A molecular and genetic analysis of otosclerosis

Joanna Lauren Ziff

Submitted for the degree of PhD

University College London

January 2014

Declaration

I, Joanna Ziff, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Where work has been conducted by other members of our laboratory, this has been indicated by an appropriate reference.

Abstract

Otosclerosis is a common form of conductive hearing loss. It is characterised by abnormal bone remodelling within the otic capsule, leading to formation of sclerotic lesions of the temporal bone. Encroachment of these lesions on to the footplate of the stapes in the middle ear leads to stapes fixation and subsequent conductive hearing loss. The hereditary nature of otosclerosis has long been recognised due to its recurrence within families, but its genetic aetiology is yet to be characterised. Although many familial linkage studies and candidate gene association studies to investigate the genetic nature of otosclerosis have been performed in recent years, progress in identifying disease causing genes has been slow. This is largely due to the highly heterogeneous nature of this condition.

The research presented in this thesis examines the molecular and genetic basis of otosclerosis using two next generation sequencing technologies; RNA-sequencing and Whole Exome Sequencing. RNA–sequencing has provided human stapes transcriptomes for healthy and diseased stapes, and in combination with pathway analysis has helped identify genes and molecular processes dysregulated in otosclerotic tissue. Whole Exome Sequencing has been employed to investigate rare variants that segregate with otosclerosis in affected families, and has been followed by a variant filtering strategy, which has prioritised genes found to be dysregulated during RNA-sequencing. This has identified multiple variants predicted to be involved in splicing within genes involved in the bone disorder Osteogenesis Imperfecta, indicating a shared genetic aetiology for this condition and otosclerosis and a possible disease mechanism involving alternative splicing in the stapes.

Whilst the hereditability of otosclerosis remains elusive, the identification of new candidate genes will make a significant contribution to the current literature. It is hoped that long term, this research will help reveal disease mechanisms and thereby improve treatment options for otosclerosis patients.

Dedication

In loving memory of

Henry Perlow

1928-2009

Acknowledgements

Firstly, I thank my PhD supervisor Dr Sally Dawson for the opportunity to carry out this project and for her guidance and patience during the last three years. I appreciate the many hours of teaching and advice I have received from her throughout my PhD and during the preparation of this thesis. I also thank my friends and colleagues at UCL Ear Institute who have made my time here so memorable. In particular I thank Emily, Greg, Kiran, Lisa, Miriam, Naila, Paromita and Steph for sharing their knowledge and expertise with me and for all their assistance in the lab as well as their friendship.

I also express my gratitude to my second supervisor Professor Shakeel Saeed for his enthusiasm for this project and for his valuable input into all clinical aspects of the study. I also thank him, along with Harry, Jeremy and Sherif, for their vital assistance with patient recruitment and sample collection which was crucial for the success of the project.

I would also like to acknowledge Deafness Research UK/ Action on Hearing Loss for funding this PhD studentship and also to the UCL graduate school for funding my trip to the American Society of Human Genetics Annual conference in San Francisco.

Most importantly, I thank my family and friends for their tireless support and encouragement throughout the past three years. They have helped me through all the problems I have encountered on this journey and enabled me to keep these in perspective.

Finally, I would like to express my gratitude to the otosclerosis patients and their families who participated in this research study, without whom this project would not have been possible.

Abbreviations

dNTPs	Dideoxynucleotide triphosphates
EDTA	Ethylenediamine tetra-acetic acid
FDR	False discovery rate
GWAS	Genome wide association study
LD	Linkage disequilibrium
NGS	Next generation sequencing
PCR	Polymerse chain reaction
qPCR	Quantitative polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RNA-seq	RNA sequencing
RPKM	Reads per kilobase per million mapped reads
SNP	Single nucleotide polymorphism
TAE	Tris-acetate-EDTA electrophoresis buffer
UV	Ultraviolet
WES	Whole exome sequencing
WGS	Whole genome sequencing

Table of Contents

DECLARATION **2**

ABSTRACT **3**

DEDICATION **4**

ACKNOWLEDGEMENTS **5**

ABBREVIATIONS **6**

TABLE OF CONTENTS **7**

LIST OF FIGURES **10**

LIST OF TABLES **13**

1 INTRODUCTION **15**

1.1 Deafness and hearing loss 15

1.2 Structure and function of the ear 16

1.3 Otosclerosis 20

1.4 Development of the stapes 28

1.5 Bone remodelling in the otic capsule 29

1.6 Animal models for otosclerosis 32

1.7 The disease aetiology of otosclerosis 33

1.8 Aims and objectives of this thesis 58

2 MATERIALS AND METHODS **59**

Materials 59

2.1 Chemicals and Equipment 59

2.2 Abbreviations 59

2.3 Stock Solutions 59

Methods 60

2.4 Tissue collection 60

2.5 cDNA preparation 62

2.6 Gene Expression Assays 64

2.7 Genomic DNA preparation from blood and saliva samples 67

2.8 Detection of genomic variation 67

2.9 Agarose gel Electrophoresis 72
2.10 Purification of PCR products 72

3 COHORT CHARACTERISTICS 74

3.1 Overview 74
3.2 Results 74
3.2.1 Collection of DNA samples from study participants 74
3.2.2 Characteristics of the otosclerosis cohort 75
3.2.3 Environmental factors to which cohort members have been exposed 79
3.2.4 Incidence of disease in cohort members 83
3.3 Discussion of cohort characteristics 84

4 WHOLE EXOME SEQUENCING IDENTIFIES NOVEL GENETIC VARIANTS THAT SEGREGATE WITH OTOSCLEROSIS IN FAMILIES 92

4.1 Introduction 92

PART I IDENTIFYING RARE VARIANTS THAT SEGREGATE WITH FAMILIAL OTOSCLEROSIS

4.2 Identification of ten individuals for WES 95
4.3 WES in ten familial otosclerotic cases 104
4.4 Systematic step-wise prioritisation of variants 106
4.5 Variant filtering for segregation analysis 112
4.6 Rationale for segregation analysis and confirmation of a causal role 117
4.7 Genotyping in Family A 118
4.8 Genotyping in Family B 125
4.9 Genotyping in Family C 130
4.10 Genotyping in Family D 135

PART II ANALYSING PRIORITISED VARIANTS THAT SEGREGATE WITH FAMILIAL OTOSCLEROSIS

4.11 Evaluation of *TRIM17* as an otosclerosis candidate gene in family C 141
4.12 Evaluation of *mir-183* and *COL1A2* as candidate genes in family D 144
4.13 Evaluation of *SERPINF1* as a family B candidate gene 149
4.14 Discussion of WES 170

5 WHOLE TRANSCRIPTOME EXPRESSION ANALYSIS IDENTIFIES GENES DYSREGULATED IN OTOSCLEROTIC STAPES 174

5.1 Introduction 174

5.2 Results and Discussion 177

5.2.1 Selection of stapes for RNA-seq 177

5.2.2 Overview of RNA-seq data 179

5.2.3 Identification of a gene expression profile in human stapes suprastructures ..181

5.2.4 Identification of a gene expression profile in human stapes footplates 183

5.2.5 Identifying molecular pathways in the stapes gene expression transcriptome 184

5.2.6 Selection of candidate genes for follow up gene expression studies 189

5.2.7 qRT-PCR assays on the six genes selected for follow up analysis 194

5.3 Discussion of RNA-seq 199

6 INVESTIGATING *RELN* EXPRESSION IN OTOSCLEROTIC STAPES 203

6.1 Introduction 203

6.2 Results 204

6.2.1 Selection of stapes for gene expression analysis 204

6.2.2 An alternatively spliced variant of *RELN* is detected in human stapes 205

6.2.3 Expression of *RELN* in human stapes is confirmed by RNA-seq 209

6.3 Discussion of expression of *RELN* in human stapes 210

7 GENERAL DISCUSSION 217

7.1 Could a shift in isoform expression of otosclerosis candidate genes be involved in disease pathogenesis? 217

7.2 A possible link between heparan sulphate modulation of TGFβ signalling and the otosclerosis disease process? 219

7.3 Osteogenesis Imperfecta and Otosclerosis: A shared genetic aetiology? 222

7.4 Conclusions 224

APPENDIX 226

BIBLIOGRAPHY 236

Figures

Figure 1.1 Anatomy of the Mammalian Ear 18
Figure 1.2 Middle ear anatomy....19
Figure 1.3 Bright field microscope image of a human stapes bone....20
Figure 1.4 Photographs by Politzer 1894 depicting stapes fixation in otosclerosis.....24
Figure 1.5 Normal hearing and otosclerosis audiograms.....25
Figure 1.6 The bone remodelling cycle 30
Figure 1.7 Molecular interactions involved in bone remodelling....31
Figure 1.8 A spectrum of genetic variation 34
Figure 1.9 Transforming growth factor-β (TGFβ) involvement in bone remodelling 53
Figure 3.1 Bar graph showing numbers of stapes and genomic DNA samples collected 75
Figure 3.2 Bar graph showing age of hearing loss onset of questionnaire respondents 76
Figure 3.3 Bar graph showing age of hearing loss onset of male and female questionnaire respondents....76
Figure 3.4 Pie chart showing ethnicity of questionnaire respondents 77
Figure 3.5 Bar graph showing family history of hearing loss in male and female questionnaire respondents....78
Figure 3.6 Pie charts showing proportion of respondents with bilateral hearing loss, tinnitus and vertigo....79
Figure 3.7 Bar graph showing percentage of respondents born in each decade vaccinated for and infected by measles....80
Figure 3.8 Pie charts showing number of pregnancies in female respondents, number of months for which respondents had breastfed, and proportion of respondents who developed hearing loss during or after pregnancy 82
Figure 3.9 Pie chart showing number of years for which female respondents had taken oral contraceptives 82
Figure 4.1 Variant prioritisation following WES....93
Figure 4.2 Pedigree of Family A 97
Figure 4.3 Pedigree of Family B 99
Figure 4.4 Pedigree of Family C 100
Figure 4.5 Pedigree of Family D 102
Figure 4.6 Schematic showing the step-wise variant prioritisation process employed following WES. 106

