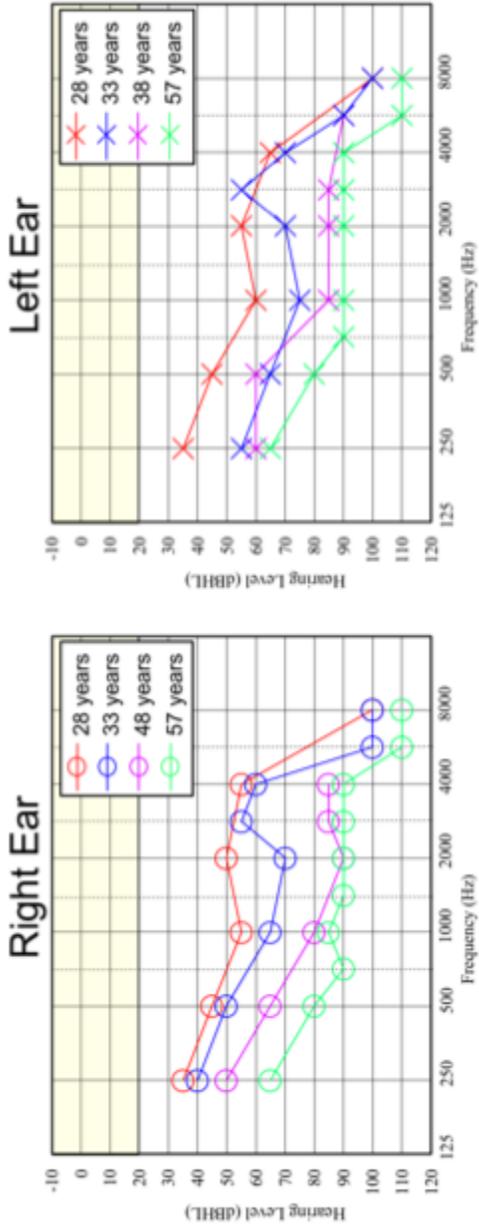
Audiological Data of **PID V-1**: Yellow area on audiogram indicates normal hearing levels



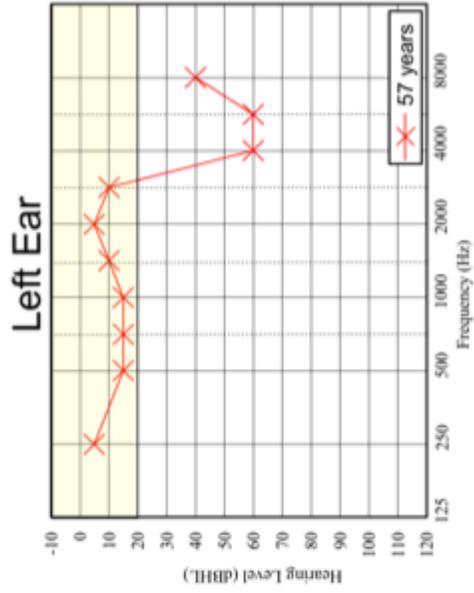
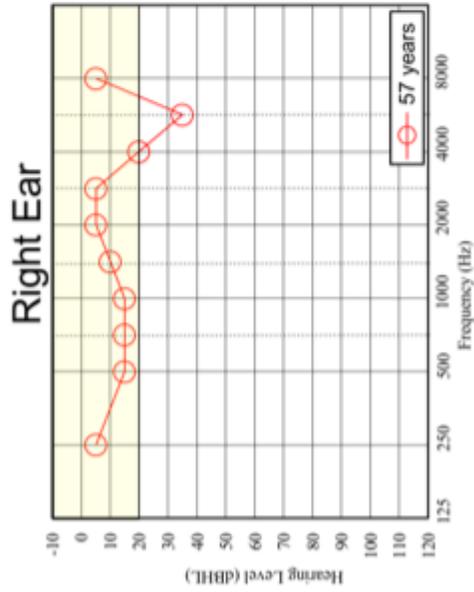
Family R2070 Pedigree indicating location of PID IV-6



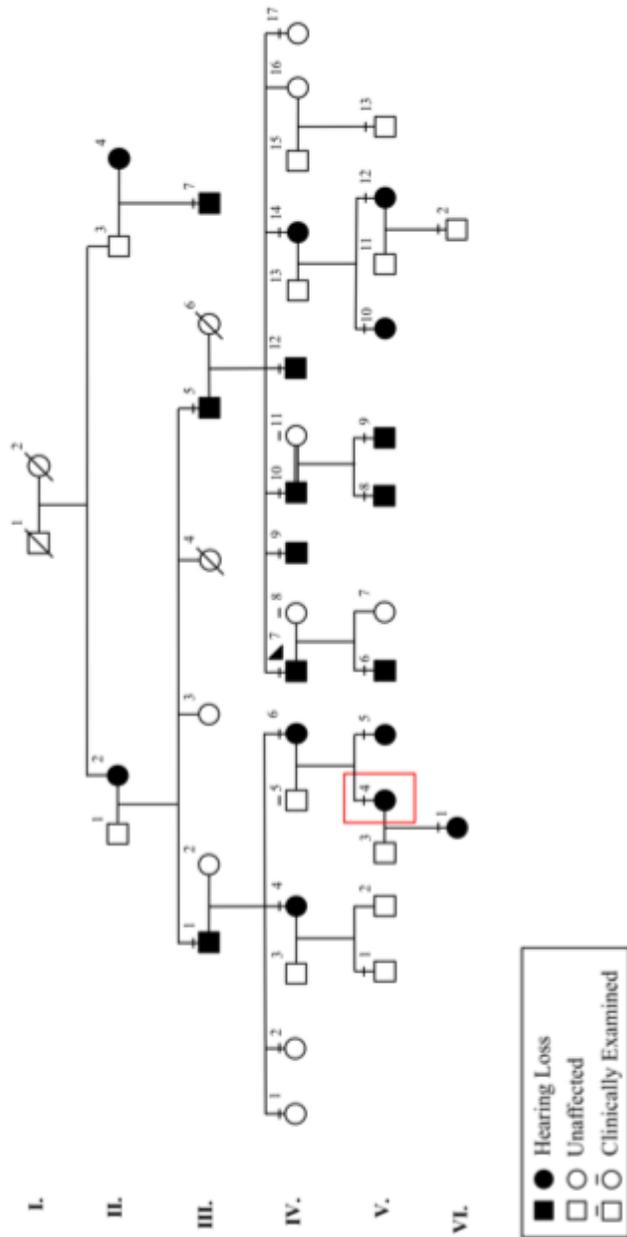
Audiological Data of PID IV-6: Yellow area on audiogram indicates normal hearing levels



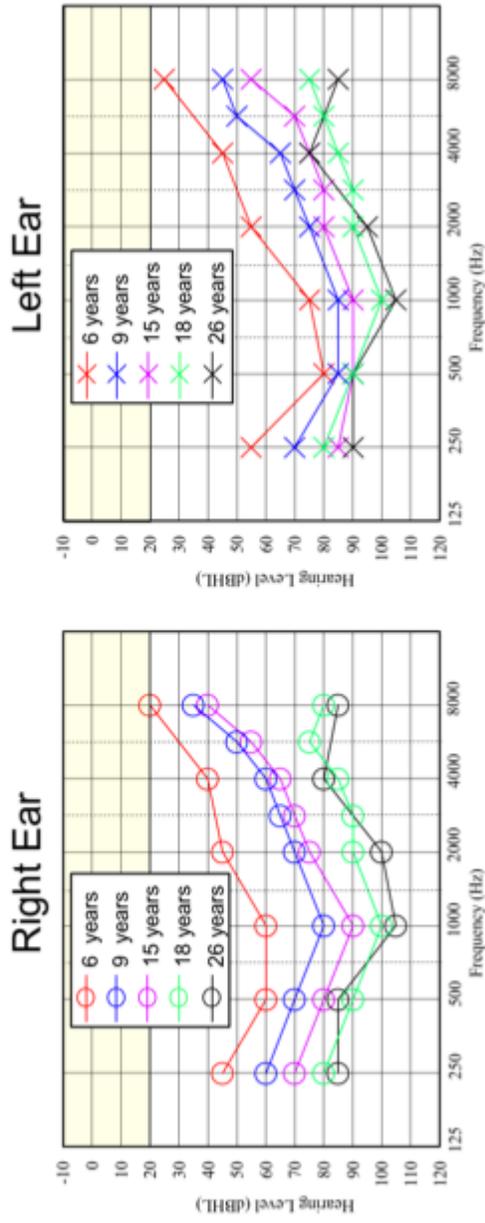
Audiological Data of PID IV-5: Yellow area on audiogram indicates normal hearing levels



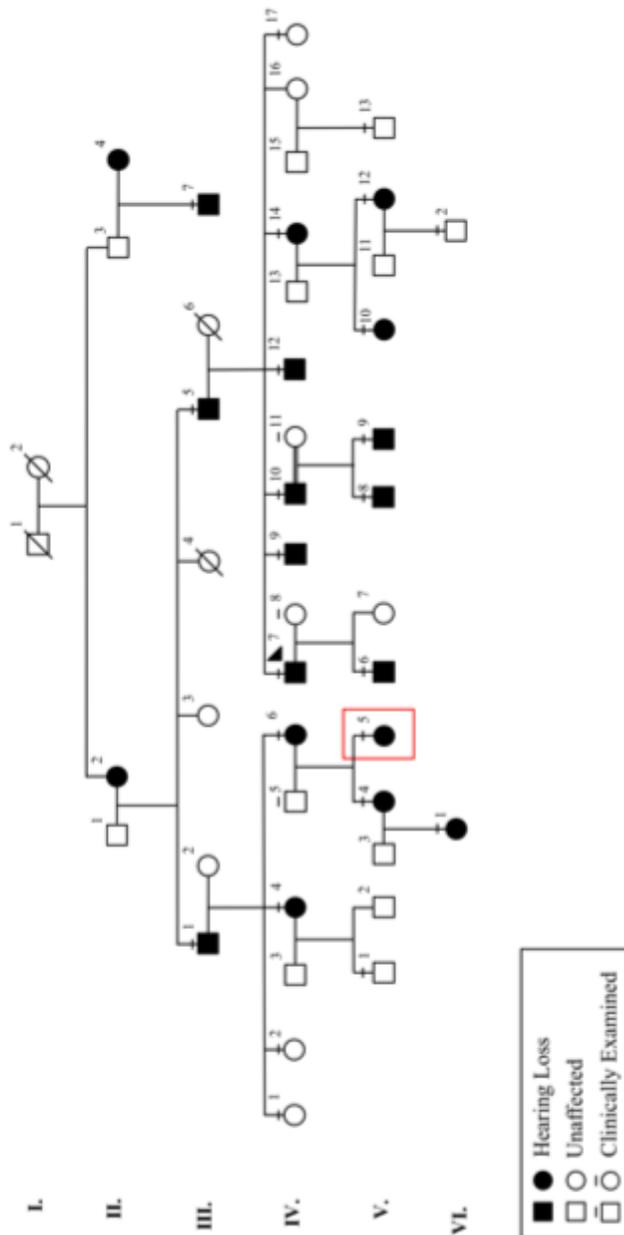
Family R2070 Pedigree indicating location of PID V-4



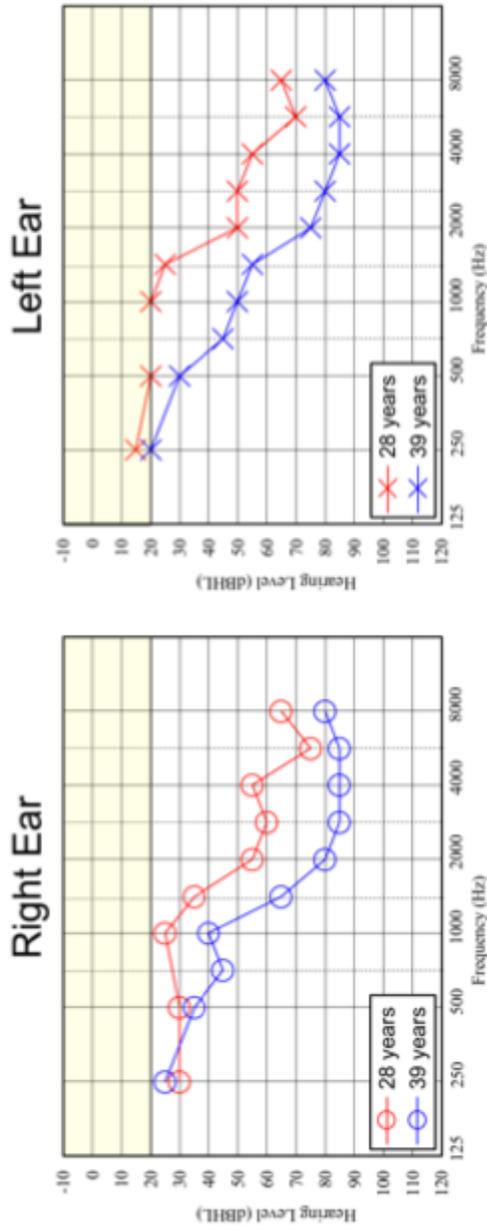
Audiological Data of PID V-4: Yellow area on audiogram indicates normal hearing levels



