

# **The Genetic Aetiology of Otosclerosis in the Population of Newfoundland and Labrador**

By

© Nelly Abdelfatah

A thesis submitted to the  
School of Graduate Studies  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Discipline of Genetics, Faculty of Medicine  
Memorial University of Newfoundland

Oct, 2014

St. John's

Newfoundland and Labrador

## **Abstract**

### **Background**

Otosclerosis is a common form of conductive and mixed hearing loss in Caucasian populations, with an estimated prevalence of 0.3-0.4%. Since 1998, eight loci have been mapped to otosclerosis in families with apparent autosomal dominant (AD) otosclerosis but none of the causative genes have been identified.

### **Objective**

As no otosclerosis gene has yet been identified, the main objective of this thesis was to identify otosclerosis-disease causing genes by studying Newfoundland (NL) families.

### **Methods**

Families with familial otosclerosis were identified and characterized clinically. Those which fit the diagnostic criteria for otosclerosis were recruited for this study. Molecular genetic analyses of these families were carried out by genotyping, haplotyping, Sanger sequencing of candidate genes in linked regions and exome sequencing.

### **Results**

One Family (2081) was solved through identification of a pathogenic variant (*FOXL1*c.976\_990hetdel) in the *FOXL1* gene at chromosome (Chr) 16q that was present in all affected individuals. The 15 base pair (bp) deletion was also identified in a second family from Ontario (ON) and the possible pathways involving *FOXL1* in the pathogenesis of otosclerosis were suggested. In the second otosclerosis family, three

candidate variants were identified through exome sequencing of the candidate regions under a dominant model.

## **Conclusion**

I have identified the first otosclerosis gene, *FOXL1*, a transcription factor involved in the disease pathogenicity. I also identified three possible candidate mutations for a second otosclerosis family. This finding will have a major impact on molecular genetic studies of other otosclerosis families and it will allow for genetic counselling and the possibility for gene therapies in the future.

## **Acknowledgments**

First, I would like to take this opportunity to thank my supervisor Dr. Terry Lynn Young for her generosity, support, guidance, teaching and patience. Words are not enough to thank her. When I came to her lab, I had a little experience in experimental design, critical thinking and scientific writing. Dr. Terry Young taught me how to be a great scientist. Thank you for believing in me and giving me the opportunity of my life.

Thanks to my supervisory committee, Drs Kathy Hodgkinson, Jane Green and Ban Younghusband for their encouragement, guidance and helpful advice throughout the completion of my project.

Thanks to The Young Lab, for being great, supportive and amazing co-workers. Special thanks to Dante Galutira and Jim Houston for your help and assistance, and for your patience and generosity. Thank you Tammy Benteau and Catherine Street for your support. Thank you Nancy Merner and Annika Haywood for being valuable members of the lab, and for your help and support. I spent a great and enjoyable time working with Lance Doucette and I wish that we could continue working with each other. Thank you David McComiskey, Jessica Squires, Cindy Penney and Andrew Bullen for being excellent peers and making learning fun. Special thanks for Jim Houston, Dante Galutira and Tammy Benteau for editing my thesis.

I would like to thank Genome Canada (Atlantic Medical Genetics and Genomics Initiative), CIHR (Canadian Institutes of Health Research), RDC NL (Research &

Development Corporation of Newfoundland and Labrador) and Memorial University Graduate Studies for funding. Also I would like to thank CIHR, RDC NL for my fellowship.

I would like to thank the Journals: European Journal of Human Genetics and Human Mutation for publishing my two papers.

I dedicate this work to my lovely Mom for her support, love and concern from abroad and in memory of my beloved Dad. Thank you to my husband Ahmed Mostafa for his continued support, help and encouragement. Without your support, I would not be able to finish my PhD. Thank you to my lovely adorable kids, Marwan and Merna, for your understanding that Mom is busy; for your laughter, love, support and encouragement which gave me the energy to continue.

## Table of Contents

<b>The Genetic Aetiology of Otosclerosis in the Population of Newfoundland and Labrador</b> .....	I
<b>Abstract</b> .....	II
<b>Acknowledgments</b> .....	IV
<b>List of Tables</b> .....	X
<b>List of Figures</b> .....	XI
<b>Chapter 1: General introduction</b> .....	1
1.1 Purpose of the study and significance.....	1
1.2 Anatomy and function of the ear.....	2
1.3 How sound is transmitted.....	4
1.4 Hearing loss.....	5
1.4.1 Classification of hearing loss.....	5
1.5 Gene discovery in non-syndromic hearing loss using NL families .....	12
1.5.1 Ascertainment of families.....	12
1.5.2 Complications of family ascertainment.....	15
1.5.2.1 Penetrance.....	15
1.5.2.2 Variable expression.....	16
1.5.2.3 Phenocopy.....	17
1.5.3 Genetic analysis of families with hearing loss.....	17
1.5.3.1 Exclusion of known loci or genes.....	18
1.5.3.2 Candidate gene and genome-wide association approaches.....	18
1.5.3.3 Genome-wide linkage analysis.....	20
1.5.3.4 Positional candidate gene sequencing.....	25
1.5.3.5 Mutation identification and validation .....	26
1.5.3.6 Next generation sequencing.....	30
1.6 Otosclerosis.....	32
1.6.1 Overview.....	32
1.6.2 Bone remodeling and otic capsule .....	35
1.6.3 History and stages of otosclerosis.....	38
1.6.4 Diagnosis of otosclerosis.....	40
1.6.5 Genetic causes of otosclerosis.....	44
1.7 Founder population.....	51
1.8 The NL population .....	51
1.8.1 Gene discovery in the NL population.....	52

1.9	Hypothesis and objectives.....	53
1.10	Co-authorship statement.....	54
<b>Chapter 2: Test for linkage of NL Families to published <i>OTSC</i> loci and associated candidate genes.....</b>		<b>55</b>
2.1	Introduction.....	56
2.2	Subjects and methods.....	56
2.2.1	Working definition of clinical otosclerosis .....	56
2.2.2	Study population.....	58
2.2.3	Genotypes and haplotypes for Families 2081 and 2114.....	58
2.2.4	Sequencing genes within the candidate region.....	61
2.2.5	Functional assay.....	67
2.2.5.1	Reverse transcriptase (RT) PCR.....	67
2.2.6	Two-point linkage analysis.....	67
2.3	Results.....	69
2.3.1	Pedigree structure and clinical analysis .....	69
2.3.1.1	Pedigree structure of Family 2081.....	69
2.3.1.2	Clinical analysis of Family 2081 .....	72
2.3.1.3	Pedigree structure of Family 2114.....	76
2.3.1.4	Clinical analysis of Family 2114 .....	78
2.3.1.5	Proband of the Family 2194.....	80
2.3.1.6	Proband of the Family 2197.....	80
2.3.1.7	Proband of the Family 2203.....	80
2.3.1.8	Proband of the Family 2066.....	81
2.3.1.9	Proband of the Family 2126.....	81
2.3.1.10	Proband of the Family 2200 .....	81
2.3.1.11	Proband of the Family 2209 .....	82
2.3.2	Test for linkage to published otosclerosis loci.....	84
2.3.2.1	Exclusion of Family 2081 from previously mapped loci and genes .....	84
2.3.2.2	Exclusion of Family 2114 from previously mapped loci and genes .....	108
2.3.3	Fine mapping of <i>OTSC4</i> locus in Family 2081 .....	129
2.3.4	Shared haplotype across four unrelated otosclerosis probands .....	132
2.3.5	Gene screening of functional candidate genes .....	134
2.3.5.1	Variant analysis.....	134
2.3.5.2	Functional assessment of the <i>ZFHX3</i> c.10557_10558insGGC .....	136
2.3.5.3	Segregation analysis.....	138
2.3.5.4	Allele frequencies of <i>ZFHX3</i> variants in NL controls .....	138
2.3.6	Recruitment of a new family member and linkage exclusion of Family 2081 from <i>OTSC4</i> .....	147

2.3.7	Two point linkage analysis .....	149
2.4	Discussion.....	152
2.5	Conclusion.....	155
<b>Chapter 3: Fifteen bp deletion in the <i>FOXL1</i> gene causes familial otosclerosis in one from NL and one from ON .....</b>		<b>156</b>
3.1	Introduction.....	157
3.2	Subjects and methods.....	157
3.2.1	Study population and clinical analysis .....	157
3.2.2	Genotypes and haplotype analysis of Family 2081 .....	157
3.2.3	Sequencing genes within the candidate region .....	158
3.2.3.1	Mutation screening.....	158
3.2.4	Whole exome sequencing of five family members of Family 2081.....	159
3.2.4.1	Mutation screening panel .....	159
3.3	Results.....	160
3.3.1	Pedigree structure and clinical analysis .....	160
3.3.2	Genetic analysis.....	160
3.3.2.1	Family 2081 is linked to 9.7 Mb region downstream of <i>OTSC4</i> .....	160
3.3.2.2	Mutation screening of functional candidate genes by Sanger sequencing.....	164
3.3.2.2.1	Variant analysis.....	164
3.3.2.3	Mutation screening of functional candidate genes by whole exome sequencing.....	172
3.3.2.3.1	Variant analysis.....	172
3.3.2.4	Identification of the effect of the <i>FOXL1</i> in NL and ON families.....	178
3.3.3	Functional studies for the 15 bp <i>FOXL1</i> deletion.....	181
3.4	Discussion... ..	182
3.5	Conclusion.....	188
<b>Chapter 4: Family 2114: Genome-wide linkage analysis and exome sequencing.....</b>		<b>189</b>
4.1	Introduction .....	190
4.2	Subjects and methods .....	191
4.2.1	Study population and clinical analysis.....	191
4.2.2	Genome-wide linkage analysis.....	191
4.2.2.1	Genome-wide SNP typing.....	191
4.2.2.2.1	Theoretical Maximum LOD Scores.....	191
4.2.2.2.2	SNP filtration.....	192
4.2.2.2.3	Multipoint linkage analysis.....	192
4.2.3	Genotyping and haplotype analyses .....	193
4.2.4	Sequencing of positional candidate genes within the linked regions.....	193

4.2.4.1	Sanger sequencing of candidate genes .....	193
4.2.4.2	Whole exome sequencing of Family 2114.....	194
4.2.4.3	Variant analysis .....	194
4.3	Results.....	196
4.3.1	Pedigree structure and clinical analysis.....	196
4.3.2	Genome-wide linkage analysis.....	196
4.3.3	Genetic analysis of Family 2114 under recessive mode of inheritance ....	198
4.3.3.1	Haplotype segregation across Chr17 linked region .....	198
4.3.3.2	Sanger sequencing of positional candidate genes.....	201
4.3.3.3	Exome sequencing and targeted analysis of the Chr17 under recessive mode.....	206
4.3.3.3.1	Selection of negative controls for filtering NGS data at Chr17 .....	206
4.3.3.3.2	Variants analysis of Family 2114 under a recessive mode of inheritance .....	209
4.3.3.4	Genetic analysis of Family 2114 under dominant inheritance.....	212
4.4	Discussion.....	235
4.5	Conclusion.....	240
<b>Chapter 5: General discussion .....</b>		<b>241</b>
5.1	Implication of identifying the first gene for familial otosclerosis .....	242
5.2	A possible second gene for familial otosclerosis .....	244
5.3	Overcoming challenges identifying the genes associated with otosclerosis.....	247
5.3.1	Issues with clinical diagnosis.....	247
5.3.2	Recruitment limitations.....	248
5.3.3	Comprehensive traditional and new technologies.....	248
5.4	Future direction.....	250
5.4.1	Benefits to patients and their families .....	250
5.4.2	Benefits to biomedical science .....	242
5.4.3	Gene and drug therapies.....	253
<b>Bibliography .....</b>		<b>255</b>
<b>Appendices.....</b>		<b>266</b>

## List of Tables

<b>Table 1.1:</b> Types of hearing loss based on severity .....	10
<b>Table 1.2:</b> Types of hearing loss based on configuration.....	11
<b>Table 1.3:</b> Currently mapped AD otosclerosis loci and their respective genomic locations .....	45
<b>Table 2.1:</b> Microsatellite markers used for genotyping.....	56
<b>Table 2.2:</b> Shared alleles across <i>OTSC4</i> among NL otosclerosis probands.....	133
<b>Table 2.3:</b> Summary of allele frequency for variants identified in the <i>ZFHX3</i> that segregated with otosclerosis in Family 2081 .....	146
<b>Table 2.4:</b> Two point LOD score in Family 2081 .....	150
<b>Table 3.1:</b> Family 2081 affected members co-segregate a 9.7 Mb haplotype .....	163
<b>Table 3.2:</b> Allele frequencies of variants reside on the light green haplotype .....	168
<b>Table 3.3:</b> Allele frequencies of NGS variants preserved on the disease haplotype.....	176
<b>Table 3.4:</b> Shared ancestral haplotype between affected probands from NL and ON ....	178
<b>Table 3.5:</b> Variants identified in <i>FOXLI</i> in NL otosclerosis probands.....	181
<b>Table 4.1:</b> Regions with an observed LOD score >1 under a dominant model .....	197
<b>Table 4.2:</b> Boundaries, size and gene numbers of the five linked regions .....	199
<b>Table 4.3:</b> Sequencing variants identified in six out of 11 candidate genes examined in the Chr17 linked region. ....	204
<b>Table 4.4:</b> Sequencing variants detected in exomes of four affected siblings in Family 2114 across Chr17 under both recessive and dominate modes of inheritance	218
<b>Table 4.5:</b> Sequencing variants detected in exomes of four affected siblings in Family 2114 across Chr7.....	220
<b>Table 4.6:</b> Sequencing variants detected in exomes of four affected siblings in Family 2114 across Chr10.....	222
<b>Table 4.7:</b> Sequencing variants detected in the exomes of four affected siblings in Family 2114 across Chr16.....	225

## List of Figures

<b>Figure 1.1:</b> Structure of the ear .....	3
<b>Figure 1.2:</b> Pure tone audiogram showing air conduction .....	7
<b>Figure 1.3:</b> Pure tone audiogram showing SNHL.....	8
<b>Figure 1.4:</b> Pure tone audiogram showing mixed hearing loss .....	9
<b>Figure 1.5:</b> Degree of informativeness of markers within a pedigree.....	22
<b>Figure 1.6:</b> Pathways involved in bone remodeling.....	37
<b>Figure 1.7:</b> Histological findings in different stages of otosclerosis.....	38
<b>Figure 1.8:</b> Audiogram profile of an otosclerosis patient .....	43
<b>Figure 2.1:</b> Flowchart showing filtration steps of variants detected in affected members under dominant mode of inheritance .....	65
<b>Figure 2.2:</b> Partial pedigree of NL Family 2081 segregating AD otosclerosis.....	71
<b>Figure 2.3:</b> Serial audiograms for the proband of Family 2081 .....	75
<b>Figure 2.4:</b> Partial pedigree of NL Family 2114 segregating otosclerosis.....	77
<b>Figure 2.5:</b> Serial audiograms for the proband of Family 2114.....	78
<b>Figure 2.6:</b> Pedigrees of Families 2149, 2203, 2066, 2197, 2126, 2209 and 2200.....	83
<b>Figure 2.7:</b> Partial pedigree of Family 2081 showing <i>OTSC1</i> haplotypes.....	86
<b>Figure 2.8:</b> Partial pedigree of Family 2081 showing <i>OTSC2</i> haplotypes.....	88
<b>Figure 2.9:</b> Partial pedigree of Family 2081 showing <i>OTSC3</i> haplotypes.....	90
<b>Figure 2.10:</b> Partial pedigree of Family 2081 showing <i>OTSC4</i> haplotypes.....	92
<b>Figure 2.11:</b> Partial pedigree of Family 2081 showing <i>OTSC5</i> haplotypes.....	94
<b>Figure 2.12:</b> Partial pedigree of Family 2081 showing <i>OTSC7</i> haplotypes.....	96
<b>Figure 2.13:</b> Partial pedigree of Family 2081 showing <i>OTSC8</i> haplotypes.....	98
<b>Figure 2.14:</b> Partial pedigree of Family 2081 showing <i>OTSC10</i> haplotypes.....	101
<b>Figure 2.15:</b> Partial pedigree of Family 2081 showing <i>COLIA1</i> haplotypes .....	103
<b>Figure 2.16:</b> Partial pedigree of Family 2081 showing <i>COLIA2</i> haplotypes .....	105
<b>Figure 2.17:</b> Partial pedigree of Family 2081 showing <i>NOG</i> gene haplotypes .....	107
<b>Figure 2.18:</b> Partial pedigree of Family 2114 showing <i>OTSC1</i> haplotypes.....	110
<b>Figure 2.19:</b> Partial pedigree of Family 2114 showing <i>OTSC2</i> haplotypes.....	112
<b>Figure 2.20:</b> Partial pedigree of Family 2114 showing <i>OTSC3</i> haplotypes.....	114
<i>OTSC4</i> .....	115
<b>Figure 2.21:</b> Partial pedigree of Family 2114 showing <i>OTSC4</i> haplotypes.....	116
<b>Figure 2.22:</b> Partial pedigree of Family 2114 showing <i>OTSC5</i> haplotypes.....	118
<b>Figure 2.23:</b> Partial pedigree of Family 2114 showing <i>OTSC7</i> haplotypes.....	120
<b>Figure 2.24:</b> Partial pedigree of Family 2114 showing <i>OTSC8</i> haplotypes.....	122
<b>Figure 2.25:</b> Partial pedigree of Family 2114 showing <i>COLIA1</i> haplotypes .....	124
<b>Figure 2.26:</b> Partial pedigree of Family 2114 showing <i>COLIA2</i> haplotypes .....	126

<b>Figure 2.27:</b> Partial pedigree of Family 2114 showing <i>NOG</i> gene haplotypes .....	128
<b>Figure 2.28:</b> Partial pedigree of NL Family 2081 segregates <i>OTSC4</i> haplotype.....	129
<b>Figure 2.29:</b> Flowchart showing filtration steps applied to identified variants at the minimized <i>OTSC4</i> .....	134
<b>Figure 2.30:</b> Electropherogram showing heterozygous missense mutation <i>ZFHX3</i> c.10831C>T, p.H3611HY in PID III-8.....	139
<b>Figure 2.31:</b> Multiple alignment of histidine at the amino acid position 3611 in <i>ZFHX3</i> across five species.....	140
<b>Figure 2.32:</b> Electropherogram showing heterozygous <i>ZFHX3</i> c.10557_10558insGGC in PID III-8.....	141
<b>Figure 2.33:</b> Conserved domain of the <i>ZFHX3</i> gene identified in Family 2081 .....	143
<b>Figure 2.34:</b> Electropherogram showing no degradation of the <i>ZFHX3</i> c.10557_10558 insGGC allele.....	144
<b>Figure 2.35:</b> Family 2081 shows segregation of five variants detected in the <i>ZFHX3</i> on the <i>OTSC4</i> yellow haplotype.....	145
<b>Figure 2.36:</b> Partial pedigree of Family 2081 showing absence of segregation of the yellow haplotype in two affected subjects .....	148
<b>Figure 3.1:</b> Family 2081 co-segregating a 9.7 Mb haplotype telomeric to <i>OTSC4</i> mapped locus.....	162
<b>Figure 3.2:</b> Flowchart showing the steps used to analyse variants identified in Family 2081, using Sanger sequencing.....	167
<b>Figure 3.3:</b> Electropherogram showing heterozygous deletion of 15 bp in the <i>FOXL1</i> gene in PID III-8 .....	169
<b>Figure 3.4:</b> Multiple alignment of <i>FOXL1</i> p.G327_L331del across five species.....	170
<b>Figure 3.5:</b> Partial pedigree of Family 2081 segregates <i>FOXL1</i> c. 976-990 het del....	171
<b>Figure 3.6:</b> Flowchart showing filtration steps of variants identified in sequenced exomes of Family 2081 at Chr16q linked region.....	175
<b>Figure 3.7:</b> Partial pedigree of Family 2081 segregating 15 bp del in <i>FOXL1</i> and two variants in <i>PKDIL2</i> gene .....	176
<b>Figure 4.1:</b> Partial pedigree of Family 2114 showing segregation of Chr17q25.1-q25.3 haplotypes. ....	200
<b>Figure 4.2:</b> Electropherogram showing an example of a high quality sequence .....	205
<b>Figure 4.3:</b> Partial pedigree of Family 2081 showing segregation of Chr17 haplotypes in Family 2081 .....	206
<b>Figure 4.4:</b> Flowchart showing filtration steps of NGS data at Chr17q under recessive mode of inheritance.....	211
<b>Figure 4.5:</b> Flowchart showing the steps of NGS analysis at five regions under dominant model .....	217
<b>Figure 4.6:</b> Electropherogram showing heterozygous missense mutation c.1706 C> T,	

	p. P569L in the <i>TNRC6C</i> gene.....	228
<b>Figure 4.7:</b>	Conservation of <i>TNRC6C</i> c.1706 C> T, p.P569L using 13 species.....	229
<b>Figure 4.8:</b>	Electropherogram showing predicted splice site mutation c.69+11C>G in the <i>CDK14</i> gene.....	230
<b>Figure 4.9:</b>	Electropherogram showing nonsense mutation c.49 C>T, p.R17* in the <i>KPNA7</i> gene.....	231
<b>Figure 4.10:</b>	Partial pedigree of Family 2114 showing segregation of <i>KPNA7</i> c.49 C>T and <i>CDK14</i> c.69+11C>G at Chr7:73,862,412-142,344,569. ....	232
<b>Figure 4.11:</b>	Partial pedigree of Family 2114 showing segregation of three candidate variants identified in Family 2114 at Chr17. ....	233
<b>Figure 4.12:</b>	Partial pedigree of Family 2114 showing segregation of <i>CSTF2T</i> c.1143G>A at Ch.10q11.23-q21. ....	234
<b>Figure 5.1:</b>	Hypothetical figure showing the predicted role of FOXL1 and KPNA7 in osteoclastogenesis and bone remodelling. ....	246

## List of Abbreviations and Symbols

ABG	Air-bone conduction gap
ABI	Applied Biosystems Incorporation
AC	Air conduction
AD	Autosomal dominant
ALP	Alkaline phosphatase
AR	Autosomal recessive
BC	Bone conduction
<i>BIRC5</i>	Baculoviral IAP repeat-containing protein 5 isoform 1 gene
<i>BMP2</i>	Bone morphogenetic protein 2 gene
<i>BMP4</i>	Bone morphogenetic protein 4 gene
bp	Base pair
<i>C6orf148</i>	Chromosome 6 open reading frame 148 gene
<i>C6orf150</i>	Chromosome 6 open reading frame 150 gene
<i>CA5A</i>	Carbonic anhydrase 5A, mitochondrial precursor gene
<i>CCL18</i>	Chemokine (C-C motif) ligand 18 gene
<i>CCXL13</i>	Chemokine (C-C motif) ligand 13 gene
<i>CD109</i>	Cluster differentiation 109 antigen (GOV platelet alloantigens) gene
<i>CDK14</i>	Cyclin-dependent kinase 14 gene
cDNA	Complementary DNA
Chr	Chromosome
cM	CentiMorgan
<i>COG4</i>	Conserved oligomeric Golgi complex subunit 4 isoform 1 gene
<i>COL1A1</i>	Collagen, type1, alpha1 gene
<i>COL1A2</i>	Collagen, type1, alpha2 gene
<i>COTL1</i>	Coactosin-like protein gene
CT scan	Computer Tomography scan
<i>CTDNEP1</i>	CTD nuclear envelope phosphatase 1 gene
<i>CX26</i>	Connexion 26 gene
<i>CXCL10</i>	Chemokine (C-X-C Motif) ligand 10 gene
<i>CXCL12</i>	Chemokine (C-C motif) ligand 12 gene
<i>CXCL14</i>	Chemokine (C-C motif) ligand 14 gene
<i>IL8</i>	Interleukin 8 gene
<i>CXCR3</i>	Chemokine (C-X-C motif) receptor 3 gene
<i>CXCR5</i>	Chemokine (C-X-C motif) receptor 5 gene
dB	Decibels
dbSNP	SNP database
<i>DDX19A</i>	DEAD (Asp-Glu-Ala-As) box polypeptide 19A gene

<i>DDX19B</i>	DEAD (Asp-Glu-Ala-Asp) box polypeptide 19B gene
<i>DDX43</i>	DEAD (Asp-Glu-Ala-Asp) box polypeptide 43 or helicase Antigen gene
<i>DFNA15</i>	Deafness autosomal dominant 15
<i>DFNB2</i>	Deafness autosomal recessive 15
<i>DHX38</i>	pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16 gene
<i>DLX5</i>	Distal-less homeobox 5 gene
DNA	Deoxyribonucleic acid
<i>DPPA5</i>	Developmental pluripotency associated 5 gene
<i>ECAT1</i>	ES cell associated transcript 1 gene
ECM	Extracellular matrix
<i>EEF1A1</i>	Eukaryotic translation elongation factor 1 alpha 1 gene
<i>EIF3S6P1</i>	Eukaryotic translation initiation factor 3, subunit 6 pseudogene 1
ES	Mouse embryonic stem
ESE	Exonic splicing enhancers
ESS	Exonic splicing silencers
FCS	Fetal calf serum
FDR	Storey's q-value false discovery rate
<i>FEN1</i>	Flap structure-specific endonuclease 1 gene
<i>FOXC1</i>	Forkhead box C1 gene
<i>FOXD4</i>	Forkhead box protein D4 gene
<i>FOXF1</i>	Forkhead box protein F1 gene
<i>FOXJ1</i>	Forkhead box protein J1 gene
<i>FOXL1</i>	Forkhead box L1 gene
<i>FOXL2</i>	Forkhead box protein L2 gene
GAPDH	Housekeeping gene
Gb	Gigabase
<i>GJB2</i>	Gap junction protein BETA 2
GWAS	Genome-wide association
H-W	Hardy-Weinberg equilibrium
HD	Huntington's disease
HEK293A	Human embryonic kidney fibroblast cell lines
Hex	b-N-acetylhexosaminidase
hFOB 1.19	Human fetal osteoblast
HHL	Hedgehog signaling pathway
<i>HSD17B2</i>	estradiol 17-beta-dehydrogenase 2 gene
HSF	Human splicing finder
Hz	Hertz
IBD	Identical by descent
IBS	Identical by state

<i>IFIT1</i>	Interferon-induced protein with tetratricopeptide repeats 1 gene
<i>IFNB1</i>	Interferon beta 1 fibroblast gene
IgG	Imunoglobulin G
IHC	Inner hair cells
<i>IL1<math>\alpha</math></i>	Interleukin 1 $\alpha$
<i>IL<math>\beta</math></i>	Interleukin $\beta$
<i>IL10</i>	Interleukin 10 gene
<i>IL13</i>	Interleukin 13 gene
<i>IL1A</i>	Interleukin 1 alpha (IL-1 $\alpha$ ) gene
<i>IL1B</i>	Interleukin 1 $\beta$ gene
<i>IL2</i>	Intelukin 2 gene
<i>IL22</i>	Interleukin 22 gene
<i>IL29</i>	Interleukin 29 gene
<i>IL34</i>	Interleukin 34 isoform 2 precursor gene
<i>IL6</i>	Interleukin 6 gene
<i>IL8</i>	Interleukin 8 gene
IMDM	Iscove's modified Dulbecco's medium
<i>IP10</i>	Interferon gamma induced protein 10 gene
<i>IRF8</i>	Interferon regulatory factor 8 gene
ISE	Intronic splicing enhancers
ISS	Intronic splicing silencers
<i>ITGB4</i>	Integrin beta 4 isoform 3 precursor gene
<i>KCNQ5</i>	Potassium voltage-gated channel KQT-like subfamily member 5 gene
<i>KCTD2</i>	Homo sapiens potassium channel tetramerisation domain containing 2 gene
<i>KPNA7</i>	karyopherin alpha 7 (importin alpha 8) gene
LD	Linkage Disequilibrium
LFSNHL	Low frequency non-syndromic sensorineural hearing loss
<i>LOC44116</i>	Similar to RIKEN cDNA 2410146L05 gene
<i>LOC64309</i>	Similar to syndecan-binding protein (syntenin) 2 gene
<i>LOC65319</i>	Similar to chromosome 6 open reading frame 148 gene
LOD	Logarithm of the odds
M-CSF	Macrophage-colony stimulating factor
MAF	Minor allele frequency
<i>MAPK-13</i>	Mitogen-activated protein kinase13 gene
Mb	Megabase
<i>MIG</i>	Chemokine (C-X-C Motif) ligand 9 gene
<i>MITF</i>	Microphthalmia associated transcription factor gene
MAF	Minor allele frequency
MSC	Mesenchymal stem cells

<i>MSX2</i>	Msh homeobox-like 2 gene
<i>MTO1</i>	Mitochondrial translation optimization 1 homolog gene
NCBI	National Center for Biotechnology Information
ND-1000	NanoDrop spectrophotometer
NFAT-c1	Nuclear factor of activated T cells calcineurin dependent 1
NGS	Next generation sequencing
NL	Newfoundland and Labrador
NMD	Nonsense-mediated decay
<i>NOG</i>	Noggin gene
OHCs	Outer hair cells
OI	Osteogenesis imperfecta
ON	Ontario
<i>OPGL</i>	Osteoprotegerin gene
<i>OSGIN1</i>	Oxidative stress-induced growth inhibitor 1 gene
<i>OTOP2</i>	Otopetrin 2 gene
<i>OTSC1</i>	Otosclerosis 1 locus
<i>OTSC2</i>	Otosclerosis 2 locus
<i>OTSC3</i>	Otosclerosis 3 locus
<i>OTSC4</i>	Otosclerosis 4 locus
<i>OTSC5</i>	Otosclerosis 5 locus
<i>OTSC7</i>	Otosclerosis 7 locus
<i>OTSC8</i>	Otosclerosis 8 locus
<i>OTSC10</i>	Otosclerosis 10 locus
<i>PCDH13</i>	Protocadherin 20
PCR	Polymerase Chain Reaction
PID	Personal ID
<i>PKD1L2</i>	Polycystic kidney disease 1 like2 gene
<i>PLCG2</i>	Phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma 2 gene
<i>PMFBP1</i>	Polyamine-modulated factor 1-binding protein 1 isoform b gene
<i>POU3F4</i>	POU domain class3 transcription factor 4 gene
<i>POU4F3</i>	POU domain class4 transcription factor 3
PTA	Pure Tone Audiometry
QPCR	Quantitative PCR
<i>RNF157</i>	RING finger protein 157 gene
<i>RPL38</i>	60S ribosomal protein L38 gene
<i>RPL39P3</i>	Ribosomal protein L39 pseudogene 3
RT	Reverse transcriptase
<i>RUNX2</i>	Runt-related transcription factor 2 gene
<i>SCL17A5</i>	Solute carrier family 17 (anion/sugar transporter), member 5 gene

SDS-PAG	Sodium dodecyl sulfate polyacrylamide
<i>SF3B3</i>	Splicing factor 3b, Subunit3, 130kDa gene
SIFT	Sorting Intolerant From Tolerant
<i>SLC38A8</i>	Putative sodium coupled neutral amino acid transporter 8 gene
<i>SLC7A5</i>	Large neutral amino acids transporter small subunit 1 gene
<i>SLC9A3R1</i>	Solute carrier family 9 subfamily A member 3 regulator 1 gene
SNHL	Sensorineural hearing loss
<i>SNP</i>	Single nucleotide polymorphism
<i>SP4</i>	Sp4 transcription factor gene
SSF	Splice site finder
SYM1	Proximal symphalangism
SYNS1	Multiple synostoses syndrome
TCAG	The Centre for Applied Genomics
TD54	Touchdown 54
<i>TGFB</i>	Transforming growth factor beta gene
<i>TIMP2</i>	Metalloproteinase inhibitor 2 precursor gene
<i>TMPRSS3</i>	Transmembrane protease serine3
TM	Tympanic membrane
<i>TNRC6C</i>	Trinucleotide repeat containing 6C gene
UCSC	Genome Browser homepage at UCSC
<i>USH1G</i>	Usher syndrome type-1G protein gene
WFS1	Wolframin syndrome 1
WHO	World Health Organization
<i>WWP2</i>	WW domain containing E3 Ubiquitin protein gene
<i>ZDHHC7</i>	Palmitoyltransferase ZDHHC7 isoform 1 gene
<i>ZFHX3</i>	Zinc finger homeobox protein 3 isoform A gene
<i>ZNF19</i>	Zinc finger protein 19 gene
<i>ZNF23</i>	Zinc finger protein 23 gene
θ	Recombination fraction

## List of Appendices

<b>Appendix 1</b> : Hearing loss medical questionnaire .....	266
<b>Appendix 2</b> : DNA extraction from whole blood .....	273
<b>Appendix 3</b> : PCR setup protocol .....	275
<b>Appendix 4</b> : TD54 Thermocycling Program .....	276
<b>Appendix 5</b> : Genotyping protocol.....	277
<b>Appendix 6</b> : Microsatellite markers sequence and PCR conditions .....	277
<b>Appendix 7</b> : List of primer sequences and melting temperature <sup>TM</sup> of genes sequenced in <i>OTSC4</i> critical region .....	290
<b>Appendix 8</b> : Cycle Sequencing protocol .....	311
<b>Appendix 9</b> : Protocol for Cycle Sequencing.....	311
<b>Appendix 10</b> : Cycle Sequencing DNA Precipitation Protocol .....	312
<b>Appendix 11</b> : Sequenced functional candidate genes at the minimized 4.2 Mb of <i>OTSC4</i> .....	314
<b>Appendix 12</b> : Sequencing variants in 12 genes across <i>OTSC4</i> .....	316
<b>Appendix 13</b> : List of primer sequences used for genes sequenced in the critical region on Chr16q24 in Family 2081 .....	322
<b>Appendix 14</b> : Pipeline report of NGS.....	328
<b>Appendix 15</b> : Sequenced functional candidate genes at the extended newly linked 9.6 Mb region .....	334
<b>Appendix 16</b> : Variants detected from exome/Sanger sequencing of genes in 9.7 Mb ...	336
<b>Appendix 17</b> : Functional studies to identify the effect of the <i>FOXL1</i> deletion at the RNA and protein levels .....	343
<b>Appendix 18</b> : Real time amplification and interpretation.....	364
<b>Appendix 19</b> : Genes down regulated by FOXL1 mutant .....	366
<b>Appendix 20</b> : Genes up regulated by FOXL1 mutant .....	369
<b>Appendix 21</b> : Statistical analysis report.....	372
<b>Appendix 22</b> : List of primer sequences used for genes sequenced in the critical region on Chr17q in Family 2114 .....	378
<b>Appendix 23</b> : Positional candidate genes on Chr 17q .....	388
<b>Appendix 24</b> : Primers used to amplify three variants identified by NGS .....	389

## **Chapter 1: General introduction**

### **1.1 Purpose of the study and significance**

Hearing loss is the most common sensory defect worldwide. It has a severe effect on language acquisition and the learning process when it develops early in life and affects the social well being and career goals of affected individuals when it develops later in life. Otosclerosis is the most common cause of progressive conductive and mixed hearing loss in the Caucasian population. Otosclerosis is a complex condition that can be treated surgically (a stapedectomy) but this can be challenging in some cases. Stapedectomy can in some cases treat the conductive component of hearing loss but in others, unsuccessful surgery can worsen the hearing loss. Identification of the relevant otosclerosis gene in a particular family can help with screening family members leading to early diagnosis and provision of an appropriate hearing aid. Providing genetic diagnosis allows accurate genetic counselling, and carrier testing for relatives. Also genetic diagnosis may have a role for planning in therapies such as a cochlear implant. Identifying mutations that cause otosclerosis can also help to understand the molecular mechanism underlying the development of otosclerosis and to elucidate the pathways involved in disease development, which ultimately may provide new knowledge for drug therapies. Development of a drug therapy for otosclerosis would be beneficial due to the risks associated with surgery, occasionally resulting in further hearing loss. Many genetic studies have been carried out to identify the genetic cause of otosclerosis in familial and sporadic cases. To date, eight loci have been mapped for otosclerosis and none of the

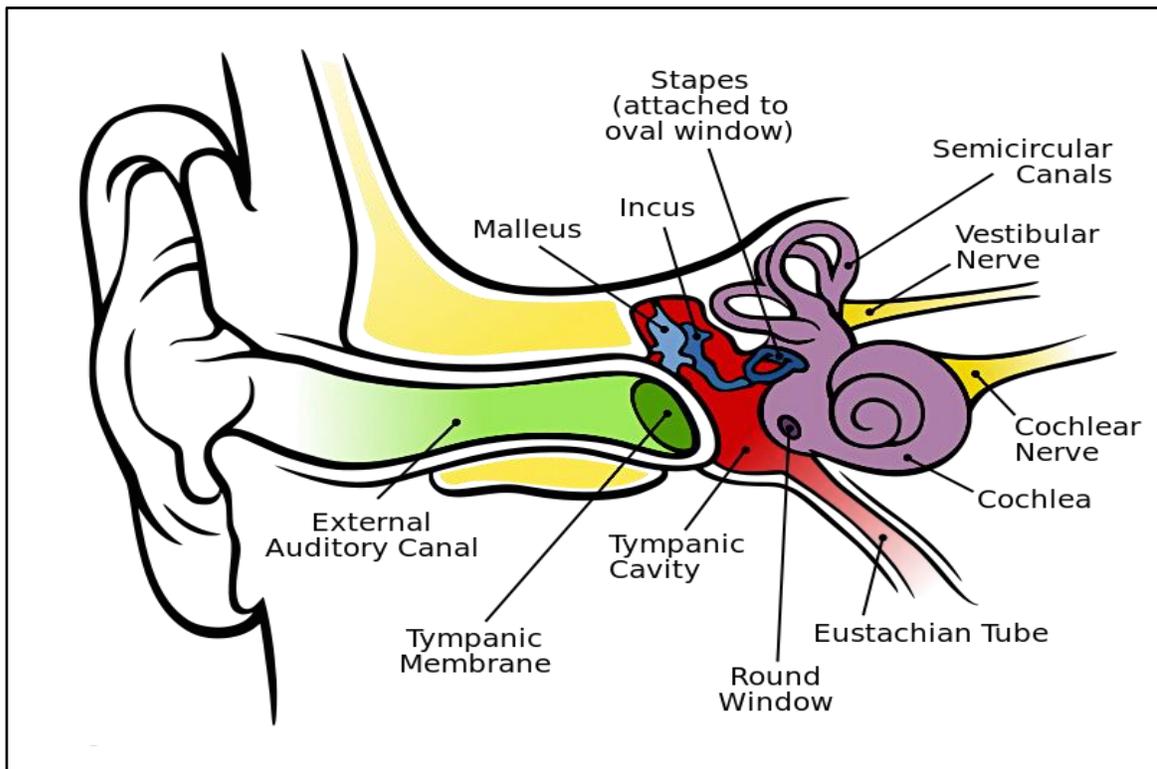
causative genes had been identified when this study began. The ultimate goal of this thesis was to identify otosclerosis genes that segregate in Newfoundland families.

## **1.2 Anatomy and function of the ear**

The ear is divided into three parts: outer ear, middle ear and inner ear. The outer ear consists of the pinna and the auditory canal. The auditory canal captures sound and passes it to the tympanic membrane. The middle ear consists of the tympanic membrane and three small ossicle bones (ossicular chain): malleus (hammer), incus (anvil) and stapes (stirrup). The stapes is attached to the oval window of the inner ear and is characterized by its free movement, which is important for the transmission of sound from the middle ear into the inner ear (Figure 1.1). The inner ear consists of the membranous labyrinth that is embedded in the bony labyrinth located within the temporal bone. The inner ear is divided into the semicircular canal, vestibule, and the snail shaped cochlea. The cochlea is divided into three compartments, the scala vestibule, and scala tympani that are filled with perilymph and scala media (vascularise) that is filled with endolymph. The most important scala is the scala media, which contains the organ of Corti <sup>1</sup>. The organ of Corti consists of one row of inner hair cells (IHC), which are responsible for the transmission of hearing signals into the brain, and three rows of outer hair cells (OHC). The tectorial membrane is located on top of the organ of Corti and is made up of a massive extracellular matrix. It helps deflection of the stereocilia of the hair cells, which is important for the mechano-electrical transduction of sound signals <sup>2</sup>.

### Figure 1.1: Structure of the ear

Ear consists of outer, middle, and inner ear compartments. The outer ear consists of the pinna and external auditory canal. The middle ear consists of three ossicles, the malleus, incus and stapes bones which is attached to the oval window. The tympanic membrane is situated between the external and middle ear. The inner ear consists of the semicircular canals, cochlea and round window. The Eustachian tube is connected to the tympanic cavity where the three middle bones are located. This Figure is taken from reference 3<sup>3</sup>.



### **1.3 How sound is transmitted**

Sound travels from the outer ear to the middle ear through the external auditory canal.

In the middle ear, sound vibrates the three middle ear bones, is transmitted into the cochlea, and moves the fluid in the three scala<sup>4</sup>. Sound traveling fluid in the cochlear compartments leads to movement of the basilar membrane, which activates the sensory cells (OHC and IHC) to convert the mechanical signals into electrical signals. When sound is transmitted into the cochlea, the fluid in the scala media and the tectorial membrane on top of the hair cells move, and the stereocilia are deflected which opens the gate of the ion channels that are located at the stereocilia tips. Opening the gate of these ion channels leads to flow of potassium ions from the potassium-rich endolymph into the hair cells causing depolarization of the cells. As a result of cell depolarization, the voltage-gated calcium channels on the IHC open causing a calcium influx, which triggers the release of neurotransmitters in the auditory nerve<sup>5</sup>.

## **1.4 Hearing loss**

### **1.4.1 Classification of hearing loss**

Hearing loss is classified based on phenotype presented by the patient. Hearing loss is divided into syndromic hearing loss (SHL) and non-syndromic hearing loss (NSHL). SHL is accompanied by extra-auditory clinical symptoms. It is estimated that about 30 % of all hearing loss is SHL and this is mostly caused by mutations in single genes. NSHL is isolated and accounts for about 70 % of the genetic forms of hearing loss<sup>6</sup>. Hearing loss is divided into categories based on the onset of the hearing loss, severity and configuration of the loss.

Hearing loss is classified as pre-lingual or post-lingual based on age of onset of the hearing loss. Pre-lingual hearing loss manifests before speech development. It is important to keep in mind that all congenital (present at birth) hearing loss is pre-lingual, but not all pre-lingual hearing loss is congenital. Post-lingual hearing loss develops after speech acquisition.

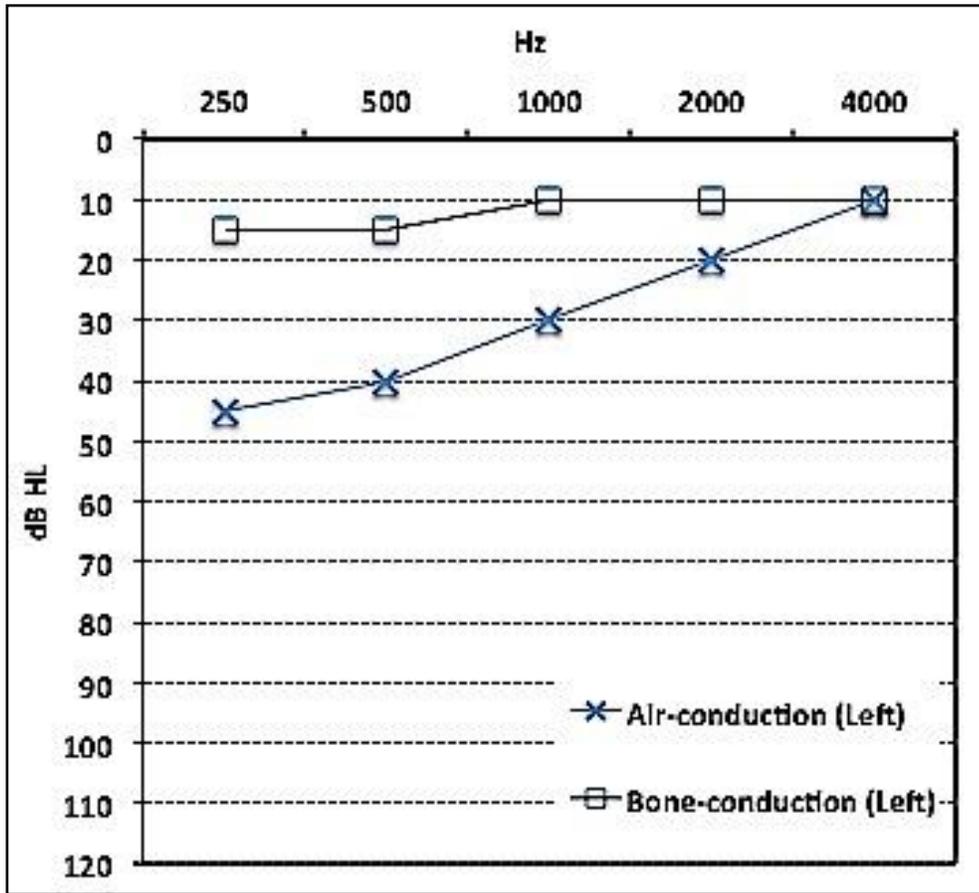
Hearing loss is classified into conductive hearing loss, SNHL and auditory neuropathy. SNHL results from abnormal function of the inner ear, which transforms sound waves to electrical impulses into the auditory cortex. This type of hearing loss is usually permanent and has no medical or surgical treatment. Conductive hearing loss develops as a result of an obstruction that interferes with the conduction of sound from the external and middle ear canal into the inner ear (e.g. wax in the external ear canal, perforation of the ear drum, fluid or infections in the middle ear or fixation of the ossicles to the oval window in otosclerosis). This type of hearing loss is often medically or

surgically treatable. Mixed hearing loss, which is a combination of conductive and sensorineural hearing loss and central auditory dysfunction that results from damage at the level of the cranial nerve, brain stem, or cerebral cortex are, another types of hearing loss<sup>7,8</sup>. These types of hearing loss are defined by measuring the hearing threshold (measured by decibel (dB)) across several frequencies (measured by hertz (Hz)). Pure Tone audiometry (PTA) is the standard technique for measuring the hearing threshold and it is represented on a chart called audiogram (Figure 1.2). It records the air conduction signals and bone conduction signals on the audiogram. Air conduction measures how much of the sound travels through the external and middle ear into the cochlea and then to the brain. Bone conduction measures how well sound travels through the bone to the cochlea, bypassing the middle and external ear. Examples of presentation of conductive and SNHL on audiogram are shown in Figures (1.2, 1.3 and 1.4)<sup>8</sup>.

The degree of hearing loss is measured by PTA and classified according to severity of the loss as shown in Table 1.2. Another classification based on the shape of the air conduction audiogram across different frequencies as shown in Table 1.2. This classification is helpful, in some cases, in identifying the underlying aetiologies. For example, mutation in *TECTA* gene cause cookie-bite configuration.

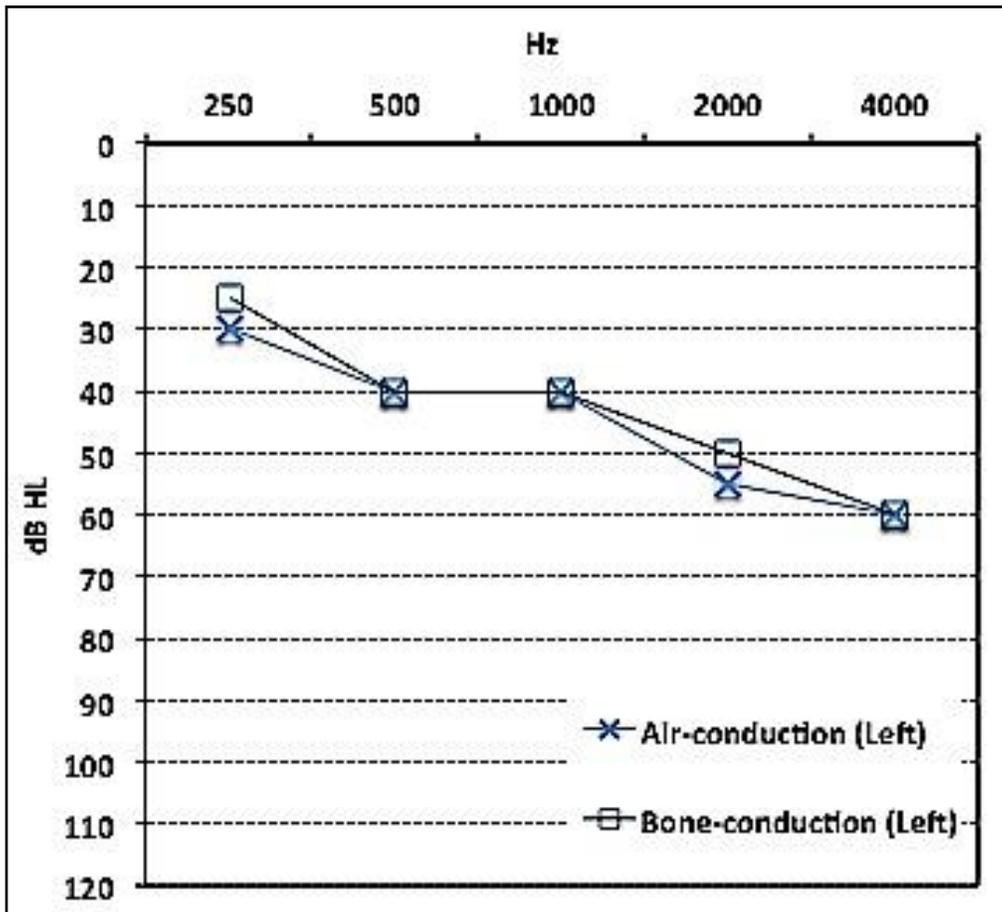
**Figure 1.2: Pure tone audiogram showing air conduction**

Pure tone audiogram from a case with conductive hearing loss. X- axis =frequency; y-axis =intensity). X=air conduction; □=bone conduction. 250-500 Hz are low-pitched frequencies. 2000-8000 Hz are high-pitched frequencies. The audiogram shows differences between air and bone conduction, which is known as air bone gap (ABG). This figure is taken from reference 8<sup>8</sup>.



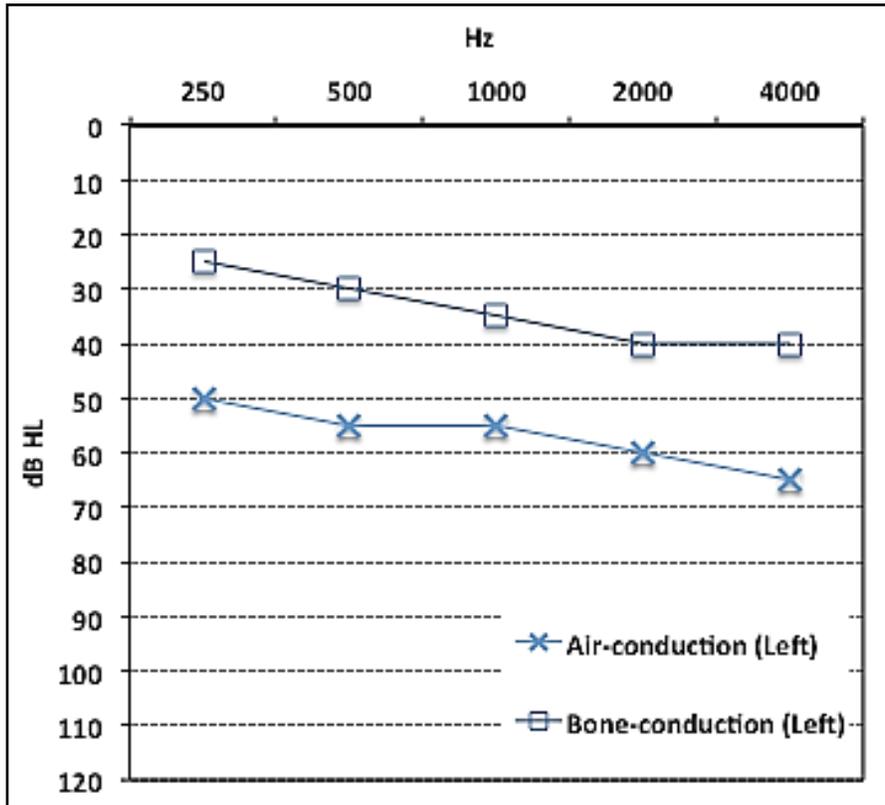
**Figure 1.3: Pure tone audiogram showing SNHL**

Pure tone audiogram from a case with SNHL. X-axis = frequency; y-axis = intensity; X = air conduction; □ = bone conduction. 250-500 Hz are low-pitched frequencies and 2000-8000 Hz are high-pitched frequencies. The audiogram shows no difference in the ABG and drop in all frequencies. This figure is taken from reference 8<sup>8</sup>.



**Figure 1.4: Pure tone audiogram showing mixed hearing loss**

Pure tone audiogram shows a case mixed hearing loss. X axis =frequency; y-axis =intensity). X=air conduction; □=bone conduction. 250-500 Hz are low-pitched frequencies and 2000-8000 Hz are high-pitched frequencies. The audiogram shows both a difference in the ABG and a drop in high frequency. This figure is taken from reference 8<sup>8</sup>.



**Table 1.1: Types of hearing loss based on severity**

<b>Pure tone average (dB)</b>	<b>Degree of Hearing Loss (Hz)</b>
<b>&gt;15</b>	Normal
<b>16-25</b>	Slight
<b>26-40</b>	Mild
<b>41-55</b>	Moderate
<b>56-70</b>	Moderately severe
<b>71-90</b>	Severe
<b>≥ 90</b>	Profound

**Table 1.2: Types of hearing loss based on configuration**

<b>Term</b>	<b>Description</b>
<b>Flat</b>	Hearing loss across all frequencies
<b>Sloping (high frequencies)</b>	Hearing loss affects high frequencies
<b>Rising (low frequencies)</b>	Hearing loss affects lower frequencies
<b>Notch (Carhart, cookie-bite)</b>	Sharply drop at one frequency

Based on mode of inheritance, hearing loss is classified into dominant, recessive, X linked and mitochondrial. To date, 137 loci and 71 genes have been identified for NSHL<sup>9</sup>. All loci associated with hearing loss have been designated as *DFN*, with a further letter determining the mode of inheritance. *DFNA* is the designation for AD inheritance, *DFNB* for autosomal recessive (AR), and *DFNX* for X-linked. Other classifications include *OTSC* for otosclerosis and recently *DFNM* for modifiers and *AUNA* for auditory neuropathy. There is no current designation for hearing loss caused by mitochondrial mutation. The numbers following the identified loci represent the chronological order at which each locus was identified; for example, *DFNA1* represents the first locus to be identified for AD hearing loss.

## **1.5 Gene discovery in non-syndromic hearing loss using families**

### **1.5.1 Ascertainment of families**

The first step in the genetic study of hearing loss in families is to construct a pedigree for the family of interest, define the clinical phenotype, and determine the pattern of inheritance. Carefully and correctly determining the clinical phenotype of all family members is an important step in order to succeed in gene discovery. Ascertainment of a family for a study usually starts with the proband, which is the first person (i.e. Index case) that participates in the study. In order to identify the clinical phenotype in a family, there needs to be a complete family history, prenatal and postnatal medical history for the proband and the proband's family. In the family history, it is important to determine for all the relatives the age of onset of the hearing loss and their relationship in order to determine the mode of inheritance when the full pedigree is constructed. Establishing a history of consanguinity is imperative in families with hearing loss because this suggests a recessive mode of inheritance.

Exclusion of syndromic form of hearing loss is important in apparently isolated hearing loss cases. This requires a physical examination, which includes an eye examination to identify, for example Usher syndrome, which is characterized by both retinitis pigmentosa and hearing loss. Physical examination by a clinical geneticist to identify any dysmorphic features is also required. In cases with a history of renal or thyroid functional abnormalities, laboratory investigation should be carried. For example, the thyroid function test should be ordered to determine the possibility of Pendred syndrome.

Determining the mode of inheritance is an important step in studying the genetics of hearing loss. Hearing loss can be developed as result of an inherited mutation or *de novo* mutation. *De novo* mutation is present in one family member as a result of a mutation in a germ cell (sperm or ovum) of one of the parents. *De novo* mutations have been reported as causative mutations in hearing loss; for example, an R75W *de novo* mutation in *GJB2* (gap junction protein BETA 2) gene was identified in a sporadic case of isolated profound hearing loss<sup>10</sup>.

The inherited form of the hearing loss is transmitted in many ways, including AD, AR, X-linked, and mitochondrial inheritance<sup>11</sup>. AD inheritance passes from generation to generation in a vertical pattern. Only one mutated allele is needed to develop the phenotype. In this case, the offspring have a 50 % chance of inheriting the mutated allele and being affected. In the AR mode of inheritance, two mutated alleles are needed to develop hearing loss. AR hearing loss is characterized by horizontal transmission; both parents are carriers of a mutated allele, and each offspring has a 25 % chance of inheriting both mutated alleles and developing hearing loss. A history of consanguinity suggests that it is likely that hearing loss in a family segregates in an AR pattern. In isolated populations, and when the phenotype is relatively common, the recessive disease may be seen in successive generations (pseudo-dominant inheritance) resulting from marriage of an affected individual with another gene carrier. AR hearing loss accounts for about 80 % of the cases of hereditary hearing loss and the most common gene that causes AR deafness is *GJB2*. AD hearing loss accounts for about 20% of hereditary hearing loss cases<sup>12</sup>.

X-linked inheritance involves genes on the X-chromosome. Males inherit their X chromosome from the mother and their Y chromosome from their father (no male to male transmission). All daughters of an affected father are carriers and transmit the disease gene to 50 % of their children. The most characteristic features of X-linked inheritance are that (i) the males are severely affected and (ii) they develop the disease early in life as they have only one copy of the X chromosome. Females with an X-linked phenotype show a wide range of clinical severity, which has been explained by varying degrees of X-inactivation. X-inactivation is a phenomenon that occurs during early embryogenesis in females. One X chromosome, either paternal or maternal, is inactivated to create balance in gene dosage between males and females<sup>13</sup>. The X-inactivation process occurs randomly; half of the female cells silence the X chromosome inherited from the father and the other half silence the X-chromosome inherited from the mother. If the disease gene is located on the X chromosome inherited from the father, the other copy of the X chromosome inherited from the mother will still be active and produce normal protein. In some cases, this process is not random and one of the chromosomes, paternal or maternal, gets inactivated completely. This phenomenon is known as Skewed X-inactivation<sup>14,15</sup>. In this case, if the disease gene is located on the inactive chromosome, the female will not develop hearing loss. On the other hand, if the disease gene is located on the active chromosome, severe phenotype will develop as result.

In addition to the main modes of inheritance, there are two other modes of inheritance depending on the gender of the carrier. Y-inheritance is characterized by male to male transmission. To date, there are no hearing loss diseases linked to the Y-

chromosome. Mitochondrial inheritance is another mode of inheritance. Since mitochondrial DNA is transmitted from the mother, every child has 50 % chance of inheriting the disease gene. X-linked and mitochondrial hearing loss make up the remaining percentage (about 5 %) of hearing loss<sup>16</sup>. It is estimated that about 50 % of all hearing loss has a genetic basis (a monogenic condition)<sup>17</sup>. To date, 71 genes have been identified where mutations cause hearing loss<sup>12</sup>.

## **1.5.2 Complications of family ascertainment**

### **1.5.2.1 Penetrance**

Penetrance is mainly a feature of dominantly inheritance genetic disorders.

Penetrance is defined as the probability of patients to express the phenotype when they have inherited a disease mutation. Penetrance can be 100%, which means that every person with the mutation will express the phenotype. For example, the missense mutation c.2005 C>T in exon 8 of Wolframin (*WFS1*) gene causes low frequency non-syndromic sensorineural hearing loss (LFSNHL) and is 100 % penetrant; every individual with this mutation develops low frequency hearing loss<sup>18,19</sup>. One hundred percent penetrant conditions are relatively rare and reduced penetrance is more common. For example, if 25 % of those with a mutation express the phenotype, the penetrance is said to be 25 % and this trait shows incomplete penetrance. AD inherited traits could be mistaken as recessive inheritance if a person who has a disease causing mutation is non-penetrant.

Age-related penetrance describes a situation when phenotypes are not clinically detected until adulthood, or have a variable age of onset. The genotype is constant but the phenotype can express at a later age. For example, Huntington's disease is an AD trait

characterized by late onset neurological deterioration and it is one of the most well known examples of a dominant trait of age-related penetrance. Affected individuals may show clinical symptoms at any age from childhood to old age but individuals usually present between the ages of 35 and 45 years. Full penetrance may not be seen until age 65 or later<sup>19,20</sup>. Age-related penetrance is well documented in hearing loss and it complicates the diagnosis in some time<sup>21</sup>. For example, age related penetrance of some manifestation of syndromic hearing loss could obscure the difference between syndromic and nonsyndromic hearing loss. For example, Pendred (PS) syndrome and non syndromic hearing loss with enlargement of vestibular aqueduct (EVA) are similar in the presentation of hearing loss and the presence of EVA. The difference between both disease is the development of goiter in the PS. Age related penetrance of the goiter phenotype can be complicated in the diagnosis. Further family history and laboratory/clinical examinations are important for clear diagnosis<sup>22</sup>.

#### **1.5.2.2 Variable expression**

Variable expression is defined as difference in the degree of phenotype expression caused by a mutation in a gene. For example, neurofibromatosis type 2 (NF2) is an AD disease caused by mutations in the neurofibromin gene and it is characterized by progressive hearing loss, Café-au-lait (coffee with cream-colored) spots on the skin, cataract, neurological and skeletal problems<sup>19,23</sup>. The range of the expressivity of these manifestations is variable. Patients may exhibit a range of symptoms ranging from freckles and pale brown palm to severely disfiguring neurofibromatosis over their body.

The variability in the expression of the full manifestation of a disease obscures the diagnosis and lead to misdiagnosis of the condition.

### 1.5.2.3 Phenocopy

A phenocopy is referred to an affected individual who had developed the same phenotype segregating in the family due to a different etiology<sup>24</sup>. For example, in genetic studies, a phenocopy is defined as an individual marked as being affected however the underlying etiology is different from the other affected individuals in the family.

In many cases, the phenocopies are caused by environmental factors rather than genetic factors but in fact can be due to either<sup>25</sup>. Phenocopies can prove problematic in gene hunting studies. They are an example of heterogeneity within a pedigree that weakens the LOD scores by creating false recombinants. The phenocopy phenomenon is common in genetic studies of hearing loss primarily because the hearing loss is common. For example, genetic analysis of the *DFNA15* kindred identified a linked region on Chr5 that was inherited by all affected in the family except one affected member. Close examination of this person found that their hearing loss was different from that seen in other affected family members and mutational analysis of the *POU4F3* gene revealed that this family member did not have the *POUF4F3* mutation associated with the hearing loss in other members of the family<sup>26</sup>. Another study carried by Abdelfatah et al has identified a novel in-frame deletion in *KCNQ4* (*DFNA2A*) in 10/23 deaf relatives. Further examination of audio profiles of the non-deletion carriers deaf relatives revealed different audiometric profiles from that of the deletion-carriers, which most likely represents phenocopies<sup>27</sup>.

### **1.5.3 Genetic analysis of the families**

There are several approaches to genetically analyzing families with hereditary disorders and I am going to discuss those used in this thesis.

#### **1.5.3.1 Exclusion of known loci or genes**

Exclusion of previously mapped loci for a certain disease is considered an important step that can save time and money and help to find the causative gene. However, this procedure can be lengthy and impractical when the disease is common, when it has been mapped to many regions in the genome, or is caused by many genes, as exemplified by hearing loss.

In AR hearing loss, there are 76 mapped loci with 39 cloned genes, with the most common mutated locus being *DFNB2 (GJB2)*. The most effective way to study AR hearing loss is to screen for the most commonly mutated genes first. In AD hearing loss where 54 loci have been mapped and 25 genes cloned, classifying hearing loss based on audio profiles provides success in identifying causative genes<sup>27,28</sup>. AudioGene is a tool which can provide a list of genes predicted to cause AD hearing loss based on patient audio profiles<sup>29</sup>. Typical audio profiles of patients with mutations in 16 known AD loci have been identified. In X-linked hearing loss, there are only 5 loci with 3 cloned genes. The lower number of these loci makes them more practical to screen individually<sup>30</sup>.

#### **1.5.3.2 Candidate gene and genome-wide association approaches**

Candidate gene (CG) and genome-wide association (GWA) studies are methods aimed to uncover the genetic aetiology of complex diseases by demonstrating an association between genotype and disease susceptibility across one or more markers using

cohorts of case and control (affected cases and unaffected controls)<sup>31</sup>. For each association study, two cohorts have to be carefully designed with clear phenotype of the cases and exclusion criteria for the control. The primary cohort used for an association study is called a discovery cohort and the secondary cohort that is used for replication and validation of the primary results is called a replication cohort.

The main difference between CG and GWA studies is that CG examines association based on candidate genes, while GWAs examine the whole genome for an association. Both techniques show ability to identify susceptibility genes, with some advantage for each one over the other<sup>32,33</sup>. CGs lack the ability to identify novel susceptibility genes due to limitations in choosing candidate genes, but manage to provide a high statistical power. On the other hand, GWA studies are capable of identifying novel susceptibility genes, but their statistical capabilities are of low power. The statistical power of the GWAs is increased when millions of the single nucleotide polymorphism (SNP) are used.

GWA studies have many limitations that interfere with their success. Of these limitations, lack of the properly defined case control cohort, especially in complex diseases with no clear phenotype (many manifestations). Insufficient sample size, population stratification and massive number of statistical tests, which increase the possibility of false positives, are other limitations encountered in GWAS. The major obstacle facing the GWAs is reproducibility. Failure to replicate the results of an association study in another population, leading to doubt in the accuracy of the findings. This failure can be

result of poor definition of the phenotype used in the association studies or difficulty in attributing genotypes<sup>34</sup>.

### **1.5.3.3 Genome-wide linkage analysis**

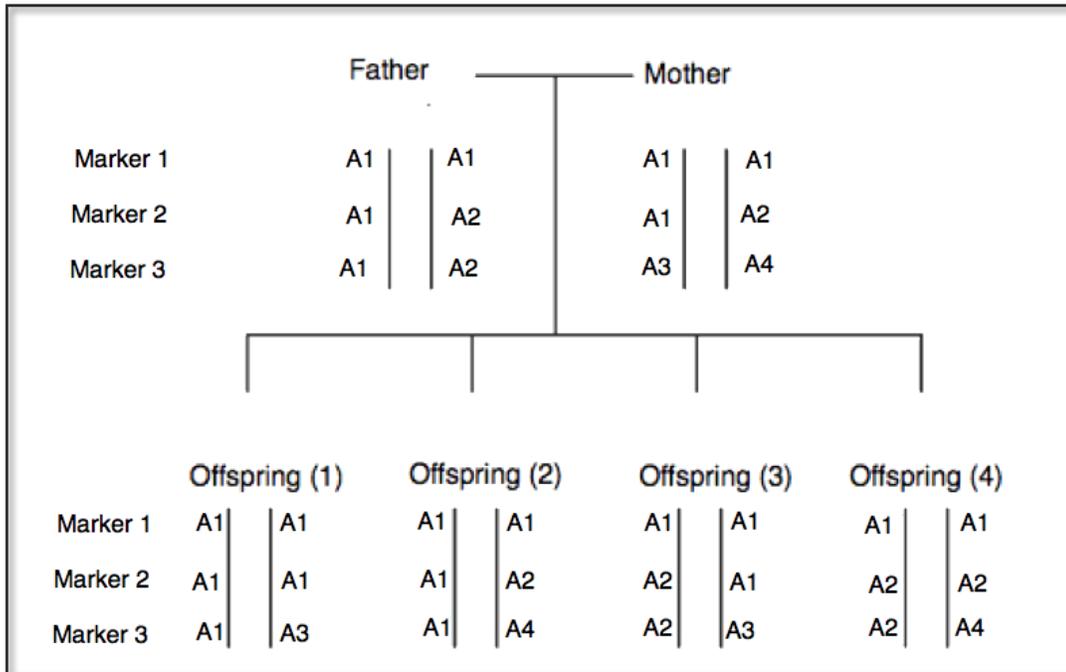
Multiplex families with strong pedigree structures and with genomic DNA available from two or more generations are suitable for genome-wide linkage analysis. Genomic DNA from available individuals (including affected and unaffected) is genotyped with polymorphic markers that span each of the 22 human autosomes. Each marker has a specific variation called allele. There are two types of markers that have been used for genome-wide scans; microsatellite markers and SNPs. Microsatellite markers are characterized by repetitive sequences that vary in length between individuals. These microsatellite markers are mainly informative because they are poly-allelic and result in informative genotypes<sup>35</sup>. SNPs are bi-allelic, giving limited genotype variability but they are more frequent and highly distributed in the genome<sup>36</sup>.

An informative marker is defined as a marker that has a high heterozygosity rate in a population. In other words, informative markers are markers that give a heterozygous genotype (two different alleles). The degree of informativeness of a marker is measured by how many alleles it gives at a certain locus within a family. The main aim of using informative markers is to obtain informative genotype (i.e. two different alleles) and therefore, informative offspring. For example, if both parents are homozygous at the marker locus (i.e. one allele), none of the offspring are informative; if both parents are heterozygous for the same alleles at the same locus, half of the offspring are informative. On the other hand, if both parents are heterozygous for different alleles at the same locus

(i.e. four different alleles), all offspring are informative (Figure 1.5). In most cases, genotypes for a single DNA marker are not informative, but because DNA tends to be inherited in large segments (haplotypes), looking at genotypes for several markers across the DNA region can be very informative. Haplotypes, which are groups of alleles that are inherited together from a single parent, have a great role in pedigree analysis. Pedigree analysis contributes significantly in the field of genetics by playing an important role in the identification of disease-genes.

**Figure 1.5: Degree of informativeness of markers within a pedigree**

An example of marker informativeness within a pedigree. Parents and all offspring are homozygous for marker 1 locus. For marker 2 locus, parents and 50 % of the offspring are heterozygous for the same allele. Each parent inherited a different allele at marker 3 locus and all offspring are heterozygous. A1, A2, A3 and A4 are alleles. Both markers 2 and 3 are examples of an informative markers in this family, however marker 1 is not informative.



The main goal of performing genome-wide genotyping in a family with a certain disease is to identify a specific region in the genome that is shared by affected relatives, but not shared by unaffected relatives, which gives a rough location of the disease gene. Identifying of this specific region in the genome requires identifying genetic recombination within the family. Recombination is a process in which paternal and maternal chromosomes exchange parts. Recombination occurs in the germline cells and passed down from generation to generation. Recombination commonly occurs between largely distant markers. If the markers are very close together, recombination is less likely.

Mapping a disease gene requires measuring the recombination fraction ( $\theta$ ). The recombination fraction measures the distance between genotyped markers and the likelihood of a crossover event to occur between two markers. The closer the markers, the less likely that a recombination can occur so they tend to segregate together<sup>37</sup>. If markers are tightly linked and no recombination occurs, the recombination fraction is 0. If markers can separate from each other by a crossover, the recombination fraction will have a value range from 0 to 0.5. In other words, if two markers are not close, the frequency of recombination will be 50 % (half of the offspring will be recombinant) and if these two markers are close, the frequency of recombination will be less than 50 %. The frequency of recombination is equal to the number of recombinant offspring divided by the total numbers of offspring. Loci/markers are unlinked if they are located on different chromosomes or lie far apart on the same chromosome. Conversely, they are linked if they are located close to each other<sup>38</sup>. The frequency of the recombination is an important

point in haplotype analysis within pedigrees. The less recombination detected, the more accurate the haplotypes. To construct haplotypes from the genotype data, all possible distinct haplotype possibilities have to be considered and haplotypes with least recombination are chosen<sup>39</sup>. In some situations, even using the rule of least number of recombination, the numbers of recombinant haplotypes are still higher than the expected. It has been reported that the rate of recombination is variable across the genome which can explain the high rate of recombination in some regions of the genome, especially the telomeric end of chromosomes<sup>40</sup>.

Linkage analysis is a statistical method that measures how likely the phenotype in question is linked to a certain region in the genome. There are two types of linkage analysis: parametric and non-parametric. In parametric linkage analysis, all parameters have to be defined such as the mode of inheritance, allele frequency, penetrance, and the phenocopy rate. In contrast, non-parametric linkage analysis does not require defined parameters and is based on identifying identical-by-descent (IBD) alleles that are shared between affected individuals in a pedigree<sup>41</sup>. In non-parametric linkage analysis, two likelihoods are considered: i) The likelihood that the phenotype segregating in the family is due to genetic linkage between loci, and ii) the likelihood that the phenotype segregating in the family is not a result of linkage between markers ( $\theta=0.5$ ). The ratio between these two likelihoods gives the odds of linkage versus no linkage. The logarithm of the odds ratio is the LOD score (Z). The LOD score is calculated at different  $\theta$  ratios for each marker. A LOD score of 3 is the threshold for accepting linkage. A LOD score

below or equal to -2 is considered a minimal threshold for linkage exclusion. LOD scores above 1 are considered suggestive of linkage.

Once the linked region has been determined, the boundaries of this region represent the disease interval. Usually the identified region is large because the markers used for whole genome genotyping are widely spread across the genome. In order to decrease the size of the linked region, markers (SNPs or microsatellite markers) that are spaced close to each other across the linked region are genotyped in all available relatives. Close spacing of genotyped markers helps to identify any critical recombination that reduces the genetic interval. The newly minimized region is called candidate region. Decreasing the size of the linked region, decreases the number of genes needed to be screened<sup>42</sup>.

#### **1.5.3.4 Positional candidate gene sequencing**

All annotated genes within the disease candidate region are called positional candidate genes. It is important to examine all the annotated genes in the critical region. All positional genes in a candidate region are candidate genes and they have to be screened because there are many genes in the genome of unknown function and expression. We can prioritize some genes to be screened first based on their expression profile or involvement in a relevant pathway. In hearing loss, genes that are expressed in the ear would be prioritized. Hearing loss is divided into many types and for each type, certain genes would be prioritized. Thus, genes that are expressed in the cochlea, hair cells, tectorial membrane, or any part of the inner ear would be prioritized for the SNHL. Similarly, genes that are expressed in the bone of the ear, or have a role in inflammation

or the immune system, would be a good candidate for conductive hearing loss caused by overgrowth of the bone cells (otosclerosis).

In the past decade, Sanger sequencing has been the gold standard method for screening genes. The Sanger sequencing technique is very suitable for detecting small mutations, including point mutations, and small insertions and deletions (indels)<sup>43 44</sup>. Traditionally, Sanger sequencing was utilized to find mutations in the coding region of a gene along within intron-exon boundaries.

#### **1.5.3.5 Mutation identification and validation**

In gene screening, a screening panel is designed using genomic DNA from a small number of confirmed clinically affected family members and unaffected family members who do not have the phenotype. A mutation screening panel is a quick screening tool for candidate genes. It is then sequenced for the positional candidate genes in the region of interest. Variants that are present in affected members but absent in unaffected members are designated for follow up. Frequency of the candidate variants can be primarily determined through SNP database (dbSNP) and the 1000 genomes project<sup>45,46</sup>. Rare variants are more likely to be pathogenic but with the development of next generation sequencing techniques (NGS) to sequence the genomes of many individuals, dbSNP and the 1000 Genomes Project databases have been hugely expanded with variants of unknown frequency or low frequency (less than 1 %) <sup>45,46</sup>. For known variants, minor allele frequency (MAF) of each variant is listed in dbSNP. MAF refers to the frequency of least common allele in certain population. Variants that are in dbSNP/1000 Genomes Project database with a low MAF are candidates for follow-up.

There are many types of variants that can be identified through Sanger sequencing. These variants include missense mutations, silent variants, splicing variants, deletions and insertions. Variants located in highly conserved coding regions and/or causing dramatic changes in amino acids (e.g.: hydrophilic to hydrophobic) are more likely to be pathogenic. Intronic variants could affect the donor and acceptor splice site. They also can affect branch points, Intronic Splicing Enhancers (ISE) and Intronic Splicing Silencers (ISS). Not only intronic variants are candidates for aberrant splicing defects, but synonymous (silent) and benign missense variants are also possible candidates for splicing because they could affect Exonic Splicing Enhancers (ESE) and Exonic Splicing Silencers (ESS). Consequently, they can enhance or suppress the accurate splicing of the pre-mRNA. Variants within these sequences could affect the splicing efficiency at the splice donor/acceptor and could activate cryptic donor/acceptor sites. Variants within the ESE and ESS motifs could be predicted as silent or benign missense variants at the protein level, which do not affect the protein drastically. Unfortunately, despite predictions that these variants are benign, functional assessment is necessary to rule out whether there are major drastic effects on RNA splicing that could not be predicted using prediction programs. Hence, silent variants and missense mutations that are predicted to be benign should be checked for possible splicing effect using a variety of splicing prediction programs.

Different bioinformatics analysis prediction programs are available to help predict the outcome of nucleotide change on protein and also predicted for possible splice site at the variation position. Two of these prediction programs, Sorting Intolerant From Tolerant

(SIFT)<sup>47</sup>, and PolyPhen-2<sup>48</sup> are used for missense mutations prediction. Human Splicing Finder (HSF), MaxEntScan, NNSPLICE, GeneSplicer, SpliceSiteFinder (SSF) and other known constitutive signals splice prediction tools are used for predicting the effect of the splice site mutations on the splicing efficiency<sup>49-55</sup>.

The SIFT prediction program is based on measuring the degree of conservation of an amino acid residues across closely related sequence through searching for the protein sequence homology to the protein of interest. Substitutions with scores less than 0.05 are predicted to be deleterious by SIFT. The Polyphen program predicts the possible impact of an amino acid change on the physiochemical properties of the protein product. The effect of amino acid substitutions is categorized by Polyphen according to the following scores: scores of 0.00–0.99 are classified as benign, 1.00–1.24 as borderline, 1.25–1.19 as potentially damaging, 1.50–1.99 as possibly damaging, and  $\geq 2$  as damaging<sup>56</sup>.

Splicing prediction programs are aimed to assess if the identified variants are located within putative donor and acceptor splice site, branch points and cis-acting elements ESE, ESS, ISE and ISS. The splicing prediction programs use algorithms to calculate the consensus values of potential splice sites and search for branch points and each program has a particular score value as follows: SSF [0-10], HSF [0-100], MaxEntScan [0-12], NNSPLICE [0-1], GeneSplicer [0-15]. The scores calculated by these programs give an indication of the possible splicing effect of the mutation based on the measured differences in the scores between the wild and mutant forms. The higher the difference in the score between the wild and mutant forms, the more likely the prediction is true.

By using previously listed prediction program, variants with deleterious effect on the protein level as well as variants that affect splice site will be of interest. Not to mention, deletion and insertion (indel) variants are highly candidate variants. There are no available programs to predict the effect of indel variants on the protein level. Sequencing analysis produce a huge number of interesting variants. In order to determine which one of these variants is the disease causing variant, segregation analysis of these variants across large family and determine the allele frequency of candidate variants in ethnically matched population will help in identifying mutation causing disease. Also, identifying additional families with the same mutation, or with different mutations within the same gene, and performing functional studies in order to show pathogenicity, are important steps to validate potentially interesting findings.

Variants are shared across families in two ways: 1) In case of a founder mutation, families inherit variants that are identical by descent (IBD) and this variant is placed on a founder haplotype. Founder haplotypes are segments of chromosomes shared by several individuals from the founder population and inherited from a shared ancestor; 2) In case of a recurrent mutation, families inherit the same variant but this variant occurs at distinct haplotype, identical by state (IBS). In other words, IBD alleles refer to alleles that are identical copies of ancestral allele. IBS refers to alleles that are the same due to coincidental co-occurrence of these alleles and not driven from a known common ancestor.

### 1.5.3.6 Next generation sequencing

Exploring the sequence of the human genome is essential for genetic studies. In the last decade, Sanger sequencing has provided great genetic information for many diseases. The limitations of Sanger sequencing regarding throughput, speed and genome coverage, have prevented scientists from obtaining the required genetic information in a reasonable time and has led to the development and adoption of a new technology. NGS is a new genomic technique, which sequences the whole genome in parallel with a very rapid throughput<sup>57</sup>. The development of NGS, especially exome sequencing, has made a huge impact to the field. Exome sequencing sequences all the coding regions of the genome (~1 % of the genome) excluding the intronic and intragenic regions<sup>58</sup>. Exome sequencing has proven utility in the discovery of a variety of disease gene mutations, including those for infantile hepatopathy<sup>59</sup>, non-small cell lung carcinoma<sup>60</sup>, and inflammatory bowel disease<sup>61</sup>. In the last few years, there are many competitive platforms with the same principle; procedure and different technologies were released. The principle behind NGS is based on sequencing a small fragment of DNA and identifies the base of each fragment from analyzing the signal released from each fragment. This procedure is applied to millions of fragments and produces hundreds of gigabase of data in a single run. Fragmentation of genomic DNA (gDNA) into a library of small fragments is considered the first step of NGS process. The constructed library is further enriched and sequenced in millions of reactions. The sequenced stretch of bases, called reads, is then realigned to a reference sequence. The final output of the sequence is a full set of aligned reads that cover each chromosome in the sequenced sample. The output of the NGS is hugely increased in the last few years from one gigabase (Gb) in 2007 to terabase in

2011<sup>62</sup>. The huge improvement in the sequencing output allowed researchers to move with their study forward. With the decreased cost of NGS, researchers can now sequence many samples in a single run. The differences between platforms are based on technology under the reaction, run time, read length/coverage and throughput per run. Of these factors, the depths of coverage and uniform distribution of reads are critical to achieve the maximum sequencing efficiency. For example, the Illumina sequencing platform uses bridge amplification technique to amplify the target library and to perform a library cluster. Sequencing technology used by Illumina platform is sequencing by synthesis that produces a fluorescent signal from incorporation of fluorescent-labeled nucleotides into the newly synthesized strand<sup>63</sup>.

Analysis of NGS data requires many steps beginning with mapping of the raw data to a reference sequence, This is followed by extensive use of a wide variety of bioinformatics tools, such as software for variant calling in the detection of SNPs and indels<sup>64</sup>. As a significantly large amount of data is provided by one NGS run, it is advantageous to sequence large numbers of well defined affected and unaffected family members which will significantly help reduce the number of potentially causative variants.

First, variants present in the unaffected exomes, as well as common variants with high frequencies from SNP/ 1000 genome database, and the exome variant server are removed<sup>46</sup>. Only novel variants and variants with low frequencies are included for further analyses which are subsequently filtered based on pathogenicity effects, i.e., nonsense, missense mutations, or potential splice variants. Validation and segregation analysis of

these potential pathogenic variants involves Sanger sequencing using all available members of the study family. The use of NGS technology reduces processing time while increasing data output. However, due to the huge amount of data produced, analysis requires a high powered computer system<sup>65</sup>. The combined use of NGS and Sanger sequencing technologies provides powerful tool in gene discovery studies.

## **1.6 Otosclerosis**

### **1.6.1 Overview**

Otosclerosis is the most common cause of progressive conductive and mixed hearing loss<sup>66 67</sup>. It is thought that otosclerosis develops as a result of abnormalities in bone remodeling in the endochondral layer of the otic capsule. Abnormalities in bone remodeling in the otic capsule leads to development of sclerotic foci and the footplate of the stapes bone is the common site affected<sup>68</sup>. The development onset of sclerotic foci ranges between the third and sixth decade, and although the effect can be unilateral, 80 % of those with otosclerosis have both ears affected<sup>69</sup>.

Otosclerosis is classified into clinical otosclerosis and histological otosclerosis. Clinical otosclerosis is characterized by conductive hearing loss which develops as a result of fixation of the stapes bone to the oval window by sclerotic foci. In addition to conductive loss, a SNHL component may develop later. SNHL is present as a first sign in 10 % of cases<sup>70</sup>. In contrast, histological otosclerosis is a type of otosclerosis discovered through post-mortem examination of the temporal bones of people with no proclaimed hearing loss but with an identified otosclerotic lesion. In Caucasian population, clinical

otosclerosis has a mean prevalence of 0.3-0.4 %<sup>71</sup> while histological otosclerosis has a mean prevalence of 3.5-4.5% in the same population<sup>70,72</sup>.

Conductive hearing loss but not residual SNHL is treatable with great success by stapedectomy surgery<sup>73</sup>. The cause of the SNHL component is unknown, though it may be a consequence of abnormal remodelling of the bony labyrinth which lies in close proximity to the cochlea, and damage due to lysosomal enzyme release<sup>74 75</sup>. Another explanation for SNHL is hyalinization of the spiral ligament caused by adjacent otosclerosis foci<sup>75,76</sup>.

Ethnicity is considered an important factor in the prevalence of otosclerosis. For example, the prevalence of clinical otosclerosis in Japanese and Chinese populations is similar and is estimated to be half as common in these populations as in Caucasian populations<sup>77 78</sup>. While these prevalence studies provide an estimation of the disease frequency in a certain population, the small number of participants and lack of clinical data can give a biased interpretation of results. Most of previous studies carried to estimate the percentage of histological otosclerosis was of high-risk bias. This high risk bias arose from including all specimens with otosclerotic temporal bone, without distinction. Of these specimens collection, about 15 % showed evidence of stapes fixation<sup>70</sup>. These selections bias the actual estimation and can overestimate the histological prevalence. On the other hand, studies carried by Declau et al and Schuknecht et al, were based on using unselected series of temporal bone and exclusion of all specimen with evidence of stapes fixation. The prevalence estimate of histological otosclerosis by their studies was lower than the previous reported studies. It ranged

between 2.5% and 4.4% by Declau et al and Schuknect et al respectively<sup>70</sup>. The prevalence of otosclerosis in the NL population is not known and as immigrants from England and Ireland settled NL originally, we can speculate that the prevalence of otosclerosis in NL may be similar to that of other Caucasian populations especially, North European. Developing a study that can estimate the actual prevalence of otosclerosis in NL population will be of great help in genetic studies of otosclerosis.

Otosclerosis is considered a non-syndromic form of conductive hearing loss. Syndromic forms of conductive hearing loss that affect bone remodelling, include osteogenesis imperfecta (OI), and congenital stapes fixation. OI is a condition characterized by bone fragility and is associated with conductive hearing loss similar to otosclerosis<sup>19,79</sup>. Approximately 90 % of OI cases are caused by mutations in *COL1A1* or *COL1A2* genes that encode alpha 1 and alpha 2 chains of type I collagen. The histopathological changes in the ears of OI patients are very similar to those observed in otosclerosis patients, and nearly half of all individuals with OI develop conductive hearing loss occurring in the second and third decade<sup>80,81</sup>.

Congenital stapes fixation is a non-progressive condition that occurs in the first decade along with other skeletal manifestations, such as proximal symphalangism (SYM1), multiple synostoses syndrome (SYNS1) and stapes ankylosis with broad thumb and toes. It is caused by mutations in the *NOG* gene<sup>82</sup>. Previous genetic studies in otosclerosis have screened *COL1A1*, *COL1A2* and *NOG* genes in families with isolated otosclerosis, but no mutations have so far been identified<sup>83 84-90</sup>.

### **1.6.2 Bone remodeling and otic capsule**

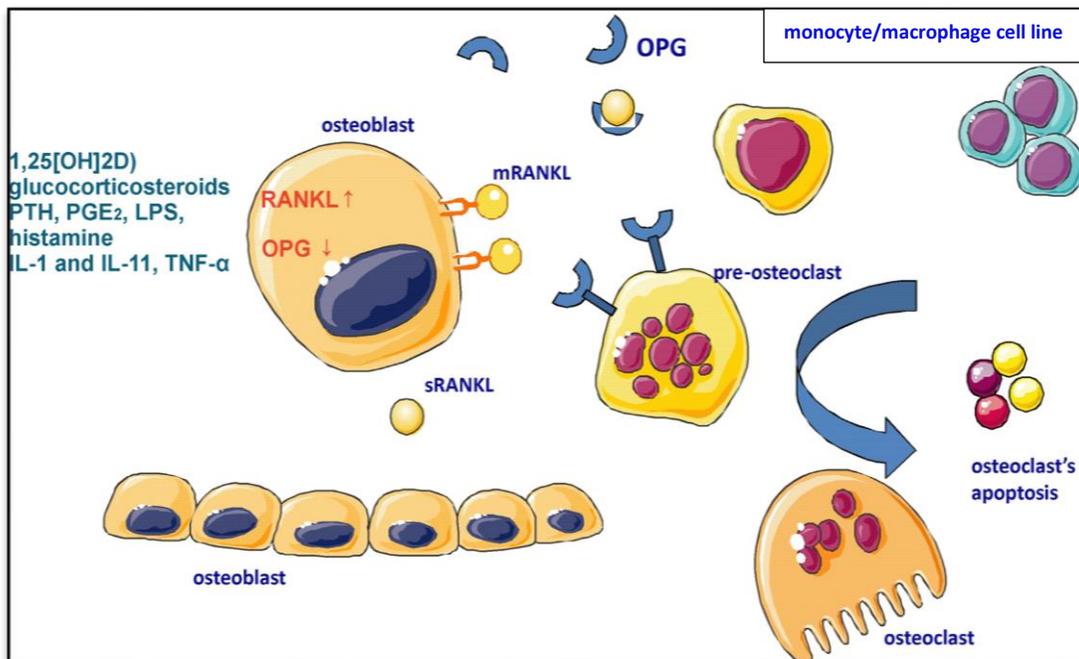
The otic capsule is trilaminar with a middle endochondral layer. The main characteristic feature of the endochondral layer of the otic capsule is that bone ossification is completed after the first year of life. Once it calcifies, the bone in the otic capsule undergoes very slow bone remodeling each year<sup>91</sup>. In comparison, the rate of bone remodeling for most of the human skeleton is high. The difference in the rates is thought to be due to the production of osteoprotegerin (OPG) by the cochlea, which antagonizes the RANKL ligand (activation of bone remodeling) and maintains the static state of otic capsule bone<sup>92</sup>.

Sclerotic lesions (foci) develop as a result of imbalance of bone remodeling in the otic capsule. Bone remodeling is a normal physiological process that is controlled by both osteoclasts (bone resorbing cells) and osteoblasts (bone forming cells). Human skeleton integrity depends on the balance between bone resorption by osteoclasts and bone formation by osteoblasts<sup>93</sup>. A detailed knowledge about osteoclasts and osteoclastogenesis helps to understand how otosclerosis develops. Osteoclastogenesis is a process that produces osteoclasts. Osteoclastogenesis is initiated by mononuclear cells, and requires expression of macrophage-colony stimulating factor (M-CSF), Osteoprotegerin (OPG) and other transcription factors. In the presence of receptor activator of nuclear factor kappa-B ligand (RANKL), produced by osteoblasts, Osteoclasts undergo differentiation after binding the RANKL ligand to the RANK receptor on the immature osteoclast (Figure 1.6). The expression of many factors is required for osteoclast maturation including the nuclear factor of activated T cells, calcineurin dependent 1

(NFAT-c1); microphthalmia-associated transcription factor (MITF); and AP-1 transcription factor c-fos expression<sup>94</sup>. The lysosomal enzymes required for bone resorption are secreted by the mature active multinucleated giant osteoclasts. New bone matrix is deposited by osteoblasts in the site where bone destruction by osteoclasts occurs. Pre-osteoblasts are differentiated from mesenchymal stem cells (MSC) in the presence of runt-related 2 (RUNX2), distal-less homeo box-5 (DLX5) and MSH homeobox-like 2 (MSX2) transcription factors, as well as glucocorticosteroids, parathormone (PTH), vitamin D (1,25-(OH)<sub>2</sub>D), prostaglandins (PGE<sub>2</sub>), lipopolysaccharides (LPS), histamine and pro-inflammatory cytokines: interleukins (IL-1 and IL-11) and tumor necrosis factor alpha (TNF- $\alpha$ )<sup>97</sup>. Collagen type1, osteoclastin, and alkaline phosphates are secreted into the matrix by mature osteoblasts. Some of the osteoblast cells turn into osteocytes which function to maintain the mineralization of the bone<sup>95 96</sup>.