Figure 4.7 Chromatograms showing c.1224_1226dupCAT in *SMAP1* 121
Figure 4.8 Chromatogram showing heterozygote for variant c.601G>A in *SERPINF1* 128
Figure 4.9 Chromatogram of a heterozygote for variant c.808G>A in COL1A2 140
Figure 4.10 Chromatogram of a heterozygote for variant c.915C>G in *TRIM17* 142
Figure 4.11 Chromatogram of a heterozygote for variant c.817_825dupGCTGGTCCT in *COL1A2* 145
Figure 4.12 Secondary structure predictions for pre-microRNA *mir-183* 146
Figure 4.13 Chromatogram of a heterozygote for variant c.441G>C in *SERPINF1* 149
Figure 4.14 Illustration of variant locations in *SERPINF1* 151
Figure 4.15 Chromatograms showing variants in *SERPINF1* 153
Figure 4.16 Image illustrating all 12 *SERPINF1* transcripts. 155
Figure 4.17 Graphical representation of read count in *SERPINF1* from RNA-seq data in human stapes 158
Figure 4.18 Image illustrating the process of splicing 159
Figure 4.19 Image illustrating the location of the upstream and downstream Taqman assay probes for all *SERPINF1* transcripts. 161
Figure 4.20 Graphs showing the relative expression of *SERPINF1* transcripts in otosclerotic and control stapes suprastructures 163
Figure 4.21 Hypothetical model of expression of *SERPINF1* in normal and otosclerotic human stapes in the presence and absence of known variants in *SERPINF1*. 167
Figure 5.1 Flow chart showing the four stages involved in RNA-seq 175
Figure 5.2 Venn diagram illustrating overlapping dysregulated genes from stapes suprastructure and footplate analyses 183
Figure 5.3 Bar graph showing five highest scoring networks from IPA analysis and their respective network scores. 184
Figure 5.4 Bar graphs showing top biological functions identified during IPA analysis 185
Figure 5.5 Venn diagram showing the dysregulated genes known to be involved in arthritis, osteosarcoma and pigmented villonodular synovitis 188
Figure 5.6 Bar graph showing the RPKM fold difference and p values as well as the z ratios calculated for each of the four linked genes located within the cluster on chromosome 9q22 193
Figure 5.7 Graph showing relative expression of 3 down-regulated genes in control and otosclerotic stapes 195
Figure 5.8 Graph showing relative expression of 2 up-regulated genes in control and otosclerotic stapes 196

Figure 6.1 Structure of the C-terminal region of the *RELN* gene and its three mRNA isoforms illustrating the positions of the Taqman® gene expression assay amplicon and isoform-specific primers 206
Figure 6.2 A 2% agarose gel showing *RELN* transcript variant 2 RT-PCR amplicons in human stapes cDNA. 207
Figure 6.3 *RELN* relative quantification graph. 208
Figure 6.4 Structure of the three *RELN* mRNA isoforms, illustrating the positions of primers designed by Schrauwen et al. and Csomor et al. 211
Figure 6.5 Haploview plot depicting linkage disequilibrium between the rs7341475 SNP associated with microexon skipping and 15 SNPs associated with otosclerosis 215
Figure 7.1 Schematic showing the possible effect of *SERPINF1* variants on TGFβ signaling via heparan sulphate mediated modulation of latent TGFβ. 220

List of Tables

CHAPTER 1

Table 1.1 Table of otosclerotic loci identified by linkage analysis 40
Table 2.1 Table showing numbers of human stapes samples collected 63
Table 2.2 Primers used during semi quantitative RT-PCR assays 65
Table 2.3 Table showing Applied Biosystems assay ID numbers for each Taqman® assay performed.....66
Table 2.4 Primers used for RT-PCR during genotyping and Sanger sequencing.....70
Table 2.5 Table showing predicted fragment sizes from restriction endonuclease digests 72
Table 3.1 Table of the number of questionnaire respondents infected by and vaccinated for measles mumps and rubella viruses 80
Table 3.2 Table of number and percentage of respondents who have lived outside the UK for over five years.....81
Table 3.3 Table showing number of years for which questionnaire respondents had taken HRT 83
Table 3.4 Table of number and percentage of respondents who reported history of autoimmune conditions 84
Table 4.1 Table showing demographics of members of family A 97
Table 4.2 Table showing demographics of members of family B 99
Table 4.3 Table showing demographics of members of family C 101
Table 4.4 Table showing demographics of members of family D 103
Table 4.5 Table of annotated data fields for WES variants 105
Table 4.6 Table illustrating the number of variants retained per family at each stage of the filtering process 109
Table 4.7 Table of websites used during annotation of variants and genes 112
Table 4.8 Table showing variants selected for segregation analysis in families A and B 115
Table 4.9 Table showing variants selected for segregation analysis in families C and D 116
Table 4.10 Table illustrating the 4 variants prioritised for segregation analysis in family A.....118
Table 4.11 Table showing genotypes of all members of family A 120
Table 4.12 Table illustrating the 8 variants prioritised for segregation analysis in family B.....125
Table 4.13 Table showing genotypes of all members of family B 127
Table 4.14 Table illustrating the 4 variants prioritised for segregation analysis in family C.....130

Table 4.15 Table showing genotypes of all members of family C.....132
Table 4.16 Table illustrating the 7 variants prioritised for segregation analysis in family D.....135
Table 4.17 Table showing genotypes of all members of family D 137
Table 4.18 Table showing the 10 variants identified in *SERPINF1*in the otosclerosis cohort....152
Table 4.19 Table showing read count data for *SERPINF1* exons.....156
Table 4.20 Table showing predicted impact on splicing of *SERPINF1* variants.....160
Table 5.1 A table illustrating the twelve stapes selected for RNA sequencing 178
Table 5.2 Table showing a summary of RNA-seq statistics for each sample.....180
Table 5.3 Table showing the number of up and down-regulated genes identified during RNA-seq.....182
Table 5.4 Table of the three top scoring gene clusters for both up and down-regulated genes from the RNA-seq dataset.....186
Table 5.5 Table showing the number of genes known to be involved in bone abnormalities, hearing anomalies and connective tissue disorders.....190
Table 5.6 Table illustrating the 6 genes prioritised for follow up analysis in the remainder of the cohort.....191
Table 5.7 Table showing the normalised read count data from RNA-seq analysis for each of the three up and down-regulated genes selected for qRT-PCR analysis and number and percentage of control and affected stapes samples in which each transcript was detected.....198
Table 6.1 Table showing the total number of human stapes suprastructure and footplate samples used for PCR assays and RNA-seq 204
Table 6.2 Table showing *RELN* transcript detection in human stapes 208
Table 6.3 Table showing the number of replicates in which *RELN* was detected in qRT-PCR assays in human stapes suprastructures as well as MG-63 cells.....209
Table 6.4 Table showing RPKM values from RNA-seq for *RELN* in four control and eight otosclerotic stapes 210
Table 6.5 Table showing the number of samples in which detection of *RELN* was consistent in both qRT-PCR assays and semi-quantitative PCR assays in human stapes suprastructures.....214

1 Introduction

1.1 Deafness and hearing loss

Hearing loss is the most common human sensory impairment, affecting over 5% of the world's population (www.who.org). It can manifest at any age and is thought to affect approximately six in every 1000 new born infants and about one third of those aged 65 and over (www.who.org). In the UK, it is estimated that ten million people are living with a hearing impairment (www.actiononhearingloss.org.uk) and it is expected that this number will increase by 10-15% in population terms over the next 15 years with the growing ageing demographic of the population (Davis et al., 2012).

A large proportion of those affected by hearing loss live in low and middle-income countries, which is partly due to the contribution of environmental factors to the development of deafness. Environmental contributors include noise exposure, particularly involving those working in manufacturing industries, bacterial and viral infection, and ototoxicity as a result of wide-spread use of certain antibiotics such as aminoglycosides. Whilst these environmental factors remain considerable problems in developing countries, the shift away from manufacturing industry in the UK and the development of new medications with reduced ototoxicity has resulted in a reduction of exposure to these agents. Despite this, incidence of hearing loss remains high in developed countries. This is in part a result of the genetic factors which contribute to hearing loss. It is estimated that two thirds of all cases of prelingual deafness are caused by known genetic factors, whilst unidentified genetic factors and the environment contribute to the remaining third (Hilgert et al., 2009). Deafness is genetically heterogeneous, caused by mutations in a multitude of genes which encode proteins with a variety of functions. To date, more than 50 deafness genes and 80 additional genetic loci have been identified (Dror and Avraham, 2010).

Despite recent advances in the identification of genes involved in hereditary deafness, there has been a lack of progress in elucidating the genetic basis of one of the most common forms of hereditary hearing loss; otosclerosis. This is an acquired form of adult-onset deafness caused by fixation of the stapes bone in the middle ear. The focus of this thesis is to investigate the genetic factors involved in the development of this condition. In this introductory chapter, the structure and function of the ear is discussed followed by an introduction to otosclerosis

and its clinical characteristics along with currently available treatment options. The development of the middle ear will then be outlined followed by an introduction to bone remodelling before an overview of progress made in investigating both environmental and genetic factors that contribute to disease pathogenesis. Finally I will summarise the key aims and objectives of this thesis, where for the first time, two Next Generation Sequencing technologies will be used in combination with one another in the hope of identifying genetic factors involved in the development of otosclerosis.