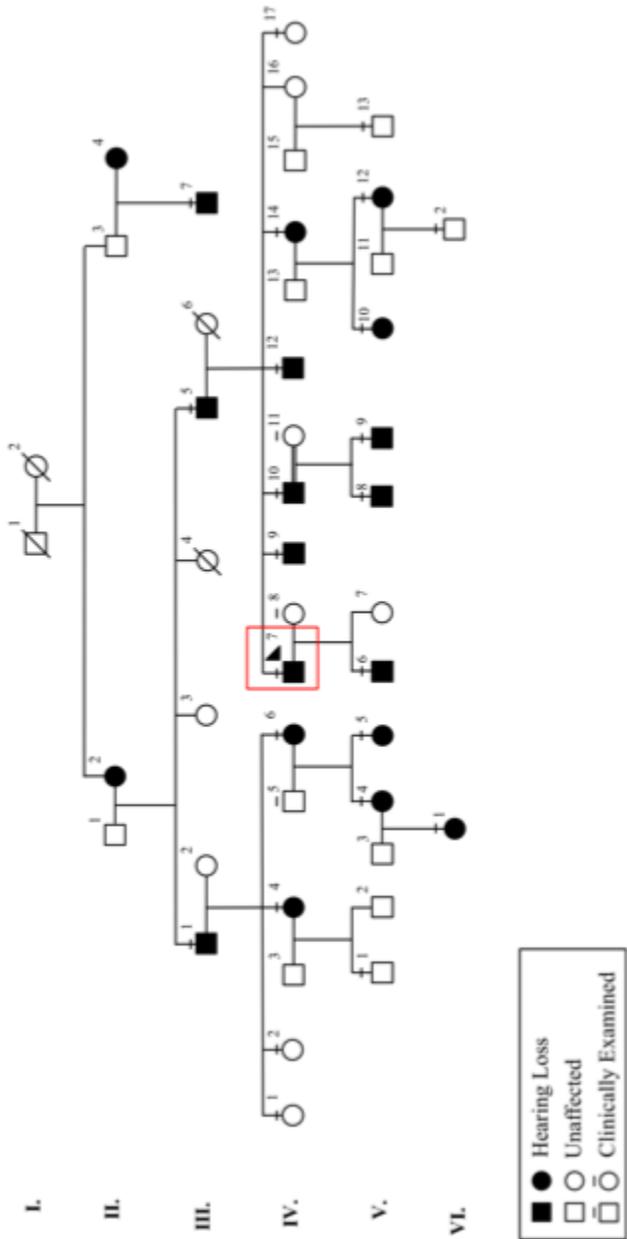
Family R2070 Pedigree indicating location of PID V-5



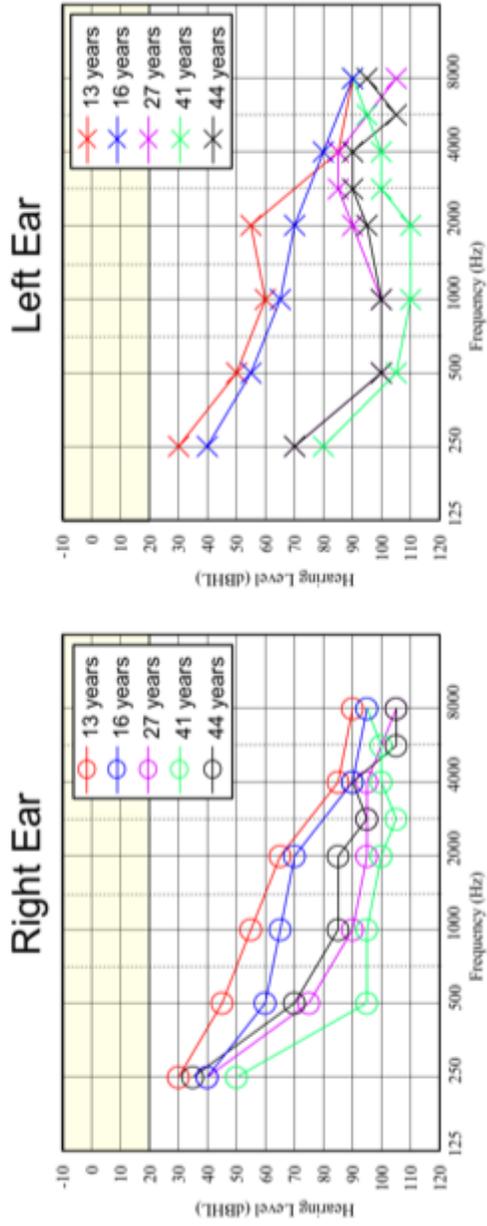
Audiological Data of PID V-5: Yellow area on audiogram indicates normal hearing levels



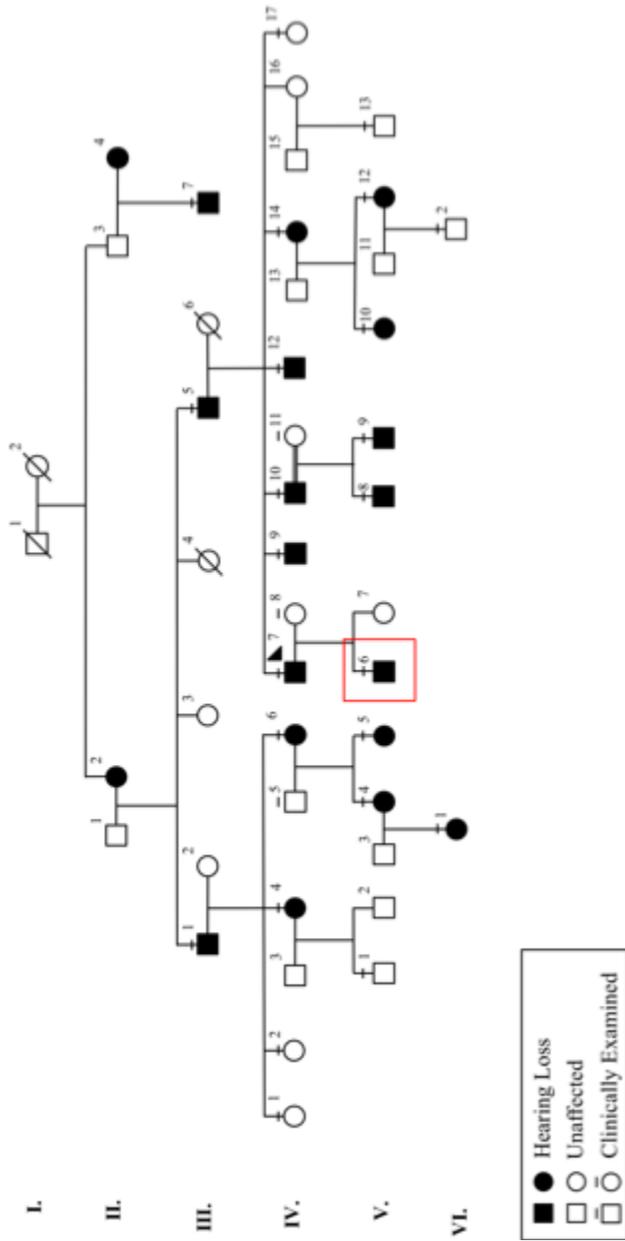
Family R2070 Pedigree indicating location of PID IV-7



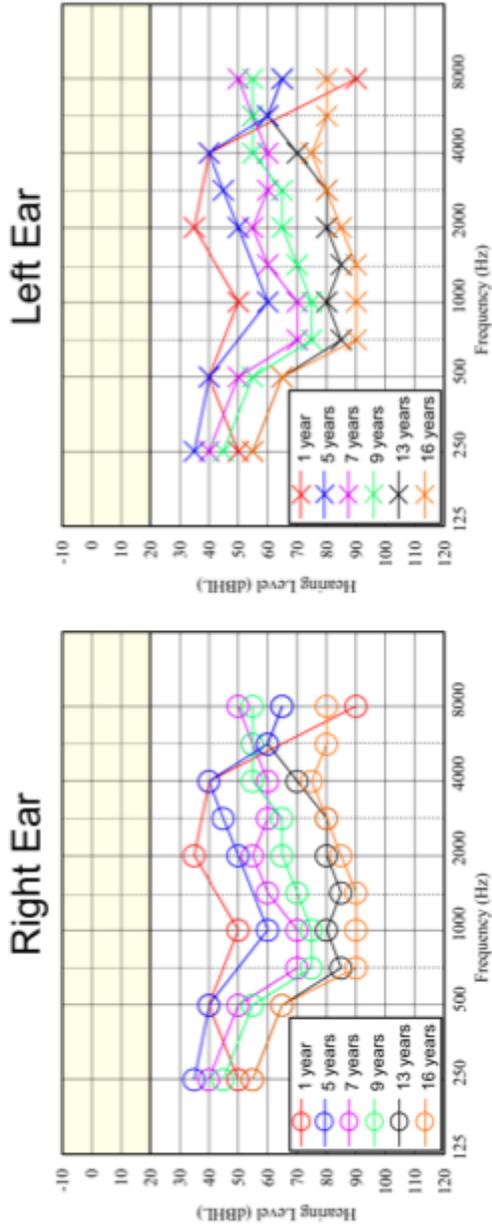
Audiological Data of PID IV-7: Yellow area on audiogram indicates normal hearing levels



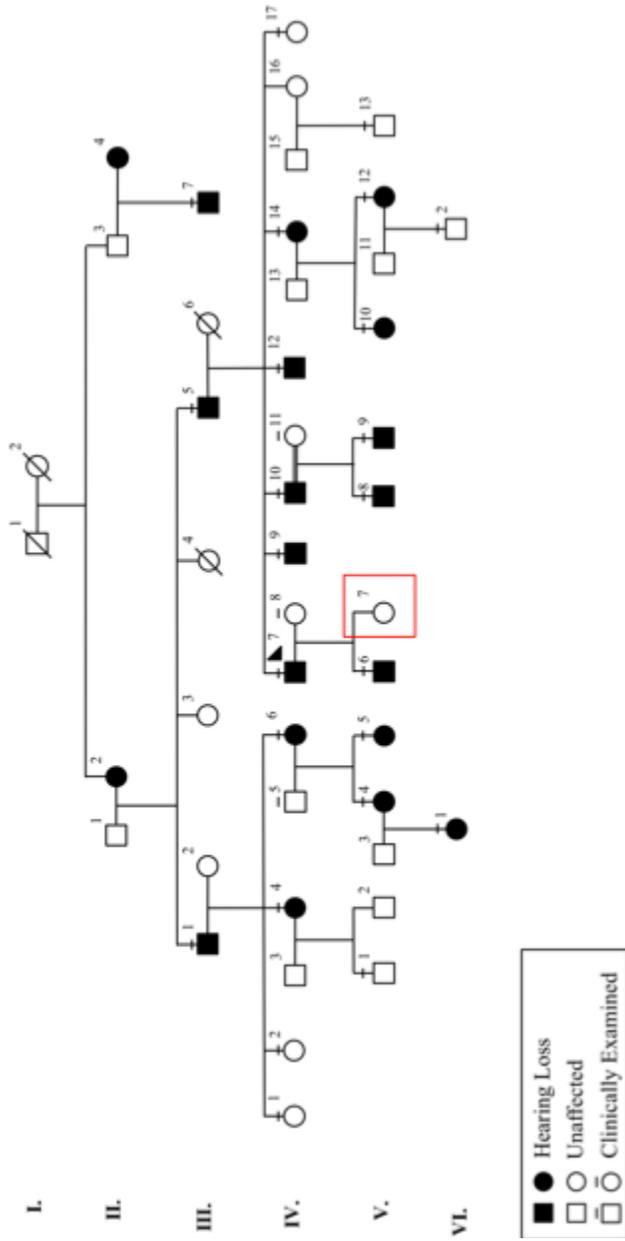
Family R2070 Pedigree indicating location of PID V-6