### Figure 1.6: Pathways involved in bone remodeling

Osteoclast and osteoblast cells with RANKL ligand on the surface of the osteoblasts. Osteoclasts carry the RANKL receptors. Upon binding the RANKL ligand to the RANKL receptor, pre-osteoclasts undergo maturation and develop into mature osteoclasts. mRANKL = membrane-bound RANKL; sRANKL = Soluble form of RANKL. This figure is taken from reference 97<sup>97</sup>.



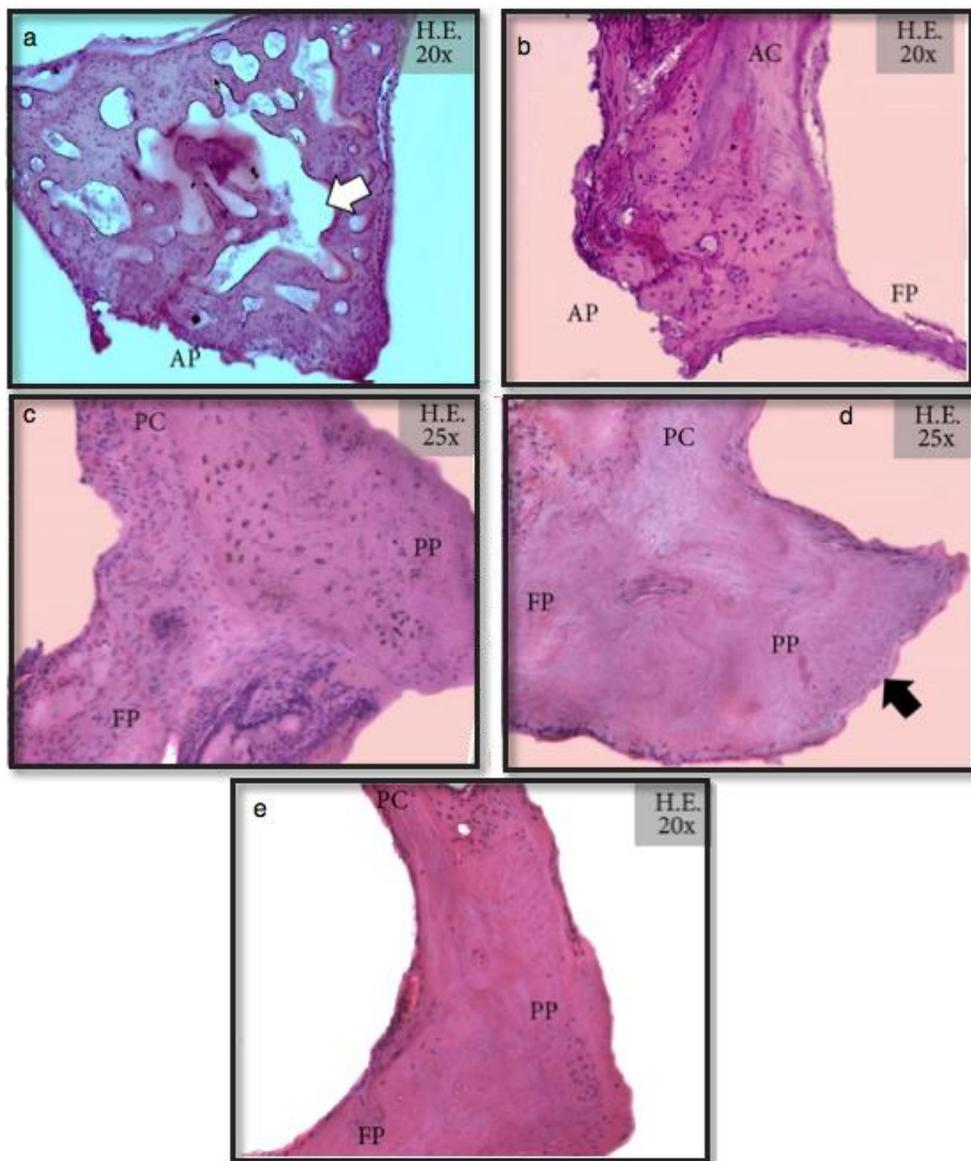
### 1.6.3 History and stages of otosclerosis

In humans, otic capsule calcification is complete by the first year of life and there is no activity of osteoblasts and osteoclasts detected in the otic capsule at the adult age<sup>98</sup>. In otosclerosis, osteoclast cells resorb the endochondral bone of the otic capsule followed by deposition of new bone by osteoblasts, which results in a poorly organized region that doesn't resemble the normal structure of an otic capsule. Otosclerosis pathology runs in two stages: i) active stage (otospongiosis) which is characterized by hypercellularity where osteoclasts and osteoblasts can be seen in large numbers with fewer collagen fibers in the extracellular matrix, and wide pseudovascular space "Swiss cheese" pattern (Figure 1.7)<sup>104</sup>; ii) inactive stage (sclerosis) which is characterized by hypocellularity and abundant deposition of new, woven bone, which increases thickness of the stapes bone footplate. Otosclerosis foci can become inactive at any time but also can become reactivated at any time. Both active and inactive otosclerosis foci can be seen in histological specimens from an otosclerotic region<sup>99</sup>. The active area in the otic capsule appears as hypodense regions surrounding the inner ear on a CT-scan. In later stage of otosclerosis, this active region is replaced by fibrous tissue and forms a dense sclerotic bone<sup>100</sup>.

Valsalva was the first to report otosclerotic lesions in 1741<sup>101</sup> but it was Politzer who, in 1894, identified otosclerosis as a common cause of hearing loss<sup>102</sup>. In 1912, Seibenmann first described the active form of otosclerosis (otospongiosis) as characterized by a high vascularized region in the otic capsule with activated macrophages and osteoclasts<sup>103</sup>.

**Figure 1.7: Histological findings in different stages of otosclerosis**

Different histological stages of otosclerosis. (a) Active stage with hypercellularity and increased number of osteoclasts and osteoblasts at the anterior pole (AP) of the stapes bone. Arrow points to the wide pseudovascular spaces “Swiss cheese” pattern. (b) Inactive otosclerosis foci at the AP of stapes bone with hypocellularity, hypovascularity and woven bone (line). (c) Advanced stage of otosclerosis with intense eosinophilic staining. (d) Sclerosis, hypocellularity, hypovascularity and increased thickness of the stapes footplate pointed by black arrow. (e) Histological image of normal stapes. PP: posterior pole; AC: anterior crus; PC; posterior crus; FP: footplate. This figure is taken from reference104<sup>104</sup>.



#### **1.6.4 Diagnosis of otosclerosis**

Commonly, patients with otosclerosis present with slowly progressive hearing loss in their 20s or 30s. In some cases, associated symptoms are reported including vertigo or dizziness, otorrhea and tinnitus. The most definitive diagnosis of otosclerosis is actual visualization of the stapes fixation during the stapedectomy surgery<sup>105</sup>. Pre-operative diagnosis of otosclerosis is based on audiological evaluation of the hearing loss. The main aim of the audiological tests is to establish the conductive component of the hearing loss. In addition to audiological evaluation, visualization of the tympanic membrane (TM) by otomicroscopy is essential in otosclerosis diagnosis. In majority of patients, TM appears normal. In 10% of cases, red hue may be seen in area with hypervascularity in the active phase of otosclerosis (Schwartz sign)<sup>106</sup>.

The audiological evaluation is based on pure tone audiometry, tympanometry, and acoustic reflexes, which are the most important tests to diagnosis otosclerosis and arranging for further treatment. Tympanometry is appreciated in otosclerosis, especially in the advanced stage as it measures the compliance of the TM. In the early stage of otosclerosis, the middle ear pressure is not changeable and therefore, the peak of the tympanogram is always in the normal range. The compliance of the tympanic membrane is reduced and results in drop of the height of the peak as otosclerosis progresses.

Acoustic reflex is helpful in diagnosis of otosclerosis as it measures the change in the compliance of the middle ear in response to sound stimuli. Change in the compliance of the middle ear is caused by contraction of stapedius muscles, stiffness of the ossicular chain and lead to decrease the sound transmission to the vestibule. Acoustic reflex is able

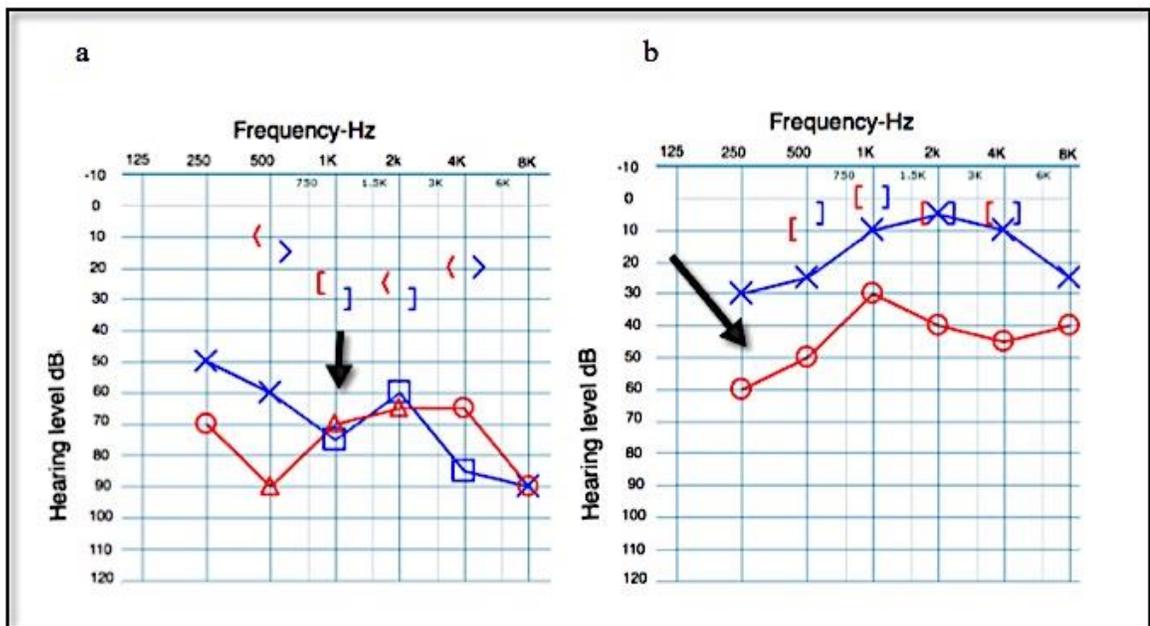
to detect otosclerosis in early stage, even before the conductive component, by the characteristic diphasic reflex pattern. Also, it detects the progression of otosclerosis over time by its characteristic pattern. As otosclerosis progresses, the acoustic reflex is reduced followed by elevation of ipsilateral (same ear) then contralateral (opposite ear) threshold and ended by absence of reflexes.

Pure tone audiometry measures the air and bone conduction, either masked or unmasked. The masking technique, which is occupying the untested ear with sound, is used to ensure that the result obtained is true for the tested ear. Normal or low thresholds of bone conduction with a higher threshold of air conduction suggests a conductive hearing loss. Conductive hearing loss is diagnosed when ABG is greater than 10 dB (Figure 1.8A). Pure tone audiometry has a better characterization of the severity of otosclerosis. The first characteristic audiometric sign of otosclerosis is a decrease in air conduction in the low frequency and is referred to as “stiffness tilt”. It is believed that it is caused by a decrease in the compliance secondary to stapes fixation. The Carhart notch, the characteristic audiologic sign of otosclerosis, is characterized by a decrease in the bone conduction thresholds at different frequencies, especially 2000Hz (Figure 1.8A). It is believed that it appears as a result of distraction of the normal ossicular resonance by stapes fixation, and is not a true representation of the cochlear dysfunction. In other words, the drop in the high frequencies is not caused by a dysfunction of the cochlea, but is the result of change in the compliance of the middle ear caused by stapes fixation. This hypothesis is supported by the fact that the Carhart notch disappears after stapedectomy.

In spite of the superiority of pure tone audiometry over tympanometry and acoustic reflexes in detecting disease progression, the measurement of pure-tone air and bone conduction thresholds has a limitation based on individual variations of test sensitivity. The variation in the test sensitivity emerges from the difference between the energy delivered by a vibrator for obtaining bone conduction thresholds, which is much greater in magnitude, and that required by headphones to deliver stimuli for air conduction threshold. It is important to apply masking technique to remove contribution of the untested ear. A combination of pure tone audiometry, acoustic reflex and tympanometry is important for accurate pre-operative diagnosis of otosclerosis<sup>106</sup>.

**Figure 1.8: Audiogram profile of an otosclerosis patient**

(a) Audiogram of a patient with late stage of otosclerosis showing a difference in the ABG. Rt ear = O, Lt ear =X, bone conduction Rt ear = [, bone conduction Lt ear =]. Characteristic Carhart notch denoted by an arrow. (b) Low frequency loss observed in a patient with early stage of otosclerosis. This loss is known as "Stiffness tilt" and is denoted by an arrow. The audiograms presented below are taken from original audiograms of one of the family probands studied in this thesis.



### 1.6.5 Genetic cause of otosclerosis

The cause for the development of sclerotic foci in the ear is unknown, though genetic factors were suggested in conjunction with the first description of familial conductive hearing loss in 1841<sup>107</sup>. Albrecht recognized AD inheritance of otosclerosis in 1922, and low penetrance (40 %) of otosclerosis was reported by Larson and Morrison in 1960 and 1967 respectively<sup>108 109 110</sup>. The most compelling evidence supporting underlying genetic factors in otosclerosis emerged in 1966 from studies of concordant monozygotic twins<sup>111,112</sup>. The most prevalent mode of inheritance is the dominant mode as suggested in previous studies<sup>113, 114, 86, 87,88, 115, 90</sup>. However, Bauer and Stain suggest a recessive mode of inheritance based on a study of 94 families with otosclerosis members<sup>116</sup>. This work has been criticized by Morrison, who claim that the data in the study lack adequate diagnostic criteria and risks the possibility of including members with other cause of hearing loss<sup>110</sup>. A digenic mode of inheritance was suggested by one study, while X-linked inheritance was proposed by another<sup>117</sup>. No mitochondrial mode of inheritance was suggested for otosclerosis transmission.

Previous genetic studies of familial form of otosclerosis have been carried out through traditional linkage analysis. The aim of these studies was to map the otosclerosis segregating in families to a region in the genome and to identify the causative mutation. To date, a literature search reveals a total of eight AD loci (*OTSC1*, *OTSC2*, *OTSC3*, *OTSC4*, *OTSC5*, *OTSC7*, *OTSC8* and *OTSC10*) that have been mapped, originally in eight large, unrelated families with different ethnic origins (Table 1.3). However, not a single gene has been identified since the first *OTSC* locus was mapped sixteen years ago<sup>113, 114, 86, 87,88, 115, 90</sup>.

**Table 1.3: Currently mapped AD otosclerosis loci and their respective genomic location**

<b>Locus name</b>	<b>Locus Ch.</b>	<b>Country of the original families</b>	<b>Subsequent families mapped to the same locus</b>	<b>Citation</b>
<i>OTSC1</i>	15q25-q26	India		113
<i>OTSC2</i>	7q34-q36	Belgium		114
<i>OTSC3</i>	6p21.3-22.3	Cyprus	Tunisian Families (n=2)	86 118
<i>OTSC4</i>	16q22.1q23.1	Israel		87
<i>OTSC5</i>	3q22-24	Dutch		88
<i>OTSC6</i>	Reserved			
<i>OTSC7</i>	6q13-16.1	Greek	Dutch Family	115
<i>OTSC8</i>	9p13.1-9q21	Tunisia		90
<i>OTSC9</i>	Reserved			
<i>OTSC10</i>	1q41-44	Dutch		80

The first advance in our understanding of the genetic aetiology of otosclerosis using traditional linkage analysis was reported in 1998, through the identification of a locus (*OTSC1*) on Chr15 which co-segregated with otosclerosis in a large family from India. This family included five of 16 affected relatives with surgically confirmed otosclerosis, and other siblings who had conductive hearing loss. Tomek et al, 1998 genotyped 160 polymorphic markers evenly dispersed across the human genome. Analysis yielded a single locus on Chr15q25-26 (*OTSC1*) that generated a logarithm of odds (LOD) score of 3.4. Construction of disease-associated haplotypes identified a minimal critical region of 14.3 cM containing 33 candidate genes<sup>113</sup>.

The second otosclerosis gene was mapped several years later in a 4 generation Belgian family with AD otosclerosis, including 10 family members with surgically confirmed otosclerosis. Multipoint linkage analysis of 391 genotyped markers yielded positive LOD scores for five candidate regions. With additional genotyping across the five regions, there was suggestive linkage to the Chr 7q34-36 region and extension of the haplotype identified the *OTSC2* critical region containing 152 genes<sup>114</sup>. In 2002, a genome-wide study on a large Cypriot family segregating otosclerosis identified *OTSC3* at Chr 6p21.3 after excluding linkage to *OTSC1* and *OTSC2* loci<sup>86</sup>. The *OTSC3* locus overlaps with the HLA region on 6p21.3-22.3 and contains 488 genes. More recently, two unrelated Tunisian families with AD otosclerosis were identified in which haplotype segregation was consistent with linkage to *OTSC3* and haplotype sharing suggested a common ancestor and single founder mutation<sup>118</sup>.

In 2006, an independent lab mapped *OTSC4* in a five generation Israeli family of Yemenite Jewish origin with AD otosclerosis and 12 affected people. After excluding *OTSC1-3*, a genome-wide scan using 400 microsatellite markers were carried out and positive LOD scores at three loci were identified. Additional genotyping in each of the three candidate regions showed that two of them no longer supported linkage. However, the third locus on Chr16q21-23.2 gave a LOD score greater than 3<sup>87</sup>. The fifth locus, *OTSC5*, at 3q22-24 was mapped in a four generation Dutch family with AD otosclerosis after linkage exclusion to previous *OTSC1*, *OTSC2*, *OTSC3* and *OTSC4* loci<sup>88</sup>. The sixth locus, *OTSC6*, has been reserved but its location in the genome has not been published<sup>12</sup>.

The seventh locus, *OTSC7*, was mapped to Chr 6q13-16.1 in a Greek family, in a genomic region spanning 66 genes. A subsequent Dutch family was identified with suggestive linkage to this locus. All genes (*KCNQ5*, *C6orf148*, *LOC653194*, *EIF3S6P1*, *DPPA5*, *ECAT1*, *LOC441161*, *RPL39P3*, *DDX43*, *C6orf150*, *MTO1*, *EEF1A1*, *SCL17A5* and *CD109*) in the overlap region between the Greek and Dutch families were sequenced, but no mutations were found<sup>115</sup>. The *OTSC7* authors suggest that either the families have independent loci or that the mutation is in a regulatory region so that deep sequencing of the region is required to solve it. This region is the first of the otosclerosis candidate regions to have the coding regions of all annotated genes sequenced.

The next locus (*OTSC8*) was mapped to the pericentric region of Chr 9 with a minimal critical region of 34.16 Mb. *OTSC8* was mapped in another Tunisian family with seven affected individuals, using genome wide scan after linkage exclusion to previously mapped *OTSC1*, *OTSC2*, *OTSC3*, *OTSC4*, *OTSC5* and *OTSC7* loci, and three candidate

genes (*COLIA1*, *COLIA2* and *NOG*)<sup>90</sup>. Recently, two additional loci have been added to the list of mapped otosclerosis loci. The first locus has been reserved for *OTSC9* but its location in the genome has not been published. The second locus (*OTSC10*) has been mapped to a region on Chr1q41-44 with a maximum LOD score of 3.3. This was confirmed in a large Dutch family after excluding linkage to the previously mapped *OTSC1*, *OTSC2*, *OTSC3*, *OTSC4*, *OTSC5*, *OTSC7*, and *OTSC8* loci<sup>12,84</sup>.

Many association studies have been carried out in search of association between candidate genes, genome-wide SNPs and otosclerosis. Association studies between each of *COLIA1* and *COLIA2* and otosclerosis were conducted previously due to similarities in the hearing loss profile and bone pathology between OI caused by mutations in *COLIA1* and *COLIA2* genes, and otosclerosis<sup>119</sup>. Significant associations between *COLIA1* and otosclerosis were reported by McKenna et al. (2004), and validated in American and German, but not in Spanish groups<sup>120 121</sup>. The association between *COLIA1* and otosclerosis was further supported in a study by Schrauwen et al who employed a large Belgian-Dutch cohort of case and controls<sup>122</sup>. No association between the *COLIA2* gene and otosclerosis were found but targeted deletion of *COLIA2* in a mouse was found to cause development of mild hearing loss and thickness in the stapes footplate<sup>119, 121</sup>. Several other genes have been implicated as a result of association studies. Transforming growth factor beta 1 (TGF- $\beta$ 1) is one important factor in bone remodeling and it helps initiation of embryogenesis of the otic capsule<sup>123</sup>. Most recently, a case-control association study on transforming growth factor beta 1 (TGF- $\beta$ 1) in a large Belgian-Dutch kindred identified a SNP (rs1800472; c.788C>T, p.T263I) that showed a

significant signal at the T allele and which was found to be underrepresented in otosclerosis patients and therefore, thought to be protective<sup>124</sup>. Genetic association with 13 candidate genes including additional members of the TGF- $\beta$ 1 pathway (*BMP2* and *BMP4*) was carried out in the same population used for detection of the TGF- $\beta$ 1 association. Two SNPs in *BMP2* and *BMP4* have shown association with otosclerosis cases<sup>125</sup>. In addition to candidate gene approach, a GWA study carried by Schrauwen et al identified seven SNPs in the *RELN* gene in which six were highly associated with otosclerosis in 694 unrelated cases from Belgian-Dutch population with otosclerosis<sup>126</sup>.

All previous association studies excluding that of Morrison et al were conducted by the same group using the same Belgian-Dutch case-control cohort population to detect an association. The cohort consisted of 694 cases of which 97.8 % (n=679) were diagnosed with otosclerosis based on surgical confirmation and a similar number (n = 694) of unrelated controls. Case and control cohorts were sex, age, and ethnically-matched. Of the 1388 subjects, only 302 cases and 302 controls were involved in the primary analysis (discovery cohort). The remaining cases and controls were used for secondary analysis (replication cohort). The sample size of the discovery cohort is considered small for an association study, especially for GWA studies. The control cohort was also randomly chosen from a DNA bank with no phenotypic characterization that could affect the statistical power of the studies. On the other hand, some of this finding was replicated in different populations. For example, the association between otosclerosis and *RELN* gene was subsequently replicated in French and Tunisian cohorts<sup>127</sup>. The replication cohorts of these studies were small in size and lacked the phenotypic assessment of the unaffected

controls. All previously listed association studies have bias regarding the sample size and phenotypic characterization of the control cohort which lower the statistical power.

Therefore, designing an association study with large sample size and proper phenotypic evaluation of the unaffected cohort will help in identifying a high quality association with low risk bias.

Despite efforts to identify the genetic factors involved in the disease pathogenicity, the disease pathophysiology remains poorly understood. Otosclerosis is genetically heterogeneous, as demonstrated by genetic mapping studies in multiplex families with an AD form of the disease. These families provide the most promise for elucidating the mechanism of this common bone remodeling disorder, yet they are very rare. The rarity of families with AD otosclerosis and the large size of the disease intervals (large number of candidate genes) are obstacles to success in identifying otosclerosis genes. The aim of this thesis was to identify causative gene for otosclerosis segregating in NL families.

## 1.7 Founder population

A founder population forms from the migration of a small group to a new area that remains isolated. The new population grows from this small group, which eventually creates a population with less genetic variation than that of the source populations<sup>128</sup>. The founder effect is potentially increased by the restriction of mating between different groups due to religion, culture, tradition, language and geographic isolation within the new population<sup>129</sup>. Founder populations increase the likelihood of identifying genes that cause genetic disease. There are many isolated populations that have helped in the discovery of novel mutations that cause hearing loss. For example, the Ashkenazi Jews population has been an isolated group in Europe for over a thousand years. The c.167delT mutation in *GJB2* was highly prevalent in this population<sup>130</sup>. Population genetic isolates are a great source for identifying some of the genetic aetiology of hearing loss<sup>131, 132, 133</sup>.

## 1.8 The NL population

NL is a Canadian province which includes the island of Newfoundland, which has many distinct genetic isolates, especially in outport communities. In NL, the majority of original settlers were from southwest England (Protestant), and southern Ireland (Roman Catholic) who settled along the coastal regions of the island for fishing purposes. Despite a lack of further immigration, the NL population increased from 19,000 to 550,000 between 1830 and early 1980s, due to large family sizes. In the mid-20th century, the average sib-ship size was still more than eight<sup>134</sup>. Religious differences between the original Protestant and Catholic settlers, and geographical distance between communities led to the formation of multiple founder clusters around the island<sup>135</sup>.

### 1.8.1 Gene discovery from the NL population

NL population has shown to be suitable for gene discovery and NL families were studied to find genes for hereditary hearing loss. For example, a mutation in the (Wolframin) *WFS1* gene that causes low-frequency hearing loss was identified using NL families. This gene was previously known to cause AR Wolframin syndrome<sup>21</sup>. Identification of the *WFS1* gene helped to identify more than 12 additional families with AD low frequency hearing loss with *WFS1* causal variants<sup>136,137</sup>. The second family that helped to find novel mutations for hearing loss in the *TMPRSS3* gene was a family from the south coast which presented with AR non-syndromic hearing loss. Two mutations in *TMPRSS3*, c.207delC and 782+3delGAG, were identified through the candidate gene approach. People with hearing loss were found to be homozygous for c.207delC or compound heterozygous for c.207delC and c.782+3delGAG<sup>138</sup>. A third Newfoundland family with AR, congenital profound, non-syndromic hearing loss had a role in the genetic identification of a novel mutation c.1583T>A in *PCDH13*<sup>139</sup>. Recently, two studies were carried out on larger extended families from NL and identified two causative genes. The first study was carried out on a multiplex family from NL and identified a novel in-frame deletion in *KCNQ4* gene<sup>27</sup>. The second study was carried out on an extended family from NL segregating X-linked hearing loss and identified a novel point deletion (c.99delC) in the *SMPX* gene that cause a frameshift and a premature stop<sup>140</sup>.

## 1.9 Hypothesis and objectives

Previous genetic studies of otosclerosis mapped ten otosclerosis loci to distinct regions in the genome. In spite of that, none of the causative genes for otosclerosis were identified. NL is a great source for genetic studies, it has shown a great success previously and many hearing loss genes were identified using large extended families from this population. Gene screening methods have been hugely improved with the development of NGS, aiming to sequence the whole genome in short period and provide answers for many disease causation. With the great progress in the new genetic technologies and availability of great foundation of multiplex families with founder origin, I hypothesised that the first otosclerosis gene can be identified in this thesis.

### Objectives

1. Examine two NL families with confirmed otosclerosis for genetic linkage to eight published loci and three candidate genes (*COL1A1*, *COL1A2* and *NOG* gene);
2. Based on evidence collected from objective 1, linked regions will be fine mapped and genes within candidate regions will be sequenced by combination of conventional Sanger sequencing and NGS.
3. Perform genome-wide linkage analysis for families that have been excluded from linkage to previously mapped loci, to identify a new region and subsequently causative gene using the same technologies mentioned in objective 2.

### **1.10 Co-authorship statement**

Nelly Abdelfatah, author of this thesis, designed the research proposal of all the genetic research sections with the guidance of Dr. Terry-Lynn Young. Nelly Abdelfatah carried out all of the experimental design, data collection and analysis of the gene discovering experiments in Chapters two, three and four. All genotyping and sequencing results were confirmed by Dante Galutira and Jim Houston (research assistants). Jim Houston sequenced four genes in Chapter 2. Post-doctoral fellow Ahmed Mostafa carried out all functional studies for Chapter 3 (Appendix 17). Nelly Abdelfatah reviewed patient clinical records under help of Anne Griffin (Audiologist) and Dr. Sue Stanton (Assistant Professor in the school of communication; Department of Science and Disorders). Anne Griffin and Sue Stanton developed the diagnostic criteria for otosclerosis. Dr. Pingzhao Hu at The Center for Applied Genomics, University of Toronto carried out statistical analysis in Chapter 4.

## **2 Chapter 2**

**Test for linkage of NL Families to published *OTSC* loci and associated candidate genes**

## **2.1 Introduction**

To date, eight loci have been mapped to the otosclerosis phenotype. Also, three genes (*COL1A1*, *COL1A2* and *NOG*) were checked for linkage to otosclerosis in previous studies (Table 1.1, page 20). To determine if otosclerosis in the NL probands was due to mutations at previously mapped/candidate genes, I used a directed approach that included genotyping for polymorphic markers within each of the published *OTSC* critical intervals, developed haplotypes at each *OTSC* locus, and/or did linkage analysis. Also, in this Chapter, I narrowed down the size of the *OTSC4* disease interval and sequenced functional candidate genes within the minimized region.

## **2.2 Subjects and methods**

### **2.2.1 Working definition of clinical otosclerosis**

Based on available otosclerosis diagnostic evaluation in the literature and in consultation with our clinical hearing scientist (Dr. Susan Stanton) and team audiologist (Anne Griffin), it was decided that a graded system could be useful in the context of familial otosclerosis. For each family member, a clinical diagnosis of otosclerosis was only assigned if the clinical presentation met the minimal diagnostic criteria as described below.

Diagnostic criteria classified as follows:

#### **Grade 1 =Otosclerosis definite: (diagnoses by surgical confirmation)**

Direct stapes fixation in the middle/inner ear at surgery.

**Grade 2 = Otosclerosis probable: (clinical diagnosis by otolaryngology)**

Diagnosis based on clinical evidence. The following evidence from clinical assessment provides well-substantiated reasons to suspect otosclerosis:

1. Audiogram with conductive component > 10 dB after 2 or more tests and at least one of the following criteria:

- Absent stapedial reflexes of affected ear ipsilateral (same ear) and contralateral (opposite ear), not explained by other types of middle ear dysfunction or solely by degree of hearing loss
- Schwarz sign on otoscopy
- Carhart's Notch on audiogram
- Low compliance of tympanic membrane on tympanogram

Each family member was required to meet minimal diagnostic criteria to be diagnosed as affected. Due to the variable age of onset of otosclerosis, family members who were at least 60 years old and had a normal hearing threshold were considered unaffected (normal).

**Hearing loss uncategorized:**

Diagnosis of hearing loss uncategorized was assigned based on the absence of any of the previously listed criteria.

### **2.2.2 Study population**

The recruitment of otosclerosis probands was done by contacting ear nose and throat (ENT) Doctors, distributing posters and circulation of a monthly magazine that was sent to recruited members and their families. Each recruited member had to complete a medical questionnaire that listed questions to cover all possible medical conditions that could help in hearing loss diagnosis (Appendix 1).

Genomic DNA from nine NL probands (index cases) with otosclerosis was available. Two probands were from multiplex families (2081 and 2114) and as genomic DNA of affected and unaffected relatives was readily obtained I started my analysis on these families first. Informed consent was obtained according to the ethical standards of the research. Of the remaining seven families, only the proband was recruited to the study. As we only had one proband for each of the seven otosclerosis families, these were used to narrow down linked regions, but not to test linkage. Ethics board of Memorial University (#01.186), and the Research Proposals Approval Committee (RPAC) of Eastern Health, St. John's, NL, and Canada.

### **2.2.3 Genotypes and haplotypes for Families 2081 and 2114**

Genomic DNA was extracted from peripheral leukocytes of all participants according to a standard method (Appendix 2). Genotyping of the microsatellite markers that spanned the critical intervals of the eight published *OTSC* loci and of the three candidate genes is shown in Table 2.1. All markers were labelled with blue fluorescent dye (6-FAM). DNA amplification PCR was done on the ABI GeneAmp 9700 thermocycler using a touchdown program (Appendices 3 and 4). After amplification, 0.5

ul of the PCR product was added to 0.5 ul of GeneScan™ LIZ ® size standard and 9 ul of Hi-Di Formamide as described in Appendix 5. PCR products were size fractionated on an ABI PRISM model 3130xl analyzer and alleles were determined using Gene Mapper version 4 software. All reagents and software were purchased from Applied Biosystems / Life Technology. All data were independently analyzed (by me and a Research Assistant; Dante Galutria) and the results compared. Primer sequences of the microsatellite markers used in this study are given in Appendix 6. Progeny (v.7) software was used to draw pedigrees. Haplotypes were built manually for the genotyped subjects according to “least recombination rules”. For Family 2081, genomic DNA(s) from seven relatives were genotyped for all the markers. The seven relatives represented four affected members with grade I otosclerosis (PIDs III-4, III-5, III-8, and III-11), one affected member with grade II otosclerosis (PID IV-5), one relative with no history of hearing loss (PID IV-10) and one relative with a history of uncategorized hearing loss (PID IV-9).

For Family 2114, three of four affected with grade I otosclerosis and one sibling with no history of hearing loss (PID III-6) were genotyped for all markers. To check for possible linkage of Family 2114 to *OTSC2* and *COL1A2* loci and to determine the maternal and paternal haplotypes, other relatives from Family 2114 (PIDs II-5, II-6, III-4 and III-5) were genotyped with microsatellite markers that spanned *OTSC2*, and *COL1A2*. To test for segregation and fine map the *OTSC4* haplotype, 15 extra DNA samples from Family 2081 (PIDs III-1, III-2, III-6, III-7, III-9, III-12, III-13, III-14, III-15, III-16 IV-1, IV-2, IV3, IV6 and IV12) and DNA samples from seven unrelated otosclerosis probands, were genotyped with the markers that spanned *OTSC4*.

**Table 2.1: Microsatellite markers used for genotyping**

Table listed microsatellite markers genotyped within the critical regions of the 8 published *OTSC* loci and within the vicinity of 3 otosclerosis associated genes (*COL1A1*, *COL1A2* and *NOG*). Boundaries markers (Boded) were taken from the original published papers. Size of each region is listed in Mb.

<b>Locus Designation</b>	<b>Location Chr</b>	<b>Size of the locus Mb</b>	<b>Markers Genotyped</b>
<i>OTSC1</i>	15q25-q26	14.5	<b><i>D15S127</i></b> , <i>D15S652</i> , <i>D15S649</i> , <i>D15S1004</i> , <i>D15S157</i> , <b><i>D15S657</i></b>
<i>OTSC2</i>	7q34-q36	16	<b><i>D7S495</i></b> , <i>D7S684</i> , <i>D7S2202</i> , <i>D7S2513</i> , <i>D7S676</i> , <i>D7S1798</i> , <i>D7S2442</i> , <i>D7S2426</i> , <b><i>D7S1827</i></b>
<i>OTSC3</i>	6p21.3-22.3	17.4	<b><i>GAAT3A06</i></b> , <i>D6S1660</i> , <i>D6S1545</i> , <i>D6S464</i> , <i>D6S273</i> <i>D6S1568</i> , <i>D6S291</i> , <i>D6S1602</i> , <b><i>D6S1680</i></b>
<i>OTSC4</i>	16q22.1q23.1	10	<b><i>D16S3107</i></b> , <i>D16S3025</i> , <i>D16S3095</i> , <i>D16S752</i> , <i>D16S3106</i> , <i>D16S3139</i> <i>D16S3018</i> , <i>D16S3115</i> , <b><i>D16S3097</i></b>
<i>OTSC5</i>	3q22-24	15.5	<b><i>D3S1292</i></b> , <i>D3S3641</i> , <i>D3S1576</i> , <i>D3S3586</i> , <i>D3S3694</i> , <i>D3S1593</i> , <i>D3S3627</i> , <b><i>D3S1744</i></b>
<i>OTSC7</i>	6q13-16.1	13.4	<b><i>D6S467</i></b> , <i>D6S280</i> , <i>D6S1596</i> , <i>D6S456</i> , <i>D6S1589</i> , <i>D6S460</i> <i>D6S1652</i> , <i>D6S1595</i> <i>D6S1613</i> , <b><i>D6S450</i></b>
<i>OTSC8</i>	9p13.1-9q21	34.6	<b><i>D9S970</i></b> , <i>D9S1844</i> , <i>D9S1862</i> , <i>D9S1879</i> , <i>D9S166</i> , <b><i>D9S1799</i></b>
<i>OTSC10</i>	1q41-44	26.1	<b><i>DIS2621</i></b> , <i>D1S439</i> , <i>D1S2800</i> and <b><i>DIS2811</i></b>
<i>COL1A1</i>	17q21.31	-	<b><i>D17S797</i></b> , <i>D17S1795</i> , <i>D17S941</i> , <i>D17S809</i> , <b><i>D17S788</i></b>
<i>COL1A2</i>	7q21.1	-	<b><i>D7S644</i></b> , <i>D7S657</i> , <i>D7S2430</i> , <i>D7S821</i> , <b><i>D7S651</i></b>
<i>NOG</i>	17q24.31	-	<b><i>D17S790</i></b> , <i>D17S1607</i> , <i>D17S1606</i> , <b><i>D17S1161</i></b>

#### **2.2.4 Sequencing genes within the candidate region**

The UCSC Genome Browser homepage<sup>141</sup> and the March 2006 assembly (NCBI build 36.1) were used to identify the genomic interval of each candidate region and to identify functional candidate genes<sup>142</sup>. Priority screening of the positional candidate genes was based on gene function, expression and plausible role in the pathogenesis of otosclerosis (e.g., expression in connective tissue, immune system, bone and involvement in bone remodeling). In this Chapter, a total of 12 genes within the minimized *OTSC4* were bidirectional sequenced. Primers sequences for the eleven genes are listed in Appendix 7.

##### **(i) Primer design and reaction amplification**

For each candidate gene, the genomic structure and alternative splice forms were determined and primer sets for all coding exons and intron-exon boundaries were designed using Primer3<sup>143</sup>. In order to determine the optimal amplification conditions for these primer sets, trial PCRs were performed. Each primer was tested using two different reaction cocktails; one with betaine and the other without betaine. Betaine acts as an enhancer and helps in the amplification, especially in regions of high GC content<sup>144</sup>. Three test DNA samples and one water control was used for each trial. Amplification reactions were run first using the touch down 54 (TD54) program (Appendices 3 and 4). PCR products from both reactions were separated and visualized on a 1 % agarose gel made with 1X TBE (Tris/ Borate/EDTA) buffer, and 5 ul of ethidium bromide or SYBER safe (from a stock solution of 10 mg/ml) per 100 ml of gel solution for a final

concentration of 0.5 ug/ml. For each reaction, 5 ul of PCR product and 1ul of bromophenol blue/xylene-cyanol dye were added to each well. A 100 bp ladder (from Invitrogen/ Life technologies) was used for determining amplicon size and the banding pattern was visualized with UV light using the Kodak Molecular Imaging System. The PCR products were purified using sephacryl HR300 (Amersham Biosciences) and MultiScreen HTS filter plates (Milipore Corporation). Purified PCR products were then bidirectionally (forward and reverse) sequenced using Big Dye Terminator V3.1 cycle sequencing kit on an ABI PRISM 3130XL DNA Analyzer, as described in Appendices 8, 9 and 10.

#### **(ii) Mutation screening panel**

The mutation screening panel included two members who inherited the presumed disease haplotype (PIDs III-4 and III-8), two members who did not, (PIDs III-9 and IV-12) and one water control. The panel was screened for the following functional genes: *ZNF19*, *ZNF23*, *COG4*, *DDX19B*, *DHX38*, *DDX19A*, *PMBP1* and *ZFH3*. Another panel included three members with the presumed disease haplotype (PIDs III-5, III-11, and IV-1), one unaffected member (PID IV-5), and one water control. This panel was screened for the following functional genes: *CABL2*, *IL34*, *SF3B3* and *WWP2*.

### **(iii) Allele frequency**

As each positional candidate gene was sequenced, the allele frequencies of the variants that segregated with the disease haplotype in the screening panel were checked. It is important to determine the allele frequency of the clinical otosclerosis before starting variant analysis. Usually frequency of the disease allele is estimated from the disease prevalence in a given population. Clinical otosclerosis has a mean prevalence of 0.5 % in the Caucasian population. Taking this number into consideration, we expect the allele frequency of otosclerosis causative variants (heterozygous variants under dominant mode of inheritance and homozygous variants under recessive mode of inheritance) would be 0.5 % in the general population.

As we do not have an estimation of the prevalence of otosclerosis in the NL population we set an allele frequency of  $\leq 2\%$  as a cut-off for variant analysis. Therefore, variants with frequencies of 2 % or less were subject to further analysis. Allele frequencies of the previously annotated variants were determined using the SNP/1000 database and allele frequencies of novel variants were determined using NL population controls which were obtained through random digit phone dialing, as part of a large colorectal cancer study. In the SNP/1000 genome database, allele frequencies of variants were listed for different populations. In this study the European population was used as reference population for allele frequency. For each variant, the sample size of the population controls analysed for this variant is provided and the allele frequency is given as a percentage of the variant positive alleles. For example, if an allele has a frequency of

45 % and the tested samples were 120 samples (240 alleles) that mean that allele was identified in 54 subjects (108 alleles) out of total 120 samples ((240 alleles).

#### **(IV) Bioinformatics analysis**

The pathogenicity of variants with an allele frequency of 2 % or less was checked using bioinformatic prediction programs. The effect of the missense mutations was predicted using SIFT<sup>47</sup> and PolyPhen-2<sup>48</sup>. Human Splicing Finder (HSF), MaxEntScan, GeneSplicer and Known constitutive signals, were used as splice prediction tools for splice site mutations. Conservation of the amino acid residue across species for the variant of interest was determined using Clustal W and WebLogo<sup>145</sup>. The effect of the insertion and deletion variants was assessed experimentally in Chapter 2.

#### **(V) Segregation analysis**

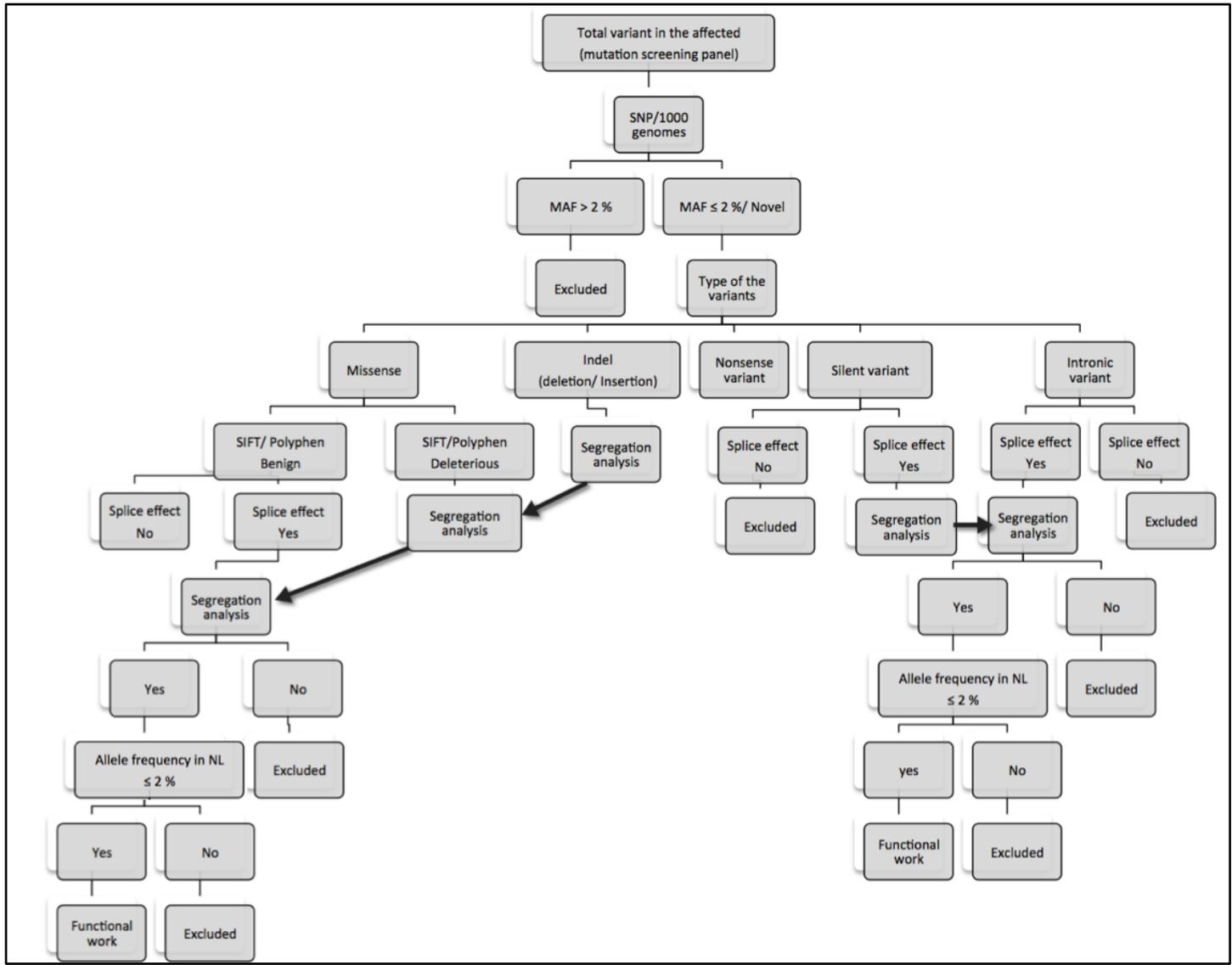
Variants predicted to have a deleterious effect on the RNA and translated proteins were checked for segregation using all available DNA in the family being studied.

## **(VI) Allele frequency in NL population**

Allele frequency of variants that fully segregated with the disease in the family of the study was estimated using NL population controls<sup>146</sup>. A summary of the variant analysis steps is illustrated in Figure 2.1.

### **Figure 2.1 : Flowchart shows filtration steps of variants detected in affected members under a dominant mode of inheritance**

Figure shows various allele frequencies of identified variants which were determined using publicly available SNP database and 1000 genomes. Pathogenicity of missense mutations was evaluated using SIFT and Polyphen prediction programs. Three splicing prediction programs (MaxEntScan, NNSPLICE, and GeneSplicer) were used for splice site prediction.



## **2.2.5 Functional assay**

### **2.2.5.1 Reverse transcriptase (RT) PCR**

To test the effect of an insertion variant on the mRNA, total RNA was extracted from Epstein-Barr virus-transformed lymphocytes from three insertion carriers (PID III-4, III-5 and III-8). RNA was isolated using Trizol reagent (Applied Biosystems / Life Technologies) and was followed by treatment with TURBO DNA-free DNase treatment (Ambion® / Life Technologies). RNA(s) were evaluated and quantified using a 2100 Bioanalyzer (Agilent Technologies) and samples with a RNA value greater than 8.5 were used. Complementary DNA (cDNA) synthesis was performed using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and Reverse transcriptase PCR was carried using primers surrounding the c.10557-10558 het insGGG in the *ZFHX3* gene. Amplification reaction was carried out for all the reactions by the ABI PCR GeneAmp 9700 thermocycler. Primers were amplified using the TD 54 program (Appendices 1 and 2) and the size fragmented on 1 % agarose gel stained with SYBER safe to confirm amplification.

### **2.2.6 Two-point linkage analysis**

For Family 2081, a two point LOD score was calculated at three microsatellite markers for each locus. The LOD score was calculated using MLINK, version 5.1<sup>147</sup>. To run MLINK program, two files were required; a pedigree file and a data file. The pedigree file included the affection status for each subject expressed in numbers; number 1 indicated unaffected, number 2 affected and 0 unknown (e.g. for a subject who has not

been examined). Also, the pedigree file included the gender code expressed as a number. For example, male =1 and female=2. The data file contained the genotype information of each marker for each subject. After loading this information into the MLINK program, the LOD scores were generated in a file called outfile data. AD with 99 % penetrance was assumed. Gene frequency was set at 0.001 and the LOD score was calculated at different recombination fraction ( $\theta$ ) frequencies ranging from 0.000 to 0.5000.

## **2.3 Results**

### **2.3.1 Pedigree structure and clinical analysis**

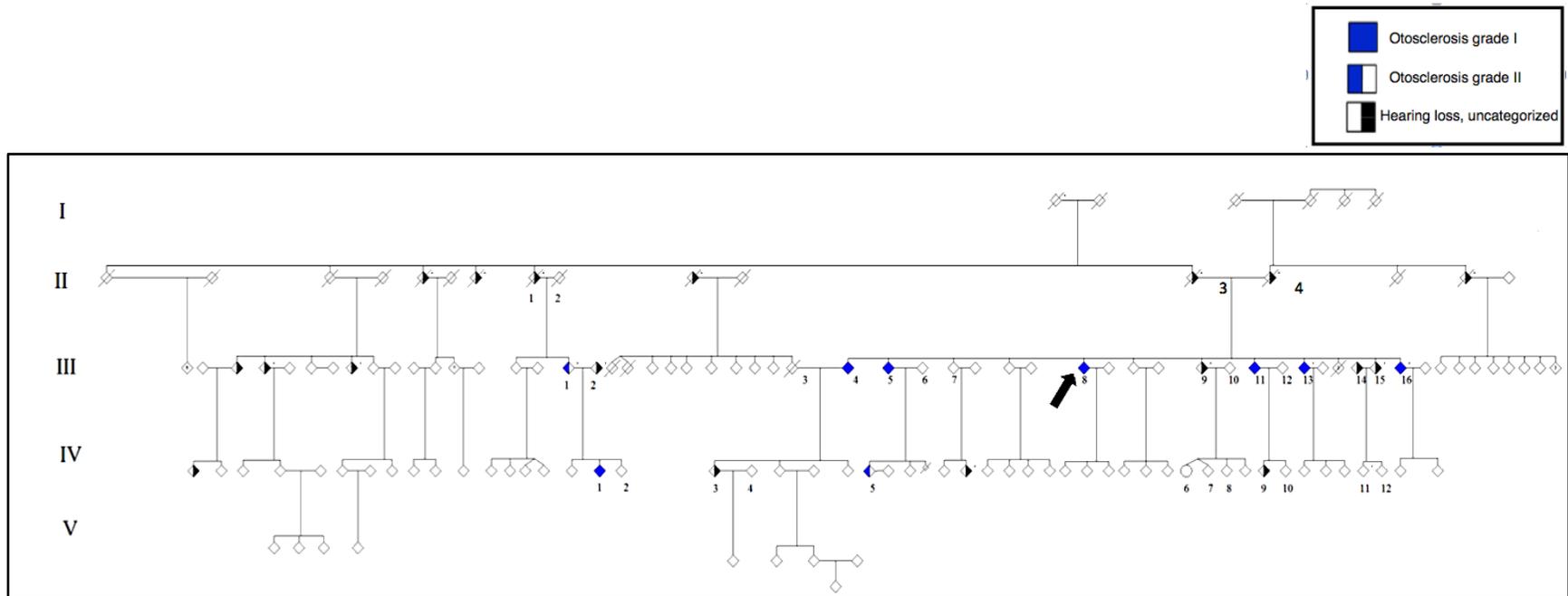
#### **2.3.1.1 Pedigree structure of Family 2081**

Family 2081 is a multiplex family with affected relatives spanning three generations where diagnosis of otosclerosis has been confirmed in both genders (Figure 2.2). A case of non-paternity in this family was identified as a result of haplotype analysis and therefore, for ethical reasons, we decided to neutralize the gender symbols in this family. The proband is PID III-8 and the proband's sibship consists of six cases of otosclerosis out of a total of 12 offspring (6/12). One of the proband's parents, PIDII-4, has a history of hearing loss and one sibling of PID II-4 also has hearing loss, although there are not confirmed cases of otosclerosis from the parental side of PID II-4. The proband's parent, PID II-4, is deceased and no further information is available from that side of the family. The other parent, PID II-3, is from a sib ship of six: four of the siblings (4/6) also have hearing loss. There is a confirmed case of otosclerosis, from subject PID II-3 side, who is the proband's cousin, PID IV-1. It is evident from the pedigree structure that otosclerosis is transmitted in a vertical pattern across generation II, III and IV. However, there is no clear evidence of X-linked inheritance. For example, the otosclerosis phenotype is similar in both males and females, as males do not appear to be more severely affected or have earlier onset. As well, if the father is the putative gene carrier, then there are documented cases of male-to-male transmission. Therefore, otosclerosis in Family 2081 appears to be inherited as an AD trait and either of the

proband's parents (PID II-3 or II-4) could be the putative gene carrier. In addition to the dominant inheritance, pseudodominant inheritance could not be excluded because both parents have a history of hearing loss within the context of the genetically isolated population of NL.

**Figure 2.2 Partial pedigree of NL Family 2081 segregating AD otosclerosis**

Figure 2.2 shows that the five generation family 2081 segregates AD otosclerosis. Genders are unidentified  $\blacklozenge$  in this pedigree for confidential reason. The arrow is points to the proband. Roman numerals on the left side of the pedigree indicate numbers of generations. The blue solid symbol refers to otosclerosis grade I. Half blue and white symbol refers to hearing loss grade II. Half black and white symbol refer to hearing loss, uncategorized. Circle =female, square =male. Crossed symbol indicates deceased subject.



### **2.3.1.2 Clinical analysis of Family 2081**

All the clinical reports were interpreted with the help of Anne Griffin and Dr Sue Stanton. The proband (PID III-8) presented at age 25 with conductive loss in the right ear, with only a slight low frequency dip in the left (Figure 2.3.A). Audiological tests were further analysed due to availability of pre and post audiological tests for both ears. At age 51, hearing loss, mostly conductive, was evident in both ears which had advanced to moderately severe in the left to severe levels in the right. The bone conduction thresholds showed a dip in the mid frequencies, known as the “Carhart’s notch”, often seen in patients with otosclerosis (Figure 2.3.B). After right stapedectomy, there was a significant improvement in the right ear and the two audiograms (post-right and pre-left stapedectomy) showed near complete resolution of the air bone gap (Figures 2.3.C and 2.3.D). After the left stapedectomy, both ears showed only a borderline-mild sensorineural loss throughout mid and low frequencies. In the high frequencies, the remaining hearing loss is greater (Figures 2.3.E and 2.3.F). At age 51, prior to both stapedectomies (Figure 2.3A), there was profound hearing loss at 8000 Hz in both ears and this did not improve at all after surgery. This could be due to otosclerosis in the cochlea just on the other side of the round window, which would affect high frequency responses the most.

The diagnosis of otosclerosis in this family was based on grading scales. Seven affected subjects in this family (PIDs III-4, III-5, III-8, III-11, III-13, III-16 and IV-1) were classified as grade I. Subject IV-5 was diagnosed with otosclerosis as grade II based

on the hearing profile, which is conductive loss with a large ABG, and the absence of a stapedial reflex with low compliance of the tympanic membrane. Family member IV-5 had developed mild to moderate left and moderate to moderately severe right conductive loss by age 36. Otosclerosis presented as conductive hearing loss in 7 members (PIDs III-4, III-8, III-11, III-13, IV-1, and III-16) and as a pure SNHL in one member (PID III-5).

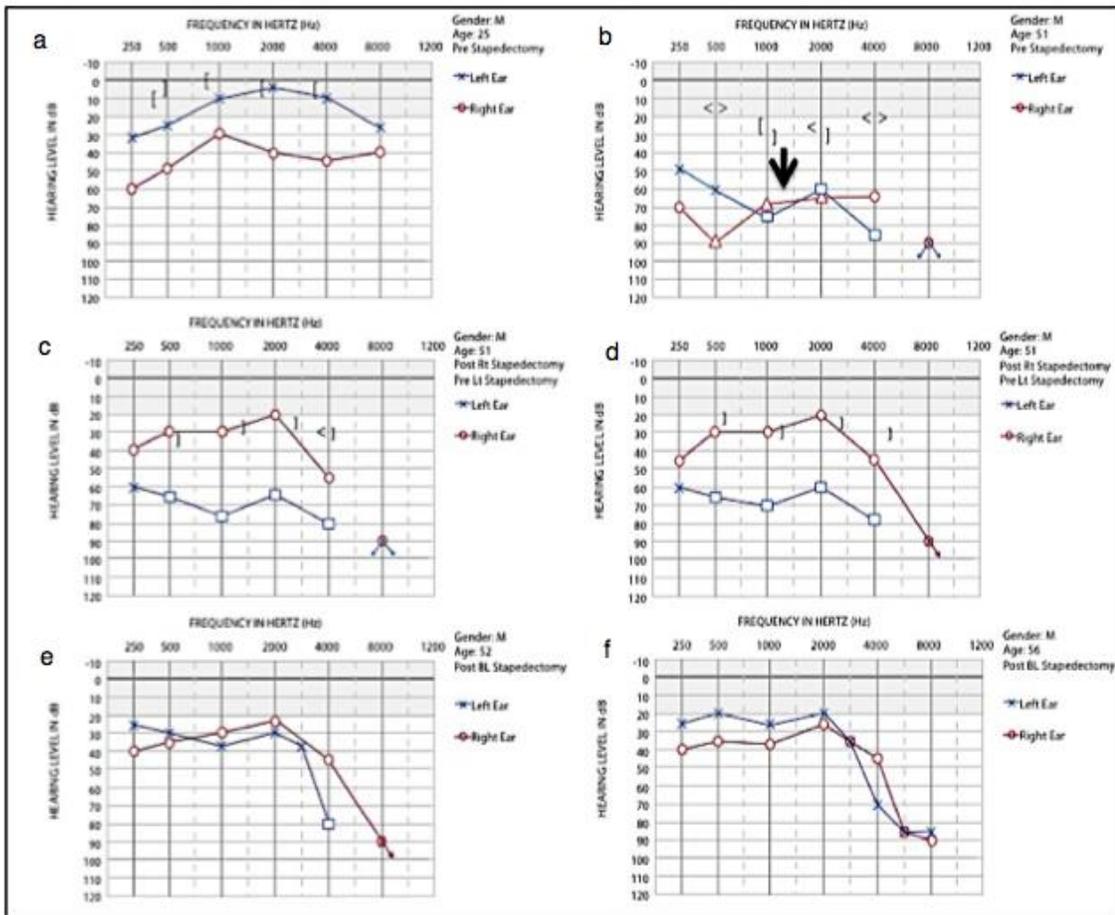
Family history indicated that the age of onset in the proband and five siblings (PIDs III-4, III-5, III-11, III-13, III-16 and IV-5) with confirmed otosclerosis ranges between mid-teens to early twenties. Audiograms of the three affected siblings (III-11, III-13 and III-16) closely resemble the proband's progression of loss, successful resolution of most conductive loss through stapedectomy in one or both ears and overall good hearing stability post-surgery. Family member (PID IV-5) who presented with hearing loss by mid-teens had a scheduled stapedectomy at 17 years. This subject had a complicated history with cholesteatoma in addition to the otosclerosis, which aggravated the degree of hearing loss. PID IV-5 received a radical tympanomastoidectomy as a treatment for the cholesteatoma and stapedectomy surgery was postponed until recovery from the tympanomastoidectomy surgery. Cholesteatoma is a pocket formed in the eardrum as result of negative pressure behind the eardrum caused by improper opening of the Eustachian tube. Similar to the skin tissue, dead skin cells slough off and fill the sac. Excess cell accumulation leads to expansion of the sac and in untreated cases, perforation of the middle ear could occur. Usually cholesteatoma is associated with conductive hearing loss, similar to otosclerosis, and treated by tympanomastoidectomy. Cholesteatoma, in some cases, is misdiagnosed as otosclerosis. Surgical exploration of the

middle ear is one way to differentiate between both conditions as stapes fixation is the definitive diagnosis of otosclerosis<sup>148</sup>. Family member PID III-4 appears to have similar early progression of conductive loss as the proband, but stapedectomy of the right ear at age 36 years was unsuccessful. Significant all frequency sensorineural loss developed along with conductive loss resulting in severe to profound mixed loss with 50 dB air bone gap bilaterally. At age 75 implantation of a vibrant sound bridge was also unsuccessful and the patient subsequently received a cochlear implant.

Family member PID III-5 reported hearing loss onset by the mid-teens but significantly differs from other affected subjects in that hearing loss was dominantly sensorineural from the start. Bilateral stapedectomies were carried out, but they were not successful. Hearing aids helped initially but were no longer effective once hearing loss developed to a profound degree bilaterally: cochlear implantation on the right was very successful. Family member IV-3 at age forty-eight showed only mild left high frequency sensorineural hearing loss which would appear consistent with noise damage incurred from a history of right handed shooting during hunting. Family member IV-10 at age thirty shows no hearing loss in either ear. Family members PID III-9, III-14 and IV-9 have history of uncategorized hearing loss. Family member PID III-7 has no history of hearing loss at 70 years old (no audiological reports were available for this subject).

**Figure 2.3: Serial audiograms for the proband of Family 2081**

Figure 2.3 shows an audiogram series for Family 2081 proband. Audiogram a indicates a conductive hearing loss bilaterally at age 25 years, progressing to a mixed hearing loss with a significant conductive component bilaterally at age 51 (Audiogram b). The improvement in post-operative hearing thresholds following right (Audiogram c) and then left ear stapedectomy (Audiogram d) indicates a successful reduction of the conductive component bilaterally as a result of stapes surgery. Serial audiograms at age 56 (Audiogram e) and 63 years (Audiogram f) show stable post-operative hearing. The arrow points to Carhart notch. O = Air conduction Rt ear, X = Air conduction Lt ear, ] = Masked bone conduction Lt ear, [ = Masked bone conduction Rt ear, < = Unmasked bone conduction Rt ear, < = Unmasked bone conduction Lt ear. Δ = Masked air conduction Rt ear, □ = Masked air conduction Lt ear.

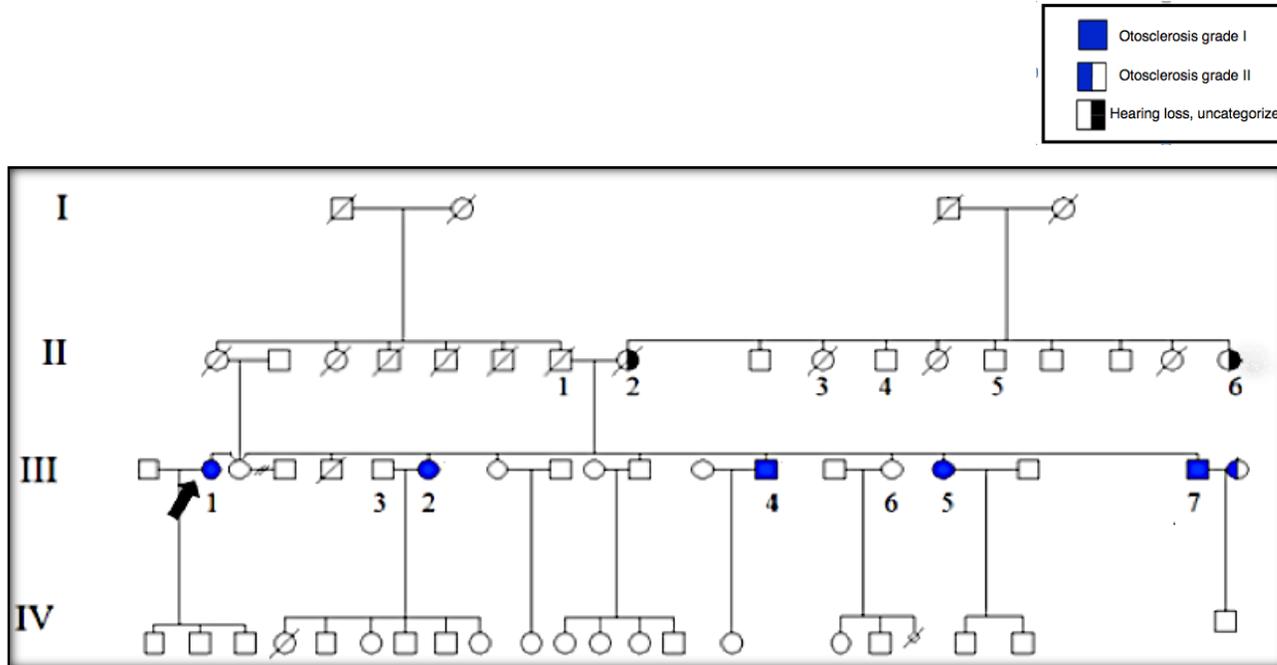


### **2.3.1.3 Pedigree structure of Family 2114**

Family 2114 is a multiplex family (Figure 2.4). The proband is PID III-1 and the proband's sibship consists of five cases of otosclerosis out of nine offspring (5/9). There is a presumed AD mode of inheritance with a history of hearing loss on the proband's mother's side. The proband's mother (PID II-2) and mother's sister (PID II-6) also have hearing loss, although not confirmed cases of otosclerosis. However the mother and three of her nine siblings are deceased and no further information is available. Although the father (PID II-1) did not have a history of hearing loss, less is known about his side of the family. Despite the fact that there are no cases of male to male transmission, the phenotype is similar across both genders so there is little evidence to support X-linked inheritance. Putting all of this information together, otosclerosis in Family 2114 appears to be inherited as an AD trait with the proband's mother the putative gene carrier. However, because we only have clinical information available on the proband's generation, we cannot rule out a recessive mode of inheritance, especially pseudodominance given the isolated NL population.

**Figure 2.4 Partial pedigree of NL Family 2114 segregating AD otosclerosis**

Figure 2.4 shows that four generation family 2114 pedigree segregates otosclerosis. Arrow is pointing to the proband of Family 2114. Roman numerals on the left side of the pedigree indicate the generation. Blue solid symbol refers to otosclerosis grade I. Half blue and white symbol refers to hearing loss grade II. Half black and white symbol refers to hearing loss, uncategorized. Crossed symbol indicates deceased subject. Circle =female, Square =male



#### **2.3.1.4 Clinical analysis of Family 2114**

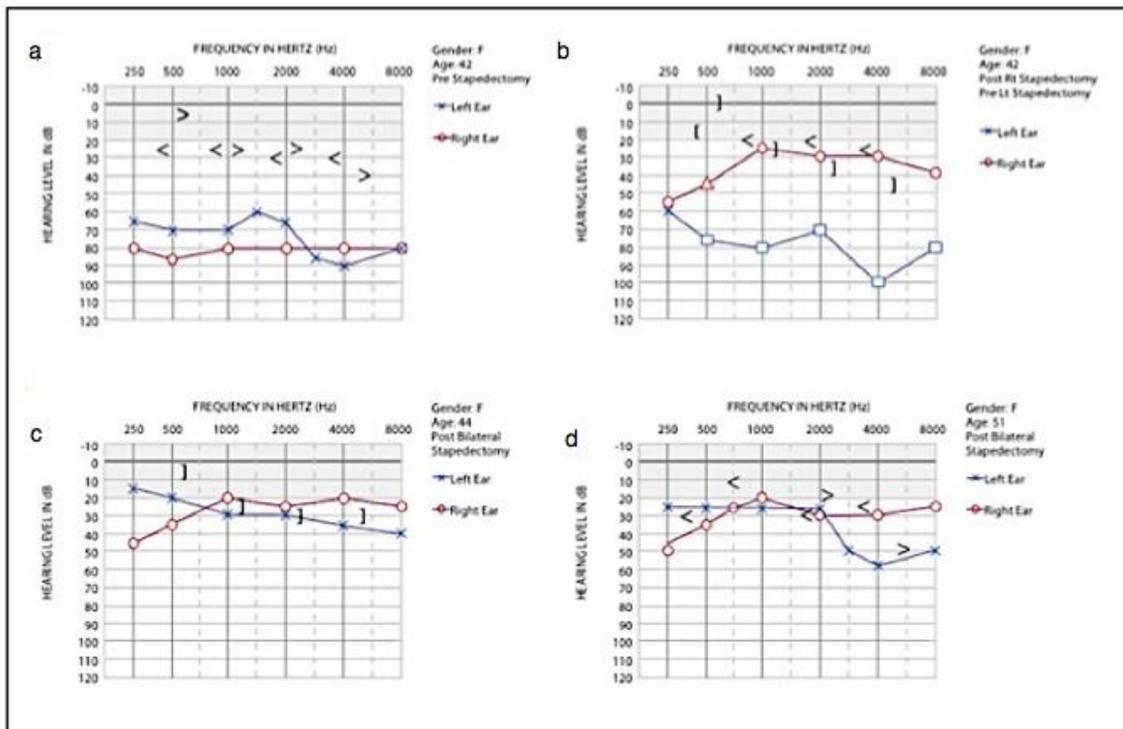
All the clinical reports were interpreted with the help of Anne Griffin and Dr Sue Stanton. The proband of Family 2114 (PID III-1, Figure 2.4) was a 42 old female, at the time of diagnosis, with severe hearing loss due to otosclerosis in both ears. Based on information taken from the medical questionnaire, she had experienced hearing loss as a teenager. An audiogram performed at the age of forty two showed that the proband's hearing loss is mostly conductive, with borderline cochlear (sensorineural) loss, which is often seen with otosclerosis (Figure 2.5.A). The proband underwent surgery at the age of 42. After undergoing stapedectomy of the right ear, hearing improved and the ABG was mainly resolved. However, some hearing loss remains, especially in the low frequencies in the right ear (Figure 2.5.B). The proband also underwent stapedectomy of her left ear at age 44, and hearing showed great improvement throughout all frequencies with only a mild sensorineural loss remaining in the mid to high frequencies (Figure 2.5.C). At age 51, the proband's left low and mid frequencies were good but left high frequencies showed moderate loss at 2000 Hz to moderately severe at 4000 Hz and 8000Hz, likely due to cochlear loss from otosclerosis on the cochlear side of the round window (Figure 2.5.D).

Otosclerosis in the five affected siblings in this family was diagnosed based on surgical visualization of stapes fixation (grade 1). Hearing loss in this family started in the teens in PIDs III-1, III-2, III-4, III-5, and around adulthood in PID III-7 (based on medical questionnaire, as no early audiograms were available). The conductive loss in all the family members was restored after stapedectomy surgery. The unaffected family

member PID III-6 has been signed as unaffected based on audiological report of no hearing loss at later age (approximately 55 years old).

**Figure 2.5: Serial audiograms for the proband of Family 2114**

Figure 2.5 shows an audiograms series for the Family 2114 proband. Audiogram a indicates a conductive hearing loss bilaterally at age 42 years with borderline sensorineural loss. Audiogram b indicates the improvement in post-operative hearing thresholds following right and then both ear stapedectomy (Audiogram c). Audiogram d indicates a successful reduction of the conductive component bilaterally as a result of stapes surgery with remaining left SNHL. O = Air conduction Rt ear, X = Air conduction Lt ear, ] = Masked bone conduction Lt ear, [ = Masked bone conduction Rt ear, < = Unmasked bone conduction Rt ear, > = Unmasked bone conduction Lt ear, Δ = Masked air conduction Rt ear, □ = Masked air conduction Lt ear.



#### **2.3.1.5 Proband of the Family 2194**

The proband of Family 2194 presented with hearing loss as a teenager (based on medical questionnaire) (Figure 2.6). Her audiological report stated that middle ear examination showed bilateral absence of acoustic reflex, which is consistent with the grade II diagnosis. The proband's mother was diagnosed with otosclerosis based on surgical confirmation of otosclerosis (grade I) of the left ear and conductive hearing loss with absence of acoustic reflex (grade II) in the right ear. Otosclerosis in this family seems to be inherited in an AD model with the proband's mother presumed to be the gene carrier. The proband is the only member of this family recruited to this study.

#### **2.3.1.6 Proband of the Family 2197**

The proband of this Family was diagnosed with otosclerosis based on surgical confirmation of otosclerosis (grade I) (Figure 2.6). Only the proband of this family was diagnosed with otosclerosis. Multiple family members were reported to have hearing loss, but no clinical information or DNA was collected.

#### **2.3.1.7 Proband of the Family 2203**

Hearing loss in the proband of Family 2203 was recorded by the second decade from the medical questionnaire (Figure 2.6). She was diagnosed with otosclerosis based on visual confirmation during surgery (grade I). Bilateral conductive components of the loss were resolved by bilateral stapedectomy, which improved her hearing acuity. Multiple subjects with uncategorized hearing loss were recognized in this family but no clinical information or DNAs were collected.

### **2.3.1.8 Proband of the Family 2066**

The proband of this Family expressed difficulty in hearing at age 16 (based on medical questionnaire) (Figure 2.6). The proband was diagnosed with otosclerosis based on surgical confirmation (grade 1). Four siblings (three deceased) of the proband were diagnosed as grade II otosclerosis. Both proband's parents have history of uncategorized hearing loss. There are multiple subjects across generations III and IV with uncategorized hearing loss. Only the proband of this family was included in the study.

### **2.3.1.9 Proband of the the Family 2126**

Hearing loss in the Family 2126 proband started bilaterally as mild to moderate conductive loss in her early twenties, which progressed to profound mixed loss. The diagnosis of otosclerosis was based on surgical confirmation (grade1) (Figure 2.6). Several subjects with uncategorized hearing loss were noticed in the family pedigree but clinical information and DNA were not collected.

### **2.3.1.10 Proband of the Family 2200**

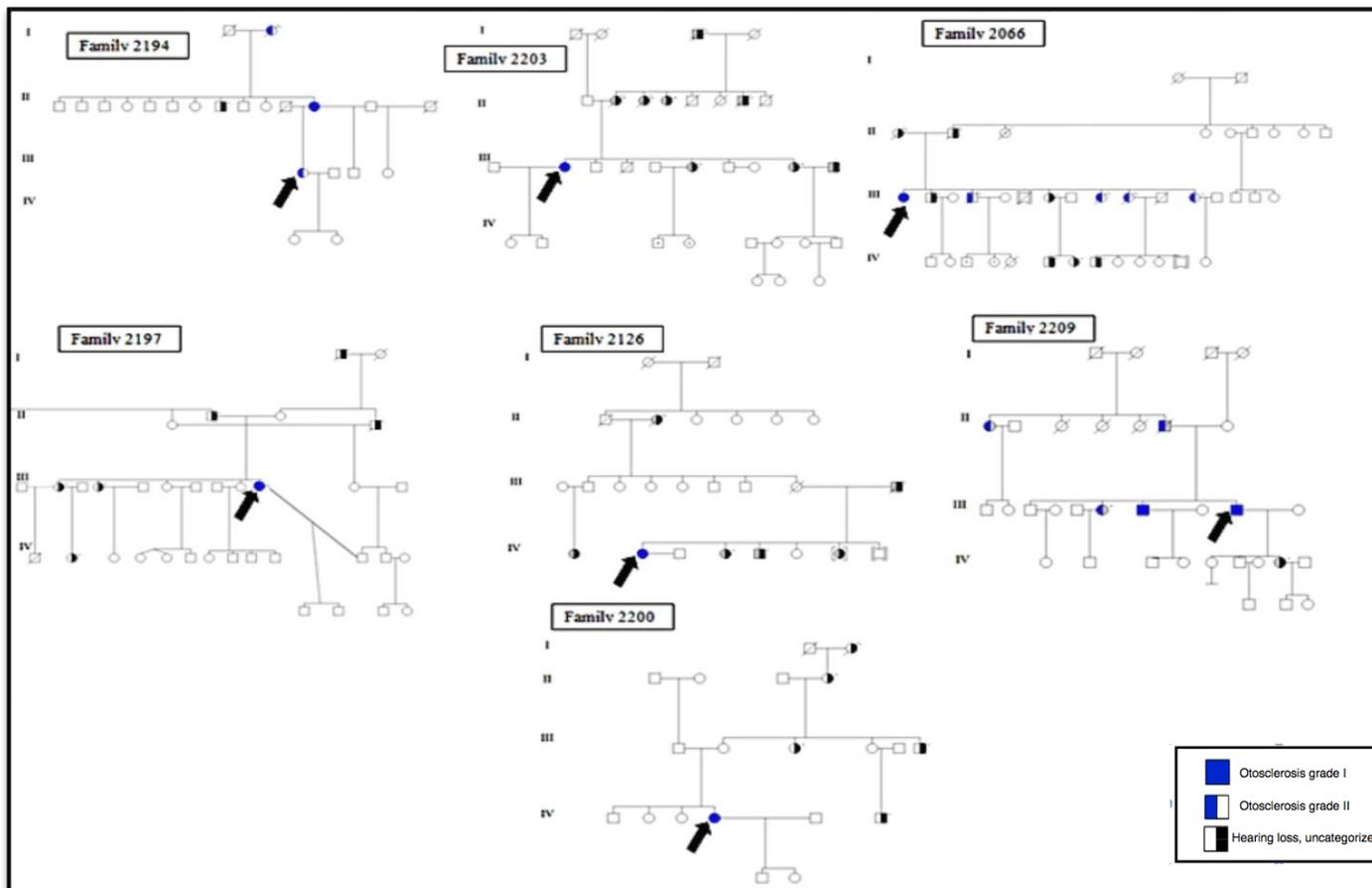
The proband of Family 2200 complained of hearing loss since age 12 (based on medical questionnaire) (Figure 2.6). Diagnosis of otosclerosis was based on surgical confirmation (grade 1). The hearing loss did not improve after unsuccessful stapedectomy surgery. The mother of the proband had many relatives with uncategorized hearing loss. Only the proband in this family was recruited to the study.

### **2.3.1.11 Proband of the Family 2209**

The proband of Family 2209 expressed hearing loss during adulthood. The proband was diagnosed with otosclerosis based on surgical confirmation (grade 1) (Figure 2.6). Otosclerosis in this family seems to be inherited in a dominant mode with the proband's father presumed to be the disease carrier, as he has been diagnosed with grade II otosclerosis. One of the proband's siblings was diagnosed with grade I otosclerosis and another sibling and aunt were diagnosed with grade II otosclerosis. We were unable to recruit other family members in this study and so only the proband is included.

**Figure 2.6: Pedigrees of Families 2149, 2203, 2066, 2197, 2126, 2209 and 2200**

Arrows point to the proband of each family. Roman numerals on the left side of the pedigree indicate the generation. Blue solid symbol refers to otosclerosis grade I. Half blue and white symbol refers to hearing loss grade II. Half black and white symbol refers to hearing loss, uncategorized. Crossed symbol indicates deceased subject. Circle =female, Square =male.



## **2.3.2 Test for linkage to published otosclerosis loci**

### **2.3.2.1 Exclusion of Family 2081 to previously mapped loci and genes**

Genetic analysis of Family 2081 was carried out by genotyping four grade I affected members with otosclerosis (PIDs III-4, III-5, III-8, and III-11), one grade II affected family member (PID IV-5), one relative with no history of hearing loss (PID IV-10) and one relative with history of uncategorized hearing loss (PID IV-9) with polymorphic markers that span the eight previously published loci and three associated genes. Haplotypes were constructed across each locus. The mode of inheritance of otosclerosis in this family is presumed to be AD with PID II-3 or PIDII-4 thought to be disease carriers. A single disease haplotype shared by grade I affected siblings PIDs III-4, III-5, III-8, III-11 was expected under a dominant mode of inheritance.

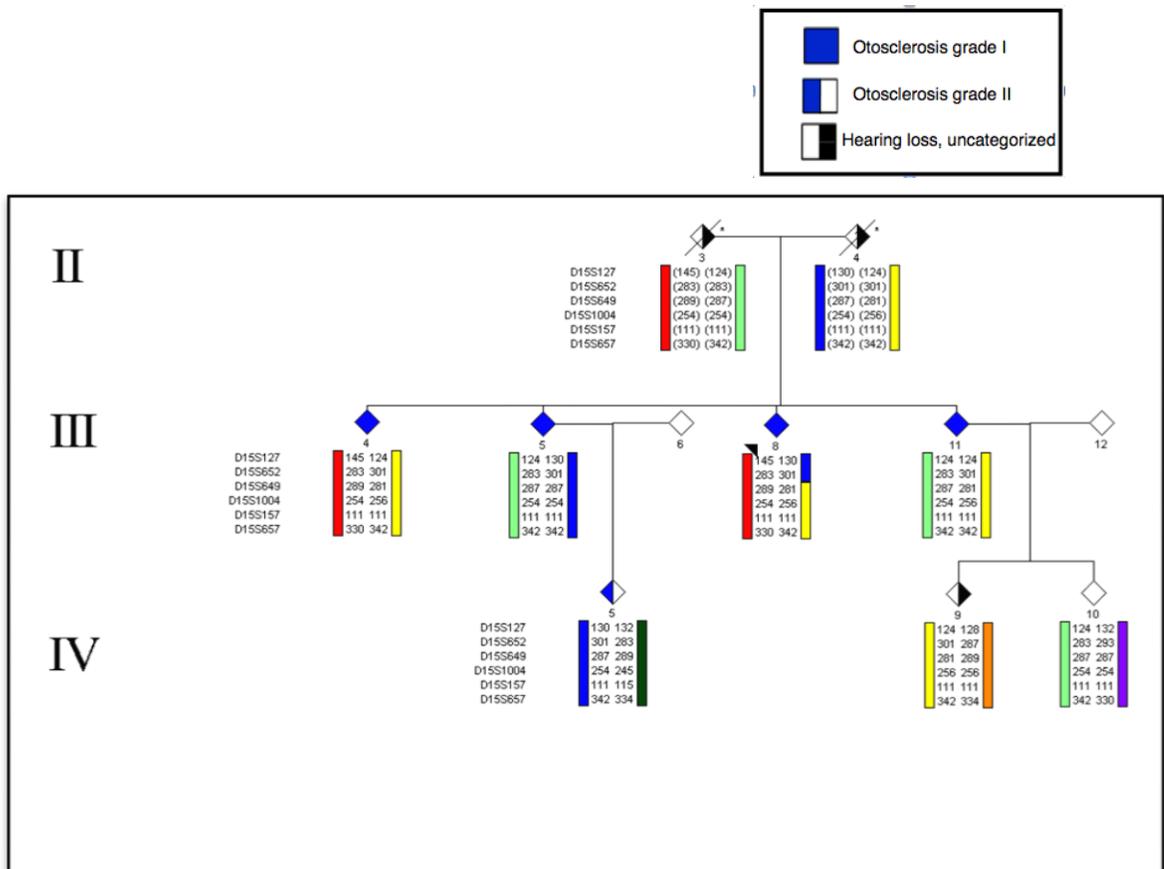
Affected siblings did not share a disease haplotype at *OTSC1*, *OTSC2*, *OTSC3*, *OTSC5*, *OTSC7*, *OTSC8*, or *OTSC10* loci nor at the three candidate genes (*COL1A1*, *COL1A2* or *NOG*) loci. The four affected siblings with grade I otosclerosis shared a presumed disease (yellow) haplotype across *OTSC4*. Below are descriptions of each locus and how the exclusion or inclusion decision was made.

## *OTSC1*

*OTSC1* is the first identified otosclerosis locus in 1998. It spans about 14.5 cM at Chr15q25-q26. For Family 2081, genomic DNA was available for seven relatives in generations III and IV (haplotypes in generation II are inferred). Six microsatellite markers were used for genotyping across this locus. All the markers were fully informative except *D15S157*, which was homozygous for the same allele in six out of seven genotyped relatives. Four distinct parental haplotypes (red, yellow, light green and blue) were created using the six informative markers (Figure 2.7). Using the rule of least number of recombinations, one recombinant haplotype on Chr15q was identified in the proband (PID III-8). All four relatives with grade I otosclerosis inherited different haplotypes, as we can distinguish between the four parental haplotypes at the *OTSC1* locus. For example, the proband's sibling PID III-4 inherited the red and yellow haplotypes. On the other hand, the proband's sibling PID III-5 inherited the light green and blue haplotypes. This locus was ruled out because the four siblings with the grade I otosclerosis did not share a disease haplotype across *OTSC1*.

**Figure 2.7: Partial pedigree of Family 2081 showing *OTSC1* haplotypes**

No shared haplotype was detected between affected members. Family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. Blue solid symbol refers to otosclerosis grade I. Half blue and white symbol refers to hearing loss grade II. Half black and white symbol refers to hearing loss, uncategorized. Crossed symbol indicates deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in the brackets are inferred.

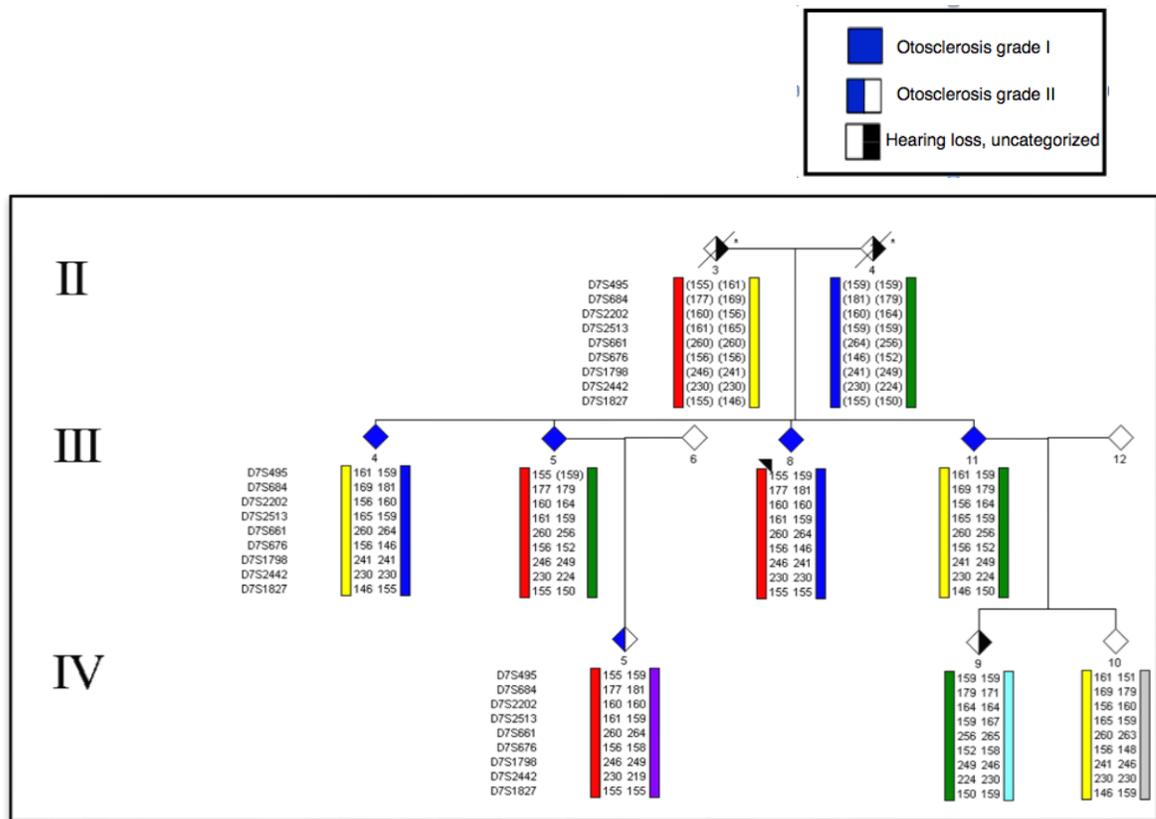


## ***OTSC2***

*OTSC2* is the second identified locus in 2001. It spans about 16 cM interval at Chr7q34-q36. Nine fully informative microsatellite markers were used for genotyping at this locus. Four distinct parental haplotypes (red, yellow, blue and green) were created (parental haplotypes were inferred and no recombinations invoked) (Figure 2.8). All four relatives with grade I otosclerosis inherited different haplotypes at the *OTSC2* locus. For example, the proband PID III-8 inherited the red and blue parental haplotypes. On the other hand, proband's sibling PID III-11 inherited the green and yellow haplotypes. Also, the three affected family members (PID III-5, III-8 and IV-5) inherited the parental red haplotype, but the remaining two affected members (PID III-4 and III-11) did not. Affected siblings did not share a disease haplotype across *OTSC2*, therefore, linkage of Family 2081 to this locus was ruled out.

**Figure 2.8: Partial pedigree of Family 2081 showing *OTSC2* haplotypes**

No shared haplotype was detected between affected members. Family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. Blue solid symbol refers to otosclerosis grade I. Half blue and white symbol refers to hearing loss grade II. Half black and white symbol refers to hearing loss, uncategorized. Crossed symbol indicates deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in the brackets are inferred.

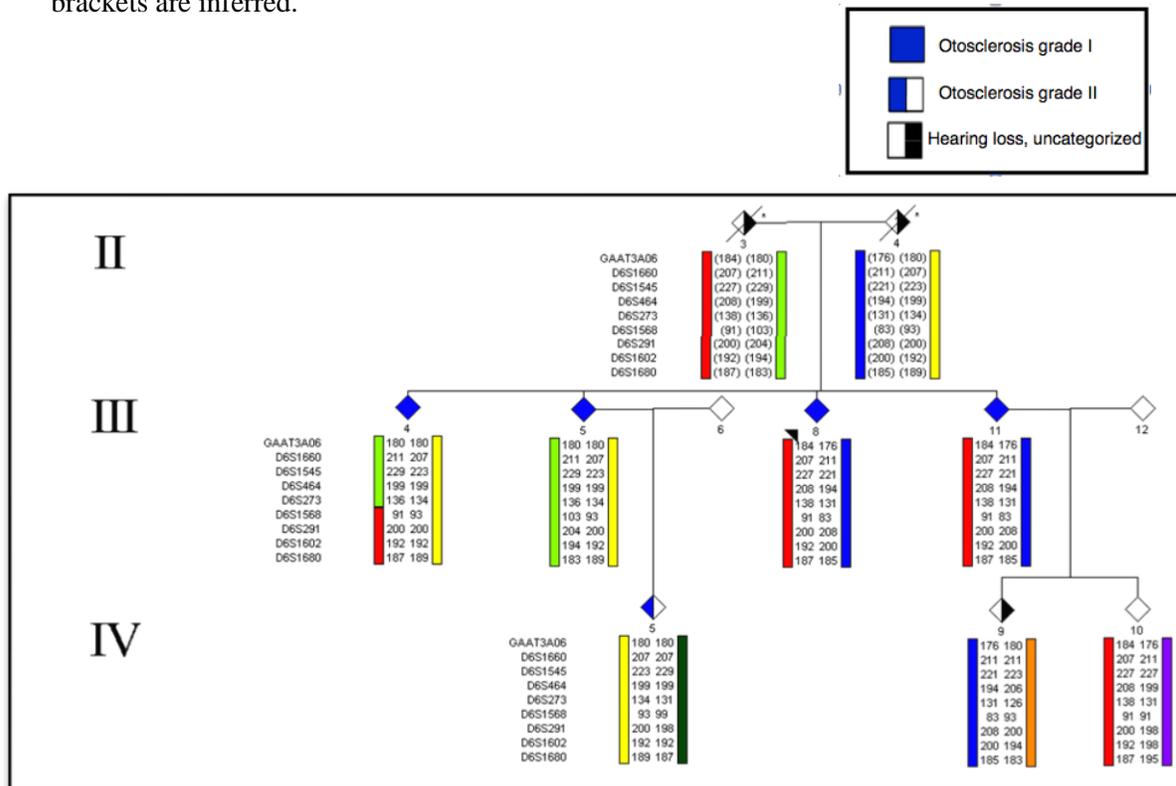


### ***OTSC3***

*OTSC3* is the third identified locus in 2002. It spans 17.4 cM interval at Chr 6p21.3-22.3. Nine microsatellite markers were used for genotyping at this locus and four distinct parental haplotypes (red, blue, light green and yellow) were created (Figure 2.9). Using the rule of least number of recombinations, one recombination was detected between *D6S273* and *D6S1568* in the proband sibling (PID III-4). Comparing haplotypes between the four affected siblings showed that they inherited different haplotypes. For example, the family proband PID III-8 inherited the blue and red parental haplotypes. On the other hand, the proband sibling PID III-5 inherited the light green and yellow haplotypes. Also, only three (PID III-4, III-5 and IV-5) out of five affected members inherited the parental yellow haplotype. The four affected siblings in Family 2081 did not share a disease haplotype across *OTSC3* and they were therefore excluded from linkage to this locus.