1.2 Structure and function of the ear

The mammalian ear is a complex organ which is composed of three main segments; the external, middle and inner ear. The role of the external ear, consisting predominantly of the auditory meatus and pinna, is to collect sound, amplify it, and direct it into the auditory canal. When the sound waves reach the tympanic membrane, which forms a barrier between the external and middle ear, it sets the membrane in motion resulting in the transfer of sound waves into mechanical movement (Lee, 1999). One of three small auditory ossicles known as the malleus which is located in the air-filled chamber of the middle ear, is attached firmly to the tympanic membrane. Motion of the membrane results in transmission of mechanical energy through the ossicular chain, passing from the malleus to the incus and finally to the stapes which is attached laterally to the incus and medially to the oval window. The stapes transfers its mechanical energy through the oval window and into the inner ear. In the inner ear, mechanical energy is transformed into electrical stimuli which are then transmitted via the auditory nerve to the auditory cortex of the brain (Figure 1.1 A).

1.2.1 Structure and function of the inner ear

The inner ear consists of the vestibular and auditory systems, which are required for balance and perception of sound respectively. These systems are located within a bony labyrinth known as the otic capsule. The auditory system comprises the cochlea, a snail-shaped fluid-filled tube which is coiled around a central bony modiolus. Three compartments known as the scala media, scala vestibuli and scala tympani make up the cochlear canal. The scala media is filled with endolymph, which has a unique electrolyte composition rich in potassium and low sodium ions (Wangemann, 1996) and is separated from the scala tympani by the basilar membrane (Figure 1.1 B).

The scala media contains the sensory epithelium of the cochlea known as the organ of Corti which sits on top of the basilar membrane. Sensory hair cells are located within the organ of Corti and are arranged in four rows; three rows of outer hair cells (OHCs) and one row of inner hair cells (IHCs). These are separated from one another by two rows of pillar cells and three rows of Deiter's cells, which together with Hensen's cells and the inner sulcus cells make up the organ of Corti's supporting cells. These surround the hair cells in order to support and maintain them (For review see Lim, 1986). When vibrations from the middle ear are transmitted to the inner ear, movement is generated in the fluid of the cochlea, activating the hair cells and converting the mechanical stimulation into an electrical signal (Figure 1.1 C).

Both inner and outer hair cells have projections of actin-rich filaments known as stereocilia on their apical surface. These are immersed in the endolymph of the scala media and are known as hair bundles. Mechanical activation of the stereocilia results in deflection of the hair bundles, which leads to the opening of mechano-electrical transduction channels located at the stereocilia tips. The synchronised opening of these channels results in an influx of potassium ions which depolarise the cell resulting in release of neurotransmitter at the base of the hair cells (Dallos, 1996). This generates an action potential in the nerves connected to the hair cells, which pass along the auditory nerve pathway to the brainstem (Figure 1.1D). When hair cells become damaged, the ability of the inner ear to transmit signals to the brain is impaired, leading to hearing loss) (For review see Guinan et al., 2012).

Loss and damage of hair cells occurs naturally during the process of ageing, leading to the progressive hearing condition presbycusis, more commonly known as age-related hearing loss. Presbycusis is the most common form of adult onset deafness affecting a third of people aged 65 and over (www.who.org). Hair cell damage can also be caused by a range of environmental factors including exposure to loud noise or ototoxic drugs, head trauma and viral infection. Genetic factors are also known to contribute to hair cell damage and may cause congenital deafness as well as childhood and adult-onset forms of the condition. Since hair cell loss is irreversible in humans, the resulting hearing impairment is permanent. Hearing loss due to hair cell damage, as well as damage to the auditory nerve which may be caused by benign tumours or head trauma, is known as sensorineural hearing loss. This is the most prevalent form of hearing impairment, responsible for approximately 90% of all cases and is associated with damage to the inner ear. The remaining 10% of cases are caused by conductive hearing loss, which occurs when sound waves are impeded from reaching the cochlea, and is associated with damage to the middle ear.

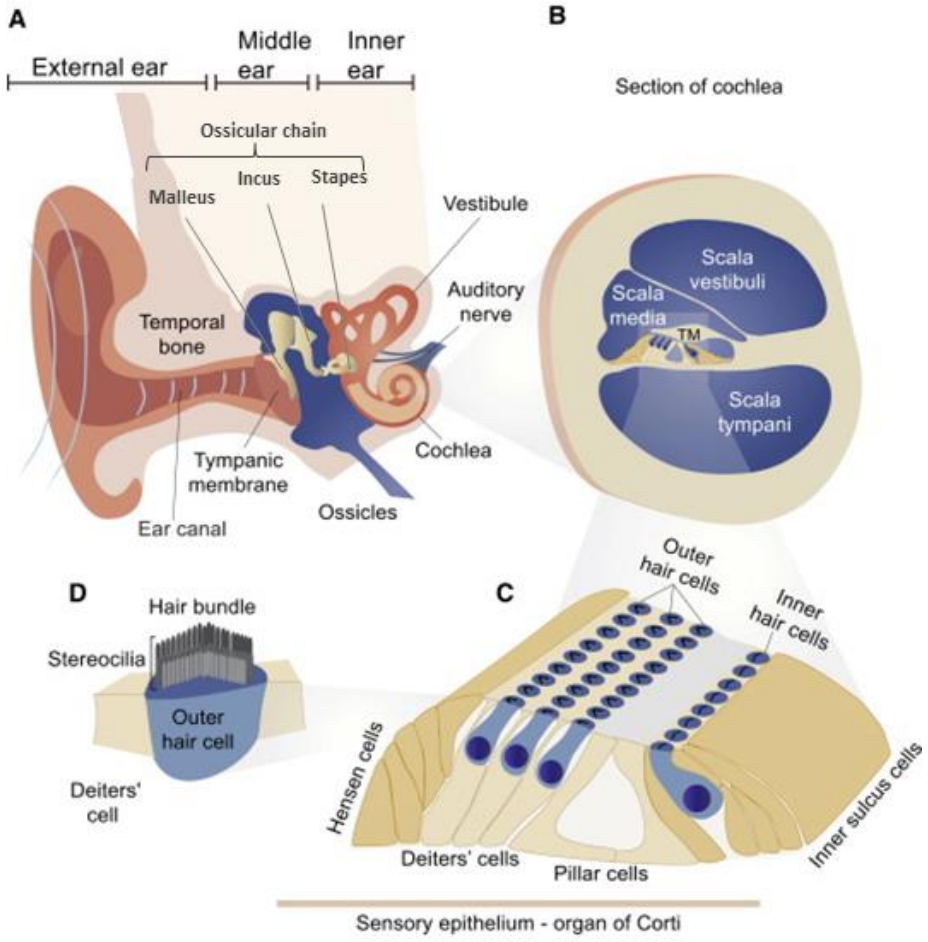


Figure 1.1 Anatomy of the Mammalian Ear

(A) The ear is composed of the external, middle, and inner ear. The cochlea, responsible for hearing, and the vestibule, responsible for balance, make up the inner ear. (B) A cross-section of the cochlear duct reveals the scala media, scala tympani and scala vestibuli that are filled with fluids over the organ of Corti. (C) An enlargement of the organ of Corti, the cochlear sensory epithelium, showing three rows of OHCs and one row of IHCs, flanked by various types of supporting cells. (D) A diagram depicting the outer hair cells and the hair bundles of stereocilia on their apical surface (Adapted from Dror and Avraham, 2010).

1.2.2 Structure and function of the middle ear

The role of the middle ear is to transmit and amplify sound from the external to inner ear. All animals relay sound via the middle ear, an air filled cavity containing the suspended auditory ossicles whose role it is to pass sound from the tympanic membrane to the basilar membrane. Unlike non-mammalian vertebrates which have a single ossicle called the columella, vertebrates have three ossicles arranged in an ossicular chain, which serve the same function. These ossicles are called the malleus, the incus and stapes (Figure 1.2). The malleus is inserted into the tympanic membrane at the distal end of the chain and is linked to the incus by the incudo-malleal joint. The incus acts as a bridge to the stapes, which link together at the incudo-stapedial joint (Amin and Tucker, 2006). The footplate of the stapes is inserted into the oval

window of the otic capsule and is held in place via the annular ligament, which allows the stapes to move. The stapes acts as piston, transducing sound waves from the air-filled cavity to the endolymph of the scala media, where mechanotransduction occurs. Since the tympanic membrane has a much larger surface area than the oval window, pressure is exerted on the stapes footplate which results in amplification of the signal (for review see Chapman, 2011).