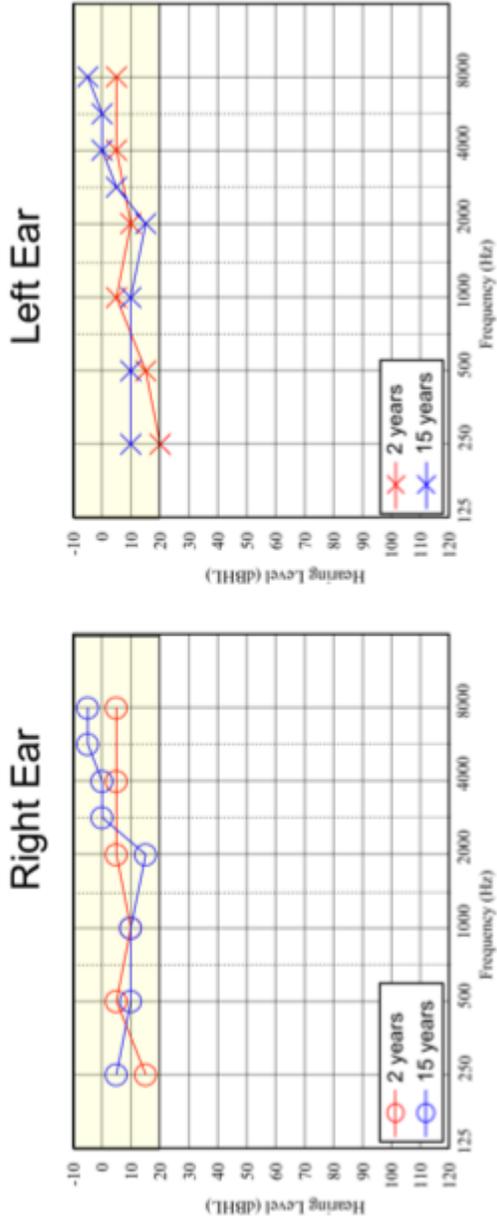
Audiological Data of PID V-6: Yellow area on audiogram indicates normal hearing levels



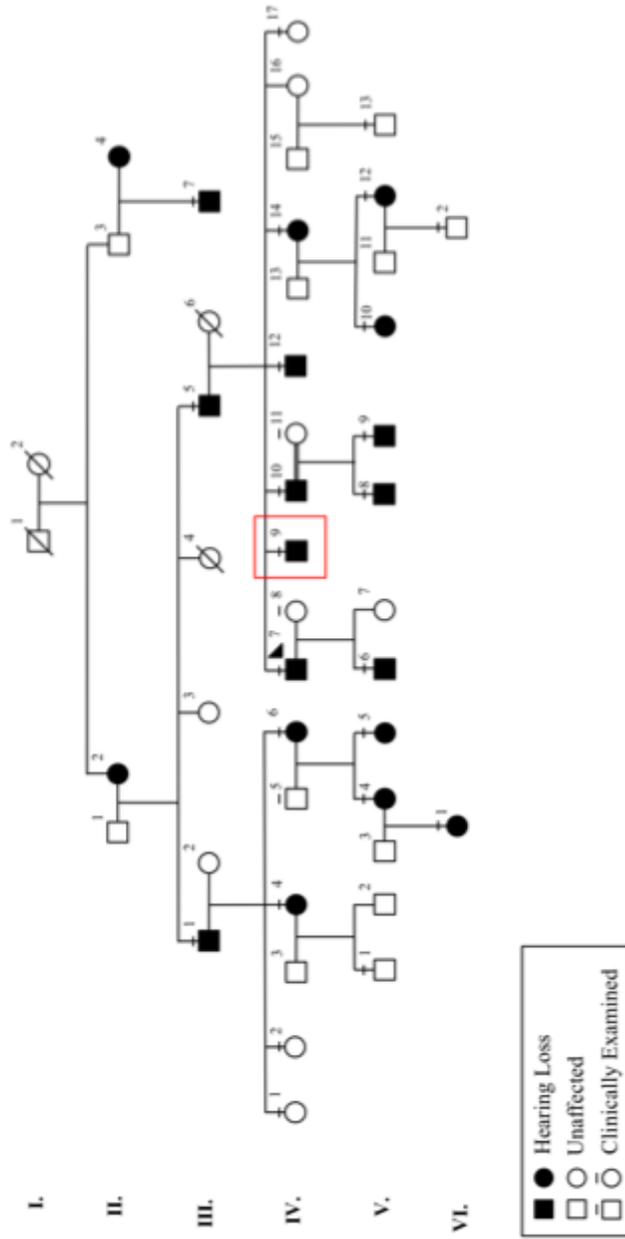
Family R2070 Pedigree indicating location of PID V-7



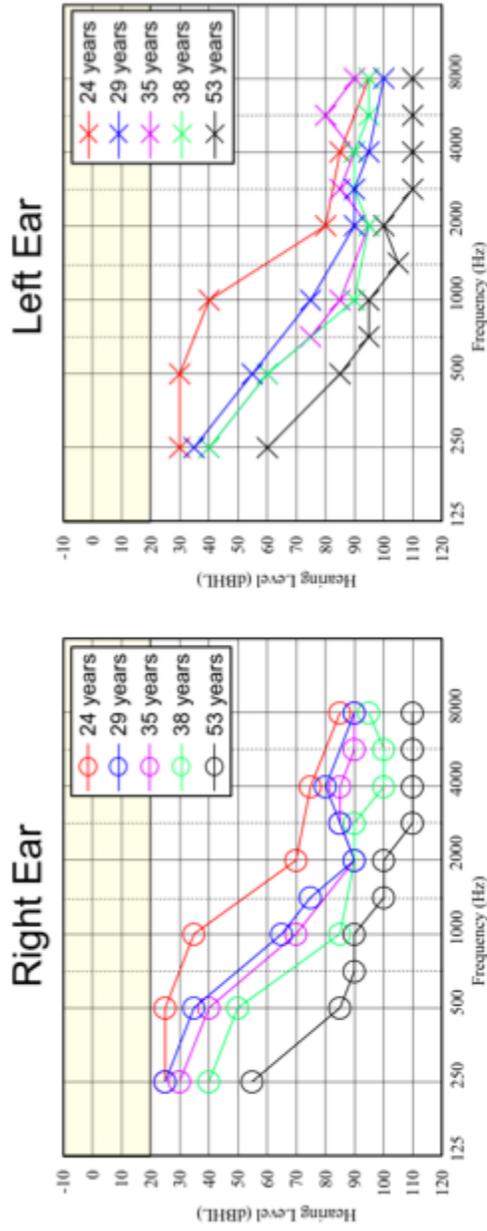
Audiological Data of **PID V-7**: Yellow area on audiogram indicates normal hearing levels



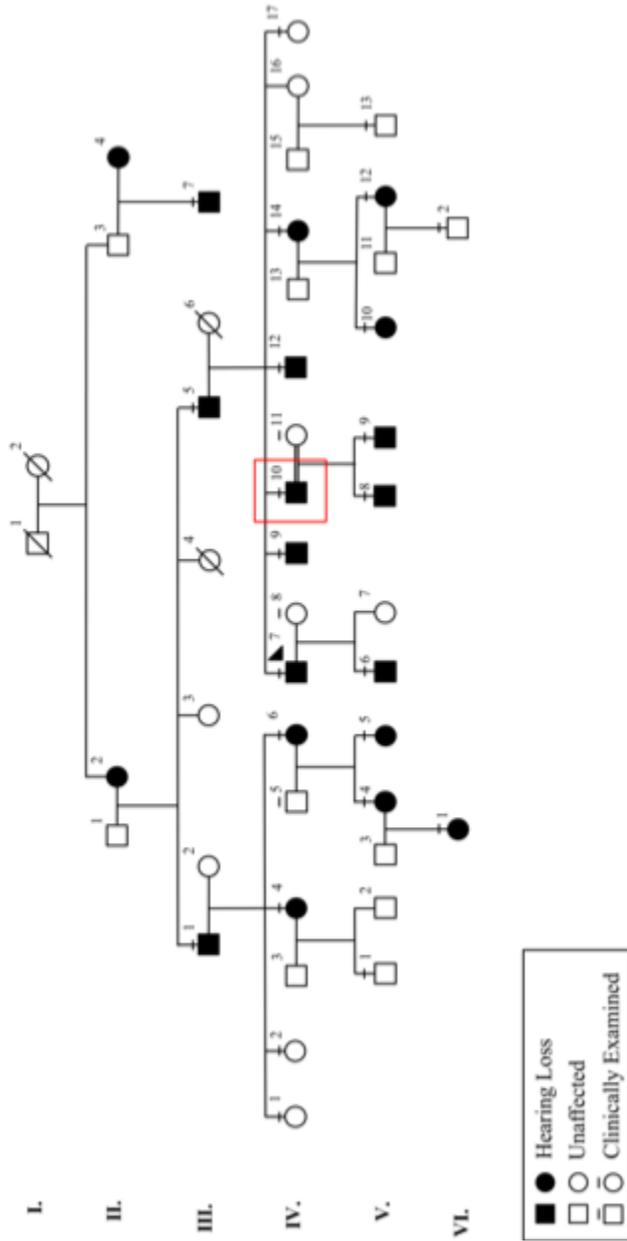
Family R2070 Pedigree indicating location of **PID IV-9**



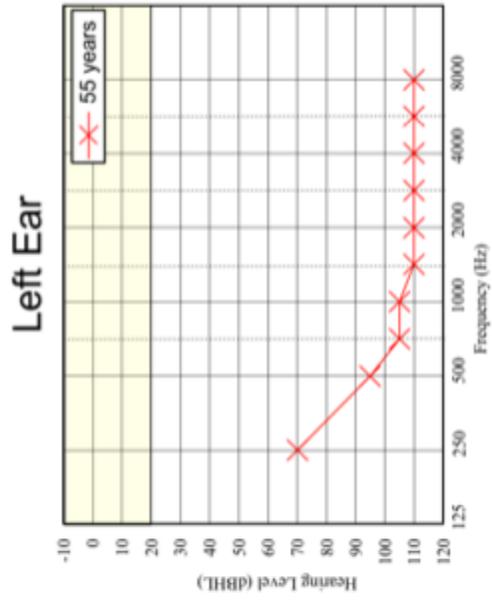
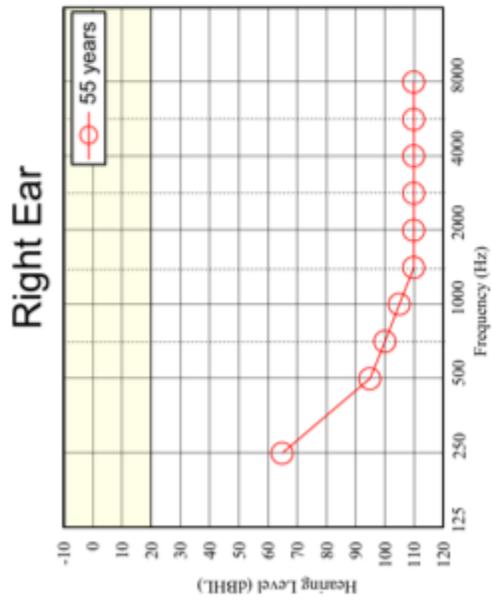
Audiological Data of PID IV-9: Yellow area on audiogram indicates normal hearing levels



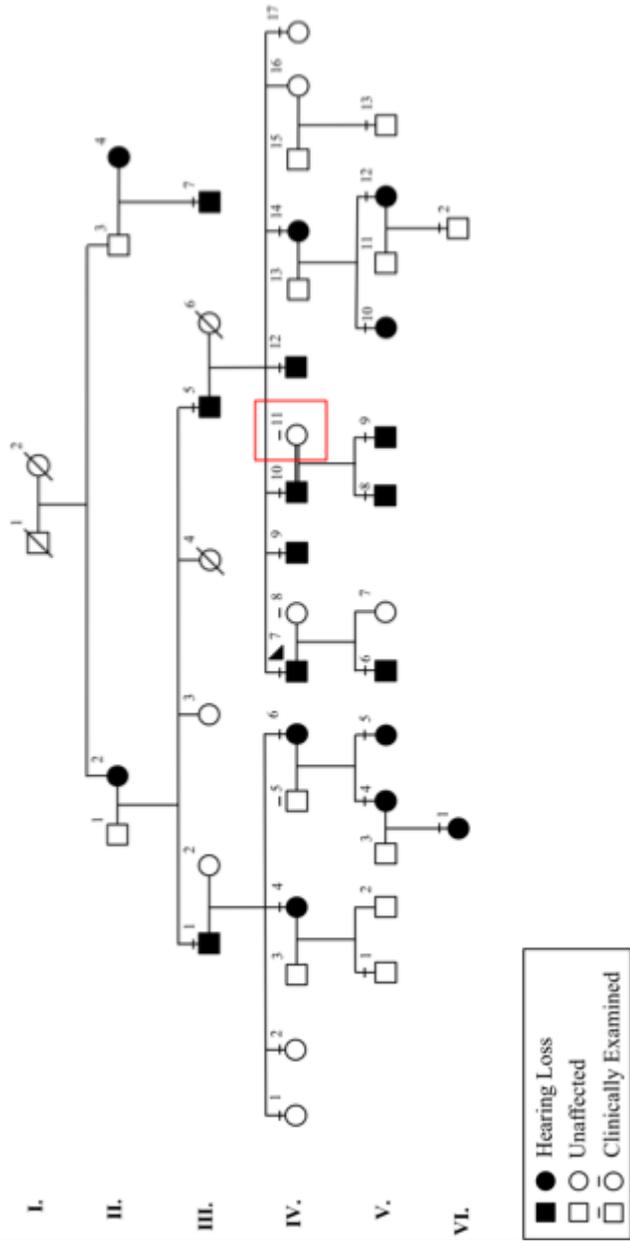
Family R2070 Pedigree indicating location of PID IV-10