**Figure 2.9: Partial pedigree of Family 2081 showing *OTSC3* haplotypes**

Affected members did not share a single haplotype. The family proband is PID III-8 Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in the brackets are inferred.

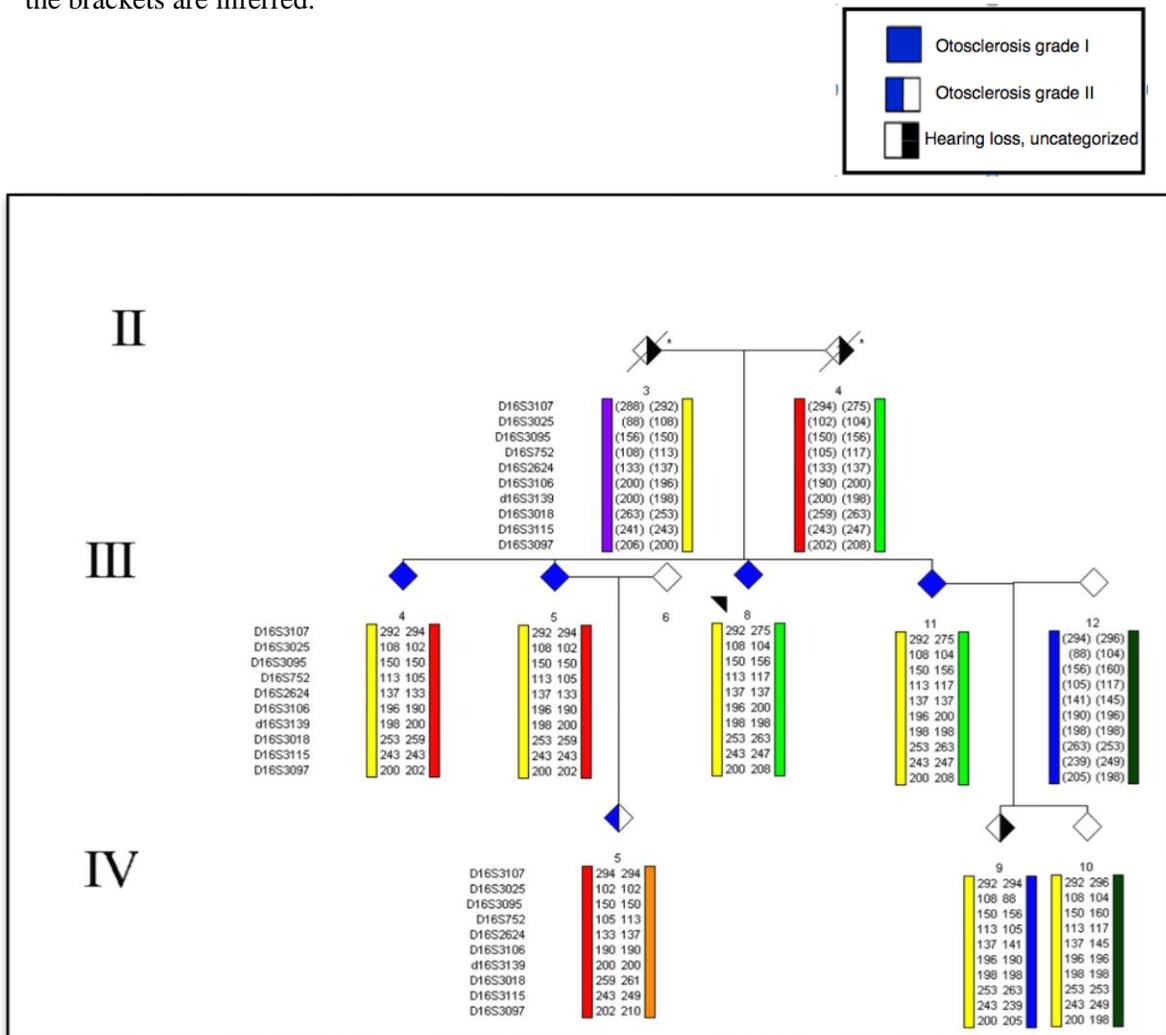


### ***OTSC4***

*OTSC4*, the fourth locus, was mapped in 2006. It spans a 9-10 Mb region at Chr16q22.1-q23. Ten microsatellite markers genotyped for this locus resulted in four distinct parental haplotypes (purple, light green, yellow and red). Parental haplotypes were inferred and no recombination was invoked (Figure 2.10). Comparing *OTSC4* haplotypes across the four grade I affected siblings showed that they shared a yellow haplotype across *OTSC4*. This locus was then subjected to further fine mapping analysis.

**Figure 2.10: Partial pedigree of Family 2081 showing *OTSC4* haplotypes**

Affected siblings in generation III share a yellow haplotype. The family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in the brackets are inferred.

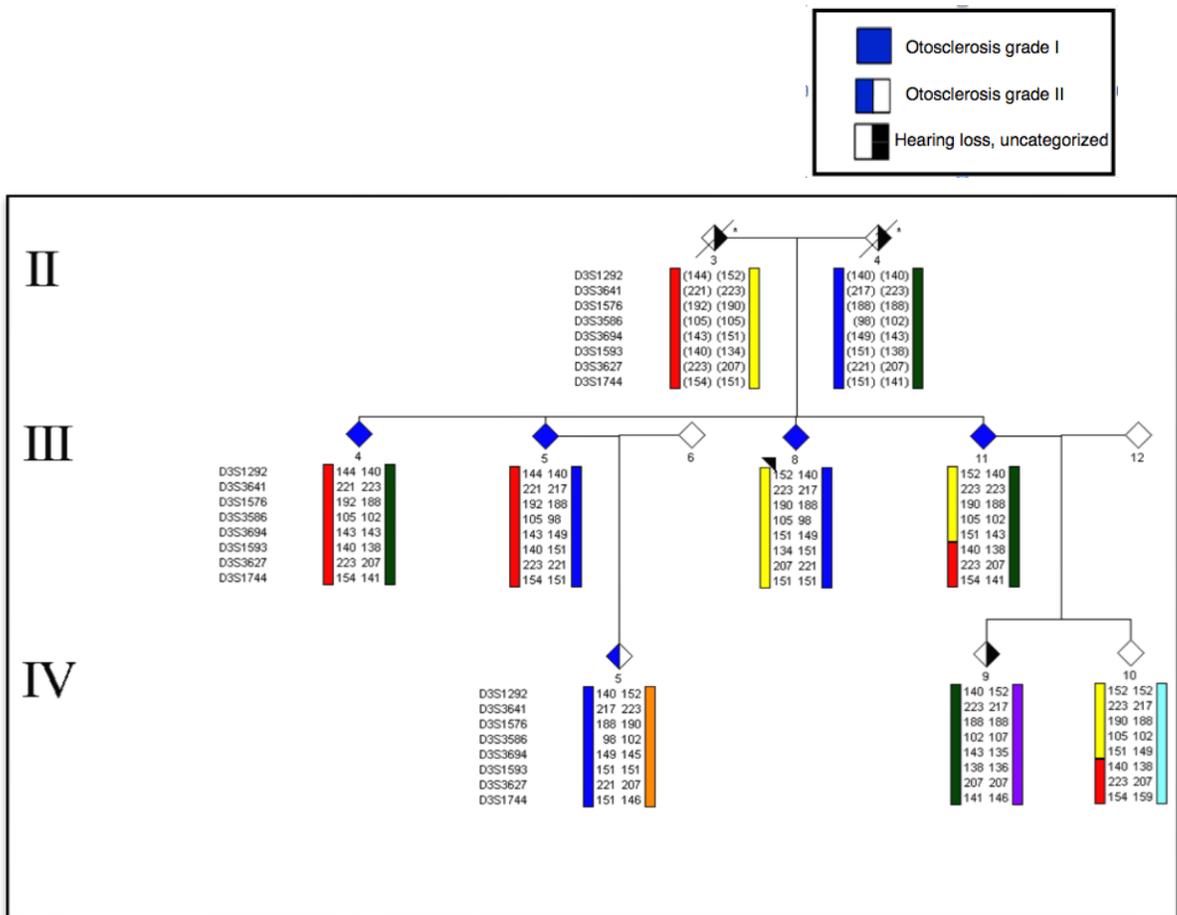


### ***OTSC5***

*OTSC5*, the fifth locus, was mapped in 2004. It spans a 15.5 Mb region at Chr 3q22-q24. Eight informative microsatellite markers were used for genotyping at this locus and four distinct parental haplotypes (red, blue, dark green and yellow) were created (Figure 2.11). Using the rule of least number of recombinations, one recombination between *D3S3694* and *D3S1593* was identified in the Family member PID III-11 which was passed down to the family member PID IV-10. The four affected siblings inherited different haplotype. For example, the proband PID III-8, inherited the yellow and blue haplotypes, while the proband's sibling PID III-4, inherited the dark green and red haplotype. The grade I affected siblings did not share a disease haplotype across this locus, and therefore, it was excluded from linkage.

**Figure 2.11: Partial pedigree of Family 2081 showing *OTSC5* haplotypes**

No shared haplotype was detected between affected members. The family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in the brackets are inferred.

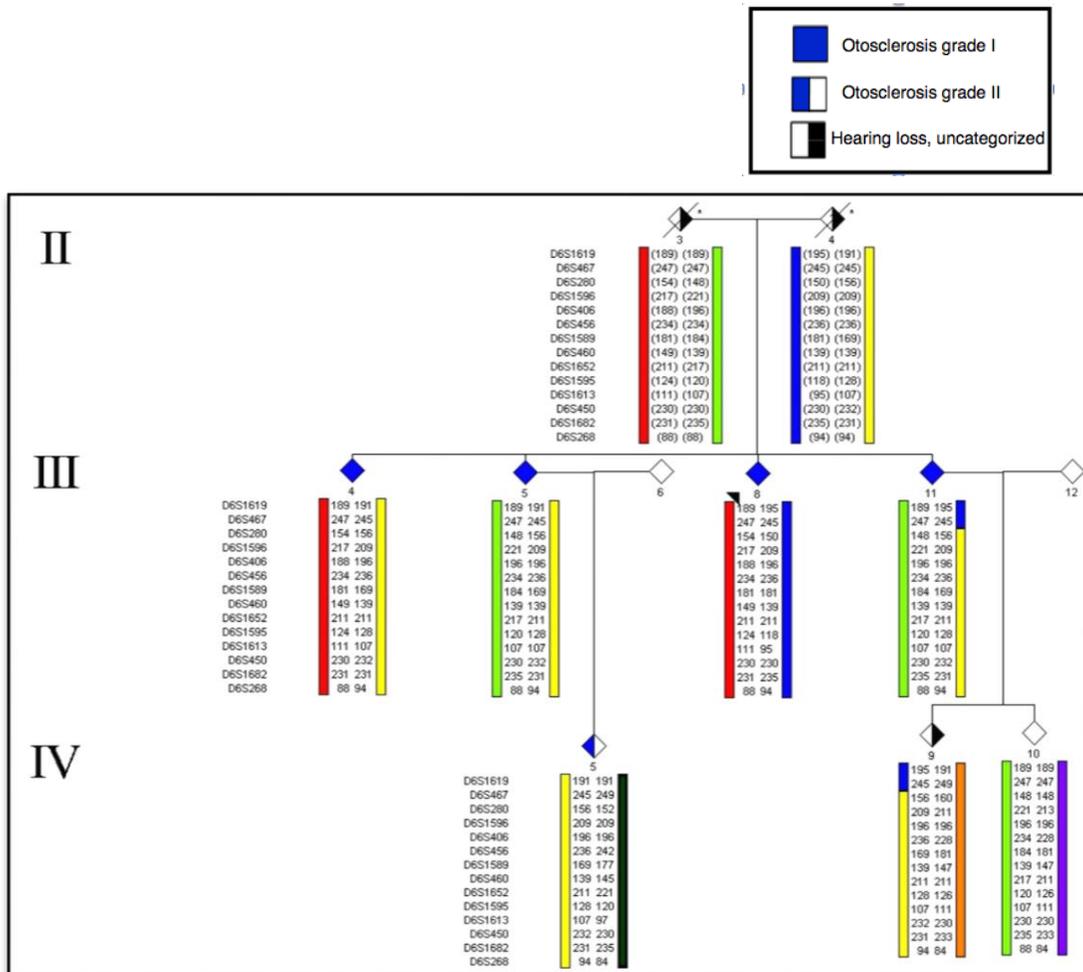


### ***OTSC7***

*OTSC7*, the seventh locus, was mapped in 2007. It spans a 13.47 Mb region at Chr 6 q13-16. Fourteen markers were genotyped for this locus and four distinct parental haplotypes (red, light green, blue and yellow) were created (Figure 2.12). Using the rule of least number of recombinations, one recombination between *D6S467* and *D6S280* was identified in PID III-11 and passed to PID IV-9. The four affected siblings inherited different haplotypes. For example, the proband PID III-8 inherited the blue and red haplotypes, while the proband's sibling PID III-5 inherited the light green and the yellow haplotypes. The three affected family members (PID III-4, III-5, and IV-5) inherited the yellow parental haplotype, affected member (PID III-11) inherited a large portion of the yellow haplotype and a small portion of the blue haplotype in a recombination but the proband (PID III-8) did not. As the four affected siblings with grade I otosclerosis did not share a disease haplotype across this region, *OTSC7* was excluded from linkage to the otosclerosis segregating in Family 2081.

**Figure 2.12: Partial pedigree of Family 2081 showing *OTSC7* haplotypes**

Affected members did not share a disease haplotype. The family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in the brackets are inferred.

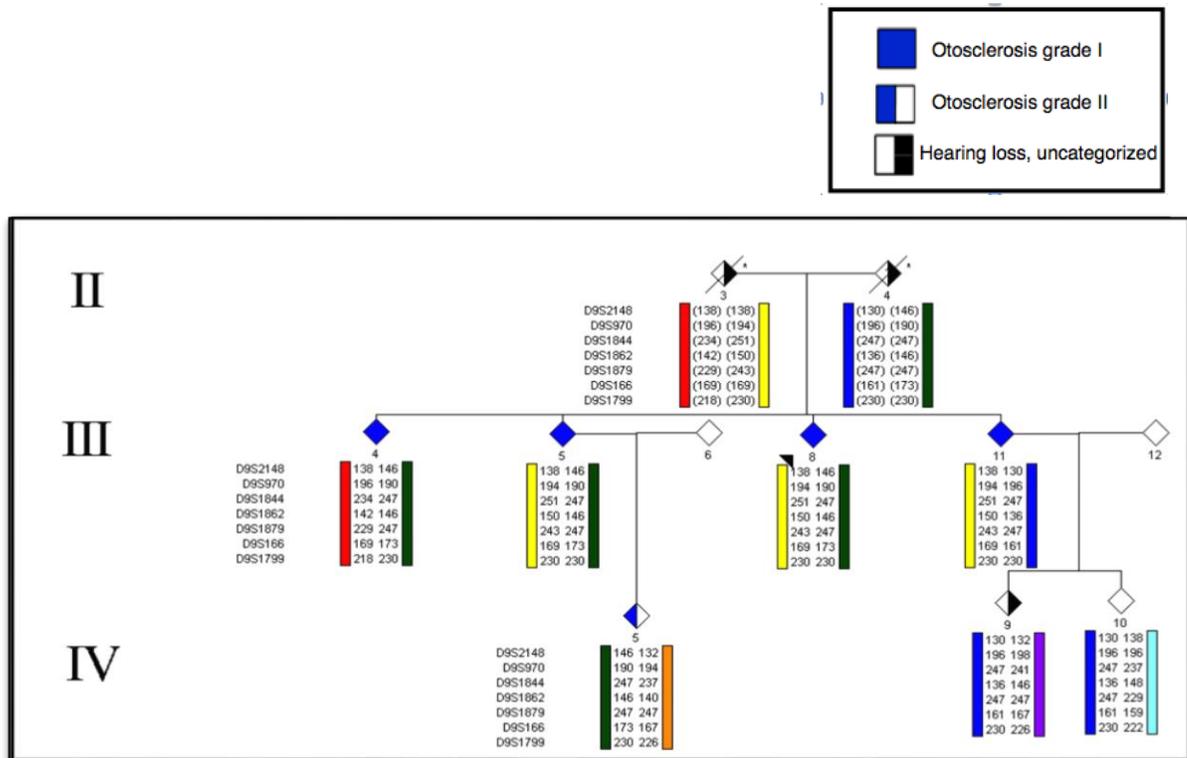


### ***OTSC8***

*OTSC8*, the eighth locus, was mapped in 2008. It spans a 34.16 Mb region at Chr 9p13.1-9q21. Seven microsatellite markers were genotyped at this locus and four distinct parental haplotypes (red, blue, dark green and yellow) were created with no recombinations detected (Figure 2.13). At this locus, three affected siblings (PIDs III-4, III-5, III-8 and IV-5) inherited the parental green haplotype but not the fourth affected sibling, PID III-11. Also, the three affected siblings, PIDs III-5, III-8 and III-11, inherited the yellow haplotype but not the affected sibling, PID III-4. As there was no shared disease haplotype identified among the four grade1 affected siblings this locus was excluded from linkage to otosclerosis segregating in Family 2081.

**Figure 2.13 Partial pedigree of Family 2081 showing *OTSC8* haplotypes**

No shared haplotype was detected between affected members. The Family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.



## ***OTSC10***

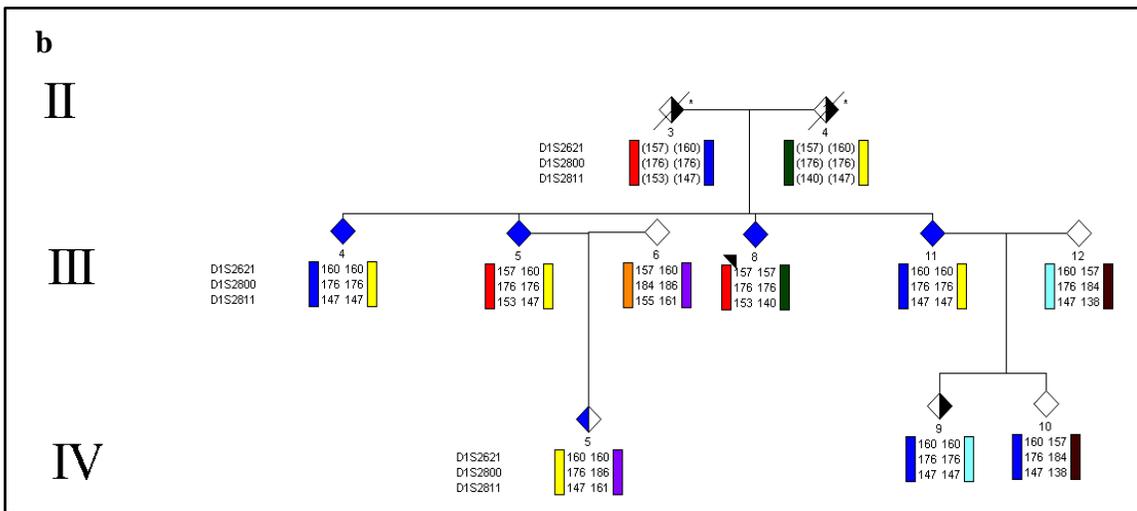
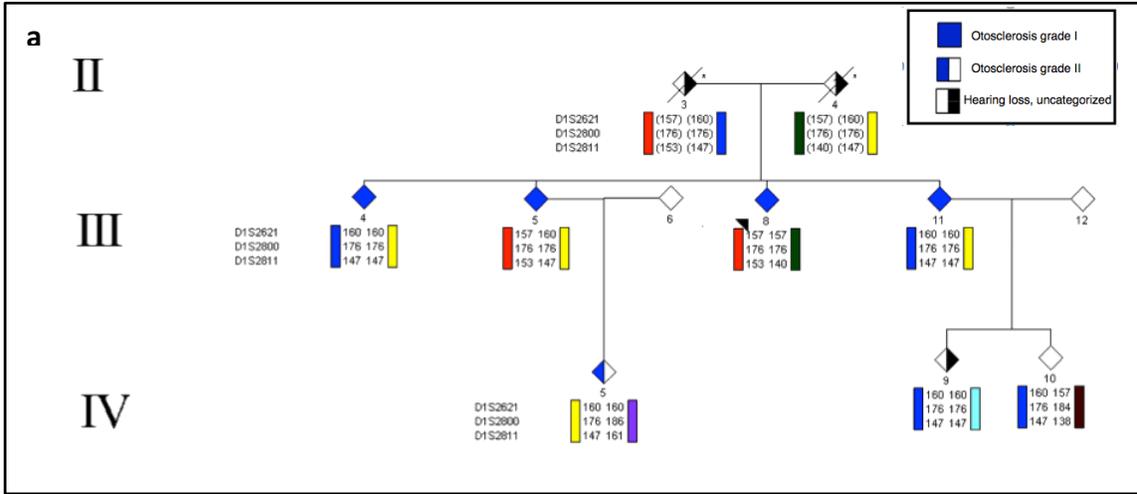
*OTSC10* is the most recently mapped locus (2010). It spans a 26.1 Mb region at Chr1q41-44. Three markers were genotyped for this locus. *DIS2621*, *DIS2811* and *DIS2800* were homozygous for the same allele in five of seven genotyped relatives. PID III-4, III-5, III-8, III-11, IV-5, IV-9 and IV-10 were genotyped for these markers and four distinct parental haplotypes (blue, red, dark green and yellow) were created (Figure 2.14A). It is obvious that the yellow and blue haplotypes have the same allele calls for the genotyping using the three markers. In order to phase each of the yellow and the blue haplotypes, two spouses PIDs III-6 and III-12 were genotyped for the three markers.

Genotyping of PID III-6 helped to confirm the yellow haplotype of PID III-5 as it showed that the yellow haplotype of PID IV-5 was inherited from PID III-5 (Figure 2.14 B). In the case of the PID III-11, III-12, IV-9 and IV-10, it was not possible to differentiate between the yellow and the blue haplotypes as PID III-11 inherited both haplotypes (yellow and blue) and the offspring PID IV-9 and IV-10 inherited either the blue or the yellow haplotype. Genotyping of PID III-12 gave two distinct haplotypes (brown and light blue). Interestingly, the light blue haplotype was similar to the blue/yellow haplotype of PID III-11. This result suggested that the yellow, blue, and light blue haplotypes are probably copies of a common haplotype in the NL population and it would not be a good candidate for being presumed as a disease haplotype. Even if we considered either haplotype (yellow or blue) as a disease haplotype, the four affected siblings did not share the yellow/ blue haplotypes. For example, the proband, PID III-8, inherited the red and dark green haplotypes, while the proband's sibling, PID III-4, inherited the blue and

yellow haplotypes. Also, it was noticed that the four affected family members, PID III-4, III-5, III-11 and IV-5, inherited the yellow haplotype but the proband did not. In summary, affected siblings of Family 2081 did not share a single haplotype across *OTSC10* and therefore this locus was excluded from linkage.

**Figure 2.14: Partial pedigree of Family 2081 showing *OTSC10* haplotypes**

The family proband is PID III-8. (a) Segregation of *OTSC10* in the core Family 2081. (b) Two spouses (III-6 and III12) are included in the analysis. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.

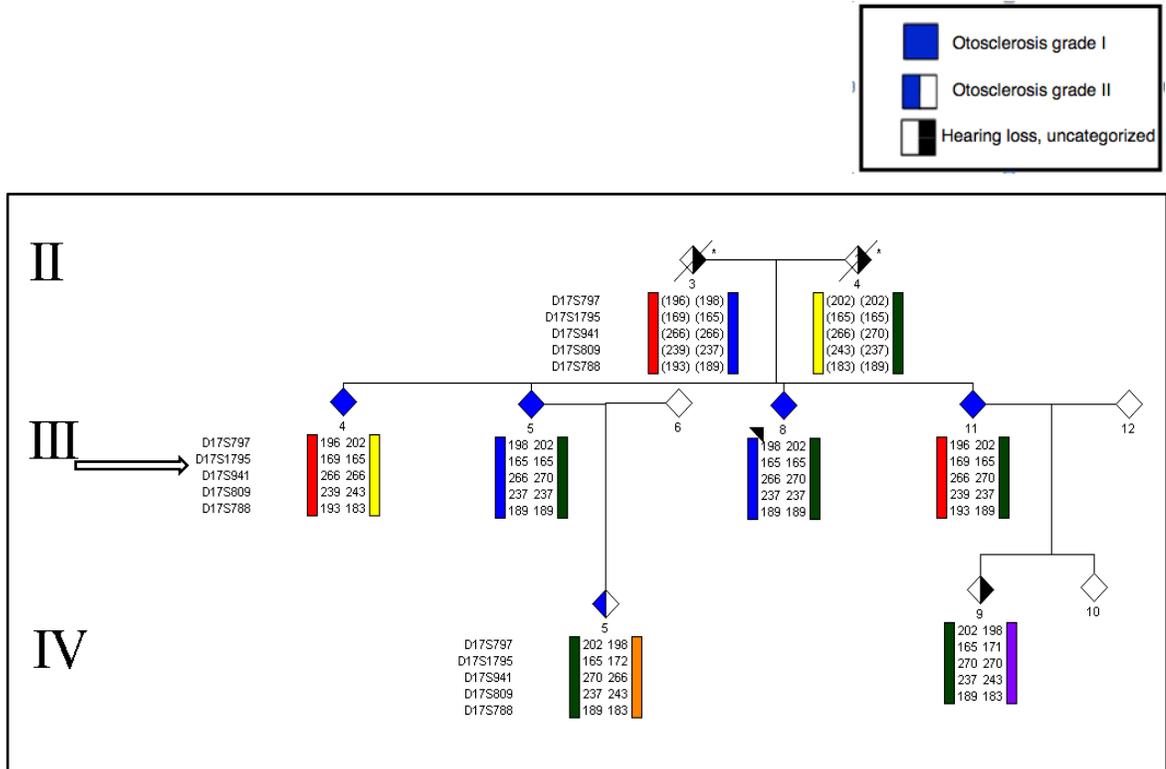


### *COL1A1*

*COL1A1*, one of the candidate genes for otosclerosis, is a large gene containing 51 exons located at Chr17q21.31. Five microsatellite markers that covered the region surrounding the *COL1A1* were genotyped and four distinct parental haplotypes (blue, red, dark green and yellow) were constructed (Figure 2.15). *COL1A1* is located between *D17S1795* and *D17S941*. All four affected siblings inherited different haplotypes across this region. For example, the proband, PID III-8, inherited the blue and dark green haplotypes, as did PID III-5 whereas, the proband's sibling, PID III-4, inherited the yellow and red haplotypes. Also, the four affected siblings, PID III-5, III-8, III-11 and IV-5, inherited the parental dark green haplotype, whereas the affected member, PID III-4, did not. As the four affected siblings did not share a single haplotype across *COL1A1*, linkage of this gene to otosclerosis segregating in Family 2081 was ruled out.

**Figure 2.15: Partial pedigree of Family 2081 showing *COL1A1* haplotypes**

No shared region across *COL1A1* among affected members. The family proband is PID III-8. The arrow points to the location of the *COL1A1* gene within the 17q haplotype. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.



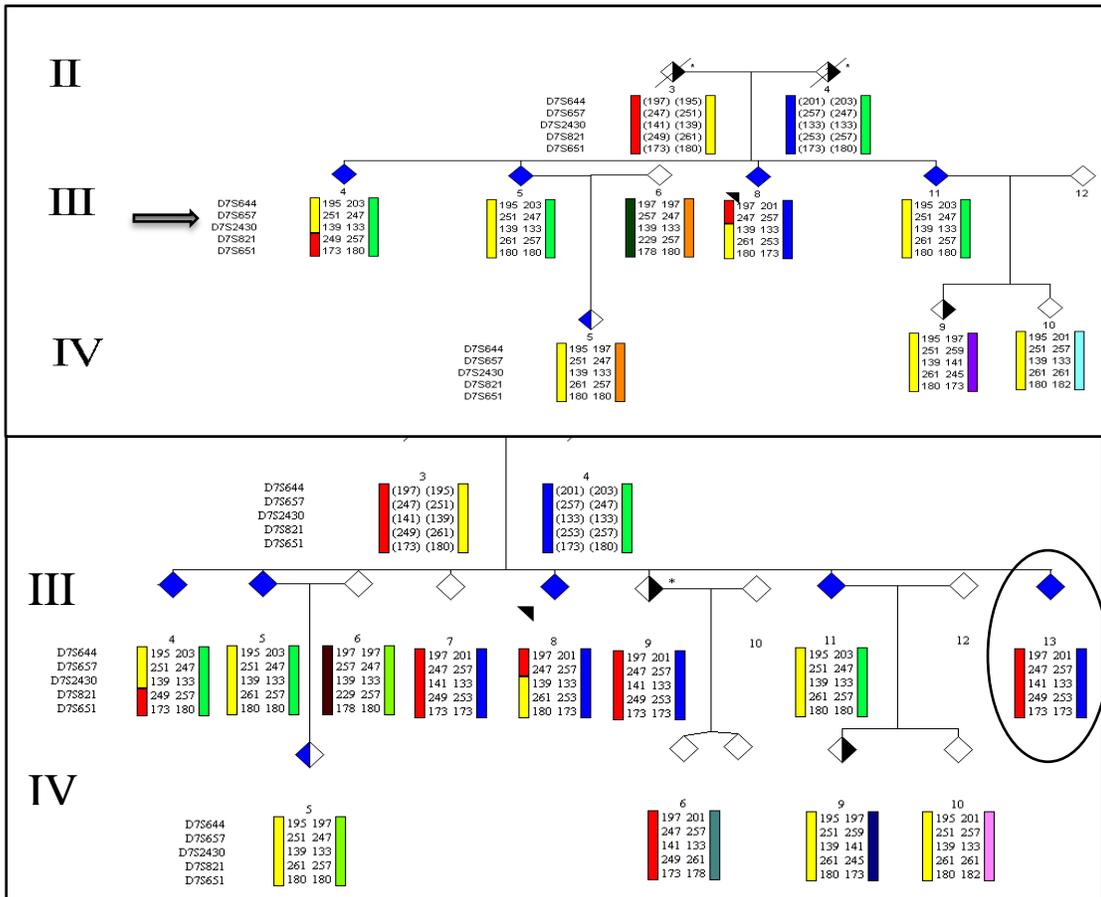
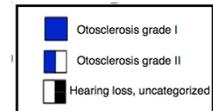
## *COLIA2*

*COLIA2*, another candidate gene for otosclerosis, is a large gene containing 52 exons located at Chr7q21.1. Five microsatellite markers covering the region surrounding *COLIA2* were used for genotyping and four distinct parental haplotypes (red, yellow, light green and blue) were constructed (Figure 2.16A). Using the rule of least number of recombinations, a recombination between *D16S2430* and *D16S821* in PID III-4, and a second recombination between *D7S657* and *D16S2430* in PID III-8 was detected. The four-affected siblings, PIDs II-4, III-5, III 8 and III-11, shared a portion of the yellow haplotype between two markers (*D7S657* and *D7S821*) where the *COLIA2* gene is located.

To validate this result, one unaffected family member PID III-7, unaffected spouse PID III-6, one family member with uncharacterized hearing loss PID III-9, and one family member with grade I otosclerosis PID III-13, were genotyped with the five markers that cover the *COLIA2* gene. Haplotype construction determined that PID III-13 inherited the parental red and blue haplotypes and did not share the yellow haplotype with the other four affected siblings (Figure 2.16B). This excluded linkage of Family 2081 to *COLIA2*.

**Figure 2.16: Partial pedigree of Family 2081 showing *COLIA2* haplotypes**

Upper panel shows segregation of *COLIA2* haplotypes in eight members. The lower panel shows segregation of *COLIA2* haplotypes within the extended pedigree. The affected member (PID III-13) who did not inherit the yellow haplotype is circled. The arrow points to the location of the *COLIA2* gene within the 7q haplotype. The family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates deceased subject. Circle = female, Square = male. Allele sizes are given in base pairs and alleles in brackets are inferred.



### ***NOG gene***

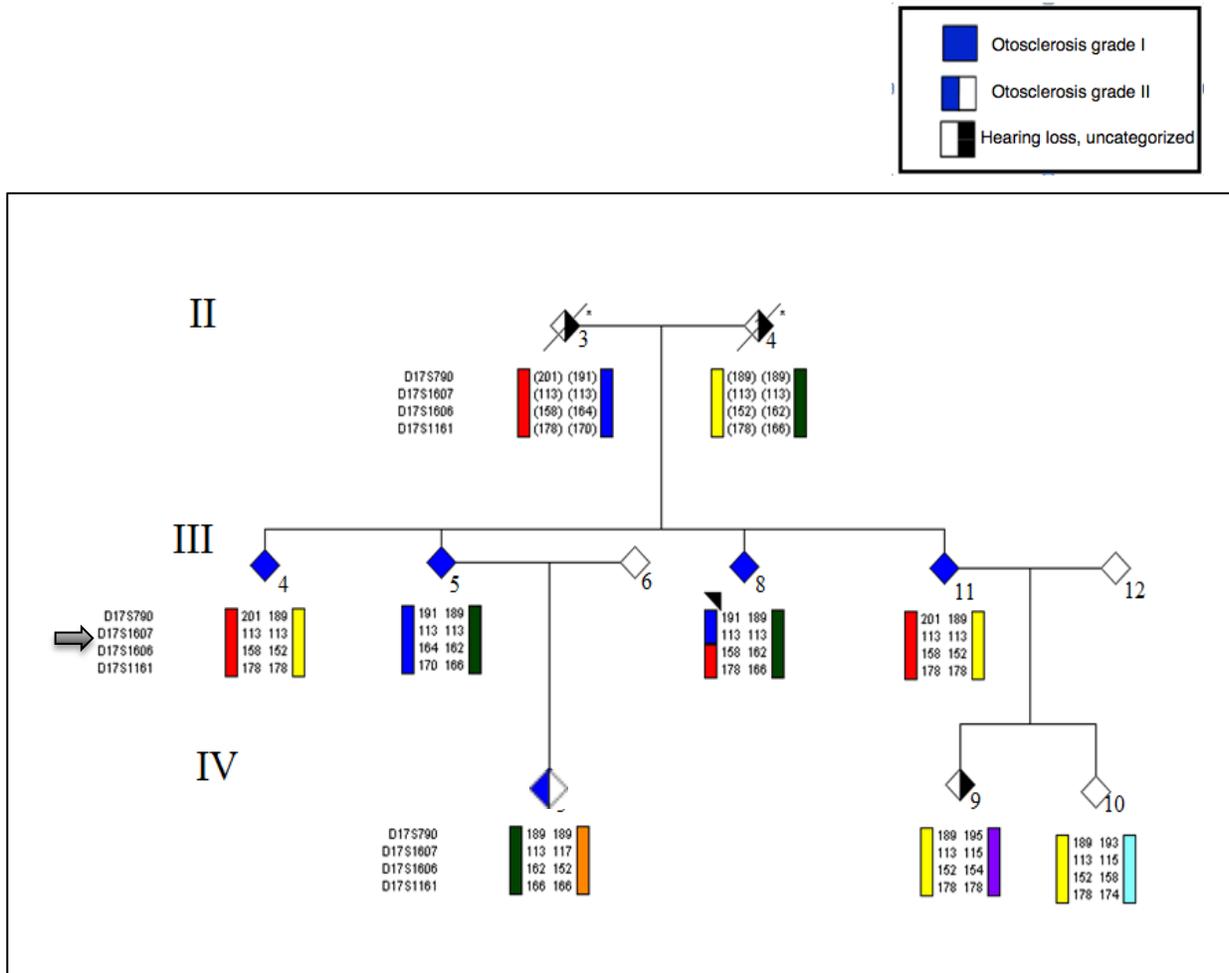
*NOG*, another otosclerosis candidate gene, is located at Chr17q24.31. Four markers (three of which were fully informative) surrounding the *NOG* gene were genotyped and four distinct parental haplotypes (blue, red, dark green and yellow) were constructed.

*D17S1607* was homozygous for the same allele in four of seven genotyped relatives.

Using the rule of least number of recombinations, one recombination between *D16S1607* and *D16S1606* was detected in the proband PID III-8 (Figure 2.17). Affected siblings with grade I otosclerosis inherited different haplotypes across this locus. For example, PID III-4 inherited the parental yellow and red haplotype, while PID III-5 inherited the parental blue and dark green haplotypes. The four affected siblings did not share a disease haplotype across the *NOG* locus and therefore, linkage of otosclerosis in Family 2081 to the *NOG* gene was ruled out.

**Figure 2.17: Partial pedigree of Family 2081 showing *NOG* gene haplotypes**

No shared region across *NOG* among affected members. The Family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.



### 2.3.2.2 Exclusion of Family 2114 from previously mapped loci and genes

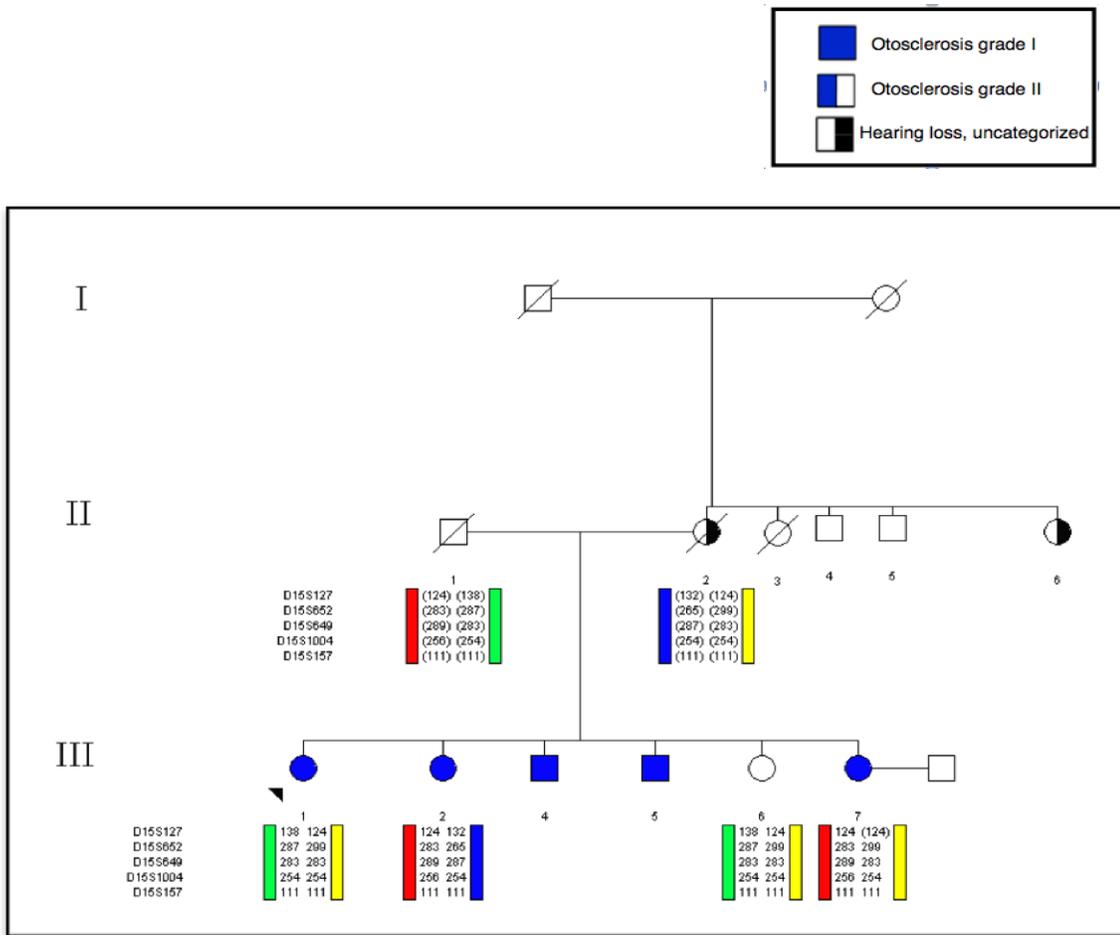
Genetic analysis of Family 2114 was carried out through genotyping of three affected siblings, PIDs III-1, III-2 and III-7, and one unaffected sibling, PID III-6, with polymorphic markers that spanned *OTSC1*, *OTSC2*, *OTSC3*, *OTSC4*, *OTSC5*, *OTSC7* and *OTSC8* and three candidate genes (*COLIA1*, *COLIA2* and *NOG*). Haplotypes of the three affected siblings were constructed across each locus. The mode of inheritance of otosclerosis in this family is presumed to be AD and the mother is thought to be the disease carrier. However, because we have only two generations affected, a recessive mode of inheritance cannot be excluded. For that reason, haplotypes of affected members were analyzed under both modes of inheritance (AD, AR). A single disease haplotype shared by affected siblings but not shared by the unaffected sibling is expected under a dominant mode of inheritance. Two disease haplotypes co-segregating in all affected siblings were expected under a recessive mode of inheritance model. Assuming a dominant mode of inheritance and presuming the mother to be the disease carrier, affected siblings did not share any disease associated haplotypes across *OTSC1*, *OTSC4*, *OTSC3*, *OTSC5*, *OTSC7*, *OTSC8* or the *COLIA1* and *NOG* genes. Whereas, assuming a dominant mode of inheritance and presuming the father to be the disease carrier, affected siblings shared a haplotype across *OTSC2* and *COLIA2*. A summary of the genotype/haplotype data is given below (Figures 2.18-2.27). Haplotypes for Family 2114 were created using the same microsatellite markers used for Family 2081.

## ***OTSC1***

Four distinct parental haplotypes (blue, light green, red and yellow) were created using five microsatellite markers. Parental haplotypes were inferred and no recombinations were invoked (Figure 2.18). The three affected siblings (PIDs III-1, III-2 and III-7) inherited different haplotypes across this region. For example, the proband PID III-1 inherited the parental light green and yellow haplotypes, whereas, the proband's sibling, PID III-2 inherited the blue and red haplotype. As affected siblings did not share a disease haplotype across this locus it was excluded from linkage to *OTSC1*.

**Figure 2.18: Partial pedigree of Family 2114 showing *OTSC1* haplotype.**

No shared allele between affected siblings across *OTSC1*. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.



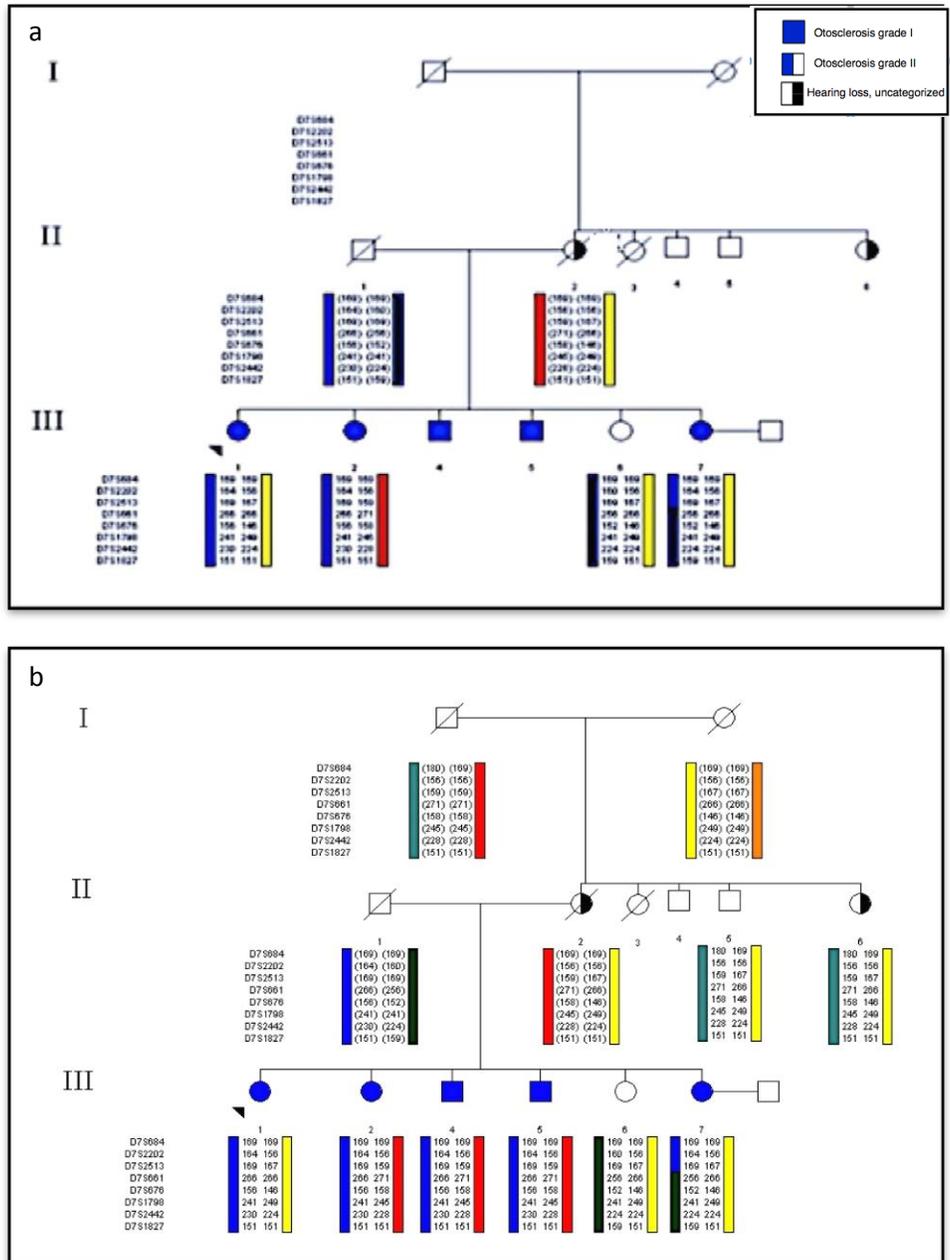
## ***OTSC2***

Eight microsatellite markers were used for genotyping this locus and four distinct parental haplotypes (dark green, red, blue and yellow) were created (Figure 2.19). First, four genomic DNA samples comprising the three affected siblings, PID III-1, III-2, and III-7, and one unaffected sibling, PID III-6 were genotyped for the markers that covered *OTSC2*. Using the rule of least number of recombinations, a recombination between the two informative markers *D7S2513* and *D7S661* was detected in PID III-7. As a result, PIDs III-1, III-2, and III-7 shared one allele of the four informative markers (*D7S684*, *D7S2202*, *D7S513*, and *D7S 661*) that covered a 5 Mb region of the blue haplotype (Figure 2.19 A, blue haplotype).

In order to phase and test segregation of the blue haplotypes within Family 2114 DNA from two extra siblings with confirmed otosclerosis, PIDs III-4 and III-5, and two of the mother's siblings, PIDs II-5 and II-6, were collected and genotypes, and haplotypes were constructed for the *OTSC2* locus. Construction of haplotypes for PIDs III-4 and III-5 showed that these two affected siblings inherited the full blue haplotype (Figure 2.19 B, blue haplotype). Construction of haplotypes for the mother's siblings, PIDs II-5 and II-6 helped to verify the inferred mother's yellow haplotype and determined that the shared blue haplotype was inherited from the father. Assuming the father is the disease carrier and a dominant mode of inheritance haplotype segregation across this locus shows that otosclerosis in Family 2114 is linked to a 5 Mb region at *OTSC2* that covers four microsatellite markers (*D7S684*, *D7S2202*, *D7S513*, and *D7S 661*).

**Figure 2.19: Partial pedigrees of Family 2114 showing *OTSC2* haplotypes**

The family proband is PID III-1. (a) Haplotypes across *OTSC2* in four siblings. (b) *OTSC2* haplotypes fully segregate with otosclerosis in Family 2114. The dark blue haplotype is the presumed disease haplotype. The Family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle = female, Square = male. Allele sizes are given in base pairs and alleles in brackets are inferred.

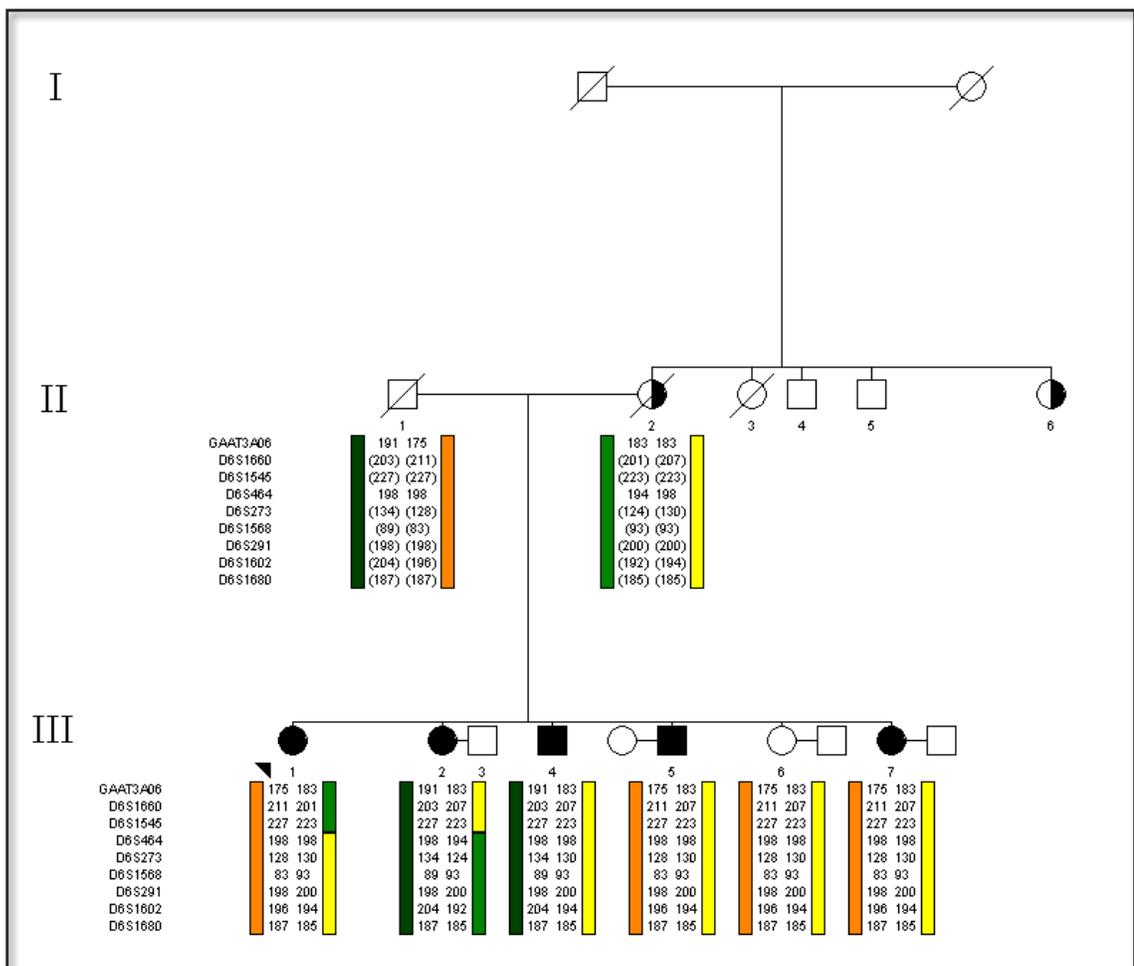
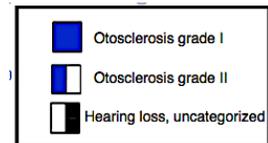


### ***OTSC3***

Nine microsatellite markers were used for genotyping at this locus and four distinct parental haplotypes (dark green, green, yellow and orange) were created (Figure 2.20). Using the rule of least number of recombinations, two recombinations were detected between *D6S1568* and *D6S439* in subject III-1 and III-2. Affected siblings, as well as the unaffected subject, PID III-6 share a 10.7 Mb region of the yellow disease haplotype between *D16S1568* and *D16S439*. This excludes linkage of Family 2114 to *OTSC3*.

**Figure 2.20: Partial pedigree of Family 2114 showing *OTSC3* haplotypes**

Affected siblings did not share a haplotype across *OTSC3*. The family proband is PID III-1 Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.

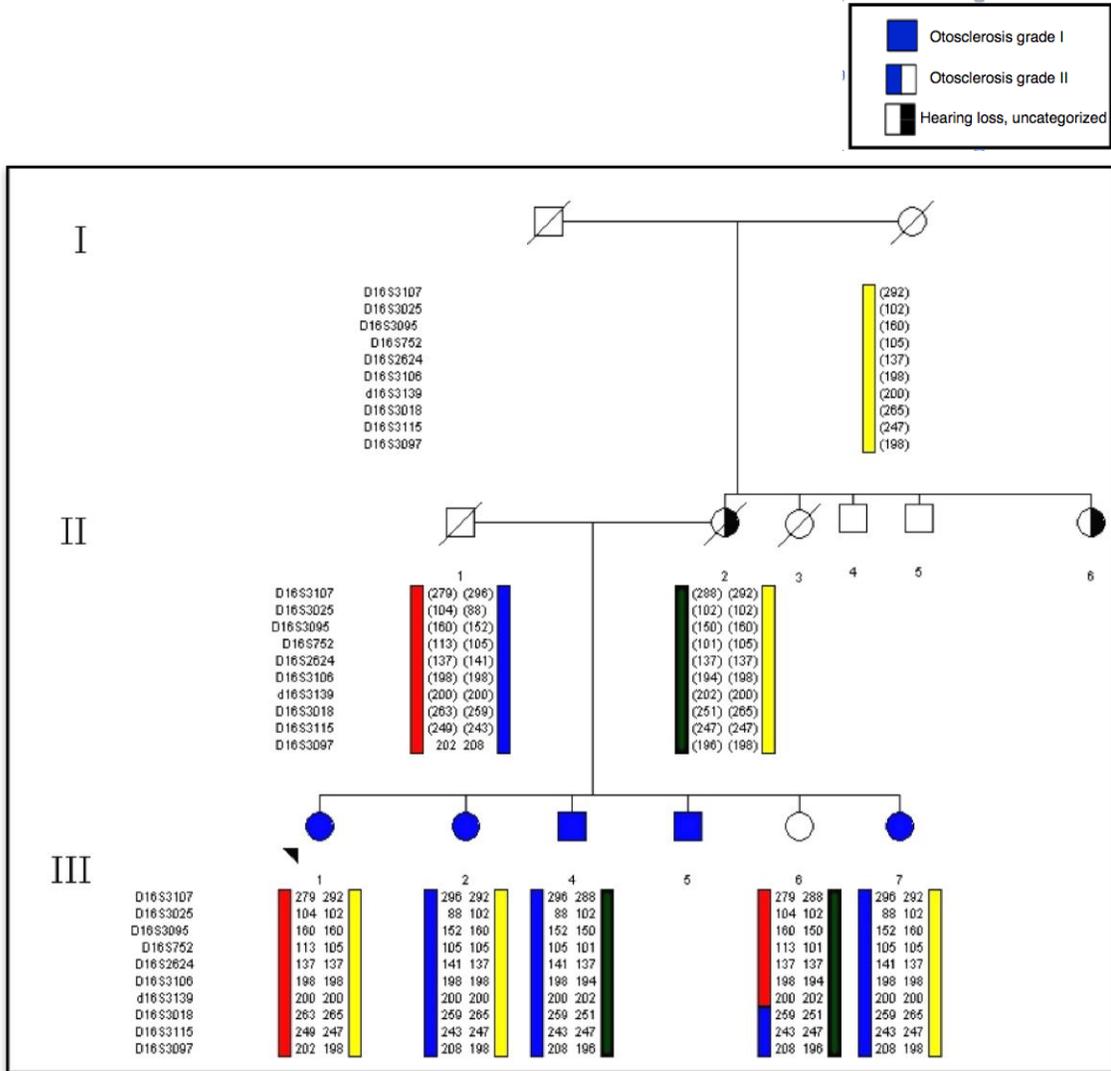


### ***OTSC4***

Ten microsatellite markers were genotyped at this locus and four distinct parental haplotypes (dark green, red, yellow and blue) were created (Figure 2.21). Using the rule of least number of recombinations, a recombination was detected between *D16S3139* and *D16S3018* in PID III-6. The affected siblings did not share a disease haplotype across the *OTSC4* locus. For example, the family proband, PID III-1, inherited the parental yellow and red haplotypes whereas, PID III-4 inherited the blue and the dark green haplotypes. This ruled out linkage of otosclerosis in Family 2114 to the *OTSC4* gene.

**Figure 2.21: Partial pedigree of Family 2114 showing *OTSC4* haplotypes.**

No shared allele was detected between affected siblings. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.

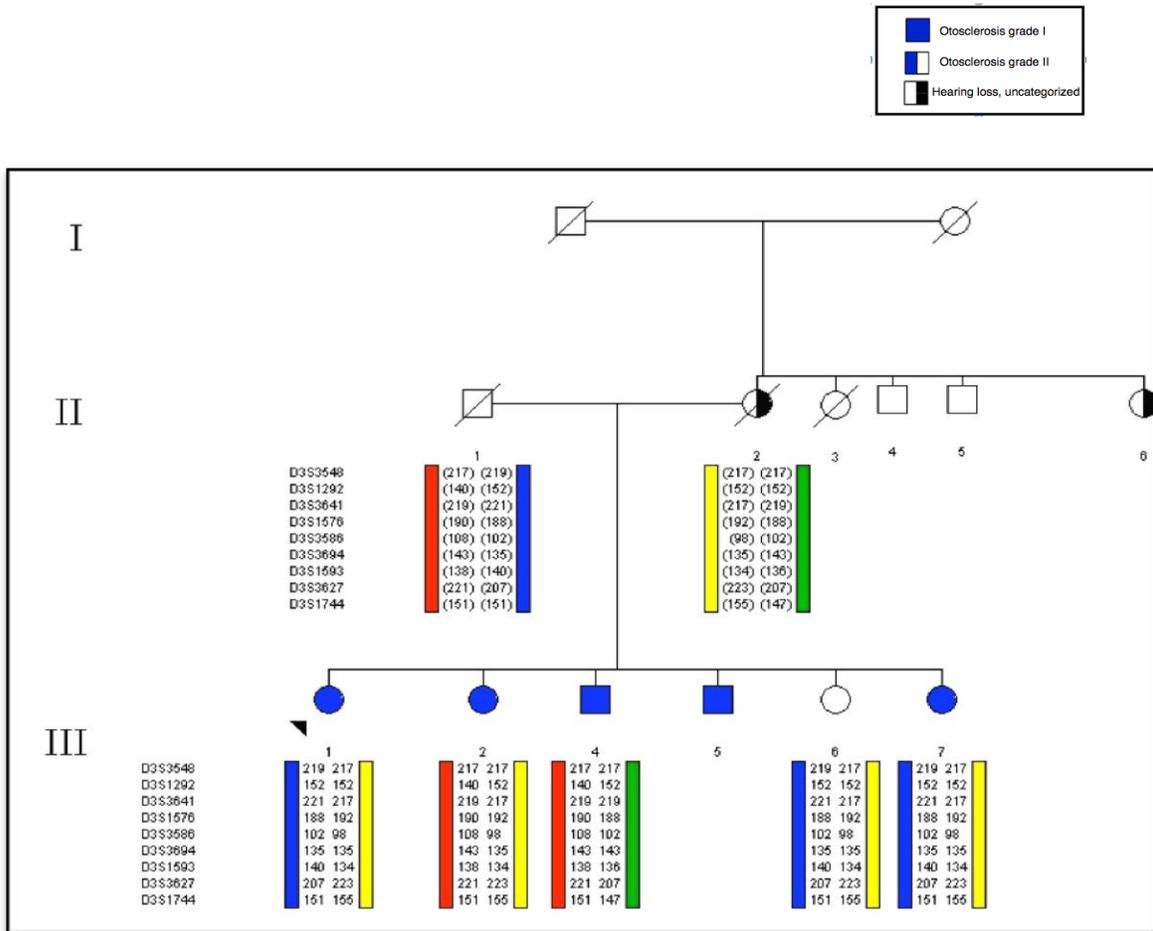


## *OTSC5*

Nine informative microsatellite markers were used for genotyping at this locus. Using the rule of least number of recombinations, no recombination was detected and four distinct parental haplotypes (green, red, yellow and blue) were inferred (Figure 2.22). The 4 affected family members did not share a disease haplotype across the *OTSC5* locus under either mode of inheritance. For example, PID-III-1 inherited the parental yellow and blue haplotypes whereas PID III-4 inherited the green and red haplotypes. This excluded *OTSC5* as the causative locus for Family 2114.

**Figure 2.22: Partial pedigree of Family 2114 showing *OTSC5* haplotypes**

Affected siblings did not share a single haplotype across *OTSC5*. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.

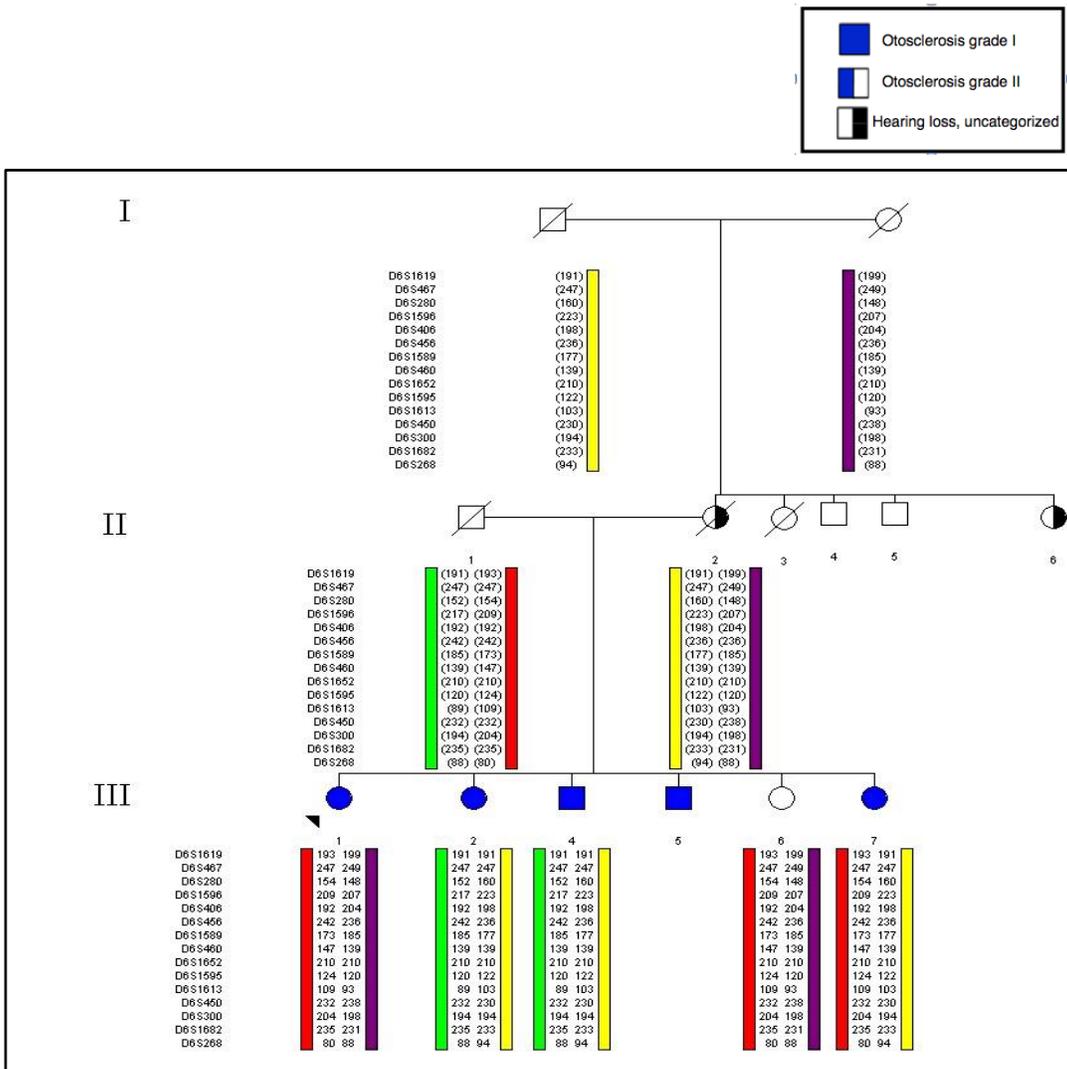


### *OTSC7*

Fifteen markers were genotyped at this locus and four distinct parental haplotypes (dark purple, red, yellow and light green) were created. Parental haplotypes were inferred and no recombination was invoked (Figure 2.23). Affected siblings did not share a disease haplotype across this locus. For example, the proband (PID III-1) and the unaffected PID III-6 shared the red and purple haplotypes. PID III-6 also shared the red haplotype with the affected sibling, PID III-7, but did not share any haplotypes with the affected siblings PIDs III-2 and III-4. These results exclude linkage of Family 2114 to *OTSC7*.

**Figure 2.23: Partial pedigree of Family 2114 showing *OTSC7* haplotypes.**

No shared haplotype was detected between affected siblings. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.

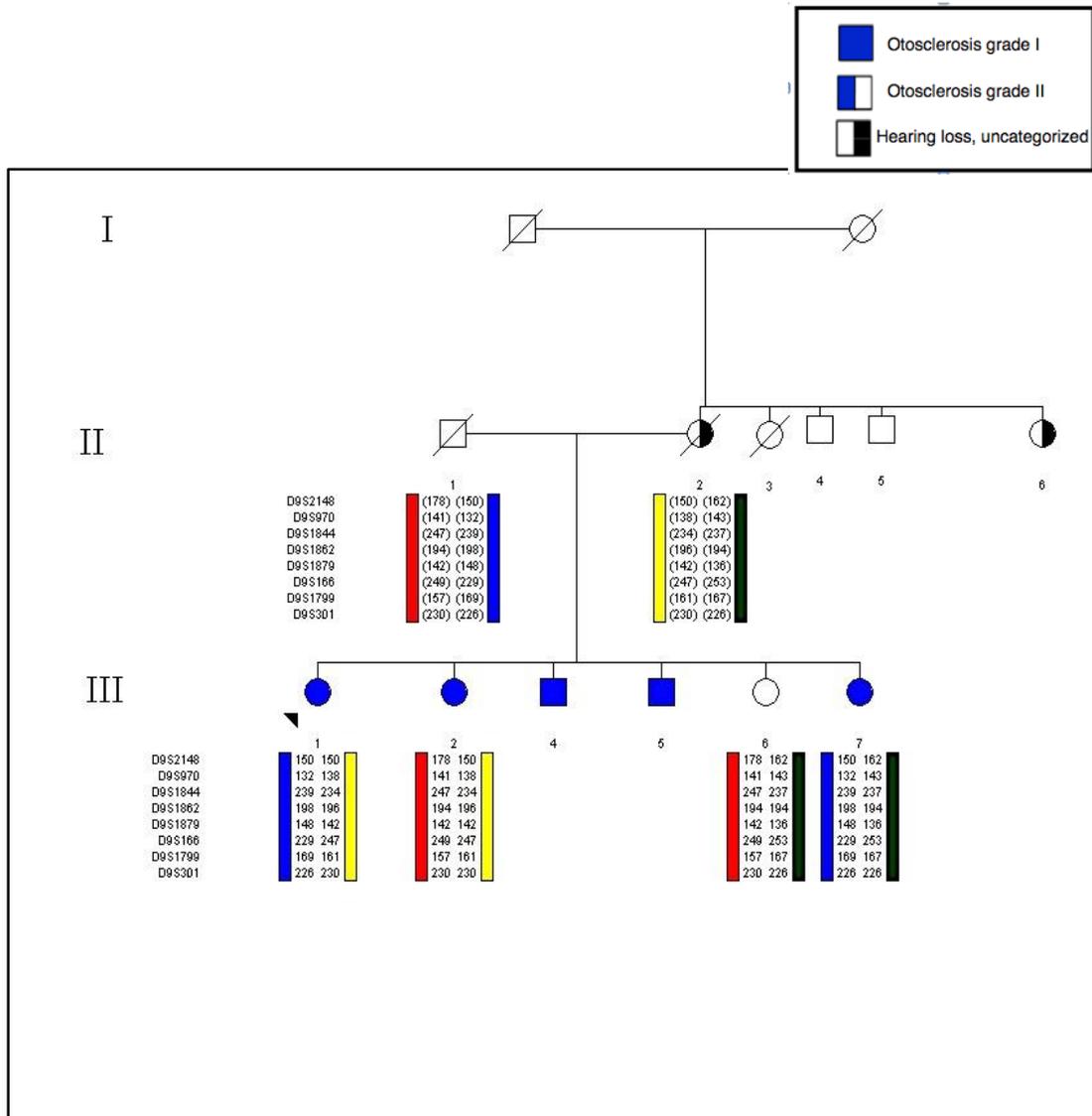


### ***OTSC8***

Eight microsatellite markers were genotyped at this locus, four distinct parental haplotypes (dark green, blue, red and yellow) were created and no recombinations were detected (Figure 2.24). The three affected siblings (PID III-1, III-2 and III-7) did not share a disease haplotype across this locus. For example (PID III-1) inherited the yellow and blue haplotypes. Subject PIDIII-1 shared the yellow haplotype with PID III-2 and the blue haplotype with PID III-7. In summary, as there was no shared haplotype detected between the affected siblings across *OTSC8*, it was excluded as a causative locus in Family 2114.

**Figure 2.24 Partial pedigree of Family 2114 showing *OTSC8* haplotypes**

No shared haplotype was recognized between affected siblings. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.

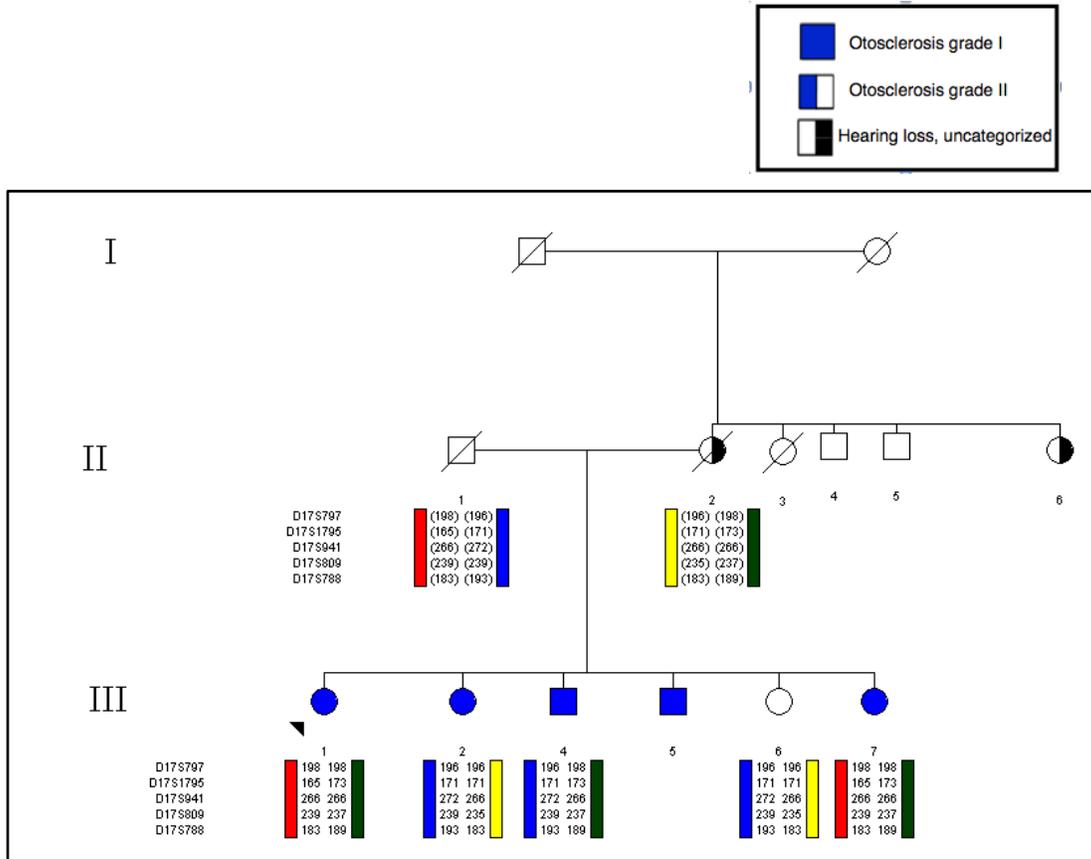


## *COL1A1*

Five microsatellite markers covering the region surrounding the *COL1A1* gene were genotyped. Using the rule of least number of recombinations, no recombination was invoked and four distinct parental haplotypes (red, blue, dark green and yellow) were constructed (Figure 2.25). The four affected siblings, PID III-1, III-2, III-4 and III-7 did not share a disease haplotype across the *COL1A1* region. For example, PID III-1, III-4 and III-7 inherited the parental green haplotype but PID III-2 did not. On the other hand, PID III-2 and III-4 shared the blue haplotype with the unaffected sibling PID III-6 but not with the affected siblings PID III-1 and III-7. This excluded linkage of otosclerosis in Family 2114 to *COL1A1*.

**Figure 2.25: Partial pedigree of Family 2114 showing *COL1A1* haplotypes**

Affected siblings did not share haplotype across *COL1A1*. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.



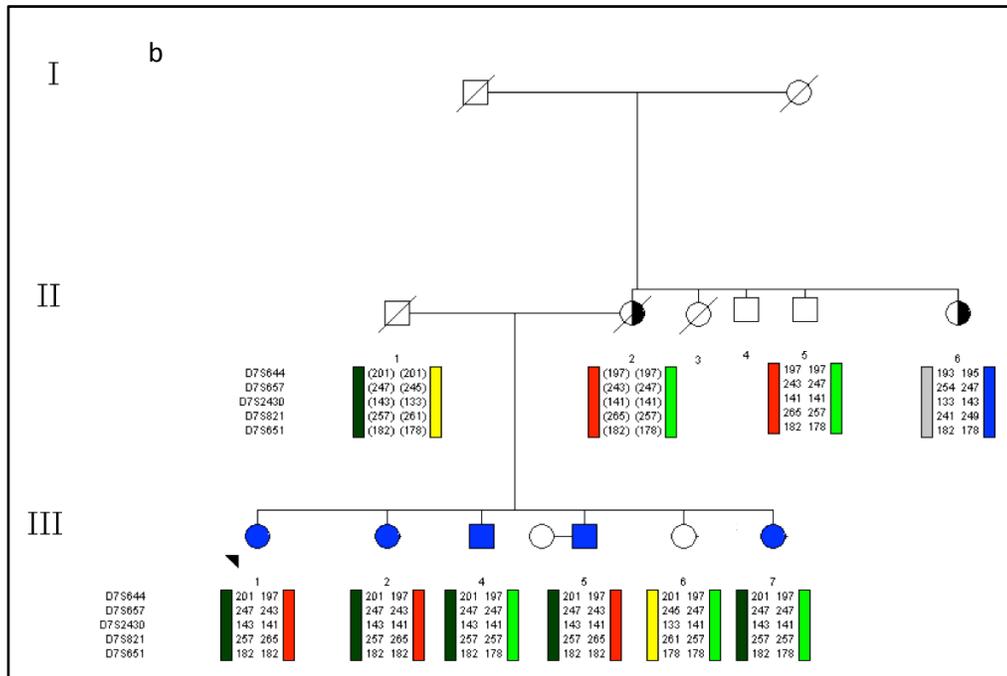
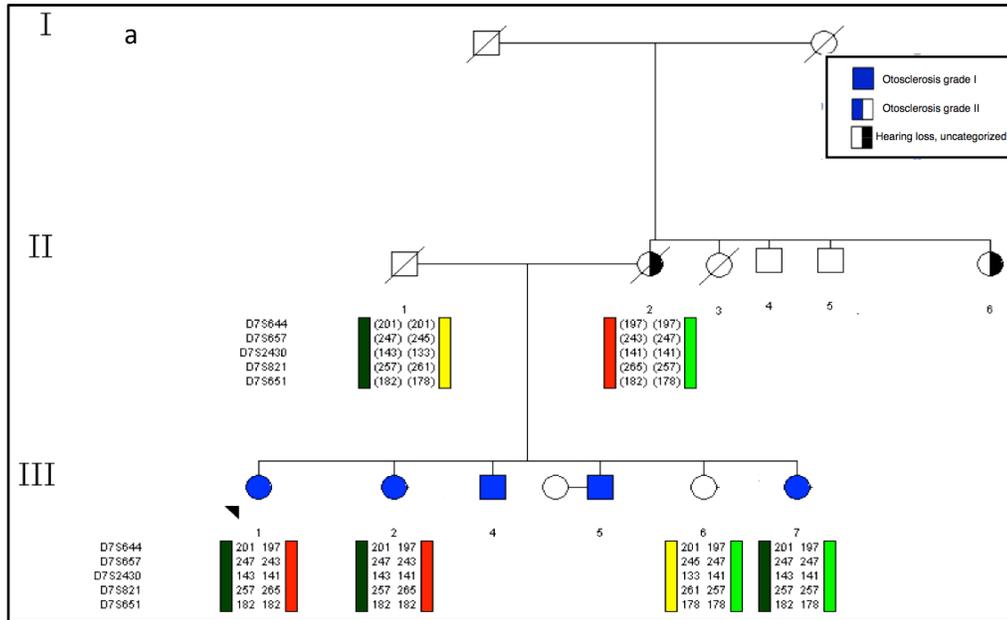
## *COLIA2*

Five microsatellite markers that covered the region surrounding *COLIA2* were genotyped and four distinct parental haplotypes (dark green, yellow, red and light green) were constructed (Figure 2.26 A). The three affected siblings, PIDs III-1, III-2, and III-7, shared the parental dark green haplotype which was not shared by the unaffected sibling PID III-6. This preliminary result suggested that Family 2114 was linked to *COLIA2* under a dominant mode of inheritance.

To confirm this, DNA from the other two affected siblings, PIDs III-4 and III-5, and from the proband's mother's siblings, PIDs II-5 and II-6, were genotyped using the same five microsatellite markers. Construction of haplotypes for the mother's sibling PID II-5 helped to infer the mother's light green and red haplotypes and indicated that the shared dark green haplotype was inherited from the father, who has no history of hearing loss (Figure 2.26 B). Also, constructed haplotypes for the affected siblings, PIDs III-4 and III-5, showed that they both inherited the dark green (disease) haplotype. Assuming the father was the disease carrier this result links Family 2114 to the *COLIA2* region.

**Figure 2.26: Partial pedigree of Family 2114 showing *COLIA2* haplotypes**

The Family proband is PID III-1. (a) Segregation of dark green haplotype in three affected (III-1, III-2 and III-7). (b) Segregation of dark green haplotype across *COLIA2* with otosclerosis in Family 2114. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. Crossed symbol indicates a deceased subject. Circle =female, Square =male. Allele sizes are given in base pairs and alleles in brackets are inferred.

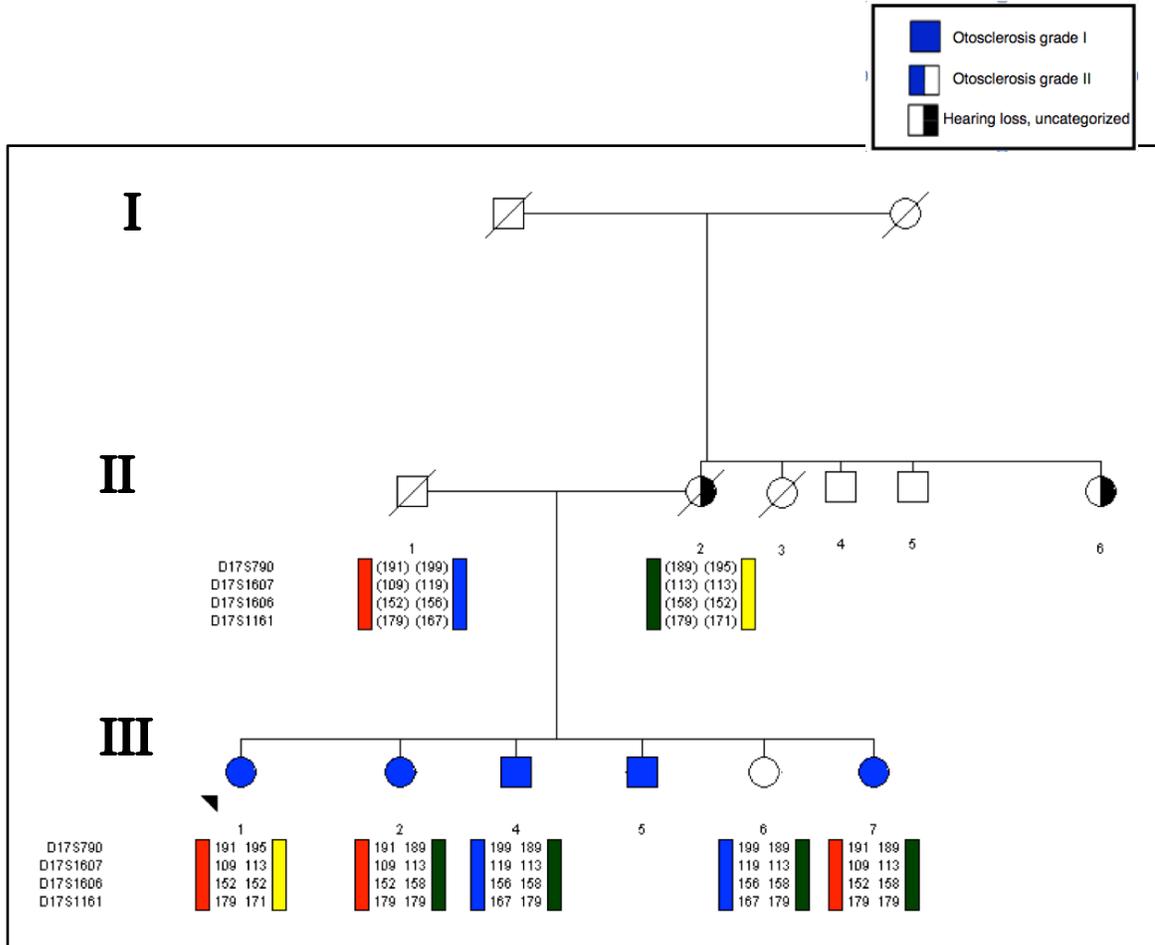


### ***NOG gene***

Four markers surrounding the *NOG* gene were genotyped and four distinct parental haplotypes (blue, yellow, dark green and red) were constructed. The four affected siblings, PID III-1, III-2, III-4, and III-7, inherited different haplotypes (Figure 2.27). For example, PID III-1 inherited the parental red and yellow haplotypes, while the PID III-4 inherited the parental blue and dark green haplotypes. Also, the three affected siblings PID III-2, III-4, PID III-7 and the unaffected sibling PID III-6 shared a dark green haplotype which the affected proband PID III-1 did not have. The four haplotyped affecteds did not share a disease haplotype across the *NOG* gene locus and this ruled out linkage of otosclerosis in Family 2114 to the *NOG* gene.

**Figure 2.27: Partial pedigree of Family 2114 showing *NOG* gene haplotypes**

No shared haplotype was detected between affected siblings. The family proband is PID III-1. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle = female, Square = male. Allele sizes are given in base pairs and alleles in brackets are inferred.



### 2.3.3 Fine mapping of *OTSC4* locus in Family 2081

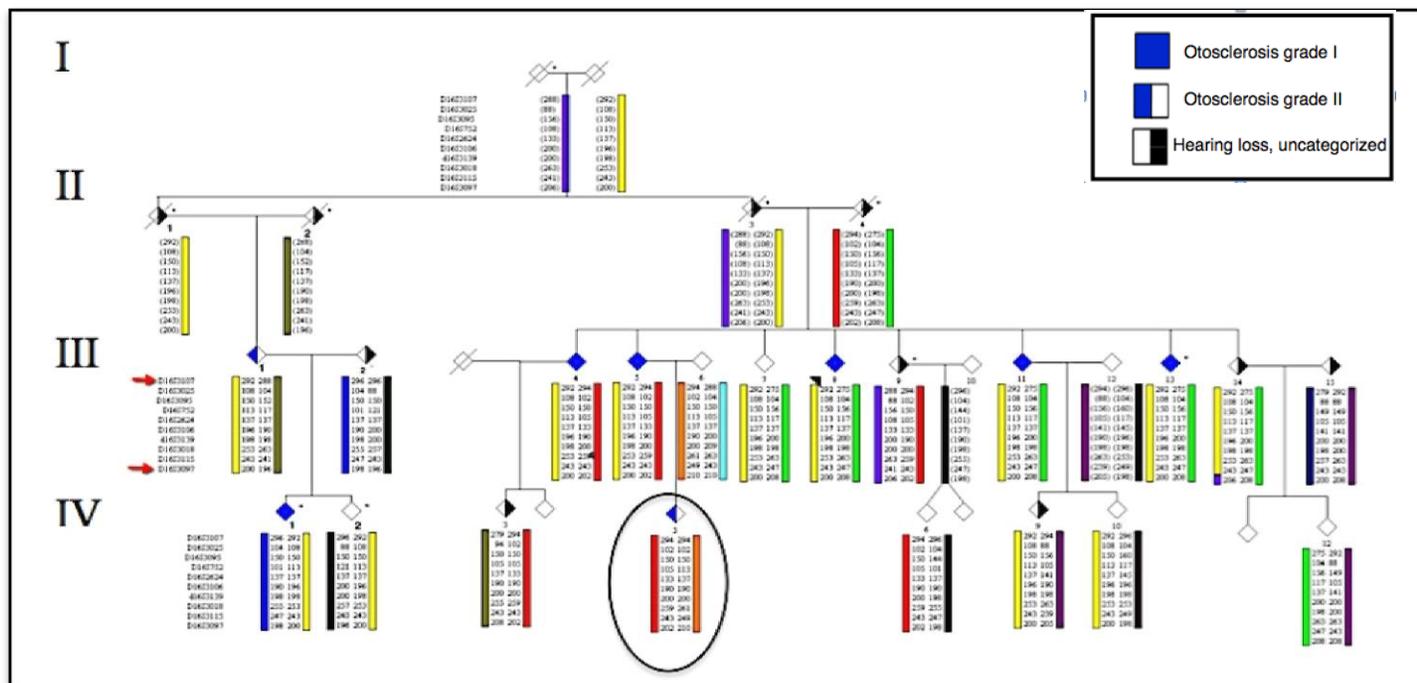
Haplotype analysis at the *OTSC4* locus showed that the four affected siblings PIDs III-4, III-5, III-8 and III-11 shared a presumed disease (yellow) haplotype (Figure 2.10). In order to further investigate the potential role of *OTSC4* in Family 2081, an extra 14 members of the family PIDs III-1, III-2, III-6, III-7, III-9, III-12, III-13, III-14, III-15, IV-1, IV-2, IV-3, IV-6 and IV 12, including the proband's cousin PID IV-1 (who was diagnosed with grade I otosclerosis), were genotyped and phased for markers that spanned the *OTSC4* interval. Segregation analysis showed that the proband, proband's affected siblings (PIDs III-4, III-5, III-11, and III-13) and affected cousin (PID IV-1) inherited the presumed disease (yellow) haplotype (Figure 2.28). In addition, two unaffected subjects in the third and fourth generation (PIDs III-7 and IV-10) and two members with a history of uncategorized hearing loss (PID III-14 and IV-9) inherited the disease (yellow) haplotype. As well, the unaffected proband's cousin, PID IV-2, and one parent, PID III-1, inherited the yellow haplotype. Family member PID IV-5 who has grade II otosclerosis did not inherit the disease (yellow) haplotype. Segregation of the *OTSC4* presumed disease haplotype in family 2081 raised the possibility of linkage of Family 2081 to *OTSC4*.

At the *OTSC4* locus, we only have one apparent recombination within several generations, which consolidates this haplotype. In addition to this, all family members with grade I otosclerosis in generation III and their cousin (subject IV-1) shared a presumed disease (yellow) haplotype. However, absence of the inheritance of the yellow haplotype from the grade II affected family member IV-5 did not fit the pattern. Subject

PID IV-5 was considered a phenocopy because the diagnosis of otosclerosis was based on the audiological profile. At this stage, we considered *OTSC4* a candidate region for Family 2081 and subject PID IV-5 a phenocopy.

**Figure 2.28: Partial pedigree of NL Family 2081 segregates *OTSC4* haplotype**

Affected members share a yellow haplotype across *OTSC4*. The subject IV-5, a phenocopy, is circled. Allele sizes are given in base pairs and alleles in brackets are inferred. Red arrows show boundary markers *D16S3107* and *D16S3097* for the *OTSC4* critical region. The family proband is PID III-8. Roman numerals on the left side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The crossed symbol indicates a deceased subject. Circle = female, Square = male. Allele sizes are given in base pairs and alleles in brackets are inferred.



#### **2.3.4 Shared haplotype across four unrelated otosclerosis probands**

The *OTSC4* mapped region is 9.7 Mb and contains 129 annotated genes. In addition to families 2081 and 2114, we recruited probands from seven unrelated NL families. Because these probands originated from NL, we hypothesized that they could share a founder haplotype. Therefore, six out of seven of these probands (Family # 2197, 2203, 2066, 2126, 2103 and 2200) were checked for a founder genotype across *OTSC4*. It was found that the proband of Family 2200 shared alleles of four microsatellite markers (*D16S3095*, *D16S752*, *D16S3106* and *D16S3139*) with the proband of Family 2081. The probands of Families 2203, 2197, and 2126 shared a portion of the presumed disease haplotype that spanned the three microsatellite markers *D16S752*, *D16S3106* and *D16S3139* (Table 2.2). The probands of Families 2200 and 2203 also shared an allele at *D16S3115*. Probands of Families 2103 and 2066 shared an allele at the *D16S3139* marker. Of these seven probands, including proband of Family 2081, four probands shared alleles at *D16S752*, *D16S3106* and *D16S3139*, which minimized the *OTSC4* candidate region from 9.7 Mb to 4.2 Mb.

**Table 2.2: Shared alleles among NL otosclerosis probands across *OTSC4***  
 Four families proband (Families 2200, 2203, 2197 and 2126) share alleles of three markers (*D16S752*, *D16S3106* and *D16S3139*) along with the proband of Family 2081.