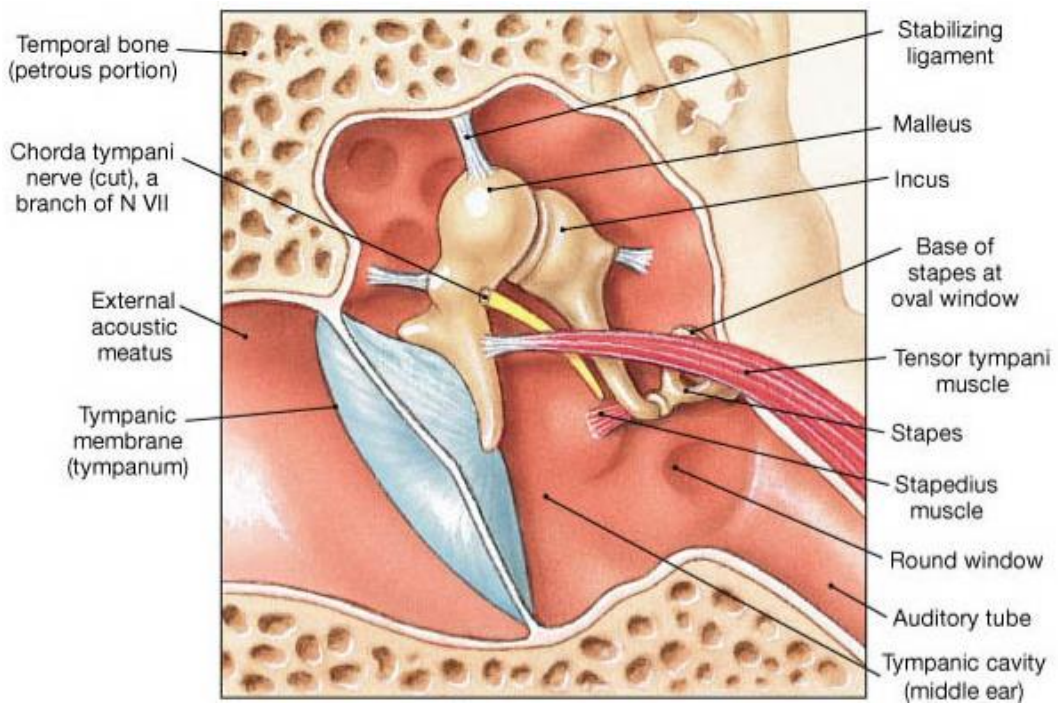


Figure 1.2 Middle ear anatomy

The middle ear transmits movement of the tympanic membrane to the inner ear via the three auditory ossicles; the malleus, incus and stapes, which are held in position by the ligaments and muscles of the inner ear. (From http://droualb.faculty.mjc.edu)

Conductive hearing loss occurs when the middle ear is unable to transfer sound from the outer to the inner ear. The most common form of adult-onset conductive hearing loss is otosclerosis which is caused by abnormal bone remodelling within the otic capsule leading to fixation of the stapes (For review see Ealy and Smith, 2011). Another common cause of conductive hearing loss is the presence of a physical barrier which obstructs the transmission of sound waves, such as a build-up of fluid or ear wax in the external ear canals, or due to chronic ear infection such as otitis media (For review see Pichichero, 2013). In addition, a number of rare syndromic conditions may also cause conductive hearing loss by affecting bone remodelling within the ear, including deafness due to Osteogenesis Imperfecta (Verstreken et al., 1996) and stapedial ankylosis. Conductive hearing loss may also be caused by a congenital defect as a result of abnormal development of the external and middle ear in the developing foetus, and affects approximately one in every 3,000 to 5,000 new-borns (Chapman, 2011).

1.3 Otosclerosis

Otosclerosis is a condition that was first described by Adam Politzer in 1893 and is characterised by abnormal bone remodelling within the otic capsule, the skeletal element that encloses the inner ear (Nager, 1969). It occurs when foci of spongy, irregular bone, invade the stapediovestibular joint, interfering with free motion of the stapes bone in the middle ear. As the otosclerotic foci advance, the stapes becomes fixed in place, resulting in progressive conductive hearing loss. The stirrup-shaped stapes is the smallest and lightest bone in the human body and is composed of two distinct parts; the stapes suprastructure and footplate, which sits in the oval window niche. The suprastructure consists of two arms (crus) which surround the stapedial foramen, joining at the base to form the stapedial footplate (Figure 1.3).

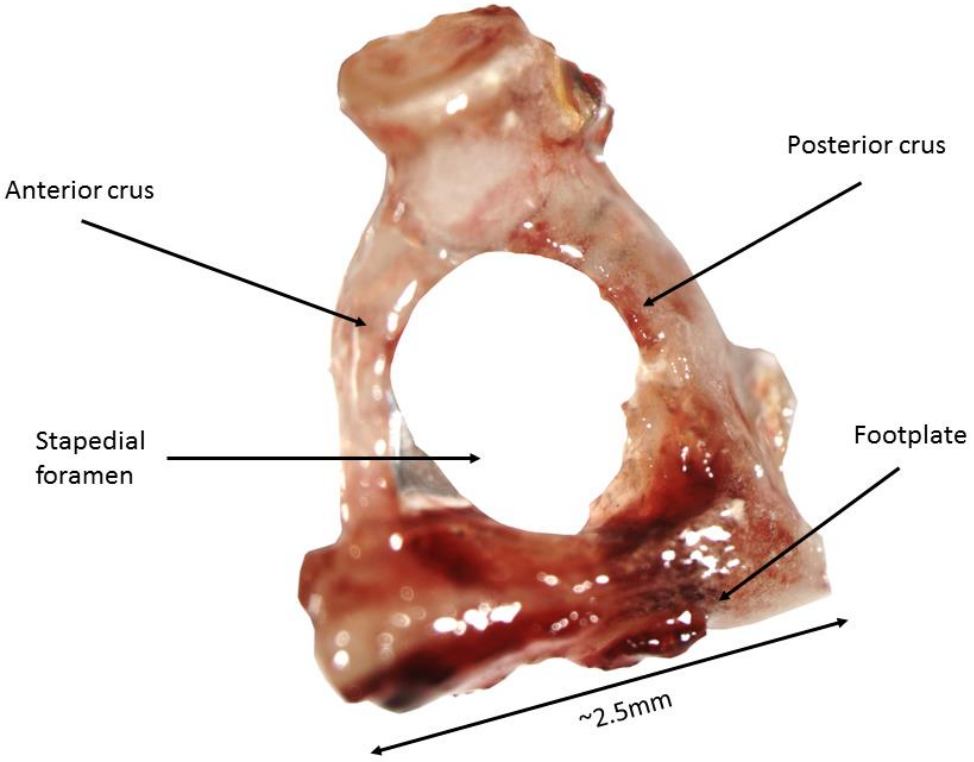


Figure 1.3 Bright field microscope image of a human stapes bone

Image depicts the stapes footplate and the suprastructure. The suprastructure consists of the anterior and posterior crus which join the footplate, forming the stapedial foramen through the centre of the stapes bone. (Produced for the purpose of this thesis in July 2013.)

1.3.1 Clinical characteristics of otosclerosis

Otosclerosis is the most common form of adult-onset conductive hearing loss, affecting approximately 0.3-0.4% of the Caucasian population (Pearson et al., 1974, Gapany-Gapanavicius, 1975, Declau et al., 2007b). There is considerable variation in incidence figures

cited by different investigators, which may be due to the varying incidence of this condition between different ethnic groups. Otosclerosis most commonly affects Caucasians, whilst the incidence is lower in black African and Oriental populations (Declau et al., 2007a). A particularly high incidence has been noted in the Todas population in Southern India (Kapur and Patt, 1967). Another possible reason for the discrepancy in incidence figures is that over the last two decades, it has become apparent that the number of patients seeking treatment for otosclerosis has fallen dramatically (Arnold, 2007). This has led some clinicians to conclude that the incidence of the condition is falling. However, a conflicting clinical opinion is that after the introduction of stapes surgery in the late 1950s, the large 'back-log' of otosclerosis patients seeking treatment was cleared. Therefore, the current situation is probably a more clear reflection of the true incidence of the disease (Saeed, 2002).

The age of onset of otosclerosis is variable, but typically the first symptoms present in the third decade of life. Although rare cases have been reported in early childhood, this is very uncommon. Development of otosclerosis after the age of 50 is unusual but not unheard of, and cases have been reported into the sixth decade of life (Gapany-Gapanavicius, 1975).

For many years, clinicians observed fewer cases of otosclerosis in males than females and an extensive study by Davenport et al. in the 1930s concluded that otosclerosis was twice as common in women as in men (Davenport, 1933). However, subsequent investigations in the 1960s found that the condition was equally common in both sexes (Larsson, 1960, Morrison, 1967). It is now thought that the condition exhibits a sex ratio of 1:1, although females are more likely to be seen in clinical practice due to a higher incidence of bilateral disease and progression of hearing loss during pregnancy due to hormonal changes (Browning and Gatehouse, 1992). Other theories include that women are more likely to consult their General Practitioner regarding hearing loss, resulting in more referrals to the clinic, and that women are less likely to work in environments where they are exposed to noise induced damage and so are more likely to notice and report a deterioration in their hearing (Browning and Gatehouse, 1992, Smyth, 1997).

Hearing loss is bilateral in approximately 70-85% of otosclerosis cases and is usually asymmetrical, developing initially in one ear before progressing to the other (Emmett, 1993). The progression of hearing loss may be interrupted by periods of stability or rapid deterioration. The majority of patients with otosclerosis experience solely conductive hearing loss as a result of stapes fixation, however, approximately 10% of clinical cases also have a sensorineural component to their deafness (Browning and Gatehouse, 1984), which is more difficult to treat. It has been suggested that sensorineural hearing loss occurs when toxic

proteolytic enzymes are released from the otosclerotic foci, affecting the inner ear (Schrauwen and Van Camp, 2010). It is thought that sensorineural hearing loss occurs in advanced stage disease, although this belief is somewhat contentious. About 80% of otosclerosis patients experience tinnitus, which has been attributed to cochlear involvement. It is possible that tinnitus is a secondary symptom to otosclerosis caused by amplification of sound following hearing loss. Imbalance and vertigo are less common than tinnitus in individuals with otosclerosis, but may be caused by the proteolytic enzymes from otosclerotic foci affecting the vestibular system (Smyth, 1997).