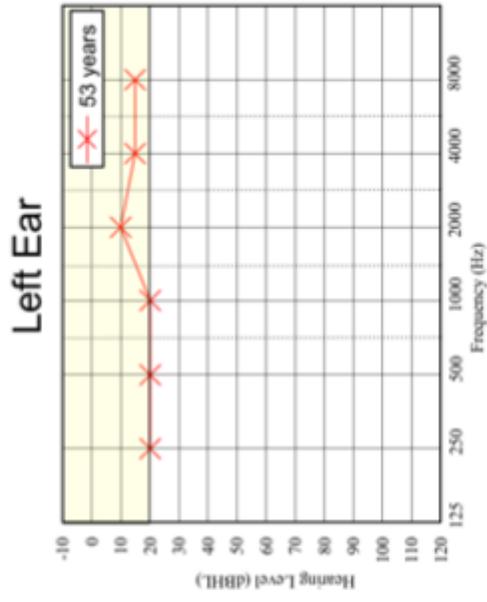
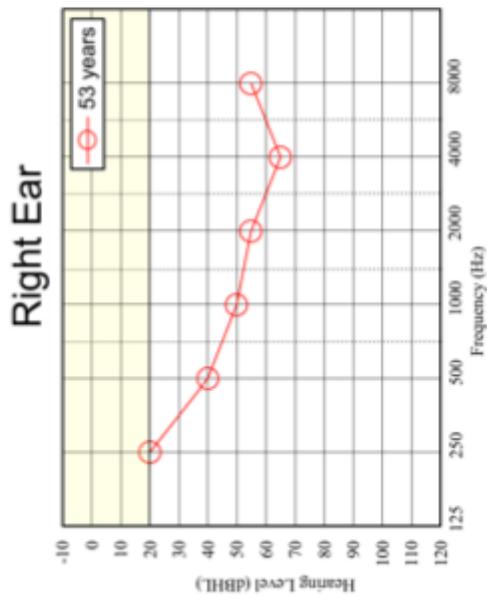
Audiological Data of PID IV-10: Yellow area on audiogram indicates normal hearing levels



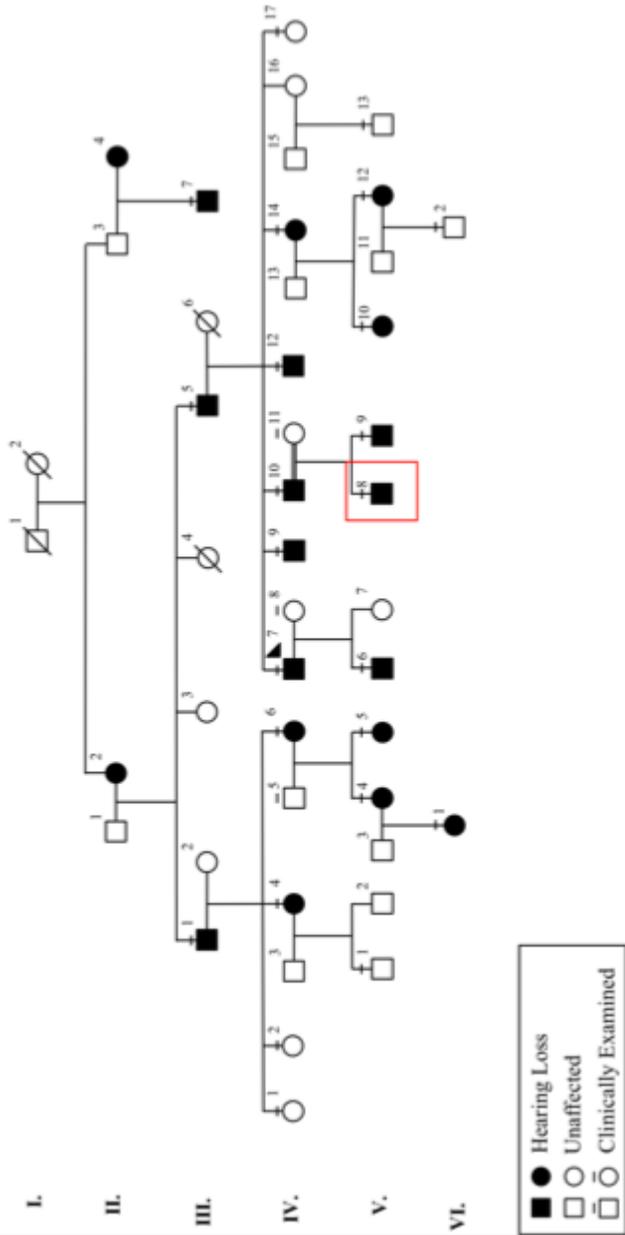
Family R2070 Pedigree indicating location of PID IV-11



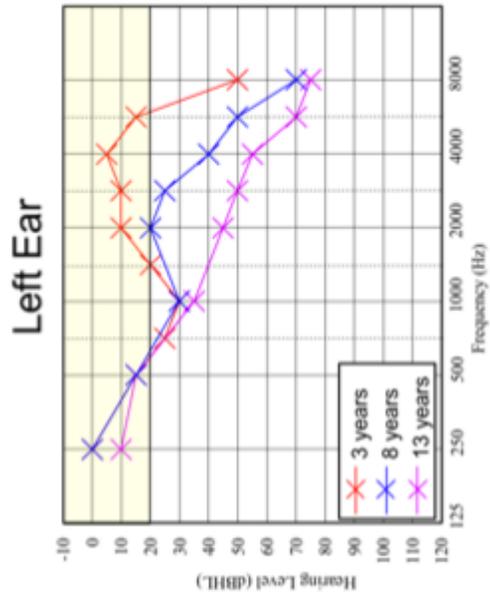
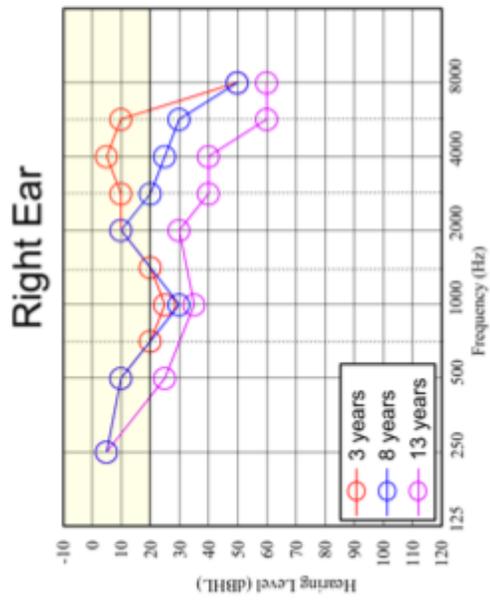
Audiological Data of PID IV-11: Yellow area on audiogram indicates normal hearing levels



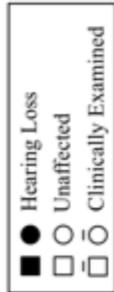
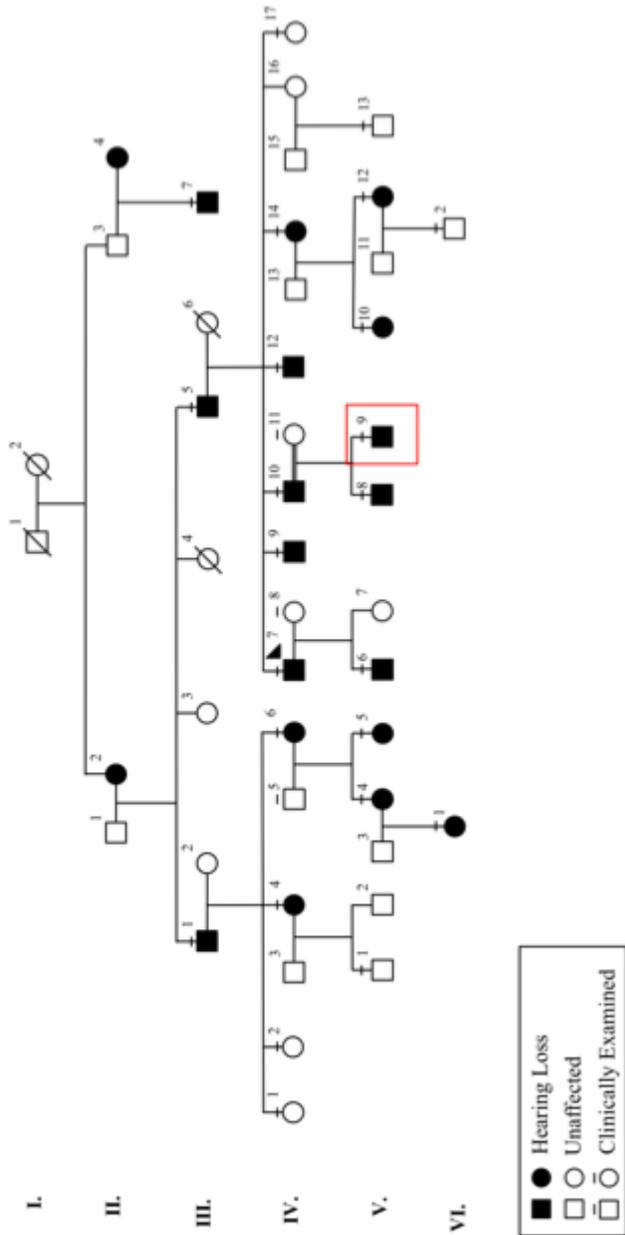
Family R2070 Pedigree indicating location of PID V-8



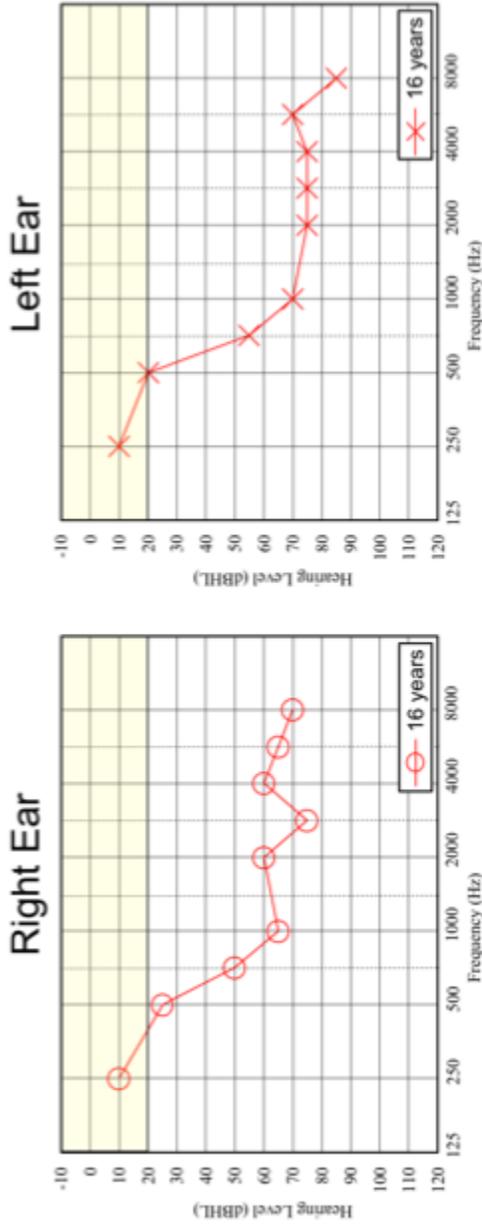
Audiological Data of PID V-8: Yellow area on audiogram indicates normal hearing levels