<i>OTSC4</i>	Distance Mb	Family 2081	Family 2200	Family 2203	Family 2197	Family 2126	Family 2103	Family 2066
<i>D16S3107</i>	66.2	<b>292</b>	na	na	na	287	288	294
<i>D16S3025</i>	67.1	<b>108</b>	na	92	88	102	104	106
<i>D16S3095</i>	68.5	<b>150</b>	150	152	156	160	156	156
<i>D16S752</i>	69.8	<b>113</b>	113	113	113	113	109	105
<i>D16S3106</i>	70.7	<b>196</b>	196	196	196	196	194	194
<i>D16S3139</i>	71.2	<b>198</b>	198	198	198	198	198	198
<i>D16S3018</i>	72.7	<b>253</b>	263	263	225	265	263	259
<i>D16S3115</i>	73.8	<b>243</b>	243	243	247	245	245	247

### **2.3.5 Gene screening of functional candidate genes**

The minimized *OTSC4* (4.2 Mb) region contains 49 positional candidate genes, which is a large number of genes to screen using Sanger sequencing. Therefore, genes were prioritized according to their function and expression, as described in Appendix 11. In total, 12 genes were sequenced and 96 variants were identified.

#### **2.3.5.1 Variant analysis**

Under a dominant mode of inheritance, we are expecting to identify variants that are present in the affected members and absent in the controls. Variant analysis was carried out as shown in Figure 2.1.

Sequencing of 12 functional candidate genes in the minimized *OTSC4* identified 96 variants in the mutation screening panel (Figure 2.29 and Appendix 12). Of these 96 variants, only seven variants, five in the *ZFHX3* and two in the *IL34* gene, exclusively segregated with the disease haplotype. To determine the possible pathogenicity of these seven variants, allele frequencies were checked using the SNP/1000 genome database. As a result, the two variants in the *IL34* gene were shown to have high frequencies; the C allele of c.107G>C has a frequency of 0.458/120 (45%) and the A allele of c.639 C>A has a frequency of 0.112/224 (11.2 %). These variants were excluded from further follow-up due to their high frequency in the general population (Appendix 10B).

Of the five identified variants in the *ZFHX3* gene, four variants were SNPs with no frequencies listed in the SNP/1000 genome database and one was novel. The four

SNPs with no frequency data included three intronic, one missense and the novel variant was a 3 bp insertion. To determine the effect of each variant on the mRNA and protein coded by the *ZFHX3* gene, the effect on the protein of the missense variants were checked using SIFT and Polyphen prediction programs (pg. 28). The splicing effects of the three intronic variants were checked using HSF, MaxEntScan, NNSPLICE, GeneSplicer, SSF and Known constitutive signals splice prediction tools. No splicing effects were predicted.

c.10831C>T, p.H3611HY variant (Figure 2.30) was predicted to be deleterious by SIFT (score 0.00) and probably damaging by Polyphen (score 0.6666). The conservation of the histidine residue at position 3611 in the *ZFHX3* gene was found to be conserved in four of five species (dog, mouse, rat and chicken) using Clustal W and Weblogo (Figure 2.31).

The *ZFHX3* c.10557-10558insCCG (Figure 2.32 A) was predicted to cause an in-frame duplication of glycine (p.Gly 3527 dup). The glycine insertion was placed within a string of a 14-glycine repeat (Figure 2.32 B). Conservation of the 14-glycine repeats was checked across five species. The examined species have fewer than 14 glycine repeats (12 in Dog and Mouse, 13 in Rat, two in Chicken and two in Zebrafish) (Figure 2.33). We further examined the conservation of this region across NL population controls and found that the normal population controls harbour less than 14 glycine repeat at this position. From the conservation study of the 14 glycine repeat sequence across species and normal NL population controls, it appears that the glycine sequence across this region was not conserved across these species and controls. These results suggested that the 14 glycine

repeat did not have any important function and deletion of glycine residues would not have a drastic effect on either the mRNA or the translated protein. In other words, the variability seen in the length of the glycine string repeats between species and in the NL population is benign and gene function is retained.

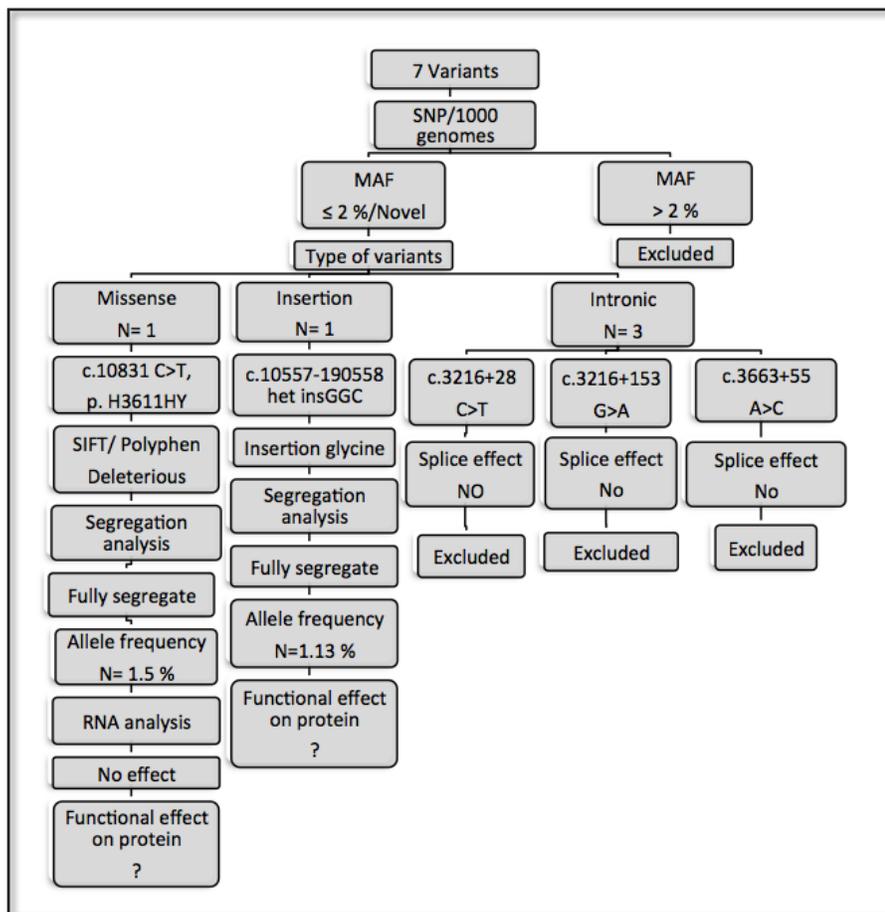
#### **2.3.5.2 Functional assessment of the *ZFH3*. c.10557\_10558ins GGC**

We hypothesized that if the deletion of glycine residue ('s) does not have a drastic effect, then maybe the insertion of glycine residue ('s) could cause toxicity and therefore cause degradation of the mRNA. To test this hypothesis, cDNA of the proband PID III-8 and two siblings (PID III-4 and II-5) were prepared and sequenced. If our hypothesis was right and the mRNA produced by the mutated form (insGGC) of the *ZFH3* gene was degraded, we would expect to see a normal electropherogram (no frame shift).

Sequencing of the cDNA that covered the 3 bp insertion in the proband and two siblings showed that the mutated allele was not degraded as recorded by the presence of a frame shift in the electropherogram (Figure 2.34). This result refuted our hypothesis that the mRNA would render the mRNA transcript less stable.

**Figure 2.29** Flowchart of filtration steps applied to identified variants at the minimized OTSC4

Missense mutation and 3 bp insertion in the *ZFHX3* gene segregates with disease in Family 2081. Allele frequencies of identified variants were determined using publicly available SNP database and 1000 genomes. Pathogenicity of missense mutations was evaluated by using SIFT and Polyphen prediction programs. Splicing prediction programs SSF, HSF, MaxEntScan, NNSPLICE, GeneSplicer were used to check for possible splice site mutations.



### **2.3.5.3 Segregation analysis**

Of the five variants identified in the *ZFHX3* gene, only the missense and the 3 bp insertion variants were subjected to segregation analysis. In order to determine which of the two sequencing variants in *ZFHX3*, were fully segregating with the disease haplotype in Family 2081, all members with available DNA were genotyped for these variants and haplotypes were created (Figure 2.35). As a result, these four variants were seen to exclusively segregate with the yellow haplotype in Family 2081; every subject who inherited the yellow haplotype harboured these two variants. Segregation analysis in Family 2081 determined that these two variants resided on the presumed disease (yellow) haplotype.

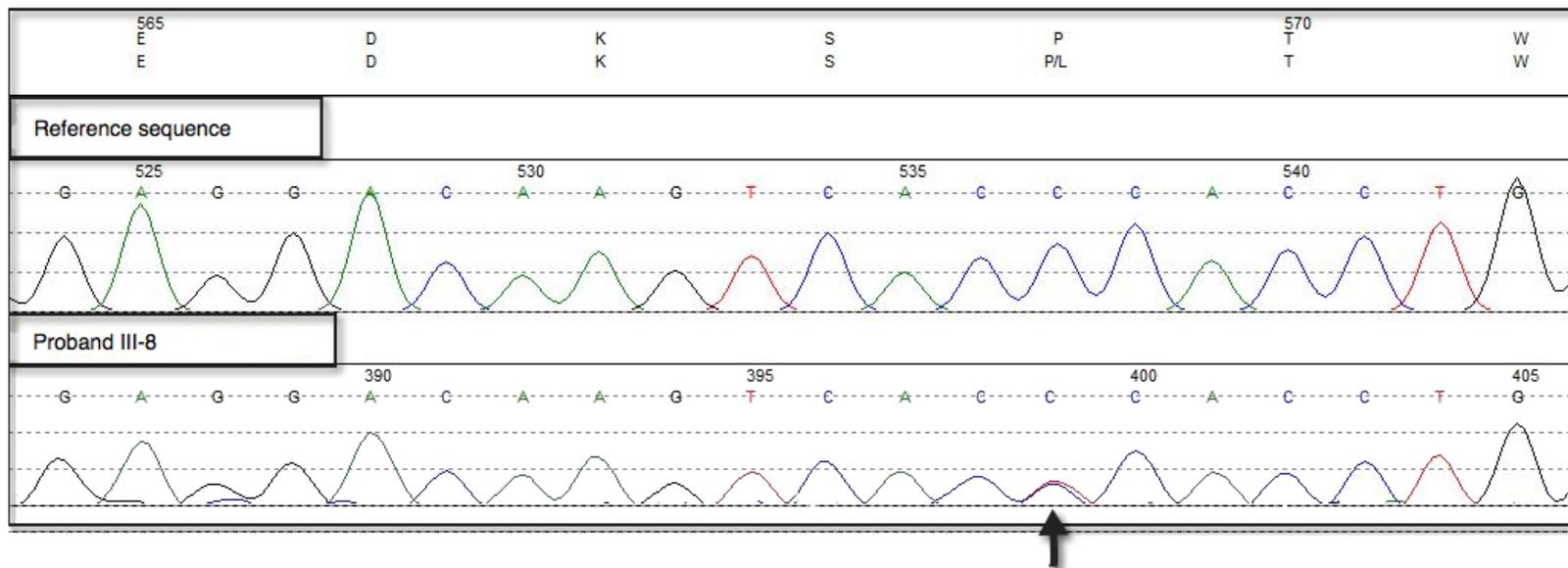
### **2.3.5.4 Allele frequencies of the *ZFHX3* variants in NL controls**

Allele frequencies of the two *ZFHX3* variants were determined using ethnically matched NL population controls. The frequencies of c.10557\_10558 insGGC, and c.1083C>T, p.H3611Y were 1.5% and 1.13% respectively (Table 2.3).

To test the role of the missense and insertion variants in other otosclerosis families, eight NL otosclerosis probands, including the four used to fine map *OTSC4*, were screened. If our hypothesis was correct and the four probands shared a founder haplotype, we would expect the four probands to harbour these two variants. None of the four probands harboured these variants which excluded the founder haplotype based on the assumption that these haplotypes were not really identical.

**Figure 2.30: Electropherogram shows heterozygous missense mutation *ZFHX3* c.10831C>T, p.H3611HY in PID III-8.**

The upper panel is the reference sequence of *ZFHX3*. The lower panel is the sequence of PID III-8 that harbours the c.10831 C>T variant. The arrow points to the c.10831 C>T mutation.



**Figure 2.31: Multiple alignment of histidine at the amino acid position 3611 in *ZFHX3* across five species.**

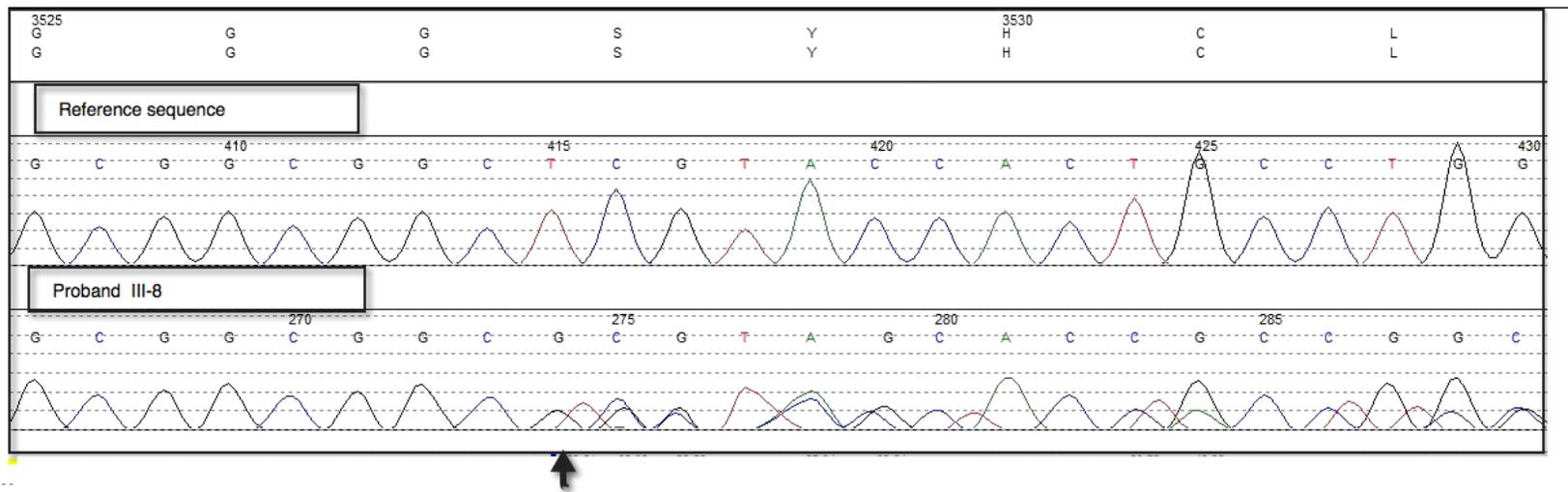
Clustal W was used to align orthologs from human (NP\_008816.3), dog (XP\_546849.2), mouse (NP\_031522.2), rat (XP\_226464.3), chicken (XP\_414230.2), and Zebrafish (XP\_688934.3). Histidine at position 3611 (indicated by the arrow) is conserved across human and four species (dog, mouse, rate and chicken).

Human	<a href="#">NP_008816.3</a>	3581	PSTASTSQSAAHSNDSPPPPSAAAPSSASPHASRKSWPQVVS	3626
Dog	<a href="#">XP_546849.2</a>	3595	PSTASTSQSAAHSNDSPPPPSAAAPSSASPHASRKSWPQVVS	3642
Mouse	<a href="#">NP_031522.2</a>	3601	PSTASTSQSAAHSNDSPPPPSAAAPSSASPHASRKSWPPVGS	3646
Rat	<a href="#">XP_226464.3</a>	3608	PSTASTSQSAAHSNDSPPPPSAAAPSSASPHASRKSWPPVGS	3653
Chicken	<a href="#">XP_414230.2</a>	3539	PSTASTSQSAAHSNDSPPPPRPPPA---SPHASRKPWPPAAPR	3581
Zebrafish	<a href="#">XP_688934.3</a>	2473	PNTASTSQSATHSNNTSPPPTTSATTPSSSTASSCKSWSQPFS	2522

**Figure 2.32 Electropherogram shows heterozygous *ZFHX3* c.10557\_10558insGGC in PID III-8**

A. The upper panel is the reference sequence of *ZFHX3*. The lower panel is the reference sequence from PID III-8 showing the 3 bp insertion and resultant frameshift (indicated by the arrow).

**A**



**B. Reference sequence of NM\_006885 (*ZFHX3*).**

Capture of the *ZFHX3* cDNA from Genome browser<sup>141</sup>. At block nine (Exon 9) of the *ZFHX3*, a string of 14 glycine codons is shown (highlighted green). Insertion of the 3 bp led to insertion of an extra glycine in the highlighted string.

```
acaaggttgt ccagccctga gcctggtggg atacttctgt tggaaatgtt 72822875
cgttttgaga ggatccttca agccttgaca tctcaaactg tgtgccagtg 72822825
acaccggtaa aacgtcttga gcttcagccc taccctgccc cagactgacc 72822775
ggtgatttct gtctccattc ctttcagCTT TAACGTCTCC TAAGCCGAAC 72822725
TTGATGGGTC TGCCCAGCAC AACTGTTCTC TCCCCTGGCC TCCCCACTTC 72822675
TGGATTACCA AATAAACCGT CCTCAGCGTC GCTGAGCTCC CCAACCCAG 72822625
CACAAAGCCAC GATGGCGATG GGCCCTCAGC AACCCCCCA GCAGCAGCAG 72822575
CAGCAGCAGC AACACAGGT GCAGCAGCCT CCCCCGCCGC CAGCAGCCCA 72822525
GCCGCCACCC ACACCACAGC TCCCCTGCA ACAGCAGCAG CAACGCAAGG 72822475
ACAAAGACAG TGAGAAAGTA AAGGAGAAGG AAAAGGCACA CAAAGGAAA 72822425
GGGGAACCCC TGCCTGTCCC CAAGAAGGAG AAAGGAGAGG CCCCCACGGC 72822375
AACTGCAGCC ACGATCTCAG CCCCCTGCC CACCATGGAG TATGCGGTAG 72822325
ACCCTGCACA GCTGCAGGCC CTGCAGGCCG CGTTGACTTC GGACCCACA 72822275
GCATTGCTCA CAAGCCAGTT CCTTCCTTAC TTTGTACCAG GCTTTTCTCC 72822225
TTATTATGCT CCCCAGATCC CTGGCGCCCT GCAGAGCGGG TACCTGCAGC 72822175
CTATGTATGG CATGGAAGGC CTGTTCCCCT ACAGCCCTGC ACTGTGCGAG 72822125
GCCCTGATGG GGCTGTCCCC AGGCTCCCTA CTGCAGCAGT ACCAGCAATA 72822075
CCAGCAGAGT CTGCAGGAGG CAATTTCAGCA GCAGCAGCAG CGGCAACTAC 72822025
AGCAGCAGCA GCAGCAAAAA GTGCAGCAGC AGCAGCCCAA AGCAAGCCAA 72821975
ACCCAGTCC CCCCCGGGGC TCCTTCCCCA GACAAAGACC CTGCCAAAGA 72821925
ATCCCCAAA CCAGAAGAAC AGAAAAACAC CCCCCTGAG GTGTCCCCC 72821875
TCCTGCCGAA ACTCCCTGAA GAGCCAGAAG CAGAAAGCAA AAGTGCGGAC 72821825
TCCCTACTAG ACCCTTCAT TGTTCCAAAG GTGCAGTACA AGTTGGTCTG 72821775
CCGCAAGTGC CAGGCGGGCT TCAGCGAGA GGAGGCAGCG AGGAGCCACC 72821725
TGAAGTCCCT CTGCTTCTTC GGCCAGTCTG TGGTGAACCT GCAAGAGATG 72821675
GTGCTTACAG TCCCCACCGG CGGCGGCGGC GGTGGCAGT GGGCGGCGG 72821625
GGCGGTGGC GCGGCGGCG GCGGCGGCG C TCGTACCAC TGCCTGGCGT 72821575
GCGAGAGCGC GCTCTGTGGG GAGGAAGCTC TGAGTCAACA TCTCGAGTCG 72821525
GCCTTGACAA AACACAGAAC AATCACGAGA GCAGCAAGAA ACGCCAAAGA 72821475
GCACCCTAGT TTATTACCTC ACTCTGCCTG CTTCCCGAT CCTAGCACCG 72821425
CATCTACCTC GCAGTCTGCC GCTCACTCAA ACGACAGCCC CCTCCCCCG 72821375
TCGGCCGCCG CCCCCTCCTC CGCTTCCCC CACGCCTCCA GGAAGTCTTG 72821325
GCCGCAAGTG GTCTCCCGGG CTTCGGCAGC GAAGCCCCCT TCTTTTCTC 72821275
CTCTCTCCTC ATCTTCAACG GTTACCTCAA GTTCATGCAG CACCTCAGGG 72821225
GTTACAGCCCT CGATGCCAAC AGACGACTAT TCGGAGGAGT CTGACACGGA 72821175
TCTCAGCCAA AAGTCCGACG GACCGGCGAG CCCGGTGGAG GGTCCCAAAG 72821125
ACCCAGCTG CCCCAGGAC AGTGGTCTGA CCAGTGTAGG AACGGACACC 72821075
TTCAGATTGT AAGCTTTGAA GATGAACAAT ACAACAAAT GAATTTAAAT 72821025
ACAAAATTA ATAACAAACC AATTTCAAAA ATAGACTAAC TGCAATTCCA 72820975
AAGCTTCTAA CCAAAAAACA AAAAAAATA AAAAAAGAAA AAAAAAGAAA 72820925
AGCGTGGGTT GTTTTCCCAT ATACCTATCT ATGCCGGTGA TTTTACATTC 72820875
```

**Figure 2.33: Conserved domain of the ZFH3 gene identified in Family 2081**

Clustal W was used to align orthologs from human (NP\_008816.3), dog (XP\_546849.2), mouse (NP\_031522.2), rat (XP\_226464.3), chicken (XP\_414230.2), and zebrafish (XP\_688934.3). The string of 14 glycine residues is not conserved across five species (A) or the NL population controls (B). C shows a trace of the ZFH3 protein showing 15 glycine residues.

**A: Evolutionary evidence of expanding glycine-rich region**

Human	<a href="#">NP_008816.3</a>	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
Dog	<a href="#">XP_546849.2</a>	3499	HLKSLCFFGQSVVNLQEMVLHVPTGGGGG--GGGVGGGGGGGGGGGSYHCL	3546
Mouse	<a href="#">NP_031522.2</a>	3507	HLKSLCCFGQSVVNLQEMVLHVPTGSG----GSGGGGGSGGGGGGSYHCL	3552
Rat	<a href="#">XP_226464.3</a>	3514	HLKSLCFFGQSVVNLQEMVLHVPTGSG----GGGGGGSGGGGGGSYHCL	3559
Chicken	<a href="#">XP_414230.2</a>	3455	HLKSLCFFGQSVANLQEMVLHVPTGSG-----QGSGSYQCV	3490
Zebrafish	<a href="#">XP_688934.3</a>	2387	HLKSICFFGQSVANLQEMLLRVPNSGN-----AAEGGLYDCL	2423

**B: Alleles identified in population controls at the expanding glycine- rich region**

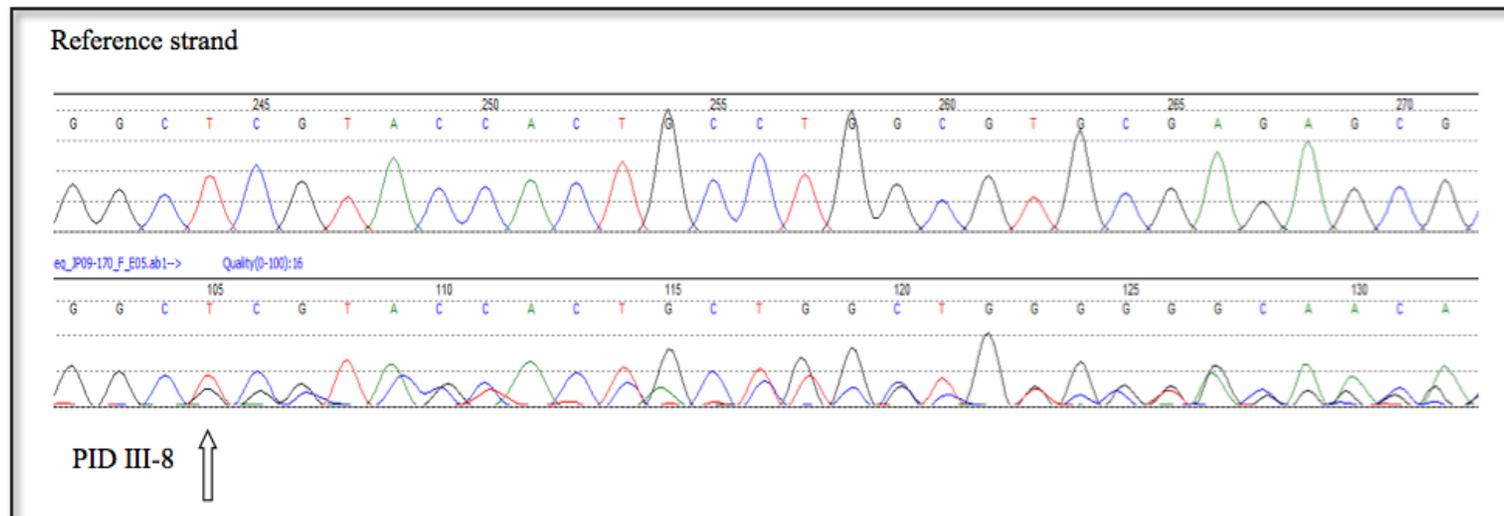
Human	<a href="#">NP_008816.3</a>	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
NL Pop Control	(del 1 Gly)	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
NL Pop Control	(del 3 Gly)	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
NL Pop Control	(del 5 Gly)	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
NL Pop Control	(del 6 Gly)	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
NL Pop Control	(del 1 Gly)	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
NL Pop Control	(del 1 Gly)	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532
NL Pop Control	(del 3 Gly)	3483	HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL	3532

**C: Alleles identified in patients with otosclerosis show insertion of a glycine residue**

NL OTSC proband (ins 1 Gly) 3483 HLKSLCFFGQSVVNLQEMVLHVPTGGGGGGSGGGGGGGGGGGGGGSYHCL 3532

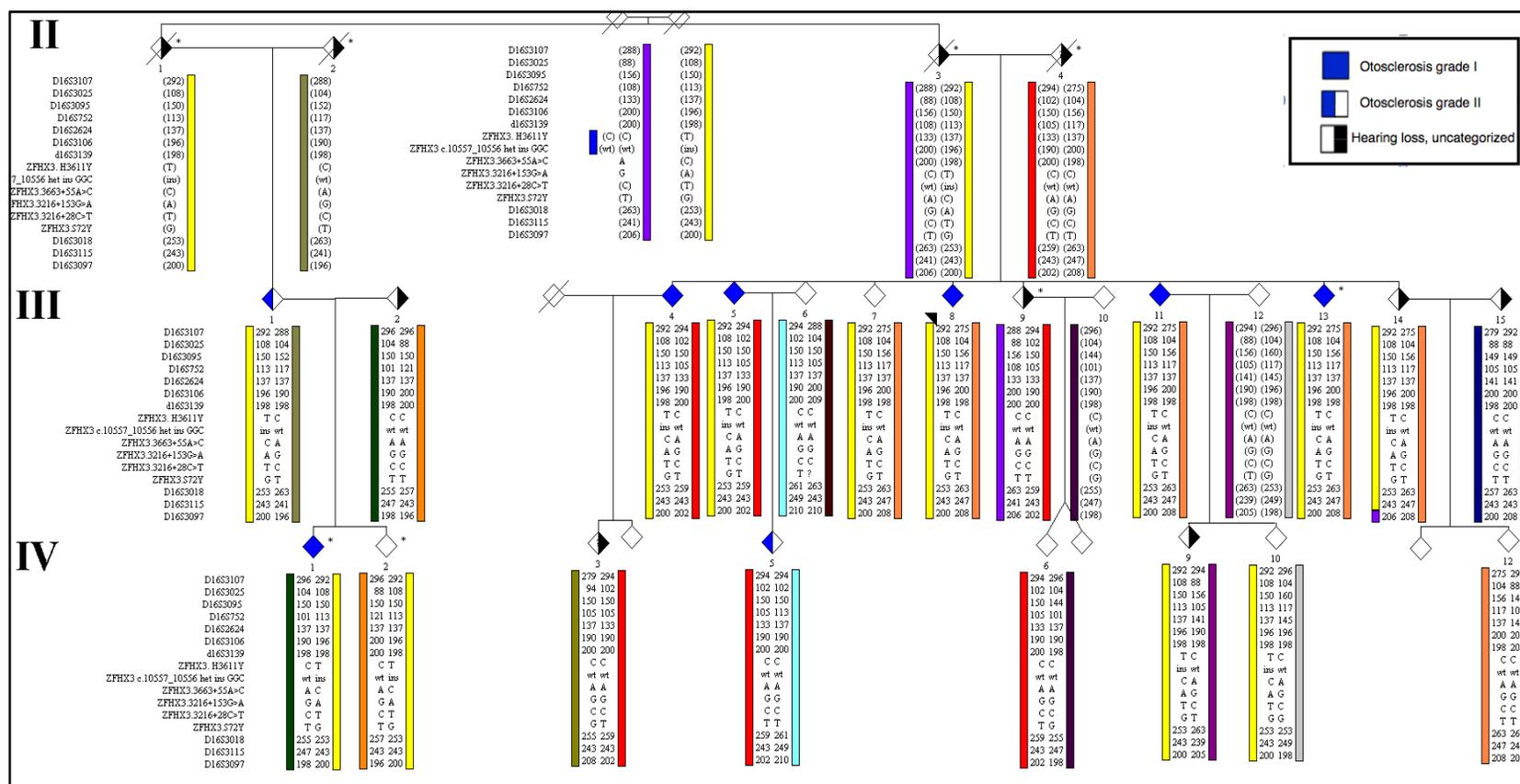
**Figure 2.34 : Electropherogram shows no degradation of the *ZFH3* c.10557\_10558insGGC allele**

The upper panel is the reference sequence of *ZFH3*. The lower panel is the cDNA sequence trace from PID III-8 showing the 3 bp insertion and resultant frameshift (shown by the arrow). The allele with the 3 bp insertion has been transcribed and therefore, did not undergo degradation.



**Figure 2.35: Family 2081 shows segregation of five variants detected in *ZFHX3* on the *OTSC4* yellow haplotype**

The five variants (c.3216+28C>T, c.3216+153G>A, c.10831C>T, p. H3611HY and c.10557-10558insGGC) found in the *ZFHX3* gene reside on the yellow disease haplotype. Roman numerals on the side of the pedigree indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. Circle = female, square = male. Alleles sizes are given in base pairs and alleles in brackets are inferred.



**Table 2.3: Summary of allele frequency for variants identified in *ZFHX3* that segregate with otosclerosis in Family 2081**  
 Frequencies of these variants were determined in 314 and 176 NL population controls for c.10557\_10558het insGGc and c.10831C>T respectively.

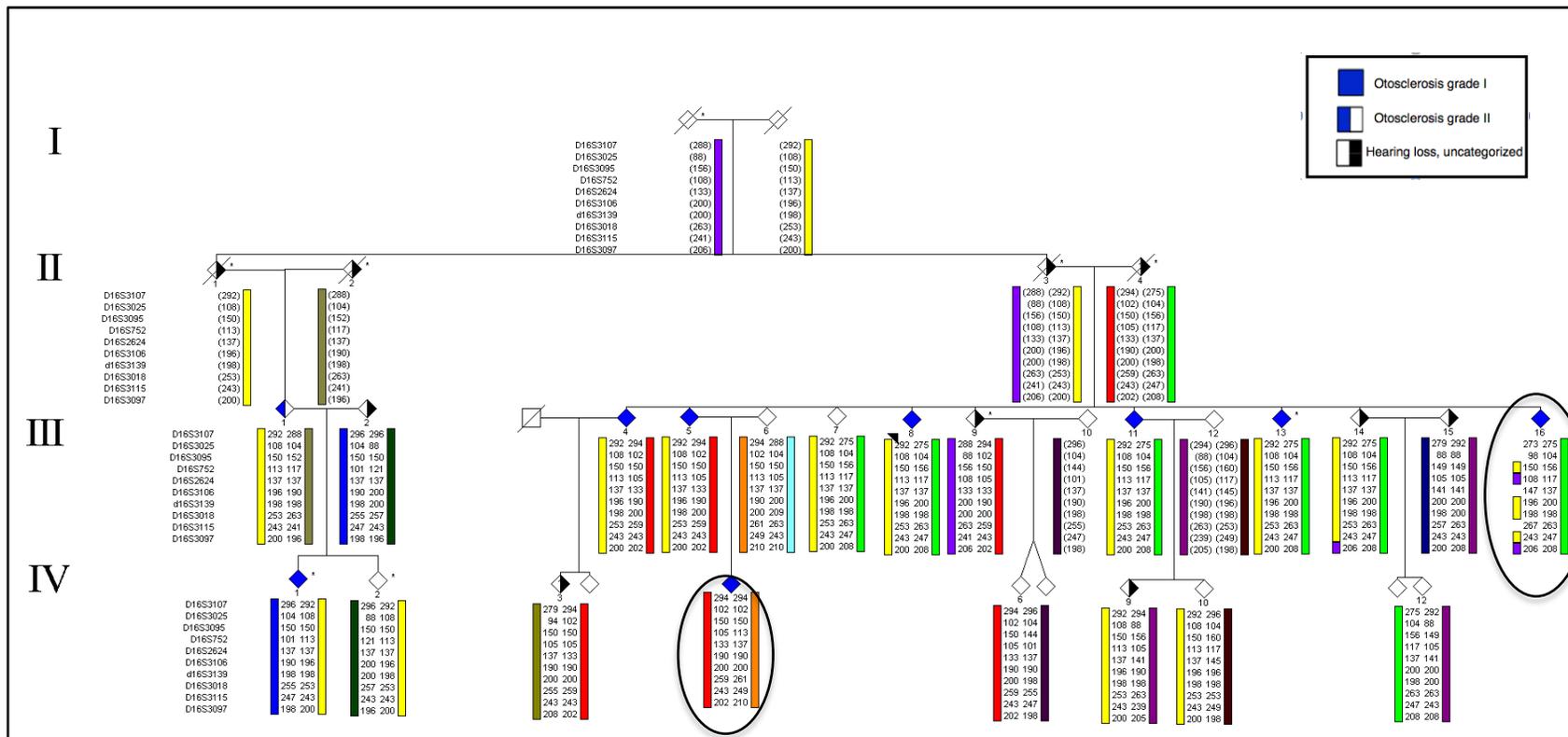
<b>Genomic Change</b>  (NM_006885)	<b>Protein</b>  <b>Level</b>	<b>SNP database</b>  <b>Update Feb2/2010</b>	<b>Freq in NL controls</b>  <b>Ratio (%)</b>	<b>SIFT Prediction</b>	<b>Polyphen</b>  <b>Prediction</b>
<b>c.10557_10558insGGC</b>	p.Gly3527_ser3528insGly	Novel	5/314 (1.5)	n/a	n/a
<b>c.10831C&gt;T</b>	p.H3611Y	rs200992486	2/176 (1.13)	Deleterious	Probably damaging

### **2.3.6 Recruitment of a new family member and excluding linkage of Family 2081 to *OTSC4***

In the process of the study, PID III-16 with surgically confirmed otosclerosis was recruited and screened for the segregation of the presumed disease haplotype (yellow haplotype) across *OTSC4* and for the presence of the two candidate variants. The family member PID III-16 did not inherit the presumed disease (yellow) haplotype or the two candidate variants. In addition, PID III-16 did not inherit a complete set of paternal haplotypes and thus has been assigned as a case of non-paternity (Figure 2.36). At this stage, we also received a confirmation of otosclerosis in PID IV-5, who had been diagnosed as grade II, but based on visualization of stapes fixation during surgery. This report has changed the diagnosis of PID IV-5 from grade II to grade I. At this stage, two affected family members with grade I otosclerosis (PIDs III-16 and IV-5) did not inherit the previously assigned disease (yellow) haplotype, which decreases the possibility that *OTSC4* was the disease locus for Family 2081.

**Figure 2.36: Partial pedigree of Family 2081 showing absence of segregation of the yellow haplotype in two affected subjects**

Two subjects IV-5 and III-16 (circled) did not inherit the yellow haplotype. Roman numerals on the side indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. Circle = female, square = male. Allele sizes are given in base pairs) and alleles in brackets are inferred.



### **2.3.7 Two point linkage analysis**

The exclusion of the previously mapped loci and the region spanning the three candidate genes was based on haplotypes. To determine the statistical significance of our exclusion data, two point linkage analyses were carried out for the otosclerosis loci (*OTSC1*, *OTSC2*, *OTSC3*, *OTSC4*, *OTSC5*, *OTSC7*, *OTSC8*, *OTSC10*) and for the three genes (*COLIA1*, *COLIA2* and *NOG*) using three microsatellite markers for each locus for Family 2081. At each locus, a negative LOD score was obtained at all recombination fractions (ranging from 0.000 to 0.5000). Statistically significant linkage exclusion of *OTSC1*, *OTSC2*, *OTSC3*, *OTSC4*, *OTSC5*, *OTSC7*, *OTSC8* and *OTSC10* loci and three genes was demonstrated in Family 2081 because the LOD scores were negative at almost all examined loci (Table 2.4).

**Table 2.4: Two point LOD score in Family 2081**

Two point LOD scores generated by the MLINK program between otosclerosis and three markers within each of the published *OTSC* loci and three otosclerosis associated genes.

Locus	Markers	Recombination fraction					
		0.000	0.100	0.200	0.300	0.400	0.500
<i>OTSC1</i>	<i>D15S127</i>	-4.096936	-0.887153	-0.387585	-0.151425	-0.035454	0.000000
	<i>D15S649</i>	-4.096936	-0.264114	-0.108875	-0.048774	-0.014267	0.000000
	<i>D15S657</i>	-0.300990	-0.099513	-0.029093	-0.005594	-0.000348	0.000000
<i>OTSC2</i>	<i>D7S495</i>	0.301029	0.212668	0.130215	0.062729	0.016819	0.000000
	<i>D7S1798</i>	-4.096936	-0.887153	-0.387585	-0.151425	-0.035454	0.000000
	<i>D7S1827</i>	-13.494847	-1.33082	-0.581401	-0.227145	-0.053183	0.000000
<i>OTSC3</i>	<i>GAAT3A06</i>	-3.795907	-0.887043	-0.387574	-0.151424	-0.035454	0.000000
	<i>D6S273</i>	-3.795907	-0.887043	-0.387574	-0.151424	-0.035454	0.000000
	<i>D6S1680</i>	-3.795907	-0.653891	-0.226226	-0.062029	-0.007427	0.000000
<i>OTSC4</i>	<i>D16S3107</i>	-4342935	-0.766770	-0.300033	-0.086144	0.002510	0.000000
	<i>D16S3106</i>	-4342935	-0.636432	-0.200119	-0.038453	0.003295	0.000000
	<i>D16S3097</i>	-4342935	-0.598769	-0.218146	-0.063158	0.001094	0.000000
<i>OTSC5</i>	<i>D3S3548</i>	0.301025	0.210206	0.118188	0.044152	0.006429	0.000000
	<i>D3S1593</i>	-4.096936	-0.755097	-0.299465	-0.104434	-0.021215	0.000000
	<i>D3S1744</i>	-3.795907	-0.653881	-0.226224	-0.062028	-0.007427	0.000000
<i>OTSC7</i>	<i>D6S1619</i>	-3.795906	-0.638142	-0.218710	-0.059726	-0.007201	0.000000
	<i>D6S1595</i>	-4.096936	-0.264114	-0.108875	-0.048774	-0.014267	0.000000
	<i>D6S268</i>	-3.795906	-0.638142	-0.218710	-0.059726	-0.007201	0.000000
<i>OTSC8</i>	<i>D9S2148</i>	-4.045783	-0.199700	-0.047532	-0.007852	-0.000417	0.000000
	<i>D9S1862</i>	-13.494847	-0.672523	-0.254094	-0.086982	-0.018424	0.000000
	<i>D9S1799</i>	-0.300990	-0.099513	-0.029093	-0.005594	-0.000348	0.000000
<i>OTSC10</i>	<i>DIS2621</i>	-4.096936	-0.264114	-0.108875	-0.048774	-0.014267	0.000000
	<i>DIS2800</i>	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	<i>DIS2811</i>	-4.096936	-0.264114	-0.108875	-0.048774	-0.014267	0.000000
<b>COLIA1</b>	<i>D17S797</i>	-0.300990	-0.248075	-0.151917	-0.069146	-0.017365	0.000000
	<i>D17S809</i>	-4.096936	-0.443629	-0.193798	-0.075713	-0.017727	0.000000
	<i>D17S788</i>	-4.096936	-0.443629	-0.193798	-0.075713	-0.017727	0.000000
<b>COLIA2</b>	<i>D7S644</i>	-3.795907	-0.013984	0.073263	0.053196	0.016338	0.000000
	<i>D7S821</i>	-3.795902	0.025915	0.139021	0.118035	0.047724	0.000000
	<i>D7S651</i>	0.602028	0.465356	0.318042	0.170249	0.049214	0.000000
<b>NOG</b>	<i>D17S790</i>	-4.698990	-0.691749	-0.345732	-0.144865	-0.035094	0.000000
	<i>D17S1606</i>	-4.096936	-0.755097	-0.299465	-0.104434	-0.021215	0.000000
	<i>D17S1161</i>	-4.096936	-0.755097	-0.299465	-0.104434	-0.021215	0.000000

## 2.4 Discussion

Eight otosclerosis loci have been published and no causative genes have yet been identified. Large multigenerational Families (2081 and 2114) with apparent AD otosclerosis, and seven unrelated otosclerosis probands from NL, were investigated for linkage to the eight previously mapped otosclerosis loci and to three candidate genes (*COL1A1*, *COL1A2* and *NOG* genes).

Otosclerosis segregating in Family 2081 was excluded from linkage to *OTSC1*, *OTSC2*, *OTSC3*, *OTSC5*, *OTSC6*, *OTSC7*, *OTSC8*, *OTSC10*, *COL1A1*, *COL1A2* and the *NOG* gene through haplotype analysis and two-point linkage analysis. In the initial phase of the genetic study of Family 2081, individuals diagnosed with otosclerosis showed apparent segregation of a presumed disease (yellow) haplotype inherited from the parent II-3. Assuming that this was a single founder disease haplotype, six unrelated otosclerosis probands were genotyped across *OTSC4*, which minimized the region from approximately 9 Mb to 4.2 Mb. Twelve genes in this region were sequenced and a missense and a 3 bp insertion variant in the *ZFHX3* gene were considered candidate variants for otosclerosis in Family 2081 based on the allele frequencies and pathogenicity prediction. Further investigation on the effect of the glycine insertion in the *ZFHX3* gene did not suggest any drastic effect at the mRNA level. Of these two variants, the missense mutation seems to be the candidate variant over the 3 bp insertion. To investigate the role of candidate variants in NL families with otosclerosis, nine probands with otosclerosis were screened and none of them harboured either the missense mutation or the insertion. The four otosclerosis probands used to narrow down the *OTSC4* region did not harbour any of the two variants which suggested that these probands did not share a founder

haplotype with Family 2081 and the shared haplotype was explained as a common haplotype in the NL population.

Recruitment of a new member of Family 2081 (PID III-6) with grade I otosclerosis who, along with subject IV-5, did not inherit the disease haplotype, combined with the negative LOD score across *OTSC4*, excluded linkage of Family 2081 to *OTSC4*. Additionally, subject III-16 did not inherit a complete paternal haplotype, which suggested a case of non-paternity and served to exclude linkage of family 2081 to the originally mapped *OTSC4*. The identified variant in *ZFHX3* could cause unrelated disease. The *ZFHX3* gene was chosen for sequencing based on its role as a transcription factor and involvement in cytoskeleton organization. Variants in this gene have also been associated with atrial fibrillation and may be causative in families with cardiac arrhythmia<sup>149 150</sup>.

Family 2114 is a large family from NL with a presumed dominant mode of inheritance. Haplotype studies of Family 2114 excluded linkage of this family to seven previously mapped loci and to the loci of three associated genes, presuming the mother was the disease carrier. In the process of the study, we noticed that individuals with a confirmed diagnosis of otosclerosis in Family 2114 shared a region from 86 Mb to 143 Mb of (57.3 Mb) (86Mb-143 Mb) at Chr 7q21.11-q35, which overlaps with the two loci *OTSC2* and *COLIA2*, inherited from the father's side. The father of this family (PID II-1) had no history of hearing loss. If this clinical finding for this subject was true, this would exclude linkage of this family to the seven mapped loci and the three associated genes from both sides. On the other hand, the absence of a history of hearing loss would not exclude non-penetrance, or a *de novo* mutation in the father's germline cells. These

findings suggested the presence of at least one novel otosclerosis locus for this family if the mother was the disease carrier. Alternately, the otosclerosis gene could be in the 57.3 Mb region at Chr7 (q21.11-q35) which overlaps *OTSC2* and *COLIA2*.

The presence of phenocopies in a pedigree affects linkage analysis. In the present study we faced a diagnostic problem with subject IV-5 of Family 2081. This subject was diagnosed with otosclerosis based on their audiological profile. Lack of a definitive diagnosis (actual visualization of stapes fixation) of otosclerosis in subject IV-5 allows us to exclude the genetic analysis of this subject in our final exclusion/Inclusion decision. Only subjects with otosclerosis confirmed by a stapedectomy were considered for the analysis. By doing so, *OTSC4* appears to be the disease locus for Family 2081, when subject IV-5 is considered a phenocopy. This decision was subsequently reversed after later confirmation that subject IV-5 has grade 1 otosclerosis.

Otosclerosis has been characterized by reduced penetrance in previously mapped families, which complicates diagnosis<sup>86,115,151</sup>. Reduced penetrance can be attributed to age-dependent expression, environmental and/or genetic modifiers. Otosclerosis usually develops between the third and sixth decade (late onset hearing loss). The severity of hearing loss in otosclerosis depends on the progression of the otosclerosis foci.

To determine the frequency of possible otosclerosis disease alleles in the NL population we used population samples from a colorectal cancer study cohort. People who are recruited for colorectal cancer studies are not evaluated for hearing loss (otosclerosis) and the main drawback of using this cohort is under/over estimation of the allele frequency of a variant. The most acceptable cohort for determining the allele

frequency of possible otosclerosis variants would include people who were assessed audiologically and had no hearing loss.

## **2.5 Conclusion**

In this chapter, linkage of both Family 2081 and 2114 to previously linked loci and three candidate genes was performed using a positional mapping technique. We are reporting exclusion of linkage of Family 2081 to any of the previously mapped loci. Also, Family 2114 was excluded from linkage to the seven previously mapped loci and three candidate genes presuming the mother is the disease carrier. This result raises the possibility of identifying a novel locus for both families. We also have illustrated the obstacles that complicate the study of otosclerosis and how this study has dealt with these obstacles.

### **3 Chapter 3**

**A fifteen bp deletion in the *FOXL1* gene causes familial otosclerosis in one family from NL and one from ON**

### **3.1 Introduction**

In Chapter 2, Family 2081 was excluded from linkage to previously published loci and three candidate genes using haplotype and two point linkage analyses. The key recombination that minimized the *OTSC4* loci in the original publication was detected in a family member with mild conductive hearing loss which was not surgically confirmed. This raised the possibility of incorrect boundaries for *OTSC4*. In this Chapter, the distal boundary of *OTSC4* was extended by 21.1 Mb through genotyping and haplotype analysis. Twelve functional candidate genes were sequenced and five samples were sent for whole exome sequencing. Targeted analysis of variants identified in the candidate region at Chr16q24 was carried out. Functional analysis was also done on the 15 bp deletion in the *FOXLI* gene (Appendix 17).

### **3.2 Subjects and methods**

#### **3.2.1 Study population and clinical analysis**

Family 2081 is a multiplex family with seven confirmed cases of otosclerosis segregating in an AD mode of inheritance. Twenty-one DNA samples from Family 2081 and 81 Canadian otosclerosis probands (41 probands from NL and 40 probands from Western ON, contributed by Dr. Stanton) were available for the study. Otosclerosis in family 2081 is classified according to the grading system (Chapter 2; pages 56-57).

#### **3.2.2 Genotype and haplotype analysis of Family 2081**

Genomic DNA was extracted from peripheral leukocytes of all participants according to a standard method (Appendix 2). The *OTSC4* haplotype was expanded by genotyping 21 members of Family 2081 with nine polymorphic markers (*D16S518*,

*D16S3094, D16S3098, D16S2625, D16S520, D16S422, D16S413, D16S3023* and *D16S3026*) that span 12.1 Mb, downstream of the lower boundary of the originally mapped *OTSC4* locus. The proband from the ON family (IV-3) was genotyped with the nine polymorphic markers. All markers were labelled with blue fluorescent dye (6-FAM). DNA amplification and genotyping were carried out as described in Chapter 2 (page 58-59).

### **3.2.3 Sequencing genes within the candidate region**

The UCSC Genome Browser homepage<sup>141</sup> and the March 2006 assembly (NCBI build 36.1) were used to define the genomic interval of each candidate region and to identify functional candidate genes. For primer design and PCR reaction, please refer to Chapter 2 (pages 60 and 61). Primer sequences for the 12 genes are outlined in Appendix 13.

#### **3.2.3.1 Mutation screening**

The mutation screening panel for this locus was comprised of three affected siblings who carried the presumed disease haplotype (PID III-5, III-11, IV-5), one non-disease carrier spouse (PID III-6), and one water control. This panel screened for the following genes *PLCG2, IRF8, SCL38A8, ZDHHC7, SLC7A5, HSD17B2, COTL1, FOXF1, FOXL1, FOXC1, CA5A* and *OSGIN1*. For segregation analysis, variant allele frequency and bioinformatics analysis please refer to Chapter 2 (pages 62).

### **3.2.4 Whole exome sequencing of five family members of Family 2081**

After Sanger sequencing of candidate genes in the linked region, five samples from family 2081 were sent for whole exome sequencing to validate the result and to fully cover the linked region at Chr16q24. Whole exome sequencing was chosen as the most cost-effective approach.

#### **3.2.4.1 Mutation screening panel**

Whole exome sequencing was carried out for five affected siblings of Family 2081 (PIDs III-4, III-8, III-11, III-13, III-16) and two unaffected population controls which were assigned as unaffected, based on audiogram reports of normal hearing (at ages 55 and 60). The Illumina platform was chosen due to higher coverage of the coding region. Samples were sent to Genome Quebec for library preparation, enrichment and whole exome sequencing. DNA libraries were prepared using the TrueSeq DNA Sample Prep Kit. Exome enrichment was carried on prepared libraries. Paired end reads of 100 bp were sequenced on an Illumina HiSeq following standard manufacturer's protocol. Approximately 50 to 150 million 100 bp paired-end reads from the Illumina HiSeq 2000 sequencer were generated. All reads were required to have a length of at least 32 bp. The filtered reads were then aligned to the 1000 genome project reference sequence, which is currently used by Genome Quebec. Variants were annotated for dbSNP and for the variants effect on the gene (amino acid change). Finally the sample files were integrated into global files. The full pipeline report for NGS is attached in Appendix 14. Primers used for validating variants of interest are listed in Appendix 24.

### **3.3 Results**

#### **3.3.1 Pedigree structure and clinical analysis**

Family 2081 is a multiplex family with seven members affected across two generations. Otosclerosis in Family 2081 appears to be inherited in an autosomal dominant mode. For further information on the pedigree structure and clinical analysis, refer to Chapter 2 (pages 68-69 and 71-73).

#### **3.3.2 Genetic analysis**

##### **3.3.2.1 Family 2081 is linked to a 9.7 Mb region downstream of *OTSC4***

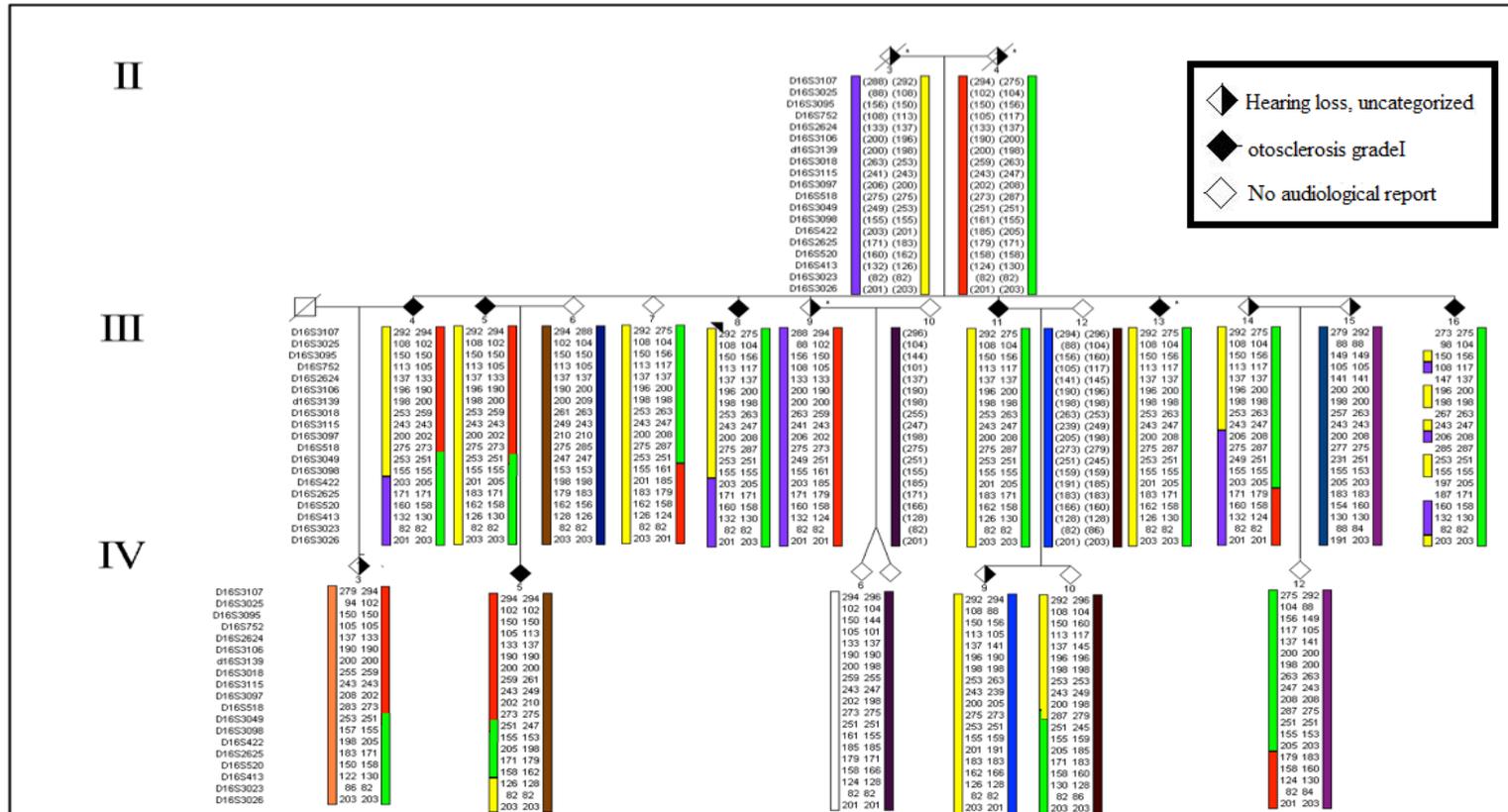
Family 2081 was excluded from linkage to the eight previously mapped loci (*OTSC1*, *OTSC2*, *OTSC3*, *OTSC4*, *OTSC5*, *OTSC7*, *OTSC8* and *OTSC10*) and three candidate genes (*COL1A1*, *COL1A2* and *NOG*) through haplotype and linkage analyses (Chapter 2). Initially, haplotype analysis in Family 2081 across the *OTSC4* locus suggested linkage to *OTSC4* but haplotype analysis of an extra Family member (PID III-16) and a negative LOD score excluded this possibility. In the previous Chapter, we tested the linkage of Family 2081 to the fine mapped region of *OTSC4*, not to the entire reported *OTSC4* locus. In the original study by Brownstein et al, distal recombination in one subject with mild conductive hearing loss changed the boundary of *OTSC4* from *D16S3091* (82.7 Mb) to *D16S3097* (77.1 Mb), which excluded 5.6 Mb from analysis<sup>87</sup>. As the distal recombination was identified in a person with mild conductive hearing loss and not confirmed surgically, I decided to expand the lower boundary of *OTSC4*. To do that, in addition to the 10 microsatellite markers used to genotype the *OTSC4* locus, all available family members of Family 2081 were genotyped with nine markers that covered

12.1 Mb downstream of the distal boundary (*D16S3097*) of the minimized *OTSC4* locus. Four distinct parental haplotypes were created (purple, yellow, red and light green) and the parental haplotypes of II-3 and II-4 were inferred (Figure 3.1).

All affected members shared a 9.7 Mb region of the light green haplotype between *D16S518*, at position 78.1Mb and *D16S413*, at position 87.8 Mb (Figure 3.1 and Table 3.1). Two Family members (PIDIII-14 with uncategorized hearing loss and PID IV-12 with no history of hearing loss) inherited the upper portion of the green haplotype between *D16S518* and *D16S2625*. PID III-14 shared a 5.1Mb region of the 9.7 Mb linked green haplotype that covered 3 markers *D16S3049*, *D16S3098* and *D16S422* (Table 3.1). If we presume that PIDIII-14 and IV-12 are non disease carriers, the size of the 9.7 Mb linked region would reduce to 4.6 Mb. Otosclerosis is characterized by a wide range of clinical expression (SNHL, conductive, mixed hearing loss) and reduced penetrance. For that reason, we did not designate PIDIII-14 and IV-12 as non-disease carriers and therefore, the 9.7 Mb region was fully examined.

**Figure 3.1: Family 2081 co-segregating a 9.7 Mb haplotype telomeric to *OTSC4* mapped locus**

All affected members share a 9.7 Mb region of the light green haplotype. A critical recombination between *D16S413* and *D16S520* in PID IV-5 minimized the linked region from 11.35 to 9.7 Mb. Alleles sizes are given in base pairs and alleles in brackets were inferred. The gender of family members is hidden for confidentiality. Roman numerals on the side of the pedigree represent the generation number.



**Table 3.1: Family 2081 affected members co-segregate a 9.7 Mb haplotype**  
 Bolded markers are the markers shared by all grade1 affected members of Family 2081.  
 Surgically confirmed affected individuals are shaded black. All shaded black subjects share a region between markers *D16S518* and *D16S413*.

Markers	Position Mb						IV-3		IV-10		III-14	IV-12
<i>D16S3107</i>	66.2	294	294	275	275	275	294	294	292	275	275	275
<i>D16S3025</i>	67.12	102	102	104	104	104	102	102	108	104	104	104
<i>D16S3095</i>	68.50	150	150	156	156	156	150	150	150	156	156	156
<i>D16S752</i>	70.74	105	105	117	117	117	105	105	113	117	117	117
<i>D16S2624</i>	71.73	133	133	137	137	137	133	133	137	137	137	137
<i>D16S3106</i>	71.26	190	190	200	200	200	190	190	196	200	200	200
<i>D16S3139</i>	72.7	200	200	198	198	198	200	200	198	198	198	198
<i>D16S3018</i>	73.08	259	259	263	263	263	259	259	253	263	263	263
<i>D16S3115</i>	74.4	243	243	247	247	247	243	243	243	247	247	247
<i>D16S3097</i>	77.3	202	202	208	208	208	202	202	200	208	208	208
<i>D16S518</i>	78.3	273	273	288	288	288	273	273	287	288	288	288
<b><i>D16S3049</i></b>	<b>78.9</b>	251	251	251	251	251	251	251	251	251	251	251
<b><i>D16S3098</i></b>	<b>81.4</b>	155	155	155	155	155	155	155	155	155	155	155
<b><i>D16S422</i></b>	<b>82.9</b>	205	205	205	205	205	205	205	205	205	205	205
<i>D16S2625</i>	83.2	171	171	171	171	171	171	171	171	171	179	179
<b><i>D16S520</i></b>	<b>85.0</b>	158	158	158	158	158	158	158	158	158	158	158
<i>D16S413</i>	87.7	130	130	130	130	130	130	126	130	130	124	124
<i>D16S3023</i>	88.4	82	82	82	82	82	82	82	82	82	82	82
<i>D16S3026</i>	89.3	201	201	201	201	201	201	201	201	201	201	201

### **3.3.2.2 Mutation screening of functional candidate genes by Sanger sequencing**

According to UCSC genome browser (Feb2009 /hg19), the new locus located at Chr16q23.1-q24.2 contains 79 annotated genes. Due to the large number of positional candidate genes in the linked region, genes were prioritized for sequencing according to their expression and function related to bone remodelling. Genes that are expressed in the bone or immune system were prioritized for screening. In total, 12 candidate genes were sequenced. The function of each gene is described in Appendix 15. All twelve genes were screened using the following mutation panel: three affected siblings (PIDs III-5, III-11, and IV-5) and a spouse with no history of hearing loss (PID III-6).

#### **3.3.2.2.1 Variant analysis**

Variant analysis of Family 2081 at this region was done under a dominant mode of inheritance. Under this mode of inheritance we would expect affected siblings (PIDs III-5, III-11, and IV-5) to carry a heterozygous mutation that is inherited from the parent PID II-4 and absent from the unaffected spouse PID III-6.

Sequencing of the 12 genes using these four samples revealed 92 variants (Figure 3.2 and appendix 16A). Of these ninety-two variants, 53 variants were present in the three affected siblings (PIDs III-5, III-11 and IV-5) as well as the unaffected spouse control (PID III-6). For example, c.26C>T in the *CA5A* gene was detected in the three affected siblings PIDs III-5, III-11 and IV-5 as well as the unaffected spouse (PID III6) of Family

2081. Ten variants were present in only two out of three affected siblings. For example, c.770+125G>A in *SLC7A5* was detected in two affected siblings PID III-5 and III-11 but it was not detected in the third affected sibling PID IV-5. Eleven variants were identified in one( PID IV-5), of 3 affected members and in one unaffected member (PID III-6). For example, c.218-104C>T in *OSGIN1* exon 3 was found in the affected member PIDIV-5 and in the unaffected member PID III-6 of Family 2081. Ten variants were found in the unaffected member PID III-6. For example, c.1332C>A in exon 7 of *OSGIN1* was found in the unaffected control member PIDIII-6. In total, eight variants were found in the three affected siblings (PIDs III-5, III-11 and IV-5) but not in the unaffected control PID III-6.

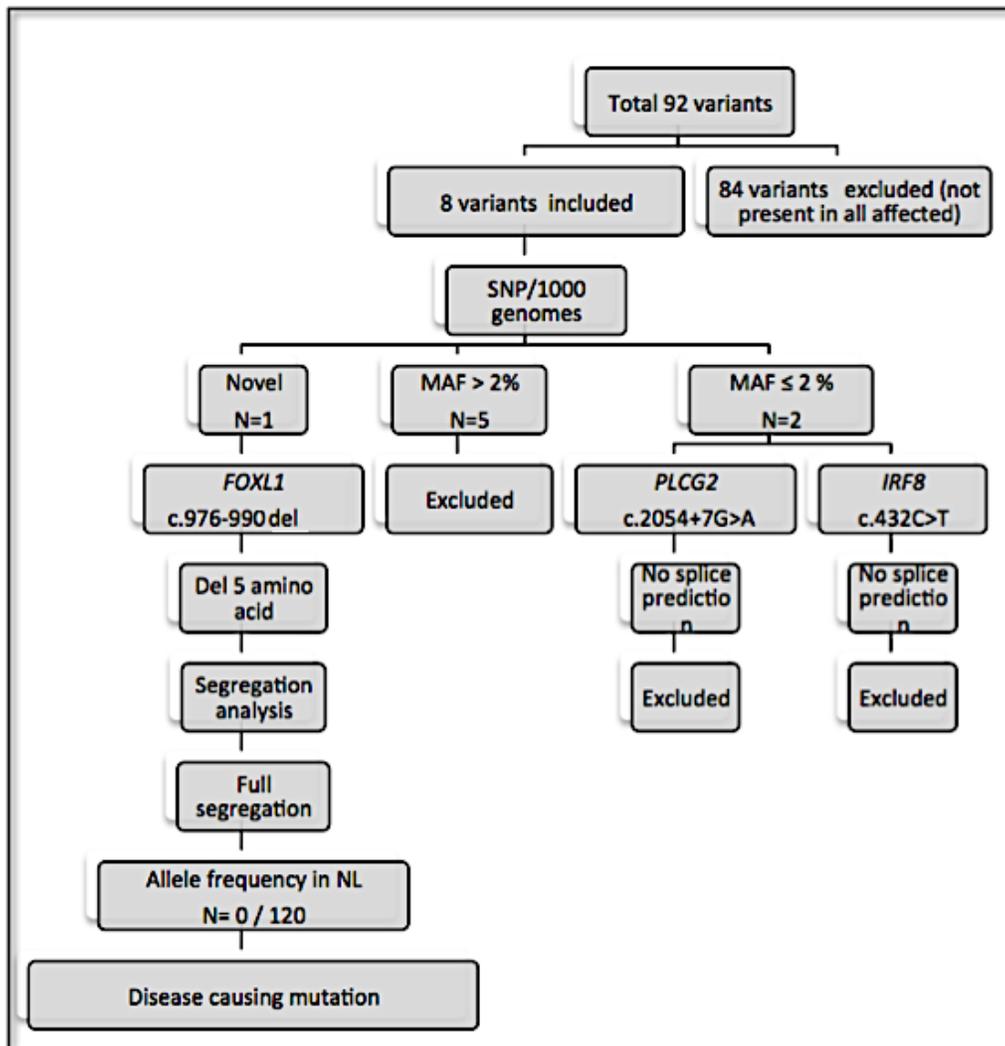
To determine candidate variants, allele frequencies of the eight variants were checked using SNP/1000 genomes project database. Two variants; c.2054+7G>A in the *PLCG2* gene and c.432C>T, p. D144 in the *IRF8* gene had frequencies of 2.1 and 1.5 % respectively (Table 3.2). Five were SNPs with frequencies ranging between 11 and 86 %. Only *FOXL1* c.976-990del in *FOXL1* was novel (Figure 3.3). The variants c.2054+7G>A in *PLCG2*, c.432C>T in *IRF8*, and c.976-990 het del in *FOXL1* were subjected to further analysis.

The novel *FOXL1* c.976-990del, p. p.G327\_L331del was predicted to cause an in-frame loss of five amino acid residues (glycine, isoleucine, proline, phenylalanine and leucine), which were highly conserved across five species (Figure 3.4). The silent and intronic variants were not predicted to effect splice sites. Only the 15 bp deletion in the *FOXL1* gene was subjected to segregation analysis.

All members of Family 2081 with available DNA were tested for segregation of the 15 bp deletion in *FOXL1*. This showed that the 15 bp deletion in the *FOXL1* gene was segregating with the lower portion of the light green haplotype between *D16S 520* and *D16S413* (Figure 3.5) in the seven grade I otosclerosis members, PID IV-10 and IV-3. Two family members (PID III-14 and IV-12), who did not inherit the lower part of the green haplotype, did not harbour the *FOXL1* 15 bp deletion. To check the frequency of the 15 bp deletion in the NL population, 116 ethnically matched population controls were screened and all were wild type for the deletion.

**Figure 3.2: Flowchart showing the steps used to analyse variants identified in Family 2081, using Sanger sequencing**

Only the 15 bp deletion in the *FOXL1* gene fully segregated with otosclerosis in Family 2081. Allele frequencies of identified variants were determined by using the publicly available SNP database and 1000 genomes project. Pathogenicity of missense mutations was evaluated by using SIFT and Polyphen prediction programs. Five prediction programs (SSF, HSF, MaxEntScan, NNSPLICE, and GeneSplicer) were used to predict any splicing effect.

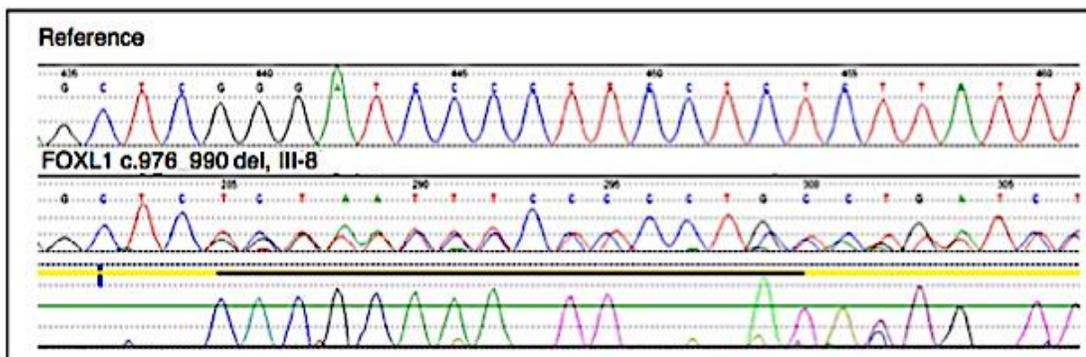


**Table 3.2: Allele frequency of variants residing on the light green haplotype**  
 Frequencies of the identified variants within the linked region were estimated by SNP/1000 genomes databases. The novel variant in the *FOXLI* gene is bolded.

Variant information				Allele frequencies using SNP database		
Gene name	Position	Variants	SNP status	Allele frequency	Sample size	Frequency
<i>PLCG2</i>	Int 20	c.2054+7G>A	<i>rs138158454</i>	A=0.021	46	2.1 %
<i>PLCG2</i>	Ex10	802C>T, p.R28RW	<i>rs1143687</i>	T=0.11	184	11%
<i>PLCG2</i>	Ex12	987-97G>C	<i>rs4889430</i>	C=0.113	2184	11.3%
<b><i>FOXLI</i></b>	<b>Ex1</b>	<b>976_990del,</b> p. G327_L331del	<b>Novel</b>	<b>Novel</b>	<b>0</b>	<b>0</b>
<i>IRF8</i>	Ex4	432C>T, p.D144	<i>rs16939945</i>	T=0.015	33	1.5%
<i>ZDHHC7</i>	Ex3	315+55G>A	<i>rs93173</i>	A=0.545	22	54.5%
<i>ZDHHC7</i>	Ex7	538-11C>T	<i>rs66772728</i>	T=0.867	120	86%
<i>HSD17B2</i>	Ex1	-133T>C	<i>rs4445895</i>	C=0.472	286	47.2%

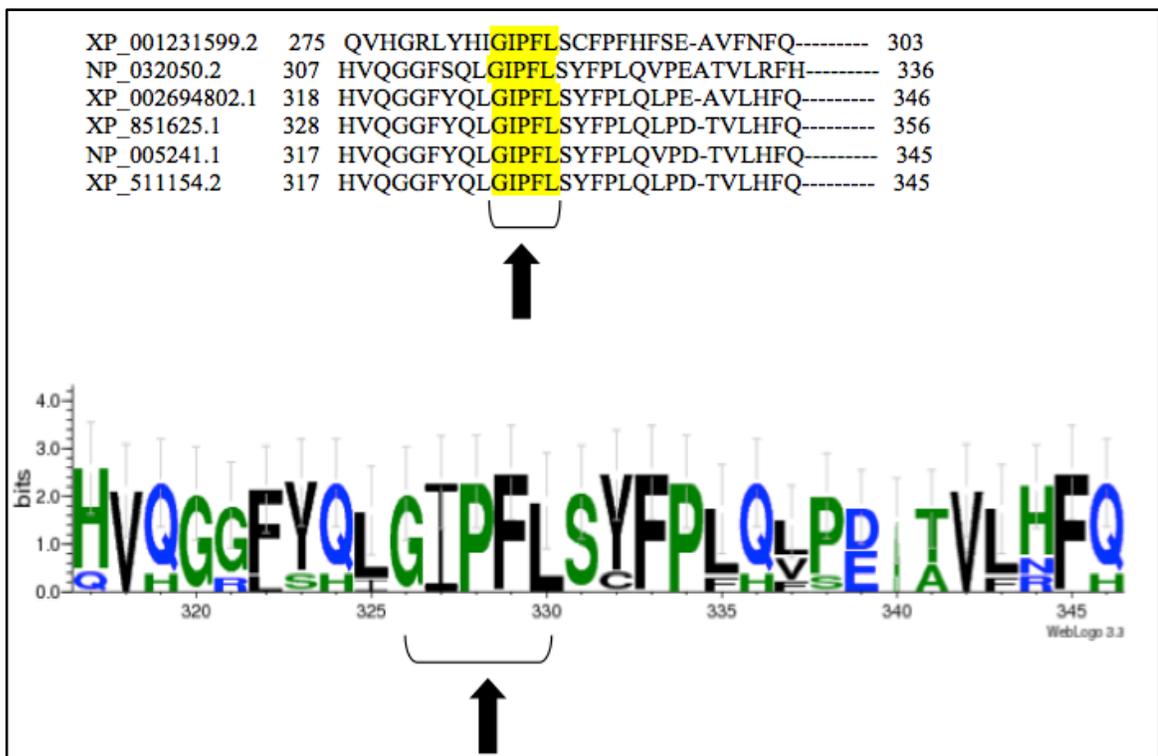
**Figure 3.3: Electropherogram showing the heterozygous deletion of 15 bp in the *FOXL1* gene in PID III-8**

The upper panel represents the reference sequence of the *FOXL1* gene. The lower panel shows a trace sequence in subject III-8 with the deletion underlined.



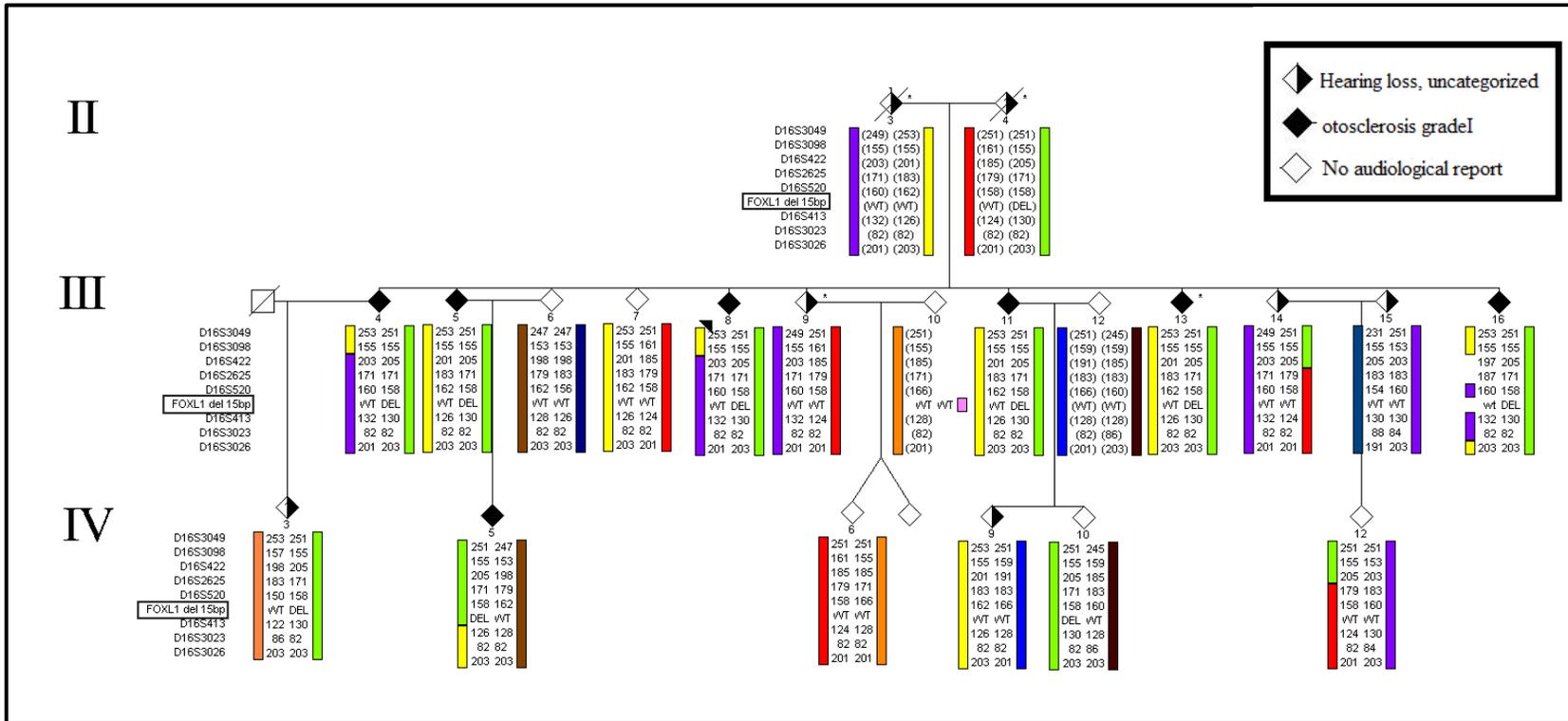
**Figure 3.4: Multiple alignment of FOXL1 p.G327\_L331del across five species**

Upper panel; Clustal W was used to align orthologs from human; NP\_005241.1, Gallus; XP\_001231599.2, M.musculus; NP\_032050.2, B.taurus; XP\_002694802.1, C.lupus; XP\_851625.1, p.troglodytes; XP\_511154.2. Lower panel; Weblogo image shows conservation of the five deleted amino acid across human and five species. Vertical axis represents the conservation scale; horizontal axis represents the amino acids positions.



**Figure 3.5: Partial pedigree of Family 2081 segregates *FOXL1* c. 976-990 het del**

*FOXL1* c. 976-990del resides on the disease haplotype (light green) and segregates with otosclerosis in Family 2081. Roman numerals on the side indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. The gender of family members in this pedigree is hidden for confidentiality. Allele sizes are given in base pairs and alleles in brackets were inferred.



### **3.3.2.3 Mutation screening of functional candidate genes by whole exome sequencing**

Sequencing of 12 genes in the Chr16 candidate region identified the 15 bp deletion in the *FOXL1* gene as the most promising causative mutation. To validate this result, full sequencing of the remaining annotated positional candidate genes (79 genes) in the Chr16 linked region was carried out through whole exome sequencing and targeted analysis of the Chr16 linked region. Whole exome sequencing was carried out for five affected siblings of Family 2081 (PID III-4, III-8, III-11, III-13, III-16) and two unaffected controls.

#### **3.3.2.3.1 Variant analysis**

Variant analysis of this region was carried for Family 2081 under an AD mode of inheritance. We would expect affected exomes (PID III-4, III-8, III-11, III-13, III-16) to carry a heterozygous mutation that is absent from the two unaffected control samples. Results from Genome Quebec were received as an excel file output. Variants identified in genes within the newly linked region at Chr16q23.1-q24.2 were analyzed. The mean coverage of the sequenced variants was 65 %.

Sequencing the coding region of the 79 annotated positional candidate genes revealed a total of 40 variants in the five affected exomes of Family 2081 (Figure 3.6 and Appendix 16B). Variants that were not exclusively found in the five affected exomes (present in 1, 2, 3 or 4 out of 5) were removed. For example, the M200V variant in the *USP10* gene was found in two affected exomes (PID III-13 and III-11) but not in the remaining 3 affected (PID III-8, III-4 and III-16) exomes (Appendix 16B). Removing these variants decreases the number of variants from forty to nine. These nine variants

were exclusively found in the affected exome of Family 2081 and of these nine variants, three in *PLCG2*, *FOXL1* and *IRF8* were previously identified through Sanger sequencing.

Allele frequencies of these nine variants were checked using SNP/1000 genome database. One synonymous variant c.432C>T, p.D144 in the *IRF4* gene has a frequency of 1.5 % (Table 3.3). Five variants have allele frequencies ranging between 5.8 % and 11 %. Two variants (c.749A>G, p. Q250R and c.658 C>T, p. Q220\* in the *PKDIL2* gene) were SNPs with no frequency reported in the SNP/1000 genome database and one variant (*FOXL1* c.976\_990del) was novel. Of the nine variants, four were subjected to further analysis.

The novel *FOXL1* 15 bp deletion was predicted to cause an in-frame loss of five amino acid residues (glycine, isoleucine, proline, phenylalanine and leucine). The missense mutation (c.749A>G, p. Q250R in the *PKDIL2* gene) was predicted to be benign by SIFT (score 0.57) and Polyphen (score 0.004), but it was predicted to create a new donor site. No splicing effect was detected for the silent variant c.432C>T in *PLCG2*. The c.658 C>T, p. Q220\* in the *PKDIL2* gene was predicted to create a premature stop codon at position 220. Of the four variants, three (c.749A>G, p. Q250R, c.658 C>T, p. Q220\* and c.976\_990del) were subjected to segregation analysis.

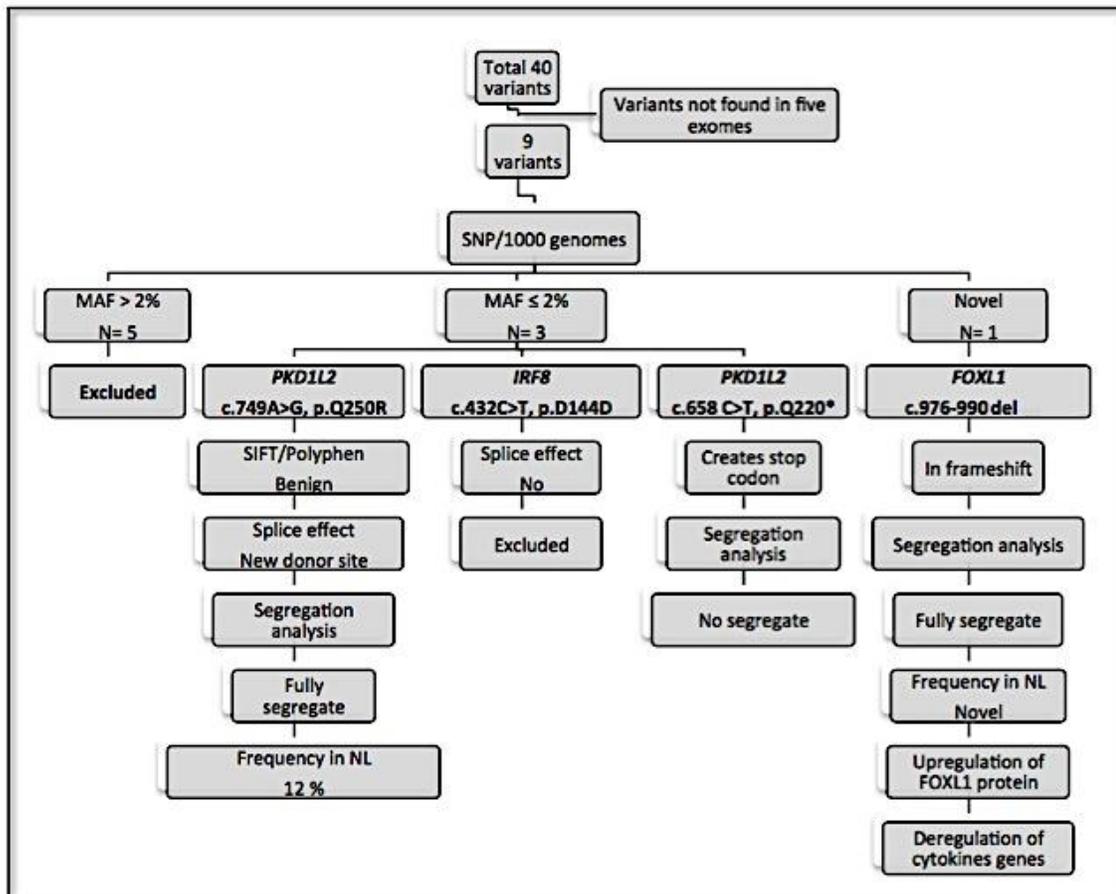
As shown previously, the 15 bp deletion was fully segregating with the lower part of the light green haplotype between markers *D16S520* and *D16S413*. The c.658 C>T, p. Q220\* and c.749A>G, p. Q250R were segregating with the upper part of the light green haplotype between markers *D16S3049* and *D16S3098* (Figure 3.7). In addition, the c.658 C>T, p. Q220\* variant was also detected in an unaffected spouse (PID III-10) and two

subjects (PID III-2 and IV-2) from the left side of the family who did not inherit the light green haplotype. This showed that the c.658 C>T, p. Q220\* did not segregate with otosclerosis in Family 2081 and excluded it from further analysis.

Of the three variants, two (c.749A>G, p. Q250R and c.976\_990hetdel) were segregating with the light green haplotype; c.749A>G, p. Q250R segregates with upper part and c.976\_990hetdel segregates with the lower part. The c.976\_990del was absent from the SNP/1000 genomes database and from 116 NL population controls. The c.749A>G had no frequency data reported in the SNP/1000 genome databases. To determine the frequency of this variant in the NL population, 80 population control samples were screened and 10 samples of the 80 had this variant which gives a frequency estimate of  $(0.125/80)$  12.5% in the NL population. Since the estimated frequency of this variant was high in NL population, this variant was excluded from further analysis. Only the *FOXL1* deletion passed all the filtration criteria. A combination of Sanger and exome sequencing of the annotated positional candidate genes in the Chr16 linked region identified the 15 bp deletion in the *FOXL1* as the causative mutation for otosclerosis in Family 2081.

**Figure 3.6: Flowchart showing the filtration steps of variants identified in sequenced exomes of Family 2081 at the chr16q linked region.**

Of three variants, *FOXL1* c.976\_990del was novel and fully segregates with otosclerosis in Family 2018. Allele frequencies of identified variants were determined by using publicly available SNP database and 1000 genomes. Pathogenicity of missense mutations was evaluated by using SIFT and Polyphen prediction programs. Five splicing prediction programs were used to check possible splice site effects: SSF, HSF, MaxEntScan, NNSPLICE, GeneSplicer.

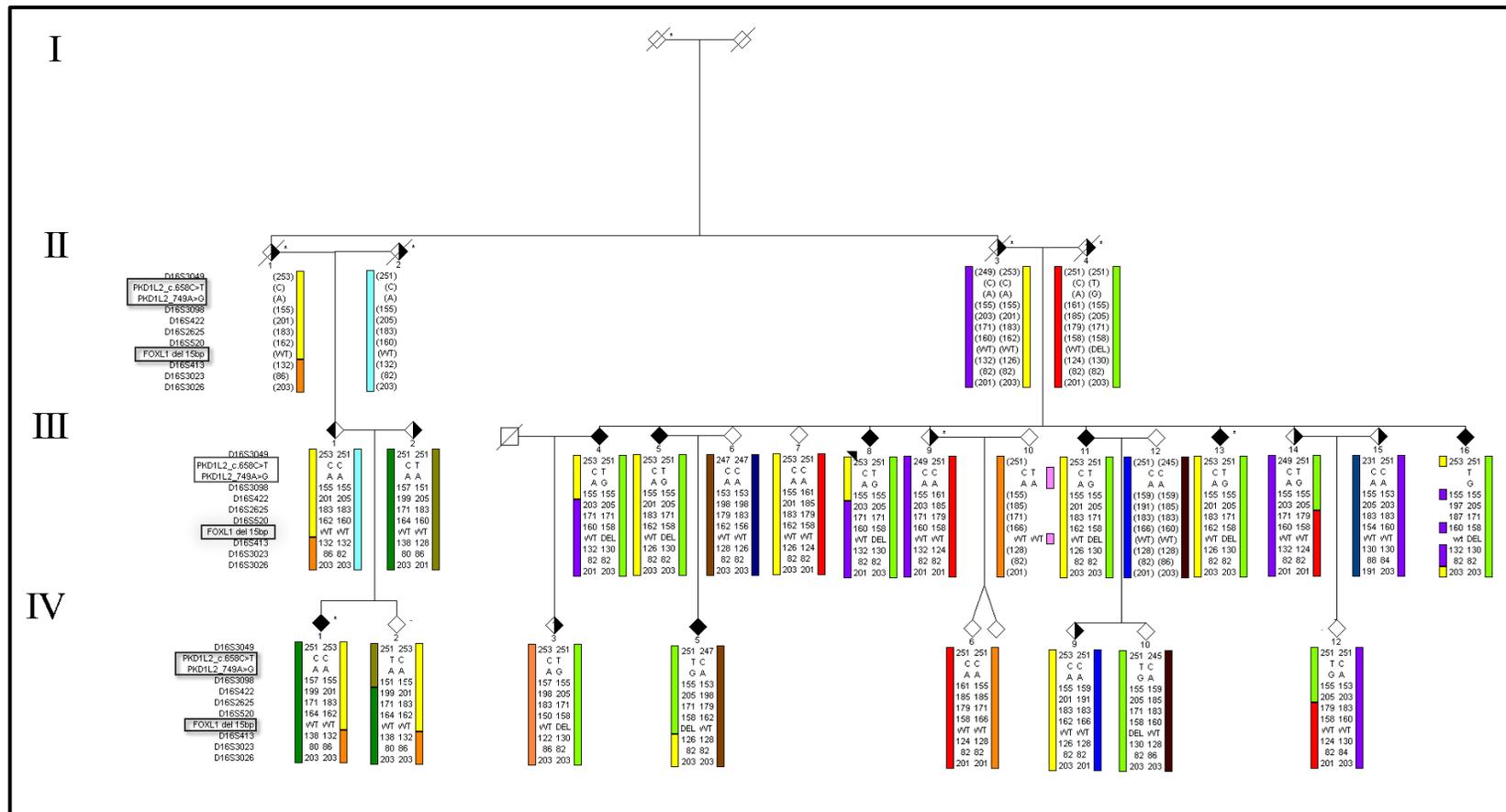


**Table 3.3: Allele frequencies of NGS variants preserved on the disease haplotype**  
 Allele frequency represented by percentage. No frequency means no data was available in SNP/1000 genomes database. Bolded variants were subjected to segregation analysis.

Gene name	Genomic Position	Variants	SNP ID	SNP status	Allele frequency in SNP/1000 genomes		
<i>DYNLRB2</i>	80583497	c.196C>T, p.L66	<i>rs11866734</i>	Silent	T=0.27	224	27%
<b><i>PKD1L2</i></b>	<b>81242107</b>	<b>c.749A/G, p.Q250R</b>	<b><i>rs143480612</i></b>	<b>Missense</b>	<b>No frequency</b>		
<b><i>PKD1L2</i></b>	<b>81242198</b>	<b>c.658 C&gt;T, p.Q220*</b>	<b><i>rs7499011</i></b>	<b>Non-sense</b>	<b>No frequency</b>		
<i>PLCG2</i>	81922813	802C>T, p.R268RW	<i>rs1143687</i>	Missense	T=0.11	184	11%
<i>ATP2C2</i>	84474484	c.1231G>A, p.G411S	<i>rs2303853</i>	Missense	A=0.067	120	6.7%
<i>KLHL36</i>	84691433	c.1020C>T, p. V340	<i>rs72797514</i>	Silent	T=0.058	120	5.8%
<b><i>IRF8</i></b>		<b>432C&gt;T, p.D144</b>	<b><i>rs16939945</i></b>	<b>Synonymous</b>	<b>T=0.015</b>	<b>33</b>	<b>1.5%</b>
<i>MTHFSD</i>	86565826	c.940G>C, p.G152R	<i>rs3751803</i>	Missense	C=0.078	116	7.8%
<b><i>FOXLI</i></b>	<b>86613301</b>	<b>c.976_990hetdel</b>	<b>Novel</b>	<b>In frame deletion</b>	<b>Novel</b>	<b>0</b>	<b>0</b>

**Figure 3.7: Partial pedigree of Family 2081 segregates a15 bp del in *FOXL1* and two variants in the *PKD1L2* gene.**

Figure shows segregation of *FOXL1* c. 976\_990del, *PKD1L2* c.749A>G, p. Q250R and c.658 C>T, p. Q220\* in Family 2081. Roman numerals on the side indicate the generation. The blue solid symbol refers to otosclerosis grade I. The half blue and white symbol refers to hearing loss grade II. The half black and white symbol refers to hearing loss, uncategorized. Allele sizes are given in base pairs) and alleles in brackets were inferred. The gender of family members in this pedigree is hidden for confidentiality.