Otosclerosis patients often experience Paracusis Willisii, the ability to hear speech better in noisy surroundings, an indicator of conductive hearing loss. This contrasts with typical sensorineural deafness and probably occurs because noisy environments encourage those with normal hearing to speak louder, above the hearing threshold of the person with otosclerosis (Smyth, 1997). Otosclerosis patients often have quiet speech, as they are able to hear their own voice by bone conduction through the skull, rather than through air conduction (Emmett, 1993).

1.3.2 Histologic otosclerosis

It is crucial that a distinction is made between 'clinical otosclerosis' and 'histologic otosclerosis'. The former is the clinically relevant form of the condition where otosclerotic lesions interfere with stapes motion leading to conductive hearing loss. In contrast, the latter is identified post-mortem, when otosclerotic lesions are found in the temporal bone of symptom-free individuals (for review see Thys and Van Camp, 2009). In 1967, research by Altmann et al. concluded that approximately 8.3% of Caucasians possess histologic otosclerotic lesions at death, of which 12% are located at the stapediovestibular joint, indicative of clinical otosclerosis. However, if the statistics derived from this study are used to calculate the population frequency for clinical otosclerosis, the frequency would be estimated at 0.996%, substantially greater than the 0.3-0.4% indicated from clinical data (Declau et al., 2007b). One possible explanation for this discrepancy is that the temporal bone collections analysed during the Altmann et al. study were biased for diseased otologic tissue, resulting in an overestimate of the frequency of histologic otosclerosis. This explanation is supported by a more recent analysis of 236 temporal cadaveric bones, which concluded that histologic otosclerosis is present at a frequency of 2.5% in the population (Declau et al., 2007b). Using the statistics derived from this study, the frequency of clinical otosclerosis would be predicted to be 0.3%, which supports the clinical data.

1.3.3 Histology of otosclerosis

The terms otosclerosis and otospongiosis are often used interchangeably in the published literature. Otospongiosis is the descriptive term coined by Siebenmann in 1912 during the identification of spongiform foci in temporal bone of otosclerosis patients during post-mortem analysis. These foci are pleomorphic in nature and vary from those consisting of spongiform bone accompanied by large fibrous regions to those of dense sclerotic bone (Linthicum, 1993). In all cases, the composition of the lesion is highly disorganised in comparison to the normal bone that it replaces. Lesions can be categorised into three histopathologic groups; active, intermediate and inactive, although it is possible that all may be present within a single lesion at any one time. Active lesions, also known as spongiotic lesions, are typified by abnormal vasculature and enlarged spaces within the bone marrow, as well as finger-like projections, which extend into the surrounding bone (Linthicum, 1993). In contrast, inactive lesions are characterised by sclerotic mineralised bone, which does not appear to have undergone any recent bone remodelling and which is largely acellular in composition (Linthicum, 1993). Intermediate stage lesions have features common to both active and inactive stages. It is generally considered that inactive lesions occur in later stage disease.

In otosclerosis, these spongiform foci develop in a variety of locations around the temporal bone and middle ear ossicles, as well as in the oval window niche. Early studies in the 1940s indicated that 80-90% of otosclerotic foci occur in the oval window niche whilst the round window is affected in 30-40% of cases(for review see Bretlau and Jorgensen, 1968). The most common location for otosclerotic lesions is at the anterior portion of the oval window at a cleft known as the fissula ante fenestram. As the focus gradually expands, the anterior portion of the stapes footplate is enveloped, causing it to become fixed (Figure 1.4 A). Occasionally, a large mass of otosclerotic bone may grow so large that it fills the oval window niche and may obscure the footplate entirely. This rare form of the condition is known as obliterative otosclerosis (Figure 1.4 B).

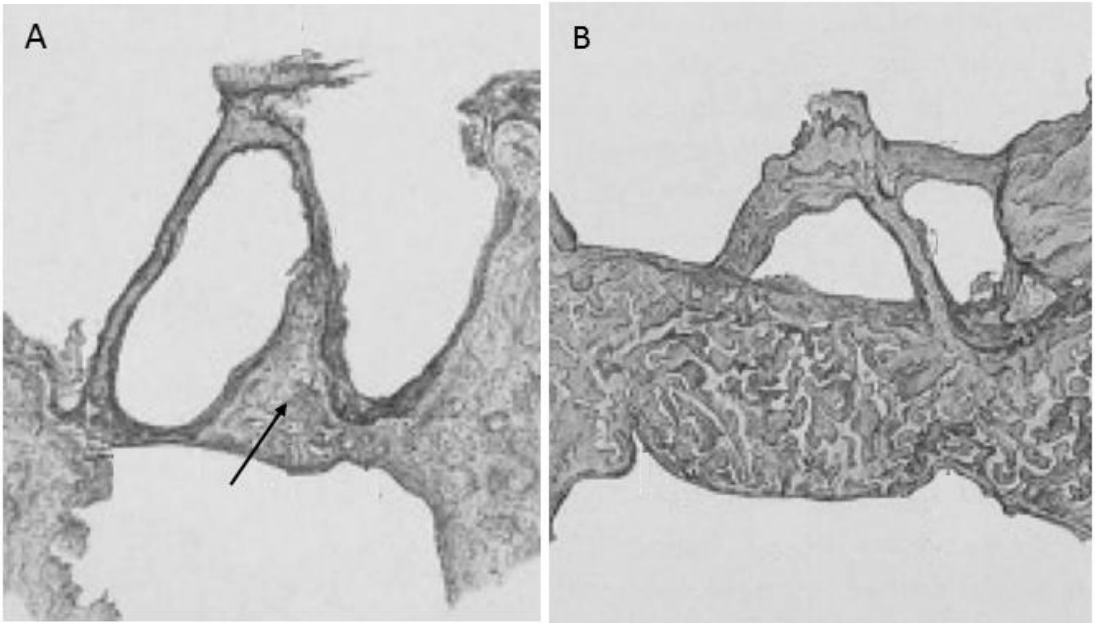


Figure 1.4 Photographs by Politzer 1894 depicting stapes fixation in otosclerosis.

(A) Otosclerotic lesion (indicated by arrow) at the fissula ante fenestrum leads to anterior fixation of the stapes bone (B) Otosclerotic lesion obscuring the stapes footplate causing obliterative otosclerosis. (Politzer, 1894).

Otosclerosis can be categorised into four clinical sub-types based on the location and the nature of the lesion in relation to the stapes footplate. A Type I footplate exhibits fixation in the absence of an otosclerotic focus, a Type II footplate is characterised by stapes fixation accompanied by a single small otosclerotic focus, a Type III footplate occurs when the lesion involves approximately 50% of the footplate, and a Type IV footplate is characterised by obliteration of the entire annular ligament and oval window niche by the invading lesion (Tange et al., 1998).

1.3.4 Diagnosis of otosclerosis

Otosclerosis is typically diagnosed following audiometric analysis together with physical examination. Audiometric analysis is used to identify the presence of the conductive hearing loss characteristic of otosclerosis, whilst physical examination is crucial in order to rule out other forms of conductive hearing impairment such as otitis media or perforation of the tympanic membrane.

The human ear is able to detect frequencies in the range of 20 Hertz (Hz) to 20 kHz, with individuals with normal hearing usually able to hear frequencies from 125 Hz to 8 kHz at intensities between 0 and 10 decibels (dB). Mild, moderate, severe and profound hearing loss

occur when the hearing threshold increases to 20-40 dB, 41-70 dB, 71-90 dB and >90 dB respectively (Figure 1.5 A).

Otosclerosis patients typically display a progressive low-frequency hearing loss that can be detected on an audiogram (Figure 1.5 B). However, they are often not aware of a problem at the early stages of disease. This is because speech perception occurs at higher frequencies. As the condition progresses and the stapes becomes fixed, higher frequencies also become affected. Otosclerosis patients therefore often do not become aware of their hearing loss until the condition has progressed. When a patient presents with symptoms characteristic of otosclerosis in the clinic, audiometric analysis is carried out. During this process both air conduction and bone conduction are analysed. Air conduction measures the sound waves which pass through the external and middle ear into the cochlea, whilst bone conduction measures the conduction of sound to the inner ear via the bones of the skull. Conductive hearing loss is characterised by a reduction in hearing thresholds of air conduction, whilst bone conduction remains unaffected, resulting in the development of an air-bone gap (ABG). As the condition progresses and the stapes becomes fixed, higher frequencies become affected causing the ABG to increase. Conductive hearing loss is diagnosed when the ABG exceeds 10 dB. For otosclerosis patients who don't have sensorineural hearing loss, the hearing level usually increases to 60-65 dB and frequently displays a characteristic Carhart's notch (Carhart, 1962) on the audiogram where the hearing threshold reduces at about 2 kHz (Figure 1.5 B).

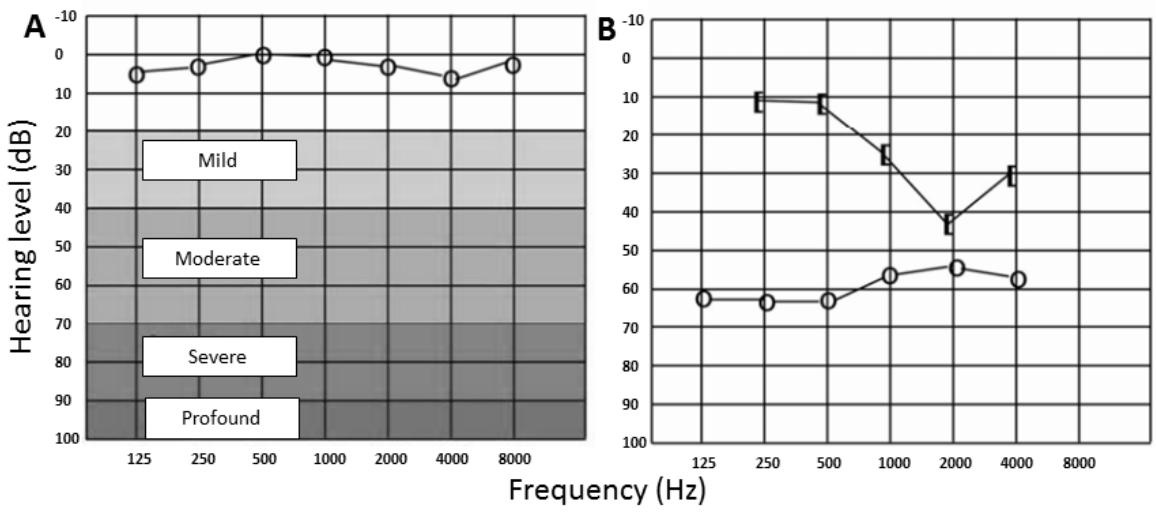


Figure 1.5 Normal hearing and otosclerosis audiograms.