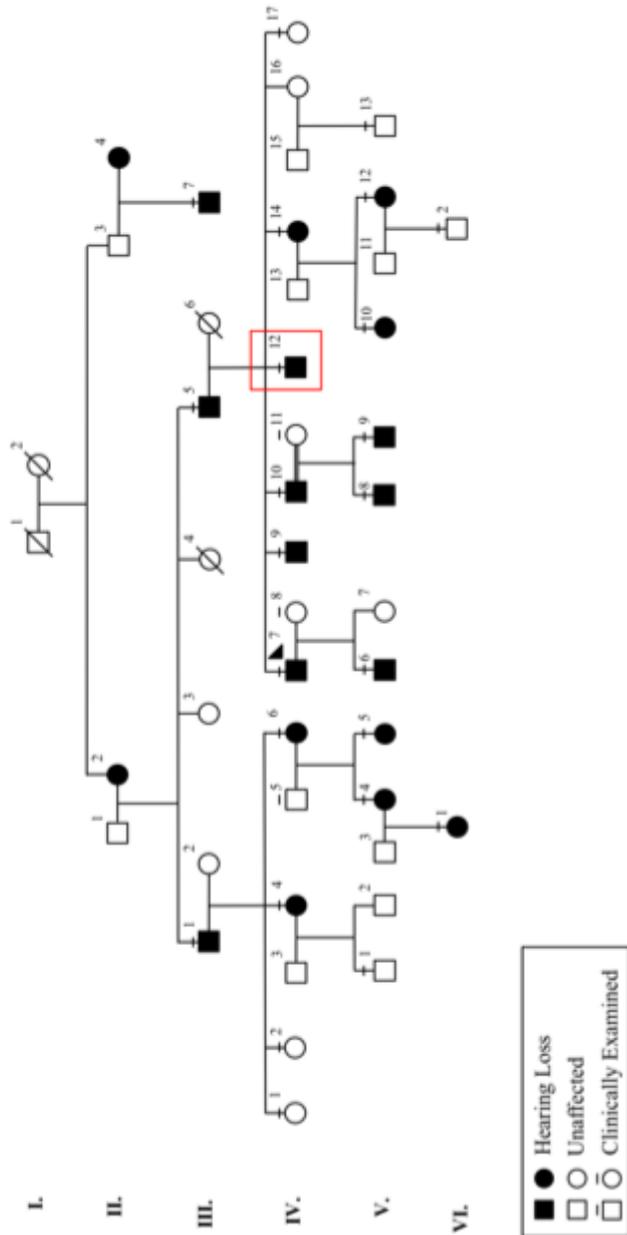
Family R2070 Pedigree indicating location of PID V-9



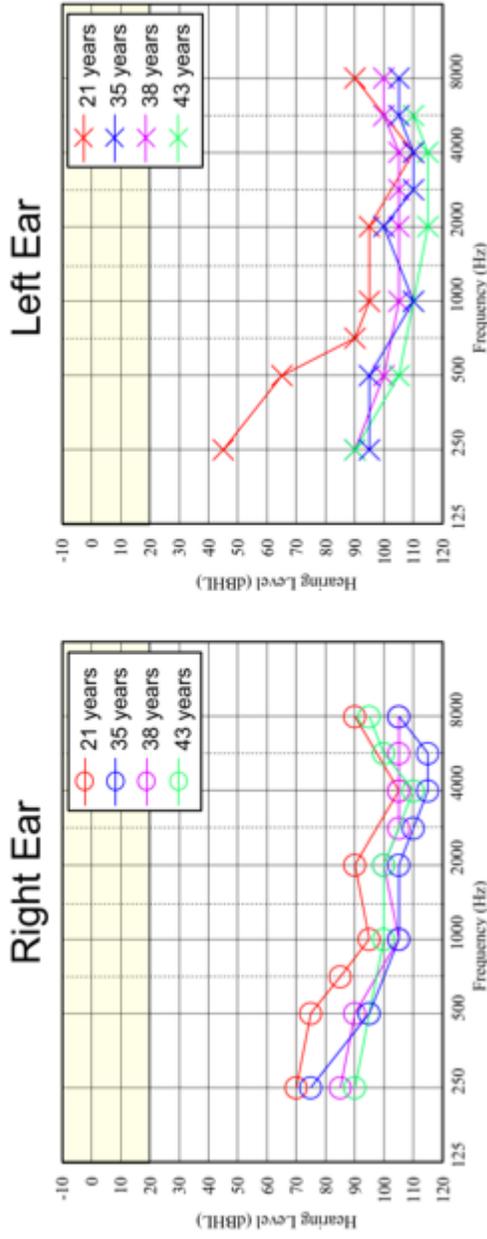
Audiological Data of PID V-9: Yellow area on audiogram indicates normal hearing levels



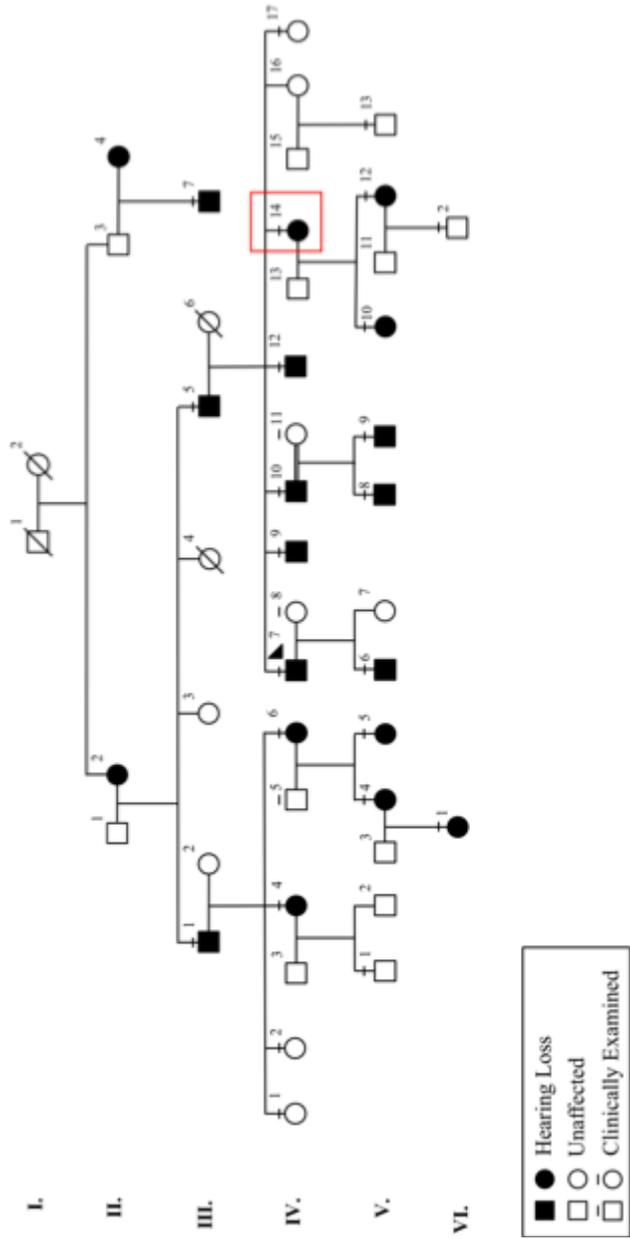
Family R2070 Pedigree indicating location of PID IV-12



Audiological Data of PID IV-12: Yellow area on audiogram indicates normal hearing levels

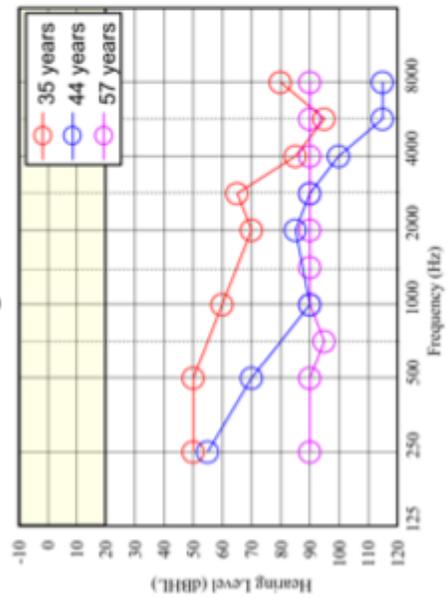


Family R2070 Pedigree indicating location of PID IV-14

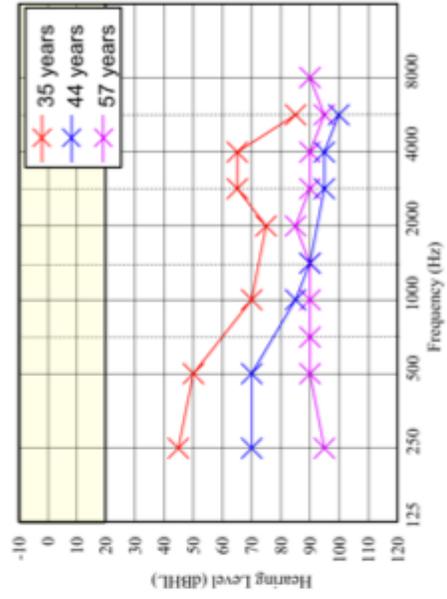


Audiological Data of PID IV-14: Yellow area on audiogram indicates normal hearing levels

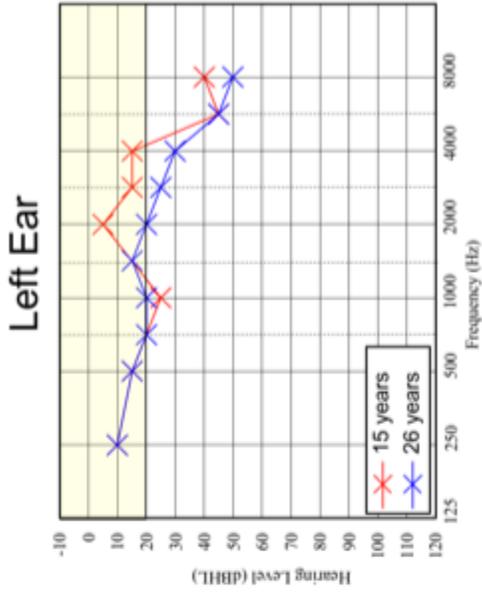
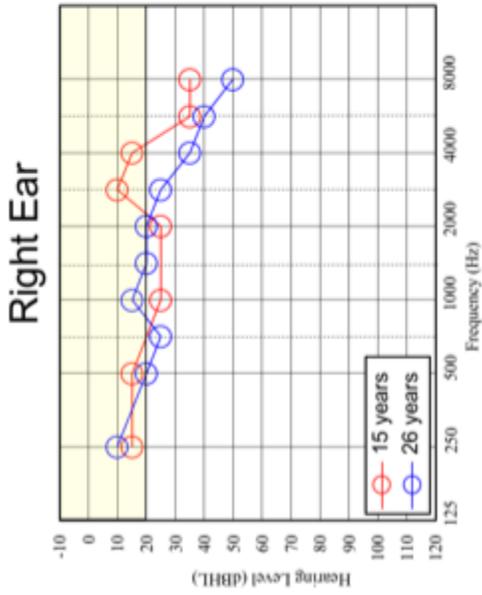
Right Ear



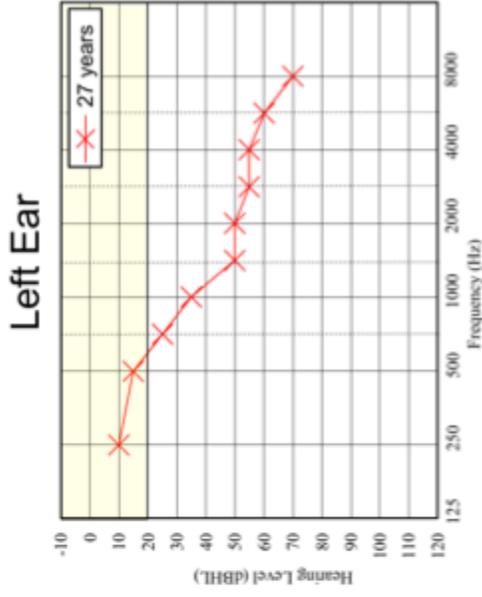
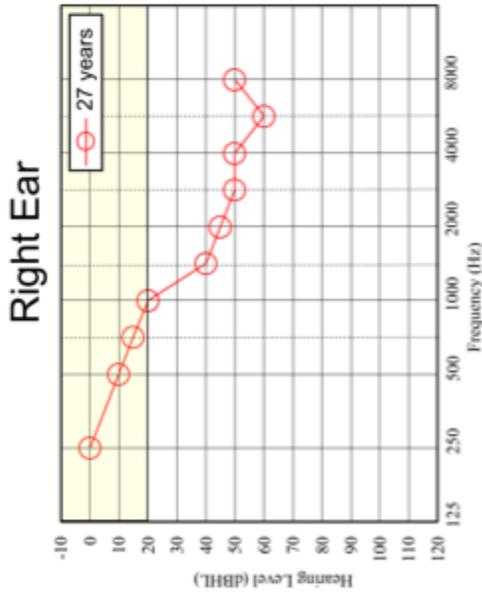
Left Ear