#### **3.3.2.4 Identify the effect of the *FOXLI* in NL and ON families**

To identify the role of *FOXLI* 15 bp deletion in other otosclerosis families, newly recruited otosclerosis probands from NL (41) and from western ON (40) were screened for the 15 bp deletion in *FOXLI*. As a result, one unrelated family from ON was identified with the 15 bp deletion. In order to determine the possible founder effect between the NL family and ON family, haplotypes across the extended *OTSC4* were created for the ON proband using intragenic SNPs and five microsatellite markers and compared with the NL proband. Comparing haplotypes between the NL and ON probands showed that both probands shared the intragenic SNP genotyped alleles and alleles of two microsatellite markers downstream of the *FOXLI* gene (Table 3.4).

To examine the role of other mutations in *FOXLI* in the recruited NL families with otosclerosis, DNA of 17 NL probands were bidirectional sequenced for the whole exon and the flanking exonic- intronic region of *FOXLI*. Full sequencing of the *FOXLI* gene in the 17 probands resulted in seven variants (Table 3.5). Of these, two (\*1216C>T and \*1902C>T) were SNPs with frequencies of 3.7 % and 6.5 % respectively. Four variants (c.47C>A, \*352C>A, \*925A>T and \*1326G>T) were novel and one (c.814G>A) was a SNP with a frequency of 0.9 %. Of these five, one was a missense and four were located in the 3UTR. The missense mutation (c.814G>A) was predicted to be benign by SIFT and polyphen. No splicing prediction for the c.814G>A variant was discovered. Segregation analysis and functional assessment of the 3UTR variants will help to determine the actual role of these variants in otosclerosis development.

**Table 3.4: Shared ancestral haplotype between affected probands from NL and ON**  
 Shared alleles of 37 SNPs and two markers between NL and ON probands are highlighted green.

SNP ID	Genomic position	Nucleotide change	Amino acid change	Family 2081 proband		Family O1 Proband	
				III-8	155	IV-3	162
<i>D16S3098</i>	8148388	-	-	155	155	162	167
<i>D16S422</i>	82911347	-	-	205	208	205	208
<i>D16S520</i>	86516112	-	-	158	160	154	160
<i>rs374262221</i>	86612362	c.33C>T	p. A11	C	C	C	C
<i>rs142168655</i>	86612378	c.49A>T	p. M17L	A	A	A	A
<i>rs151159060</i>	86612388	c.59T>C	p.L20	T	T	T	T
<i>rs377725521</i>	86612424	c.95T>C	p. F 32S	T	T	T	T
<i>rs370331025</i>	86612461	c.132C>G	p. A44	C	C	C	C
<i>rs145237316</i>	86612467	c.138C>A	p. T46	C	C	C	C
<i>rs200593453</i>	86612469	c.140C>T	p. P47L	C	C	C	C
<i>rs367924573</i>	86612473	c.144G>A	p.Q48	Q	G	G	G
<i>rs199982325</i>	86612539	c.210C>T	p. V70	C	C	C	C
<i>rs113108019</i>	86612644	c.315C>G	p. D105E	C	C	C	C
<i>rs371951781</i>	86612674	c.345G>T	p. E115	G	G	G	G
<i>rs201070716</i>	86612678	349C>A	p. P117T	C	C	C	C
<i>rs374419803</i>	86612702	c.373C>T	p. L125	C	C	C	C
<i>rs141410677</i>	86612708	c.379C>T	p. P127S	C	C	C	C
<i>rs200323545</i>	86612744	c.415C>A	p. R139	C	C	C	C
<i>rs374413539</i>	86612780	c.451G>A	p. A151T	G	G	G	G
<i>rs201954519</i>	86612784	c.455C>T	p. P152T	C	C	C	C
<i>rs112864402</i>	86612872	c.543C>G	p. P181	C	C	C	C
<i>rs202031875</i>	86613074	745C>T	p. R249C	C	C	C	C
<i>rs372247336</i>	86613102	c.773C>T	p. S258F	C	C	C	C
<i>rs62051072</i>	86613143	c.814G>A	p. G272S	G	G	G	G
<i>rs374181632</i>	86613160	c.831A>T	p. S277	A	A	A	A

<i>rs200642220</i>	86613171	c.842T>C	p. L281P	<b>T</b>	T	<b>T</b>	T
<i>rs188450396</i>	86613173	c.844C>A	p. L281I	<b>C</b>	C	<b>C</b>	C
<i>rs181086974</i>	86613179	c.850G>A	p. G284S	<b>G</b>	G	<b>G</b>	G
<i>rs368421556</i>	86613204	c.875G>A	p. R292H	<b>G</b>	G	<b>G</b>	G
<i>rs376672255</i>	86613206	c.877C>T	p. L293	<b>C</b>	C	<b>C</b>	C
<i>rs138954668</i>	86613221	c.892C>T	p. L298	<b>C</b>	C	<b>C</b>	C
<i>rs372235566</i>	86613226	c.897C>T	p. A299	<b>C</b>	C	<b>C</b>	C
<i>rs369054293</i>	86613242	c.913C>T	p. RC	<b>C</b>	C	<b>C</b>	C
<i>rs371495333</i>	86613273	c.944A>G	p. D315G	<b>A</b>	A	<b>A</b>	A
<i>rs376500241</i>	86613274	c.945C>T	p. D315	<b>C</b>	C	<b>C</b>	C
<i>rs185528186</i>	86613283	c.954C>T	p. V318	<b>C</b>	C	<b>C</b>	C
<i>rs143767765</i>	86613292	c.963C>T	p. G321	<b>C</b>	C	<b>C</b>	C
<b>Novel</b>	<b>86613301</b>	<b>c.976_990hetdel</b>	<b>p. G325_L329del</b>	<b>del</b>	<b>wt</b>	<b>Del</b>	<b>Wt</b>
<i>rs200555965</i>	86613324	c.995A>T	p. Y332F	<b>A</b>	A	<b>A</b>	A
<i>rs138392858</i>	86613328	c.999C>T	p. F333	<b>C</b>	C	<b>C</b>	C
<i>D16S413</i>	87893836	-	-	<b>130</b>	132	<b>130</b>	142
<i>D16S3023</i>	88521327	-	-	<b>82</b>	82	<b>82</b>	79

**Table 3.5: Variants identified in *FOXL1* in NL otosclerosis probands**

Seven variants were identified as result of full sequencing of the *FOXL1* gene in 17 NL probands with otosclerosis.

Variants	SNP ID	SNP/ 1000 genomes database frequency	Prediction SIFT/ Polyphen
c.814G>A, p.G272S	<i>rs62051072</i>	0.009 (0.9%)	Benign
c.47C>A, p.P16H	-	Novel	Benign
*352C>A	-	Novel	Not predicted
*925A>T	-	Novel	Not predicted
*1216C>T	<i>rs2288016</i>	0.065 (6.5%)	Not predicted
*1326G>T	-	Novel	Not predicted
*1902C>T	<i>rs4843173</i>	0.037 (3.7%)	Not predicted

### 3.3.3 Functional studies for the 15 bp *FOXL1* deletion

Functional studies to test the effect of the *FOXL1* 15 bp deletion at the RNA and protein level were carried out by a post-doctoral fellow (Ahmed Mostafa) as described in Appendix 17. This work showed that the 15 bp deletion did not affect the RNA expression; however, it did result in up regulation of the *FOXL1* protein expression. A microarray study which was done to investigate the effect of the 15 bp deletion on downstream genes showed deregulation of many genes, especially those involved in inflammation and immunity.

### 3.4 Discussion

I used a positional mapping approach on a large extended NL family (2081) to identify a 15 bp deletion in the *FOXL1* gene that was then shown to segregate with otosclerosis in this family. Family 2081 is a large, multiplex, five generation family which was excluded from linkage to previously mapped loci by our previous study (Chapter 2) . In this Chapter, extension of the *OTSC4* haplotype suggested linkage of Family 2081 to a 9.7 Mb region at Chr16q23.1-q24.2. This 9.7 Mb region contains approximately 79 positional candidate genes. Bidirectional Sanger sequencing of 12 candidate genes, exome sequencing and targeted analysis of the linked region identified a 15 bp deletion (c.976\_990hetdel) in *FOXL1* that segregated with the disease haplotype and was absent from the SNP/1000 genomes project database and 116 NL population control chromosomes. This 15 bp deletion results in an inframe deletion of five highly conserved amino acid residues located at the C-terminus end of the FOXL1 protein.

This Chapter describes the mapping of a new locus at Chr16q23.1-q24.2 in Family 2081. The *FOXL1* gene is located 9 Mb downstream of the lower boundary of the previously linked locus *OTSC4*. Affected members of Family 2081 shared a single haplotype across the new locus at Chr16q23.1-q24.2. This haplotype was shown to be inherited from the parent PIDII-4. A high rate of recombination events was evident across the studied region at Chr16. This region is located toward the telomere of chromosome 16. It has been reported that the rate of meiotic recombination is higher at telomeric regions<sup>40,152,153</sup>.

To examine the role of the 15 bp in *FOXL1* in other families with otosclerosis, otosclerosis probands (representing 81 families) from NL and ON were screened; one of the families from ON had the deletion. Haplotyping of the ON proband using genotyped microsatellite markers spanning the Chr16q23.1-q24.2 region and *FOXL1* intragenic SNPs indicated a common ancestor. Interestingly, sequencing of *FOXL1* by *OTSC4* group didn't reveal any deleterious mutations which suggest that the locus identified in this study is a new one (personal communication).

The Human Forkhead (FKH) box (FOX) gene consists of 43 members. They were named after the discovery of the *Drosophila* FKH gene in 1989. FOX proteins act as transcription factors and are characterized by the sharing of common elements that are formed from approximately 110 amino acid DNA binding sites, called the fork head domain (FHD). The *FOX* gene transcription factor family has been shown to be important in embryogenesis and cell differentiation<sup>154,155</sup>. FOX proteins also play an important role in the regulation of other genes in different species, from yeast to human, and are involved in different biological processes such as tumorigenesis and speech acquisition in humans<sup>156 157 158</sup>.

The *FOXL1* gene (Accession number; NM\_005250) consists of one coding exon, a small 3UTR and a huge 5UTR. The FOXL1 protein, like other members of the FOX family, contains the conserved sequence domain FHD DNA binding site. FOXL1 is highly expressed in otic vesicles in vertebrates and is important in their development by acting as a negative regulator to Sonic hedgehog (SHH) signalling and in embryonic

transcriptional regulation<sup>159</sup>. Previous in vivo studies in mice have shown that *Foxf1* (Forkhead transcription factor f1) and *Foxl1* (Forkhead transcription factor l1) genes are the targets of the Hedgehog (HHL) signalling pathway in stomach development as their expression is controlled by Gli2 and Gli3 transcription factors, which are components of the HHL pathway. Their mutant forms show reduced expression of *BMP4* in the gut. *BMP4* and *BMP2* genes have been previously associated with otosclerosis in human<sup>125,160</sup>. The HHL signalling pathway is important in regulating many cell processes, including cell proliferation, differentiation, angiogenesis, cellular matrix remodelling, and stem cell homeostasis, and deregulation of this pathway leads to the development of different forms of tumours<sup>161</sup>. Another study carried out in mice has shown that mutant Foxl1 causes up regulation of syndecan1 adhesion molecules, which suggests that Foxl1 has a role in decreasing bone formation and increasing bone reabsorption by increasing expression of *IL11* and *RANKL* genes<sup>162,163</sup>.

*FOXL1* and *FOXC2* (Forkhead Transcription Factor C 2) genes are mapped to Chr16q24 and both have conserved GLI binding, which suggests they have a common regulatory mechanism. It has been found (Appendix 17) that the FOXL1 15 bp deletion did not cause any change at the RNA level but resulted in an up regulation of the FOXL1 mutant protein. Up regulation of the mutant protein could be caused by accumulation of the misfolded protein, which would cause a gain of function. Deletion of an amino acid, or an incorrect amino acid could cause destabilization of the protein, resulting in incorrect folding of the protein. It has been reported that an incomplete or incorrectly folded

protein may be degraded resulting in a loss of function or may accumulate and cause a gain of function<sup>165</sup>.

The microarray expression analysis in appendix 17 was done to identify genes deregulated by the *FOXL1* deletion. The expression of 31000 genes in mutant and wild cell lines was compared and numerous deregulated genes were identified. Interestingly, we noticed that there are many cytokines including: *IL1A*<sup>166</sup>, *IL8*, *CCXL10*<sup>167 168 169 170</sup>, *IFNBI*<sup>171</sup>, *IL29*<sup>172</sup> and *IFITI*<sup>173</sup> that have a role in inflammation and osteoclastogenesis were deregulated by the mutant *FOXL1*. A previous study that used prediction algorithms to construct a gene regulatory network of systemic inflammation predicted that *FOXL1* regulates many cytokines including *IL1A*, *IL6*, *IL1B*, *CXCL12*, *CXCL14*, *CCL18*, *IL22* and *IL8* but because of a lack of information about the expression of the *FOXL1* gene, the authors of the study weren't able to verify their findings<sup>174</sup>. Another study of gene expression between normal and otosclerotic human stapedial footplates identified many genes that were differentially expressed in otosclerosis. These genes were predicted to be involved in cytokine signaling and inflammation pathways as well as others<sup>173</sup> and it was hypothesized that the deregulation of the cytokine signaling pathway has a role in abnormal bone remodeling. Our finding agrees with these studies and suggests that cytokine pathways have a role in the development of otosclerosis.

The study outlined in appendix 17 validates the change in expression of *IL29* and *CXCL10* as a result of the *FOXL1* 15 bp deletion in mutant compared to wild type cell lines. *IL29* levels show a 2.52 fold increase in the *FOXL1*<sup>-Mut</sup> as compared to *FOXL1*<sup>-WT</sup>

transcription factors. *IL29*, known also as interferon lambda, has a role in modulating the immune response. *IL29* has a role in inhibition of T-helper 2 cells (Th2) and changes the Th1/Th2 cell response through down-regulation of Interleukin 13 (IL13)<sup>175</sup>. It has been shown that *IL29* up-regulates the level of *IL6*, and *IL10*<sup>176</sup> and that *IL29* induces secretion of *IP-10*, *MIG* and *IL8* cytokines in the peripheral blood cells, which have a role in systemic lupus erythematosus pathogenicity<sup>177</sup>. Recently, it was shown that *IL29* was deregulated in a patient with rheumatoid arthritis which would indicate a role in inflammation, one of the signs of otosclerosis<sup>172</sup>.

We have shown that *CCXL10* expression decreases by 2.85 fold in *FOXLI*<sup>-Mut</sup> compared to *FOXLI*<sup>-WT</sup>. *CCXL10* is an interferon inducing protein and is expressed in osteoblasts at different levels during osteoclast differentiation<sup>178</sup>. A previous study has shown the importance of *CCXL10* and *CCXL13* in osteoblastic activity by examining the expression of chemokine receptors in isolated human osteoblasts. During bone inflammation, mononuclear cells invade bone stroma and secrete chemokines and cytokines that interact with specific receptors on the osteoblast. Analysis of expression of chemokine receptors in osteoblasts showed *CXCR3* and *CXCR5* receptors to be highly expressed in the osteoblast and also identified *CCXL10* and *CCXL13* as ligands for these receptors. Binding of *CCXL10* and *CCXL13* to their receptors on the osteoblast stimulates and increases the b-N-acetylhexosaminidase (Hex) and alkaline phosphatase (ALP) activity. Hex is an enzyme known to have a role in endochondral ossification and bone remodeling. ALP is an important enzyme for bone calcification<sup>179</sup>. Bone specific ALP is reported to be increased in cases of bone loss (osteoporosis)<sup>180</sup>. Changing levels of the

expression of *CCXL10* at any stage of osteoblast differentiation could affect secretion of ALP and Hex enzymes, which would affect bone remodeling.

We have identified a novel founder deletion in the *FOXLI* gene that causes otosclerosis in a family from NL and a family from ON. We believe that the study in appendix 17 identified two possible downstream genes (*IL29* and *CCXL10*) that possibly have a role in otosclerosis development.

### **3.5 Conclusion**

This study has identified a novel founder 15 bp deletion in the *FOXL1* gene which segregated with the otosclerosis phenotype in a NL family in a dominant mode of inheritance. The mutation was also found in an ON family. Functional studies have shown that the c.976\_990hetdel in *FOXL1* upregulates the FOXL1 protein. A microarray study of four *FOXL1* transfected cell lines has identified numerous differentially expressed genes and a number of pathways that potentially have a role in otosclerosis pathogenicity. This study is the first to identify a mutation in a transcription factor causing otosclerosis, and identifies the downstream effects of the mutation. This result provides an opening to understanding the mechanism underlying the development of otosclerosis. The differentially expressed genes identified can be considered as candidates for screening otosclerosis patients with no genetic etiology.

## **4 Chapter 4**

### **Family 2114: Genome-wide linkage analysis and exome sequencing**

## 4.1 Introduction

In Chapter 2, otosclerosis segregating in Family 2114 was excluded from linkage to the six reported *OTSC* loci (*OTSC1*, *OTSC3*, *OTSC4*, *OTSC5*, *OTSC7*, and *OTSC8*) and the two associated candidate genes (*COLIA1* and *NOG*) through haplotype analysis. Interestingly, affected members of Family 2114 shared a region that covered 57 Mb (86 Mb-143 Mb) at Chr 7q21.11-q35, overlapping with both the *OTSC2* locus and *COLIA2* gene. However, absence of DNA from the proband's parents interfered with phasing the haplotypes across these two loci. Subsequent recruitment of maternal relatives helped to resolve the proband's maternal haplotype and subsequently excluded linkage of Family 2114 to both *OTSC2* and *COLIA2* from the mother's side, suggesting that it may be possible to identify a new otosclerosis locus by conducting further studies on this family. In this Chapter, I use both SNP genotyping and exome sequencing to look for a novel otosclerosis locus/gene.

## **4.2 Subjects and methods**

### **4.2.1 Study population and clinical analysis**

To recap, Family 2114 is a multiplex family with five confirmed cases of otosclerosis, apparently segregating as an AD trait. Genomic DNA from eight relatives across two generations was available and a grading system was used to categorize otosclerosis in Family 2114 (refer to Chapter 2, pages 55-57 for details).

### **4.2.2 Genome-wide linkage analysis**

#### **4.2.2.1 Genome wide SNP genotyping**

To identify a new locus, genomic DNA from the five siblings confirmed to have otosclerosis (PIDs III-1, III-2, III-4, III-5 and III-7), one unaffected sibling (PID III-6), a maternal aunt with a history of hearing loss (PID II-6), and a maternal uncle with no history of hearing loss (PID II-5) was sent to the Genome Centre at McGill University for genotyping 610,000 SNPs on a 610K Illumina SNP array. Nine genomic DNA samples from unrelated NL subjects with normal hearing were included in the analysis as population controls. The genotyping data obtained from the Illumina SNP array was analyzed by Dr. Pingzhao Hu at The Center for Applied Genomics, University of Toronto.

#### **4.2.2.2 Genome wide linkage analysis**

##### **4.2.2.2.1 Theoretical Maximum LOD Scores**

Dr. Hu performed simulated analysis in order to obtain the theoretical maximum LOD score for Family 2114 using a two-point comparison under both dominant and

recessive models. The LOD score calculated from this simulation indicated the maximum possible score that the pedigree structure can achieve.

#### **4.2.2.2.2 SNP filtration**

Genotype results from Genome Quebec were sent to The Center for Applied Genomics (TCAG) in Toronto for this analysis. First, the genotype calls in the family were exported from Genome Studio software (V2010.3). Minor allele frequencies were estimated from the genotypes of nine unrelated hearing loss NL samples and the genetic distance of the SNPs was obtained from the TCAG report.

#### **4.2.2.2.3 Multipoint linkage analysis**

Linkage analysis was carried out on the Family 2114 pedigree with Merlin 1.12<sup>181</sup> using the observed genotypes. Multipoint linkage analysis requires estimation of the disease allele frequency and penetrance. The disease allele frequency is usually estimated from the prevalence of the disease in the population. The prevalence of clinical otosclerosis was estimated to be 0.3-0.5 % in the Caucasian population<sup>71</sup>. I assumed that the NL population had the same prevalence and the multipoint linkage analysis with disease allele frequencies of 0.07 and 0.1, corresponding to disease prevalence of 0.5 % and 1 % for the AR model was carried out. For the AD disease, the disease allele frequencies used were 0.0025 and 0.005, corresponding to prevalence of 0.5 % and 1 %. For both dominant and recessive models, multipoint linkage analysis was performed assuming otosclerosis is 100% penetrant (Appendix 21).

### **4.2.3 Genotyping and haplotypes analysis**

Four microsatellite markers (*D17S1829*, *D17S1807*, *D17S1839* and *D17S1822*) that span the newly linked region at Chr17q25.1-q25.3 and five markers (*D16S422*, *D16S2625*, *D16S520*, *D16S413* and *D16S3023*) at Chr16q24 were used for genotyping members of Family 2114. Genomic DNA from members of Family 2081 was also genotyped with markers at Chr17q25.1-q25.3 and used as controls for NGS analysis. PCR amplification and genotyping were carried out as mentioned in Chapter 2 (pages 57-58).

### **4.2.4 Sequencing of positional candidate genes within the linked regions**

#### **4.2.4.1 Candidate genes sequencing by Sanger sequencing**

The UCSC Genome Browser homepage<sup>141</sup> and the March 2006 assembly (NCBI build 36.1) were used to define the genomic interval of each critical disease region and to identify candidate genes. From these, a subset of genes was chosen for sequencing based on knowledge of gene function and gene expression. For example, otosclerosis is a bone disease and it is hypothesized that otosclerosis develops as a result of abnormal bone remodeling in the otic capsule. For that reason, genes involved in bone metabolism and bone remodelling, and genes that are expressed in bone cells, connective tissue or the immune system, would make excellent functional candidate genes to be prioritized for sequencing. Primer design, amplification and sequencing were carried out as given in Chapter 2 (pages 60-61). Primers for all functional candidate genes are listed in Appendix 22.

The proband of Family 2114 was primarily sequenced for the following 11 physical candidate genes: *BIRC5*, *CTDNEP1*, *TIMP2*, *RPL38*, *SLC9A3R1*, *USH1G*,

*OTOP2*, *KCTD2*, *RNF157*, *ITGB4*, and *FOXJ1*. Details regarding the function of these genes are provided in Appendix 23.

#### **4.2.4.2 Whole exome sequencing of Family 2114**

Whole exome sequencing was carried out on four affected siblings (PID III-1, III-2, III-4 and III-7). Two control exomes from the NL population with no hearing loss (based on late audiograms at ages 55 and 60) were also sequenced. Whole exome sequencing was carried out as described in Chapter 3 (pages 158-159). Primers used to validate variants of interest are listed in Appendix 23. The data received from Genome Quebec provided us with the nucleotide change, genomic position, and variant MAF from SNP/1000 genome database. The full pipeline report for NGS is attached in Appendix 14.

#### **4.2.4.3 Variant analysis**

An allele frequency of 2 % was used as the cut-off limit for heterozygous variant analysis under the dominant mode of inheritance. Variants with population frequencies of 2 % or less were subjected to further analysis. For detailed information about the variant analysis, please refer to Chapter 2 (page 51-53). As I could not rule out recessive inheritance, I needed to estimate the frequency of heterozygous variants in the general population. In this case, the frequency of the heterozygous carrier would be higher than that of the homozygous affected patients. To determine the frequency of the heterozygous carrier under the recessive mode of inheritance, Hardy-Weinberg (H-W) equilibrium<sup>182</sup> was carried out, considering the disease allele (q) and the normal allele (p) and using the following equation  $p^2 + 2pq + q^2 = 1$ .

*Hardy – Weinberg Equilibrium*

*Frequency of otosclerosis = 0.5%; therefore,  $q^2 = 0.005$*

$$q = \sqrt{q^2} = \sqrt{0.005} = 0.07$$

$$p + q = 1$$

$$p = 1 - q$$

$$p = 1 - (0.07)$$

$$p = 0.93$$

$$p^2 = (0.93)^2 = 0.865$$

*2 x p x q to get 2pq*

$$2pq = 2 (0.93) (0.07) = 0.13$$

*Therefore:*

*The frequency of the homozygous normal genotype:  $p^2 = 0.865$*

*The frequency of the homozygous mutant genotype:  $q^2 = 0.005$*

*The frequency of heterozygous carrier:  $2pq = 0.13$*

About 0.13 (13 %) of the population will be carriers of the otosclerosis allele and approximately 0.5 % will be homozygous affected. Heterozygous variants with frequencies of 13 % or less in the SNP/1000 genomes database were included in the analysis under recessive inheritance. The allele frequency of homozygous variants was expanded to 2 %, similar to dominant heterozygous variant allele frequency. Homozygous variants with frequencies of 2 % or less were subjected to further analysis. Homozygous variants under the recessive mode of inheritance and heterozygous variants under the dominant mode of inheritance were analyzed as shown in Figure 2.1. The allele frequency of identified variants was checked first using the SNP/1000 genome database.

Variants with an allele frequency of 2 % or less were subjected to further analysis.

Variants with an allele frequency of more than 2% were excluded from further analysis.

## **4.3 Results**

### **4.3.1 Pedigree structure and clinical analysis**

Family 2114 is a multigenerational family with five cases of otosclerosis.

Otosclerosis in Family 2114 appears to have a dominant mode of inheritance. Although the proband's mother was presumed to be the disease carrier (because she has a history of hearing loss), a recessive mode of inheritance could not be ruled out because of the lack of clinical information on the paternal side of the proband's family. For further information on pedigree structure and clinical analysis of Family 2114, please refer to Chapter 2 (pages 63-66).

### **4.3.2 Genome wide linkage analysis**

Although it appears from the pedigree structure that otosclerosis is inherited as an AD trait, genotype analysis was carried out under both the dominant and recessive modes of inheritance for reasons listed above. However, linkage simulation yielded a max LOD of 2.5 under a recessive model and the results from the multipoint analysis yielded the theoretical maximum at Chr17q25.1-q25 under a recessive model. For the dominant model, the maximum theoretical LOD score was 1.73 and linkage analysis yielded a maximum observed LOD score of 1.4 at each of five distinct genomic loci (Table 4.1), including the same Chr17 locus identified under a recessive model.

It was important to first see if any of the five loci (7q, 10p, 10q, 16q, and 17q) overlapped with previously mapped or associated otosclerosis genes/loci. For both 10p and 10q, there were no mapped *OTSC* loci or associated genes. Chr 7q, 16q and 17q, however, harboured otosclerosis loci. Comparing the linked regions giving positive LOD scores with critical otosclerosis boundaries (Table 4.1) revealed that the Chr7 critical region overlapped with both *OTSC2* and *COLIA2* (as identified in Chapter 2) and the Chr16q locus overlapped with the new otosclerosis locus identified in Chapter 3.

**Table 4.1 Regions with an observed LOD score >1 under a dominant model**  
All regions have a maximum LOD score of 1.4. Boundaries SNPs are reported from TCAG.

Chr	Regions with LOD>1		Known <i>OTSC</i> loci	
	Position (Mb)	Size (Mb)	Name	Position (Mb)
<b>7q</b>	73.8-143.03	69.7	<i>OTSC2</i> <i>COLIA2</i>	85.05-152.05 94.02 -94.06
<b>10p</b>	0.19-3.1	2.91	-	-
<b>10q</b>	50.04-54.12	4.08	-	-
<b>16q</b>	82.6-89.02	6.3	<i>OTSC4</i>	67.6-77.3
			New locus (Family 2081)	78.1-87.8
<b>17q</b>	71.9-77.7	5.79	<i>NOG</i>	54.6-54.6

### **4.3.3 Genetic analysis of Family 2114 under a recessive mode of inheritance**

Because Chr17q gave the highest LOD score under a recessive model, genetic analysis for this region was carried out first, which involved examining the segregation of haplotypes across this region and identifying paternal and maternal chromosomes. This was followed by Sanger sequencing of functional candidate genes and targeted analysis of the exome variants.

#### **4.3.3.1 Haplotype segregation across Chr17 linked region**

Under the assumption of a recessive mode of inheritance, we would expect that affected members of Family 2114 would inherit two disease haplotypes, one from each carrier parent. If parents were related (in the case of consanguinity) the two disease haplotypes would be identical by descent and all affected offspring would harbour a homozygous mutation. On the other hand, if parents were not related, the two disease haplotypes would likely not be identical and all affected offspring would harbour two different mutations in the same otosclerosis gene (compound heterozygotes). Also, if parents were not related and the recessive allele was common in the NL population, the two disease haplotypes might not be identical and the affected offspring would harbour two identical mutations on different disease haplotypes (IBS).

Haplotypes were constructed using 23 SNPs, including the reported boundary SNPs *rs9909316* and *rs2587507* (Table 4.2), and three microsatellite markers. Although parental haplotypes were inferred, four distinct parental haplotypes (brown, green, red

and yellow) (Figure 4.1) were identified. Constructing haplotypes across Chr17q revealed that four out of the five affected members (PID III-1, III-2, III-5 and III-7) inherited the same parental haplotypes (yellow and brown). The other affected sibling (PID III-4) inherited the yellow haplotype, and a portion of the brown haplotype as a recombinant haplotype from his father. The only unaffected sibling (PID III-6) inherited the red and green haplotypes. Construction of the haplotype contributed by the proband's uncle (PID II-5) helped to infer the mother's yellow haplotype and led to the conclusion that the brown haplotype was inherited from the proband's father.

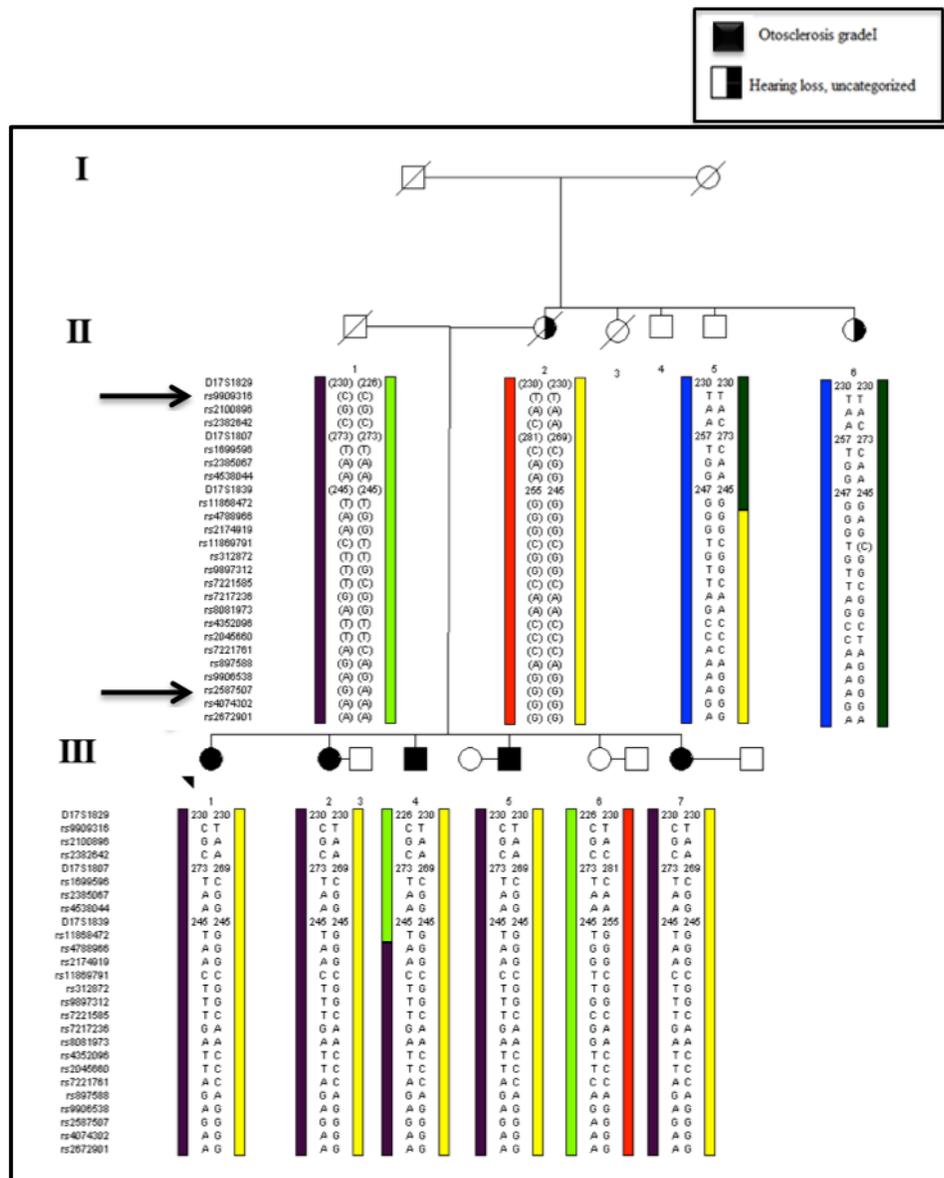
**Table 4.2: Boundaries, size and gene numbers of the five linked regions**

Start and end SNPs were taken from TCAG statistical report.

Family 2114 linked regions					
Chr	Start SNP	End SNP	Boundaries position	Size Mb	Gene numbers
7q	<i>rs2158797</i>	<i>rs2103193</i>	73.8-143.03	69.7	284
10p	<i>rs10903451</i>	<i>rs3829916</i>	0.19-3.1	2.91	12
10q	<i>rs7922169</i>	<i>rs1194683</i>	50.04-54.12	4.08	41
16q	<i>rs11646011</i>	<i>rs13329773</i>	82.6-89.02	6.3	64
17q	<i>rs9909316</i>	<i>rs2587507</i>	71.9-77.7	5.79	141

**Figure 4.1 Partial pedigree of Family 2114 showing segregation of Chr17q25.1-q25.3 haplotypes.**

The yellow haplotype and a portion of the brown haplotype are segregating in Family 2114. Roman numerals on the side indicate the generation level. The black solid symbol refers to otosclerosis grade I. The half black and white symbol refers to hearing loss, uncategorized. Arrows point to boundaries SNPs. ( ) refers to inferred alleles. Circle = female, square = male.



#### 4.3.3.2 Positional candidate genes sequencing by Sanger sequencing

From linkage analysis, the boundaries of Chr17 critical region were reported between *rs9909316* and *rs2587507*. These boundaries cover a region spanning 5.79Mb with 141 annotated genes according to Genome Browser (March 2006 assembly, NCBI build 36.1). In total, 11 positional candidate genes were chosen and sequenced using genomic DNA from the proband (PID III-1).

To determine the possible outcome of the sequencing analysis, haplotypes were constructed using provided genotyped SNPs. Haplotype analysis showed that affected members of Family 2114 inherited the same haplotypes (yellow and a portion of the brown haplotypes) from their parents. If the yellow and brown parental haplotypes are assumed to be the disease haplotypes, we would expect that the proband's parents are heterozygous for two disease alleles and the proband inherits two different mutations in the putative disease gene (compound heterozygote). Conversely, both parents may be heterozygous for the same mutation, and in that case the proband inherits 2 copies of the same disease mutation. The full length of the linked region at Chr17 was analyzed under a recessive mode of inheritance primarily and then under a dominant mode. To cover both modes of inheritance and to detect any common variants, the region was not decreased in size using the recombination in subject III-4.

Direct sequencing of the 11 genes yielded high quality sequences (Figure 4.2). In total, eight variants were identified in the six positional candidate genes (*TIMP2*, *BIRC5*, *FOXJ1*, *ITGB4*, *OTOP2* and *RNF157*) in the proband PID III-1 (Table 4.3). No variants

were identified in the remaining five genes (*CTDNEP1*, *SLC9A3R1*, *USH1G*, *RPL38* and *KCTD2*).

Of these eight variants, no homozygous variants were identified. Three heterozygous variants (missense, intronic and synonymous) were identified in the *ITGB4* gene and the remaining five variants were identified in five different genes. Under recessive inheritance, it is conceivable that two of the three variants identified in *ITGB4* could be pathogenic and cause otosclerosis. The first step in analyzing these three variants was to determine their allele frequencies. We estimated the frequency of otosclerosis heterozygous carriers to be 13 % under a recessive model using the Hary-Weinberg Equilibrium principle. Taking this number into consideration, allele frequencies of these variants were checked, and as result, these three variants (c.4311C>G, c.4844-25C>G and c.5126T>C) were excluded from further analyses because their frequencies were higher than 13 % (48%, 79% and 25%, respectively).

Because Chr17q11.23-q34 gave a positive LOD score under the dominant model, allele frequencies of the remaining five heterozygous variants were determined, four of which had frequencies of greater than 10 %. Interestingly, we identified a silent variant in the *FOXJ1* gene with a frequency of 4 %. Because *FOXJ1* is a member of the FOX family of transcription factors, and otosclerosis in Family 2081 was caused by a deletion in *FOXL1*, I thought it was important to follow-up on this particular variant. This variant is not predicted to affect splicing and was further analyzed using NGS sequencing data.

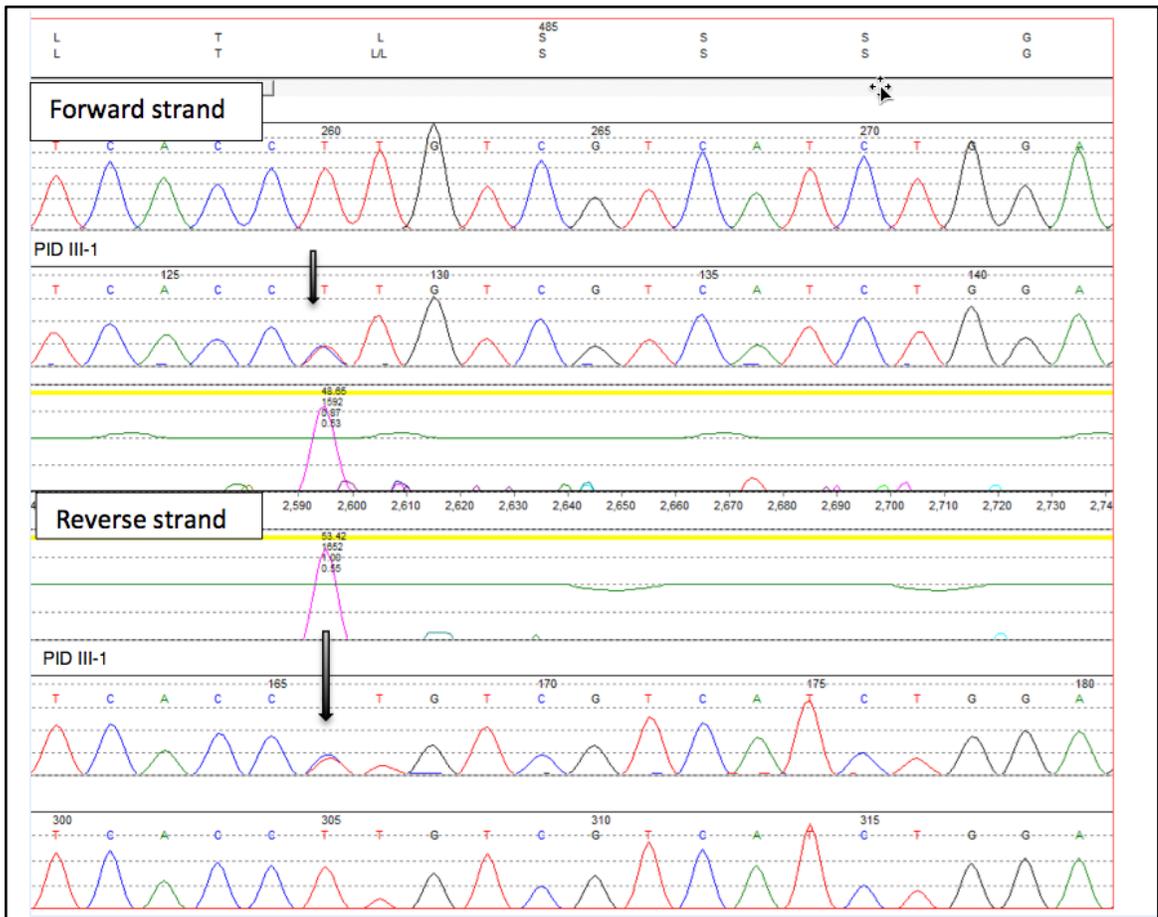
The selection of functional candidate genes in the critical region on Chr17 did not yield a putative otosclerosis gene under recessive or dominant models. Of the 141 annotated genes, we only looked at 11/141. Therefore, to provide a complete examination of the 5.79 Mb critical region, whole exome sequencing and target analysis of the entire critical region was carried out.

**Table 4.3: Sequencing variants identified in six out of 11 candidate genes examined in the Chr17 linked region.**  
Frequencies of the six identified variants are higher than 2%.

<i>Gene Name</i>	NM number	Exon number	Variants	SNP status	SNP ID	Allele frequency	Sample size	Frequency %	Splice prediction
<i>TIMP2</i>	NM_003255.4	Ex3	c.303G>A, p. S101	Silent	rs2277698	A=0.100	184	10%	-
<i>BIRC5</i>	NM_001012271.1	Ex 5	c.454G>A,p. E152K	Missense	rs2071214	A=0.776	1094	77%	-
<i>FOXJ1</i>	NM_001454	Ex3	c.726G>A,p. T242	Silent	rs894542	A=0.043	1482	4%	No splicing prediction
<i>ITGB4</i>	NM_001005619	Ex32	c.4311C>G,p. p.P1437	Silent	rs8669	G=0.483	120	48%	-
<i>ITGB4</i>	NM_001005619	Int34	c.4844-25C>G	Intronic	rs2249613	G=0.792	120	79%	-
<i>ITGB4</i>	NM_001005619	Ex38	c.5126T>C, p.L1779P	Missense	rs871443	C=0.258	120	25%	-
<i>OTOP2</i>	NM_178160.2	EX6	c.1393G>T,p. G465W	Missense	rs6501741	T=0.667	120	66%	-
<i>RNF157</i>	NM_052916	EX14	c.1450T>C,p. L484	Silent	rs881502	C=0.342	120	34%	-

**Figure 4.2: Electropherogram showing an example of a high quality sequence**

The c.1450T>C, p. L484LL variant identified in the *RNF157* gene. Upper panel represents reference sequence of the forward strand of the *RNF157* gene and a trace sequence with the missense mutation c.1450 T>C in PID III-1. Lower panel represents reference sequence of the reverse strand of the *RNF157* gene and a trace sequence with the missense mutation c.1450 T>C in PID III-1.



#### **4.3.3.3 Exome sequencing and targeted analysis of Chr17 under a recessive mode**

In order to complete sequencing of the 141 positional candidate genes at Chr17 linked region, four exomes of Family 2114 (PID III-1, III-2, III-4 and III-7), along with two unaffected controls were sent for whole exome sequencing. Results from Genome Quebec were received in an excel file output. Variants identified in genes within the new linked region at Chr17q25.1-q25.3 were fully analyzed under a recessive mode of inheritance. To help reduce the number of variants under consideration, extra negative controls were chosen.

##### **4.3.3.3.1 Selection of negative controls for filtering NGS data at Chr17**

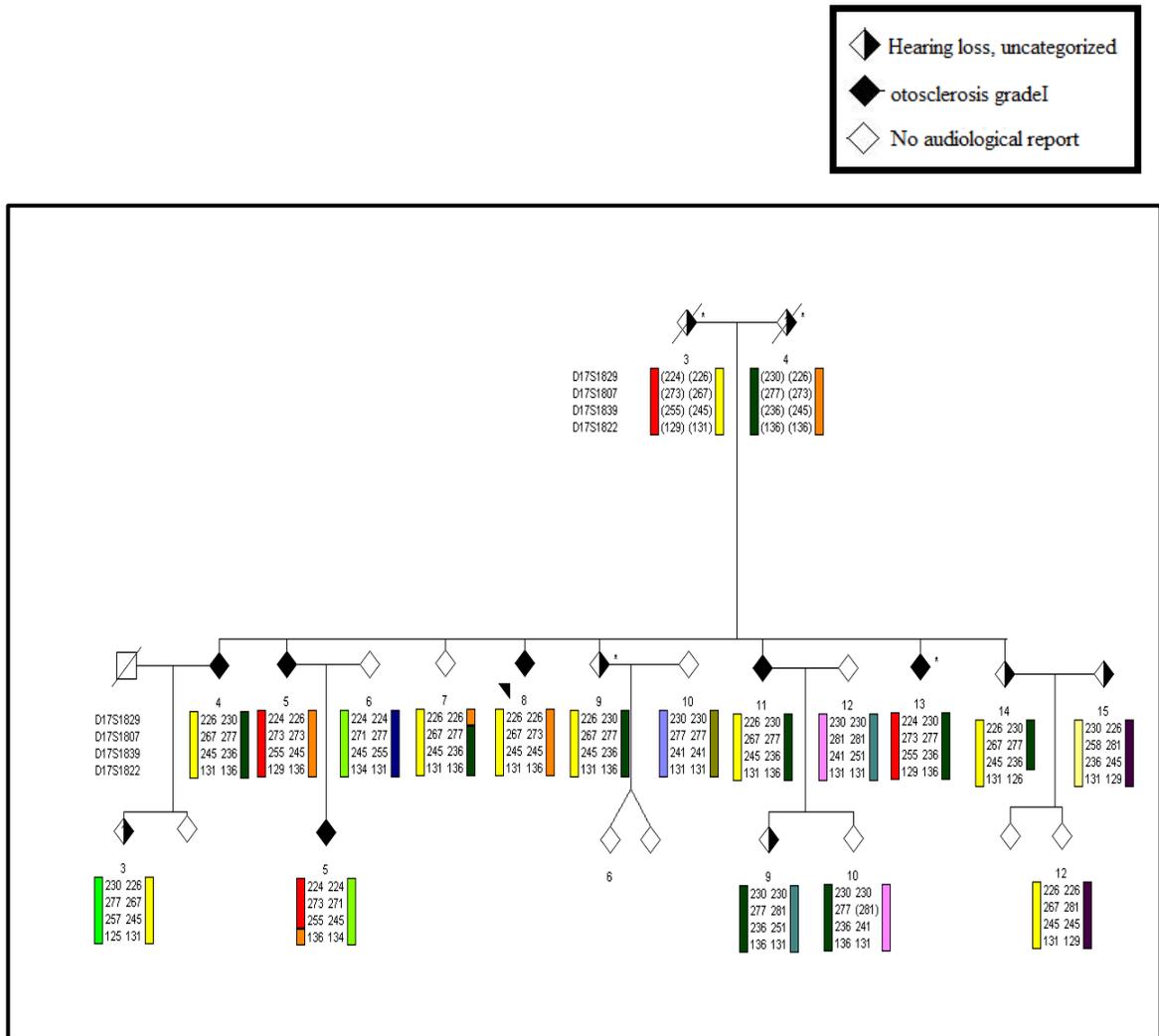
In Chapter 3, DNA samples from Family 2081 were also sent for whole exome sequencing and targeted analysis of Chr16 was carried out. Because family 2081 was solved (as described in Chapter 3) by identifying a 15 bp deletion in the *FOXL1* at Chr16 and both families originated from NL, NGS from family 2081 could be used as a negative control.

As Family 2081 was mapped to a new locus on Chr16, it was important to rule out linkage of Family 2081 to this new 17q locus. To do this, genotyping and haplotype analysis using four fully informative microsatellite markers that span the critical region (*D17S1829*, *D17S1807*, *D17S1839* and *D17S1822*) was carried out. Although the proband's maternal and paternal haplotypes were inferred, four distinct parental haplotypes (red, yellow, dark green and orange) were created (Figure 4.3). Using the rule of least recombination, two recombinations were detected between *D17S1829* and

*D17S1807* (PID III-7) and between *D17S1839* and *D17S1822* (PID IV-5). Affected siblings of Family 2081 did not share a disease haplotype across this locus. For example, the proband (PID III-8) inherited the yellow and the orange parental haplotypes. On the other hand, the proband's affected sibling (PID III-13) inherited the dark green and the red parental haplotypes. Because two other affected siblings inherited two different parental haplotypes (and thus did not share any region on Chr17q25.1-q25.3), I concluded that otosclerosis in Family 2081 was excluded from linkage to Chr17q25.1-q25.3. Family 2081 could then be confidently used as population-based controls for analyzing NGS at Chr17.

**Figure 4.3: Partial pedigree of Family 2081 showing segregation of Chr17 haplotypes in Family 2081**

No shared haplotype was detected between affected members of Family 2081 across Chr17q. Roman numerals on the side indicate the generation level. Blue solid symbol refers to otosclerosis grade I. Half blue and white symbol refers to hearing loss grade II. Half black and white symbol refers to hearing loss, uncategorized. Arrows point to boundaries SNPs. ( ) refers to inferred alleles.



#### 4.3.3.3.2 Variants analysis of Family 2114 under a recessive mode of inheritance

Sequencing of the coding regions of 141 positional candidate genes revealed a total of 49 heterozygous variants in Family 2114 affected exomes (Figure 4.4). Seven variants were removed because they were also identified in affected members of Family 2081. Furthermore, variants that were not inherited by all affected siblings of Family 2114 were also removed. For example, the Q1133K variant in *RNF213* was detected in the proband and two of her affected siblings (PID III-2 and III-1) but not in her other affected siblings (PID III-4 and III-7). Removing all of these variants reduced the number to 17, which were distributed across 11 genes. However, none of these variants were found to be homozygous in the affected siblings. Of these 17 variants, two were identified in *TEN1* and six were identified in *EVPL*. The remaining nine were heterozygous variants in nine separate genes. Interestingly, the silent variant identified in *FOXJ1* was found in three of the proband's four affected siblings, and was also detected in two affected members of Family 2081 (Appendix 22).

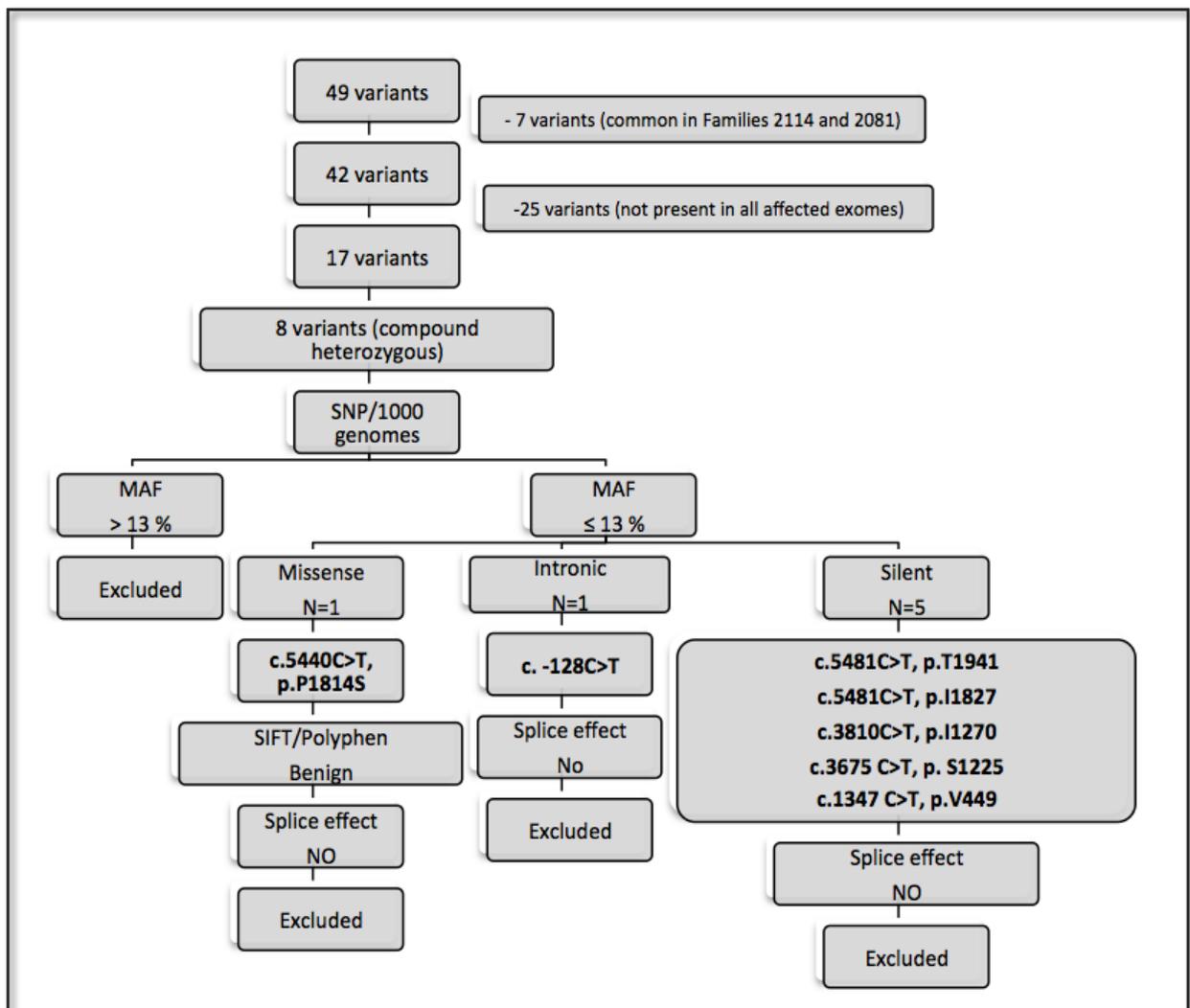
To examine the possibility of compound heterozygous mutations in the *TEN1* and *EVPL* genes, these eight variants were analyzed under a recessive mode of inheritance. The frequencies of the eight variants were checked using SNP/1000 genome database and as a result, one was found to be a known SNP (*rs10852767*) with a frequency of 15.6 % (Figure 4.5A and Table 4.4). The seven remaining variants (five silent and one missense variant in *EVPL*; one intronic variant in *TEN1*) were subjected to further analysis. Splicing and pathogenicity effects of the silent, intronic and missense variants were checked and as a result the missense variant was predicted to be benign. No splicing

effect was predicted for the five silent, intronic and benign missense variants. Therefore, these seven variants were excluded from further analysis. In conclusion, the possibility of compound heterozygosity of the variants identified in *TEN1* and *EVPL* could be excluded.

Complete analysis of the exome of 141 genes in the Ch17 critical region under a recessive mode of inheritance did not identify any homozygous or pathogenic compound heterozygous mutations. Therefore, I decided to analyze this region and the other four regions identified by linkage under a dominant mode of inheritance.

**Figure 4.4 : Flowchart showing filtration steps of NGS data at Chr17q under recessive mode of inheritance**

No pathogenic heterozygous variants were detected. Allele frequencies of identified variants were determined by using a publicly available SNP database and 1000 genomes. Pathogenicity of missense mutations was evaluated using SIFT and Polyphen prediction programs. Five splicing prediction programs were used to check splice site prediction: SSF, HSF, MaxEntScan, NNSPLICE, and GeneSplicer.



#### 4.3.3.4 Genetic analysis of Family 2114 under dominant inheritance

Pedigree simulation gave a maximum theoretical LOD score of 1.7 under a dominant mode of inheritance. The experimental result yielded a positive LOD score of 1.4 for each of five distinct loci (Table 4.1). Under dominant inheritance, we would expect affected siblings to carry one copy of the disease allele (heterozygous mutation) that is inherited from the mother (PID II-2) and unaffected siblings to not carry the disease allele.

The five distinct loci were distributed across four chromosomes (Table 4.3). All together, this adds up to 542 positional candidate genes. Across the five regions, exome sequencing revealed a total of 301 variants after removing variants from two unaffected NL controls (Figure 4.5). Family 2081 could be used as a negative control for both Chr7 and Chr17 because it had been previously excluded from linkage to these regions. The benefit of using exome data from Family 2081 was that it resulted in a reduction of variants from 301 to 264 (in other words, 37 variants were also identified in exomes of Family 2081, the family with a causative *FOXL1* deletion). Moreover, 153 variants could be removed because they were not found in all affected siblings, further reducing the number to 111 variants (Tables 4.4, 4.5, 4.6, and 4.7). Of these, 68 variants had a frequency of greater than 2 %, reducing the number to 43 (24 silent; 15 missense; one intronic; one nonsense; one deletion; and one 3 UTR).

The 14 missense mutations were predicted to be benign by SIFT and Polyphen, and were not predicted to cause any splicing effects by HSF, MaxEntScan, GeneSplicer and known constitutive signals prediction programs. The c.1706 C>T, p. P569L variant in *TNRC6C* (Figure 4.6) was predicted to be deleterious by SIFT and damaging by Polyphen. As well, conservation of the 569 proline residue in the *TNRC6C* gene is highly conserved (Figure 4.7). The majority (23/24) of the silent variants were predicted to have no effect on splicing. However, the c.1143C>T, p. R381R variant in *CSTF2T* was predicted to cause the loss of a splice acceptor site. In addition, the deletion (c.1578 het delC) in *C17orf80* was predicted to create a premature stop codon 10 codons upstream of the normal stop codon. The intronic c.69+11C>G variant in *CDK14* (Figure 4.8) was predicted to create a new splice donor site. As well, the nonsense variant c.49C>T in *KNAP7* (Figure 4.9) was predicted to interrupt the reading frame and create a premature stop codon. Although mutations in the non-coding region of genes are difficult to interpret, it is possible that the 3' UTR, \*709 variant in *SEPT* has a deleterious effect on gene expression. In summary, variants in six genes (*TNRC6C*, *C17orf80*, *SEPT*, *CSTF2T*, *CDK14* and *KNAP7*) were potentially pathogenic and therefore subjected to segregation analysis.

Segregation analysis was done on a chromosome by chromosome basis. For example, in order to test segregation of c.69+11C>G in *CDK14* and c.49C>T in *KNAP7*, Chr7 haplotypes were constructed using 13 microsatellite markers and 19 genotyped SNPs, including the boundary SNPs. Although the haplotypes of the proband's parents were inferred, four distinct parental haplotypes (red, light green, dark green and orange) could

be created (Figure 4.10). Haplotype construction of the proband's aunt (PID II-6) and uncle (PID II-5) helped to infer the proband's maternal haplotypes (light green and a portion of the red haplotypes in Figure 4.10). We observed that affected siblings shared a haplotype (dark green) that covered the region between *rs2158797* and *rs2103193*, apparently inherited from their father (PID II-1). Two variants in *CDK14* and *KPNA7* were tested for segregation and as result both variants segregated with otosclerosis in Family 2114 and resided on the dark green paternal haplotype.

For Chr17, haplotypes were constructed as mentioned on page 200 and three variants were tested. Segregation analysis of the heterozygous frameshift deletion in *C17orf80* showed that it was not present in the affected sibling PID III-5 (Figure 4.11). Segregation analysis of c.1706 C>T, p. P569L in *TNRC6C* gene and \*709 in the *SEPT9* gene revealed that these variants segregate with otosclerosis in family 2114.

For Chr10q, haplotypes were created using nine SNPs. Four distinct parental haplotypes (red, yellow, dark green and blue) were created (Figure 4.12). Haplotype construction of the proband's aunt (PID II-6) and uncle (PID II-5) helped to reconstruct the maternal contribution (PID II-2). Affected members all shared the dark green haplotype that appears to be inherited from their mother. One variant was tested for segregation, and absence of segregation of p. R381R (*CSTF2T*) was noted, as one of the affected siblings (PID III-5) was wild type.

Of the six tested variants, only four segregated with otosclerosis. The allele frequency of these variants in NL population controls was tested and as a result, c.1706 C>T in

*TNRC6C*, C.69+11C>G in *CDK14*, and c.49C>T in *KPNA7* were not detected in 125, 64 and 67 NL population controls, respectively. The c. \*709 C>T in the *SEPT9* gene has a frequency of 10.3 % in the NL controls.

To summarize, exome sequencing and genetic analysis, under a recessive model and looking at the Chr17 region, did not yield any likely pathogenic mutations (either in the homozygous or compound heterozygous state). On the other hand, targeted analysis under a dominant model revealed three relatively rare, functionally significant variants in *KPNA7*, *CDK14* and *TNRC6C* at Chr7, and Chr17 co-segregated with otosclerosis in Family 2114. Interestingly, none of the deregulated cytokines (*IL29* and *CCXL10*) identified in Chapter 3 were located in the linked regions and, therefore, were not screened.

Of the three variants, the *KPNA7* (IPOA8) is the most promising candidate. This gene belongs to the importin alpha family, and although there was not much information about the function of this gene, it is known to act as a nuclear protein transporter and has been predicted to be involved in cytokine signalling in the immune system. The non-sense mutation in *KPNA7* is predicted to be highly deleterious causing a prematurely truncated protein due to a frameshift within the coding region of the gene.

The second most promising candidate is the splicing variant in the *CDK14* gene. Cyclin-Dependent Kinase 14 (*CDK14*) is a member of the CDC-related protein kinase family. A previous study reported abundant expression of CDK14 protein in the bone through examining osteoclast cell line<sup>183</sup> and it has a role in the regulation of cell cycle

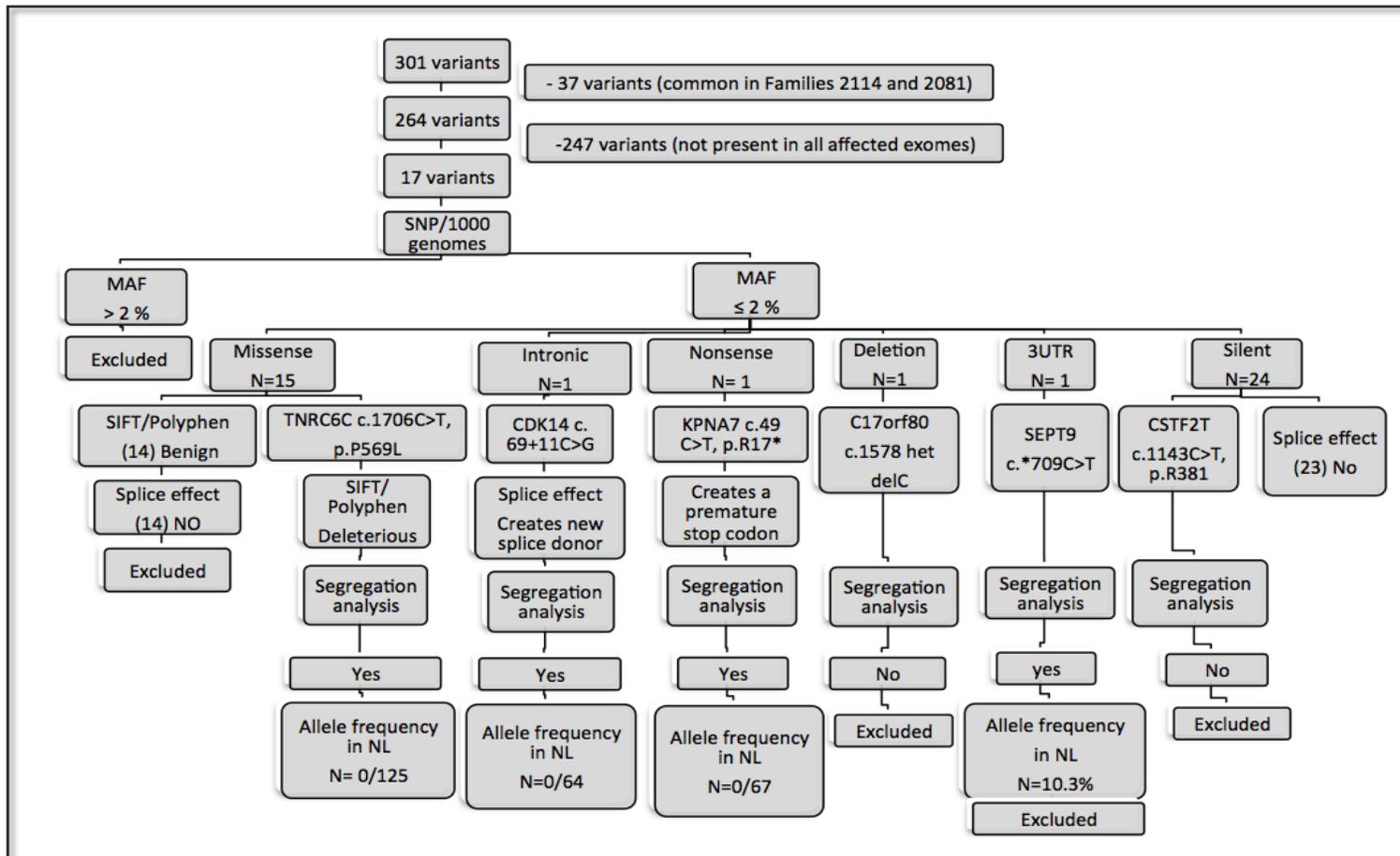
progression and proliferation through interaction with *CCDN3*<sup>184</sup>. It is also involved in activation of the Wnt signalling pathway through phosphorylation of *LRP6*<sup>185</sup> and it has been reported by previous studies that osteoclastogenesis is regulated by the Wnt pathway<sup>186</sup>. *CDK14* has been reported to interact with genes involved in cytokinesis<sup>187</sup> and in negative regulation of osteogenesis<sup>188</sup>.

The third variant identified was in the *TNRC6C* gene. *TNRC6C* is one component of the TNRC6 complex, a protein family involved in silencing mRNA<sup>189</sup>. Specifically, the TNRC6 complex is involved in miRNA induced silencing complexes (miRISCs). The *TNRC6C* interacts with other genes (*AGO2*, *DICER*, and *DGCR8*) involved in miRNA homeostasis. Previous studies have shown that gene silencing of Dicer or Argo2 by small interfering RNAs prevents the expression of osteoclast transcription factors (*PU.1*, *MITF*, *c-Fos*, and *NFATc1*) and therefore, decreases osteoclastogenesis and decreases bone resorption<sup>190</sup>. The missense mutation identified in *TNRC6C* was located in a highly conserved region and was predicted to be deleterious.

In summary, the three genes that harboured three variants were considered good candidate genes for otosclerosis development. Of these three, the nonsense variant in *KNAP7* appeared to be the most deleterious as it created a stop codon at c.49 position in exon 1, which probably would lead to nonsense mediated decay for the truncated gene product. Further functional and segregation analysis is required to validate this assumption and confirm the pathogenic gene/mutation causing otosclerosis in family 2114.

**Figure 4.5: Fowchart showing the steps of NGS analysis at five regions under the dominant model**

Three heterozygous variants in *KPNA7*, *TNRC6C* and *CDK14* passed all filtration steps and segregate with otosclerosis in Family 2114. Allele frequencies of identified variants were determined by using a publicly available SNP database and 1000 genomes. Pathogenicity of missense mutations was evaluated by using SIFT and Polyphen prediction programs. Five splicing prediction programs were used to check splice site prediction: SSF, HSF, MaxEntScan, NNSPLICE, and GeneSplicer.



**Table 4.4: Sequencing variants detected in exomes of four affected siblings in Family 2114 across Chr17 under both recessive and dominate mode of inheritance.**

Under the recessive model, variants with MAF of higher than 13 % were excluded. Bolded variants represent compound heterozygotes. Under the dominant model, variants with MAF of >2 % were excluded. SIFT and Polyphen were used for prediction of the missense mutations. Human Splicing Finder (HSF), MaxEntScan, GeneSplicer and known constitutive signals prediction programs were used for splice site prediction. (-) indicates no prediction was carried out.

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<i>CI7orf80</i>	c.1578 het delC	-	Frame shift	No frequency			-	-	-
<i>CD300E</i>	c.472G>A, p.G158R	<i>rs1878061</i>	Missense	No frequency			No splice prediction	Benign	Benign
<i>NUP85</i>	c.186T>C, p.D62	<i>rs34126097</i>	Silent	C=0.59	128	5.9%	No splice prediction	-	-
<i>ACOX1</i>	c.301G>A, p.G101S	<i>rs3744032</i>	Missense	No frequency			No splice prediction	Benign	Benign
<i>TEN1</i>	<b>c. -128C&gt;T</b>	<i>rs3760132</i>	<b>Intronic</b>	<b>T=0.074</b>	<b>162</b>	<b>7.4%</b>	<b>No splice prediction</b>	-	-
<i>TEN1</i>	<b>c.5823C&gt;T, p.H65</b>	<i>rs10852767</i>	<b>Silent</b>	<b>T=0.156</b>	<b>340</b>	<b>15.6%</b>	<b>No splice prediction</b>	-	-
<i>EVPL</i>	<b>c.5481C&gt;T, p.T1941</b>	<i>rs17886642</i>	<b>Silent</b>	No frequency			<b>No splice prediction</b>	-	-

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<i>EVPL</i>	c.548C>T, p.I1827	<i>rs7342882</i>	Silent	No frequency			No splice prediction	-	-
<i>EVPL</i>	c.5440C>T, p.P1814S	<i>rs7342883</i>	Missense	No frequency			No splice prediction	Benign	Benign
<i>EVPL</i>	c.3810C>T, p.I1270	<i>rs7225323</i>	Silent	No frequency			No splice prediction	-	-
<i>EVPL</i>	c.3675 C>T, p. S1225	<i>rs75019874</i>	Silent	No frequency			No splice prediction	-	-
<i>EVPL</i>	c.1347 C>T, p.V449	.	Silent	No frequency			No splice prediction	-	-
<i>RNF157</i>	c.186T>C, p.F62	<i>rs61760884</i>	Silent	No frequency			No splice prediction	-	-
<i>QRICH2</i>	c.1808 A>G, p.H603R	<i>rs143901735</i>	Missense	No frequency			No splice prediction	Benign	Benign
<i>PRPSAPI</i>	c.1044A>G, p.Q245	<i>rs10459</i>	Silent	G=0.148	188	14.8%	-	-	-
<i>SEPT9</i>	c.*709C>T	<i>rs146084702</i>	3UTR	0.0045	3974	0.45%	-	-	-
<i>TNRC6C</i>	c.1706 C>T, p.P569L	<i>rs200406635</i>	Missense	0.0028	8356	0.28%	-	Deleterious	Probably damaging

\* No frequency means no data were available in SNP/1000 genome databa

**Table 4.5: Sequencing variants detected in exomes of four affected siblings in Family 2114 across Chr7: 73,862,412-142,344,569.**

Under the dominant model, variants with MAF of > 2 % were removed from further analysis. Bolded variants have MAF of  $\leq 2$  % or no frequency estimated and (-) indicates no prediction was carried out.

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<i>ZNF804B</i>	c.1614G>A, p.T538	<i>rs148827448</i>	Silent	A=0.15	8596	15%	-	-	-
<i>CDK14</i>	<b>c.69+11C&gt;G</b>	-	<b>Splice site</b>	<b>No frequency</b>			<b>creates a new donor site</b>	-	-
<i>PEX1</i>	c.2088A>G, p.Iqq3M	<i>rs35996821</i>	Missense	G=0.038	78	3.8%	-	-	-
<i>RBM48</i>	c.588G>A, p.P196	<i>rs79531673</i>	Slent	A=0.023	51	3%		-	-
<i>TRRAP</i>	<b>c.10731C&gt;T, p.I3606</b>	<i>rs56290902</i>	<b>Silent</b>	<b>T=0.009</b>	<b>20</b>	<b>.09%</b>	<b>No splice prediction</b>	-	-
<i>KPNA7</i>	<b>c.49C&gt;T, p.R17*</b>	-	<b>Stop_gain (non sense mutation)</b>	<b>Novel</b>			-	-	-
<i>ZNF3</i>	c.55+56G>A	<i>rs34312198</i>	Intronic	A=0.05	109	5%	-	-	-
<i>ZNF3</i>	c.-231G>A	<i>rs113683483</i>	Intronic	A=0.023	51	3%	-	-	-
<i>MCM7</i>	c.1083C>T, p.I361	<i>rs12267</i>	Silent	T=0.446	166	44%	-	-	-
<i>C7orf59</i>	c.202+141G>T	<i>rs3736590</i>	Intronic	T=0.625	176	6.25%	-	-	-
<i>MUC12</i>	c.512G>A, p.R171Q	<i>rs61746936</i>	Missense	A=0.0669	3182	6.6%	-	-	-
<i>MUC17</i>	c.4669A>G, p.T1557A	<i>rs74209688</i>	Missense	G=0.24	537	24%	-	-	-
<i>LRWD1</i>	c.1778C>T, p.P593L	<i>rs148981077</i>	Missense	T=0.446	13004	44%	-	-	-
<i>FAM185A</i>	c.56A>T, p.Q19L	<i>rs147397718</i>	Missense	T=0.088	193	8.8%	-	-	-

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<i>LRRIC17</i>	c.560 G>A, p.G187A	<i>rs1057066</i>	Missense	A=0.173	376	17%	-	-	-
<i>NAPEPLD</i>	c.454G>T, p.S152A	<i>rs12540583</i>	Missense	T=0.933	120	93%	-	-	-
<i>SPAMI</i>	c.1352 A>T, p.D451V	<i>rs75168651</i>	Missense	T=.008	17	0.8%	-	Benign	Benign
<i>SLC13A4</i>	c.1351C>T, p.P451S	<i>rs36004833</i>	Missense	No frequency			-	Benign	Benign
<i>MGAM</i>	c.3822T>C, p.D1151	<i>rs2960758</i>	Silent	No frequency			No splice prediction	-	-
<i>FAM115A</i>	c.2220G>A, p. G740	<i>rs3735380</i>	Silent	No frequency			No splice prediction	-	-
<i>GIMAP7</i>	c.247C>T, p.R83C	<i>rs3735080</i>	Missense	T=0.169	2184	16.9%	-	-	-

**Table 4.6: Sequencing variants detected in exomes of four affected siblings in Family 2114 across Chr10: 49715462-69.77254365.**

Under the dominant mode of inheritance, variants with MAF of > 2 % were excluded. Bolded variants have MAF of  $\leq$  2 % or no frequency estimated. SIFT and Polyphen were used for prediction of missense mutations. Human Splicing Finder (HSF), MaxEntScan, GeneSplicer and known constitutive signals prediction programs were used for splice site prediction. (-) indicates no prediction was carried out.

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<i>GTPBP4</i>	c. -8 C>T	<i>rs12772979</i>	Intronic	T=0.1769	8600	17.6%	-	-	-
<i>C10orf71</i>	c.1998C>G, p.H666Q	<i>rs10857469</i>	Missense	G=0.097	211	9.7%	-	-	-
<i>CSTF2T</i>	<b>c.1263 C&gt;T, p.R421</b>	-	<b>Silent</b>	<b>No frequency</b>	-	-	<b>No splice prediction</b>	-	-
<i>CSTF2T</i>	<b>c.1143C&gt;T, p.R381</b>	-	<b>Silent</b>	<b>No frequency</b>	-	-	<b>Predicted to loss acceptor site</b>	-	-
<i>CSTF2T</i>	<b>c.312T&gt;C, p. L104</b>	<i>rs2292828</i>	<b>Silent</b>	<b>C=0.075</b>	<b>168</b>	<b>0.75%</b>	<b>No splice prediction</b>	-	-
<i>TUBB8</i>	-48T>A	<i>rs6560829</i>	Intronic	G=0.97	224	9.7%	-	-	-
<i>ZMYND11</i>	c.1356A>G, p.G537	<i>rs1017361</i>	Silent	A=.334	224	3.3%	-	-	-
<i>DIP2C</i>	c.1866C>T, p.G622	<i>rs10904083</i>	Silent	A=.339	224	3.35%	-	-	-
<i>DIP2C</i>	<b>c.1248 G&gt;A, p.P416</b>	<i>rs6560837</i>	<b>Silent</b>	<b>No frequency</b>	-	-	<b>No splice prediction</b>	-	-
<i>DIP2C</i>	<b>c.970C&gt;T, p.L324</b>	<i>rs4881274</i>	<b>Silent</b>	<b>No frequency</b>	-	-	<b>No splice prediction</b>	-	-
<i>GTPBP4</i>	<b>c.750G&gt;A, p.A250</b>	<i>rs2306409</i>	<b>Silent</b>	<b>No frequency</b>	-	-	<b>No splice prediction</b>	-	-

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<b>WDR37</b>	c.726+22T>C	rs10794716	Intronic	C=0.89	118	89%	-	-	-
<i>ADARB2</i>	c.1876G>A, p.A626T	rs2271275	Missense	A=0.665	226	66.5%	-	-	-
<i>PFKP</i>	1281A>G, p.A419	rs1052309	Silent	G=0.625	112	62%	-	-	-
<i>PFKP</i>	c.1359C>T p.F453	rs4881086	Silent	T=0.252	226	25%	-	-	-
<i>PITRM1</i>	c.3021A>G, p.R909S	rs11414	Missense	G=0.606	226	60%	-	-	-
<i>FRMPD2</i>	c.2735G>A p.G912E	rs200503523	Missense	No frequency			-	Benign	Benign
<i>FRMPD2</i>	c.2720A>G, p.L534E	.	Missense	No frequency			-	Benign	Benign
<i>FRMPD2</i>	1600A>G, p.L534E	rs1864345	Missense	No frequency			-	Benign	Benign
<i>ARHGAP22</i>	c.591A>G, p.197	rs3853761	Silent	No frequency			No splice prediction	-	-
<i>ARHGAP22</i>	c.330C>T, p. A110	rs4080665	Silent	T=0.850	226	85%	-	-	-
<i>WDFY4</i>	c.4582T>C, p.458P	rs2170132	Missense	C=0.274	226	27%	-	-	-
<i>WDFY4</i>	c.5447G>A, p.R1816E	rs7097397	Missense	A=0.375	120	37%	-	-	-
<i>WDFY4</i>	c.6396A>G, p.Q2132	rs3747874	Silent	No frequency			No splice prediction	-	-
<i>WDFY4</i>	c.7580G>A, p.S2527N	rs2663046	Missense	No frequency			-	Benign	Benign
<i>WDFY4</i>	c.9216C>T, p.C3072	rs2271565	Silent	T=0.334	728	33.4%	-	-	-
<i>WDFY4</i>	c.9353C>T, p.p3118L	rs2292584	Missense	T=0.339	870	33.9%	-	-	-

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<i>VSTM4</i>	c.873A>G, p.E291	<i>rs1913525</i>	Silent	G=0.89	116	89%	-	-	-
<i>VSTM4</i>	c.203T>C, p.F68S	<i>rs13088</i>	Missense	C=0.397	224	39%	-	-	-
<i>C10orf71</i>	c.1014T>C, p.L338	<i>rs7093235</i>	Silent	C=0.89	112	89%	-	-	-
<i>C10orf71</i>	2093T>C, p.F698S	<i>rs7921186</i>	Missense	C=0.89	116	89%	-	-	-
<i>PGBD3</i>	c.1477C>T, p.L961	<i>rs4253073</i>	Silent	T=0.902	224	90%	-	-	-
<i>PGBD3</i>	c.1145G>A, p.R850L	<i>rs4253072</i>	Missense	A=0.983	116	98%	-	-	-
<i>ERCC6</i>	c.135C>G, p.L45	<i>rs2228524</i>	Silent	G=0.617	180	61%	-	-	-
<i>SLC18A3</i>	c.1559C>A, p.A520E	<i>rs8187730</i>	Missense	A=0.899	120	89%	-	-	-
<i>CHAT</i>	c.1027G>A, p.M343V	<i>rs4838544</i>	Missense	A=0.989	118	98%	-	-	-
<i>CHAT</i>	c.1287T>C, p.H429	<i>rs8178992</i>	Silent	C=0.885	226	88%	-	-	-
<i>C10orf53</i>	c.216C>T, p.F72	<i>rs1133837</i>	Silent	T=0.483	120	48%	-	-	-
<i>OGDHL</i>	<b>c.564C&gt;T, p.L131</b>	<i>rs1258184</i>	<b>Silent</b>	<b>No frequency</b>			<b>No splice effect</b>		
<i>AGAP7</i>	<b>c.904T&gt;A, p.Y302N</b>	<i>rs76363375</i>	<b>Missense</b>	<b>No frequency</b>			-	<b>Benign</b>	<b>Benign</b>
<i>NCOA4</i>	c.22T>G, p.F8V	<i>rs10761581</i>	Missense	0.492	120	48%	-	-	-
<i>TIMM23</i>	<b>c.25A&gt;G, p.N8D</b>	<i>rs4935253</i>	<b>Missense</b>	<b>No frequency</b>			-	<b>Benign</b>	<b>Benign</b>

**Table 4.7: Sequencing variants detected in the exomes of four affected siblings in Family 2114 across Chr16: 73,862,412-142,344,56.**

Variants with a frequency of higher than 2 % were excluded and did not predicted for pathogenicity effect and replace by (-). Bolded variants are analyzed under a recessive mode of inheritance. SIFT and Polyphen were used for prediction of the missense mutations. Human Splicing Finder (HSF), MaxEntScan, GeneSplicer and known constitutive signals prediction programs were used for splice site prediction.

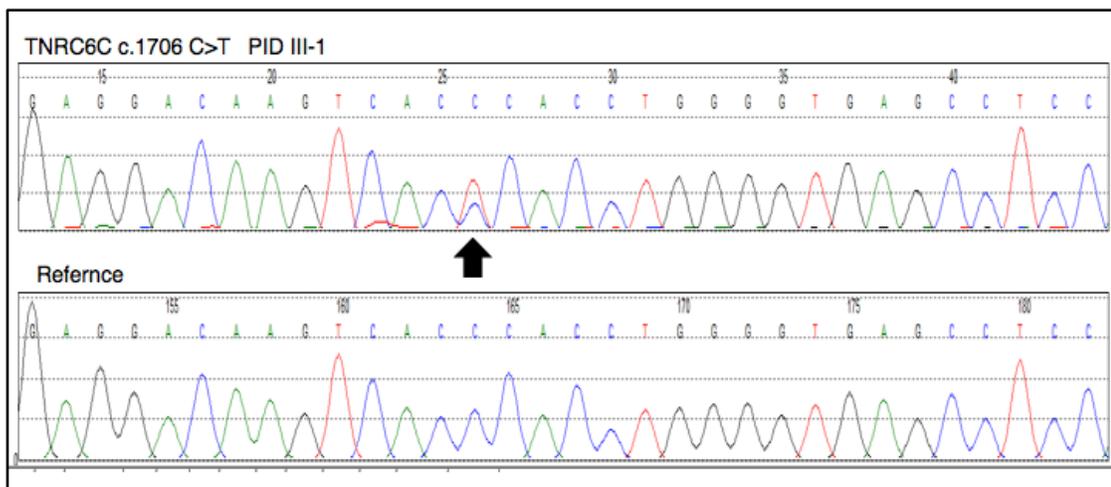
Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<b>BCMO1</b>	c.678A>G, p.P226	rs7202895	Silent	G=0.124	270	12.4%	-	-	-
<b>MBTPSI</b>	c.1407C>T, p.L469	rs12933523	Silent	T=0.135	74	13.5%	-	-	-
<b>DNAAF1</b>	c.1898T>C, p.L633S	rs2288020	Missense	C=0.303	660	30%	-	-	-
<b>DNAAF1</b>	c.1976T>A, p.L659P	rs2288022	Missense	C=0.302	658	30%	-	-	-
<b>DNAAF1</b>	c.2024G>C, p.S675T	rs2288023	Missense	C=0.303	659	30%	-	-	-
<b>TAF1C</b>	c.1590G>T, p.R768	rs1804500	Silent	T=0.196	46	19.6%	-	-	-
<b>TAF1C</b>	c.1047G>A, p.V349	rs2230130	Silent	A=0.396	164	39.6%	-	-	-
<b>TAF1C</b>	c.727T>A, p.L243M	rs2230129	Missense	A=0.393	1488	39.6%	-	-	-
<b>TAF1C</b>	c.545A>G, p.A332	rs2230127	Silent	G=0.403	62	40%	-	-	-
<b>ADAD2</b>	c.348C>T, p.A116	rs79224535	Silent	0.029	64	2.9%	-	-	-
<b>ADAD2</b>	c.1212C>T, p.A404	rs2303239	Silent	0.296	645	29%	-	-	-

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction	SIFT	Polyphen
<i>ADAD2</i>	<b>c.1688C&gt;T, p.T563TI</b>	.	Missense	-	-	-	-	Benign	Benign
<i>KCNG4</i>	c.973G>A, p.G325R	<i>rs7196482</i>	Missense	G=0.795	78	79%	-	-	-
<i>KIAA1609</i>	<b>c.107C&gt;T, p.N357</b>	<i>rs62640938</i>	Silent	No frequency			No splice effect		
<i>CA5A</i>	c.807A>T, p.A269	<i>rs72816311</i>	Silent	0.826	4550	82%	-	-	-
<i>CA5A</i>	<b>c.453C&gt;T, p.P151</b>	<i>rs7186698</i>	Silent	No frequency			No splice effect		
<i>SPG7</i>	c.1507A>G, p.T503A	<i>rs2292954</i>	Missense	G=0.119	259	11%	-	-	-
<i>SPG7</i>	c.2063G>A, p.R688Q	<i>rs12960</i>	Missense	A=0.118	257	11%	-	-	-
<i>SPG7</i>	<b>c.2292C&gt;T, p.I764</b>	<i>rs61747711</i>	Silent	<b>T=0.016</b>	<b>35</b>	<b>0.16%</b>	No splice effect	-	-
<i>RPL13</i>	c.141C>T, A47	<i>rs174035</i>	Silent	T=0.154	335	15%	-	-	-
<i>MCIR</i>	c.178G>T, p.V60L	<i>rs1805005</i>	Missense	T=0.051	111	6%	-	-	-
<i>CTU2</i>	C.994G>A, p.V332I	<i>rs4782321</i>	Missense	A=0.178	118	17%	-	-	-
<i>CTU2</i>	<b>c.1280G&gt;A, p.R427Q</b>	<i>rs11549835</i>	Missense	<b>A=0.066</b>	<b>144</b>	<b>0.6%</b>	-	Benign	Benign
<i>PIEZO1</i>	c.2159G>C, p.R720P	<i>rs35265318</i>	Missense	A=0.178	118	17%	-	-	-
<i>PIEZO1</i>	c.2130C>G, p.D710E	<i>rs35353288</i>	Silent	T=0.074	68	7.4%	-	-	-
<i>PIEZO1</i>	c.2031C>A, p.V677	<i>rs34908386</i>	Silent	A=0.138	58	13%	-	-	-

Gene name	Variants	SNP ID	SNP status	Allele frequency	Sample size	Frequency %	Splice prediction
<i>PIEZO1</i>	c.1540C>T, p.L514	rs78905828	Silent	No frequency			No splice effect
<i>PIEZO1</i>	c.1365G>A, T455	rs61742623	Silent	No frequency			No splice effect

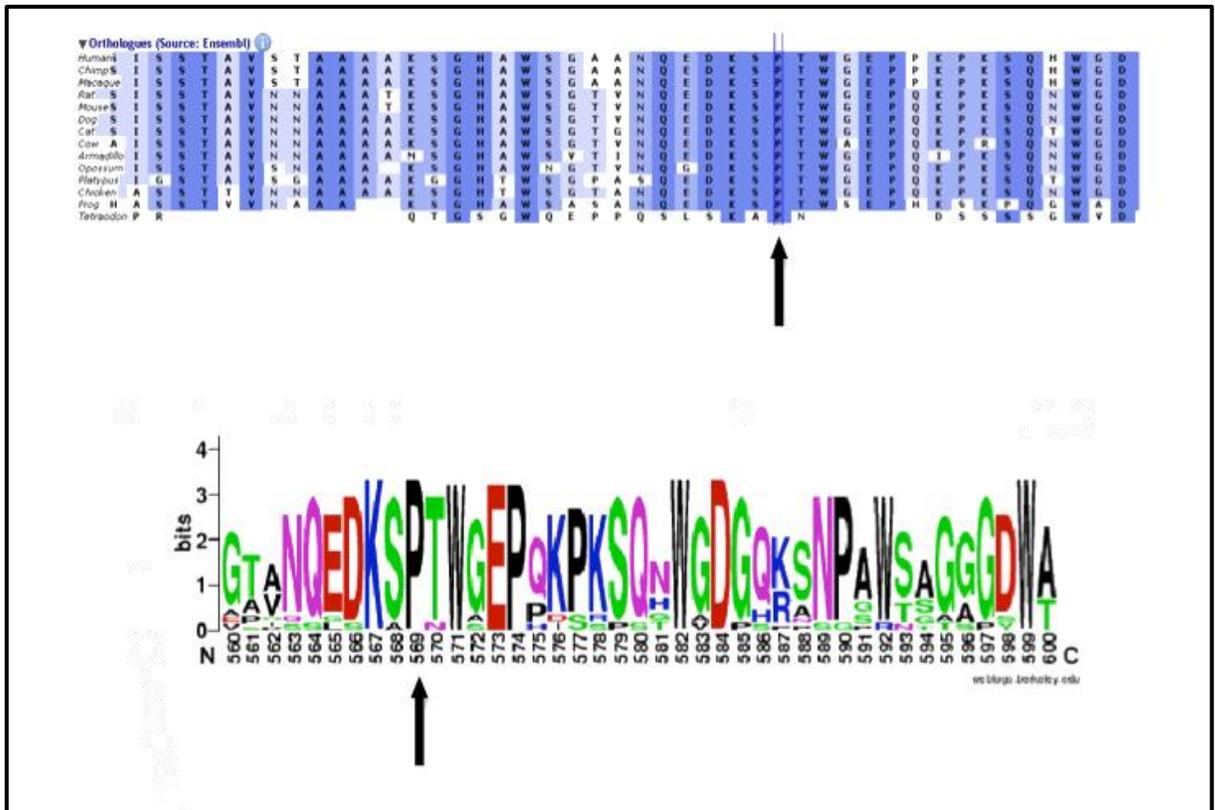
**Figure 4.6: Electropherogram shows heterozygous missense mutation c.1706 C> T, p. P569L in the *TNRC6C* gene. Arrow pointed to the missense mutation.**

Upper panel shows a trace sequence with the missense mutation c.1706 C>T in PID III-1, indicated by an arrow. Lower panel represents the reference sequence of the *TNRC6C* gene.