(A) Hearing loss can be classified as mild, moderate, severe, or profound. A normal hearing person typically can hear frequencies spanning 125 Hz to 8,000 Hz at intensities between 0 and 10 dB. (B) Typical audiogram of an individual with otosclerosis. A gap between air conduction (o) and bone conduction ([ ) is seen in these individuals. A characteristic Carhart's notch is evident at the 2,000 Hz frequency (Audiogram based upon image from Uppal et al., 2009).

Upon physical examination, the tympanic membrane may display a characteristic pink tinge due to the vascular nature of the lesion below. This is often referred to as the Schwartz sign. Although physical examination combined with audiometric analysis is usually effective at diagnosing otosclerosis, the most definitive diagnosis is confirmation during stapes surgery.

1.3.5 Treatment for otosclerosis

There are several treatment options for a patient once diagnosed with otosclerosis including both non-surgical and surgical options. The choice of treatment will be determined by the patient's personal circumstances, age and general health and will be decided upon during a consultation between the otosclerosis patient and an otolaryngologist.

1.3.5.1 Non-surgical management of otosclerosis

Some patients may choose not to actively treat their hearing loss, but instead to monitor its progression. This is most likely to be suitable for elderly patients or those in poor health. It may also be suitable for patients with unilateral disease who do not find that their hearing loss considerably impacts their quality of life. These patients will be regularly monitored for signs of progression of hearing loss, and in the case of those with unilateral disease, for deterioration in their other ear. If progression of symptoms were to occur, these patients may choose to discuss alternative treatment options with their otolaryngologist.

Other otosclerosis patients may opt for a hearing aid. This is predominantly suitable for those with losses of 60dB or less and with purely conductive hearing loss. Although patients with a sensorineural component to their deafness may experience problems with amplification when using hearing aids, some may benefit from their use (Johnson, 1993). Compression-type hearing aids may also be used by patients following successful stapes surgery. The advantage of using hearing aids over other treatments is the lack of associated side-effects. However despite this, few otosclerosis patients in the UK opt for hearing aids, favouring alternative treatments.

Fluoride therapy is another option for patients who cannot or choose not to undergo surgery. In the 1960s, it was suggested that treatment with sodium fluoride may slow the progression of otosclerosis symptoms (Shambaugh and Scott, 1964) and a number of case-control studies in the 1980s indicated that there was less hearing deterioration in otosclerosis patients who had been treated with sodium fluoride (Bretlau et al., 1985, Bretlau et al., 1989). However this

correlation has not been supported by all studies (Vartiainen and Vartiainen, 1997a, Vartiainen and Vartiainen, 1997b). A recent review of the literature published since the 1960s concluded that there is some evidence to suggest that sodium fluoride may be of benefit to preserve hearing in otosclerosis patients but that the quality of evidence in some of these studies is weak (Cruise et al., 2010). Currently, sodium fluoride treatment is not common practice in the UK although it is recommended in some rare cases.

Bisphosphonates, which are known to affect bone remodelling, have also been used in the treatment of otosclerosis, specifically for patients with a sensorineural component to their hearing loss. The benefits of first-generation bisphosphonates such as etidronate were investigated in the early 1990s, but were not found to have a significant impact on progression of otosclerosis (Kennedy et al., 1993). However, over the past several years, nitrogen-containing bisphosphonates such as risedronate and zoledronate which offer more potent bone remodelling inhibition than early generation bisphosphonates have been developed. A recent study using these drugs indicated that progression of sensorineural hearing loss in otosclerosis was reduced following treatment (Quesnel et al., 2012). However, like fluoride therapy, bisphosphonate treatment is not common practice in the UK.

1.3.5.2 Surgery for otosclerosis

In the UK, stapes surgery is the most common treatment option for conductive otosclerosis. Most surgeons opt for laser-assisted stapedotomy surgery, which involves the surgical removal of the stapes suprastructure and insertion of a piston-like prosthesis which connects the incus to the inner ear. Stapedectomy surgery is a less common alternative where the whole stapes bone is excised and replaced with a micro prosthesis. This technique was used predominantly prior to the introduction of stapedotomy and remains common practise in some countries. Although the choice of technique is dependent on the surgeon and patient, evidence suggests that stapedotomy is less likely to result in complications such as uncontrolled leakage of perilymph and damage to the inner ear than stapedectomy (Thamjarayakul et al., 2010). Other complications associated with stapes surgery include dizziness following the procedure, which is usually transient though occasionally may be permanent, and in rare cases total loss of hearing may occur in the operated ear, although this occurs in less than 1% of cases. There is also a minimal risk of facial palsy due to the close proximity of the facial nerve and the stapes, which should be discussed with patients prior to surgery. Recent studies have suggested that the stapedotomy procedure can be further improved by use of a CO_2 laser in conjunction with a scanner system, which reportedly increases accuracy of the technique and reduces the risk of

irradiation of inner ear structures (Albers et al., 2011). Although stapes surgery is very effective and in most cases improves or fully restores hearing, disease recurrence may occur if complications are encountered. In such cases, revision surgery may be required. In otosclerosis patients with bilateral hearing loss, the surgeon will usually choose to operate on the most severely affected ear first and will wait until the outcome of the surgery is known before deciding whether or not to proceed with stapes surgery on the second ear.

1.4 Development of the stapes

The middle ear is a composite organ formed from all three primary germ layers; the endoderm, mesoderm and ectoderm of the developing embryo, as well and the neural crest, a pluripotent cell population which is derived from the neural plate during early embryogenesis (Couly et al., 1993). During embryonic development, the neural crest cells disperse from the surface of the neural tube giving rise to a variety of differentiated cell types. At the initiation of craniofacial development, the neural crest cells begin to populate the branchial arches which give rise to various skeletal structures of the head and neck including the jaw, larynx, trachea and the auditory ossicles. The way in which the auditory ossicles develop in the middle ear of the human embryo is unclear and this is particularly true of the stapes (Rodriguez-Vazquez, 2005). Whilst it is thought that the malleus and incus develop from the first branchial arch, the stapes is thought to develop from the second.

A number of theories have been proposed to explain stapedial development in the embryo. The one most widely accepted is that the stapes develops from a unique source of cartilage within the second branchial arch known as Reichart's cartilage (Hamilton, 1975). However, it has also been proposed that the stapes originates from a double source, where the stapes suprastructure develops from Reichert's cartilage and the footplate from the tissue of the otic capsule (Cauldwell, 1942). This theory was supported by a recent study which used transgenic reporter mice to demonstrate the dual origin of the stapes. This study found that the stapes suprastructure and base of the stapes develops from tissues of the neural crest, whilst the outer rim of the stapes footplate develops from the mesoderm (Thompson et al., 2012). This supports the concept of a dual origin for the stapes, indicating that development of the stapes footplate is intricately linked with development of the otic capsule.

1.4.1 Development of the otic capsule

The otic capsule is the cartilaginous structure which surrounds and protects the mechanism of the inner ear. It initially develops through the process of chondrogenesis in response to growth factors secreted from the developing embryonic otocyst, such as transforming growth factor beta 1 (*TGFβ1*) and other members of the transforming growth factor beta superfamily including bone morphogenetic proteins 2 and 4 (*BMP2* and *BMP4*) (Thompson et al., 2012). Chondrogenesis of the otic capsule is followed by endochondral ossification, where bone develops via the calcification of cartilage. Ossification begins to appear around the round window at gestational week 16.5 and by 22 weeks of gestation covers the cochlea and vestibule (Declau et al., 1989). Remnants of embryonic cartilage known as *globuli interossei* remain in the otic capsule into adulthood (Schuknecht and Gacek, 1993). These *Globuli interossei* express collagen type II genes (Niedermeyer et al., 2007) and have been proposed to predispose the otic capsule to development of otosclerotic lesions (Wang et al., 1999).

1.5 Bone remodelling in the otic capsule

Bone remodelling is the dynamic process responsible for turnover and renewal of the skeleton which is coordinated by osteoblast and osteoclast cells, which form and resorb bone respectively. The remodelling cycle involves removal of mineralised bone which is carried out by osteoclasts, followed by production of new bone matrix by osteoblasts which subsequently become mineralised resulting in the production of osteocytes, the most commonly found cell in mature bone. The cycle consists of a series of overlapping phases including osteoclast recruitment, bone resorption, and bone formation. The adult human skeleton experiences bone turnover at a rate of approximately 10% per year which is crucial for maintenance and to repair damage to the bone matrix. However, unlike the other bones in the human skeleton, the otic capsule normally undergoes limited remodelling after development, with turnover rate of just 0.13% per year (Frisch et al., 2000).