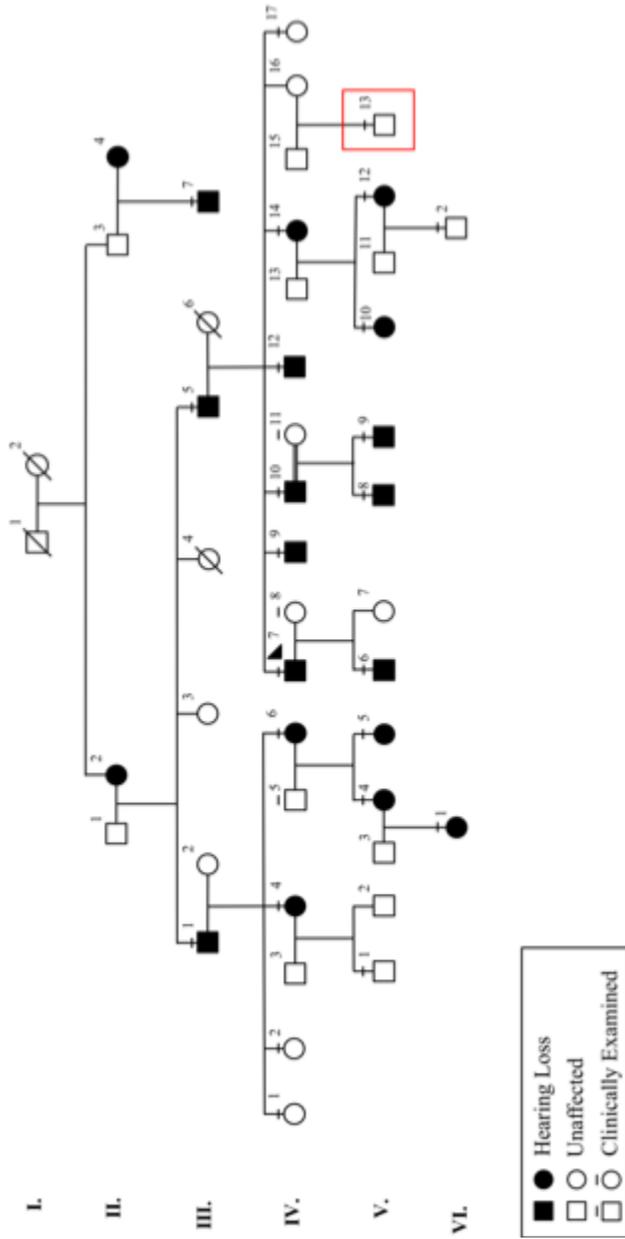
Audiological Data of PID V-10: Yellow area on audiogram indicates normal hearing levels



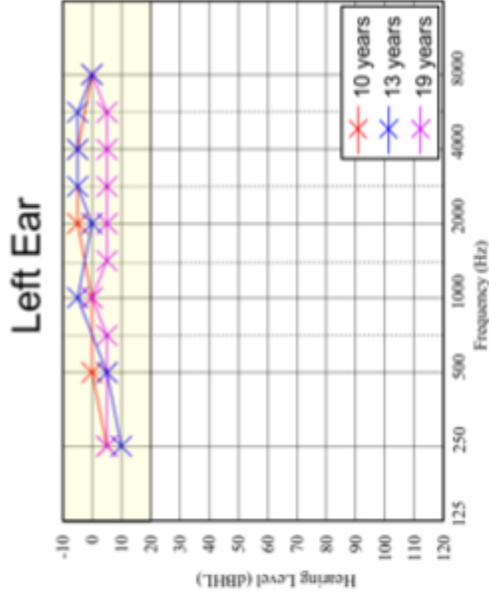
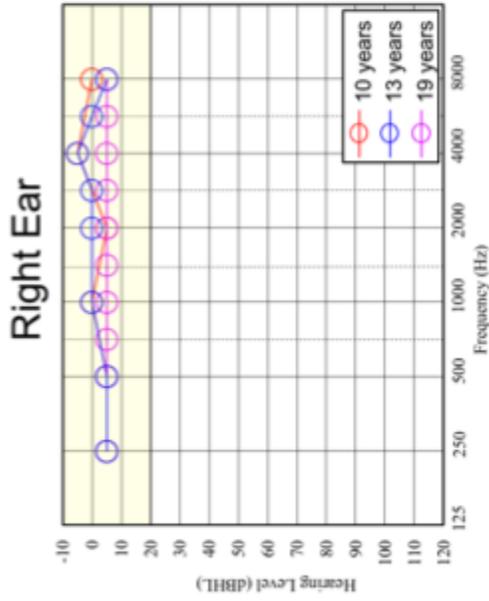
Audiological Data of PID V-12: Yellow area on audiogram indicates normal hearing levels



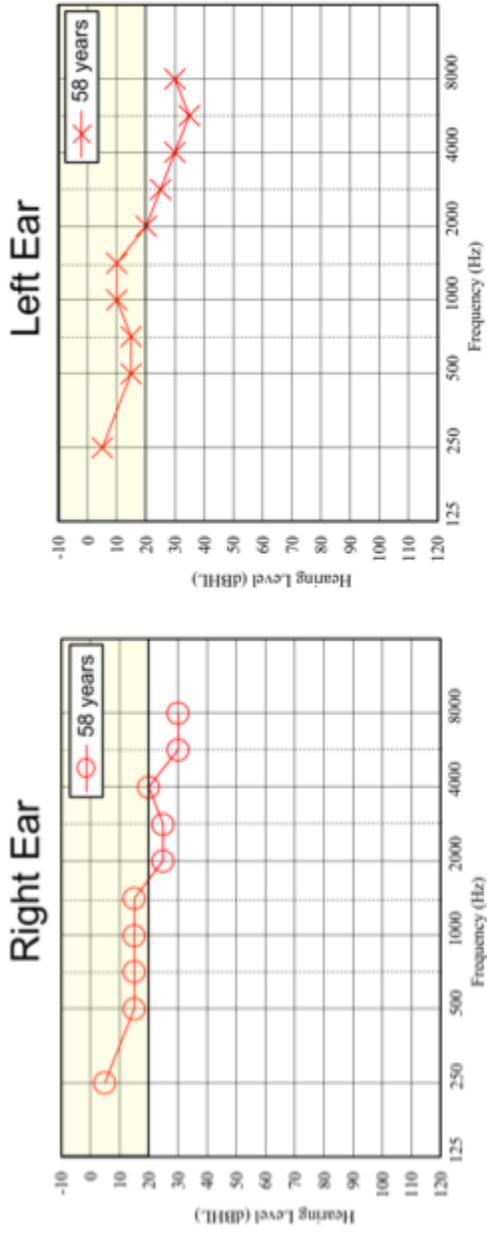
Family R2070 Pedigree indicating location of PID V-13



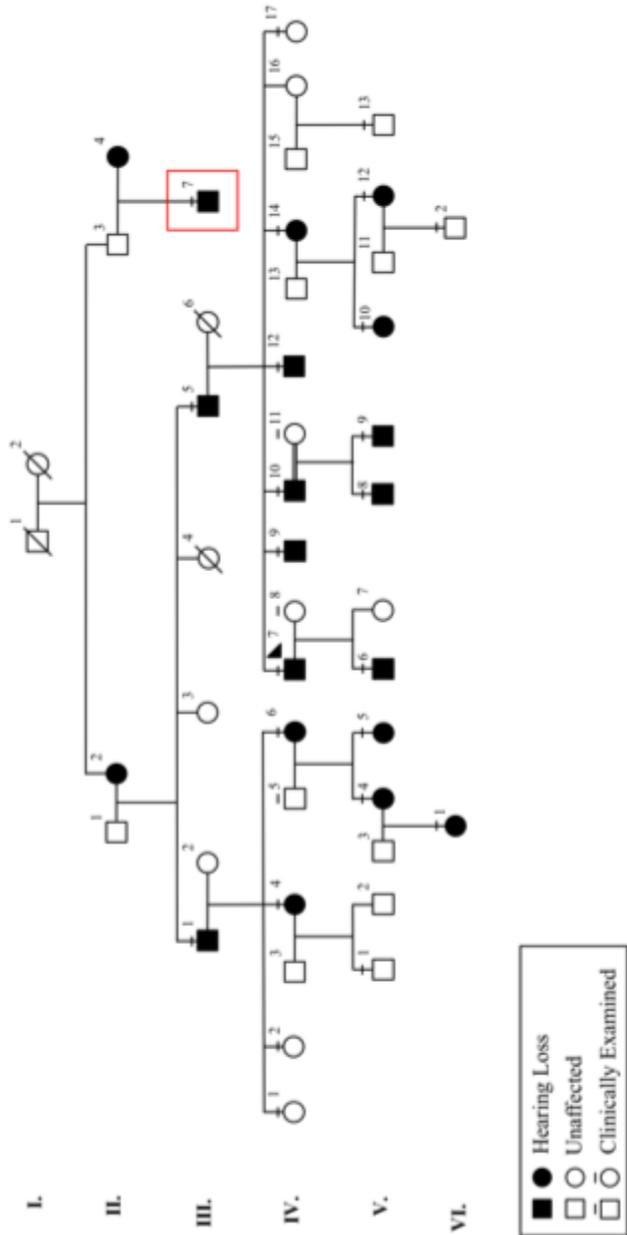
Audiological Data of PID V-13: Yellow area on audiogram indicates normal hearing levels



Audiological Data of PID IV-17: Yellow area on audiogram indicates normal hearing levels

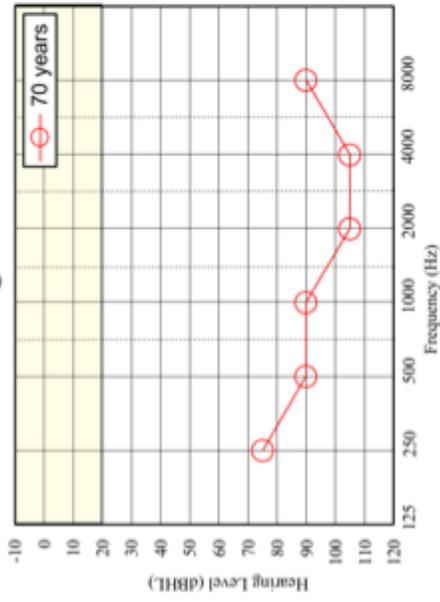


Family R2070 Pedigree indicating location of PID III-7

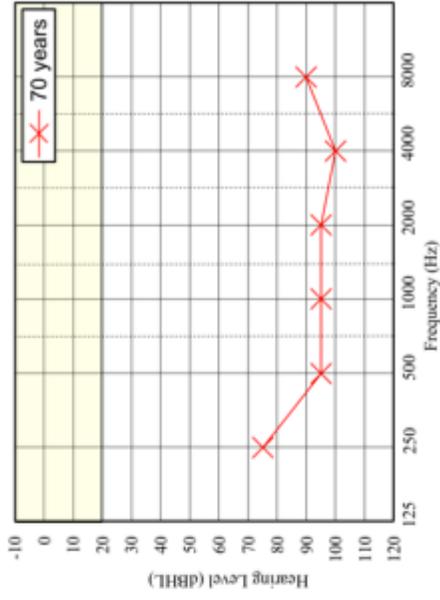


Audiological Data of PID III-7: Yellow area on audiogram indicates normal hearing levels

Right Ear



Left Ear



Appendix J–Audioprofiling Report

OVERVIEW

Age of onset:

- This is probably the most difficult area to understand in this family. The earliest audiograms on file for most family member already exhibit advanced stages of hearing loss.
- PID V-6 and PID V-8 (NO DNA) present very similar when comparing their diagnostic audiogram. The only difference is that PID V-8 presents with a milder phenotype at age 3, compared to a more advanced stage in PID-V-6 at age 1. The configuration of the audiogram is virtually identical.
- PID V-4 presents with low frequency hearing loss at 6 years of age.

Severity:

- There appears to be differences in hearing loss severity within the family.
 - In the sixth decade, PID IV-4 presents with good hearing thresholds in the lower frequencies, relative to other family members. This trend is also similar in PID V-5. Again, this is hard to ascertain due to the lack of early audiological data.
 - In addition, a milder type of hearing loss is present in PID V-10 and PID V-12. These women have a similar audiogram configuration, when compared to the right ear of their married-in consanguineous aunt.

Type of loss:

Sloping losses:

- These family members are the only ones that have enough longitudinal data to form an opinion on. This is just my interpretation.
 - PID IV-4, PID IV-6, PID V-5, PID IV-7, PID IV-9, PID V-9, PID V-12, PID V-10, PID V-8

Low frequency losses:

- PID V-4

Mid frequency losses:

- ID V-8 (NO DNA): starts off with normal hearing thresholds at 0.25 and 0.50 kHz, with mild hearing loss at 1 kHz. The 3, 4, and 6 kHz thresholds come back into the normal range, but deteriorate into the moderate range at 8kHz. Not 100% if this would be classified as a mid frequency loss, but the audiogram configuration does not seem to fit the overall picture of what is going on in the other family members.

Flat losses

- PID V-5 – similar to PID V-8 with a dip at 1 kHz frequency and more significant deterioration at the higher frequencies. Not 100% if this would be classified as a flat loss, but the audiogram configuration does not seem to fit the overall picture of what is going on in the other family members.