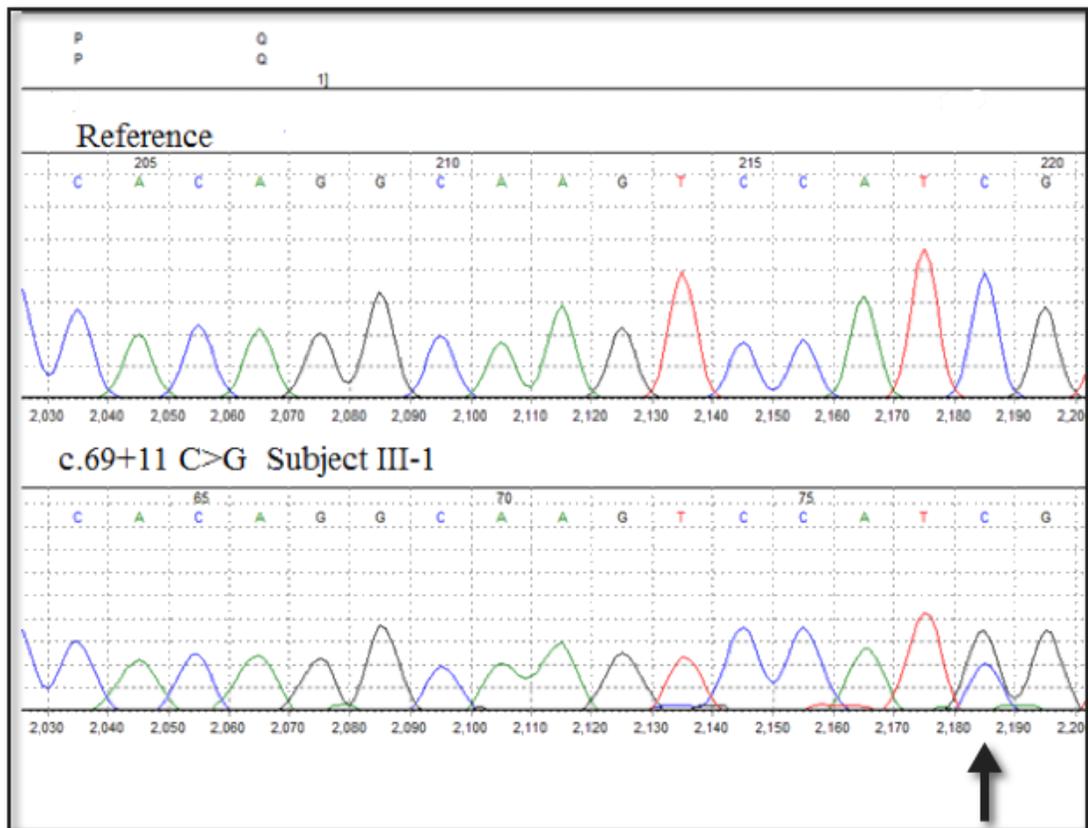
**Figure 4.7: Conservation of *TNRC6C* c.1706 C> T, p. P569L using 13 species**

Upper panel, orthologous from Homo sapien, Pan troglodytes, Macaca mulatta , Rattus norvegicus, Mus musculus, Canis familiaris, Felis catus, Bos Taurus, Dasypus novemcinctus, Monodelphis domestica , Ornithorhynchus anatinus, Gallus gallus, Xenopus tropicalis and Tetraodon nigroviridis aligned and conserved at p. P569L, as indicated by the arrow. Highlighted amino acids are highly conserved. Lower panel, Weblogo image shows conservation of the proline at amino acid position 569 across human and 13 otherspecies. Vertical axis represents the conservation scale; horizontal axis represents the amino acid position.



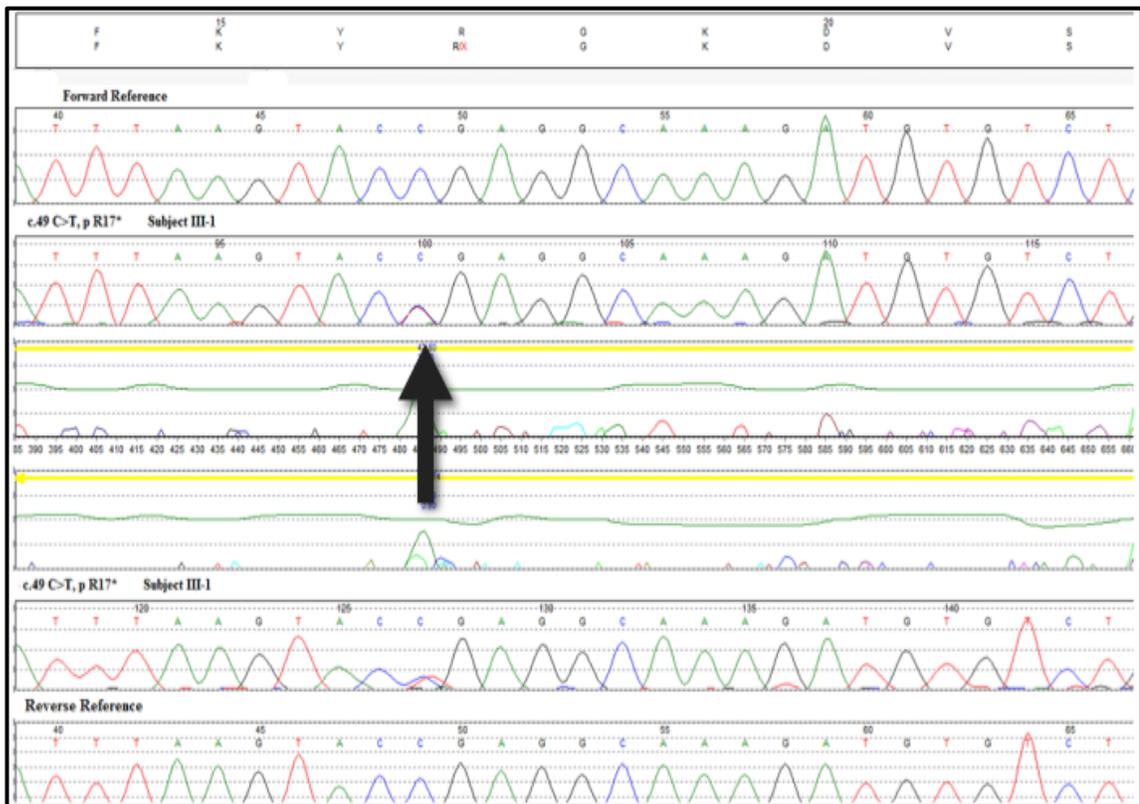
**Figure 4.8: Electropherogram shows predicted splice site mutation c.69+11C>G in the *CDK14* gene.**

Upper panel represents the reference sequence of the *CDK14* gene. Lower panel shows a trace sequence with the splice site c.69+11 C>G in PID III-1, indicated by the arrow.



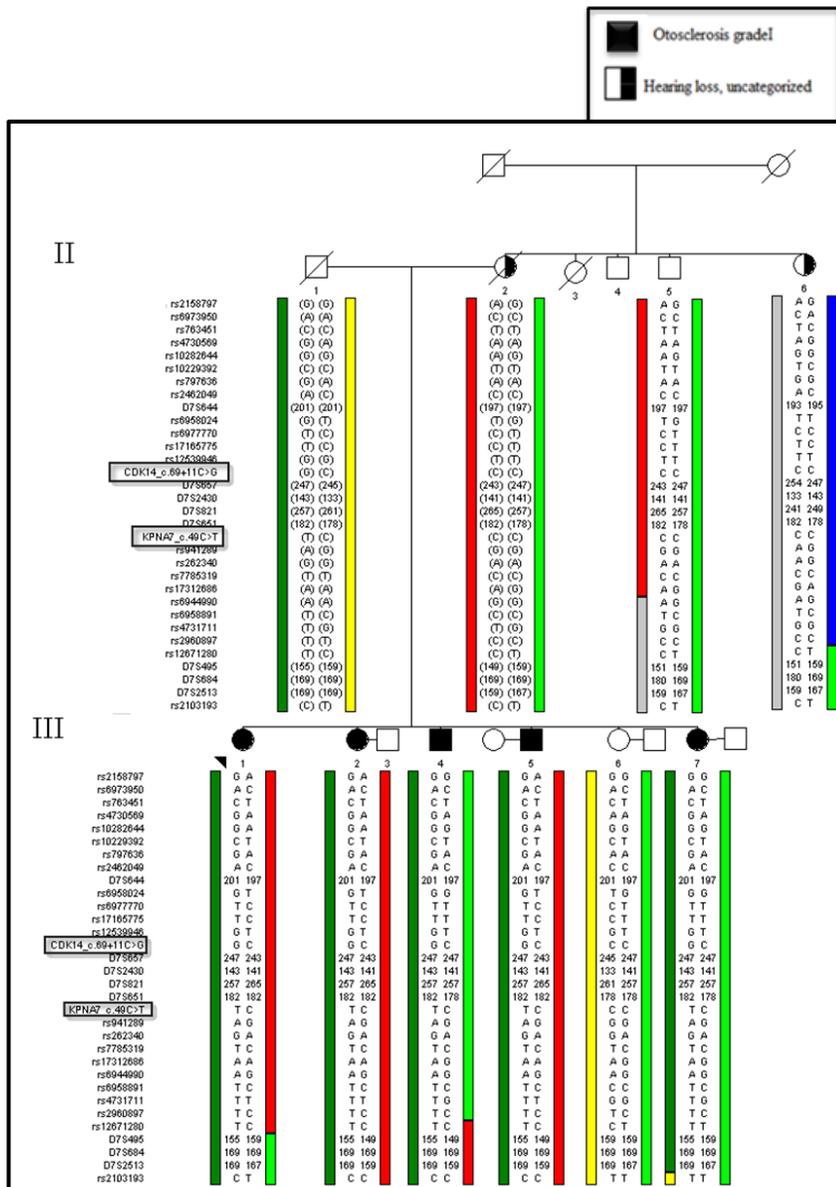
**Figure 4.9: Electropherogram shows nonsense mutation c.49 C>T, p.R17\* in the *KPNA7* gene.**

Upper panel represents the reference sequence of forward strand of the *KPNA7* gene and a trace sequence with the nonsense mutation c.49 C>T in subject III-1. Lower panel represents reference sequence of the reverse strand of the *KPNA7* gene and a trace sequence with the nonsense mutation cc.49 C>T in subject III-1. Arrow points to the nonsense mutation. Amino acids are listed at the top of the caption.



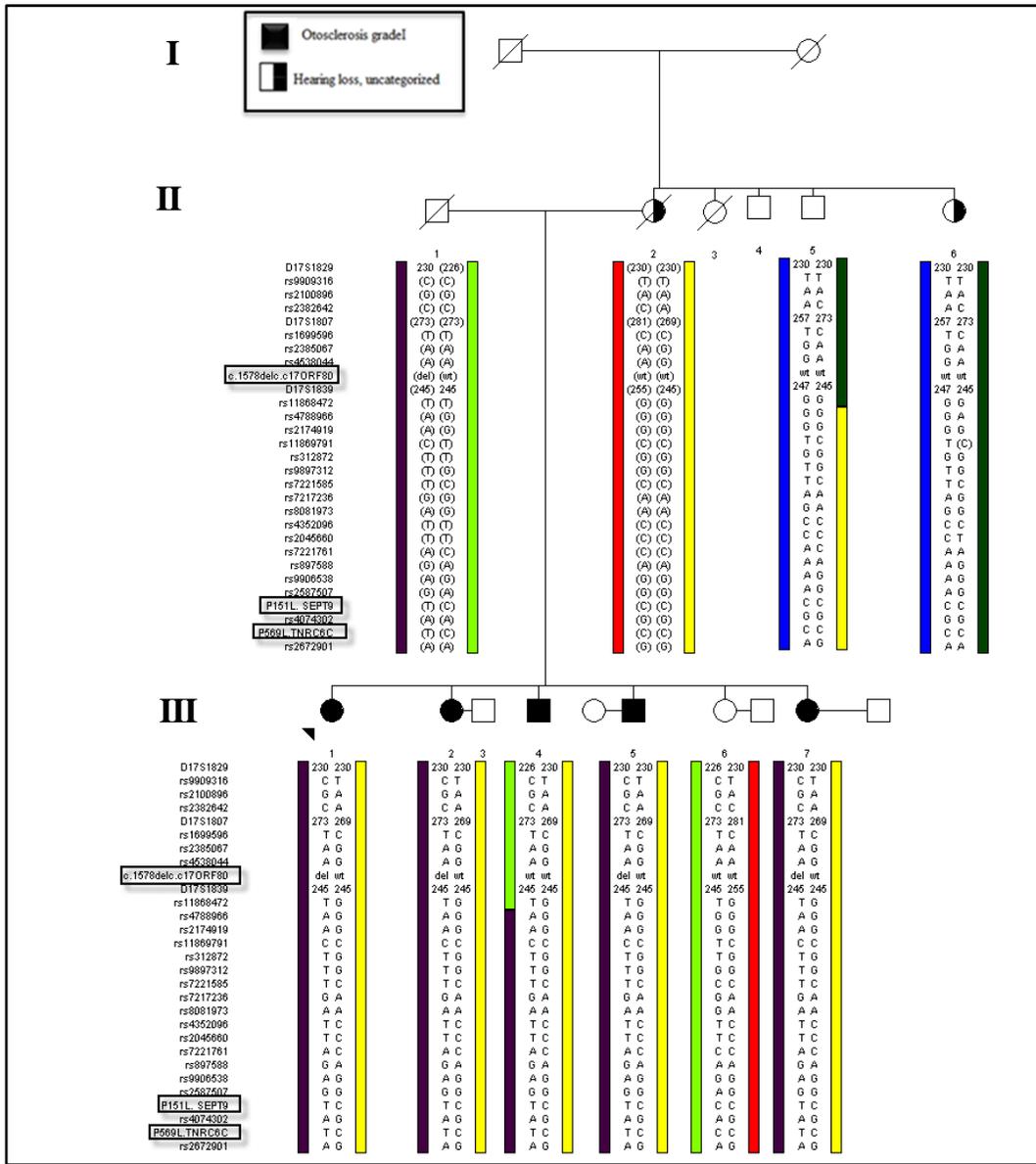
**Figure 4.10: Partial pedigree of Family 2114 shows segregation of *KPNA7* c.49 C>T and *CDK14* c.69+11C>G at Chr7:73,862,412-142,344,569.**

c.49C>T and c.69+11C>G mutations reside on the dark green haplotype that segregates with otosclerosis in Family 2114. Roman numerals on the side indicate generation level. Black solid symbol refers to otosclerosis grade I. Half black and white symbol refers to hearing loss, uncategorized. ( ) refers to inferred alleles. Circle =female, square = male. Cross symbol means deceased.



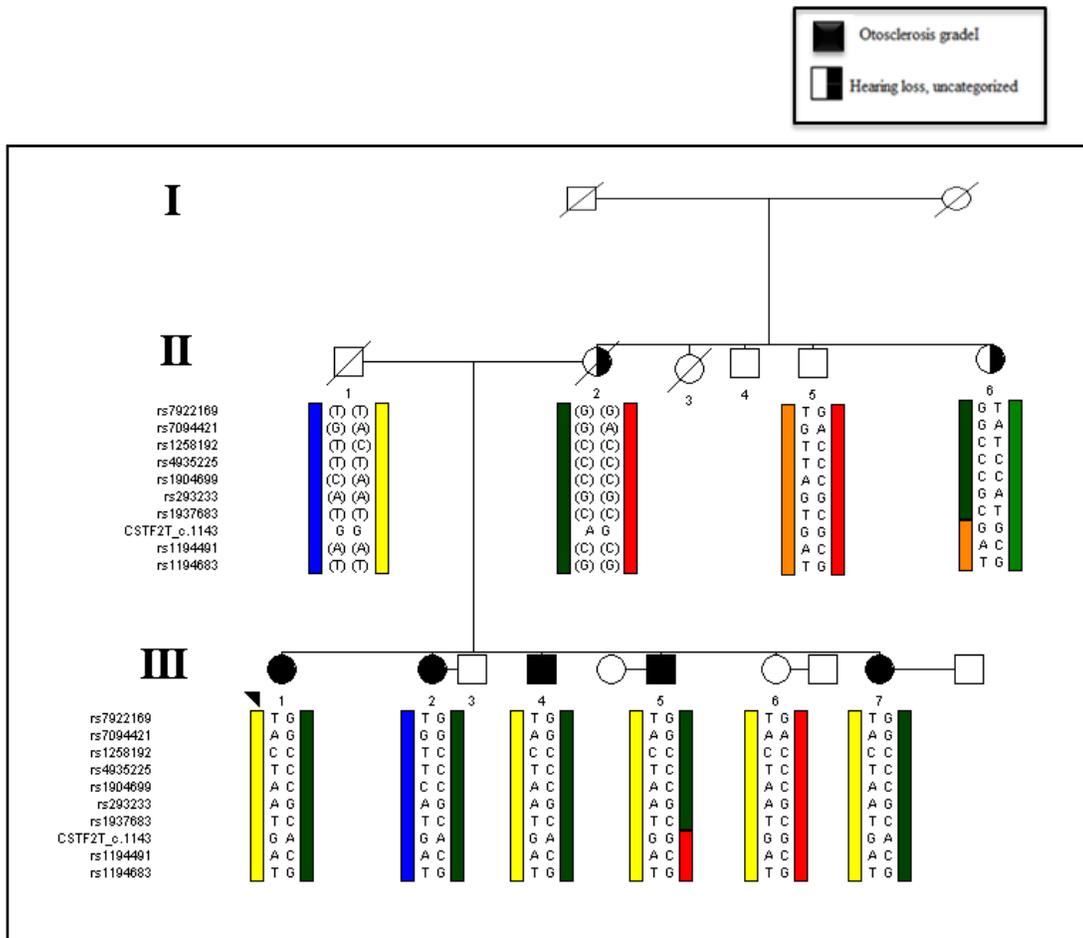
**Figure 4.11: Partial pedigree of Family 2114 showing segregation of three candidate variants identified in Family 2114 at Chr17.**

*C17orf80* c.1578, *SEPT9* c. \*709C>T and *TNRC6C* c.1706 C>T, p. P569L reside on the brown haplotype that segregates with otosclerosis in Family 2114. Roman numerals on the side indicate generation level. Black solid symbol refers to otosclerosis grade I. Half black and white symbol refers to hearing loss, uncategorized. ( ) refers to inferred alleles. Circle = female, square = male. Cross symbol means deceased.



**Figure 4.12: Partial pedigree of Family 2114 showing segregation of *CSTF2T* c.1143 G>A at Ch10q11.23-q21.**

c.1143G>A did not segregate with otosclerosis in Family 2114. Recombination in PID III-5 is detected between rs1937683 and rs1194491. Roman numerals on the side indicate generation level. Black solid symbol refers to otosclerosis grade I. Half black and white symbol refers to hearing loss, uncategorized. Arrows point at boundaries SNPs. ( ) refers to inferred alleles. Circle = female, square = male. Cross symbol means deceased.



#### 4.4 Discussion

Otosclerosis is a complex disease with evidence of Mendelian inheritance<sup>191</sup>. The mode of inheritance in otosclerosis is not clear, however previous studies have reported AD inheritance as the most common mode of inheritance<sup>84,86-90,114</sup>. In Family 2114, the proband's mother was presumed to be the disease carrier under a dominant inheritance model. The disease allele was expected to be inherited from the mother (PID II-2), was found in the 5 affected siblings, and was absent in the unaffected sibling. The DNA of the proband's parents was unavailable, but creating haplotypes for the mother's siblings helped in inferring the mother's haplotypes across the identified regions. Inferring the mother's haplotype helped to identify three excellent candidate pathogenic mutations (in *KPNA7*, *CDK14* and *TNRC6C*) passed down through the father. Two variants located in *KPNA7* and *CDK14*, respectively, reside on the same putative otosclerosis-associated haplotype. The nonsense mutation is located at the first exon of *KPNA7*, which consists of a total of 10 exons. The position of the nonsense mutation was predicted to interrupt the reading frame and create a premature stop codon at the beginning of the gene, which may result in nonsense-mediated decay (NMD) of the gene. NMD is one of the protective mechanisms that selectively degrade the mRNA harbouring nonsense mutations. Without this mechanism, these mRNA would produce a truncated protein which would have a dominant negative effect on the normal allele<sup>192</sup>. Furthermore, *KPNA7* has been predicted to be involved in cytokine signalling in the immune system<sup>193-195</sup>. I had suggested in Chapter 3 that the cytokine-signalling pathway was one that may be deregulated as a result of the *FOXL1* deletion, as this pathway is involved in otosclerosis

pathology. Degradation of the *KPNA7* mRNA by NMD would interfere with its function as a component in the cytokine signalling pathway. The mechanism underlying development of otosclerosis as a result of the nonsense mutation in *KPNA7* is not understood. Further exploration of the role of *KPNA7* in bone remodeling, whether it acts as an activator or inhibitor, will help to identify the possible mechanism acting on the development of otosclerosis.

Cyclin-Dependent Kinase 14 (*CDK14*) gene has only one isoform, NM\_012395, and contains one protein kinase domain. *CDK14* interacts with many interesting proteins, such as *CCDN3*, which regulates cell cycle progression and proliferation<sup>184</sup>. It also interacts with Septin (*SEPT8*), a protein predicted to have a role in cytokinesis, and the *YWHAB* protein, which has a role in negative regulation of osteogenesis<sup>187,188</sup>.

*TNRC6C* (trinucleotide repeat containing 6C) is a large gene with two isoforms (NM\_018996.3 and NM\_001142640.1). NM\_018996.3 consists of 21 exons and encodes 1690 amino acids, while NM\_001142640.1 consists of 22 exons and encodes 1726 amino acids. The *TNRC6C* protein consists of an N-terminal rich with glycine and tryptophan and contains an ago-hook domain that is responsible for binding to Argonaute protein. The c-terminal consists of a GW rich region that is referred to as the silencing domain. The function of *TNRC6C* is not fully understood. The *TNRC6* group of proteins is involved in miRNA induced silencing complexes (miRISCs) through interaction between their N-terminal and Argonaute protein<sup>196</sup>. Disease causing mutations in these genes have not been previously reported. The possible function of the *KPNA7* in cytokine regulation and the high impact of the non-sense mutation make this variant the most promising candidate

mutation. The other variants could have a modifier role in the disease phenotype. Further functional work may be needed to assign causation to one of these three mutations. Also, the ability to test for segregation of these variants in subsequent generations (fourth and fifth generations) may help determine the causative variant. On the other hand, sequencing genomic DNA from other otosclerosis families may also be informative.

The linkage analysis of otosclerosis in Family 2114 identified five regions under both dominant and recessive inheritance. Targeted exome analysis of these regions identified three variants that were inherited from the father and no variants inherited from the mother (presumed disease carrier). This finding suggested that perhaps otosclerosis was inherited from the father's side of the pedigree. As low penetrance is one of characteristics of otosclerosis, the father may be a case of non-penetrance. In another scenario, this variant could have arisen *de novo* in the father's germline and then been passed down to his offspring. Further exploration of the father's side of the pedigree may help in testing segregation of the other variants. In this study, we were able to infer both parental haplotypes but absence of their DNA samples makes it difficult to rule out *de novo* mutations.

Previous linkage analysis (Chapter 2) excluded linkage of this family to the seven previously mapped otosclerosis loci and to the loci of three associated genes (assuming dominant inheritance and transmission from the maternal side of the pedigree). However, there was evidence of possible linkage to a 57.3 Mb region on Chr7q21.11-q35 that overlapped *OTSC2* and *COL1A2* under the assumption of transmission from the paternal side of the pedigree. In order to identify a new otosclerosis locus, genome wide SNP

genotyping, linkage and targeted analyses of the linked regions were carried out. The highest LOD score achieved overall for the pedigree was 2.5 at Chr17q25.1-q25.3 under a recessive model. Conversely, under a dominant mode, linkage analysis gave a LOD score of 1.4 at each of five regions on different chromosomes. Further genotyping of extended family members could decrease the number of the linked regions. However, we were not able to recruit any new family members and therefore, these five regions were fully analyzed, with priority given to the Chr17q25.1-q25.3 region.

In spite of the fact that the otosclerosis trait in Family 2114 appears to be inherited in a dominant pattern, the highest LOD score was achieved at Chr17q under a recessive mode of inheritance (compared to a LOD score of 1.4 under dominant mode of inheritance). This region is 5.79 Mb long and contains 141 genes. Targeted analysis of this region identified a heterozygous variant in the *TNRC6C*. No homozygous or pathogenic compound heterozygous mutations were identified under a recessive mode of inheritance. These findings could indicate that: 1) otosclerosis in Family 2114 is following a recessive mode of inheritance, but also has a second mutation located in a non-coding (intronic or promoter) region not covered by Sanger or exome sequencing; 2) the mode of inheritance in this family is dominant and the high LOD score (2.5) was detected by chance.

Under a dominant mode of inheritance, Family 2114 showed suggestive linkage to five regions. Out of these five regions, one was overlapping with previously mapped *OTSC2* and *COLIA2* regions at Chr7. Mutations in *COLIA2* have been reported to cause OI (a syndromic form of otosclerosis). *COLIA2* is a huge gene, making it impractical to

Sanger sequence. Availability of the exome sequencing data allows us to check for mutations in this gene that could play a role in otosclerosis development. Analyzing exome data for *COLIA2* did not identify any variant that segregated with the affected exomes of Family 2114.

Also, another region at Chr16 was overlapping with the new locus identified in Family 2081 at Chr16. The haplotype inherited by Family 2114 affecting siblings at Chr16-linked locus was different than the linked haplotype of Family 2081. We hypothesized that if Family 2114 was linked to the new locus identified in Family 2081, both families would harbour mutations in the same gene. Because Family 2114 inherited a different haplotype than 2081, we would expect that Family 2114 would have a different mutation than Family 2081 in the same gene or it could have the same mutation in case of a recurrent mutation. Targeted analysis of the exome sequencing at the Chr16 linked region did not identify any variant in Family 2114 that was shared by affected members of Family 2081 or a different variant in the genes of identified variants in Family 2081. This finding could be explained by non-linkage of Family 2114 into the new locus at Chr16q23.1-q24.2. The exome sequencing data of Family 2081 helped exclude variants across Chr7 and Chr17 regions, which significantly decreased the number of the candidate variants.

## 4.5 Conclusion

This study has identified three likely pathogenic mutations that co-segregate with otosclerosis in Family 2114. According to our previous finding on the role of the cytokine pathways in otosclerosis development and the deleterious effect of the non-sense variant in *KPNA7*, it is probable that *KPNA7* will be disease causing in Family 2114. Further functional and segregation studies of these variants are required to determine if this is the case.

## 5 Chapter 5

General discussion

## 5.1 Implication of identifying the first gene for familial otosclerosis

Since 1998, 10 otosclerosis loci have been mapped, but despite recent advances in genomic technologies such as NGS, no *OTSC* gene has been identified. The most recent locus, *OTSC10*, was mapped 4 years ago<sup>84</sup>. This thesis reports the first gene (*FOXLI*) identified as causing familial otosclerosis. To accomplish this, we applied the following guideline: 1) we employed strict diagnostic criteria for otosclerosis, based on the visual inspection of sclerotic foci during surgery; 2) we recruited extended family members and population controls from the homogeneous NL founder population which greatly helped in filtering the NGS variant data; 3) we completed a thorough and careful interpretation of Family 2081 pedigree, taking into account the possibility of both dominant and pseudodominant models, especially in the context of the NL founder population; and 4) we ensured close communication between researchers and clinician which allowed timely access to recruited probands in NL and ON.

The identification of *FOXLI* provides the first opportunity to better understand the pathways that may be involved in the pathogenicity of otosclerosis. For example, our microarray study showed that many cytokines were deregulated by the effect of the *FOXLI* deletion. Cytokines have a major role in many processes that are involved in inflammation, the immune system, and many cytokines are involved in osteoclastogenesis as well as play a role in osteoclast efficiency<sup>197</sup>. From the known histology and pathology of otosclerosis, it appears that the bone cells are important players in bone remodelling and therefore otosclerosis development<sup>198</sup>.

From both literature review and our microarray findings, we speculate that the otosclerosis in Family 2081 developed as a result of cytokine deregulation as an indirect effect of the *FOXL1* deletion. We can now postulate that the causative gene(s) for the previously mapped otosclerosis loci could be a member of the FOX gene family or in fact any gene that is involved in cytokine regulation. Reviews of positional candidate genes in the previously mapped *OTSC* loci have revealed several possible candidates. For example, the mitogen-activated protein kinase 13 (*MAPK-13*) gene, which is found in the *OTSC3* locus encodes protein p38 and is activated by pro-inflammatory cytokines. Similarly, the Mitogen-activated protein 3 kinase 7 (*MAP3K7*) gene located in the *OTSC7* locus has a role in transcriptional regulation activity and apoptosis. In response to interleukin (*IL1*), the *MAP3K7* gene forms a kinase complex that is required for the activation of nuclear factor kappa B, a proinflammatory transcriptional factor<sup>199</sup>. In *OTSC5*, there are two possible candidate genes: Interleukin-2 (*IL2*), a cytokine and Forkhead Box protein L2 (*FOXL2*), a transcriptional factor. In *OTSC8*, there are four forkhead Box protein D4 genes (*FOXD4L2*, *FOXD4L4*, *FOXD4L5*, *FOXD4L6*) while in *OTSC10*, Transforming growth factor, beta 2 gene (*TGFB2*) that encodes members of the transforming growth factor beta (*TGFB*) family of cytokines is located. Based on their related functions, we recommend screening the above genes in families linked to previously mapped loci.

Clinical reports for *FOXL1* family members (Family 2081) indicated variability in the onset of hearing loss and in the degree of severity. The onset of hearing loss in the proband and proband's generation occurred in their early twenties and the progression of

hearing loss was consistent in most of the affected siblings in the proband's generation. In six out of seven sibs, hearing loss was conductive, bilateral and progressive while a seventh sib suffered from a sensorineural loss. In contrast, hearing loss in one of the proband's niece presented by mid-teens and in addition to otosclerosis had a cholesteatoma. Interestingly, two family members, a 40 year old relative with noise-induced hearing loss and a 28 year old with no hearing loss, harbour the *FOXL1* deletion. It is probable these two members may be too young to express the clinical phenotype (an example of age-related penetrance). Recruitment of family members who carry the *FOXL1* deletion from both the right side in Family 2081 pedigree and from the ON family will allow us to conduct a detailed study of the phenotype within each family.

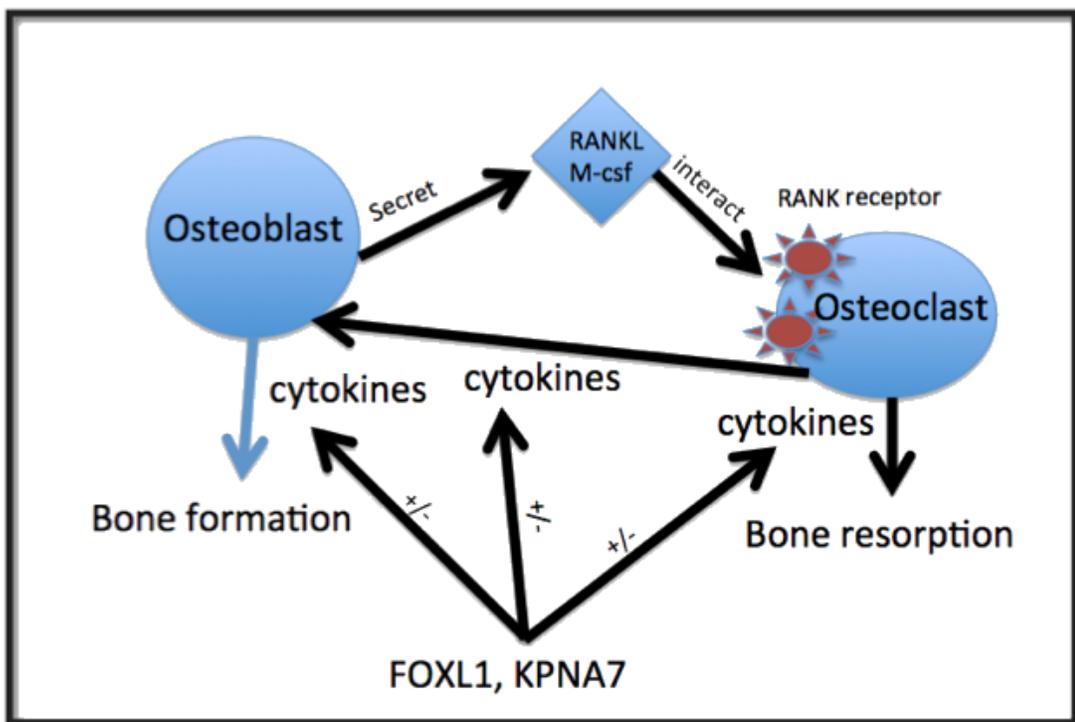
## **5.2 A possible second gene for familial otosclerosis**

In addition to the *FOXL1* mutation, we identified three mutations segregating with otosclerosis in Family 2114. Of these three, the nonsense mutation in the *KNAP7* gene appears to be the most plausible cause of otosclerosis. *KNAP7* gene is a member of the importin family of nuclear importin receptors. A recent study indicates that the bovine *KNAP7* is expressed specifically in the embryo, which indicates a role for *KNAP7* in early development similar to *FOXL1*<sup>195</sup>. *KNAP7* binds to importin  $\alpha$  and importin  $\beta$  and enhances import of the nuclear localization signal (NLS). It has been proposed that changes in the amino acid sequence of *KNAP7* could affect its binding to importin  $\alpha$  and importin  $\beta$  which then could interfere with transport of the NLS out of the nucleus<sup>200</sup>. Like *FOXL1*, *KNAP7* protein is predicted to be involved in the interferon and cytokine signalling pathways<sup>195</sup>. Thus, the role of *KNAP7* in cytokine regulation makes it a good

candidate gene and in conjunction with the *FOXL1* transcription factor will provide an opportunity to study the mechanism(s) underlying the development of otosclerosis and the pathway(s) involved (Figure 5.1). Further recruitment and genotyping of additional members of Family 2114 along with probands from ON otosclerosis families may help us find the gene causing this disease. Confirmation of the identity of the actual variant responsible for otosclerosis in Family 2114 will allow us to conduct a detailed phenotype study of otosclerosis caused by this variant across families and also help in differentiating the phenotype of families with otosclerosis caused by *FOXL1* mutations versus families with otosclerosis caused by other mutations.

**Figure 5.1: Hypothetical figure shows the predicted role of *FOXL1* and *KPNA7* in osteoclastogenesis and bone remodeling.**

Osteoblast secretes RANKL and M-csf. RANKL interacts with the RANK receptor on osteoclasts. This process leads to maturation of the osteoclast and initiates bone resorption process and stimulate osteoblasts. Cytokines have roles in bone resorption and bone formation. *FOXL1* is suggested as having a role in the regulation of cytokines. *KPNA7* has a predicted role in cytokine regulation. Both genes could be components of the same pathway that are involved in cytokine regulation. +/- = Activation/suppression.



## **5.3 Overcoming challenges identifying the genes associated with otosclerosis**

### **5.3.1 Issues with clinical diagnosis**

The most accurate diagnosis of otosclerosis is surgical confirmation of otosclerosis. Conductive hearing loss is common and it is caused by several causes other than otosclerosis. Therefore, when designing a study, it is prudent to include only otosclerosis patients who have otosclerosis confirmed through surgery. Otherwise, linkage analysis may be hindered which may lead to misinterpretation of data. For example, in the first stage of this study, subject IV-5 of Family 2081 (see Chapter 2) was thought to be a phenocopy primarily because the patient had a cholesteatoma in one ear and the surgical report confirming or refuting otosclerosis was unavailable. With the availability of the surgical report, subject IV-5 was re-considered to have grade I otosclerosis. Concluding that subject IV-5 in Family 2081 was a phenocopy resulted in a suggestive linkage to *OTSC4*. Fortunately, confirmation of otosclerosis in subject IV-5 along with subsequent recruitment of other family members in the last year of the study changed the study direction and led to the identification of the *FOXL1* gene. Thus, the use of strict diagnostic criteria for otosclerosis helped us confirm a diagnosis of otosclerosis and further refine the critical disease region. In our study, only members who were diagnosed with otosclerosis through surgery were used. The two unaffected adult members that were originally used as controls were audiologically assessed to have no hearing loss.

### **5.3.2 Recruitment limitations**

Large, extended families with monogenic otosclerosis are rare and this rarity is reflected in the low number of published studies. There are eleven previously reported families with otosclerosis (Table 1.3, Chapter 1.1). Collaboration with Dr. Sue Stanton provided access to 40 otosclerosis probands from ON which helped us in identifying a second otosclerosis family harbouring the *FOXL1* gene mutation. This underscores the positively huge impact that participation rate has in the success of any gene discovery study. The majority of members of the *FOXL1* families were open to participation in the study. Although initially, we were not able to recruit members from the father's side of Family 2114, the family has recently become more co-operative and additional recruitment is in progress.

### **5.3.3 Comprehensive traditional and new technologies**

In early otosclerosis papers, the reported critical disease regions spanned 10-34.16 Mb containing hundreds of genes. The advent of NGS, has helped in the rapid acquisition of data allowing analysis of exomes and whole genomes. The limitations faced by using NGS have primarily been the lack of good depth of coverage and the huge numbers of variants detected<sup>201</sup>. The depth of coverage has improved with the introduction of each new version of advanced chemistry, protocols and instruments resulting in increased sensitivity and higher rate of coverage up to 100 and more. To illustrate, the TruSeq Nano DNA and TruSeq DNA PCR-free Kits used in the Illumina NGS platform exhibits 150 % coverage improvement compared to the original TruSeq DNA Kit<sup>202</sup>. Hundreds of thousands of variants can be identified by NGS, but the filtering and annotation involved

in the process is time consuming. Hundreds of variants which are of interest must then be validated, for example by Sanger sequencing before testing for segregation in families. Fortunately, in cases of Mendelian inheritance, the decreasing cost of NGS is providing an opportunity to sequence a greater number of affected and unaffected family members thereby significantly decreasing the number of variants to be filtered and annotated. By carefully selecting and eliminating variants common to both affecteds and unaffecteds, the number of putative disease-causing variants is reduced which leads to a smaller number of candidate variants for validation and segregation analysis. In our study, the exomes of five out of seven affected members of Family 2081 and four out of five affected members of Family 2114 were sequenced, thus aiding in the elimination of sequencing variants. The NGS data of Family 2081 was also used to reduce the number of variants identified in Family 2114 across linked regions. Consequently, the number of variants to be genotyped in order to test for segregation in extended family members was reduced. In summary, we recognized that sequencing large number of family members has aid to filter out variants and significantly reduce the number of candidate variants.

## **5.4 Future direction**

Identification of the first gene for otosclerosis will have a great impact in the area of otosclerosis.

### **5.4.1 Benefits to patients and their families**

Otosclerosis is a disease of the middle ear but is fortunately a treatable condition. Surgical therapy for otosclerosis has been an established procedure for a few decades. The surgical treatment of otosclerosis is based on either the removal of the stapes footplate (stapedectomy) or the creation of a hole in the stapes footplate (stapedotomy), which helps the circulation of the fluid from the middle ear into the inner ear. However, despite the general availability of treatment for otosclerosis, there are some cases which are not treatable. Furthermore, in many cases, complications may develop as a result of the stapedectomy. Stapedectomy can cause unexplained unilateral SNHL in about 1-2% of the cases<sup>203</sup> and perforation of the eardrum and injury of the facial nerve occurs in about 1 in 1,000. For a minority of the cases, vestibular symptoms may increase or develop<sup>203</sup>.

Identification of the otosclerosis gene can result in early diagnosis of otosclerosis thus allowing early intervention which may prevent or slow the progression of the disease while minimizing the need for surgery which may have undesirable side effects. Otosclerosis is a disease of the middle-aged and the first sign is a loss in low frequency hearing, which may not be noticeable by those affected. Identification of the otosclerosis gene will have many positive benefits with respect to the establishment of screening protocols for otosclerosis. Otosclerosis screening may be extended to newborns in

families with a history of otosclerosis thus helping with early diagnosis. Early diagnosis of otosclerosis by gene screening followed by early treatment of patients based on their mutation results will have huge impact on patients and their families.

It has been reported by many studies that treatment with sodium fluoride (used as a dietary supplement) in early stages of otosclerosis has a beneficial effect in slowing the progression of the disease<sup>204,205</sup>. The main aim of using sodium fluoride is to stop the osteospongiosis phase of otosclerosis. The idea of using sodium fluoride was not acceptable since many side effects were reported. Side effects include stomach upset, allergic itching, and joint pain that can lead to arthritis.

#### **5.4.2 Benefit to biomedical science**

Identification of a gene causative for otosclerosis “opens the gate” for studying the mechanisms underlying the disease. The human fetal osteoblast (hFOB 1.19) cell line is the most acceptable and available for in vitro study. Obtaining a stapes footplate specimen from surgery is the most powerful target for studying mechanism underlying otosclerosis. Many models have been used to study the effect of gene mutation in vivo. Using a mouse model to study the effect of a mutation on hearing organs was a great success. For example, the tectorin alpha (Tecta) mutant mouse helped researchers to better understand the function of the tectorial membrane and its important role in transforming sound waves into a mechanical stimulus. These mechanical stimuli then stimulate sensory hair cells converting the mechanical signal into an electrical signal that is sent to the brain through the cochlear nerve<sup>206</sup>.

A previous study created a mouse model to investigate the role of the *FOXII* gene in the gastrointestinal tract. Homozygous knock out mice showed epithelial cell hyperplasia with distortion of the architectures of the stomach and small intestine <sup>207</sup>. *FOXII* is one of the targeted genes that will be studied by the International Mouse Phenotypic Consortium (IMPC) which is composed of 17 research institutions and their aim is to, by using the embryonic stem cell (ES) of the mouse, develop targeted knockout mice for the 20,000 known or predicted mouse genes. In otosclerosis, conditional deletion is the most acceptable way to test the effect of the mutation on the otic capsule of the mouse model from day zero. This procedure starts by creating target vectors which carry the mutated *FOXLI* gene, inserting this vector into mouse (ES) cells, growing these cells in culture, and then injecting these cells into a developing mouse embryo. The developing embryo will carry the cells that harbour the *FOXLI* deletion. Collection of histological specimens of the temporal bone of the transgenic mice from day zero and across wide range of time will help to determine the onset of the sclerotic foci development and to study of the disease progression, which can eventually help in treating otosclerosis.

### 5.4.3 Gene and drug therapies

Gene therapy has been used experimentally in the treatment of hearing loss in mice<sup>208</sup>. The concept involving gene therapy is to provide mutant cells with the required normal protein by inserting the normal copy of the gene into the disease specific tissue through the use of a specialized vector. Introduction of the normal gene copy allows the diseased tissue to recover and subsequently, to have its normal function restored. As an example, the introduction of normal *Atoh1* gene into the cochlea of young mice stimulates the generation of new sensory hair cells, resulting in the improved production of electrical signals.

It is believed that the development of otosclerosis is a result of abnormal bone remodeling in the otic capsule, which in turn leads to the formation of otosclerotic foci. Fixation of the stapes by the formed sclerotic foci is one of the common events which cause conductive hearing loss. Otosclerotic foci can also affect the cochlea and cause SNHL. In this study, a 15 bp deletion in the *FOXL1* gene has been proposed to cause otosclerosis. Deletion of the *FOXL1* gene results in up-regulation of the FOXL1 protein in the cell nucleus. The mechanism responsible for the development of otosclerosis by *FOXL1* deletion is unknown. We speculate that up-regulation in the FOXL1 protein is a result of accumulation of non-functioning misfolded protein. Consequently, gene therapy can benefit patients with otosclerosis with the introduction of the normal copy of the *FOXL1* gene into the otic capsule. If the therapy is successful, the normal *FOXL1* allele will provide the diseased cells with the required FOXL1 protein, thereby restoring the function of the otic capsule and subsequently improving the hearing function.

In addition to gene therapy, drug therapy has been considered and used in treating genetic diseases. Therapies using anti-cytokines (anti-TNF) have been approved in the United States for the treatment of rheumatoid arthritis (RA)<sup>209</sup>. Rheumatoid arthritis is characterized by ankylosis of the joints, which results from abnormal bone remodeling. The similarity between otosclerosis and rheumatoid arthritis regarding pathogenicity suggests that both diseases might share a common pathway. The results of the microarray study reported in Chapter 3 of this thesis suggest that the cytokine pathway is one of the pathways controlled by *FOXL1* and may therefore play a role in regulation of bone remodeling. Anti-cytokines have been approved as treatment for rheumatoid arthritis. The association between cytokines and *FOXL1* shown in this study warrants further investigation into the possibility of using anti-cytokines in the treatment of otosclerosis.

## 6 Bibliography

1. Dror, A.A. et al. Hearing impairment: a panoply of genes and functions. *Neuron* **68**, 293-308 (2010).
2. Lukashkin, A.N. et al. Multiple roles for the tectorial membrane in the active cochlea. *Hear Res* **266**, 26-35 (2010).
3. Chittka, L. & Brockmann, A. (2005) Perception space--the final frontier. *PLoS Biol* **3**(4), e137. doi:10.1371/journal.pbio.0030137
4. Lawrence, M., D. Wolsk, and W.B. Litton,. Circulation of the inner ear fluids. *Ann Otol Rhinol Laryngol* **70**, 753-76. (1961).
5. Dallos, P., ed. Overview. Cochlear neurophysiology. In *Springer Handbook of Auditory Research: The Cochlea.*, 1-43 (1996).
6. Cabanillas Farpon, R. & Cadinanos Banales, J. [Hereditary hearing loss: genetic counselling]. *Acta Otorrinolaringol Esp* **63**, 218-29 (2012).
7. Alford, R.L. et al. American College of Medical Genetics and Genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. *Genet Med* **16**, 347-55 (2014).
8. Swanepoel, D. Classification of hearing loss. Open access Guide to Audiology and Hearing Aids for Otolaryngologists. Retrived [May 20, 2014], from <http://www.entdev.uct.ac.za/guides/open-access-guide-to-audiology-and-hearing-aids-for-otolaryngologists/>
9. Van Camp G, S.R. Hereditary Hearing Loss Homepage. open access website, Retrived [May 20, 2014]from, <http://hereditaryhearingloss.org>
10. Janecke, A.R. et al. De novo mutation of the connexin 26 gene associated with dominant non-syndromic sensorineural hearing loss. *Hum Genet* **108**, 269-70 (2001).
11. Simon Angeli, X.L.X.Z.L. Genetics of Hearing and Deafness. *he Anatomical Record* **295**, 1812-1829 (2012).
12. Smith RJ, C.G. Deafness and Herediatory Hearing loss Overview In: Gene Reviews at Gene Tests : Medical Genetics Information Resource (database online). . 2008,2007
13. Dobyns, W.B. The pattern of inheritance of X-linked traits is not dominant or recessive, just X-linked. *Acta Paediatr Suppl* **95**, 11-5 (2006).
14. Stanton, S.G. et al. X-linked hearing loss. Two gene mutation examples provide generalizable implications for clinical care. *Am J Audiol* (2014).
15. Van den Veyver, I.B. Skewed X inactivation in X-linked disorders. *Semin Reprod Med* **19**, 183-91 (2001).
16. Hertzano R, A.K. Developmental genes associated with human hearing loss. n: Kelly M, Wu D, Popper AN, Fay RR (eds) *Deveopment of the Inner Ear*. New York: Springer. , p204-232 (2005).
17. Schrijver, I. Hereditary non-syndromic sensorineural hearing loss: transforming silence to sound. *J Mol Diagn* **6**, 275-84 (2004).
18. Tsai, H.T. et al. A novel mutation in the WFS1 gene identified in a Taiwanese family with low-frequency hearing impairment. *BMC Med Genet* **8**, 26 (2007).
19. Online Mendelian Inheritance in Man, O.J.H.U., Baltimore, MD. MIM Number: [# 143100]: {04/25/2013}: . World Wide Web URL: <http://omim.org/>.
20. McNeil, S.M. et al. Reduced penetrance of the Huntington's disease mutation. *Hum Mol Genet* **6**, 775-9 (1997).

21. Young, T.L. *et al.* Non-syndromic progressive hearing loss DFNA38 is caused by heterozygous missense mutation in the Wolfram syndrome gene WFS1. *Hum Mol Genet* **10**, 2509-14 (2001).
22. Napióntek, U. *et al.* Intrafamilial variability of the deafness and goiter phenotype in Pendred syndrome caused by a T416P mutation in the SLC26A4 gene. *J Clin Endocrinol Metab* **89**, 5347-51 (2004).
23. Riccardi, V.M. Neurofibromatosis: clinical heterogeneity. *Curr Probl Cancer* **7**, 1-34 (1982).
24. Rannala, B. & Reeve, J.P. High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *Am J Hum Genet* **69**, 159-78 (2001).
25. Lescai, F. & Franceschi, C. The impact of phenocopy on the genetic analysis of complex traits. *PLoS One* **5**, e11876 (2010).
26. Vahava, O. *et al.* Mutation in transcription factor POU4F3 associated with inherited progressive hearing loss in humans. *Science* **279**, 1950-4 (1998).
27. Abdelfatah, N. *et al.* 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**, 1112-9 (2013).
28. Hildebrand, M.S. *et al.* Audioprofile-directed screening identifies novel mutations in KCNQ4 causing hearing loss at the DFNA2 locus. *Genet Med* **10**, 797-804 (2008).
29. Taylor, K.R. *et al.* AudioGene: predicting hearing loss genotypes from phenotypes to guide genetic screening. *Hum Mutat* **34**, 539-45 (2013).
30. Hildebrand, M.S. *et al.* A contemporary review of AudioGene audioprofiling: a machine-based candidate gene prediction tool for dominant nonsyndromic hearing loss. *Laryngoscope* **119**, 2211-5 (2009).
31. Amos, W., Driscoll, E. & Hoffman, J.I. Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proc Biol Sci* **278**, 1183-8 (2011).
32. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881-5 (2007).
33. Pericak-Vance, M.A. *et al.* Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* **48**, 1034-50 (1991).
34. Riancho, J.A. Genome-wide association studies (GWAS) in complex diseases: advantages and limitations. *Reumatol Clin* **8**, 56-7 (2012).
35. Gulcher, J. Microsatellite markers for linkage and association studies. *Cold Spring Harb Protoc* **2012**, 425-32 (2012).
36. Koboldt, D.C., Miller, R.D. & Kwok, P.Y. Distribution of human SNPs and its effect on high-throughput genotyping. *Hum Mutat* **27**, 249-54 (2006).
37. J., O. Analysis of Human Genetic Linkage. Baltimore, M.T.J.H.U.P., (1991).
38. Strachan T, R.A. Chapter 11, Genetic mapping of Mendelian characters. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK7560/>. *Human Molecular Genetics. 4th edition.*New York: Wiley-Liss, 1999 (2010).
39. Qian, D. & Beckmann, L. Minimum-recombinant haplotyping in pedigrees. *Am J Hum Genet* **70**, 1434-45 (2002).
40. Nachman, M.W. Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev* **12**, 657-63 (2002).
41. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* **58**, 1347-63 (1996).

42. Strachan T, R.A. Chapter 15, Identifying human disease genes. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK7561/>. *Human Molecular Genetics. 2nd edition. New York: Wiley-Liss* (2010).
43. Sanger, F., S. Nicklen, and A.R. Coulson, . DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology*, 1992. 24: p. 104-8. (1992).
44. Sanger, F., S. Nicklen, and A.R. Coulson, . DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, **74**, p.5463-7 (1977).
45. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-11 (2001).
46. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
47. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
48. Ivan Adzhubei, D.M.J., and Shamil R. Sunyaev. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics* **7.20.1-7.20.41**, January 2013(2013).
49. Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. & Krainer, A.R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**, 3568-71 (2003).
50. Desmet, F.O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**, e67 (2009).
51. Fairbrother, W.G., Yeh, R.F., Sharp, P.A. & Burge, C.B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007-13 (2002).
52. Houdayer, C. *et al.* Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* **33**, 1228-38 (2012).
53. Pertea, M., Lin, X. & Salzberg, S.L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* **29**, 1185-90 (2001).
54. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**, 377-94 (2004).
55. Reese, M.G., Eeckman, F.H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J Comput Biol* **4**, 311-23 (1997).
56. Xi, T., Jones, I.M. & Mohrenweiser, H.W. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. *Genomics* **83**, 970-9 (2004).
57. Kircher, M. & Kelso, J. High-throughput DNA sequencing--concepts and limitations. *Bioessays* **32**, 524-36 (2010).
58. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-6 (2009).
59. Casey, J.P. *et al.* Identification of a mutation in LARS as a novel cause of infantile hepatopathy. *Mol Genet Metab* **106**, 351-8 (2012).
60. Geurts-Giele, W.R. *et al.* Molecular diagnostics of a single multifocal non-small cell lung cancer case using targeted next generation sequencing. *Virchows Arch* **462**, 249-54 (2013).
61. Christodoulou, K. *et al.* Next generation exome sequencing of paediatric inflammatory bowel disease patients identifies rare and novel variants in candidate genes. *Gut* **62**, 977-84 (2013).

62. Illumina. (2013). An introduction to next generation sequencing technology. Available/Retrieved from [http://res.illumina.com/documents/products/illumina\\_sequencing\\_introduction.pdf](http://res.illumina.com/documents/products/illumina_sequencing_introduction.pdf)
63. Quail, M.A. *et al.* A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **5**, 1005-10 (2008).
64. Grada, A. & Weinbrecht, K. Next-generation sequencing: methodology and application. *J Invest Dermatol* **133**, e11 (2013).
65. Pabinger, S. *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform*, 2-23 (2013).
66. e Souza C, G.M. Otosclerosis and stapedectomy. . *Thieme, Stuttgart*, (2004).
67. Somers, T.D., F. . Kuhweide, R. and Robillard, Th., Otosclerosis. *B-ENT*, 3, Suppl. 6, 3-10, 2007.
68. Cureoglu, S. *et al.* Otosclerosis: etiopathogenesis and histopathology. *Am J Otolaryngol* **27**, 334-40 (2006).
69. Niedermeyer, H.P. & Arnold, W. Etiopathogenesis of otosclerosis. *ORL J Otorhinolaryngol Relat Spec* **64**, 114-9 (2002).
70. Schuknecht, H.F. & Kirchner, J.C. Cochlear otosclerosis: fact or fantasy. *Laryngoscope* **84**, 766-82 (1974).
71. Gordon, M.A. The genetics of otosclerosis: a review. *Am J Otol* **10**, 426-38 (1989).
72. Declau, F. *et al.* Prevalence of otosclerosis in an unselected series of temporal bones. *Otol Neurotol* **22**, 596-602 (2001).
73. Karimi Yazdi, A., Sazgar, A.A., Motiee, M. & Ashtiani, M.K. Improvement of bone conduction after stapes surgery in otosclerosis patients with mixed hearing loss depending from surgical technique. *Eur Arch Otorhinolaryngol* **266**, 1225-8 (2009).
74. Nager, G.T. Histopathology of otosclerosis. *Arch Otolaryngol* **89**, 341-63 (1969).
75. Causse, J.R.a.L.G.C. Sensorineural hearing loss due to cochlear otospongiosis: etiology. *Otolaryngol Clin North Am*, **11**, 125-34 (1978).
76. Abd el-Rahman, A.G. Cochlear otosclerosis: statistical analysis of relationship of spiral ligament hyalinization to hearing loss. *J Laryngol Otol* **104**, 952-5 (1990).
77. Joseph, R.B. & Frazer, J.P. Otosclerosis Incidence in Caucasians and Japanese. *Arch Otolaryngol* **80**, 256-62 (1964).
78. Huang, T.S. & Lee, F.P. Surgically confirmed clinical otosclerosis among the Chinese. *Arch Otolaryngol Head Neck Surg* **114**, 538-44 (1988).
79. Srivastava, T.P. & Gupta, O.P. Otosclerosis and osteogenesis imperfecta. *J Laryngol Otol* **83**, 1195-204 (1969).
80. Willing, M.C. *et al.* Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. *Am J Hum Genet* **55**, 638-47 (1994).
81. Riedner, E.D., Levin, L.S. & Holliday, M.J. Hearing patterns in dominant osteogenesis imperfecta. *Arch Otolaryngol* **106**, 737-40 (1980).
82. Brown, D.J. *et al.* Autosomal dominant stapes ankylosis with broad thumbs and toes, hyperopia, and skeletal anomalies is caused by heterozygous nonsense and frameshift mutations in NOG, the gene encoding noggin. *Am J Hum Genet* **71**, 618-24 (2002).
83. Usami, S. *et al.* Mutations in the NOG gene are commonly found in congenital stapes ankylosis with symphalangism, but not in otosclerosis. *Clin Genet* **82**, 514-20 (2012).
84. Schrauwen, I. *et al.* A new locus for otosclerosis, OTSC10, maps to chromosome 1q41-44. *Clin Genet* **79**, 495-7 (2011).

85. Alzoubi, F.Q., Ollier, W.R., Ramsden, R.T. & Saeed, S.R. No evidence of linkage between 7q33-36 locus (OTSC2) and otosclerosis in seven British Caucasian pedigrees. *J Laryngol Otol* **121**, 1140-7 (2007).
86. Chen, W. *et al.* Linkage of otosclerosis to a third locus (OTSC3) on human chromosome 6p21.3-22.3. *J Med Genet* **39**, 473-7 (2002).
87. Brownstein, Z., Goldfarb, A., Levi, H., Frydman, M. & Avraham, K.B. Chromosomal mapping and phenotypic characterization of hereditary otosclerosis linked to the OTSC4 locus. *Arch Otolaryngol Head Neck Surg* **132**, 416-24 (2006).
88. Van Den Bogaert, K. *et al.* A fifth locus for otosclerosis, OTSC5, maps to chromosome 3q22-24. *J Med Genet* **41**, 450-3 (2004).
89. Pauw, R.J. *et al.* Phenotype description of a Dutch otosclerosis family with suggestive linkage to OTSC7. *Am J Med Genet A* **143A**, 1613-22 (2007).
90. Bel Hadj Ali, I. *et al.* A new locus for otosclerosis, OTSC8, maps to the pericentromeric region of chromosome 9. *Hum Genet* **123**, 267-72 (2008).
91. Frisch, T., Sorensen, M.S., Overgaard, S. & Bretlau, P. Estimation of volume referent bone turnover in the otic capsule after sequential point labeling. *Ann Otol Rhinol Laryngol* **109**, 33-9 (2000).
92. Zehnder, A.F., Kristiansen, A.G., Adams, J.C., Merchant, S.N. & McKenna, M.J. Osteoprotegerin in the inner ear may inhibit bone remodeling in the otic capsule. *Laryngoscope* **115**, 172-7 (2005).
93. Raisz, L.G. Physiology and pathophysiology of bone remodeling. *Clin Chem* **45**, 1353-8 (1999).
94. Raggatt, L.J. & Partridge, N.C. Cellular and molecular mechanisms of bone remodeling. *J Biol Chem* **285**, 25103-8 (2010).
95. Robling, A.G., Castillo, A.B. & Turner, C.H. Biomechanical and molecular regulation of bone remodeling. *Annu Rev Biomed Eng* **8**, 455-98 (2006).
96. Eriksen, E.F. Cellular mechanisms of bone remodeling. *Rev Endocr Metab Disord* **11**, 219-27 (2010).
97. Ewa Stepień (2011). Acceleration of New Biomarkers Development and Discovery in Synergistic Diagnostics of Coronary Artery Disease, Coronary Angiography - Advances in Noninvasive Imaging Approach for Evaluation of Coronary Artery Disease, Prof. Baskot Branislav (Ed.), ISBN: 978-953-307-675-1, InTech, DOI: 10.5772/18940. Available from: <http://www.intechopen.com/books/coronary-angiography-advances-in-noninvasive-imaging-approach-for-evaluation-of-coronary-artery-disease/acceleration-of-new-biomarkers-development-and-discovery-in-synergistic-diagnostics-of-coronary-art>
98. John B., *et al.* Disease of the temporal bone. In: *Scott-Brown's Otolaryngology, sixth edn, edited by J B Booth, A G Kerr, Vol. 3. London: Butterworth's. pp. 15/1-15/52. (1997).*
99. Chole, R.A. & McKenna, M. Pathophysiology of otosclerosis. *Otol Neurotol* **22**, 249-57 (2001).
100. Harold., S. Pathology of the Ear. *Otology & Neurology*. **22**, 113-122 (2001).
101. Valsalva A. *Valsalvae Opera et Morgagni Epistolae*. Venetiis, Italy. Franciscus Pitteri (1741).
102. Politzer. *Über primäre Erkrankung der knöchernen Labyrinthkapsel*. *Zeitschrift für Ohrenheilkunde* **25**:309-327.
103. Seibenmann, F. Totaler knöcherner Verschleiss beider Labyrinthfester und Labyrinthitis serosa infolge progressiver Spongiosierung. *Verhandlungen Deutschen Otologischen Gesellschaft* **6**:267-283, . (1912).

104. Liktor, B., Csomor, P. & Karosi, T. (2013). Detection of otosclerosis-specific measles virus receptor (cd46) protein isoforms. *ISRN Otolaryngol* **2013**, 6. Available /Retrieved from <http://www.hindawi.com/journals/isrn/2013/479482/>
105. Ross Roeser, M.V. AUDIOLOGY Diagnosis. 90 (2007).
106. Probst, R. Audiological evaluation of patients with otosclerosis. *Adv Otorhinolaryngol* **65**, 119-26 (2007).
107. Toynbee, J. The Diseases of the Ear. *Their Nature, Diagnosis, and Treatment. With a supplement by James Hinton. London* (1868).
108. Albrecht, W., Uber der vererbung der hereditaren labyrinth-schwerkorigkeitund der otosclerose. *Arch. ohrenheilk. Nas. Kehlkopfheilk* 110: p. 15–48. (1922).
109. Larsson, A. Otosclerosis. A genetic and clinical study. *Acta Otolaryngol Suppl* **154**, 1-86 (1960).
110. Morrison, A.W. Genetic factors in otosclerosis. *Ann R Coll Surg Engl* **41**, 202-37 (1967).
111. Fowler, E.P. Otosclerosis in identical twins. A study of 40 pairs. *Arch Otolaryngol* **83**, 324-8 (1966).
112. Hernandez Orozco, F. & Torrescourtney, G. Otosclerosis in Identical Twins; a Genetic and Clinical Study. *Ann Otol Rhinol Laryngol* **74**, 252-9 (1965).
113. Tomek, M.S., et al., . Localization of a gene for otosclerosis to chromosome 15q25-q26. . *Hum Mol Genet*, **7**, 285-90 (1998).
114. Van Den Bogaert, K. *et al.* A second gene for otosclerosis, OTSC2, maps to chromosome 7q34-36. *Am J Hum Genet* **68**, 495-500 (2001).
115. Thys, M. *et al.* A seventh locus for otosclerosis, OTSC7, maps to chromosome 6q13-16.1. *Eur J Hum Genet* **15**, 362-8 (2007).
116. Bauer J., S.C. Vererbung und Konstitution bei ohrenkrakheiten Z. Konstitutionslehr. **10**, 483-545 (1925).
117. Hernandez-Orozco, F. & Courtney, G.T. Genetic Aspects of Clinical Otosclerosis. *Ann Otol Rhinol Laryngol* **73**, 632-44 (1964).
118. Ali, I.B. *et al.* Clinical and genetic analysis of two Tunisian otosclerosis families. *Am J Med Genet A* **143A**, 1653-60 (2007).
119. McKenna, M.J., Kristiansen, A.G., Bartley, M.L., Rogus, J.J. & Haines, J.L. Association of COL1A1 and otosclerosis: evidence for a shared genetic etiology with mild osteogenesis imperfecta. *Am J Otol* **19**, 604-10 (1998).
120. McKenna, M.J., Nguyen-Huynh, A.T. & Kristiansen, A.G. Association of otosclerosis with Sp1 binding site polymorphism in COL1A1 gene: evidence for a shared genetic etiology with osteoporosis. *Otol Neurotol* **25**, 447-50 (2004).
121. Rodriguez, L., Rodriguez, S., Hermida, J., Frade, C., Sande, E., Visedo, G., Martin, C., and C. Zapata, . Proposed association between the COL1A1 and COL1A2 genes and otosclerosis is not supported by a case–control study in Spain. *Am. J. Med.Genet* **A128** (2004).
122. Schrauwen, I. *et al.* COL1A1 association and otosclerosis: a meta-analysis. *Am J Med Genet A* **158A**, 1066-70 (2012).
123. Frenz, D.A., Galinovic-Schwartz, V., Liu, W., Flanders, K.C. & Van de Water, T.R. Transforming growth factor beta 1 is an epithelial-derived signal peptide that influences otic capsule formation. *Dev Biol* **153**, 324-36 (1992).
124. Thys, M., et al., . The coding polymorphism T263I in TGF-beta1 is associated with otosclerosis in two independent populations. *Hum Mol Genet*, **16**, 2021-30 (2007).
125. Schrauwen, I. *et al.* Association of bone morphogenetic proteins with otosclerosis. *J Bone Miner Res* **23**, 507-16 (2008).

126. Schrauwen, I. *et al.* A genome-wide analysis identifies genetic variants in the RELN gene associated with otosclerosis. *Am J Hum Genet* **84**, 328-38 (2009).
127. Khalfallah, A. *et al.* Genetic variants in RELN are associated with otosclerosis in a non-European population from Tunisia. *Ann Hum Genet* **74**, 399-405 (2010).
128. Arcos-Burgos, M. & Muenke, M. Genetics of population isolates. *Clin Genet* **61**, 233-47 (2002).
129. Neuhausen, S.L. Founder populations and their uses for breast cancer genetics. *Breast Cancer Res* **2**, p. 77-81. (2000).
130. Morell, R.J. *et al.* Mutations in the connexin 26 gene (GJB2) among Ashkenazi Jews with nonsyndromic recessive deafness. *N Engl J Med* **339**, 1500-5 (1998).
131. Delmaghani, S. *et al.* Mutations in the gene encoding pejvakin, a newly identified protein of the afferent auditory pathway, cause DFNBS9 auditory neuropathy. *Nat Genet* **38**, 770-8 (2006).
132. Lafreniere, R.G., *et al.*, Identification of a novel gene (HSN2) causing hereditary sensory and autonomic neuropathy type II through the Study of Canadian Genetic Isolates. *Am J Hum Genet*, **74**, 1064-73. (2004).
133. Walsh, T. *et al.* Genomic analysis of a heterogeneous Mendelian phenotype: multiple novel alleles for inherited hearing loss in the Palestinian population. *Hum Genomics* **2**, 203-11 (2006).
134. Mannion, J., ed. *The peopling of Newfoundland: Essays in Historical Geography.* Institute of Social and Economic Research. 1977, Memorial University: St. John's.
135. Bear, J.C. *et al.* Persistent genetic isolation in outport Newfoundland. *Am J Med Genet* **27**, 807-30 (1987).
136. Bernalova, I.N. *et al.* Mutations in the Wolfram syndrome 1 gene (WFS1) are a common cause of low frequency sensorineural hearing loss. *Hum Mol Genet* **10**, 2501-8 (2001).
137. Hildebrand, M.S., Sorensen, J.L., Jensen, M., Kimberling, W.J. & Smith, R.J. Autoimmune disease in a DFNA6/14/38 family carrying a novel missense mutation in WFS1. *Am J Med Genet A* **146A**, 2258-65 (2008).
138. Ahmed, Z.M. *et al.* 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 (2004).
139. Doucette, L. *et al.* 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**, 554-64 (2009).
140. Abdelfatah, N. *et al.* 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**, 66-9 (2013).
141. Karolchik, D., *et al.*, *The UCSC Genome Browser Database.* *Nucl Acids Res* **31**, p.51-4 (2003).
142. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
143. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115 (2012).
144. Jensen, M.A., Fukushima, M. & Davis, R.W. DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis. *PLoS One* **5**, e11024 (2010).
145. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-90 (2004).

146. Woods, M.O. *et al.* The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. *Gut* **59**, 1369-77 (2010).
147. Cottingham, R.W., Jr., Idury, R.M. & Schaffer, A.A. Faster sequential genetic linkage computations. *Am J Hum Genet* **53**, 252-63 (1993).
148. Holt, J.J. Cholesteatoma and otosclerosis: two slowly progressive causes of hearing loss treatable through corrective surgery. *Clin Med Res* **1**, 151-4 (2003).
149. Gudbjartsson, D.F. *et al.* A sequence variant in ZFX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet* **41**, 876-8 (2009).
150. Benjamin, E.J. *et al.* Variants in ZFX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet* **41**, 879-81 (2009).
151. Pauw, R.J. *et al.* The phenotype of the first otosclerosis family linked to OTSC5. *Otol Neurotol* **27**, 308-15 (2006).
152. Carneiro, M., Ferrand, N. & Nachman, M.W. Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* **181**, 593-606 (2009).
153. Jensen-Seaman, M.I. *et al.* Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res* **14**, 528-38 (2004).
154. Carlsson, P. & Mahlapuu, M. Forkhead transcription factors: key players in development and metabolism. *Dev Biol* **250**, 1-23 (2002).
155. Kaufmann, E. & Knochel, W. Five years on the wings of fork head. *Mech Dev* **57**, 3-20 (1996).
156. Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. & Monaco, A.P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519-23 (2001).
157. Nishimura, D.Y. *et al.* The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. *Nat Genet* **19**, 140-7 (1998).
158. Mirzayans, F. *et al.* Axenfeld-Rieger syndrome resulting from mutation of the FKHL7 gene on chromosome 6p25. *Eur J Hum Genet* **8**, 71-4 (2000).
159. Nakada, C., Satoh, S., Tabata, Y., Arai, K. & Watanabe, S. Transcriptional repressor foxl1 regulates central nervous system development by suppressing shh expression in zebra fish. *Mol Cell Biol* **26**, 7246-57 (2006).
160. Madison, B.B., McKenna, L.B., Dolson, D., Epstein, D.J. & Kaestner, K.H. FoxF1 and FoxL1 link hedgehog signaling and the control of epithelial proliferation in the developing stomach and intestine. *J Biol Chem* **284**, 5936-44 (2009).
161. Bale, A.E. Hedgehog signaling and human disease. *Annu Rev Genomics Hum Genet* **3**, 47-65 (2002).
162. Teng, Y.H., Aquino, R.S. & Park, P.W. Molecular functions of syndecan-1 in disease. *Matrix Biol* **31**, 3-16 (2012).
163. Perreault, N., Katz, J.P., Sackett, S.D. & Kaestner, K.H. Foxl1 controls the Wnt/beta-catenin pathway by modulating the expression of proteoglycans in the gut. *J Biol Chem* **276**, 43328-33 (2001).
164. Park, S.J. *et al.* The forkhead transcription factor Foxc2 promotes osteoblastogenesis via up-regulation of integrin beta1 expression. *Bone* **49**, 428-38 (2011).
165. Claudio, N., Unfolding the role of protein misfolding in neurodegenerative
166. Kim, J.H. *et al.* The mechanism of osteoclast differentiation induced by IL-1. *J Immunol* **183**, 1862-70 (2009).

167. Ildeu Andrade, S.R.A.T., Paulo E.A. Souza. Inflammation and Tooth Movement: The Role of Cytokines, Chemokines, and Growth Factors. *Seminar in Orthodontics* **18**, 257-269 (2012).
168. Lisignoli, G. *et al.* Human osteoblasts express functional CXC chemokine receptors 3 and 5: activation by their ligands, CXCL10 and CXCL13, significantly induces alkaline phosphatase and beta-N-acetylhexosaminidase release. *J Cell Physiol* **194**, 71-9 (2003).
169. Gasper, N.A., Petty, C.C., Schrum, L.W., Marriott, I. & Bost, K.L. Bacterium-induced CXCL10 secretion by osteoblasts can be mediated in part through toll-like receptor 4. *Infect Immun* **70**, 4075-82 (2002).
170. Lee, E.Y. *et al.* Potential role and mechanism of IFN-gamma inducible protein-10 on receptor activator of nuclear factor kappa-B ligand (RANKL) expression in rheumatoid arthritis. *Arthritis Res Ther* **13**, R104 (2011).
171. Takayanagi, H. *et al.* RANKL maintains bone homeostasis through c-Fos-dependent induction of interferon-beta. *Nature* **416**, 744-9 (2002).
172. Wang, F. *et al.* Interleukin-29 modulates proinflammatory cytokine production in synovial inflammation of rheumatoid arthritis. *Arthritis Res Ther* **14**, R228 (2012).
173. Ealy, M. *et al.* Gene expression analysis of human otosclerotic stapedial footplates. *Hear Res* **240**, 80-6 (2008).
174. Chen, B.S., Yang, S.K., Lan, C.Y. & Chuang, Y.J. A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. *BMC Med Genomics* **1**, 46 (2008).
175. Srinivas, S. *et al.* Interferon-lambda1 (interleukin-29) preferentially down-regulates interleukin-13 over other T helper type 2 cytokine responses in vitro. *Immunology* **125**, 492-502 (2008).
176. Jordan, W.J. *et al.* Modulation of the human cytokine response by interferon lambda-1 (IFN-lambda1/IL-29). *Genes Immun* **8**, 13-20 (2007).
177. Lin, S.C., Kuo, C.C., Tsao, J.T. & Lin, L.J. Profiling the expression of interleukin (IL)-28 and IL-28 receptor alpha in systemic lupus erythematosus patients. *Eur J Clin Invest* **42**, 61-9 (2012).
178. Grassi, F. *et al.* Human osteoclasts express different CXC chemokines depending on cell culture substrate: molecular and immunocytochemical evidence of high levels of CXCL10 and CXCL12. *Histochem Cell Biol* **120**, 391-400 (2003).
179. Lisignoli, G.T.S., Piacentini A. Human Osteoblasts Express Functional CXC Chemokine Receptors 3 and 5: Activation by Their Ligands, CXCL10 and CXCL13, Significantly Induces Alkaline Phosphatase and b-N-Acetylhexosaminidase Release. *JOURNAL OF CELLULAR PHYSIOLOGY* **194**, 71-79 (2002).
180. Kyd, P.A., Vooght, K.D., Kerkhoff, F., Thomas, E. & Fairney, A. Clinical usefulness of bone alkaline phosphatase in osteoporosis. *Ann Clin Biochem* **35** ( Pt 6), 717-25 (1998).
181. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).
182. Weinberg, W. Über den Nachweis der Vererbung beim Menschen. *ahreshefte des Vereins Varterländische Naturkdunde in Württemberg* **64**, 369-382 (1908).
183. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28**, 1248-50 (2010).
184. Shu, F. *et al.* Functional characterization of human PFTK1 as a cyclin-dependent kinase. *Proc Natl Acad Sci U S A* **104**, 9248-53 (2007).
185. Davidson, G. *et al.* Cell cycle control of wnt receptor activation. *Dev Cell* **17**, 788-99 (2009).

186. Qiang, Y.W. *et al.* Characterization of Wnt/beta-catenin signalling in osteoclasts in multiple myeloma. *Br J Haematol* **148**, 726-38 (2010).
187. Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* **2011**, 9 (2011).
188. Liu, Y., Ross, J.F., Bodine, P.V. & Billiard, J. Homodimerization of Ror2 tyrosine kinase receptor induces 14-3-3(beta) phosphorylation and promotes osteoblast differentiation and bone formation. *Mol Endocrinol* **21**, 3050-61 (2007).
189. Chyi-Ying A. Chen, D.Z., Zhenfang Xia, and Ann-Bin Shyu. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol.* **November;16**, 1160-1166 (2009).
190. Sugatani, T. & Hruska, K.A. Impaired micro-RNA pathways diminish osteoclast differentiation and function. *J Biol Chem* **284**, 4667-78 (2009).
191. Somers, T.D., F. Kuhweide, R. and Robillard, Th. Otosclerosis. *B-ENT,3, Suppl. 6, 3-10* (2007).
192. Chang, Y.F., Imam, J.S. & Wilkinson, M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* **76**, 51-74 (2007).
193. Haw, R.A. *et al.* The Reactome BioMart. *Database (Oxford)* **2011**, 31 (2011).
194. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* **39**, 691-7 (2011).
195. D'Eustachio, P. Reactome knowledgebase of human biological pathways and processes. *Methods Mol Biol* **694**, 49-61 (2011).
196. Lazzaretti, D., Tournier, I. & Izaurrealde, E. The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. *RNA* **15**, 1059-66 (2009).
197. Lee, S.K. & Lorenzo, J. Cytokines regulating osteoclast formation and function. *Curr Opin Rheumatol* **18**, 411-8 (2006).
198. Tanaka, Y., Nakayamada, S. & Okada, Y. Osteoblasts and osteoclasts in bone remodeling and inflammation. *Curr Drug Targets Inflamm Allergy* **4**, 325-8 (2005).
199. Ishitani, T. *et al.* Role of the TAB2-related protein TAB3 in IL-1 and TNF signaling. *EMBO J* **22**, 6277-88 (2003).
200. Kelley, J.B., Talley, A.M., Spencer, A., Gioeli, D. & Paschal, B.M. Karyopherin alpha7 (KPNA7), a divergent member of the importin alpha family of nuclear import receptors. *BMC Cell Biol* **11**, 63 (2010).
201. Voelkerding, K.V., Dames, S.A. & Durtschi, J.D. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* **55**, 641-58 (2009).
202. Illumina.(2013). Comparison of TrueSeq Sample Preparation Kit. Availabe/Retrived from [http://res.illumina.com/documents/products/technotes/technote\\_truseq\\_comparison.pdf](http://res.illumina.com/documents/products/technotes/technote_truseq_comparison.pdf).
203. Mahafza, T., Al-Layla, A., Tawalbeh, M., Abu-Yagoub, Y. & Atwan Sulaiman, A. Surgical Treatment of Otosclerosis: Eight years' Experience at the Jordan University Hospital. *Iran J Otorhinolaryngol* **25**, 233-8 (2013).
204. Shambaugh, G.E., Jr. Fluoride therapy for otosclerosis. *Arch Otolaryngol Head Neck Surg* **116**, 1217 (1990).
205. Naumann, I.C., Porcellini, B. & Fisch, U. Otosclerosis: incidence of positive findings on high-resolution computed tomography and their correlation to audiological test data. *Ann Otol Rhinol Laryngol* **114**, 709-16 (2005).
206. Legan, P.K. *et al.* A targeted deletion in alpha-tectorin reveals that the tectorial membrane is required for the gain and timing of cochlear feedback. *Neuron* **28**, 273-85 (2000).

207. Kaestner, K.H., Silberg, D.G., Traber, P.G. & Schutz, G. The mesenchymal winged helix transcription factor Fkh6 is required for the control of gastrointestinal proliferation and differentiation. *Genes Dev* **11**, 1583-95 (1997).
208. Kelly, M.C., Chang, Q., Pan, A., Lin, X. & Chen, P. Atoh1 directs the formation of sensory mosaics and induces cell proliferation in the postnatal mammalian cochlea in vivo. *J Neurosci* **32**, 6699-710 (2012).
209. Caporali, R. *et al.* Treatment of rheumatoid arthritis with anti-TNF-alpha agents: a reappraisal. *Autoimmun Rev* **8**, 274-80 (2009).
210. Shanghai Genomics, I., ‡Chinese National Human Genome Center, Shanghai, Zhangjiang Hi-Tech Park, Shanghai 201203. Functional characterization of human PFTK1 as a cyclin-dependent kinase. *PNAS* **104**, 9249 (2007).
211. Gary Davidson, J.S. *et al.* Cyclin Y, a novel membrane-associated cyclin, interacts with PFTK1. *Developmental Cell*, 788-799 (2009).
212. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
213. Oshlack, A., Robinson, M.D. & Young, M.D. From RNA-seq reads to differential expression results. *Genome Biol* **11**, 220 (2010).
214. Beysen, D. *et al.* Missense mutations in the forkhead domain of FOXL2 lead to subcellular mislocalization, protein aggregation and impaired transactivation. *Hum Mol Genet* **17**, 2030-8 (2008).
215. Mounne, L. *et al.* Differential aggregation and functional impairment induced by polyalanine expansions in FOXL2, a transcription factor involved in cranio-facial and ovarian development. *Hum Mol Genet* **17**, 1010-9 (2008).
216. Cubitt, A.B. *et al.* Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* **20**, 448-55 (1995).
217. Han, W. *et al.* C-terminal ECFP fusion impairs synaptotagmin 1 function: crowding out synaptotagmin 1. *J Biol Chem* **280**, 5089-100 (2005).
218. German-Retana, S., Candresse, T., Alias, E., Delbos, R.P. & Le Gall, O. Effects of green fluorescent protein or beta-glucuronidase tagging on the accumulation and pathogenicity of a resistance-breaking Lettuce mosaic virus isolate in susceptible and resistant lettuce cultivars. *Mol Plant Microbe Interact* **13**, 316-24 (2000).

## **7 Appendices**

### **Appendix 1: Hearing loss medical questionnaire**

---

---

# **Newfoundland and Labrador Hearing Loss Study**

## **Medical Information Questionnaire**

---

---

Adapted from:

THE HARVARD CENTRE  
FOR HEREDITARY HEARING LOSS

1. Your Name \_\_\_\_\_ Date of Birth \_\_\_\_\_

Address \_\_\_\_\_

Home Phone \_\_\_\_\_ Work Phone \_\_\_\_\_

E-mail Address (if you have one) \_\_\_\_\_

2. To your knowledge, are your parents related, even distantly?  Yes  No  Don't Know  
(This may sound like a strange question, but in a genetic study, we ask it of everyone)

3. Have you ever visited any of the following doctors?

- An ENT (Ear, Nose and Throat) Doctor?  Yes  No  Don't Know  
\_\_\_\_\_

- An Audiologist? (Person performing hearing tests)  Yes  No  Don't Know  
\_\_\_\_\_

- An Eye Doctor? (Ophthalmologist)  Yes  No  Don't Know  
\_\_\_\_\_

- A Genetics Doctor? (Geneticist)  Yes  No  Don't Know  
\_\_\_\_\_

- A doctor who treats diseases of the nervous system? (Neurologist)  Yes  No  Don't Know  
\_\_\_\_\_

- A Heart Doctor? (Cardiologist)  Yes  No  Don't Know  
\_\_\_\_\_

4. Have you ever been admitted to hospital? If yes, please give name of hospital and approximate date(s) of admission.

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_



31. Fluctuating (sometimes better, sometimes worse)..... Right       Left       N/A
32. Slowly progressing (getting worse over years)..... Right       Left       N/A
33. Rapidly progressing (getting worse over weeks/months)..... Right       Left       N/A
34. Sudden hearing loss ..... Right       Left       N/A
35. Did the patient's mother have German measles (Rubella), CMV, toxoplasmosis or any other infections during pregnancy?.....  Yes  No       Don't Know  
If yes, which infections? \_\_\_\_\_
36. Did/does the patient's mother have diabetes?..... Yes  No       Don't Know  
If yes, at what age was diabetes diagnosed? \_\_\_\_\_
37. Did the patient's mother have any other illnesses during pregnancy?.. Yes  No  Don't Know  
If yes, which illnesses? What time during the pregnancy did they occur?  
\_\_\_\_\_
38. Did the patient's mother have any exposure to medication or drugs during pregnancy?..... Yes       No       Don't Know  
If yes, which substances? When during the pregnancy were they taken?  
\_\_\_\_\_
39. Did the patient's mother smoke during pregnancy?..... Yes       No       Don't Know  
If yes, how much and how often \_\_\_\_\_
40. Did the patient's mother receive any vaccinations during pregnancy?... Yes  No  Don't Know  
If yes, which vaccinations? When during pregnancy were they received?  
\_\_\_\_\_
41. Was the patient born prematurely?..... Yes       No       Don't Know  
If yes, how early? \_\_\_\_\_

42. At birth, did the patient require: If so, for what reason?
- any special or intensive care as a newborn \_\_\_\_\_
- antibiotic treatment \_\_\_\_\_
- help with breathing, e.g. ventilator \_\_\_\_\_
- prolonged hospitalization \_\_\_\_\_
- light therapy (phototherapy) as a newborn \_\_\_\_\_
43. Did the patient have any other serious illnesses at birth?.....  Yes  No  Don't Know  
If yes, what other illnesses? \_\_\_\_\_
44. Scarlet fever..... Yes  No  Don't Know
45. Measles or German measles...(circle which one)..... Yes  No  Don't Know
46. Mumps..... Yes  No  Don't Know
47. Meningitis (brain infection)..... Yes  No  Don't Know
48. Tuberculosis (TB))..... Yes  No  Don't Know
49. Repeated or chronic ear drainage..... Yes  No  Don't Know
50. Cyst (cholesteatoma) of middle ear..... Yes  No  Don't Know
51. Mastoiditis..... Yes  No  Don't Know
52. Ear Surgery (specify) \_\_\_\_\_..... Yes  No  Don't Know
53. Exposure to very loud noises..... Yes  No  Don't Know
54. Spinning, dizziness, lightheadedness, or unsteadiness.... Yes  No  Don't Know
55. Drop attacks (sudden fainting spells)..... Yes  No  Don't Know
56. Tinnitus (ringing, buzzing or other sounds in the ear(s))....  Yes  No  Don't Know
57. Unusual marks, skin tags or other abnormalities of the ears or ear lobes..... Yes  No  Don't Know
58. Unusual shape to the head or facial features.....  Yes  No  Don't Know
59. White skin patches..... Yes  No  Don't Know
60. Brown skin patches..... Yes  No  Don't Know
61. Red skin patches..... Yes  No  Don't Know

62. White patch of hair on the head..... Yes       No       Don't Know  
 Age when this appeared: \_\_\_\_\_)
63. Premature graying of hair before age 30..... Yes       No       Don't Know  
 (Not just at the temples)
64. Jerky eye movements (nystagmus)..... Yes       No       Don't Know
65. White part of the eye showing a pale blue color..... Yes       No       Don't Know
66. Colored part of the eye showing two different colors in the  
 same eye (eg. Left eye both blue and brown) ..... Yes       No       Don't Know
67. Colored part of the eye being different in each eye  
 (eg. Left eye blue, right eye brown)..... Yes       No       Don't Know
68. Brilliant blue of the colored parts of the eyes..... Yes       No       Don't Know
69. Night blindness..... Yes       No       Don't Know
70. Color blindness..... Yes       No       Don't Know
71. Severe progressive nearsightedness (able to see near objects  
 more clearly than distant ones without glasses)..... Yes       No       Don't Know
72. Cloudy vision (cataracts)..... Yes       No       Don't Know
73. Decreased peripheral vision..... Yes       No       Don't Know  
 (ability to see objects and movement outside of the direct line of vision)
74. Absent or deformed tear ducts..... Yes       No       Don't Know
75. Kidney problems..... Yes       No       Don't Know
76. Diabetes mellitus ("sugar diabetes")..... Yes       No       Don't Know
77. Thyroid problems (goiter, underactive, overactive)..... Yes       No       Don't Know
78. Impairment of smell, taste, or touch..... Yes       No       Don't Know
79. Learning impairment..... Yes       No       Don't Know
80. Is there anything else which you think we should know about the patient's medical history? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

81. Hearing test?..... Yes       No       Don't Know

If yes, please give dates and locations if you know them:

Date(s) \_\_\_\_\_

Place where test was done \_\_\_\_\_

82. Vestibular Testing (Balance Testing)  
(ENG, Rotary Chair, Posturography)?..... Yes       No       Don't Know

If yes, please give dates and locations if you know them:

Date(s) \_\_\_\_\_

Place where test was done \_\_\_\_\_

83. CT Scan/MRI scan of the head or ear?..... Yes       No       Don't Know

If yes, please give dates and locations if you know them:

Date(s) \_\_\_\_\_

Place where test was done \_\_\_\_\_

## **Appendix 2: DNA extraction from whole blood**

**Purpose:** To provide a method for the extraction, precipitation, and resuspension of DNA from (EDTA) whole blood.

**Responsibilities:** Lab personnel who have been trained in this procedure.

### **Definitions:**

- hr. = hour
- ppt = precipitate
- WBC = white blood cell

### **Equipment and Supplies:**

#### Equipment:

- centrifuge
- Biological safety cabinet
- 37° or 55 °C water bath or heat block
- Pipette aid

#### Supplies:

- 50 ml & 15ml centrifuge tubes
- 1 & 10ml pipettes
- 2 ml microwtube with o-ring seal screw cap
- “hooked” Pasteur pipettes
- 500ml plastic beaker

#### Reagents:

- RBC lysis solution (NH<sub>4</sub>CL/Tris)
- 0.85% NaCl
- Nuclei lysis buffer (10mM Tris-HCl, 400mM NaCl, 2.0 mM EDTA, pH 8.0).
- 10% SDS
- Protease solution (3mg/ml protease, 1% SDS, 2mM EDTA)
- Saturated NaCl
- Absolute ethanol
- TE buffer
- 70% ethanol
- TE buffer (10mM Tris, 1mM EDTA, pH 8.0)
- Bleach (for disinfection).

## Procedure:

**Note:** *place a 1.0 litre plastic beaker containing ~100ml of bleach in the safety cabinet for waste discard/disinfection. Decant all supernatants etc. into this container.*

1. Pre-warm RBC lysis solution to 37°C.
2. In a 50ml centrifuge tube mix 5 vols. of RBC lysis solution with 1 vol. of whole blood (to avoid aerosols add the lysis solution first then the blood). Mix by inverting and then incubate at 37°C for 5min to 1 hr.
3. Centrifuge at 1000 x g (2500 rpm) for 5min.
4. Decant the supernatant and add 10 ml of 0.85% NaCl. Vortex (vigorously) to resuspend the WBC pellet then centrifuge (as previous).
5. Decant the supernatant and add 3ml of nuclei lysis buffer
6. Vortex (vigorously) to resuspend the WBC pellet then add 0.2 ml of 10% SDS and 0.5ml of Protease solution.
7. Incubate at 55°C for 2 hr. or 37°C for 12 to 60 hr.
8. Add 1.0ml of saturated NaCl, shake/vortex vigorously for ~15sec. then centrifuge at 1000 x g (~2500 rpm) for 15min's.
9. Carefully (so as not to disturb the pellet) decant the supernatant into a clean 15ml centrifuge tube containing 2 volumes of absolute ethanol and gently mix by inversion until a DNA ppt. is visible. If the ppt. is not visible check the underside of the tube cap – sometimes the DNA lodges there!
10. “Hook” the DNA ppt. with a (“hooked”) Pasteur pipette, decant the waste from the tube, and then place the inverted pipette back into the tube.
11. Wash the DNA with a stream of (~ 2-5ml) of 70% ethanol (the ethanol will flow down the pipette and into the tube).
12. Break off the hooked end (containing the DNA) of the pipette into a 2 ml microwtube with o-ring seal screw cap. Add 1.0ml of TE buffer and store at 4°C.
13. DNA can be quantitated (UV 260/280nm) after 24 hr storage at 4°C or >2hr. at 37°C.

### Appendix 3: PCR setup protocol

1x Polymerase Chain Reaction (PCR) Protocol w/ 25% Betaine

<u>Reagent</u>	<u>Volume (uL)</u>
dH <sub>2</sub> O	9.92
KapaTaq Buffer	2
25mM Betaine	5
KapaTaq dNTPs	0.4
Forward Primer (10uM)	0.8
Reverse Primer	0.8
KapaTaq Taq Pol.	0.08
<u>10 ng/uL DNA Sample</u>	<u>1</u>
Total Volume	20

1x Polymerase Chain Reaction (PCR) Protocol w/ 25% Betaine and 5% Dimethylsulfoxide (DMSO)

<u>Reagent</u>	<u>Volume (uL)</u>
dH <sub>2</sub> O	8.92
DMSO	1
KapaTaq Buffer	2
25mM Betaine	5
KapaTaq dNTPs	0.4
Forward Primer	0.8
Reverse Primer	0.8
KapaTaq Taq Pol.	0.08
<u>10 ng/uL DNA Sample</u>	<u>1</u>
Total Volume	20

## **Appendix 4: TD54 Thermocycling Program**

94°C for 5 mins

5 cycles of:

94°C Denaturation for 30s

64°C Annealing minus 2°C per cycle for 30s

72°C Extension for 30s

30 cycles of:

94°C Denaturation for 30s

54°C Annealing for 30s

72°C Extension for 30s

72°C Final Extension for 7 mins

Hold at 4°

## **Appendix 5: Genotyping protocol**

Procedure:

1. The Rx mix consists of 0.5 ul of template\*, 0.5 ul of LIZ standard\*\*, and 9.0 ul of Hi-Di formamide.
2. Centrifuge for a few seconds at ~ 1250 rpm.
3. Denature at 95°C for 3min before loading on the analyzer.