During bone remodelling, mesenchymal stem cells are recruited to the site of damage and produce macrophage colony stimulating hormone which stimulates the differentiation of cells from the haematopoetic stem cell lineage into mononuclear cells or pre-osteoclasts (Figure 1.6). These express the receptor activator of NFκB known as RANK on their surface. Differentiation factors such as runt-related transcription factor 2 (RUNX2), cause the Mesenchymal Stem Cells to differentiate into pre-osteoblasts (Figure 1.7). These express the RANK ligand (RANKL) on their surface which interacts with the RANK membrane receptors on the surface of pre-osteoclast cells, stimulating differentiation into mature osteoclasts. This

process of osteoclastogenesis is triggered indirectly by the production of parathyroid hormone in response to damage. Parathyroid hormone stimulates the up-regulation of RANKL on the surface of pre-osteoblasts, which results in increased binding with RANK on the surface of pre-osteoclasts, activating the signalling pathways that lead to osteoclast differentiation. Mature osteoclasts are large multinucleated cells that release the lysosomal enzymes that trigger bone resorption. The mature osteoclasts resorb bone on the matrix surface before undergoing programmed cell death via apoptosis. As the bone matrix is resorbed, it is replaced by the action of osteoblasts. Osteoblasts develop from pre-osteoblasts when stimulated by osteoblastogenesis inducing factors such as bone morphogenetic proteins and Wnt ligands. These inducing factors suppress the transactivation of peroxisome proliferator-activated receptor γ (PPARγ), a transcription factor that supresses osteoblastogenesis (Takada et al., 2009) (Figure 1.7). Mature osteoblasts secrete osteoprotegerin (OPG), a member of the TNF receptor superfamily which can also bind to the RANK ligand, out-competing RANK and thus inhibiting osteoclast activation. The rate of bone turnover can therefore be controlled by the OPG/RANKL ratio. Mature osteoblasts express type 1 collagens and bone sialoprotein, which are key components of the bone matrix. As they lay down the matrix, some osteoblasts become embedded as it calcifies. These osteoblasts become osteocytes, the most common cell within bone (For review see Hadjidakis and Androulakis, 2006) (Figure 1.6).

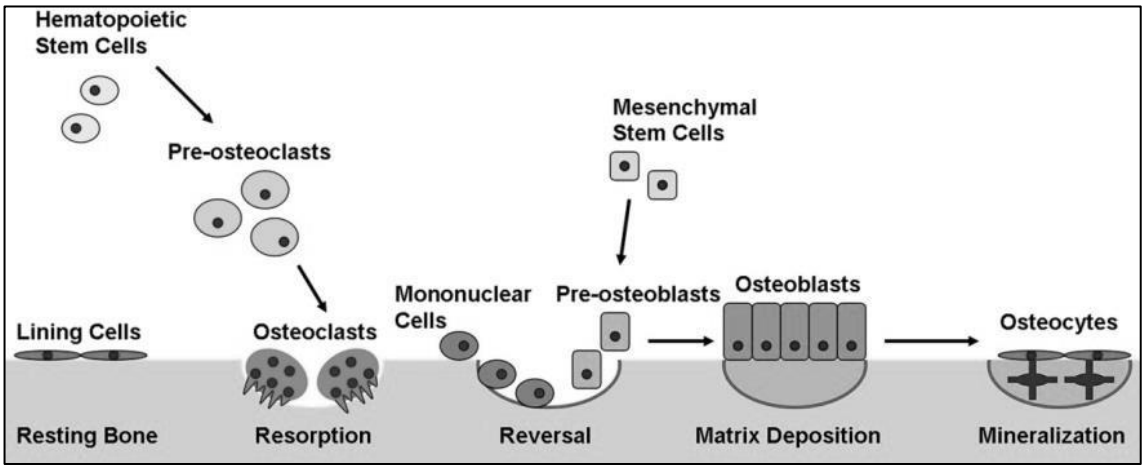


Figure 1.6 The bone remodelling cycle

Remodelling begins when osteoclasts resorb bone mineral and matrix. Mononuclear cells prepare the resorbed surface for osteoblasts, which generate newly synthesised matrix as they differentiate. Matrix mineralisation and the differentiation of some osteoblasts into osteocytes completes the remodeling cycle (From Kapinas and Delany, 2011).

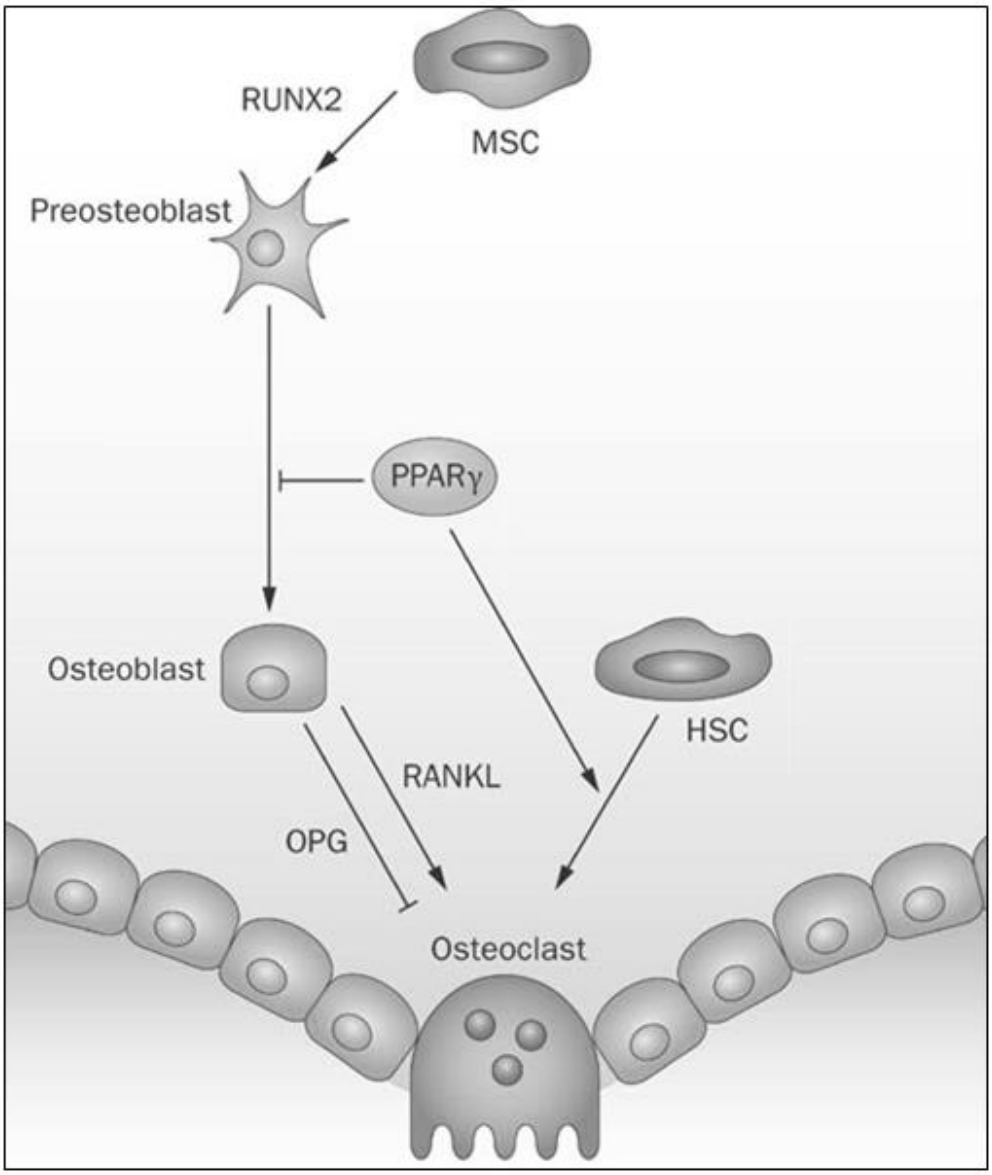


Figure 1.7 Molecular interactions involved in bone remodelling

The first step in bone remodeling involves commitment by Mesenchymal Stem Cells (MSCs) to the osteoblastic lineage. RUNX2 is required for MSC differentiation into preosteoblasts. PPARγ inhibits osteoblastogenesis, reducing the number of osteoblasts in the bone marrow. PPARγ also stimulates osteoclastogenesis from Haematopoetic Stem cell (HSCs). The OPG/RANKL ratio regulates the balance between osteoblastogenesis and osteoclastogenesis (Adapted from Kawai et al., 2009).

It has been suggested that the reduced rate of bone turnover in the otic capsule relative to other skeletal structures is due to the production of the osteoclastogenesis inhibitory factor OPG by the cochlea. It is thought that when OPG is secreted into the cochlear perilymph, it can diffuse into the otic capsule inhibiting remodelling (Zehnder et al., 2005). In the absence of OPG, mice experience activation of bone remodelling in the otic capsule and subsequent progressive hearing loss (Zehnder et al., 2006). This suggests that OPG is a crucial factor for regulation of bone turnover within the otic capsule.

1.6 Animal models for otosclerosis

Otosclerosis is a condition unique to humans affecting exclusively the human temporal bone for which no representative animal model has been proposed. The aforementioned (section 1.5) OPG knock-out mouse is possibly the best candidate animal model for otosclerosis suggested to date. These mice experience activation of bone remodelling in the otic capsule and display well- defined regions of bone resorption and deposition within the otic capsule, similar to those seen in otosclerosis. However, these mice also experience active bone remodelling throughout the skeleton including the other auditory ossicles, which is not seen in otosclerosis. Furthermore, stapes fixation is not observed in these mice (Zehnder et al., 2006). Despite the differences between the OPG knock-out mouse and otosclerosis, it is possibly the best model available for studying abnormal bone remodelling within the otic capsule, which resembles otosclerosis.