Symmetry:

- These family members had differences between their right and left ears. Unsure if this is clinically relevant.
 - PID V-5, PID IV-12, PID IV-14, PID IV-11, PID IV-7

Future Ascertainment:

- On the left side of the pedigree, there are two men with apparent hearing loss. I think it would be beneficial to determine their clinical and genetic status.
- PID V-4 has an affected daughter. She would provide a lot of insight with regards to understanding the genetic etiology of hearing loss in this family. Consenting her daughter is of particular importance because PID V-4 is actually the family member that presents with low frequency hearing loss.
- All family members that have no DNA, but have audiological data would be nice to have, especially PID V-6 as they are unaffected.

DESCRIPTION OF AUDIOLOGICAL DATA

PID III-1

- Onset can't be determined, as we only have a single audiogram on file. At 76 years of age, subject (male) has profound, flat hearing loss on all frequencies.

PID IV-1

- Subject (female) has normal hearing thresholds at 36 years of age.

PIC IV-2

- Subject (female) has normal hearing thresholds at 55 years of age.

PID IV-4

- Exact onset can't be determined. However, we do have several audiograms on file from the 4th, 5th and 6th decade.
- Subject (female) has a mild, sloping to profound loss in the 4th decade. By the 6th decade, the lower frequencies exhibit further deterioration in the right ear, as she has a

moderate sloping to profound hearing loss. The left ear actually has normal hearing thresholds sloping to profound in the 4th decade and progresses to a mild sloping to profound configuration by the 6th decade.

PID V-1

- Subject (male) has normal hearing thresholds at 24 years of age.

PID VI-6

- Exact onset can't be determined. However, we do have several audiograms on file from the 4th, 5th and 6th decade.
- Subject (female) has a mild, sloping to profound loss in the 3rd and 4th decade, progressing to a moderate, sloping to profound loss in the 5th and 6th decade.

PID IV-5

- Subject (female) has normal hearing with a noise notch affecting the 4 & 6 kHz frequencies in the right ear and the 4, 6 and 8 kHz frequencies in the left ear.

PID V-4

- Onset can't be determined, but we have audio data from the first decade. Subject presents with moderate, sloping to normal low frequency hearing loss at the age of 6. Further deterioration at all frequencies is seen at age 9, leading a profound loss at all frequencies in the third decade.

PID V-5

- Onset can't be determined, but we have audio data from the first decade. Subject presents with moderate, sloping to normal low frequency hearing loss at the age of 6.

Further deterioration at all frequencies is seen at age 9, leading a profound loss at all frequencies in the third decade.

PID IV-7 - Proband

- Onset can't be determined. We have audio data from the second, third and fifth decade. In the right ear, the proband presents with mild sloping to profound hearing loss in the second and third decade. In the fifth decade, the hearing phenotype progresses to a moderate sloping to profound loss. In the left ear, the proband presents with a mild sloping to profound hearing loss in the second decade. However, there appears to be a lack of symmetry between the right and left ears in the third and fifth decade, as the subject progresses to a moderately severe/profound loss at all frequencies.

PID V-6

- Hearing loss in this subject is likely congenital, as we have an audiogram from at one year of age, presenting with a flat, mild/moderate loss in both ears. The exception is the 8kHz hearing threshold, which is in the profound range. At age 5, subject presents with a mild loss at 0.25 and 0.50 kHz, dipping down into moderate at 1 kHz and back into mild thresholds at the 2, 3 and 4 kHz frequencies. At age 7 and 9, subject presents with moderate hearing loss at the lower frequencies, sloping to moderately severe/profound in the mid frequencies and back up into the moderate/moderately severe ranges in the high frequencies. At ages 13 and 16, subject presents with moderately severe hearing thresholds, sloping to profound in the mid frequencies. Higher frequency hearing thresholds come back into the moderately severe range at age 13 and exhibit further deterioration into the profound range by 16 years of age.

PID V-7 (NO DNA)

- Subject has normal hearing thresholds.

PID IV-9

- Onset can't be determined. We have audio data from the third, fourth and sixth decade. In the third and fourth decade, subject presents with mild sloping to profound hearing loss in the low and high frequencies, respectively. This hearing loss progresses to moderately severe sloping to profound in the sixth decade.

PID IV-10

- Onset can't be determined. We only have a single audiogram for this family member. At 55 years of age, subject presents with moderately severe hearing thresholds at 0.25kHz and profound thresholds at other frequencies.

PID IV-11

- Onset can't be determined. We only have a single audiogram for this family member. At 53 years of age, subject presents unilateral hearing loss. In the right ear, subject has a 20 dB hearing threshold at 0.25 kHz, which slopes into the moderately severe range at higher frequencies. Given that we are dealing with bi-lateral hearing loss, this phenotype most likely has nothing to do with the phenotype in question.

PID V-8

- Onset can't be determined. We have audiograms from the first and second decade. This family member does not have a DNA sample. At 3 years of age, subject presents with normal hearing thresholds at 0.25 and 0.50 kHz, with mild hearing loss at 1 kHz. The 3, 4, and 6 kHz thresholds come back into the normal range, but deteriorate into the

moderate range at 8kHz. At 8 years of age, the same audiogram configuration is present; however, the 4 and 6 kHz frequencies drop into the mild hearing loss range in the right ear. Relative to the right ear, the left ear seems to have a more significant loss at the 3, 4, 6, and 8 kHz frequencies, as it slopes from the normal to profound range. At 13, the 2, 3, 4, 6 and 8 kHz a marked deterioration and begins to resemble a mild, sloping to moderately severe (right ear) and mild sloping to profound (left ear) configuration.

PID V-9

- Onset can't be determined. We only have a single audiogram for this family member. At 16 years of age, subject presents with normal low frequency (0.25 and 0.50 kHz) hearing thresholds that slope into the profound range. Even though it's very unlikely that this hearing loss is due to *CLDN14*, this audiogram reminded of the R2010 family.

PID V-12

- Onset can't be determined. We have audiograms from the third, fourth and fifth decade. At 21 years of age, the right ear exhibits profound hearing threshold at all frequencies, while the left ear 0.25 kHz and 0.50 kHz hearing thresholds are in the moderate and moderately severe ranges, respectively. In the fourth and fifth decade, all hearing thresholds are in the profound range.

PID IV-14

- Onset can't be determined. We have audiograms from the fourth, fifth and sixth decade. In the fourth decade, this family member presents with a moderate, sloping to profound hearing loss. While this does slope, it doesn't look like hearing thresholds deteriorate as aggressively from low to high frequencies, when compared to other family members.

While all left ear hearing threshold are in the profound range in the fifth and sixth decade, lower frequencies in the right ear are in the moderately severe range in the fifth decade.

PID V-12

- Onset can't be determined. We only have a single audiogram for this family member. At 27 years of age, subject presents with normal, sloping to moderately severe hearing loss. This appears to be a more mild bilateral loss. I think it's interesting to point out that her audiogram configuration resembles the right ear of 2070.0028 (AM15-17), a married-in, consanguineous family member with unilateral hearing loss.

PID V-10

- Onset can't be determined. We have audiograms from the second and third decade. At 15 years of age, subject presents normal/borderline hearing thresholds from 0.25 to 4 kHz, while the 6 and 8 kHz frequencies are in the moderate and moderately severe ranges in the right and left ear, respectively. At 26, her hearing thresholds are normal/borderline from 0.25 to 1.5 kHz, but further deterioration is apparent at the 2, 3, 4, 6 and 8 kHz frequencies, sloping from mild to moderate hearing loss. Perhaps this is an earlier stage of her sister, 2070.A017 (KR15-14)?

PID V-13

- Subject has normal hearing thresholds at 19 years old.

PID IV-17

- Subject has mild/borderline hearing loss at the 2, 3, 4, 6 and 8 kHz frequencies. Given that she is 58 years old at the time of testing, this is most likely a case of presbycusis.

Despite the likelihood of presbycusis, an important question is yet to be determined “Is presbycusis an age-related deterioration, or is it due to complex inherited genetics?” –

Food for thought.

PID III-7

- Onset can't be determined, as we only have a single audiogram on file. At 70 years of age, subject has profound, flat hearing loss on all frequencies.

Linkage analysis in deafness pedigree 2100

Data providers: Terry Lynn Young (tlyoung@mun.ca), Nelly Abdelfatah (nellya@mun.ca), Lance Doucette (lanced@mun.ca), Tammy Benteau (tbenteau@mun.ca)

Report prepared by: Nicole Roslin (nroslin@sickkids.ca)

Last modified: 9 July 2013

Data summary

Chip: Human610-Quadv1_B.bpm

Number of markers passing QC: 573,708

Number of markers used in linkage analysis: 17,407

Family: 2100 (5 genotyped samples)

Introduction

Five individuals from family 2100, from Newfoundland and Labrador, were genotyped using the Human610-Quad chip from Illumina. The genotypes were of excellent quality, and were consistent with the pedigree structure and genders provided for this family; see the document "*Analysis of SNP genotypes from the Illumina Human610-Quad panel in AMGGI Deafness pedigrees*," dated 22 May 2013, for details.

Samples, pedigree and phenotype

Family 2100 is a multigenerational pedigree with two individuals with confirmed deafness (Figure 1), and one individual with unconfirmed deafness. The family is known to be consanguineous. The hearing loss was early onset, and slowly progressing. The two individuals with confirmed hearing loss will be considered to be affected for all analyses. Two other individuals do not have hearing loss, and will be unaffected. The individual with hearing loss that was not confirmed by audiogram analysis, DW08-63, will be considered to be affected in one analysis, and unknown in a second analysis, to determine how sensitive the linkage results are to this sample.

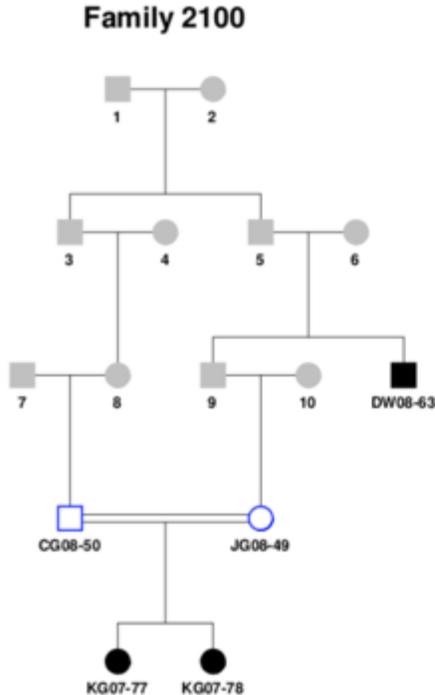


Figure 1. Family 2100. Black symbols are affected, grey symbols have unknown phenotype, and white symbols outlined in blue are unaffected for hearing loss. The deafness for DW08-63 has not been confirmed by audiogram analysis. Symbols with hyphenated labels have been genotyped.

The genetic model to be used in the linkage analysis requires the specification of a disease allele frequency. In linkage analysis, this frequency is used as an estimate of the probability that a founder carries the disease allele. Generally, the frequency is derived from the prevalence of the disease in the population. However, statistics are not available regarding the prevalence of hearing loss in Newfoundland and Labrador. Statistics from the Canadian population indicates hearing loss is found in approximately 5% of the adult population. However, there are many different types of hearing loss, with different causes, so the prevalence of the particular type of hearing loss observed in this family will be much less frequent. Therefore, a range of disease allele frequencies was used, with an upper bound corresponding to 5% prevalence.