For an “average” PCR dilute the products 1/20 (in PCR grade water).

Mix well before use.

## **Appendix 6: Microsatellite markers sequence and PCR conditions**

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D15S652</i>	FAM	F R	GCAGCACTTGGCAAATACTC CATCACTCAAGGCTCAAGGT	90,3	284-309	Yes	TD54	2.00
<i>D15S1004</i>	FAM	F R	GGCAAGACTCCATCTCAAAA GAATAAAAAGCCTGTAAACCACC	92.1	247-271	Yes	TD54	1.50
<i>D15S657</i>	FAM	F R	TCTACATTGGACAGAAATGGG GATACACATTCTGATTCATGCG	94.5	330-360	Yes	TD54	1.50
<i>D15S649</i>	FAM	F R	GGAACAGGTCCAACATCTTG CCTCATATCCCAACTCCTT	92.05	285-297	Yes	TD54	2.00
<i>D15S157</i>	FAM	F R	AGTGCATGGTGTGTC GTGAAGTCCACAGTATCTGAC	93.63	114-125	Yes	TD54	2.00
<i>D15S127</i>	FAM	F R	CCAACCACACTGGGAA AACAGTTGCCACGGT	89.1	137	Yes	TD54	1.50
<i>D7S684</i>	FAM	F R	GCTTGCAGTGAGCCGAC GATGTTGATGTAAGACTTCCAGCC	138.1	169-187	Yes	TD54	1.50
<i>D7S2513</i>	FAM	F R	GCAGCATTATCCTCAACAGC CACAAATGGCAGCCTTTC	141.05	157-181	Yes	TD54	1.50
<i>D7S661</i>	FAM	F R	TTGGCTGGCCCAGAAC TGGAGCATGACCTTGAA	143.2	252-282	Yes	TD54	1.50

<i>Markers</i>	<b>Label Dye</b>	<b>Primers</b>	<b>Genomic sequence</b>	<b>Genomic position Mb</b>	<b>Allele Size</b>	<b>Betaine (3.75M)</b>	<b>Thermo cycling File</b>	<b>MgCl2 (mM)</b>
<i>D7S2430</i>	FAM	F R	TGCAATGAGCCATGTCC GAGTGCAGTTTAATCCCCATAG	93.0	139-151	No	TD54	1.50
<i>D7S651</i>	FAM	F R	GGCTGCCTTCAAAAAGTC AGCCTGGCATGTGGAT	98.3	173-191	Yes	TD54	2.00
<i>D7S644</i>	FAM	F R	AGGTGAGAGACCCTCAGCAATAGTG CCTCCAAAGTTAGGGGTGGAAG	86.122	194-206	No	TD54	1.50
<i>D7S2202</i>	FAM	F R	TCTCTTACCCTTTGGGACCT CTTGCAGATGGCCTAATTGT	139.4	149-169	Yes	TD54	2.00
<i>D7S676</i>	FAM	F R	TGANTCTAAGCAGCCACCT GCAACATGATCCAGAAAACA	143.71	149-166	Yes	TD54	2.00
<i>D7S1827</i>	FAM	F R	CATCCATCTATCTCTGTAATCTCTC TATTTAACACACCTGTCTCAATCC	150.25	142-162	Yes	TD54	2.00
<i>D7S657</i>	FAM	F R	GTCACAGCACAGTTTTTTGG GTCAAGTAGAGATTGAGATTCC	92.64	245-264	Yes	TD54	2.00
<i>D7S821</i>	FAM	F R	ACAAAACCCCAAGTACGTGA TATGACAGGCATCTGGGAGT	95.89	238-270	Yes	TD54	2.00
<i>D7S2442</i>	FAM	F R	TGAGCCAAGATCACAGCACT CTGGAAGCAACAGATGTCACTA	147.97	226-236	No	TD54	2.00

Markers	Label Dye	Primer	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D7S1798</i>	FAM	F R	AGGTGAGTAGTATGTA AAAACACCG CACCTAGATCTAATTTGTGCTCC	145.0	251	Yes	TD54	1.50
<i>D7S495</i>	FAM	F R	TGGCATTCAATTACAATAGCC AGCACCTGGTCCAATTTTCT	137.65	150-168	Yes	TD54	1.50
<i>D6S1660</i>	FAM	F R	GAGTCTTGAGTAACTCCCACG GACAATGAGTATCCCCAC	23.42	198-210	Yes	TD54	1.50
<i>D6S291</i>	FAM	F R	CTCAGAGGATGCCATGTCTAAAATA GGGGATGACGAATTATCACTAACT	36.37	203-217	Yes	TD54	1.50
<i>D6S1680</i>	FAM	F R	AAAATTCCACCCCGC CCATCTCCCAGCAGAC	39.30	176-198	Yes	TD54	1.50
<i>D6S1619</i>	FAM	F R	GGCCTTGCTCTAACTTGCT GTCTCCTCNTCAAAGTGCTG	69.77	189-211	Yes	TD54	1.50
<i>D6S280</i>	FAM	F R	CCGGAAGTGTCCAATATCTT TCTCATTGAGATTTCTNAGTTTTA	73.80	150-164	Yes	TD54	2.00
<i>D6S406</i>	FAM	F R	CCTGGGTGACAGAGTGAGAC CACCATAGATTCTGAAGCACC	74.52	186-206	Yes	TD54	2.00
<i>D6S268</i>	FAM	F R	CTAGGTGGCAGAGCAACATA	107.86	86-100	Yes	TD54	2.00

Markers	Label Dye	Primer	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
			AAAAGGAGGTCATTTAATCG					
<i>D6S1568</i>	FAM	F R	ACATGACCAGAACTTCCCAG AGCTAGGCCAGGCCGT	34.16	84-110	Yes	TD54	2.00
<i>D6S1602</i>	FAM	F R	GATTACAGGCGTGAGCACC ACCTGAGTGGGAAATTCTGG	37.54	194-208	Yes	TD54	2.00
<i>D6S467</i>	FAM	F R	CCGAAGTCTTTGAAATGTCT TGACAAAATGAACAGAAACG	69.98	246-250	Yes	TD54	2.00
<i>D6S1596</i>	FAM	F R	GACCTCCCCAGTCACA CCATCACCTTACAGTTACCA	74.44	175-225	Yes	TD54	2.00
<i>D6S1682</i>	FAM	F R	CTCTTCAAACCTAAGGA CTGACATTCATACAATGGA	96.95	231-237	Yes	TD54	2.00
<i>D16S3139</i>	FAM	F R	TCCTGAGCAATTAGGTAAGACA CCCGTGACTGTAAAACGAC	96.95	201-209	Yes	TD54	2.00
<i>D6S273</i>	FAM	F R	GCAACTTTTCTGTCAATCCA ACCAAACCTCAAATTTTCGG	31.79	120-140	Yes	TD54	2.00
<i>D6S464</i>	FAM	F R	TGCTCCATTGCACTCC CTGATCACCTCGATATTTTAC	69.98	202-226	Yes	TD54	2.00
<i>D6S1576</i>	FAM	F R	AGAATTATTAAGGGGCAAGA CCAAAGTAGTGAATTACAGGT	27.83	186-222	Yes	TD54	2.00

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D6S439</i>	FAM	F R	GATGATTTAAGTTTCCTGTGGACC TTCAAGGACAGCCTCAGGG	35.26	272-292	Yes	TD54	2.00
<i>D6S1613</i>	FAM	F R	AGGAAGACTCCACCTCATT AGGTAACACCNACAGCAAAT	90.65	92-119	Yes	TD54	2.00
<i>D6S1645</i>	FAM	F R	CAGGAGAACCACTTGAACC CCCACCTTAGCAGACAGAGAG	149.64	226-252	Yes	TD54	2.00
<i>D6S456</i>	FAM	F R	GCATAGTTTGATTACTTCAGAACAC TAAATGGGTCTGCCCTG	76.16	232-248		TD54	1.50
<i>D6S1589</i>	FAM	F R	CCCTCCACATACAGTGAAAG ATGCTTGCTTCAGCCAAT	78.51	170-188	No	TD54	1.50
<i>D6S460</i>	FAM	F R	AATTCCTTTGAAGAAACC CAGTGGGCTTCACCC	80.40	144-166	No	TD54	1.50
<i>D6S1652</i>	FAM	F R	GAAATCCCTGGTCGGTC ATTTATGGTAGCCGAGCC	86.01	213-231	No	TD54	1.50
<i>D6S1595</i>	FAM	F R	ATATTACACCANTCCTCCCAAAGTA ACCTCTAGTTTTGCCTGCTTTTA	88.5	125-131	Yes	TD54	1.50
<i>D6S450</i>	FAM	F R	ACACATTCACCCGTTTATTTT GGAAATGAAGCACACAATG	92.07	227-239	Yes	TD54	1.50

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D6S456</i>	FAM	F R	GCATAGTTTGATTACTTCAGAACAC TAAATGGGTCTGCCCTG	76.16	232-248	Yes	TD54	1.50
<i>D6S1589</i>	FAM	F R	CCCTCCACATACAGTGAAAG ATGCTTGCTTCAGCCAAT	78.51	170-188	No	TD54	1.50
<i>D6S460</i>	FAM	F R	AATCCCATTGAAGAAACC CAGTGGGCTCTCACCC	80.40	144-166	No	TD54	1.50
<i>GAAT3A06</i>	FAM	F R	CCCATGAATGCTGAGACTTT TTGCAGTCCTTTTCAGTAAGG	21.32	177-189	No	TD54	1.50
<i>D6S1016</i>	FAM	F R	GCTTAAAAATTTAAAAGTGAGTTTCC CCTGTCAGCTAGAGAGGCAG	26.81	231-256	Yes	TD54	1.50
<i>D6S306</i>	FAM	F R	TTTACTTCTGTTGCCTTAATG TGAGAGTTTCAGTGAGCC	28.03	230-248	Yes	TD54	2.00
<i>D6S1545</i>	FAM	F R	AATCTATGCTCCTGGGTTG GAAGTTCTGGAAATACAGCCTC	25.09	223-231	Yes	TD54	2.00
<i>D6S299</i>	FAM	F R	AGGTCATTGTGCCAGG TGTCTATGTATACTCCTGAATGTCT	24.03	206-226	Yes	TD54	1.50
<i>D6S2439</i>	FAM	F R	CATTTCAAACCCTGAGTG TGGAGACAGCATGTGAATTG	24.41	218-258	Yes	TD54	1.50

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D6S1558</i>	FAM	F R	GCTACTTGGGAGGCTGGAC CTGGCAGGAGGGCTAGTG	27.13	245-259	Yes	TD54	1.50
<i>D6S394</i>	FAM	F R	GGCCAACATCTGAAACTCC GTGGTAATGGAACCTCCCA	26.24	108-338	Yes	TD54	1.50
<i>D16S310</i> 7	FAM	F R	CCAGAGTGATGGGAATA TGAGCACTGTCTCAAAAAA	66.21	286-296	Yes	TD54	1.50
<i>D16S302</i> 5	FAM	F R	TCCATTGGACTTATAACCATG AGCTGAGAGACATCTGGG	67.12	90-110	Yes	TD54	1.50
<i>D16S309</i> 5	FAM	F R	TCAGTTGGAAGATGAGTTGG TATAGTTTGTGTCCCCCGAC	68.50	134-162	Yes	TD54	1.50
<i>D16S752</i>	FAM	F R	AATTGACGGTATATCTATCTGTCTG GATTGGAGGAGGGTGATTCT	69.89	101-129	Yes	TD54	1.50
<i>D16S262</i> 4	FAM	F R	GAGACCTACAGTCTTTTGCATTTAC TTTTGAAGCTGAGCAGAAGG	70.29	143	Yes	TD54	1.50
<i>D16S310</i> 6	FAM	F R	GAGACCTACAGTCTTTTGCATTTAC TTTTGAAGCTGAGCAGAAGG	70.74	156-206	Yes	TD54	1.50
<i>D16S311</i> 5	FAM	F R	GGAGAATGGCTTTCTTGC CAACTCTATGATGGGGTTTTATTAC	73.08	242-252	Yes	TD54	1.50

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D16S3018</i>	FAM	F R	GGATAAACATAGAGCGACAGTTC AGACAGAGTCCCAGGCATT	72.73	244-270	Yes	TD54	2.00
<i>D16S3139</i>	FAM	F R	TCCTGAGCAATTAGGTAAGACA CCCGTGACTGTAAAACTGAC	71.26	201-209	Yes	TD54	2.00
<i>D16S3097</i>	FAM	F R	TGATAGCCAAAGAAGTTGGT CTTGTGGGTCAATATAGATTAATAA	75.94	194-214	Yes	TD54	1.50
<i>D16S422</i>	FAM	F R	CAGTGTAACCTGGGGGC CTTTCGATTAGTTTAGCAGAATGAG	81.46	188-212	Yes	TD54	1.50
<i>D16S2625</i>	FAM	F R	TACGCAAGTCAAAGAGCCTC GGACACATGAGACCCTGTCT	83.27	183	Yes	TD50	1.50
<i>D16S520</i>	FAM	F R	GCTTAGTCATACGAGCGG TCCACAGCCATGTAAACC	85.07	181-197	Yes	TD54	1.50
<i>D16S518</i>	FAM	F R	GGCCTTTGGCAGTCA ACCTTGGCCTCCCACC	76.69	271-290	Yes	TD54	1.50
<i>D16S413</i>	FAM	F R	ACTCCAGCCCGAGTAA GGTCACAGGTGGGTTC	86.45	131-149	Yes	TD54	1.50
<i>D16S3023</i>	FAM	F R	CTGCATTCTCATCACAGTG GAGCGCCTATGTTCGG	87.03	77-97	Yes	TD54	1.50

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D16S3026</i>	FAM	F R	CTCCCTGAGCAACAAACACC GGTCATCTATATGCGCCTGA	88.02	178-210	Yes	TD54	1.50
<i>D16S3049</i>	FAM	F R	GCAATGAAGGCAACAAAGT TTAAAAGACCTGGGGGAAT	77.47	233-255	Yes	TD54	1.50
<i>D16S3098</i>	FAM	F R	TTCCACACATAAGGTGAGTTT TTGTCTGCTTCTTACGGA	80.00	151-165	Yes	TD54	1.50
<i>D16S3040</i>	FAM	F R	TACTCCGGCAAGGACG GCTGCCTAGCACATGG	78.20	109-129	Yes	TD50	1.50
<i>D3S3641</i>	FAM	F R	TCTTTTGTCTATTAAACCTCCGTTC CCCCCATGCTCTCTTG	136.25	220-228	Yes	TD54	1.50
<i>D3S1593</i>	FAM	F R	TCAATATGGCTGGTAGCAGA ACAGTATTCTTGGTTGAAAGGTATG	146.81	137-153	Yes	TD54	1.50
<i>D3S1744</i>	FAM	F R	TTTAAGCGGAAGGAAGTGTG CTGGCCCCATCTCTCTAT	148.57	131-163	Yes	TD54	1.50
<i>D3S3694</i>	FAM	F R	AGTGTCCATCAACATGGG TCGCACAAATAACAGGATTC	143.67	124-162	Yes	TD54	2.00
<i>D3S3627</i>	FAM	F R	AAGCTAATATCAATAACAAGGC TTAAATGTAGCNTCAGATGTTAAAG	147.82	205-232	Yes	TD54	2.00

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D3S1576</i>	FAM	F R	AAGCTAATATCAATAAACAAGGC TTAAATGTAGNTCAGATGTAAAG	138.90	189-203	Yes	TD54	1.50
<i>D3S3586</i>	FAM	F R	TGGATAAATAACTGCCACC CTACACATAAGTCTGGATGAATAGG	140.47	102-114	Yes	TD54	1.50
<i>D3S1576</i>	FAM	F R	AGCTTTGGACGCAGGAAG GGCTTTTATTCATGTAGTCCTCATA	138.90	189-203	Yes	TD54	1.50
<i>D3S1292</i>	FAM	F R	TGGCTTCATCACCAGACC CAGATTC AAGAGGCACTCCA	133.11	142-166	Yes	TD54	1.50
<i>D3S3548</i>	FAM	F R	CTTCCAGGTCCAAGAGTG CAAAGGCAGCAGAATATG	132.67	217-223	Yes	TD54	1.50
<i>D9S970</i>	FAM	F R	GTCCAAGA AACTCATGAATCTC GCCACCTATAAAGGCATA	38.40	130	Yes	TD54	2.00
<i>D9S1879</i>	FAM	F R	TCAAAGAAACCCTCAGC CTAAATCAGCAACAACAGC	70.65	140-160	Yes	TD54	2.00
<i>D9S1799</i>	FAM	F R	TTGCCAACTATTTTAGCCC TGCAGTTTCAATCCACATC	72.55	139-178	Yes	TD54	2.00
<i>D9S1844</i>	FAM	F R	CATGCGAAAAACCGCTT TCCGCAGCCTCAGAGA	45.61	139-178	Yes	TD54	2.00

Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D9S166</i>	FAM	F R	AAATCATGCAATTCATTCA TCCTAATTCCTGGGAAAAC	72.37	139-178	Yes	TD54	2.00
<i>D9S301</i>	FAM	F R	AGTTTTCATAACACAAAAGAGAACA ACCTAAATGTTTCATCAAAAAGAGG	72.99	235	Yes	TD54	2.00
<i>D9S2148</i>	FAM	F R	TCAATCAACATCTGTCTATTTCATC ACATCTGGCACTCTGGAGAG	38.28	148-172	Yes	TD54	2.00
<i>D9S1862</i>	FAM	F R	CATGAGAGCACTGTATGAGGAC ACATCAGGATTGTGGGTTTC	70.33	167-207	Yes	TD54	2.00
<i>D17S1795</i>	FAM	F R	AGTGCCAGAGATATACCGTG GTCTGCAAGCAAGTTGTC	45.27	167-177	Yes	TD54	2.00
<i>D17S809</i>	FAM	F R	CAAAAAGGCAGAATGCAGTA TCCAGAGTCAAAAACACAGG	47.27	229-247	Yes	TD54	2.00
<i>D17S1607</i>	FAM	F R	CAGATAAAAAACACAAGTTTCTGAC GCTCCACCCAGACCTA	51.16	103-123	Yes	TD54	2.00
<i>D17S790</i>	FAM	F R	AAAATGAGTGGACCATACGAAGA GGGTTATTGTTTTTCTGTGTGA	50.15	187-201	Yes	TD54	2.00
<i>D17S797</i>	FAM	F	ACCTTACAGGTTGNAAAGTGC	44.90	198-204	Yes	TD54	1.50

		R	CTTCTGAGCAGTGGAGGTGG					
Markers	Label Dye	Primers	Genomic sequence	Genomic position Mb	Allele Size	Betaine (3.75M)	Thermo cycling File	MgCl2 (mM)
<i>D17S941</i>	FAM	F R	CCAAACATTGTTTCAGGTGC TATGCCAGCCGAAATCA	47.19	268-277	Yes	TD54	2.00
<i>D17S788</i>	FAM	F R	CTAGGCAGCCACTACCAAAT CAGCATCTTTGCTATAAGCATC	47.64	186-198	Yes	TD54	2.00
<i>D17S1606</i>	FAM	F R	TGGTATTCAATCCTGGAGC TGATGAGTCTTCATAGCCCC	52.68	175	No	TD54	1.50
<i>D17S1161</i>	FAM	F R	GCCAAGATAATGCCATTGCA TTCTCCCTGTGCCCTCTAA	53.68	133-175	No	TD54	1.50
<i>DIS2621</i>	FAM	F R	AAGCCTTAGTTTTACCCTATGGGAC GTCTGGCACACCTAGCAGAATG	216.41	139-163	No	TD54	1.50
<i>DIS2800</i>	FAM	F R	AACCTCTCTGGTATGAAGCC TTGTTCAAGGGTCAAATGC	232.5	178-190	No	TD54	1.50
<i>DIS2811</i>	FAM	F R	CCACTGCACTCCAACCTG GTAGTTTCTGACTGAAGG	241.175	120-164	No	TD54	1.50

## Appendix 7: List of primer sequences of genes sequenced in *OTSC4* critical region

*WWP2*, *CALB2* and *SF3B3* were sent for sequencing in Genome Quebec center

COG4						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E1	NM_015386 -F	ccattagagttcattgggcact	362	TD 54	yes (0.75)	1.5
	NM_015386 -R	ttttgtatccccagcaag				
E2	NM_015386 -F	cataagtgcaaaatgaagcattg	391	TD 54	yes (0.75)	1.5
	NM_015386 -R	cccttaggtgttcactttatcttga				
E3	NM_015386 -F	tcaaacgttcttgagaaaaagc	384	TD 54	yes (0.75)	1.5
	NM_015386 -R	gagttggggctgtagtgagc				
E4	NM_015386 -F	gggggtttgtgagattgc	384	TD 54	yes (0.75)	1.5
	NM_015386 -R	tgtgtttttattggtcgagtctg				
E5	NM_015386 -F	cacagcatctgacatgg	498	TD 54	yes (0.75)	1.5

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
	NM_015386 -R	ggcaacagggtgttgaatg				
E6	NM_015386 -F	tcaatgggaggattttgga	281	TD 54	yes (0.75)	1.5
	NM_015386 -R	catcaggctagaagggtgaga				
E7	NM_015386 -F	aagcagccaaaaggattgag	400	TD 54	yes (0.75)	1.5
	NM_015386 -R	ccagtgtcgccttagcatt				
E8	NM_015386 -F	atcttggcctgtcatcctg	375	TD 54	yes (0.75)	1.5
	NM_015386 -R	ttcatatgatacattgcactgaaaa				
E9	NM_015386 -F	gggactctctttgccctgta	300	TD 54	yes (0.75)	1.5
	NM_015386 -R	tccttaataagagatttggcctta				
E10	NM_015386 -F	aggtaggtgaaaacatggaaaa	394	TD 54	yes (0.75)	1.5
	NM_015386 -R	cctcagccatttcctcag				
E11	NM_015386 -F	aaaaggagcagagctaggg	346	TD 54	yes (0.75)	1.5
	NM_015386 -R	ggacaaaggcagaaacaaa				
E12	NM_015386 -F	ttcccctaaaagagcacct	389	TD 54	yes (0.75)	1.5
	NM_015386 -R	gacactggaggaaaagtctctagc				
E13	NM_015386 -F	tagcgcaagtctgaactcca	291	TD 54	yes (0.75)	1.5
	NM_015386 -R	caccaagataagtgtctcacca				
E14	NM_015386 -F	agctggctggaggagtagaa	286	TD 54	yes (0.75)	1.5

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
	NM_015386 -R	cacgatgagccagctaaca				
E15	NM_015386 -F	agaggggaagggatgagtcgt	237	TD 54	yes (0.75)	1.5
	NM_015386 -R	tctgttcagatggctctgct				
E16	NM_015386 -F	aggcacaggggaaggaagaat	290	TD 54	yes (0.75)	1.5
	NM_015386 -R	tcccagaaggaatgtgatg				
E17	NM_015386 -F	accctcacgttagaaactcc	300	TD 54	yes (0.75)	1.5
	NM_015386 -R	caagtctctgggggtggtg				
E18	NM_015386 -F	agtctctctctggcttctga	375	TD 54	yes (0.75)	1.5
	NM_015386 -R	tgctggcagacatggttag				
E19a	NM_015386 -F	ataggacacctccccctcc	478	TD 54	yes (0.75)	1.5
	NM_015386 -R	ATTAGGAGCCCGCTTCTCTC				
E19b	NM_015386 -F	GTCTGGCTTGGGGGAGAT	458	TD 54	yes (0.75)	1.5
	NM_015386 -R	aggctgaggtcttgggttt				
<b>DDX19A</b>						
E1	NM_018332 -F	gcccagaatgatgatgtcg	373	TD 54	yes (0.75)	1.5
	NM_018332 -R	aagagcaaatcaaggccaag				
E2	NM_018332 -F	gaacaaaccgactccctgaa	300	TD 54	yes (0.75)	1.5
	NM_018332 -R	tgcttataaacagctgaagt				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E3	NM_018332 -F	ccttgggctacaagagaaaa	299	TD 54	yes (0.75)	1.5
	NM_018332 -R	tctctgattcctctcatctctg				
E4	NM_018332 -F	aggagaaaaagatggcttca	299	TD 54	yes (0.75)	1.5
	NM_018332 -R	gtaaaattctcacaagaatctctcag				
E5	NM_018332 -F	tgacagaggtaggaggcatca	282	TD 54	yes (0.75)	1.5
	NM_018332 -R	tgttctcaggcacaaaaacg				
E6	NM_018332 -F	gggggaaatatgcctttcat	265	TD 54	yes (0.75)	1.5
	NM_018332 -R	actcctcaccagatggcact				
E7	NM_018332 -F	tgggtgccttcaggtctgt	250	TD 54	yes (0.75)	1.5
	NM_018332 -R	cgcatttcaaaattcaaacg				
E8	NM_018332 -F	gaccaagactgaggggaaca	399	TD 54	yes (0.75)	1.5
	NM_018332 -R	aggagcgtctgaagtgaag				
E9	NM_018332 -F	ctagggacttcccccaacc	489	TD 54	yes (0.75)	1.5
	NM_018332 -R	ggcatccagtctctctga				
E10	NM_018332 -F	tgacttgagcccaggacttt	371	TD 54	yes (0.75)	1.5
	NM_018332 -R	gaccatccgacttccctgta				
E11	NM_018332 -F	ttcctgcctatatgtgtgc	499	TD 54	yes (0.75)	1.5
	NM_018332 -R	gctccctgagaactccata				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E12a	NM_018332 -F	ctgggcgatagagtgagacc	538	TD 54	yes (0.75)	1.5
	NM_018332 -R	tgccaataacttaaaatgaaccaa				
E12b	NM_018332 -F	aaatgcaatgcctgtttggt	495	TD 54	yes (0.75)	1.5
	NM_018332 -R	tatgggaaaatgctggcttc				
E12c	NM_018332 -F	attccccatagtcgtgga	599	TD 54	yes (0.75)	1.5
	NM_018332 -R	GAGCACTGGCTAGCGACAT				
E12d	NM_018332 -F	ACAGCATCTTCCTCCCTCCT	499	TD 54	yes (0.75)	1.5
	NM_018332 -R	TCTCCCTCTCTGTGGACCTG				
E12e	NM_018332 -F	GCATGAGGACAGACCACAGA	397	TD 54	yes (0.75)	1.5
	NM_018332 -R	AACAAACAGCGTGAATGTGC				
<b>DDX19B</b>						
E1	NM_007242 -F	ggtctgagggaacagaatc	365	TD 54	yes (0.75)	1.5
	NM_007242 -R	cccatccagcctctaatcaa				
E2	NM_007242 -F	ggaggctctgtgattccaa	230	TD 54	yes (0.75)	1.5
	NM_007242 -R	gccctgcctgtacaaaatac				
E3	NM_007242 -F	ccccaggcctttacaaaaac	249	TD 54	yes (0.75)	1.5
	NM_007242 -R	gccaaactgggtataaaaatgaa				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E4	NM_007242 -F	tggcagttgaattagagtaaacg	363	TD 54	yes (0.75)	1.5
	NM_007242 -R	cctgggcaacagagtgagac				
E5	NM_007242 -F	tcctccagatattttgctgct	240	TD 54	yes (0.75)	1.5
	NM_007242 -R	ttgctggcctgtgaaaataa				
E6	NM_007242 -F	tgctacataatccaaggtctcc	299	TD 54	yes (0.75)	1.5
	NM_007242 -R	gcccatagctgtctttaatcc				
E7	NM_007242 -F	cagctggaggagtgctaaag	285	TD 54	yes (0.75)	1.5
	NM_007242 -R	caactattgggtgggtttga				
E8	NM_007242 -F	gagtcgagatcatgccattg	378	TD 54	yes (0.75)	1.5
	NM_007242 -R	ggaggcatctgaagtgaagg				
E9	NM_007242 -F	agctgggaaggctgtgct	391	TD 54	yes (0.75)	1.5
	NM_007242 -R	caggggatccagtcctc				
E10	NM_007242 -F	Ggatttagcaggccttcat	567	TD 54	yes (0.75)	1.5
	NM_007242 -R	agcctctgggagacacaatg				
E11	NM_007242 -F	Gactgattgcctcctgctc	440	TD 54	yes (0.75)	1.5
	NM_007242 -R	aggggagctgttcttgagat				
E12	NM_007242 -F	Tttgggctgaatgaatgat	470	TD 54	yes (0.75)	1.5
	NM_007242 -R	caggagatcctgggcaagta				

<b>DHX38</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Ampl icon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E1	NM_014003 -F	aacggaagagcttctagtcacc	637	TD 54	yes (0.75)	1.5
	NM_014003 -R	ggatttgaccagagcaagga				
E2	NM_014003 -F	tcttgagagagggagcaaa	579	TD 54	yes (0.75)	1.5
	NM_014003 -R	tctccaacaggaaaactgagaga				
E3	NM_014003 -F	agaaatgccttgggttgg	438	TD 54	yes (0.75)	1.5
	NM_014003 -R	ccetaacaaatgggttgc				
E4	NM_014003 -F	atcacattcgagcagtcag	384	TD 54	yes (0.75)	1.5
	NM_014003 -R	gtgccacctgactcagaac				
E5	NM_014003 -F	tttgggaattgactttgg	334	TD 54	yes (0.75)	1.5
	NM_014003 -R	ctgagctaagccagctctc				
E6	NM_014003 -F	CGAGATCGAGACAGgtgatg	418	TD 54	yes (0.75)	1.5
	NM_014003 -R	agcacaggaagcaggtcagt				
E7	NM_014003 -F	GGCCGATGACAGAAGACACT	379	TD 54	yes (0.75)	1.5
	NM_014003 -R	taccctcatcaaagcccagt				
E8	NM_014003 -F	ccttcttctgtggccatgt	385	TD 54	yes (0.75)	1.5
	NM_014003 -R	tcatacccactcctcttc				

Exon	Primer I.D.	Sequence	Ampl icon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E9	NM_014003 -F	tgacatctttgcagctaccag	400	TD 54	yes (0.75)	1.5
	NM_014003 -R	caatgatccaccctcctcag				
E10	NM_014003 -F	aaggcctgatgtggagacag	382	TD 54	yes (0.75)	1.5
	NM_014003 -R	tcgggcctaaggacttgata				
E11	NM_014003 -F	gaggctgaggattggaatg	353	TD 54	yes (0.75)	1.5
	NM_014003 -R	ccacaaagtcgatgggttct				
E12	NM_014003 -F	tcctgaggtcgcgatgagaacc	376	TD 54	yes (0.75)	1.5
	NM_014003 -R	gagctcctgggtaaacatgg				
E13	NM_014003 -F	ggctgttgcctccagtgaat	518	TD 54	yes (0.75)	1.5
	NM_014003 -R	CTTCAAAGCGGATGGCATAG				
E14	NM_014003 -F	gctccaccatttttgcta	473	TD 54	yes (0.75)	1.5
	NM_014003 -R	cagcagagcaagaacagcat				
E15	NM_014003 -F	ttgtcagctttggcttgg	366	TD 54	yes (0.75)	1.5
	NM_014003 -R	ggtccccaggaaatcttctc				
E16	NM_014003 -F	tctttagaggcctcgcagt	390	TD 54	yes (0.75)	1.5
	NM_014003 -R	TGCTCCACAATCTGGTCTGA				
E17	NM_014003 -F	ccttggtcacgactgtgatg	378	TD 54	yes (0.75)	1.5

	NM_014003 -R	TGGAGCctagagcaggacat				
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Ampl icon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E18	NM_014003 -F	tgggctaggaggtaggcttt	362	TD 54	yes (0.75)	1.5
	NM_014003 -R	gggttaaacaaccaagcaca				
E19	NM_014003 -F	ctgccatgtgtagcaaccag	393	TD 54	yes (0.75)	1.5
	NM_014003 -R	gaaagccactccttcttcc				
E20	NM_014003 -F	gggtggggagaagaatgag	400	TD 54	yes (0.75)	1.5
	NM_014003 -R	agcaaaccggatattgac				
E21	NM_014003 -F	CGGAGGACAACATGCTCAAC	395	TD 54	yes (0.75)	1.5
	NM_014003 -R	cagtaatacagagaagccattcc				
E22	NM_014003 -F	atgtgtggtagtgtagtg	384	TD 54	yes (0.75)	1.5
	NM_014003 -R	CTGCACCATGATGTCCTTGA				
E23	NM_014003 -F	gagtggatgagcaggat	300	TD 54	yes (0.75)	1.5
	NM_014003 -R	ccacaccaggacaaggaag				
E24	NM_014003 -F	gttctgcagatctggcttc	388	TD 54	yes (0.75)	1.5
	NM_014003 -R	tgaaaaccaacctcacatgc				
E25	NM_014003 -F	cccatgtgctcctaattgt	268	TD 54	yes (0.75)	1.5
	NM_014003 -R	ggatgaccgcaagtcttcaa				

Exon	Primer I.D.	Sequence	Ampl icon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E26	NM_014003 -F	gagaaggcggctttggtact	399	TD 54	yes (0.75)	1.5
	NM_014003 -R	ctgtcaccagtttgccctca				
E27	NM_014003 -F	ccaccatgggttagaacag	694	TD 54	yes (0.75)	1.5
	NM_014003 -R	accatgcaaaggaagtcag				
E1	NM_014500 -F	tttacatgcacacacaaaa	486	TD 54	yes (0.75)	1.5
	NM_014500 -R	AAGCCCATATAAGCGCTCTC				
E1b	NM_014500 -F	GGGCTGGCTTACACATGCT	398	TD 54	yes (0.75)	1.5
	NM_014500 -R	ggaggggtcaaatagcaaaa				
E2	NM_014500 -F	cttcgggccttgtttct	298	TD 54	yes (0.75)	1.5
	NM_014500 -R	cttgcaacccaacacgat				
E3	NM_014500 -F	agtctggcagatcagaacc	479	TD 54	yes (0.75)	1.5
	NM_014500 -R	gccctaaaatcattttctttcc				
<b>IL34</b>						
E1	NM_001172772 -F	tgggtgaagactccctct	483	TD 54	yes (0.75)	1.5
	NM_001172772 -R	ttgaaaccagacagacag				
E2a	NM_001172772 -F	ttcggacatcctgaagtcatt	386	TD 54	yes (0.75)	1.5
	NM_001172772 -R	CAGCAGCTGCCCTTATCTG				

Exon	Primer I.D.	Sequence	Ampl icon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E2b	NM_001172772 -F	GCTCCTTGGAAGGAAGACC	479	TD 54	yes (0.75)	1.5
	NM_001172772 -R	tggaatctcagagggtgc				
E3	NM_001172772 -F	gcaaatggatgatggttg	370	TD 54	yes (0.75)	1.5
	NM_001172772 -R	ttcggagaaaggcagagaaa				
E4	NM_001172772 -F	gcactccatactgggtgaca	370	TD 54	yes (0.75)	1.5
	NM_001172772 -R	gctgtggacagtggacaac				
E5	NM_001172772 -F	caggcacatgtttgccact	425	TD 54	yes (0.75)	1.5
	NM_001172772 -R	ctctgcttgcccatctt				
E6	NM_001172772 -F	tcacaggaagtctgctcttc	397	TD 54	yes (0.75)	1.5
	NM_001172772 -R	atcctggctccagaaatgg				
E7a	NM_001172772 -F	ggctacaagccattctgga	470	TD 54	yes (0.75)	1.5
	NM_001172772 -R	AGAATCCCCCTCAAAGGAAA				
E7b	NM_001172772 -F	TTTCCTTTGAGGGGGATTCT	441	TD 54	yes (0.75)	1.5
	NM_001172772 -R	ccctccacttctgcactt				
<b>PMFBP1</b>						
E1	NM_001160213 -F	atgcacctgtgcatgaag	296	TD 54	yes (0.75)	1.5

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
	NM_001160213 - R	caggatgatgcatttgaggtc				
E2	NM_001160213 -F	cttgctgagggcatagatcc	247	TD 54	yes (0.75)	1.5
	NM_001160213 - R	tggaatttgaatcataaaacatga				
E3	NM_001160213 -F	gatggacaaattccagcaa	364	TD 54	yes (0.75)	1.5
	NM_001160213 - R	tcctgaatgaacagtggaagaa				
E4	NM_001160213 -F	tggatgtgaatgactgaacaaa	426	TD 54	yes (0.75)	1.5
	NM_001160213 - R	ctgctttctagtggccttg				
E5	NM_001160213 -F	ttacaaccgcatccttctct	395	TD 54	yes (0.75)	1.5
	NM_001160213 - R	gggtaaaggttgggttcc				
E6	NM_001160213 -F	ccatgttcatggaattctgaagta	316	TD 54	yes (0.75)	1.5
	NM_001160213 - R	tttccccctcttctctcc				
E7	NM_001160213 -F	tgctgcagaaaggaacactc	280	TD 54	yes (0.75)	1.5
	NM_001160213 - R	aaaaataagtaggaaagccagaa				
E8&9	NM_001160213 -F	gtggatgctgggcatttag	543	TD 54	yes (0.75)	1.5
	NM_001160213 - R	gcaataatggtgggtgaagg				
E10	NM_001160213 -F	ggctgagacaacctctcatt	400	TD 54	yes (0.75)	1.5
	NM_001160213 - R	ggagagagaagggtgctgct				

	R					
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E11	NM_001160213 -F	gggcacaaaatgaaaggaaa	393	TD 54	yes (0.75)	1.5
	NM_001160213 -R	atgggaaatcctgtctctcc				
E12	NM_001160213 -F	ctcccctccagagagaggtt	391	TD 54	yes (0.75)	1.5
	NM_001160213 -R	ttacaggtgtgagccactgc				
E13	NM_001160213 -F	ccctgtgtgcaattgtctgt	369	TD 54	yes (0.75)	1.5
	NM_001160213 -R	gtcccctcccgtcttctctaa				
E14	NM_001160213 -F	ggtggagtggggcctttatag	297	TD 54	yes (0.75)	1.5
	NM_001160213 -R	gtgtggaggagaggtgggta				
E15	NM_001160213 -F	gcagacaggagagagccttgg	397	TD 54	yes (0.75)	1.5
	NM_001160213 -R	cccacacaagggttgagat				
E16	NM_001160213 -F	cagaggacaggaaggactg	368	TD 54	yes (0.75)	1.5
	NM_001160213 -R	cagcccaagggaatgactta				
E17	NM_001160213 -F	gctggccactgagaagaaaa	368	TD 54	yes (0.75)	1.5
	NM_001160213 -R	tggcatttgggtggtcatag				
E18	NM_001160213 -F	ttcccaggagatagaagagg	291	TD 54	yes (0.75)	1.5
	NM_001160213 -R	gcagggtgtgtgtgactaa				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E19	NM_001160213 -F	cagcagcagtcttgctgaga	272	TD 54	yes (0.75)	1.5
	NM_001160213 -R	ccccctatcaagggctaaag				
E20	NM_001160213 -F	tctgtgctgtcttcctgtg	471	TD 54	yes (0.75)	1.5
	NM_001160213 -R	tggccaagtcgctaagaaa				
E21	NM_001160213 -F	ccttctgccacctgtaa	490	TD 54	yes (0.75)	1.5
	NM_001160213 -R	ttcagaaaaacaggatgaaca				
<b>ZFH3</b>						
E0	NM_006885 -F	tgctttgcatcccaattaca	630	TD 54	yes (0.75)	1.5
	NM_006885 -R	aagagaaaagttggcgtaggg				
E1a	NM_006885 -F	ttttctcgccecgattcta	470	TD 54	yes (0.75)	1.5
	NM_006885 -R	TGTAAATGAAGCCCAACTCG				
E1b	NM_006885 -F	CGGCACGACTGTAGATGTCA	627	TD 54	yes (0.75)	1.5
	NM_006885 -R	ggaagttcagccaagactcg				
E2a	NM_006885 -F	gccctctaaccttttctt	562	TD 54	yes (0.75)	1.5
	NM_006885 -R	AGGCTCTCCACAATGTACGC				
E2b	NM_006885 -F	AGGTCACCTGCAACGAATGT	583	TD 54	yes (0.75)	1.5
	NM_006885 -R	AAGCCATCGAATTTGGACAG				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E2c	NM_006885 -F	GGCTTTCCCGAATACCTCAG	597	TD 54	yes (0.75)	1.5
	NM_006885 -R	TTCAGTACCGAGCTGGTGAG				
E2d	NM_006885 -F	ATTCGAATGGAAGGGGAGGA	584	TD 54	yes (0.75)	1.5
	NM_006885 -R	TTCGCACCATCAAAGACAAC				
E2e	NM_006885 -F	CAGCAAAAAGGACCTTGCTC	567	TD 54	yes (0.75)	1.5
	NM_006885 -R	CCGCTTTTGCAGTAGACACA				
E2f	NM_006885 -F	GACTCAAGTGCCCCAAGT	599	TD 54	yes (0.75)	1.5
	NM_006885 -R	TCTGGGCCAGGTAGTATTGG				
E2g	NM_006885 -F	GAGAAGCACATGCATAACATG A	595	TD 54	yes (0.75)	1.5
	NM_006885 -R	ttc gatgaaaccccaagtt				
E3a	NM_006885 -F	aatgtccgaggaagattg	466	TD 54	yes (0.75)	1.5
	NM_006885 -R	GTGTGTAGCGGCAGAGCTT				
E3b	NM_006885 -F	AAGTTCACGACGGACAACCT	596	TD 54	yes (0.75)	1.5
	NM_006885 -R	ggcaccagaaagagaacct				
E4	NM_006885 -F	gactgtctgtttggctgtg	473	TD 54	yes (0.75)	1.5
	NM_006885 -R	ccagctggatcagtaactct				
E5	NM_006885 -F	agagaagggggacgagtcacat	581	TD 54	yes (0.75)	1.5
	NM_006885 -R	cagegactggacctcaatag				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E6/7	NM_006885 -F	tcaccaacacaaaccctct	600	TD 54	yes (0.75)	1.5
	NM_006885 -R	cccatgttactgeatcacctt				
E8	NM_006885 -F	ctggtgaaggctaatttgct	284	TD 54	yes (0.75)	1.5
	NM_006885 -R	gaattcctgaaaaccttctcc				
E9a	NM_006885 -F	gcacttccctcatcatcaga	597	TD 54	yes (0.75)	1.5
	NM_006885 -R	AGAGTGGGGTCTCCCATTTG				
E9b	NM_006885 -F	GGCTCTGAAGAAGCACCTTG	587	TD 54	yes (0.75)	1.5
	NM_006885 -R	CATTGCTGCTGCCACTTG				
E9c	NM_006885 -F	CAGCCCAGACAACAAACCTT	564	TD 54	yes (0.75)	1.5
	NM_006885 -R	AGGAGGGTGGGGTTAAACAG				
E9d	NM_006885 -F	CAACAACAACAAGCACAAACG	585	TD 54	yes (0.75)	1.5
	NM_006885 -R	GAGCATGGAAGGCTCAGAAC				
E9e	NM_006885 -F	AAGCCAGAGAGAGAGGGACA	570	TD 54	yes (0.75)	1.5
	NM_006885 -R	GGTGCAATTGTAGGTGAGGTG				
E9f	NM_006885 -F	CAGAGCCACCACCACCTC	551	TD 54	yes (0.75)	1.5
	NM_006885 -R	CTCCAGGCTGGTGATAGGAG				
E9g	NM_006885 -F	CCCCAGTGAAGAGCAAATA	597	TD 54	yes (0.75)	1.5
	NM_006885 -R	AATCCTCATTTTGGCTGTGC				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E9h	NM_006885 -F	ATCAGGGAGAGGGCAAAGAT	577	TD 54	yes (0.75)	1.5
	NM_006885 -R	GGTAAGGGGCACTGTGGAG				
E9i	NM_006885 -F	CCGAGCAGAAGACCAACACT	554	TD 54	yes (0.75)	1.5
	NM_006885 -R	CAAACGCTTGTCTCTCTGAGG				
E9j	NM_006885 -F	TCCAACCTCCACAATGAACA	596	TD 54	yes (0.75)	1.5
	NM_006885 -R	TGGAGGAAGGGCTTAGAAGA				
E9k	NM_006885 -F	TGCGATGCTCTTAGACTGTGA	584	TD 54	yes (0.75)	1.5
	NM_006885 -R	GATCTCCGCTGTCACCAGAT				
E9l	NM_006885 -F	GCCCCGAGCTTTTATAGCA	592	TD 54	yes (0.75)	1.5
	NM_006885 -R	CTGCCAGTCCAAGGACCTC				
E9m	NM_006885 -F	CCAGCTGGACAAGGAGAAAG	341	TD 54	yes (0.75)	1.5
	NM_006885 -R	ctcagagggttggtgta				
E10a	NM_006885 -F	tctcaaactgtgtccagtg	595	TD 54	yes (0.75)	1.5
	NM_006885 -R	GAAGGAACTGGCTTGTGAGC				
E10b	NM_006885 -F	CACCATGGAGTATGCGGTAG	555	TD 54	yes (0.75)	1.5
	NM_006885 -R	TGCACCTTTGGAACAATGAA				
E10c	NM_006885 -F	CCCCAAACCAGAAGAACAGA	536	TD 54	yes (0.75)	1.5
	NM_006885 -R	GGCTGTCGTTTGTGAGC				

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E10d	NM_006885 -F	GAACAATCACGAGAGCAGCA	595	TD 54	yes (0.75)	1.5
	NM_006885 -R	CAACCCACGCTTTTTCTTTT				
E10e	NM_006885 -F	CCAGTGTAGGAACGGACACC	550	TD 54	yes (0.75)	1.5
	NM_006885 -R	TGTGAGCGATTGACCTGAGA				
E10c/d	NM_006885 -F	CACCTGAAGTCCCTCTGCTT	500	TD 54	yes (0.75)	1.5
	NM_006885 -R	GAGGTGCTGCATGAACTTGA				
<b>ZNF19</b>						
E1	NM_006961 -F	aagaggatgtctccaatcg	282	TD 54	yes (0.75)	1.5
	NM_006961 -R	cgtccccttcttgagctg				
E2	NM_006961 -F	tgccaacatagaagaaaatgc	315	TD 54	yes (0.75)	1.5
	NM_006961 -R	acgaagccgacagcctct				
E3	NM_006961 -F	tgaatgctgcttgctattgg	246	TD 54	yes (0.75)	1.5
	NM_006961 -R	tggtctaaagactgccagaa				
E4	NM_006961 -F	cctcacaggacttccctca	324	TD 54	yes (0.75)	1.5
	NM_006961 -R	tecttccactccactcc				
E5	NM_006961 -F	tcgagaaccctctctgtgg	282	TD 54	yes (0.75)	1.5
	NM_006961 -R	tgcccagcatgattcattt				
E6a	NM_006961 -F	ccttatcatttcaatgctctattatt	373	TD 54	yes (0.75)	1.5

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
	NM_006961 -R	CAAAAGGTTTCTCCCCAGTG				
E6b	NM_006961 -F	GTGGAAAAGCACCAGGACAT	477	TD 54	yes (0.75)	1.5
	NM_006961 -R	TTTCTGATGCCGAAGTAGGG				
E6c	NM_006961 -F	CAGTGGGGACAGACCCTATTAC	477	TD 54	yes (0.75)	1.5
	NM_006961 -R	TTGCTAGTCAAGGCTTTTCCA				
E6d	NM_006961 -F	CAAACTAACTCGGCACCAGA	453	TD 54	yes (0.75)	1.5
	NM_006961 -R	GCTTGAGGGATGGATCAGAG				
E6e	NM_006961 -F	AGAAGCCTGTGCTGGACATT	499	TD 54	yes (0.75)	1.5
	NM_006961 -R	AAGCCAAAACCAACCAACA				
E6f	NM_006961 -F	GGTGAGAATGAAGTCCACACA	495	TD 54	yes (0.75)	1.5
	NM_006961 -R	AATGGATTCCTGGTGCAGA				
E6g	NM_006961 -F	CTCCTCAACCCCATGATTGT	399	TD 54	yes (0.75)	1.5
	NM_006961 -R	GAAGGACACGTGGGAGAAAA				
E6h	NM_006961 -F	TGGAATGGATGCTACTCTAGGC	343	TD 54	yes (0.75)	1.5
	NM_006961 -R	tgctcattcatctgctgaca				
<b>ZNF23</b>						
E1a	NM_145911 -F	cagcgacctgctcctaga	455	TD 54	yes (0.75)	1.5

	NM_145911 -R	CCAGACCTCAGCATGTCAAA				
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Ampl icon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E1b	NM_145911 -F	TTTGTGGCTCAATAGCTTGC	500	TD 54	yes (0.75)	1.5
	NM_145911 -R	ttgctttcatgggggtatg				
E2	NM_145911 -F	gtgaggggctatgttgcctc	250	TD 54	yes (0.75)	1.5
	NM_145911 -R	aagtccagtcagcaggtg				
E3	NM_145911 -F	agagctttccaccctgtcct	240	TD 54	yes (0.75)	1.5
	NM_145911 -R	caggaccaaacgaccataca				
E4	NM_145911 -F	tctgagatctcccctgttg	389	TD 54	yes (0.75)	1.5
	NM_145911 -R	cagctctggggttctcact				
E5	NM_145911 -F	ctccccattgctgagacag	300	TD 54	yes (0.75)	1.5
	NM_145911 -R	gactggccaataaacacagtatg				
E6a	NM_145911 -F	tttaaggagcccctgatt	400	TD 54	yes (0.75)	1.5
	NM_145911 -R	CCCAATCCTGTACACCCAGA				
E6b	NM_145911 -F	AGGAAGTCCACTCAGCAGGA	316	TD 54	yes (0.75)	1.5
	NM_145911 -R	GGCTTTTCCTCAGTGTGTTCT				
E6c	NM_145911 -F	GGGAAATCCATCAGCTTTGA	346	TD 54	yes (0.75)	1.5
	NM_145911 -R	AGGCCTTCCCACACATCTT				

Exon	Primer I.D.	Sequence	Ampl icon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E6d	NM_145911 -F	TGTTCCGGAGTGTGGCAAA	454	TD 54	yes (0.75)	1.5
	NM_145911 -R	AGCTGCACCTGAAGCCTTT				
E6e	NM_145911 -F	TCAGTGTTAATGGAAGCCTAAGT	462	TD 54	yes (0.75)	1.5
	NM_145911 -R	TGTGAATTCTCTGGTGTGG				
E6f	NM_145911 -F	AGGTGCAGCTCCCAGCTTAG	361	TD 54	yes (0.75)	1.5
	NM_145911 -R	GCCGCATTAGTTTCCCATTA				
E6g	NM_145911 -F	CCTTCAGTGTCAAAGGGAAGTT	367	TD 54	yes (0.75)	1.5
	NM_145911 -R	TGCCTAGTTAGTTGGCATTG				
E6h	NM_145911 -F	CAATTCGGCAGCATCTGAG	489	TD 54	yes (0.75)	1.5
	NM_145911 -R	GCCTTTGCCACATTCCAC				
E6i	NM_145911 -F	GCCTTCAGTATCAATGCCAAA	499	TD 54	yes (0.75)	1.5
	NM_145911 -R	CCATATTCATGCCCTCTTGG				
E6j	NM_145911 -F	GCATTCAGGTTTAGCTTCCAG	398	TD 54	yes (0.75)	1.5
	NM_145911 -R	CCTTTGTAAAGTGGCCCTCA				
E6K	NM_145911 -F	CCTCCAAGAGGGCATGAATA	500	TD 54	yes (0.75)	1.5
	NM_145911 -R	ccctgtgatagttctagcttcc				

### **Appendix 8: Cycle Sequencing protocol**

<u>Reagent</u>	<u>Volume (uL)</u>
dH <sub>2</sub> O	15.68
BDT 5x Buffer	2
BDT Sequencing Mix	0.5
Primer (10uM)	0.32
<u>Purified PCR Sample</u>	<u>1</u>
Total Volume	20

### **Appendix 9: Thermocycling Protocols for Cycle Sequencing**

94°C for 1 min

25 cycles of:

96°C Denaturation for 10s

50°C Annealing for 5s

60°C Extension for 4 mins

Hold at 4°

## **Appendix 10: Cycle Sequencing DNA Precipitation Protocol**

This step was performed after Cycle sequencing was completed, as a second purification step before being placed on the ABI 3130xl or ABI 3730.

### **Step 1) DNA precipitation**

- Add 65uL of 95% Ethanol (EtOH) to each well
- Add 5uL of 125mM Ethylenediaminetetraacetic acid to each well
- Let precipitate for 15 mins to overnight in dark
  - Can place at -20°C if preferred or if not using plate for a few days.

### **Step 2) Ethanol mixture Removal**

- Place plate in centrifuge, spin at 3000 RPM for 30 mins
- Remove plate from centrifuge and decant ethanol mixture onto a dry paper towel by inverting the plate.
- Leave plate inverted on paper towel and place back in centrifuge. Spin up to 200 RPM, and then immediately stop the spinning.
- Remove and discard paper towel.

### **Step 3) Rinse step**

- Add 150uL of 70% EtOH to each well, and place in centrifuge.
- Spin plate at 3000 RPM for 5 mins.
- Remove plate from centrifuge and decant EtOH mixture onto a paper towel
- Leave plate inverted on paper towel and place back in centrifuge. Spin up to 200 RPM, and then immediately stop the spinning.

- Remove and discard paper towel. Let plate dry in dark and uncovered for 20 mins

#### **Step 4) Sample Resuspension**

- Add 15 uL of Hi-Dye Formamide (HDF) to each well
- Place plate in thermocycler on 'denat' program
  - 95°C for 2 mins
  - Hold at 4°

**Appendix 11: Sequenced functional candidate genes at the minimized 4.2Mb of *OTSC4***

Genes	Accession number	Strand	Genomic position		Function or expression
			Start	End	
<i>ZNF19</i>	NM_006961	-	71507976	71523254	<i>ZNF19</i> gene is one of zinc fingers genes that encoded zinc finger proteins. The zinc finger protein has a nucleic acid binding domain property that present in many of the transcription factor. <i>ZNF19</i> protein has a transcriptional and post-transcriptional function and it is highly expressed different tissue including temporal bone ( <a href="http://embl-ebi.org/gxa/gene/ENSG00000261611?ef=organism_part">http://embl-ebi.org/gxa/gene/ENSG00000261611?ef=organism_part</a> ).
<i>ZNF23</i>	NM_145911	-	71481503	71496117	<i>ZNF23</i> is located near <i>ZNF19</i> and has the same function as <i>ZNF19</i> .
<i>COG4</i>	NM_015386	-	70514472-70557457		<i>COG4</i> is one of the multi-protein complexes. <i>COG4</i> protein is highly expressed in different tissue including the primary osteoblast ( <a href="http://www.ebi.ac.uk/gxa/gene/MGI:2142808">http://www.ebi.ac.uk/gxa/gene/MGI:2142808</a> ).
<i>DDX19B</i>	NM_001257175	+	70333062-70367735		This gene encoded protein that is involved in many cellular processes including cellular growth and division. Dead box protein reported to be expressed in the immune system, cartilage and fibrous tissue.
<i>PMFBP1</i>	NM_001160213	-	72152996-72206349		This gene involved in the general organization of cellular cytoskeleton.
<i>DDX19A</i>	NM_018332	+	70380824-70407281		<i>DDX19A</i> is another DEAD gene. This gene encoded protein that is

				involved in many cellular processes including cellular growth and division. Dead box protein reported to be expressed in the immune system, cartilage and fibrous tissue
<i>DHX38</i>	NM_014003	+	72127615-72146811	<i>DHX38</i> is one of the DEAD box gene. This gene encoded protein that is involved in many cellular processes including cellular growth and division. Dead box protein reported to be expressed in the immune system, cartilage and fibrous tissue.
<i>ZFXH3</i>	NM_006885	-	72816786-73082274	<i>ZFXH3</i> is one of zinc finger protein. This gene encoded transcription factor. Transcription factors are important for transcription of many down stream genes. Genes that act as transcription factor or encoded a transcription factor considered candidate in this study. It also involved in cytoskeleton organization.
<i>CALB2</i>	NM_001740	+	71392616-71424342	This gene involved in the matrix formation during bone formation stage.
<i>IL34</i>	NM_001172771	+	70613798-70694585	This gene plays an important role in inflammatory process. It also involved in the regulation of osteoclast proliferation and differentiation.
<i>SF3B3</i>	NM_012426	+	70557691-70611571	This gene has a splicing function and has function in cell –cycle progression.
<i>WWP2</i>	NM_001270455	+	69796187-69944308	This gene involved in multiple processes including chondrogenesis

## Appendix 12: Sequencing variants in 12 genes across OTSC4

A) Sequencing variants detected in *ZNF19*, *ZNF23*, *COG4*, *DDX19B*, *DHX38*, *DDX19A*, and *ZNFHX3*. Colours represents corresponding colour of haplotype in figure 2.28. The disease-associated haplotype on chromosome 16 (OTSC4) is coloured yellow. Bolded variants were variants subjected for further analysis.

Gene	Location of Variant	Variant location	SNP	Affected				Controls			
				PID III-4		PID III-8		PID IV-12		PID III-9	
<b>ZNF19</b>	c.654G>C, p. Q218QH	Coding (EX6)	Rs8050872	G	C	G	G	G	G	G	C
	c.765G>C, p. T255TT	Coding (EX6)	Rs2288489	G	C	G	G	G	G	G	C
	*844A>T	Coding (EX6)	Rs34015581	A	T	A	A	A	A	A	T
<b>ZNF23</b>	c. -756G>A	Coding (EX1)	Not Listed	G	A	G	G	G	G	G	A
	c.82A>G, p. S28SG	Coding (Ex5)	rs2070832	A	G	A	A	A	A	A	G
<b>COG4</b>	c.369+38het_delT	Intron 3	Not listed	Wt	Del	DEL	Wt	Wt	DEL	Wt	DEL
	c.485C>T	Coding (EX4)	rs3931036	T	T	T	T	T	T	T	T
	544+19T>A	Intron 4	rs74324138	T	T	A	T	T	A	T	T
	c.646C>T, p. L216LL	Coding (EX5)	rs3762171	C	T	T	C	C	T	T	T
	c.1061+76T>G	Intron 8	not Listed	T	T	G	T	T	G	T	T
<b>DDX19B</b>	c.1482-25T>C	Coding (Ex12)	rs2303793	T	C	C	T	T	C	C	C
	c.296+122delT het	Intron 4	Rs35388962	Del	WT	WT	Del	Del	WT	Del	WT
	c.607-69delA het	Intron 7	Not listed	Del	WT	WT	Del	Del	WT	Del	WT
	<b>PMFBP1</b>	c.165+21C>T	Intron 3	rs3852785	C	C	C	C	T	C	C
c.2145G>A		Coding (EX5)	Not listed	G	G	G	G	A	G	A	G
c.1949C>A		Coding (Ex13)	rs34832584	C	C	C	C	A	C	A	C
c.2422-87G>T		Intron 16	rs3812987	G	T	G	G	G	G	G	T
c.2693+61A>C		Intron 18	rs4788606	C	A	A	C	A	A	C	A

<b>DHX38</b>	c.754C>G	Coding (EX1)	Not listed	C	C	G	C	C	C	G	C	C
	c.456G>GT	Coding (EX1)	Not listed	G	G	G	G	T	G	G	G	G
	c.147C>G	Coding (EX2)	rs105036 G=0.473	G	G	G	G	C	G	G	G	G
	c.418C>A	Coding (EX3)	rs1050362 A=0.432	A	A	A	A	C	A	A	A	A
	c.616+64_616+65delTC	Intron 3	rs10537374 del=0.5	Del	Del	Del	Del	Wt	del	Wt	Del	Del
	c.1308T>C	Coding ( Ex10)	Not listed	T	C	C	T	T	C	T	T	C
	c.1387-80T>C	Coding (Ex11)	rs8043606	T	C	T	T	T	T	T	C	C
	c.1698G>A	Coding (EX13)	rs2240243 A=0.42	A	A	A	A	G	A	G	A	A
	c.10999G>T	Coding (EX14)	rs12325142 T=0.49	T	T	T	T	G	T	G	T	T
	c.12071C>A	Coding (Ex16)	rs2074626 A=0.49	A	A	A	A	C	A	C	A	A
	c.2487+14T>C	Coding (Ex18)	rs42544	C	C	C	C	C	C	C	C	C
	c.2380-20T>C	Coding (EX18)	rs150617	T	C	C	T	T	C	C	T	C
	c.*45C>G	Coding (EX27)	rs7940	C	C	C	C	G	C	C	C	C
	c.*174TC>T	3UTR	rs6680 T=0.441	T	T	T	T	C	T	C	C	T

<b>DDX19A</b>	c.360C>T het	Coding (Ex5)	Rs1134074	C	T	T	T	G	C	T	A	T	C	T	T	T	T	T	T	T
	c.783-152T>C het	Coding (EX9)	Not Listed	T	T	T	T	G	C	T	A	T	C	T	T	T	T	T	T	T
	c.783-113T>C het	Coding (EX9)	Not Listed	T	T	T	T	G	C	T	A	T	C	T	T	T	T	T	T	T
	c.783-105C>T het	Coding (EX9)	Rs188463504	T	T	T	T	G	C	T	A	T	C	T	T	T	T	T	T	T
	c.783-64G>A het	Coding (EX9)	Not Listed	G	C	A	T	C	G	A	T	C	G	A	T	C	G	A	T	C
	c.783-51C>T het	Coding (EX9)	Not Listed	C	T	T	T	G	C	T	A	T	C	G	A	T	C	G	A	T
	c.783-48T>C het	Coding (EX9)	Not Listed	T	T	T	T	G	C	T	A	T	C	G	A	T	C	G	A	T
	c.783-40A>G het	Coding (EX9)	Not Listed	A	T	C	T	A	T	C	T	A	T	C	G	A	T	C	G	A
	c.783-23T>C het	Coding (EX9)	Not Listed	T	C	T	T	C	T	T	T	C	C	G	T	T	T	T	T	T
	c.783-17C>A het	Coding (EX9)	Not Listed	C	T	T	T	C	C	G	T	T	C	C	G	T	T	T	T	T
	c.876T>C het	Coding (EX9)	Rs1065167	T	T	T	T	C	C	G	T	T	C	C	G	T	T	T	T	T
	c.912T>C het	Coding (EX9)	Rs1065168	T	T	T	T	C	C	G	T	T	C	C	G	T	T	T	T	T
	c.876T>C het	Coding (EX9)	Rs1065167	T	T	T	T	C	C	G	T	T	C	C	G	T	T	T	T	T
	c.1020+12C>T	Intron 9	Rs200086133	C	C	G	T	T	C	C	G	T	T	C	C	G	T	T	T	T
	c.1020+33C>T	Intron 9	Not Listed	C	C	G	T	T	C	C	G	T	T	C	C	G	T	T	T	T
	c.*92G>A	Intron 12	Rs111798739	G	T	T	T	C	C	G	T	T	C	C	G	T	T	T	T	T
c.*265T>A	Intron 12	Rs114589283	T	T	T	T	C	C	G	T	T	C	C	G	T	T	T	T	T	
<b>ZFHX3</b>	c.185C>T, p.A62AV	Coding (EX2)	Not 1602listed	C	T	C	C	G	C	C	C	C	C	C	C	C	C	C	C	
	c.214T>G, p.S72SA	Coding (Ex2)	Not listed	G	G	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	<b>c.3663+55A&gt;C</b>	Coding (EX2)	rs9972835	A	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.1282A>C, p.T428TP	Coding (EX2)	not listed	A	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.1602C>T,p.P534PP	Coding (Ex2)	rs62640008	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.1753G>A, p.G585GS	Coding (EX2)	rs140602496	G	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.1776C>T, p.D592DD	Coding (EX2)	rs11075951	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.2330T>C, p.V777VA	Coding (EX2)	rs4788682	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.2385G>C, p.795PP	Coding (EX2)	rs10852515	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.2577C>A,p.A859AA	Coding (EX2)	not listed	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	
	c.2719+40G>A	Coding (EX2)	rs2157786	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
	c.2719+229G>A	Coding (EX2)	rs73592725	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
	c.2916G>A	Coding (EX3)	rs2228200	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
	c. 3216+28G>A	Coding (EX3)	not Listed	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
	c.3216+153G>A	Intron 3	Rs12149928	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
	<b>c.3216+28C&gt;T</b>	Intron 3	rs12149946	T	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	

<b>c.3216+153G&gt;A</b> c.4446A>T, p.A1482A c.6819G>A, p.E2273EE c.8823A>G, p.G2941G <b>c.10831C&gt;T,</b> <b>p.H3611HY</b> <b>c.10557-10558 het</b> <b>ins GGC</b>	Intron 6 Coding (EX9) Coding (EX9) Coding (EX9) Coding (EX10) Coding (EX10)	rs12149928 Rs740178 Not Listed Rs699444 <b>rs200992486</b> Not listed	<b>A</b> <b>C</b> <b>T</b> <b>G</b> <b>T</b> <b>Ins</b>	<b>G</b> <b>A</b> <b>A</b> <b>G</b> <b>C</b> <b>WT</b>	<b>G</b> <b>A</b> <b>T</b> <b>A</b> <b>C</b> <b>WT</b>	<b>A</b> <b>C</b> <b>T</b> <b>G</b> <b>T</b> <b>Ins</b>	<b>G</b> <b>A</b> <b>T</b> <b>G</b> <b>C</b> <b>Wt</b>	<b>G</b> <b>A</b> <b>T</b> <b>A</b> <b>C</b> <b>WT</b>	<b>A</b> <b>A</b> <b>T</b> <b>G</b> <b>C</b> <b>WT</b>	<b>G</b> <b>A</b> <b>A</b> <b>G</b> <b>C</b> <b>Wt</b>
--	--	--	--	---	---	--	---	---	---	---

## B) Sequencing variants detected in sequencing *CALB2*, *IL34*, *SF383* and *WWP2*.

Colours represents corresponding colour of haplotype in figure 2.28. The disease-associated haplotype on chromosome 16 (OTSC4) is coloured yellow. Bolded variants were variants subjected for further analysis.

Gene	Location of Variant	Variant location	SNP	Affected						Control	
				III-5		III-11		IV-1		IV-5	
<i>CALB2</i>	c.95-32C>T c.699+44A>G c.540+5214T>C	Intron 1 Intron 10 Intron 9	rs7203012 rs43270	C G C	T G C	C G C	C G C	C G C	T G C	C G C	
<i>IL34</i>	c.-107G>C c.-1C>G <b>c.367G&gt;C</b> <b>c.639C&gt;A</b>	Intron 2 Intron 2 Coding (EX5) Coding (EX8)	rs3813904 rs3813905 rs8046424 0.458/120 rs4985556 0.112/224	C G C A	G C G C	C G C A	G C C C	C G C A	C G C C	G C G C	
<i>SF3B3</i>	c.717A>T, c.2827-60G>A	Coding (EX6) Intron 21	rs33910368 rs7197262	A G	T A	A G	T A	A G	T A	A G	
<i>WWP2</i>	c.219-165T>G c.341-37T>G c.478+32C>T c.575+62T>G c.575+63T>G c.915-52delG c.1179+62_1179+63insTC c.1239G>A c.1522-131T>G c.1522-132C>T	Intron 4 Intron 5 Intron 6 Intron 7 Intron 7 Intron Intron 11 Coding (Ex13) Intron 14 Intron 14	Not listed rs2291959 rs2291960 Not listed Not listed Not listed Not listed rs8052727 rs78727023	T T C T T T Del InS A G C	G G C T G G WT WT A G C	T T C T T T Del ins A G C	T T C T T T WT WT G T C	T T C T T T Del Ins A G C	G G C T G G WT WT A T C	G G C T G G DEL INS A T C	

	c.1593+109A>G c.1707T>C c.1976+111T>G c.2091A>G c.2440+101A>G c.1008+31068G>A	Intron 16 Coding (EX18) Intron 19 Coding (EX19) Intron 23 Intron 24	rs2291961 rs2270841 rs2270842 rs1983016 rs2270843 -	G	G	G	A	G	G	G	G	G
				C	C	C	T	C	C	C	C	C
				G	G	G	T	G	G	G	G	G
				G	G	G	A	G	G	G	G	G
				G	A	G	A	G	A	G	G	A
				G	A	G	G	G	A	G	G	A

**Appendix 13: List of primers sequences used for genes sequenced in the critical region on chromosome Chr16q24**

*PLCG2, IRF, SLC38A8, ZDHHC7* and *SLC7A5* were sent for sequencing at Genome Quebec sequencing center.

CA5A						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E1	NM_001739 -F	atcaaagcccagtgacctga	466	TD 54	yes (0.75)	1.5
	NM_001739 -R	ggctcgggatggtctctact				
E2	NM_001739 -F	gtttcccgtgggtcagttt	397	TD 54	yes (0.75)	1.5
	NM_001739 -R	cgaagctcttgaccctgct				
E3	NM_001739 -F	ggcagagtgaatgggacaga	499	TD 54	yes (0.75)	1.5
	NM_001739 -R	ttacagcgccttctctgtag				
E4	NM_001739 -F	ccattcagaagggaagtga	465	TD 54	yes (0.75)	1.5
	NM_001739 -R	gcggatcatctaaggtcagg				
E5	NM_001739 -F	cttagatgatccgccacct	381	TD 54	yes (0.75)	1.5
	NM_001739 -R	cgcaccagctagttttgta				
E6	NM_001739 -F	gccttatctgggggaaag	400	TD 54	yes (0.75)	1.5

	NM_001739 -R	ccccaaatgcatgaaagaat				
E7	NM_001739 -F	attgatcattcgggccatc	463	TD 54	yes (0.75)	1.5
	NM_001739 -R	ctccatcccagcgttatgtt				

<i>COTLI</i>						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
1	NM_021149-Ex1F	Gccgccagagaatcaagac	498	TD 54	yes (0.75)	1.5
	NM_021149-Ex1R	GGAATAAGGCGAGGAAGACA				
2	NM_021149-Ex2F	Tgtcttctcgccttattcc	393	TD 54	yes (0.75)	1.5
	NM_021149-Ex2R	Gaggccttaaacgctcaca				
3	NM_021149-Ex3F	Aactgcaccggtctctg	398	TD 54	yes (0.75)	1.5
	NM_021149-Ex3R	Gcttgctcccaagtcacagcta				
4a	NM_021149-Ex4aF	Tcccagtggttgctga	592	TD 54	yes (0.75)	1.5
	NM_021149-Ex4aR	GGGCTCTGGAAAGCAACG				
4b	NM_021149-Ex4bF	CTCCTGCAGTGACCCCTTT	587	TD 54	yes (0.75)	1.5
	NM_021149-Ex4bR	CGTTCCCAGCTCTCCTTAAA				
4c	NM_021149-Ex4cF	GACGGAGGGAGTATTTTCAGG	545	TD 54	yes (0.75)	1.5
	NM_021149-Ex4cR	Tctgaaggc gatgatgactg				
<i>FOXF1 (NM_001451)</i>						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
1a	NM_001451-Ex1aF	atgacggcagaggTGCAG	560	TD 54	yes (0.75)	1.5
	NM_001451-Ex1aR	CCCTCCTCGAACATGAACTC				
1b	NM_001451-Ex1bF	CTCTCGCTCAACGAGTGCTT	385	TD 54	yes (0.75)	1.5
	NM_001451-Ex1bR	AGCCGCCCATGTACGAGT				
1c	NM_001451-Ex1cF	CATGATGAACGGCCACTTG	571	TD 54	yes (0.75)	1.5

	NM_001451-Ex1cR	cactcttcgaaggggagag				
2	NM_001451-Ex2F	tgctcctctgcctgaactct	575	TD 54	yes (0.75)	1.5
	NM_001451-Ex2R	TCAGCAGAATTCCTGTGTGG				
<b>FOXLI</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
1a	NM_005250-Ex1aF	ggagggaaaagcttggagtt	579	TD 54	yes (0.75)	1.5
	NM_005250-Ex1aR	TGTCGTGGTAGAAGGGGAAG				
1b	NM_005250-Ex1bF	GCCTCCCTACAGCTACATCG	514	TD 54	yes (0.75)	1.5
	NM_005250-Ex1bR	GTCACCAGCGTCCTCGTT				
1c	NM_005250-Ex1cF	GGAAGAGGAAGCCCAAGC	585	TD 54	yes (0.75)	1.5
	NM_005250-Ex1cR	GCAGGGGGAAATAAGAGAGG				
1d	NM_005250-Ex1dF	AACGAGGACGCTGGTGAC	585	TD 54	yes (0.75)	1.5
	NM_005250-Ex1dR	CCCAGGCAAAGATCATTTTA				
<b>HSD17B2</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
1	NM_002153-Ex1F	CGAGAGCGGTCAAATAGGAA	578	TD 54	yes (0.75)	1.5
	NM_002153-Ex1R	CCAGATTTGACAAAGTCCTGCT				
2	NM_002153-Ex2F	cctcaggaaccctctcttc	481	TD 54	yes (0.75)	1.5
	NM_002153-Ex2R	acacatggtaggggtcagc				
3	NM_002153-Ex3F	ctetaatccccaggagacc	386	TD 54	yes (0.75)	1.5
	NM_002153-Ex3R	ttgttccccacatagaatg				

4	NM_002153-Ex4F	ttgtctgcccaagtcactg	374	TD 54	yes (0.75)	1.5
	NM_002153-Ex4R	tggtgtgctcatctttgac				
5	NM_002153-Ex5F	ccttccaacagagacaagc	618	TD 54	yes (0.75)	1.5
	NM_002153-Ex5R	attctccatgaggctgttgc				
<b>OSGIN1</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
1	NM_182981-Ex1F	CTCTCTGCAGCCCCCTCTAAC	599	TD 54	yes (0.75)	1.5
	NM_182981-Ex1R	aatgtgcagctccaacacac				
2	NM_182981-Ex2F	gcacatacactggcacacg	371	TD 54	yes (0.75)	1.5
	NM_182981-Ex2R	gatgtgtaagaactgcatgg				
3	NM_182981-Ex3F	gtttggagcagaggctgtgt	495	TD 54	yes (0.75)	1.5
	NM_182981-Ex3R	ttgcagtgaaccaagattgc				
4	NM_182981-Ex4F	cagtgtctggcacaggctaa	389	TD 54	yes (0.75)	1.5
	NM_182981-Ex4R	tccagcaaacacacacttttg				
5	NM_182981-Ex5F	cctccagtgccagggaat	395	TD 54	yes (0.75)	1.5
	NM_182981-Ex5R	aaagggagatgccagagtcc				
6	NM_182981-Ex6F	ggcatctcccttctcatt	353	TD 54	yes (0.75)	1.5
	NM_182981-Ex6R	cgtcctcaacctggaaaaa				
7a	NM_182981-Ex7aF	ctgacactgccttgacat	596	TD 54	yes (0.75)	1.5
	NM_182981-Ex7aR	CATCTTGGGCAGCTGGTT				

7b	NM_182981- Ex7bF	CTACAACATCCCGGTGATCC	497	TD 54	yes (0.75)	1.5
	NM_182981- Ex7bR	CCTTCCTTAGCAGGGAGCTG				
7c	NM_182981- Ex7cF	AAGAGGAACCCCATTGACG	580	TD 54	yes (0.75)	1.5
	NM_182981- Ex7cR	Aatcccccttctcccttc				

## Appendix 14: pipeline report of NGS

# Exome Sequencing Pipeline

Genome Quebec and McGill Innovation Center

September 2012

[bioinformatics.genome@mail.mcgill.ca](mailto:bioinformatics.genome@mail.mcgill.ca)

This document contains the description of the current exome-seq analysis pipeline. The information presented here reflects the current state of the pipeline as of September 2012.

Program	Version	Reference
Genome Analysis Toolkit		<a href="http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit">http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit</a>
BWA		<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
igvtools		<a href="http://www.broadinstitute.org/igv/igvtools">http://www.broadinstitute.org/igv/igvtools</a>
samtools		<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
fastX		<a href="http://hannonlab.cshl.edu/fastx_toolkit/">http://hannonlab.cshl.edu/fastx_toolkit/</a>
picard		<a href="http://picard.sourceforge.net/">http://picard.sourceforge.net/</a>
snpSift/snpEff		<a href="http://snpeff.sourceforge.net/SnpSift.html">http://snpeff.sourceforge.net/SnpSift.html</a>

## Deliverables

The following files are delivered to the client:

1. A sample statistics file containing the metrics enumerated in Step 6.
2. A .csv file containing all the variants found in at least one sample, often classified by chromosome because of large file sizes
- 3.** A .csv file containing all the high impact coding variants found in at least one sample, based on the Step 9 annotations

## Sequencing

### Reads

Around 50 to 150 million 100 b.p. paired-end reads from the Illumina HiSeq 2000 sequencer. Base quality is encoded in phred 33.

### Pipeline steps

The pipeline is executed on Compute Canada clusters via unix bash commands, perl scripts and open source software.

#### Step 1: Read trimming and clipping of adapters

Reads are trimmed from the 3' end to have a phred score of at least 30. Illumina sequencing adapters are removed from the reads, and all reads are required to have a length of at least 32 b.p. Trimming and clipping are done with the fastx software [1].

#### Step 2: Aligning the reads to the genome reference

The filtered reads are aligned to a reference genome. The 1000 genome reference is currently used for human samples, and the mm9 reference is currently used for the mouse samples. The 1000 genome reference is similar to the hg19 reference, except that it uses the newer mitochondrial build (NC\_012920). Also chromosome nomenclature is slightly different ('1' vs 'chr1', non-chromosomal supercontigs are named differently e.g. GL000207 ). The alignment is done per lane of sequencing, and then merged for a complete **B**inary **A**lignment **M**ap file (.bam). The alignment software used is bwa [2], and the merging is done with the picard software <sup>210</sup>.

#### Step 3: Realignment insertions and deletions (INDELs)

Insertion and deletion realignment is performed on regions where multiple base mismatches are preferred over indels by the aligner since it can appear to be less costly by the algorithm. Such regions will introduce false positive variant calls which may be filtered out by realigning those regions properly. Realignment is done with the GATK software <sup>19</sup>.

#### **Step 4: Fixing the read mates**

Once local regions are realigned, the read mate coordinates of the aligned reads need to be recalculated since the reads are realigned at positions that differ from their original alignment. Fixing the read mate positions is done with picard software <sup>210</sup>.

#### **Step 5: Marking duplicates**

Aligned reads are duplicates if they have the same 5' alignment positions (for both mates in the case of paired-end reads). All but the best pair (based on alignment score) will be marked as a duplicate in the .bam file. Marking duplicates is done with picard software <sup>210</sup>.

#### **Step 6: Compute metrics and generating coverage track**

Multiple metrics are computed at this stage and given in the statistics file:

- Number of raw reads
- Number of filtered reads (after Step 1)
- Number of aligned reads (after Step 2)
- Number of duplicate reads (after Step 5)
- Duplicate rate (number of duplicate reads / number of raw reads. Good run max of 25%)
- Median, mean and standard deviation of insert sizes of reads after alignment
- Mean coverage over exons (mean number of reads per base position)
- Percentage of bases covered at X reads (%\_bases\_above\_50 means the % of exons bases which have at least 50 reads. A good run is typically around 50%)

A TDF (.tdf) coverage track is also generated at this step for easy visualization of coverage in the IGV browser <sup>19</sup>.

#### **Step 7: Variant calling**

Variants (SNPs and INDELS) are called using samtools mpileup and bcftools varfilter <sup>19</sup>. The following options are given to mpileup to filter for low quality variants which could introduce false positive calls: -L 1000 -E -q 1 -u -D -S, where:

-L INT           max per-sample depth for INDEL calling [250]  
 -E               extended BAQ for higher sensitivity but lower specificity  
 -q INT           skip alignments with mapQ smaller than INT <sup>210</sup>  
 -u               generate uncompress BCF output  
 -D               output per-sample DP in BCF (require -g/-u)  
 -S               output per-sample strand bias P-value in BCF (require -g/-u)

The output of mpileup is then fed to varfilter, which does an additional filtering of the variants and transforms the output into the VCF (.vcf) format. The arguments used are: -d 2 -D 1200 -Q 15 -l 0.0, where:

-d INT           minimum read depth [2]  
 -D INT           maximum read depth [10000000]  
 -Q INT           minimum RMS mapping quality for SNPs <sup>19</sup>  
 -l FLOAT         min P-value for strand bias (given PV4) [0.0001]

The final .vcf files are filtered for long 'N' INDELS which are sometimes introduced and causing excessive memory usage by downstream tools.

### **Step 8: Mappability annotation**

An in-house database identifies regions in which reads are confidently mapped to the reference genome. Generally, low mappability corresponds nicely to RepeatMasker regions and *increases substantially with read length*. A region is identified as HC = coverage too high (>400), LC = low coverage (<50), MQ = to low mean mapQ (<20) and ND = no data at the position

### **Step 9: dbSNP annotation**

The .vcf files are annotated for dbSNP using the software SnpSift [7]<sup>211</sup>.

### **Step 10: Variant effect annotation**

The .vcf files are annotated for variant effects using the SnpEff software [7]. SnpEff annotates and predicts the effects of variants on genes (such as amino acid changes).

### **Step 11: Additional SVN annotations**

Provides extra information about SVN by using numerous published databases.

1. Biomart : adds GO annotations based on gene information

2. dbNSFP: an integrated database of functional annotations from multiple sources for the comprehensive collection of human non-synonymous SNPs. It compiles prediction scores from four prediction algorithms (SIFT, Polyphen2, LRT and Mutation Taster), three conservation scores (PhyloP, GERP++ and SiPhy) and other function annotations.
3. Cosmic (Catalogue of Somatic Mutations in Cancer): Annotates SVNs which are known somatic mutations

### Step 10: Merging the .vcf files

This step merges the sample .vcf files into a global .csv file. The following columns are reported:

Column	Description
<b>chromosome</b>	chromosome number
<b>position</b>	position on chromosome
<b>ref_allele</b>	reference allele
<b>alt_alleles</b>	alternativ allele
<b>mappability</b>	Mappability flag from simulated dataset (LC = low coverage, HC = high coverage, NODATA = no data, empty = normal mappability)
<b>gene_name</b>	gene name
<b>Nb_Samples</b>	number of samples with this variant
<b>Frac_of_Samples</b>	number of samples with this variant / total number of samples
<b>Avg_Depth</b>	mean coverage of variant over all samples that have the variant
<b>Frac_of_Alt</b>	mean fraction of reads which have the alternate allele over all samples
<b>1000_genome_global_maf</b>	minor allele frequency score (-1 = no sample in the 1000 genome database has this variant, 1=all samples in the 1000 genome database has this variant)
<b>gene_description</b>	gene description generated by biomart. Derived from Ensembl
<b>go_ids</b>	all GO ids associated with gene
<b>go_terms</b>	all GO terms associated with gene
<b>impact</b>	severity of the change
<b>effect</b>	description of change
<b>effect_type</b>	type of change
<b>codon_change</b>	exact nucleotide change
<b>amino_acid_change</b>	exact amino acid change

<b>gene_type</b>	type of transcript from annotation file
<b>transcript</b>	ENSEMBL transcript ID
<b>db_snp_id</b>	snp id in the db_snp database ( <a href="http://www.ncbi.nlm.nih.gov/projects/SNP/">http://www.ncbi.nlm.nih.gov/projects/SNP/</a> )
<b>1000genome_population_mafs</b>	1000 genome MAF divided by population
<b>CosmicID</b>	cosmic ID and cancer type information from control database of somatic mutations
<b>InterPro_Domain</b>	description of domain
<b>Uniprot</b>	Uniprot gene ID
<b>GERP++_neutral_rate</b>	Neutral rates of evolution (sum of branch lengths)
<b>GERP++_RS_score</b>	RS score, the larger the score, the more conserved the site.
<b>29way_SiPhy_score</b>	SiPhy score based on 29 mammals genomes. The larger the score, the more conserved the site.
<b>Polyphen2</b>	Prediction based on HumVar, 'D' ('probably damaging'), 'P' ('possibly damaging') and 'B' ('benign') (separated by ',')
<b>SIFT_score</b>	If a score is smaller than 0.05 the corresponding NS is predicted as 'D(amaging)' otherwise it is predicted as 'T(olerated)'
<b>Normal HomRef</b>	Number of samples homozygous for reference allele
<b>Normal HomAlt</b>	Number of samples homozygous for alternative allele
<b>Normal Het</b>	Number of samples heterozygous
<b>"SampleName" Depth</b>	Depth of Coverage from sample from DP field
<b>"SampleName"</b>	INFO field from vcf file (GT:PL:DP:SP:GQ) GT: Genotype, PL: Genotype likelihood (RR, RA, AA), DP: Depth of Coverage, SP: Strand bias, GQ: Genotype Quality

**Appendix 15: Sequenced functional candidate genes at the extended newly linked 9.6 Mb region**

Genes	Accession Number	St	Genomic position	Exons	Function or expression
			Start End		
<i>PLCG2</i>	NM_002661	+	81812930 81991899	33	This gene encodes by protein which is responsible for conversion of 1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate to 1D-myo-inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG) which is important for transmitting signals from growth factors receptors and immune system receptors a cross the cell membranes
<i>HSD17B2</i>	NM_002153	+	82660339 82132139	5	It is a single -pass type membrane protein and it is expressed in the bone and bone marrow.
<i>OSGIN1</i>	NM_182980	+	83982672 83999937	7	It regulates the differentiation and proliferation of normal cells through apoptosis. Also it regulate inflammatory and anti-inflammatory molecules. Absence of this gene and it is protein leads to uncontrolled proliferation and growth.
<i>SLC38A8</i>	NM_001080442	-	84043389 84075762	10	It is a multi-pass membrane protein and is widely expressed in tissue including immune system, connective tissue and ear.
<i>COTL1</i>	NM_021149	+	84599204 84651609	4	This gene encodes one of the numerous actin-binding proteins which regulate actin cytoskeleton.

<i>IRF8</i>	NM_002163	+	85932774 85956211	9	It plays a negative regulatory role in cells of the immune system, and is expressed in connective tissue.
<i>ZDHHC7</i>	NM_001145548	-	85008067 85045141	9	It is one of the zinc finger protein . it is expressed in connective tissue and bone.
<i>FOXF1</i>	NM_001451	+	86544133 86548070	1	This gene belongs to the forkhead family of transcription factors. Function of this gene has not been determined yet. it is abundantly expressed in the connective tissue.
<i>FOXC2</i>	NM_001453	+	81610681 81614129		This gene belongs to the forkhead family of transcription factors. Function of this gene has not been determined and it has been shown a role in regulation of embryonic and ocular development. Mutation in this gene has been identified in various glaucoma phenotypes. it is abundantly expressed in the connective tissue
<i>FOXL1</i>	NM_005250	+	86612115 86615304	1	This gene belongs to the forkhead family of transcription factors. Function of this gene has not been determined yet. it is abundantly expressed in the connective tissue.
<i>SLC7A5</i>	NM_003486	-	87863629 87903100	10	It is a multi-pass membrane protein which is abundantly expressed in bone marrow and inflammatory immune systeme.
<i>CA5A</i>	NM_001739	+	87921625 87970112	7	Carbonic anhydrases are a large family of zinc metalloenzyme. They have an important role in a variety of biological processes including calcification and bone resorption .

## Appendix 16: Variants detected from exome / Sanger sequencing of genes in 9.7 Mb

### A. Variants detected from the sequencing of genes in the newly linked 9.7Mb region.

Variants present in the three affected and the control subject are bolded. Variants present in only two affected are highlighted yellow. Variants present in one affected and one unaffected subject are highlighted brown. Variants present in only unaffected are highlighted blue. Variants that were exclusively in affected are highlighted light blue. Green haplotype was constructed among the three affected siblings (highlighted green).

Gene name	Variants	Affected			Unaffected
		III-5	III-11	IV-5	III-6
CA5A EX1	c. -26C>T	<b>T</b> T	<b>T</b> T	<b>T</b> T	T T
	c.-4C>T	<b>T</b> T	<b>T</b> T	<b>T</b> T	T T
	c.(11G>A	<b>A</b> A	<b>A</b> A	<b>A</b> A	A A
	c.41-43 delTCT	<b>Del</b> del	<b>Del</b> del	<b>Del</b> del	Del del
	c.44T>G	<b>G</b> G	<b>G</b> G	<b>G</b> G	G G
	c.120A>G	<b>G</b> G	<b>G</b> G	<b>G</b> G	G G
	c.121T>C	<b>C</b> C	<b>C</b> C	<b>C</b> C	C C
	c.125A>G	<b>G</b> G	<b>G</b> G	<b>G</b> G	G G
	c.134A>G	<b>G</b> G	<b>G</b> G	<b>G</b> G	G G
	c.142+23 T>C	<b>C</b> C	<b>C</b> C	<b>C</b> C	C C
	142+26delG	<b>Del</b> del	<b>Del</b> del	<b>Del</b> del	Del del
	142+31G>A	<b>A</b> A	<b>A</b> A	<b>A</b> A	A A
	142+65G>A	<b>A</b> A	<b>A</b> A	<b>A</b> A	A A
	142+69G>T	<b>T</b> T	<b>T</b> T	<b>T</b> T	T T
	142+77G>T	<b>T</b> T	<b>T</b> /T	<b>T</b> /T	T/T

	142+85G>A	A/A	A/A	A/A	A/A
EX2	143-9T>A	A/A	A/A	A/A	A/T
EX4	c. 555+61 het delT	Del/wt	Del/wt	Del/wt	Del/wt
COTL1 Ex1	c.-136T>C	C/C	C/C	C/C	C/C
Ex3	c.*108T>G	G/G	G/G	G/G	G/G
	c.*109T>G	G/G	G/G	G/G	T/G
EX4	c.*104A>G	A/G	A/G	A/G	A/G
FOXL1	976_990Het delGGGATC CCCTTCCT C	Del/wt	Del/wt	Del wt	Wt/wt
HSD17B2 Ex1	-133T>C	T/C	T/C	T/T	T/T
IRF8 Ex4	c.432C>T,p. D144DD	C/T	C/T	C/T	C/C
OSGIN1 Ex1	c.-95C>T	T/T	T/T	T/T	T/T
	c.17C>T,p.P6 L	T/T	T/T	T/T	T/T
Ex2	C.101A>C,p. N34T	C/C	C/C	C/C	C/C
	c.169A>C,p. M67L	C/C	C/C	C/C	C/C
	c.175C>G,p. L59V	G/G	G/G	G/G	G/G
Ex3	c.218- 104C>T	C/C	C/C	C T	C T
	c.316+115T> C	T/T	T/T	T C	T C

	316+116T>G	T/T	T/T	T/G	T/G
Ex4	<b>c.317-80G&gt;A</b>	A/A	A/A	A/A	A/A
EX5	c.454-48C>A	C/C	C/C	C/A	C/A
	c.454-21C>T	C/C	C/C	C/T	C/T
Ex6	c.646-38G>A	G/G	G/G	G/A	G/A
	<b>c.646-17G&gt;T</b>	T/T	T/T	T/T	T/T
Ex7	c.1332C>A, p.L444L	C/C	C/C	C/C	C/A
	<b>c.*205T&gt;C</b>	C/T	C/T	C/T	T/C
	c.*41A>G	A/A	A/A	A/G	A/G
<b>PLCG2</b>	-109C>G	C/C	C/C	C/C	C/G
Ex2	<b>c.174T&gt;C, p.A58AA</b>	C/C	C/C	C/C	C/C
EX3	c.194-92A>C	A/A	A/A	A/C	A/C
	c.[297A>G,p. L9	A/A	A/A	A/G	A/G
	c.337+45G>C	G/G	G/G	G/C	G/C
	c.3314-23C/A	C/A	C/A	C/C	C/C
EX4	<b>c.431+37G&gt; C</b>	C/C	C/C	C/C	C/C
EX6	c.480-56A>G	A/G	A/G	A/A	A/A
	<b>c.480-22A&gt;G</b>	G/G	G/G	G/G	G/G
	c.564+87G>A	A/G	A/G	A/A	A/A
	c.648+134C> T	C/C	C/C	C/C	C/T
	c.648+171T> A	T/T	T/T	T/T	T/A
	<b>c.648+237A&gt;</b>	T/T	T/T	T/T	T/A

	<b>T</b>				
	c.648+242C> <b>T</b>	<b>T/T</b>	<b>T/T</b>	<b>T/T</b>	T/C
	c.648+319T> <b>C</b>	<b>C/C</b>	<b>C/C</b>	<b>C/C</b>	C/T
Ex8	c.692+25C>T	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	C/T
	c.692+61A>C	<b>A/A</b>	<b>A/A</b>	<b>A/A</b>	A/ <b>C</b>
<b>Ex10</b>	<b>802C&gt;T,p.R2</b> <b>68RW</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	C/C
Ex12	<b>987-97G&gt;C</b>	<b>C/G</b>	<b>C/G</b>	<b>C/G</b>	G/G
	c.1149C>T,p. <b>D383D</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	C/T
	c.1188C>G,p. T396TT	<b>C/C</b>	<b>C/C</b>	<b>C/C</b>	C/ <b>G</b>
Ex15	c.1467+38G> <b>C</b>	<b>C/C</b>	<b>C/C</b>	<b>C/C</b>	G/C
	c.147+45G>T	<b>C/G</b>	<b>G/G</b>	<b>G/G</b>	G/ <b>T</b>
Ex16	c.1497C>T],p .A499A	<b>T/T</b>	<b>T/T</b>	<b>T/T</b>	T/T
<b>Ex19</b>	<b>c.2054+7G&gt;</b> <b>A</b>	<b>A/G</b>	<b>A/G</b>	<b>A/G</b>	G/G
	c.[2054+47A >C	<b>A/A</b>	<b>A/A</b>	A/A	A/ <b>C</b>
Ex20	c.[2055-8T>C	<b>C/C</b>	<b>C/C</b>	<b>C/C</b>	C/C
Ex21	c.[2236- 14C>G	<b>G/G</b>	<b>G/G</b>	<b>G/G</b>	G/G
Ex28	c.[3093T>C,p .N1031NN	<b>C/C</b>	<b>C/C</b>	<b>C/C</b>	T/ <b>C</b>
Ex30	c.[3314- 23C>A	<b>A/A</b>	<b>A/A</b>	<b>A/A</b>	A/A
<b>SLC7A5</b>	c.345C>A,p.	<b>C/A</b>	<b>C/A</b>	<b>C/A</b>	C/A

Ex1b	<b>G115GG</b>				
	c.770+125G> A	G/A	G/A	G/G	G/G
	c.770+127C> A	C/A	C/A	C/A	C/A
Ex5	c.939+61T>G	G/T	G/T	G/G	G/G
	c.1140+85C> G	C/G	C/G	C/C	C/C
<b>SLC38A8</b>	c.89+11G>T	T/G	T/G	T/T	T/T
Ex1					
	c.[388+128A >T	A/A	A/A	A/A	A/T
Ex4	c.[531- 137T>C	C/T	C/T	C/C	C/C
Ex6	c.805+53_805 +54het_delA A	Wt/del	Wt/del	Wt/del	Wt/del
Ex8	c.[1162+87C >G	G/C	G/C	G/G	G/C
Ex9	c.[1214+107G >A	G/A	G/A	G/G	G/G
	c.[1214+116C >T	C/T	C/T	C/C	C/C
	c.[1214+174T >C	T/C	T/C	T/T	T/T
	c.[1214+194A >G	A/G	A/G	A/A	A/A
<b>ZDHHC7</b>	c.[315+55G> A	A/A	A/A	A/G	G/G
Ex3					
Ex5	c.316- 366het_delT	Wt/wt	Wt/wt	Wt/del	del wt

Ex7	<b>c.538-11C&gt;T</b>	<b>T/T</b>	<b>T/T</b>	<b>T/C</b>	C/C
Ex9	c.[906T>C],p .G302G	<b>C/C</b>	<b>C/C</b>	<b>C/T</b>	T/C
<b>HSD17B2</b> Ex1	<b>c.[-133T&gt;C</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	T/T

B: Variants detected from exome sequencing of 79 genes in the newly linked 9.7 Mb region.

gene_name	position	effect_type	codon_change	amino_acid	db_snp_id	PolyPhe	SIFT_score	III-8	III-13	III-4	III-11	III-16	control1	control2
WVVOX	78466437	MISSENSE	Cca/Gca	P282A	rs3764340	P	0.38			C/G				
MAF	79633691	MISSENSE	Gtg/Atg	V37M	.	B,B	0.21					G/A		
<b>DYNLRB2</b>	<b>80583497</b>	<b>SILENT</b>	<b>Ctg/Ttg</b>	<b>L66</b>	<b>rs11866734</b>	.	.	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>		
ATMIN	81074904	Start_Gained			rs12444747							ins/wt		
ATMIN	81076821	MISSENSE	Tcc/Ccc	S240P	rs2278022	B	0.06					T/C		
ATMIN	81077204	SILENT	ccT/ccA	P211	rs16954513							T/A		
GCSH	81124275	SILENT	ttC/ttT	F53	rs8177876	.	.					C/T		
PKD1L2	81232294	MISSENSE	Att/Gtt	I506V	rs61734110	B,B	0.39	A/G	A/G	A/G	A/G			
<b>PKD1L2</b>	<b>81242107</b>	<b>MISSENSE</b>	<b>cAg/cGg</b>	<b>Q250R</b>	.	<b>B,B</b>	<b>0.18</b>	<b>A/G</b>	<b>A/G</b>	<b>A/G</b>	<b>A/G</b>	<b>A/G</b>		
<b>PKD1L2</b>	<b>81242198</b>	<b>NONSENSE</b>	<b>Cag/Tag</b>	<b>Q220*</b>	<b>rs7499011</b>	.	<b>0.04</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>		
BCMO1	81295780	SILENT	acC/acA	T121	rs35683292							C/A		
BCMO1	81295902	MISSENSE	aTg/aCg	M162T	rs7500996			T/C	T/C	T/C	T/C			
PLCG2	81914554	SILENT	Ctg/Ttg	L230	.	.	.					C/T		
<b>PLCG2</b>	<b>81922813</b>	<b>MISSENSE</b>	<b>Cgg/Tgg</b>	<b>R268W</b>	<b>rs17537869</b>	<b>P</b>	<b>0.02</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>		
CDH13	83704419	MISSENSE	Gct/Act	A337T	rs35549391	B,B,B	0.66		G/A		G/A			
NECAB2	84012104	MISSENSE	tTa/tCa	L20S	rs925331				T/C		T/C			
SLC38A8	84050209	SILENT	ctC/ctT	L359	rs77876966				C/T		C/T			
MBTPS1	84115393	SILENT	ctC/ctT	L469	rs12933523							C/T		
KCNQ4	84270476	MISSENSE	Cgg/Tgg	R206W	rs11646443	B,B	0.17	C/T		C/T				
<b>ATP2C2</b>	<b>84474484</b>	<b>Misense</b>	<b>Ggt/Agt</b>	<b>rs2303853</b>	<b>B,B,B,B</b>	<b>B,B,B,B</b>	<b>0.19</b>	<b>G/A</b>	<b>G/A</b>	<b>G/A</b>	<b>G/A</b>	<b>G/A</b>		
ATP2C2	84438827	MISSENSE	Gtg/Atg	V102M	rs78887288	B,B,B	0.11					G/A		
ATP2C2	84485677	MISSENSE	cTg/cAg	L604Q	rs62640926	B,B,B,B	0.98		17 T/A		T/A	T/A		
<b>KLHL36</b>	<b>84691433</b>	<b>SILENT</b>	<b>gtC/gtT</b>	<b>V340</b>	<b>rs72797514</b>			<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>		
KLHL36	84695280	SILENT	ctA/ctG	L401	rs17755815							A/G		
KLHL36	84695490	SILENT	caC/caT	H471	rs12928590	.	.					C/T		
USP10	84778685	MISSENSE	Atg/Gtg	M200V	rs1862792	B,B	0.33		A/G		A/G			
USP10	84796603	SILENT	caA/caG	Q33	rs12932018	.	.	A/G		A/G				
CRISPLD2	84900645	MISSENSE	atG/atA	M127I	.	B,B,B	0.08		G/A		G/A			
CRISPLD2	84914234	SILENT	tAa/tGa	*450	rs113031416				A/G		A/G			
FAM92B	85141661	SILENT	taC/taT	Y100	rs17200833	.	.	C/T		C/T				
KIAA0182	85695279	MISSENSE	gCc/gTc	A619V	.	.	.		C/T		C/T			
<b>IRF8</b>	<b>85945249</b>	<b>SILENT</b>	<b>gaC/gaT</b>	<b>D144</b>	<b>rs16939945</b>	.	.	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>	<b>C/T</b>		
IRF8	85947779	Stoped_Gain			rs903202							ins/wt		
<b>MTHFSD</b>	<b>86565826</b>	<b>MISSENSE</b>	<b>Ggg/Cgg</b>	<b>G152R</b>	<b>rs3751803</b>	<b>B,B,B</b>	<b>0.75</b>	<b>G/C</b>	<b>G/C</b>	<b>G/C</b>	<b>G/C</b>	<b>G/C</b>		
<b>FOX1</b>	<b>86613301</b>		<b>gggatcccccttc</b>	<b>GIPFL326-</b>	.	.	.	<b>del/wt</b>	<b>del/wt</b>	<b>del/wt</b>	<b>del/wt</b>	<b>del/wt</b>		
ZCCHC14	87445270	SILENT	acA/acG	T882	rs3748399			A/G		A/G		A/G		
ZCCHC14	87493737	MISSENSE	Atc/Gtc	I54V	rs11648852	B	0.56	A/G		A/G		A/G		
ACO10536.1	87735527	SILENT	ccC/ccT	P198	rs17199037				C/T		C/T			
CASA	87925534	SILENT	ttC/ttT	F215	.	.	.	C/T		C/T		C/T		
CASA	87938398	SILENT	ccC/ccT	P151	rs7186698									

## **Appendix 17: Functional studies to identify the effect of the FOXL1 deletion on the RNA and protein levels**

**6.1 Aim of the study:** To identify the downstream effect of the 15 bp deletion in FOXL1 gene.

### **6.2 Material and Methods**

#### **6.2.1 Test expression of wild (normal) and mutant forms of FOXL1 gene from patients lymphoblastoid cell lines**

To measure the effect of the *FOXL1* deletion on the mRNA. First, total RNA was extracted from Epstein-Barr virus-transformed lymphocytes cell line from four deletion carriers (PID III-4, III-5, III-8 and III-11), and two non-deletion NL control subjects (EC09, TA09). Lymphoblastoid cell lines were maintained in a 37 C , 5% CO<sub>2</sub>, humidified chamber, with growth media consisting of Roswell Park Memorial Institute media (RPMI), 10% heat inactivated fetal calf serum (FCS) , 2 mM L-glutamine and antibiotic-antimycotic mixture (100 units/ml penicillin G sodium, 100 µg/ml streptomycin sulfate, and 0.25 µg/ml amphotericin B).

RNA was isolated using Trizol reagent (Invitrogen by Life Technologies, Carlsbad, CA) and was followed by treatment with TURBO DNA-free DNase treatment (Ambion, Austin, TX). RNA(s) were evaluated and quantified using a 2100 Bioanalyzer (Agilent Technologies; Waldbronn, Germany) and samples with a RNA value greater than 8.5 were used. Complementary DNA (cDNA) synthesis was performed using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and Reverse

transcriptase PCR was carried using primers surrounding the c. 976\_990 deletion in the *FOXL1* gene. Amplification reaction was carried out for all the reactions by the ABI PCR GeneAmp 9700 thermocycler. Primers were amplified using a touchdown 54 program (Appendix 3 and 4) and the size fragmented on 1% agarose gel stained with SYBER safe to confirm amplification.

## **6.2.2 FOXL1 deletion generation and transfection**

### **6.2.2.1 Generation of FOXL1c.976\_990het\_delGGGATCCCCTTCCTC deletion construct**

The *FOXL1* gene was not expressed in the lymphoblastoid cell line, the wild and mutant forms of the *FOXL1* gene were overexpressed. Two *FOXL1* expression vector construct (ORF; NM\_005250) (Catalogue number: EX-E0843-M02), (Catalogue number: EX-E0843-M29) containing wild type full-length *FOXL1* and two empty (mock) plasmids (Catalogue number: EX-NEG-M02), (Catalogue number: EX-NEG-M29) were purchased from GeneCopoeia company, USA. Two of the purchased plasmid (EX-E0843-M29) and (EX-NEG-M29) were tagged with the green fluorescent protein (GFP). Mutant constructs of *FOXL1* (*FOXL1* c.976\_990het\_delGGGATCCCCTTCCTC) were generated by site directed mutagenesis using the *FOXL1* expression vectors (EX-E0843-M02), (EX-E0843-M29-GFP) as a template. *FOXL1* deletion construction was generated and sequenced by Noroclone Company (biotech laboratory, Canada).

### 6.2.2.2 Cell Culture and treatment

Human fetal osteoblast (hFOB 1.19) cell lines were used for transfection and human embryonic kidney cell lines (HEK293A) were used as a positive control. Both cell lines were purchased from American Type Culture Collection (ATCC, Biocompare, Canada). HEK293A cells were maintained in a 37 C, 5% CO<sub>2</sub>, humidified chamber, with growth media consisting of Modified Dulbecco's Media (IMDM), 10% heat inactivated fetal calf serum (FCS, Hyclone,), 2 mM L-glutamine and antibiotic-antimycotic mixture (100 units/ml penicillin G sodium, 100 µg/ml streptomycin sulfate, and 0.25 µg/ml amphotericin B). hFOB 1.19 cells maintained in a 37 C, 5% CO<sub>2</sub>, humidified chamber, with growth media consisting of Modified Dulbecco's Media (IMDM) without phenol red, 10% heat inactivated fetal calf serum (FCS), 2 mM L-glutamine, antibiotic-antimycotic mixture (100 units/ml penicillin G sodium, 100 µg/ml streptomycin sulfate, and 0.25 µg/ml amphotericin B) and geneticin antibiotic (G418) was added to a final concentration of 0.3 mg/ml.

### 6.2.2.3 Transient Transfection

hFOB1.19 cells were transfected each time with one of the following six plasmids  $FOXL1^{-WT}$  (wild type),  $FOXL1^{-WT-GFP}$ , (wild type- tagged with GFP),  $FOXL1^{-Mut}$  (mutant type),  $FOXL1^{-Mut-GFP}$  (mutant type- tagged with GFP),  $FOXL1^{-Mock}$  (empty plasmid), and  $FOXL1^{-Mock-GFP}$  (empty plasmid-tagged with GFP (MOCK-GFP) vectors. Transfection of hFO1.19 cells was performed using FuGENE ® HD transfection reagent (Roche, Applied Science, USA) according to the manufacturer's instruction. Plasmids harbouring the wild and mutant forms of *FOXL1* were diluted with Opti-MEM® /Reduced serum media

(Invitrogen by Life technology) to a concentration of 0.02µg /µl. Fugene HD transfection reagent was added to the diluted plasmids in the ratio of 7:2 (Fugene HD in µl: Plasmid DNA in µg) and left for 20 minutes incubation at room temperature. Following incubation, 100ul of this mixture was used to transfect the cells in a 6-well plate and incubated for a further 48 hours at 37C.

### **6.2.3 Test RNA expression of wild and mutant forms of FOXL1 transfected hFOB1.19 cell lines and FOXL1 downstream genes**

#### **6.2.3.1 Real-time PCR (Q-PCR)**

To test the expression of the transfected *FOXL1* genes (wild type versus mutant type), total RNA(s) were extracted and cDNA(s) were prepared as mentioned above. Primers for the real time PCR were designed using primer3 as follows: left primer: CCTCCCTACAGCTACATCGC, right primer: TGTCGTGGTAGAAGGGGAA and hybrid probe: GGTCACGCTCAACGGCATCTA.

In order to test expression of the *FOXL1* downstream genes, primers for the following genes; *IL1A* (Catalogue number: Hs00174092\_m1), *CXCL8* (Catalogue number: Hs00174103\_m1), *CXCL10* (Catalogue number: Hs01124251\_g1), *IL29* (Catalogue number: Hs00601677\_g1), *IFNB1* (Catalogue number: Hs01077958\_s1), *IFIT1* (Catalogue number: Hs01911452\_s1), *FEN1* (Catalogue number: Hs00748727\_s1) and *SP4* (Catalogue number: Hs00162095\_m1) were designed and purchased from Applied Biosystem. Housekeeping gene (GAPDH) (Catalogue number: Hs03929097\_g1) was used as endogenous control. RNA negative control sample and reference sample (HEK293A cell line) were included in each experiment. Amplification reactions and

calculation were carried out for all the reactions as described in Appendix 18. Each experiment was repeated three times for validation and the mean of the three experiments was calculated.

## **6.2.4 Testing protein expression of wild and mutant forms of FOXL1 transfected hFOB1.19 cell lines**

### **6.2.4.1 Antibodies**

FOXL1 protein expression was determined using Anti-FOXL1 rabbit polyclonal IgG (Catalogue number: clone ab83000, Abcam).  $\alpha$  tubulin housekeeping proteins were detected with  $\alpha$  tubulin antibody of 200  $\mu$ g/ml concentration (Catalogue number: clone DM1A+DM1B, Abcam) and anti-nuclear matrix protein p84 antibody of 1  $\mu$ g/ml concentration (Catalogue number: clone 5E10, Abcam). Secondary antibodies used in this experiment were horseradish peroxidase conjugated mice and rabbit IgG antibodies. Secondary antibodies were purchased from Jackson Immunoresearch.

### **6.2.4.2 Western blot (Immunoblotting)**

Nuclear and cytoplasmic proteins were extracted from four hFOB1 cell lines (three hFOB1 cell lines were transfected with *FOXL1*<sup>-WT</sup>, *FOXL1*<sup>-Mut</sup>, and an *FOXL1*<sup>-Mock</sup> plasmids respectively and one hFOB1 cell line was left non-transfected as a negative control) by nuclear extract kit (ActiveMotif, USA) according to the manufacturer's protocol. Protein concentration was quantified using a Bradford protein assay kit (BioRad, USA). 0.2M of mercaptoethanol was added to the protein samples and then samples were boiled for 5 minutes. Ten micrograms of protein were loaded per lane of 8% Sodium Dodecyl Sulfate polyacrylamide (SDS-PAGE) gel. Following electrophoresis,

proteins were transferred onto nitrocellulose membranes and blocked with 5% milk powder prepared in TBS-Tween solution (5M NaCl, 1M Tris pH 7.4, 0.5% Tween 20) for 1 hour. Optimally diluted primary antibodies were added and incubated at 4 C overnight. Membranes were washed, and horseradish peroxidase conjugated secondary antibodies were used to detect primary  $\alpha$  tubulin antibody binding and rabbit IgG antibody was used to detect primary FOXL1 rabbit polyclonal IgG antibody. Signals were amplified using chemoluminescence detection agent (Millipore). Immunoreactivity was visualized and quantified by scanning densitometry using Image Quant LAS 4000 and Image GE software respectively (GE Healthcare, USA).

#### **6.2.4.3 Immunofluorescence**

To visualize the protein fluorescent intensity of the living hFOB1.19 cells transfected by, *FOXL1*<sup>WT-GFP</sup>, *FOXL1*<sup>Mut-GFP</sup>, and *FOXL1*<sup>Mock-GFP</sup>, transfected living hFOB1.19 cells were grown in 6-well plates and visualized by a fluorescent microscope (Carl Zeiss AxioObserver A.1) with settings at X20 magnification objective. Images were captured using a Zeiss AxioCam MRM3 camera with Zeiss AxioVision 4.8 software. To measure the fluorescent intensity, flow-cytometry was carried out. In flow-cytometry, GFP-transfected cells were harvested by trypsin, followed by fixation in 1.0% formaldehyde (Sigma). 10,000 cells were analyzed using a FACS Calibur flow cytometer (Becton-Dickinson, Franklin Lakes, NJ).

## **6.2.5 Testing the effect of the FOXL1 deletion on the downstream genes**

### **6.2.5.1 Microarray Analysis**

Microarray analysis was carried out by Genome Quebec using Illumina platform. Total RNA (s) were extracted from four hFOB1 cell lines (three hFOB1 cell lines were transfected with *FOXL1*<sup>-WT</sup>, *FOXL1*<sup>-Mut</sup>, and *FOXL1*<sup>-Mock</sup> plasmids respectively and one hFOB1 cell line was left non-transfected as a negative control) and quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, Inc.). RNA(s) integrity was assessed using a 2100 Bioanalyzer (Agilent Technologies). Double stranded cDNA was synthesized from 250 ng of total RNA, and in-vitro transcription was performed to produce biotin-labelled cRNA using Illumina® TotalPrep RNA Amplification Kit, according to manufacturer's instructions (Life Technologies). The biotin-labeled cRNA was hybridized on Human HT-12 v4 Expression Bead Chip and incubated in an Illumina Hybridization oven at 58°C for 14 to 20 hours at a rocking speed of 5 according to Illumina's Whole-Genome Gene Expression Direct Hybridization Assay Guide. All four samples were hybridized in triplicate. Chemoluminescence detection and image acquisition were performed using Illumina iScan Reader. Human HT-12 v4 Expression BeadChip covers greater than 31,000 annotated genes with more than 47,000 probes derived from the National Center for Biotechnology information Reference Sequence (NCBI).

### 6.2.5.2 Microarray Statistical analysis

Microarray results were analyzed by Genome Quebec using LIMMA package<sup>212</sup> from the Bioconductor project. The expression data of two color microarrays were compared using LIMMA package and expressed in a log ratio. The neqc<sup>213</sup> method was applied to the raw intensity values, which performs background adjustments using negative control probes, and performs quantile normalization followed by a log<sub>2</sub> transformation. Normalization was carried out to adjust microarray data for any effects that could be arise from variation in the technology rather than from biological differences between the RNA samples or between the printed probes. In Log<sub>2</sub> transformed data, Value of 2 equal to ratio of 4. A log<sub>2</sub> transformation helps in identifying doubling or halving in ratios.

The statistical significance of the fold change of each transcript between the four cell lines was determined by Student's t-test using log<sub>2</sub> transformed data. Storey's q-value false discovery rate (FDR) method was used to detect expected proportions of false positives. Significantly differentially expressed genes were determined using false discovery rate FDR of <0.05% and a fold change equal or greater than 2 fold. Genes reaching statistical significance were submitted to Panther for functional analysis.

## 6.3 Result

### 6.3.1 RNA expression by real time PCR

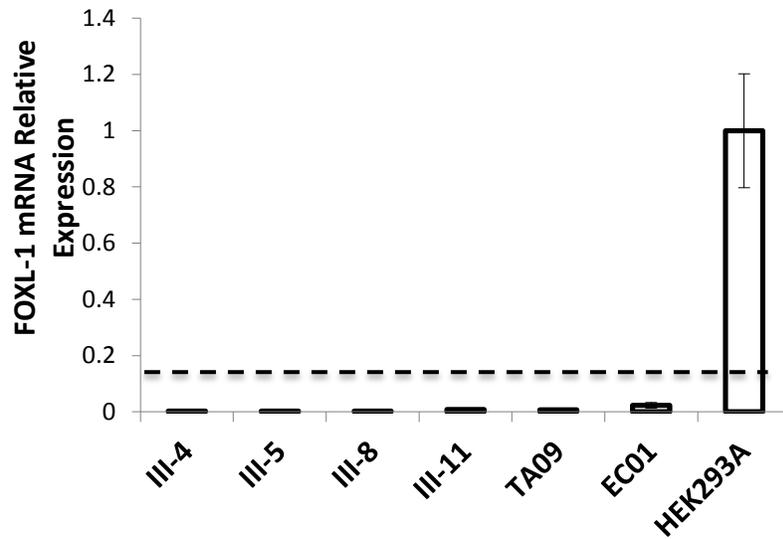
In order to test the effect of the *FOXL1* deletion on the RNA expression, levels of *FOXL1* mRNA expression were tested in four-deletion carriers (PID III-4, III-5, III-8 and III-11) and two non-deletion carriers (EC09 and TA09), using RNA extracted from lymphoblastoid cells of the tested subjects and RNA from HEK293A, as a positive control. Comparing the level of the *FOXL1* expression between the samples and the positive control showed that level of the RNA expression in the samples was less than the cut of value (0.2) (Figure 6.1). This data implied that *FOXL1* is not normally expressed in lymphocytes.

Because of the lack of *FOXL1* expression in the lymphoblastoid cells, an artificial bone cell line model system was created: *FOXL1*<sup>-WT</sup> (hFOB1.19 cells transiently transfected with the wild type *FOXL1* expression plasmid), *FOXL1*<sup>-Mut</sup> (hFOB1.19 cells transiently transfected with the *FOXL1* harboring the 15bp deletion expression plasmid), *FOXL1*<sup>-Mock</sup> (hFOB1.19 cells transiently transfected with the empty expression plasmid). *FOXL1* mRNA was relatively quantified by Q-PCR from the three-bone cell line models and untransfected hFOB1.19 cell line and compared. The hFOB1.19 cells normally expressed *FOXL1* as shown in figure 3.9 and transfection of the empty vector cause slight increase in the expression but it was not significant. Transfection of the hFOB1.19 cell line with the wild type *FOXL1* showed significant increase in the expression of the *FOXL1* and no significant difference was detected between *FOXL1*<sup>-WT</sup> and *FOXL1*<sup>-Mut</sup>

(Figure 6.2), suggesting that the *FOXLI* deletion did not alter the *FOXLI* mRNA expression.

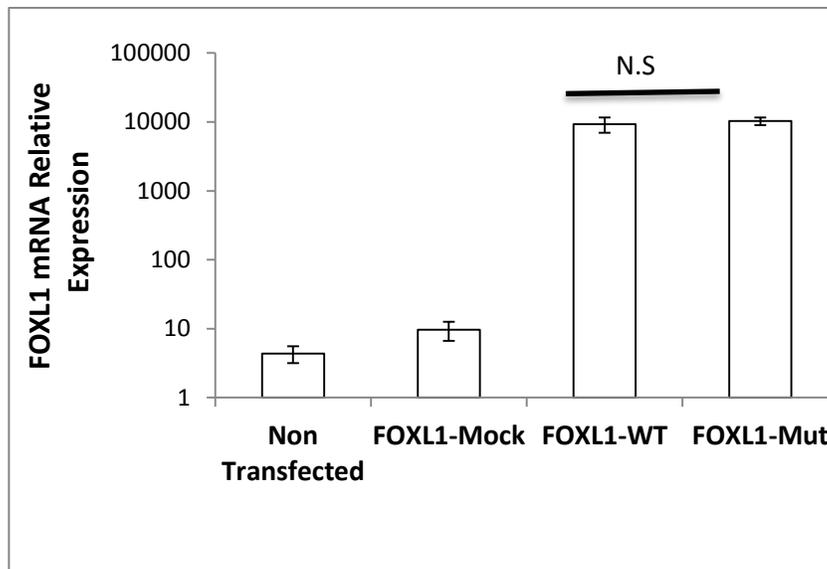
**Figure 6.1: FOXL1 is not normally expressed in the lymphoblast**

*FOXLI* mRNA was relatively quantified by real time PCR using Taqman gene expression assay. GAPDH was used as an endogenous control and the data were expressed relative to a control HEK293A cell line. Bar graphs represent the mean  $\pm$  SD of three replicate assays



**FIGURE 6.2: *FOXL1* mRNA EXPRESSION WAS NOT AFFECTED BY THE 15BP DELETION**

hFOB 1.19 cell line was either non-transfected or transiently transfected with mock, wild, or mutant types *FOXL1* expression plasmids. *FOXL1* mRNA was relatively quantified by real time PCR using Taqman gene expression assay. GAPDH was used as an endogenous control and the data was expressed relative to a HEK293A cell line. Bar graphs represent the mean  $\pm$  SD of three independent experiments. (\*\*\*)= $p$  value  $<0.001$ ). Note that *FOXL1* was originally expressed in the hFOB 1.19 cells and transient transfection of the *FOXL1* resulted in 1000-fold increase in mRNA expression of *FOXL1*.

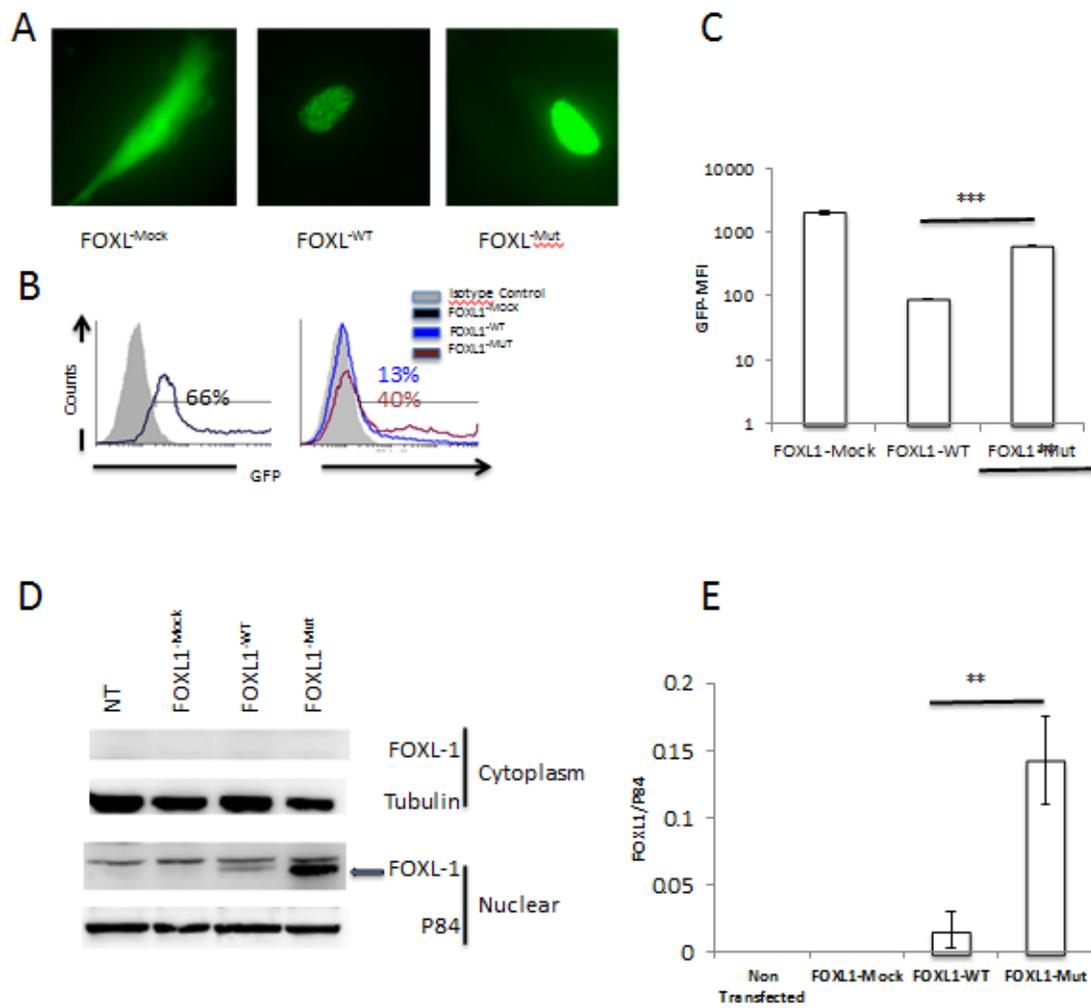


### 6.3.2 Protein expression by Western blot and immunofluorescent

Creating FOXL1 cell line models and measuring the *FOXL1* expression did not show any difference in the gene expression between the wild and mutant form of the *FOXL1* transfected cell line. Mutations in other members of the FOX family of transcription factors have been reported to be associated with cellular mislocalization and protein aggregation<sup>214,215</sup>. Thus, we hypothesized that the identified *FOXL1* deletion may affect the folding of the FOXL1 protein and subsequently hide the nuclear localization signals (NLS). To test this hypothesis, localization and protein level expression of FOXL1<sup>-WT</sup> and FOXL1<sup>-Mut</sup> were tested and we found that both FOXL1<sup>-Mut</sup> and FOXL1<sup>-WT</sup> have the same diffuse nuclear localization, but the staining intensity was higher in FOXL1<sup>-Mut</sup> (Figure 6.3 A). This finding was further confirmed by flow-cytometry through quantification of the mean fluorescence intensity (Figure 6.3 B-C). To overcome the argument that protein function can be modified by GFP-fusion<sup>216,217 218</sup>, nuclear and cytoplasmic protein extracts were prepared from the artificial bone cell lines model system (discussed above), followed by immunoblotting and probing with anti-FOXL1 antibody. Expression of the FOXL1 protein was mainly in the nuclear extract (Figure 6.3 D) and the *FOXL1* deletion resulted in a significant increase of FOXL1 protein expression (Figure 6.3 D-E). These findings suggest that mutation in *FOXL1* may result in a gain of protein function.

**Figure 6.3: mutant FOXL1 localized appropriately**

hFOB 1.19 cell line was either transiently transfected with mock, wild, or mutant type of FOXL1-GFP fusion proteins expression plasmids. A) Living cells were examined by direct fluorescence microscope using equal exposure time. B) Cells were harvested by trypsinization and fluorescence intensity was measured by flow cytometry. 40% of hFOB1.19 cell line transfected by the mutant FOXL1 were shifted to the left comparing to the hFOB transfected by the wild FOXL1. C) bar graphs represents mean fluorescence intensity (MFI)  $\pm$  SD of three independent experiments. D) hFOB 1.19 cell line was either non transfected or transiently transfected with mock, wild, or mutant types FOXL1 expression plasmids followed by cytoplasmic and nuclear extract preparation and immunoblotted using FOXL1 antibody (ab83000).  $\alpha$ -Tubulin (DM1A+DM1B) and P84 (5E10) were used as cytoplasmic and nuclear loading control respectively. E) FOXL1 levels were normalized to P84 and the average band intensity after normalization is presented in the bar graph. Error bars represent the  $\pm$  SD of three independent experiments. (\*\*=p value <0.01).



### 6.3.3 Test expression of the downstream genes using microarray analysis

Protein expression measurement showed that the FOXL1 protein expression was increased in the hFOB1.19 cells transfected by the mutant form of *FOXL1* gene which could be as result of cellular misfolding of the FOXL1 protein and protein aggregation. If our hypothesis that the 15 bp deletion in *FOXL1* gene causes misfolding of the FOXL1 protein is correct, then that means that *FOXL1* mutant could fail in binding and regulation of the downstream genes as transcription factor and that lead to deregulation of the gene expression of these genes.

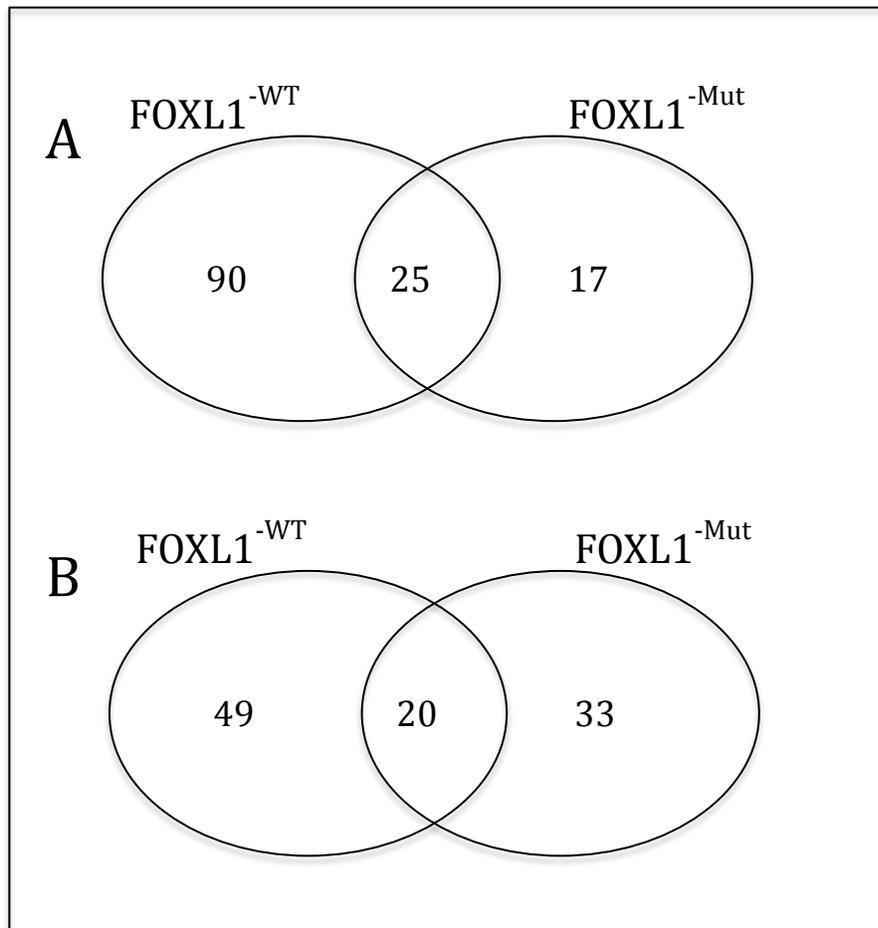
To test this hypothesis, microarray analysis was carried out to identify genes deregulated by this *FOXL1* deletion through testing the difference in the expression level of 31,000 annotated genes, between *FOXL1*<sup>-WT</sup> and *FOXL1*<sup>-Mut</sup> hFOB1.19 transfected cells and *FOXL1*<sup>-Mock</sup> cells was used as a negative control. Twelve RNA samples were analyzed: three paired RNA samples from each of *FOXL1*<sup>-WT</sup>, *FOXL1*<sup>-Mut</sup> and *FOXL1*<sup>-Mock</sup> transfected hFOB1.19 cells and untransfected hFOB1.19. Comparison of gene expression between *FOXL1*<sup>-WT</sup> and *FOXL1*<sup>-Mut</sup> cells was carried out using the false discovery rate (FDR) <0.05 and two fold increase as a significant cutoff. Using this cutoff, seventeen genes were up regulated in the *FOXL1*<sup>-Mut</sup> cells and 33 genes were down regulated in the *FOXL1*<sup>-Mut</sup> cells (Figure 6.4 A and 6.4 B and appendices 19 and 20). These genes were only deregulated in *FOXL1*<sup>-Mut</sup> cells and not in *FOXL1*<sup>-WT</sup> or *FOXL1*<sup>-MOCK</sup> which reflects the actual effect of the deletion. These results demonstrated that *FOXL1*-induced genes are differentially expressed between *FOXL1*<sup>-WT</sup> and *FOXL1*<sup>-Mut</sup> cells. The top 25

deregulated genes in *FOXL1*<sup>-Mut</sup> were selected and their differential expression was illustrated in a heat map (Figure 6.5 A-B).

Gene ontology (GO) analysis using the PANTHER classification system was performed to determine the biological process associated with these top 25 deregulated genes. The top 25 deregulated genes were involved in pathways related to inflammation, immunity, gene regulation and Eicasonoid signaling (Figure 6.5 C-D).

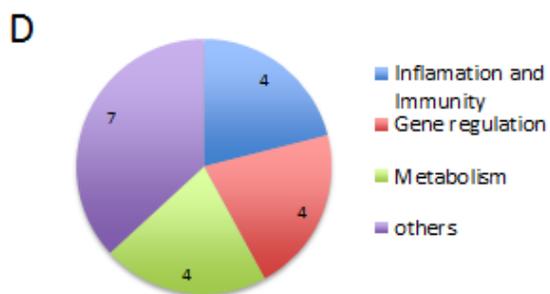
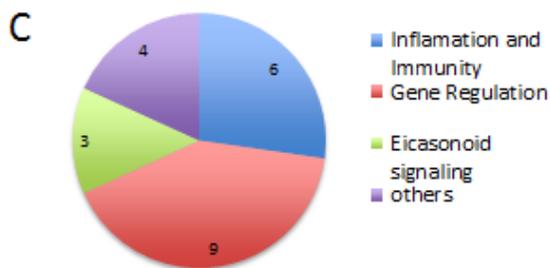
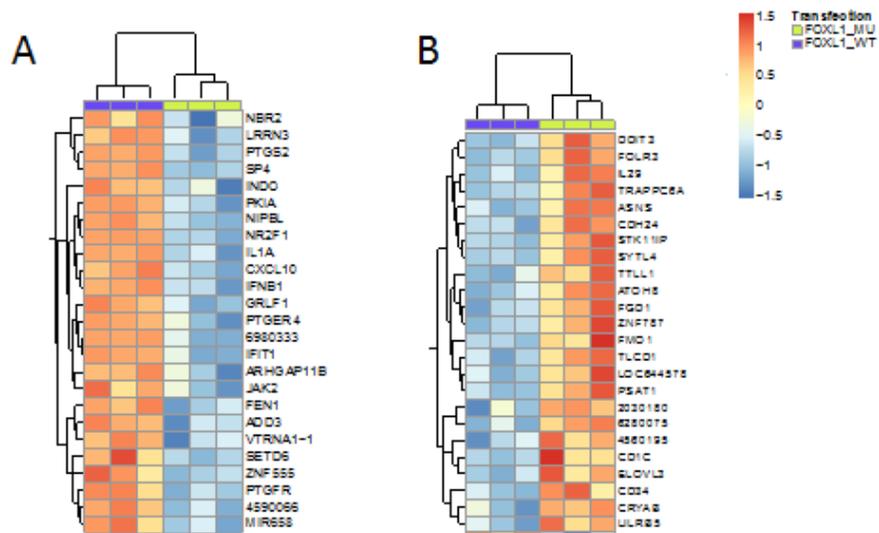
**Figure 6.4: Microarray analysis showing genes induced in FOXL1-WT and FOXL1-Mut cells.**

A) Venn diagram comparing genes up regulated by different forms of *FOXL1* transfection. First, we compared expression genes up regulated by *FOXL1* in  $FOXL1^{-WT}$  and  $FOXL1^{-Mut}$ . Ninety genes were up regulated in  $FOXL1^{-WT}$  cells while 17 genes were up regulated in  $FOXL1^{-Mut}$  cells. Twenty-five genes were up regulated commonly in both  $FOXL1^{-WT}$  and  $FOXL1^{-Mut}$  cells. B) Venn diagram comparing genes down regulated by different forms of *FOXL1* transfection. Shown are the numbers of genes significantly down regulated (FDR <0.05). Forty-nine genes were down regulated in  $FOXL1^{-WT}$  cells, thirty-three genes were down regulated in  $FOXL1^{-Mut}$  cells and 20 genes were commonly down regulated in both  $FOXL1^{-WT}$  and  $FOXL1^{-Mut}$ .



**Figure 6.5: Gene ontology analysis of genes differentially expressed in FOXL1-WT and FOXL1-Mut cells**

Heat map representation of the top twenty-five genes. A) Twenty-five genes down regulated in *FOXL1*<sup>-Mut</sup> cells. B) Twenty five genes up regulated in *FOXL1*<sup>-Mut</sup> cells. The lowest gene expression is shown in blue and the highest are shown in orange. Cut-off value =0. Differentially expressed genes were analysed using PANTHER classification system. C) Pathways down regulated in *FOXL1*-Mut cells. D) Pathways up regulated in *FOXL1*-Mut cells, each number value reflects the number of genes in each pathway. Of 25 down-regulated genes, only 22 genes were predicted to be involved in known pathways. The remaining three genes are of unknown function. The same was for the 25 upregulated genes, only 19 genes were predicted and the remaining were of unknown function.

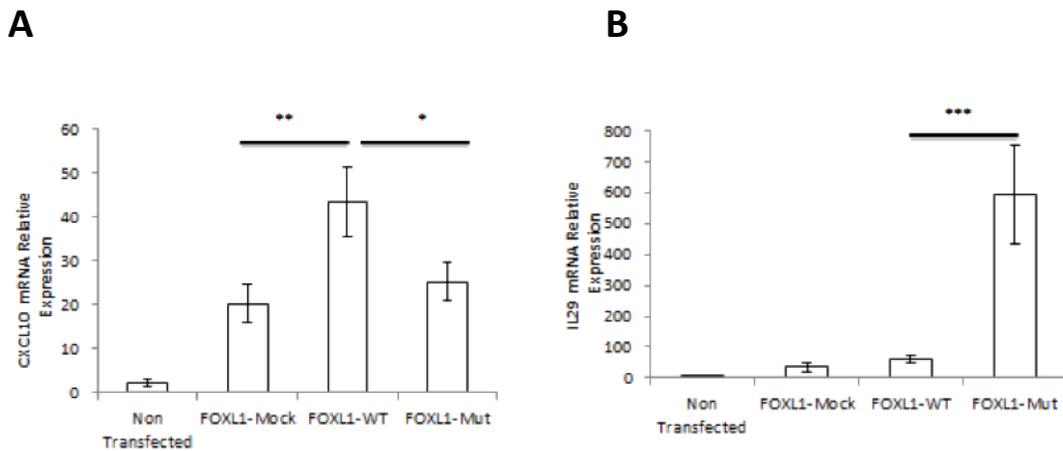


### 6.3.4 Validation of the microarray result by qPCR

We have identified many cytokines deregulated in the *FOXLI*<sup>-Mut</sup> transfected cells. Of these cytokines, there were cytokines reported to have a significant role in bone development and remodelling including, for example: *IL1A*<sup>166</sup>, *IL8*, *CCXLI0*<sup>167 168 169 170</sup>, *IFNBI*<sup>171</sup>, *IL29*<sup>172</sup> and *IFITI*<sup>173</sup> (Appendix 17). Quantitative RT-PCR (RT-qPCR) was conducted to validate the results of the microarray analysis of these six cytokines genes and 2 regulatory genes (*FEN1* and *SP4*) using *FOXLI*<sup>-WT</sup>, *FOXLI*<sup>-Mut</sup> and *FOXLI*<sup>-Mock</sup> transfected cells (Figure 6.6). Due to the difference in sensitivity and statistical power of the two different techniques, we were only able to validate the microarray results of two genes *CCXLI0* and *IL29*. Microarray analysis indicated a 2.85 fold down-regulation of *CXCL10* in *FOXLI*<sup>-Mut</sup> cells comparing with the *FOXLI*<sup>-WT</sup> cells. This was confirmed by QPCR, which showed decrease in *CXCL10* mRNA levels in *FOXLI*<sup>-Mut</sup> cells. Also, QPCR showed that *IL29* was significantly increased in *FOXLI*<sup>-Mut</sup> cells comparing with the *FOXLI*<sup>-WT</sup> cells, which was consistent with microarray results, which showed a 2.52 fold up-regulation in *FOXLI*<sup>-Mut</sup> cells.

### Figure 6.6. Validation of microarray data by quantitative RT-PCR

The level of gene expression in untransfected, *FOXLI*<sup>-Mock</sup>, *FOXLI*<sup>-WT</sup> and *FOXLI*<sup>-Mut</sup> cells was determined by quantitative RT-PCR. The relative expression was calculated compared to HEK293A positive control cells after normalisation against GAPDH. Bar graphs represent the mean  $\pm$  SD of three independent experiments. (\*\*=*p* value <0.001, \*\*=*p* value <0.01, \*=*p*<0.05). A. *CXCL10* is slightly expressed in the hFOB1.19, transfection of the empty plasmid cause increase the level of *CXCL10* expression. Transfection of the *FOXLI*<sup>-WT</sup> plasmid cause significant increase in the expression of *CXCL10* comparing to the *FOXLI*<sup>-Mock</sup>. Transfection of *FOXLI*<sup>-Mut</sup> showed significant decrease in the *CXCL10* expression comparing with the *FOXLI*<sup>-WT</sup>. B. the level of the expression of *IL29* was significantly increased in the *FOXLI*<sup>-Mut</sup> comparing with the *FOXLI*<sup>-WT</sup>.



6.4 Discussion: see pages 189-195

## **Appendix 18: Real time amplification and interpretation**

Real time PCR was performed using TaqMan® probe-based gene expression analysis kit that include pre-designed primers and probes for optimal amplification (Applied Biosystems) along with TaqMan® universal PCR master mix (Applied Biosystem). All gene expression assays have a carboxyfluorescein (FAM) reporter dye at the 5' end of TaqMan minor groove binder (MGB) probe and non-fluorescent quencher at the 3' end of the probe. All samples were assayed in triplicate to ensure accuracy with a reaction volume of 20 µl. Ten microliters of TaqMan gene expression master mix (Amplitaq® Gold DNA polymerase Ultra-Pure, Uracil DNA glycosylase, dNTP's , ROX™ which act as a passive reference) were mixed with 1 µl of specific primer, 1 µl of cDNA and 8 µl UltraPure™ DNase/RNase-Free distilled water. The optimal concentration of cDNA was predetermined using a 10-fold serial dilution to generate a standard curve. Ideally, 10-fold dilution increases the cycle threshold (CT) value about 3.33 cycles. The amplification efficiency is calculated from the slope, which should lie between 85-100%. The thermal cycler protocol included an initial 2 min at 50°C, 10 min at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 min at 60°C

Real-time PCR amplification was performed by the comparative cycle threshold ( $\Delta\Delta CT$ ) method and normalized to GAPDH. A control sample without RNA and a reference sample (RAJI, B cell line) were included in each experiment. The  $\Delta\Delta CT$  method is used to determine the relative quantity (RQ) in samples. The software

determines the RQ of target in each sample by subtracting normalized target quantity in each sample from normalized target quantity in the reference sample. Therefore, the formulas used to calculate RQ are:

$$CT_{\text{CITTA}} - CT_{\text{GAPDH}} = \Delta CT \text{ (both in samples and reference control)}$$

$$\Delta CT_{\text{sample}} - CT_{\text{reference control}} = \Delta \Delta CT$$

$$\textit{Relative quantity} = 2^{-\Delta \Delta CT}$$

After calculation of relative quantity of the samples and controls, samples and the positive control RNA expression were compared and the cut-off limit was placed at 0.2 (double the background).

## Appendix 19: Genes down regulated by FOXL1 mutant

Gene symbol	Gene Name	Function	Fold change	P-Value
<i>ADD3</i>	Adducin 3 (gamma)	Membrane-cytoskeleton-associated protein that promotes the assembly of the spectrin-actin network. Binds to calmodulin	1.39	1.55E-04
<i>ARHGAP11B</i>	Rho GTPase activating protein 11B	GTPase activation	2.15	8.01E-05
<i>CXCL10</i>	Chemokine (C-X-C motif) ligand 10	chemotaxis, cytokine, inflammation, inflammatory response, osteoclastogenesis, osteoblast proliferation	2.85	1.07E-05
<i>FEN1</i>	Flap structure-specific endonuclease 1	Acetylation, complete proteome, direct protein sequencing, Endonuclease, exonuclease, hydrolase, magnesium, metal-binding, nuclease	2.37	6.44E-07
<i>GRLF1</i>	Glucocorticoid receptor DNA binding factor 1	Alternative splicing, complete proteome, cytoplasm, dna-binding, GTPase activation, nucleus, phosphoprotein, repeat, repressor, Transcription, transcription regulation, tumor suppressor	1.57	1.55E-04
<i>IFIT1</i>	Interferon-induced protein with tetratricopeptide repeats 1	Proteome, polymorphism, repeat, tpr repeat, decreased in otosclerosis	1.82	7.11E-06
<i>IFNB1</i>	Interferon, beta 1, fibroblast	Antiviral defense, cytokine, pharmaceutical, ,Secreted, signal, maintain bone homeostasis	-2.58	2.91E-06
<i>IL1A</i>	interleukin 1, alpha	cytokine, immunoregulation, 'Inflammatory response, osteoclast activating factor, osteoclast differentiation	1.71	1.33E-04
<i>JAK2</i>	Janus kinase 2	kinase, ,SH2 domain, transferase, tyrosine-protein kinase, IFN $\gamma$ signaling, cytokine signalling	1.654172948	1.00E-04
<i>LRRN3</i>	leucine rich repeat neuronal 3	Immunoglobulin domain, leucine-rich repeat, membrane, polymorphism, repeat, signal,	2.21	1.94E-05

		transmembrane, fetal development		
<b>Gene symbol</b>	<b>Gene Name</b>	<b>Function</b>	<b>Fold change</b>	<b>P-Value</b>
<i>MIR658</i>	microRNA 658	Post-transcriptional regulation of gene	1.75	1.32E-05
<i>NBR2</i>	neighbor of BRCA1 gene 2	this gene does not appear to encode a protein	1.94	1.00E-05
<i>NIPBL</i>	Nipped-B homolog (Drosophila)	facilitates enhancer-promoter communication of remote enhancers and plays a role in developmental regulation associated with deafness	1.77	5.78E-05
<i>NR2F1</i>	nuclear receptor subfamily 2, group F, member 1	transcription factor neurogenesis and neural crest cell differentiation associated with deafness	2.15	1.34E-06
<i>PKIA</i>	protein kinase (cAMP-dependent, catalytic) inhibitor alpha	inhibitor of cAMP-dependent protein kinase activity	1.49	3.75E-05
<i>PTGER4</i>	prostaglandin E receptor 4 (subtype EP4)	one of four receptors identified for prostaglandin E2 (PGE2).  mediate PGE2 induced expression of early growth response 1 (EGR1), regulate the level and stability of cyclooxygenase-2 mRNA  Stimulation of bone formation and prevention of bone loss  PGE2 Signaling Through the EP4 Receptor on Fibroblasts Upregulates RANKL and Stimulates Osteolysis	1.6	1.44E-04
<i>PTGFR</i>	Prostaglandin F receptor (FP)	receptor forPGF2-alpha	1.59	2.89E-05

Gene symbol	Gene Name	Function	Fold change	P-Value
<i>PTGS2</i>	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	3d-structure,chromoprotein,complete proteome,dioxygenase,disulfide bond,endoplasmic reticulum,Fatty acid biosynthesis,glycoprotein,heme,iron,lipid synthesis,membrane,metal-binding,metalloprotein,microsome,oxidoreductase,peroxidase,phosphoprotein,polymorphism,prostaglandin biosynthesis,signal,	1.76	1.31E-05
<i>SETD6</i>	SET domain containing 6	down-regulate NF-kappa-B transcription factor activity by RELA6 monomethylates 'Lys-310'	1.73	9.88E-05
<i>SP4</i>	Sp4 transcription factor	Binds to GT and GC boxes promoters elements. Probable transcriptional activator	2.00	2.52E-06
<i>VTRNA1-1</i>	vault RNA 1-3; vault RNA 1-1; vault RNA 1-2	This family of RNAs are found as part of the enigmatic vault ribonucleoprotein complex.	3.34	1.97E-06
<i>ZNF555</i>	zinc finger protein 555	May be involved in transcriptional regulation	1.45	1.32E-04

## Appendix 20: Genes up regulated by FOXL1 mutant

Gene symbol	Gene Name	Function	Fold change	P-Value
<i>ASNS</i>	asparagine synthetase	Involved in the synthesis of asparagine, Associated with osteopenia	1.63	9.35E-05
<i>ATOH8</i>	atonal homolog 8 (Drosophila)	Putative transcription factor. implicated in specification and differentiation of neuronal cell lineages in the brain. participate in kidney development and involved in podocyte differentiation involved in TGF/BMP signaling	2.10	1.15 E-04
<i>CD1C</i>	CD1c molecule	Antigen-presenting protein that binds self and non-self lipid and glycolipid antigens and presents them to T-cell receptors on natural killer T-cells	5.07	5.69E-05
<i>CD34</i>	CD34 molecule	Cluster differentiation molecule, adhesion molecule It also interacts with L-selectin, important in inflammation.  Expressed in osteoblast precursors.  improved osteoblast activity and concurrently impaired osteoclast differentiation, maturation and functionality	1.62	7.47E-05
<i>CDH24</i>	cadherin-like 24	calcium dependent cell adhesion proteins upregulated in osteoblast	1.52	1.35 E-04
<i>CRYAB</i>	crystallin, alpha B	act as molecular chaperones protecting the osteoblast cytoskeleton from mechanical stress	1.81	1.27 E-04
<i>DDIT3</i>	DNA-damage-inducible transcript 3	Inhibits the DNA-binding activity of C/EBP and LAP by forming heterodimers that cannot bind DNA induces osteoblastic cell differentiation.	1.52019 8633	5.18E-05
<i>ELOVL3</i>	elongation of very long	Condensing enzyme that elongates saturated and monounsaturated very long chain fatty	2.35812 9228	8.83E-06

	chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 3	acids		
<b>Gene symbol</b>	<b>Gene Name</b>	<b>Function</b>	<b>Fold change</b>	<b>P-Value</b>
<i>FGD1</i>	FYVE, RhoGEF and PH domain containing 1	Activates CDC42, a member of the Ras-like family of Rho- and Rac proteins, by exchanging bound GDP for free GTP. Plays a role in regulating the actin cytoskeleton and cell shape  important regulator of bone development.	1.59412 4735	3.81E-05
<i>FMO1</i>	flavin containing monooxygenase 1	This protein is involved in the oxidative metabolism of a variety of xenobiotics	1.94829 9913	9.43E-05
<i>FOLR3</i>	folate receptor 3 (gamma)	Binds to folate and reduced folic acid derivatives and mediates delivery of 5-methyltetrahydrofolate to the interior of cells.  Foxl1 binding site	1.59510 7187	3.44E-05
<i>IL29</i>	interleukin 29 (interferon, lambda 1)	Cytokine with immunomodulatory activity	2.52788 3612	2.66E-05
<i>LILRB5</i>	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 5	bind to MHC class I molecules on antigen-presenting cells and inhibit stimulation of an immune response	1.71075 6739	5.39E-05
<i>PSAT1</i>	chromosome 8 open reading frame 62;	catalyzes the reversible conversion of 3-phosphohydroxypyruvate to phosphoserine and of	1.78477 6735	1.12 E-04

	phosphoserine aminotransferase 1	3-hydroxy-2-oxo-4-phosphonooxybutanoate to phosphohydroxythreonine		
<b>Gene symbol</b>	<b>Gene Name</b>	<b>Function</b>	<b>Fold change</b>	<b>P-Value</b>
<i>STK11IP</i>	serine/threonine kinase 11 interacting protein	regulate STK11/LKB1 function by controlling its subcellular localization	1.78365 1263	5.15E-05
<i>SYTL4</i>	synaptotagmin-like 4	Modulates exocytosis of dense-core granules and secretion of hormones	1.87031 0362	1.51E-04
<i>TLCD1</i>	TLC domain containing 1		1.83063 6358	1.00E-04
<i>TRAPPC6A</i>	trafficking protein particle complex 6A	play a role in vesicular transport	3.04674 7925	9.79E-05
<i>TLL1</i>	tubulin tyrosine ligase-like family, member 1	Catalytic subunit of the neuronal tubulin polyglutamylase complex.	1.64604 7548	5.65E-05
<i>ZNF787</i>	zinc finger protein 787	transcription termination factor I interacting peptide	1.65335 4835	0.0001.27-E04

## **Appendix 21: Statistical analysis report**

### **STATISTICAL CONSULTING REPORT**

**TITLE:**

Genome-wide linkage analysis

**SPONSOR:**

Dr. Terry Young

**REPORT PREPARED BY:**

Pingzhao Hu

Statistical Analysis Core Facility

The Center for Applied Genomics

The Hospital for Sick Children Research Institute

E-mail: [phu@sickkids.ca](mailto:phu@sickkids.ca)

Tel.: (416) 813-7654 ext. 6016

**TIME SPENT ON ANALYSIS:** 25 hours

**REPORT DATE:**

August 15, 2011

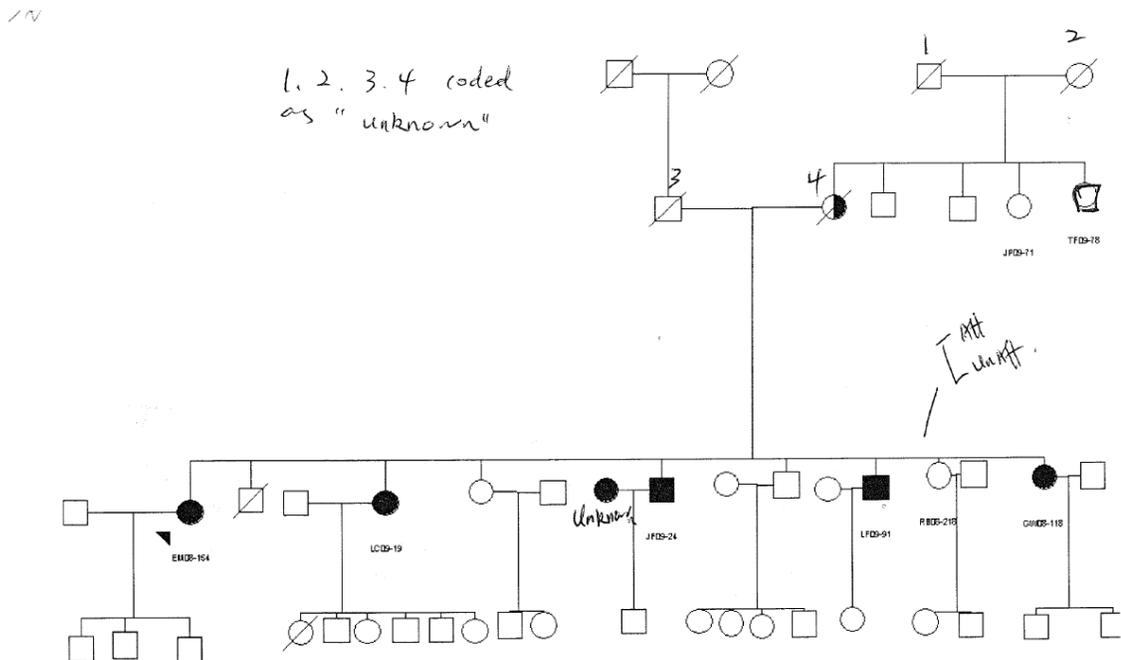
## Linkage Analysis

### Genotype Calls

The 8 samples in the family have been genotyped using Illumina Human 610 chip. Genotype calls (forward strand) in the family were exported from GenomeStudio software (V2010.3). All the 8 samples have call rate larger than 95%.

### Gender and Pedigree Check

We estimated the gender of the 8 individuals in GenomeStudio using their genotype information and confirmed the estimated gender of the individuals is matched to their clinically defined gender. We also performed pairwise IBD estimation of the individuals using PLINK (Purcell et al., 2007) and confirmed the below pedigree structure is correct.



pedigree	individualID	fatherID	motherID	sex	affection
F1	1	0	0	1	0
F1	2	0	0	2	0
<b>F1</b>	<b>TF09</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>
F1	JP09	1	2	2	1
F1	3	0	0	1	0
F1	4	0	0	2	0
F1	EM08	3	4	2	2
F1	LC09	3	4	2	2
F1	LF09	3	4	1	2
F1	GW08	3	4	2	2
F1	JF09	3	4	1	2
<b>F1</b>	<b>RB08</b>	<b>3</b>	<b>4</b>	<b>2</b>	<b>1</b>

Sex: 1-male while 2 is female

Affection: 1 is unaffected while 2 is affected and 0 is unknown

For TF09, the original gender was defined as female in the pedigree plot. We identified he should be male based on genetic information and reconfirmed with the family, now he is defined as male.

## **Genetic Linkage Analysis**

### **Minor allele frequency estimation**

We estimated minor allele frequency of the SNPs used in linkage analysis based on the 9 unrelated samples from the 7 families as bellows

TF09

BR08

DW08

AR07

RA04

EP06

LY08

SM08

PL07

### **Software for linkage analysis**

We used MERLIN software (version 1.1.2) to do the parametric linkage analysis (Abecasis et al. 2002). The software was downloaded from (<http://www.sph.umich.edu/csg/abecasis/Merlin/tour/parametric.html>).

### **Selection of markers for linkage analysis**

To select the markers only on autosomal chromosomes used in linkage analysis, we performed the following filtering steps using the 16 samples (8 in the family and 9 unrelated samples mentioned above. Please be noted that one (TF09) of the nine unrelated samples is in the family for linkage analysis):

**A: QC:** Here QC cutoffs are defined as follows (SNPs meet either condition is removed for further analysis):

- 1) Hardy Weinberg Equilibrium (Wigginton et al. 2005)  $p$  value in controls smaller than  $1 \times 10^{-5}$
- 2) Missing genotype rate (SNPs) larger than 0.05
- 3) Minor allele frequency (MAF) smaller than 0.05
- 4) Mendel error rate larger than 0.1

**B: LD-based SNP pruning using PLINK software** (<http://pngu.mgh.harvard.edu/~purcell/plink>). We performed two times pruning. In the first round pruning, the parameters used in the pruning are:

- Consider a window of 50 SNPs,
- Calculate LD between each pair of SNPs in the window,
- Remove one of a pair of SNPs if the LD is greater than 0.4,
- Shift the window 5 SNPs forward and repeat the procedure.

From the pruned data, we performed the second round pruning,

In the parameters used in the pruning are:

- Consider a window of 50 SNPs,
- Calculate LD between each pair of SNPs in the window,
- Remove one of a pair of SNPs if the LD is greater than 0.25,
- Shift the window 5 SNPs forward and repeat the procedure.

From the second round pruned data, we selected SNPs which have minor allele frequency (MAF) at least 0.45 in the selected samples described above. We got total 6073 SNPs used in final linkage analysis.

### **Genetic distance**

We obtained the genetic distance of the SNPs from <http://compgen.rutgers.edu/RutgersMap/AffymetrixIllumina.aspx>.

### **Penetrances**

We considered two genetic models: one is recessive model and another is dominant model

For autosomal recessive model, we consider penetrances 0.0001, 0.0001, 1.0 (no mutated allele, one mutated allele and two mutated alleles) and disease allele frequency 0.1.

For autosomal dominant model, we consider penetrances 0.0001, 1.0, 1.0 (no mutated allele, one mutated allele and two mutated alleles) and disease allele frequency 0.005.

### **Multi-point LOD score calculation (Parametric linkage analysis)**

We calculate the parametric LOD score for a 1-cM grid along the chromosomes. The distribution of LOD scores for each chromosome is shown in files for the 4 cases:

Dominant model \_RB08 is Unaffected

LODScore\_DominantModel\_RB08UnAff.txt (actual LOD scores)

LODScore\_DominantModel\_RB08UnAff.pdf (LOD score plot)

Recessive model \_RB08 is Unaffected

LODScore\_RecessiveModel\_RB08UnAff.txt (actual LOD scores)

LODScore\_RecessiveModel\_RB08UnAff.pdf (LOD score plot)

### **The potential linkage signals (LOD>1) in each case are:**

Dominant model \_RB08 is Unaffected (maximum LOD ~1.4)

Chr7: 85.059cM - 152.059cM

Chr10: 0.121cM - 7.121cM and 67.121cM - 69.121cM

Chr16: 112.252cM - 133.252cM

Chr17: 112.044cM - 129.044cM

All these 5 regions are in the previous analysis with the same bound.

Recessive model \_RB08 is Unaffected (maximum LOD 2.5)

Chr17: 111.044cM - 129.044cM

### **Steps to find genes/physical positions in these regions:**

1. Based on the start and end position (cM) in the linkage regions (see above in each case), you can find the start and end SNP ids in file: Nellya.map (this file has been sorted).
2. Based on the start and end SNPs in a given region, you can find the start and end positions (bp) from dbSNP – remember genome build for the positions (you can use either hg18 or hg19)
3. Based on the start and end position (bp), you can find genes from UCSC genome browser in the regions.

**Appendix 22: List of primers sequences used for genes sequenced in the critical region on chromosome in Family 2114**

<i>BIRC</i>						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E1	NM_001012271-Ex1F	gactacaactcccggcacac	344	TD 54	yes (0.75)	1.5
	NM_001012271-Ex1R	agtcacagtggcctcgcta				
E2	NM_001012271-Ex2F	ctcccctccctgcttgt	248	TD 54	yes (0.75)	1.5
	NM_001012271-Ex2R	gttcaaaacaagcccatcg				
E3	NM_001012271-Ex3F	tgagagtgtgagctaggggg	194	TD 54	yes (0.75)	1.5
	NM_001012271-Ex3R	tagtggagacggggtttcac				
E4	NM_001012271-Ex4F	actgccgtttaatcccttc	213	TD 54	yes (0.75)	1.5
	NM_001012271-Ex4R	cattgaacagggttgagca				
E5	NM_001012271-Ex5F	aaatatggtaggggaaggggg	168	TD 54	yes (0.75)	1.5
	NM_001012271-Ex5R	aacagaccctggcaaacatc				
E6-1	NM_001012271-Ex6-1F	ctgggaagctctggtttcag	461	TD 54	yes (0.75)	1.5
	NM_001012271-Ex6-11R	CTAGCAAAAGGGACACTGCC				
E6-2	NM_001012271-Ex6-2F	GGGCTCATTTTTGCTGTTTT	465	TD 54	yes (0.75)	1.5
	NM_001012271-Ex6-2R	CATCCACCTGAAGTTCACCC				
E6-3	NM_001012271-Ex6-3F	CATGGCTTTCCTATTTTGTGTTGAA	498	TD 54	yes (0.75)	1.5
	NM_001012271-Ex6-31R	TTCCAGCGAAGCTGTAACAA				
E6-4	NM_001012271-Ex6-4F	GCTGAAGTCTGGCGTAAGATG	466	TD 54	yes (0.75)	1.5
	NM_001012271-Ex6-41R	ACAGAGGCTGGAGTGCATTT				
E6-5	NM_001012271-Ex6-5F	TTTCTGCCACATCTGAGTCG	476	TD 54	yes (0.75)	1.5
	NM_001012271-Ex6-51R	CCTGAAAAATGGACATGTGG				
E6-6	NM_001012271-Ex6-6F	AATAAAGCCGTAGGCCCTTG	390	TD 54	yes (0.75)	1.5

	NM_001012271-Ex6-61R	ctgtggtcattccactgcac				
<b>CTDNEPI</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E2	NM_015343-Ex2F	ctccccgttctgcgag	407	TD 54	yes (0.75)	1.5
	NM_015343-Ex2R	ggattcccttctagccc				
E3-4	NM_015343-Ex3-4F	ggtgcacgcctgtagtcc	280	TD 54	yes (0.75)	1.5
	NM_015343-Ex3-4R	cagaggaacaagatgggctg				
E5	NM_015343-Ex5F	gtcccagagcatctgtctcc	82	TD 54	yes (0.75)	1.5
	NM_015343-Ex5R	taccattacacagcctcccc				
E6-7	NM_015343-Ex6-7F	tcctcagagacattgtgttg	345	TD 54	yes (0.75)	1.5
	NM_015343-Ex6-7R	tgcattcgacttacagcaaac				
E8	NM_015343-Ex8F	tgttgactccggtagtggg	85	TD 54	yes (0.75)	1.5
	NM_015343-Ex8R	tttgaagtaagacgacctggg				
E9-1	NM_015343-Ex9-1F	gtaggctgtgggcaattga	329	TD 54	yes (0.75)	1.5
	NM_015343-Ex9-1R	TGTGTCCATCCAGACTCCAA				
<b>FOXJ1</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E1	NM_001454-Ex1F	agaggcggaaagctgtttatg	315	TD 54	yes (0.75)	1.5
	NM_001454-Ex1R	ccacttccgttaaggacgaa				
E2-1	NM_001454-Ex2-1F	cegtacacacactgtcccct	585	TD 54	yes (0.75)	1.5
	NM_001454-Ex2-1R	CGGATTGGTGGCGTAGTC				
E2-2	NM_001454-Ex2-2F	ACCAGGTGCCAGGTTTCAG	399	TD 54	yes (0.75)	1.5
	NM_001454-Ex2-2R	gcctgcagattgggatatg				
E3-1	NM_001454-Ex3-1F	cactgacctagcggcttcctc	476	TD 54	yes (0.75)	1.5
	NM_001454-Ex3-1R	AAAGTTGCCTTTGAGGGGTT				
E3-2	NM_001454-Ex3-2F	GCATAAGCGCAAACAGCC	499	TD 54	yes (0.75)	1.5

Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
	NM_001454-Ex3-2R	ACCCTGACTTGGGCACTGT				
E3-3	NM_001454-Ex3-3F	CTCTTTGAGGCTGGGGATG	490	TD 54	yes (0.75)	1.5
	NM_001454-Ex3-3R	CCTGGGGACTCTCTCTGGAT				
E3-4	NM_001454-Ex3-4F	AACAGAACTGGGCCCTCC	495	TD 54	yes (0.75)	1.5
	NM_001454-Ex3-4R	GGTCCCAGTAGTTCCAGCAA				
E3-5	NM_001454-Ex3-5F	GGGGCAGGACAGACAGACTA	331	TD 54	yes (0.75)	1.5
	NM_001454-Ex3-5R	ctctccagaactgagccc				
<b><i>ITGB4</i></b>						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E2	NM_000213-F	gcccttggtcacattgtg	89	TD 54	yes (0.75)	1.5
	NM_000213 -R	caagttttccagccctgaag				
E3-4	NM_000213-F	ccatctctccaggtgaaggt	312	TD 54	yes (0.75)	1.5
	NM_000213 -R	cagctggctagagggctact				
E5	NM_000213-F	GAGgtgcctggtgtggg	205	TD 54	yes (0.75)	1.5
	NM_000213 -R	ggggctcaatgtccagtg				
E6	NM_000213-F	acctttgtccagcatgcaa	97	TD 54	yes (0.75)	1.5
	NM_000213 -R	tctgacctcaggtgatctgc				
E7	NM_000213-F	gtgagccaagatcgtgcata	172	TD 54	yes (0.75)	1.5
	NM_000213 -R	tgacctgtctcagggtgaa				
E8	NM_000213-F	cttgtctgtaaccagtggc	264	TD 54	yes (0.75)	1.5
	NM_000213 -R	ggaaggaaaggaggaaaag				
E9-10	NM_000213-F	ggctaagaggcaggctc	494	TD 54	yes (0.75)	1.5
	NM_000213 -R	gctaggcccaggatctg				
E11-12	NM_000213-F	gagcaggagctcatttcagg	428	TD 54	yes (0.75)	1.5
	NM_000213 -R	tagccgtttctttgattgcc				
E13	NM_000213-F	agcttgcaagtgcgaagat	203	TD 54	yes (0.75)	1.5

	NM_000213 -R	ctggaactggggattctga				
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E14-15	NM_000213-F	gagatgacttctaccccaggc	336	TD 54	yes (0.75)	1.5
	NM_000213 -R	gctccccaacctcttgc				
E16	NM_000213-F	caggaggggctaagcctg	130	TD 54	yes (0.75)	1.5
	NM_000213 -R	ccagctctggagtctctgc				
E17-18	NM_000213-F	aaggactccactctccccc	323	TD 54	yes (0.75)	1.5
	NM_000213 -R	accaggaaggggttg				
E19-20	NM_000213-F	gtgtctggggaggcactg	400	TD 54	yes (0.75)	1.5
	NM_000213 -R	catagcaccatcagcacagg				
E21-22	NM_000213-F	ctatgaacctcatgctcgg	494	TD 54	yes (0.75)	1.5
	NM_000213 -R	acaaagaggcgctgtgag				
E23	NM_000213-F	AGACCAAGTCCGgtgagtc	24	TD 54	yes (0.75)	1.5
	NM_000213 -R	acacatctgacctcccttc				
E24-25	NM_000213-F	gaagacctgcacttcttgc	421	TD 54	yes (0.75)	1.5
	NM_000213 -R	acagatggagaaccaaggc				
E26	NM_000213-F	cagaggccaatcagatagg	149	TD 54	yes (0.75)	1.5
	NM_000213 -R	gcacggccagtatccag				
E27	NM_000213-F	gtactccaccagcaagcag	205	TD 54	yes (0.75)	1.5
	NM_000213 -R	tcctaccaacgtgtatgttttac				
E28	NM_000213-F	ggtgggcaggtctgagttg	158	TD 54	yes (0.75)	1.5
	NM_000213 -R	gaccacctgtgctaggttcc				
E29-30	NM_000213-F	cacatggcagatctctcage	432	TD 54	yes (0.75)	1.5
	NM_000213 -R	gctgaccaggatgaaggc				
E31-32	NM_000213-F	gcttgccttgccttccc	404	TD 54	yes (0.75)	1.5
	NM_000213 -R	attgcattgtccaaggctg				
E33	NM_000213-F	aagggttcacctgccg	210	TD 54	yes (0.75)	1.5

	NM_000213 -R	tgctctgagcagctgtg				
E34	NM_000213-F	gagggaactggttatttaagc	240	TD 54	yes (0.75)	1.5
	NM_000213 -R	gtgtgcgtgcatgtgtgc				
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E35	NM_000213-F	gggtgagtgagttgccagc	150	TD 54	yes (0.75)	1.5
	NM_000213 -R	tctcaggctggatgggg				
E36-37	NM_000213-F	aaaccacagctagtctggg	431	TD 54	yes (0.75)	1.5
	NM_000213 -R	GGGCTGTCTCCATCCACC				
E38-39	NM_000213-F	ATGGGGATATCGTCGGCTAC	368	TD 54	yes (0.75)	1.5
	NM_000213 -R	AGGGTCAGCCCATctgtg				
<b>OTOP2</b>						
E1	NM_178160-F	ccctgcctctcgggtataa	200	TD 54	yes (0.75)	1.5
	NM_178160 -R	tggatgggaagaaaggagaa				
E2	NM_178160-F	tcagggtgacctggagttt	485	TD 54	yes (0.75)	1.5
	NM_178160 -R	gccctcgtatgtcaatctgc				
E3	NM_178160-F	gccttctctcctcctttt	244	TD 54	yes (0.75)	1.5
	NM_178160 -R	tggccaagtgttagctcctt				
E4	NM_178160-F	cgctggagttttgccatct	193	TD 54	yes (0.75)	1.5
	NM_178160 -R	agggtggaaagcctgaatct				
E5	NM_178160-F	agggtcacagcctctttgt	266	TD 54	yes (0.75)	1.5
	NM_178160 -R	atgagggcacaggagactga				
E6-1	NM_178160-F	aaaggtcacaggctagggtt	489	TD 54	yes (0.75)	1.5
	NM_178160 -R	AAACGGTAGATGATGGAGCC				
E6-2	NM_178160-F	GCAGGCCCTGGTCATCTACT	492	TD 54	yes (0.75)	1.5
	NM_178160 -R	CTTCTGGGGGTTGAGACGAT				
E6-3	NM_178160-F	ACCTCACCTTCACCAACCTG	231	TD 54	yes (0.75)	1.5
	NM_178160 -R	tcagtctcttccaagcc				

E7	NM_178160-F	agaaatgaccacaggcatc	304	TD 54	yes (0.75)	1.5
	NM_178160 -R	ttcccagagggtgtctgagt				
<b>RNF157</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E1	NM_052916 -F	TGTACCGCTACCCGCCCAAGT	147	TD 54	yes (0.75)	1.5
	NM_052916 -R	TTTCGGAGCGTCCGCAACCA				
E2	NM_052916 -F	tcagggtgacctggagttt	119	TD 54	yes (0.75)	1.5
	NM_052916 -R	gcctctgtatgcaatctgc				
E3	NM_052916 -F	gccttctctctgcctcttt	89	TD 54	yes (0.75)	1.5
	NM_052916 -R	tgccaagtgttagctcctt				
E4	NM_052916 -F	cgctggagttttgtccatct	147	TD 54	yes (0.75)	1.5
	NM_052916 -R	agggtggaagcctgaatct				
E5	NM_052916 -F	agggtcacaggcctctttgt	118	TD 54	yes (0.75)	1.5
	NM_052916 -R	atgagggcacaggagactga				
E6	NM_052916 -F	aaaggtcacaggctaggggt	67	TD 54	yes (0.75)	1.5
	NM_052916 -R	AAACGGTAGATGATGGAGCC				
E7	NM_052916 -F	GCAGGCCCTGGTCATCTACT	44	TD 54	yes (0.75)	1.5
	NM_052916 -R	CTTCTGGGGGTTGAGACGAT				
E8	NM_052916 -F	ACCTCACCTTCACCAACCTG	48	TD 54	yes (0.75)	1.5
	NM_052916 -R	tcagtcttctcccaagcc				
E9	NM_052916 -F	agaaatgaccacaggcatc	72	TD 54	yes (0.75)	1.5
	NM_052916 -R	ttcccagagggtgtctgagt				
E10-11	NM_052916 -F	TGTACCGCTACCCGCCCAAGT	468	TD 54	yes (0.75)	1.5
	NM_052916 -R	TTTCGGAGCGTCCGCAACCA				
E12	NM_052916 -F	tcagggtgacctggagttt	239	TD 54	yes (0.75)	1.5
	NM_052916 -R	gcctctgtatgcaatctgc				
E13	NM_052916 -F	gccttctctctgcctcttt	109	TD 54	yes (0.75)	1.5

	NM_052916 -R	tgccaagtgttagctcctt				
E14	NM_052916 -F	cgctggagttttgccatct	112	TD 54	yes (0.75)	1.5
	NM_052916 -R	agggtggaaagcctgaatct				
E15-16	NM_052916 -F	agggtcacaggcctctttgt	483	TD 54	yes (0.75)	1.5
	NM_052916 -R	atgagggcacaggagactga				
E17	NM_052916 -F	aaaggtcacaggctagggt	46	TD 54	yes (0.75)	1.5
	NM_052916 -R	AAACGGTAGATGATGGAGCC				
E18	NM_052916 -F	GCAGGCCCTGGTCATCTACT	111	TD 54	yes (0.75)	1.5
	NM_052916 -R	CTTCTGGGGGTTGAGACGAT				
E19-1	NM_052916 -F	ACCTCACCTTCACCAACCTG	580	TD 54	yes (0.75)	1.5
	NM_052916 -R	tcagtctcttcccagcc				
E19-2	NM_052916 -F	agaaatgacccacaggcatc	580	TD 54	yes (0.75)	1.5
	NM_052916 -R	tcccagagggtgtctgagt				
<b>RPL38</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E1	NM_001035258 -F	gcgaacgcctacatgagtct	202	TD 54	yes (0.75)	1.5
	NM_001035258 -R	ggtccccagatttcacctg				
E2	NM_001035258 -F	gccgatattcgggggag	198	TD 54	yes (0.75)	1.5
	NM_001035258 -R	ggaggctctctgccctc				
E3	NM_001035258 -F	ccgggagacgtgtctcttt	154	TD 54	yes (0.75)	1.5
	NM_001035258 -R	ctaggcaccgetcttctga				
E4	NM_001035258 -F	gcttgagtgtattgcacatgg	258	TD 54	yes (0.75)	1.5
	NM_001035258 -R	acacagcaacttccaaacc				
E5	NM_001035258 -F	agacaagcagcagaactca	179	TD 54	yes (0.75)	1.5
	NM_001035258 -R	agcattttcccttcccat				
<b>TIMP2</b>						
E1	NM_003255-F	CCCAGACAAAGAGGAGAGA	458	TD 54	yes (0.75)	1.5

	NM_003255 -R	actgacgagattcgagaagg				
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E2	NM_003255-F	tgacgtcttatccctctcc	288	TD 54	yes (0.75)	1.5
	NM_003255 -R	tagtccacaggtggccaga				
E3	NM_003255-F	gctgtaccccagctacaggt	349	TD 54	yes (0.75)	1.5
	NM_003255 -R	tacctgctcactgttctct				
E4	NM_003255-F	acagtgcacagattgttgc	372	TD 54	yes (0.75)	1.5
	NM_003255 -R	ctcccagagcctgtctta				
E5-1	NM_003255-F	gcacatgacaggtaggcaca	481	TD 54	yes (0.75)	1.5
	NM_003255 -R	CCAGACCCACAACCATGTCT				
E5-2	NM_003255-F	ctggacatgccccagag	395	TD 54	yes (0.75)	1.5
<b>KCTD2</b>						
<b>Exon</b>	<b>Primer I.D.</b>	<b>Sequence</b>	<b>Amplicon Size</b>	<b>Annealing temp.</b>	<b>Betaine (final conc. (M))</b>	<b>MgCl (final conc (mM))</b>
E1	NM_015353 -F	tctccctgccgagaaatg	599	TD 54	yes (0.75)	1.5
	NM_015353 -R	atgcctcagcatggaagaag				
E2	NM_015353 -F	ttgcaaatccagccagtaag	597	TD 54	yes (0.75)	1.5
	NM_015353 -R	tctcaccaggaagctgaac				
E3	NM_015353 -F	ttggagactgaggcaagag	551	TD 54	yes (0.75)	1.5
	NM_015353 -R	tgtaagcagcgacaatcttc				
E4	NM_015353 -F	cacagtatgctggcttgg	557	TD 54	yes (0.75)	1.5
	NM_015353 -R	tggaaagtctcgtgtgaag				
E5	NM_015353 -F	aatgatgccctgatgaagaac	559	TD 54	yes (0.75)	1.5
	NM_015353 -R	attgattggcagaggatgg				
E6	NM_015353 -F	ccccgacatctctgttttc	565	TD 54	yes (0.75)	1.5
	NM_015353 -R	acgtcttggagggtgcatag				

<i>USH1G</i>						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E1	NM_173477-F	gggtgagcgttcagatgtctg	483	TD 54	yes (0.75)	1.5
	NM_173477 -R	ggcagctcagaggagtgggga				
E2	NM_173477-F	ctgtgacagtgggaagctccc	1,399	TD 54	yes (0.75)	1.5
	NM_173477 -R	cctgaataggcagatctgtaccccc				
E2a	NM_173477-F	tctccgaggatggcgcaag	551	TD 54	yes (0.75)	1.5
	NM_173477 -R	gaggaacatgtccccggagcgg				
E3	NM_173477-F	ggtatctatggccgttatcctg	385	TD 54	yes (0.75)	1.5
	NM_173477 -R	tgcagacagacttcaaaggag				

<i>SLC9A3R1</i>						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E1	NM_004252-F	tggtctgtgctcctctctcg	810	TD 54	yes (0.75)	1.5
	NM_004252 -R	cctgcctaggatgtgtcagg				
E2	NM_004252-F	ccagaaccctctcaattgc	542	TD 54	yes (0.75)	1.5
	NM_004252 -R	caggggagtgaggtaagag				
E3	NM_004252-F	atatttctgtggccctcc	576	TD 54	yes (0.75)	1.5
	NM_004252 -R	tgccactccctacactagge				
E4	NM_004252-F	ggggctaggagtttgagacc	552	TD 54	yes (0.75)	1.5
	NM_004252 -R	tgtatggtgggagttggag				
E5	NM_004252-F	gtctgttccatcccatcc	517	TD 54	yes (0.75)	1.5
	NM_004252 -R	cattaggctccctgtgctc				
E6	NM_004252-F	gttctgtgacctgcttcc	544	TD 54	yes (0.75)	1.5
	NM_004252 -R	aggcactcagtgaggaggag				

<i>TNRC6C</i>						
Exon	Primer I.D.	Sequence	Amplicon Size	Annealing temp.	Betaine (final conc. (M))	MgCl (final conc (mM))
E1	NM_004252-F	tggtctgtgctcctctctcg	810	TD 54	yes (0.75)	1.5
	NM_004252 -R	cctgcctaggatgtgcagg				
E2	NM_004252-F	ccagaaccctctcaaattgc	542	TD 54	yes (0.75)	1.5
	NM_004252 -R	caggggagtgaggtaagag				
E3	NM_004252-F	atatttcggtggccttcc	576	TD 54	yes (0.75)	1.5
	NM_004252 -R	tgccactccctacactaggc				
E4	NM_004252-F	ggggctaggagttagagacc	552	TD 54	yes (0.75)	1.5
	NM_004252 -R	tgtatggtgggagttggag				
E5	NM_004252-F	gtctgttcccatccatcc	517	TD 54	yes (0.75)	1.5
	NM_004252 -R	cattaggctccctggctctc				
E6	NM_004252-F	gttcttgtgacctggcttcc	544	TD 54	yes (0.75)	1.5
	NM_004252 -R	aggcactcagtgaggaggag				

## Appendix 2: Positional candidate gene in the candidate region at Chr 17q

Gene name	Gene function	Accession number	Strand	Start	End	Ex n
<i>BIRC5</i>	Encoded negative regulatory proteins that prevent apoptotic cell death	NM_001012271	+	76210277	76221716	5
<i>CTDNEP1</i>	expressed in bone cells	NM_015343	+	7146906	7155259	9
<i>TIMP2</i>	The proteins encoded by this gene family are natural inhibitors of the matrix metalloproteinases, a group of peptidases involved in degradation of the extracellular matrix	NM_003255	-	76849059	76921472	5
<i>RPL38</i>	This gene encodes a ribosomal protein that is a component of the 60S subunit	NM_000999	+	72199795	72206019	3
<i>SLC9A3R1</i>	Defect in this gene causes osteoporosis type2	NM_004252	+	72744763	72765499	
<i>USH1G</i>	Mutations in this gene are associated with Usher syndrome type 1G (	NM_173477	-	70423771	70430946	3
<i>OTOP2</i>		NM_178160	+	70431965	70441601	7
<i>KCTD2</i>	Homo sapiens potassium channel tetramerisation domain containing 2	NM_015353	+	70554874	70573576	6
<i>RNF157</i>		NM_052916	-	71650129	71747985	19
<i>ITGB4</i>	Integrins mediate cell-matrix or cell-cell adhesion, and transduced signals that regulate gene expression and cell growth.	NM_000213	+	71229111	71265494	40
<i>FOXJ1</i>	Polymorphisms in this gene are associated with systemic lupus erythematosus and allergic rhinitis.	NM_001454	-	74132415	74137380	1

## Appendix 24: Primers used to amplify three variants identified by NGS

Gene name	Genomic position	Variant	Primers	Amplicon size	PCR program
<i>PKD1L2</i>	81242107	c.749A/G, p.Q250R	F. agcccagtacctggtcttca R. gtgggtggctgctgtctc	370	TD54
<i>PKD1L2</i>	81242198	c.658 C>T, p.Q220*			
<i>KPNA7</i>	98805041	c.49C>T p.R17*	F. taaccacacaatgggcact R. tgttgggattacaggcatga	244	TD54
<i>CDK14</i>	90338935	c.69+11C>G -	F. TTGTGAATTGCCTTGACAGC R. tgagaagaactcaaggacatgc	197	TD54
<i>TNRC6C</i>	75495449	c.1706 C>T, p.P569L	F. AGTATCTGGGTGGGTCAACG R. GTTCCCAGATCTCGTGGTGT	400	TD54
<i>SEPT9</i>	75495449	*709C>T	F. GCCCTGCTCCTAAGGGTAGA R. ACACTTCCTGGTCACTTGG	459	TD54