Simulations

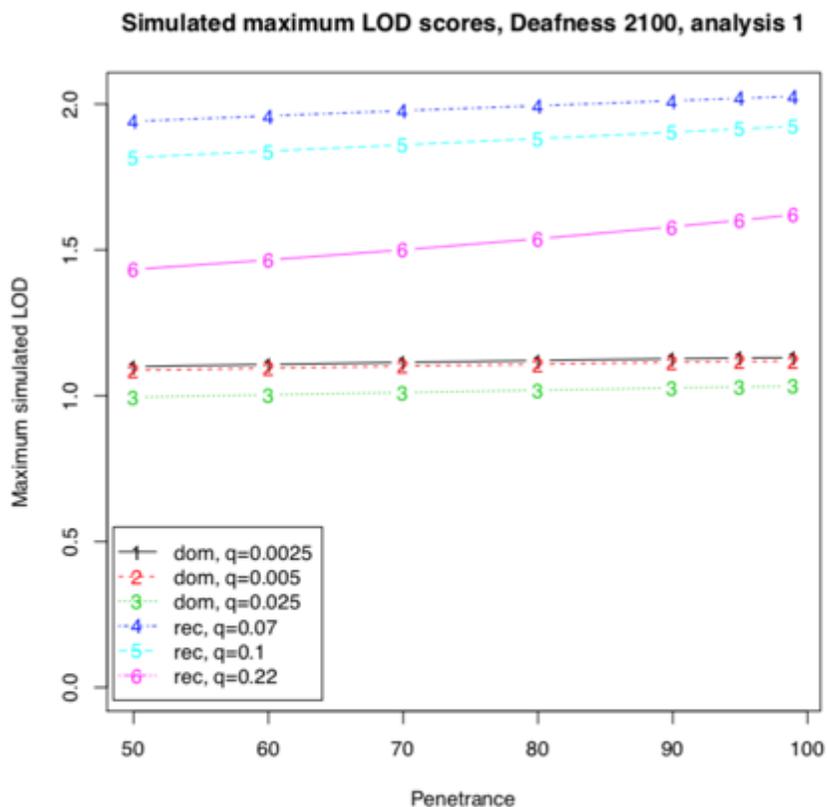
Simulations were performed under the alternate hypothesis of linkage to determine the maximum possible LOD score for the pedigree, under ideal conditions. These conditions included a disease-causing locus that was perfectly linked to the marker locus, a marker locus that provided perfect information regarding the segregation of alleles in the pedigree, and an analysis model that was correctly specified. It was also assumed that the pedigree structure and affection status were correct. Pedigrees were simulated under dominant and recessive models with a range of disease allele frequencies and penetrances using SLINK 3.02 (Schaffer, et al., 2011),

and analyzed under the same model using Merlin 1.1.2 (Abecasis, et al., 2002). The phenocopy rate was fixed at 0.2% for all models. The maximum LOD score obtained from the analysis of 1000 simulated pedigrees was declared the maximum LOD score for any particular model.

For the dominant model, disease allele frequencies of 0.025, 0.005 and 0.0025 were used, corresponding to prevalences of 5%, 1% and 0.5%, respectively, under a fully penetrant model. For the recessive model, the disease allele frequencies used were 0.22, 0.1 and 0.07 (corresponding to the same prevalences). For both dominant and recessive models, simulations were performed using penetrances of 50, 60, 70, 80, 90, 95 and 99%.

For Analysis 1 (DW08-63 affected), the maximum simulated LOD score was 2.03, under a recessive model with a disease allele frequency of 0.07. The maximum LOD scores had a similar profile when a disease allele frequency of 0.1 was used (Figure 2), although the maximum LOD score dropped to 1.92. Recessive models had consistently higher LOD scores than dominant models.

For Analysis 2, when DW08-63 was changed to have unknown affection status, the maximum simulated LOD score was 1.68 under a recessive model with a disease allele frequency of 0.07 and 99% penetrance (results not shown).



/Users/nrosin/Projects/Ybung2013/Deafness/15Sim2100_1/sim.pdf

Tue Jun 18 12:46:09 2013

Figure 2. Simulated maximum LOD scores under the alternative hypothesis of linkage, under a range of genetic models, for the pedigree shown in Figure 1. dom = dominant, rec = recessive, q = disease allele frequency.

Linkage analysis

SNP filtering

Starting with a set of 572,708 high quality SNPs from the Illumina 610Quad genotyping chip, markers were removed in order to end up with a smaller set suitable for linkage analysis. First, SNPs on chromosomes other than 1 to 22 and X were removed (mitochondrial, Y, XY and unplaced markers). Next, markers that had alleles ambiguous for strand information (A/T and G/C variants) were removed, in order to facilitate strand matching with HapMap data (Altshuler, et al., 2010). Markers that were monomorphic in all genotyped samples were also removed. At this point, approximately 550,000 markers remained. Markers that were present in both this set and the CEU and TSI samples (of western European origin) from HapMap were retained, matching on marker name. Out of this set of about 525,000 markers, SNPs with minor allele frequency >0.45 were kept, which should select for markers with the most linkage information. In order to avoid inconsistencies between the genotypes and the genetic map, only markers with unique locations on the genetic map were extracted. Arbitrarily, for a group of markers at the same genetic position, the marker with the lowest physical position was retained. Since linkage disequilibrium (LD) between the SNPs can arbitrarily inflate multipoint LOD scores (Huang, et al., 2004), the set of markers was pruned for LD between pairs of markers: only SNPs with pairwise $r^2 < 0.1$ were kept. These steps resulted in a set of 17,407 SNPs across the genome that was suitable for linkage analysis. For this set, the average intermarker distance was 0.26 cM, or 197 kb.

Multipoint linkage analysis

Linkage analysis was performed on the pedigree in Figure 1 using Merlin 1.1.2 (Abecasis, et al., 2002) using the observed genotypes. Multipoint linkage analysis was performed under a recessive model with a disease allele frequency (q) of 0.07 and penetrances of 0.2, 0.2 and 99% for 0, 1, 2 copies of the disease-causing allele, respectively. These parameters correspond to a disease prevalence of approximately 0.5%, and formed the model with the largest simulated maximum LOD score for this family (out of the models tested).

Analysis 1

Analysis 1 consisted of the phenotypes shown in Figure 1. The largest observed LOD score was 2.03, on chromosome 4 (Table 1, Figure 3, black line). Two other regions, on chromosomes 1 and 6, had LOD scores nearly as high (2.02 each). A total of 6 regions had LOD >1 . The linked regions spanned approximately 45 cM; this represents 1.25% of the genome, assuming a total genetic length of 36M. The average information content across the genome was 0.58.

Analysis 2

Analysis 2 used the same genetic model as Analysis 1. However, the affection status for DW08-63 was changed from affected to unknown, since this individual did not have his phenotype confirmed by audiogram analysis. The largest observed LOD score was 1.68, on chromosomes 3 and 5 (Table 1, Figure 3, red line). LOD scores >1 were also seen on chromosomes 4, 6 11, 15 and 20 (a total of 7 regions). These linked regions spanned approximately 54 cM, representing 1.5% of the genome.

Comparison

The two analyses differed only in the affection status of one individual, DW08-63. However, this individual appeared to influence the results of the linkage analysis in certain regions. For example, a large peak was observed on chromosome 1 when this individual was affected, but the peak disappeared when this person was unknown. The peaks on chromosome 4 and 6 remained large, relative to the simulated maximum LOD scores, in either analyses.

Table 1. Regions with maximum observed LOD scores >1 in either analysis, under a recessive model with 99% penetrance and a disease allele frequency of 0.07. A region was defined as the 1-LOD support interval around the maximum. LOD score percentages indicate the percentage of the simulated maximum that was achieved by each region. The simulated maxima are 2.03 for Analysis 1 and 1.68 for Analysis 2.

Chr	Analysis 1			Analysis 2		
	Max LOD (%)	Start SNP	End SNP	Max LOD (%)	Start SNP	End SNP
1	2.02 (99)	rs10489631	rs591540	0.58 (34)	rs3933251	rs591540
3	1.01 (50)	rs1400207	rs13084851	1.68 (100)	rs1400207	rs13084851
4	2.03 (100)	rs7671597	rs2035906	1.68 (100)	rs7671597	rs2035906
5	1.01 (50)	rs355412	rs257239	1.68 (100)	rs12110158	rs257239
6	2.02 (200)	rs1322633	rs1490388	1.67 (99)	rs1322633	rs1490388
11	0.98 (48)	rs1320211	rs224619	1.64 (98)	rs1320211	rs10833818
15	1.01 (50)	rs937302	rs11070349	1.68 (100)	rs937302	rs11070349
20	0.74 (37)	rs237417	rs4140471	1.41 (84)	rs237417	rs4140471

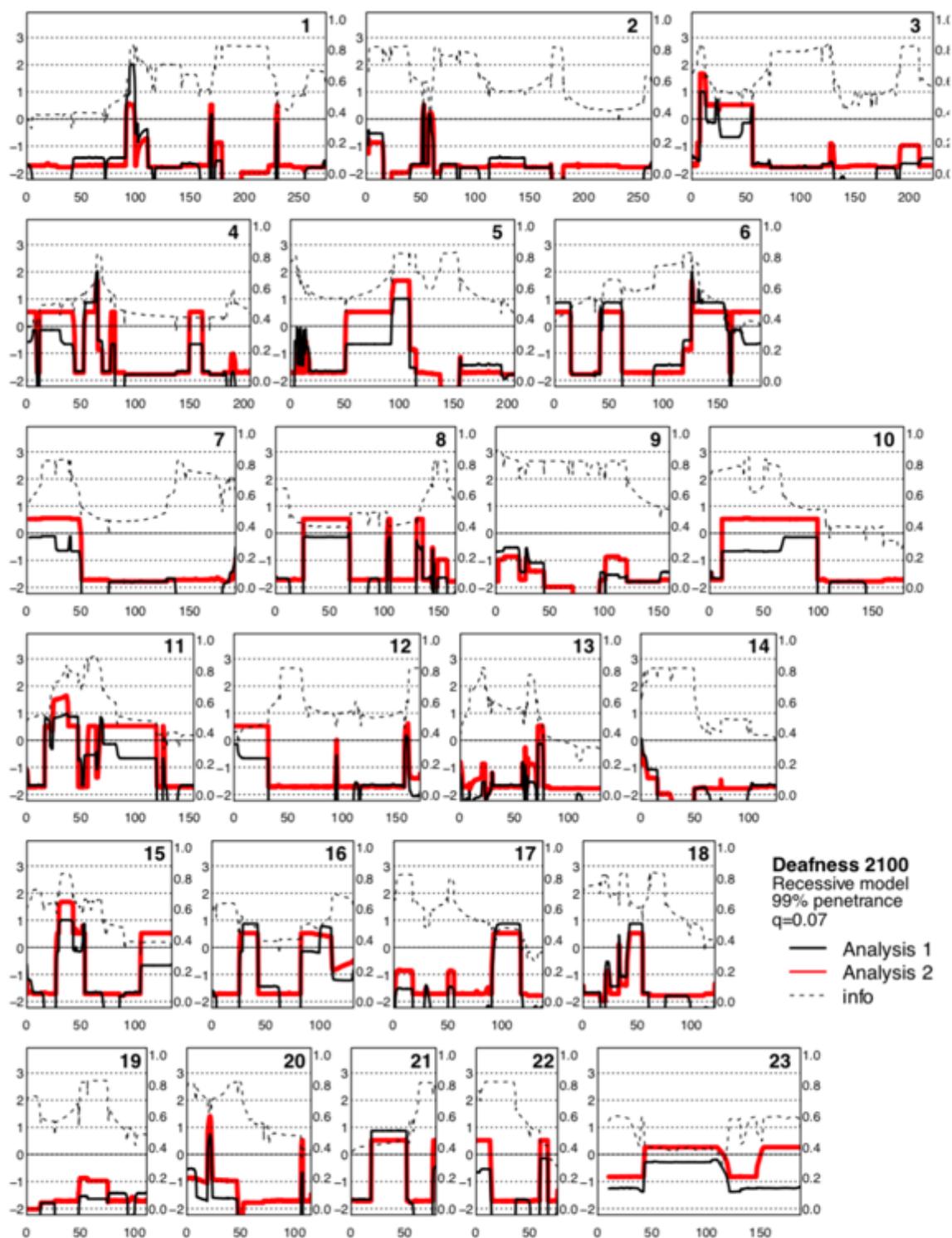


Figure 3. LOD scores (solid lines) and information content (dashed line) for the family shown in Figure 1, under a recessive model with 99% penetrance and $q=0.07$ for Analysis 1 (black line) and Analysis 2 (red line). LOD scores use the vertical axis on the left, and information content uses the vertical axis on the right. Only LOD >-2 are shown.

Summary

Multipoint linkage analysis was performed on family 2100 under a recessive model with a disease allele frequency of 0.07 and a penetrance of 99%, corresponding to a prevalence of 0.5%. Analysis 1 used affection status as shown in Figure 1, while in Analysis 2 the affection status of DW08-63 was changed to missing. In Analysis 1, 6 regions had LOD >1, spanning 45 cM. In Analysis 2, 7 regions had LOD >1, spanning 59 cM. Five of these regions overlapped the two analyses, and two of the regions had LOD scores approximately equal to the simulated maxima in both analyses.

Since DW08-63 appears to influence the linkage results, it would be ideal if the deafness in this person could be confirmed or denied by audiogram analysis. Otherwise, a judgment would have to be made as to how likely it is that this person is affected with the same type of deafness as the rest of the affected family members. Also, the two regions with the highest LOD scores in both analyses could be prioritized over the other linked regions for further study.

References

Abecasis, G.R., *et al.* (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees, *Nature Genetics*, **30**, 97-101.

Altshuler, D.M., *et al.* (2010) Integrating common and rare genetic variation in diverse human populations, *Nature*, **467**, 52-58.

Huang, Q., Shete, S. and Amos, C.I. (2004) Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis, *American Journal of Human Genetics*, **75**, 1106-1112.

Schaffer, A.A., *et al.* (2011) Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees, *Human Heredity*, **71**, 126-134.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, *17*(5), 405-424. doi:10.1038/gim.2015.30