In addition to the OPG knock-out mouse, two additional animal models which display abnormal bone remodelling within the otic capsule were proposed in the past. In the 1980s, the LP/J mouse which spontaneously develops abnormal bony lesions in the auditory ossicles and otic capsule that are histologically similar to those found in otosclerosis, was suggested as an appropriate animal model for otosclerosis. Like in many humans with otosclerosis, these mice have both conductive and sensorineural hearing loss, but in addition, they possess an accompanying inflammatory reaction that is not seen in otosclerosis. In addition, they display an autosomal recessive inheritance pattern which contrasts with the autosomal dominant inheritance pattern with incomplete penetrance that is seen in otosclerosis (Henry and Chole, 1987).

In 1990, the incisor-absent rat was also proposed as suitable model for otosclerosis. This rat displays progressive conductive hearing loss which manifests at puberty accompanied by bony abnormalities of the ossicles and oval window. These bony lesions are located at the incudostapedial joint and stapes footplate, and appear histologically similar to those seen in otosclerosis. However, although there are some similarities in abnormal otic capsule remodelling between the incisor-absent rat and otosclerosis, remodelling is not localised to the temporal bone (Kaniff et al. 1990).

Despite a certain degree of resemblance between the phenotype of otosclerosis and each of the OPG knock-out mouse, LP/J mouse and the incisor-absent rat, none represent an ideal animal model for studying this condition. A lack of an appropriate animal model to study is just one of the numerous reasons why the pathophysiology of otosclerosis is yet to be elucidated.

1.7 The disease aetiology of otosclerosis

Although otosclerosis is one of the most common forms of adult-onset hearing loss, little is understood about its pathophysiology. Since the 19th century, when familial otosclerosis was first described, otosclerosis has been known to have a genetic component (for review see Ealy and Smith, 2011). Approximately 50% of all otosclerosis cases reportedly occur in individuals with a family history of the condition, whilst sporadic incidents in individuals who lack a positive family history comprise the remainder of cases (Cawthorne, 1955). Although this figure of 50% is broadly accepted amongst clinicians as the proportion of otosclerosis cases that are familial, there is no recent evidence in the current literature to substantiate this, so it should be considered as an indication of familial incidence rather than as fact. The most plausible explanation for the occurrence of otosclerosis in both individuals with a family history of the condition and those without, is that single gene defects are predominantly responsible for the development of otosclerosis within families, whilst complex inheritance, where a combination of genetic and environmental factors interact leading to the disease phenotype, is responsible for non-familial cases (Gordon, 1989). Despite this, there is likely to be considerable overlap between the genetic and environmental factors involved in familial and non-familial disease. For example, although familial otosclerosis tends to exhibit monogenic autosomal dominant inheritance (Albrecht, 1922), other modes of inheritance have also been suggested including digenic recessive inheritance and digenic inheritance of one X-linked dominant gene and one autosomal recessive gene (Moumoulidis et al., 2007), indicating that multiple genes may be involved in development of the condition in some families. In addition, reduced penetrance of otosclerosis is common in affected families with a reported frequency of approximately 40% in some studies (Larsson, 1960, Morrison, 1967), suggesting that environmental factors or additional modifier genes may be involved in the disease process.

Monogenic inheritance and complex inheritance should therefore not be considered two distinct categories but rather a continuous spectrum. In general, genetic variants responsible for monogenic inheritance of a disease are very rare but have a large effect size which results in the disease phenotype. In contrast, variants that contribute to complex diseases are likely to be more common but tend to have smaller effect sizes, so numerous small effect variants are required for the disease phenotype to result (Figure 1.8). In otosclerosis, there is some evidence to suggest that both common and rare variants, as well as environmental factors are involved in the disease process, but the extent to which these contribute to disease pathogenesis has not been defined. A number of large families with monogenic inheritance of otosclerosis have been identified in linkage analysis studies, indicating that rare variants with

large effect sizes contribute to the disease process, although no definitive genes have been identified. In addition, common variants within otosclerosis candidate genes have been proposed through genetic association studies, although their effect sizes remain unknown.

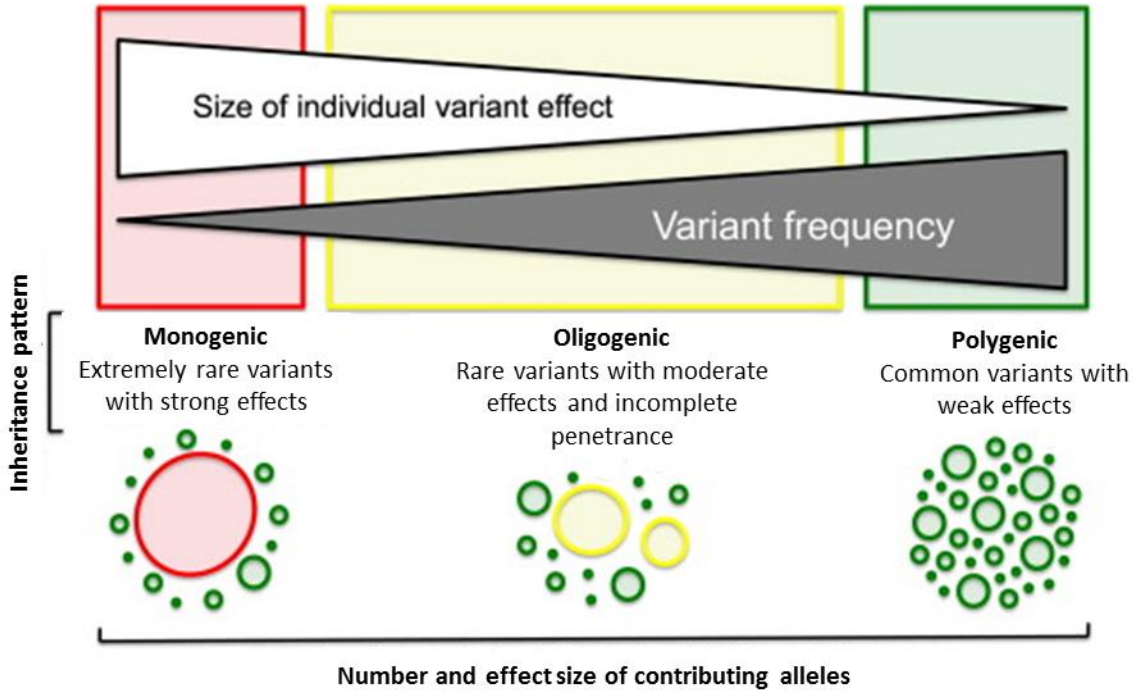


Figure 1.8 A spectrum of genetic variation

At the severe (red) end of the spectrum are extremely rare (<1% minor allele frequency) mutations that have strong effects and typically result in monogenic disorders. In the middle of the spectrum (yellow) are rare variants with moderate effects that often require a second hit (eg. genetic modifier). Lastly, at the benign (green) end of the spectrum are common (>5% minor allele frequency) variants that have weak effects and only contribute to the disease in rare circumstances when multiple hits (genetic and environmental) are present (Adapted from Giudicessi and Ackerman, 2013).

It is possible that both familial and non-familial otosclerosis are caused by genetic and environmental factors that affect the same gene regulatory network, but that in familial otosclerosis, a single gene mutation is responsible for disrupting this network, whilst in non-familial cases, a number of genetic and environmental factors make modifications to the regulation of the network, which when combined result in its disruption. Here the evidence for both environmental and genetic factors that have been implicated in otosclerosis pathogenesis is evaluated, along with an overview of possible disease mechanisms.

1.7.1 Environmental contributors to otosclerosis

A number of environmental factors have been implicated in the pathogenesis of otosclerosis, however limited progress has been made in determining the contribution of these to disease aetiology. Viral and endocrine influences, as well as exposure to environmental agents such as sodium fluoride have all been suggested to play a causative role in the development of otosclerosis, but despite a considerable body of research investigating these factors, the role that they play in disease pathogenesis is yet to be established. The lack of progress in elucidating the role of environmental contributors in the development of otosclerosis is probably largely due to the heterogeneous nature of this disorder, in which a combination of genetic and environmental factors are involved in disease aetiology. This makes it difficult to identify the precise role of each environmental contributor to the pathophysiology of otosclerosis.

1.7.1.1 Viral infection as an otosclerosis susceptibility factor

Since the late 1980s, there has been considerable interest in the role of persistent viral infection in the development of otosclerosis. A study published in 1987 detected antigens from the measles, mumps and rubella viruses in otosclerotic lesions through a series of immunohistochemistry experiments using monoclonal and polyclonal antibodies (Arnold and Friedmann, 1987), and early gene expression studies that detected measles virus protein within otosclerotic tissue supported the theory that the measles virus was implicated in otosclerosis development (McKenna et al., 1989). In addition, a number of reports in the 1990s suggested that RNA or protein from the measles virus was present in some but not all histologically positive otosclerotic lesions from formalin-fixed temporal bone specimens (Niedermeyer et al., 1994, McKenna et al., 1996), further supporting a role for viral infection in the disease process in some otosclerosis cases. However, when RNA was extracted from 28 non-formalin fixed otosclerotic tissue samples which had been frozen, measles virus RNA was not detected (Grayeli et al., 2000). This study was therefore unable to confirm the hypothesis that persistent measles viral infection was involved in otosclerosis. However, it has since been suggested that expression of anti -measles IgG is decreased in otosclerotic patients compared to controls (Lolov et al., 2001, Karosi et al., 2006), and measles virus receptor CD46 has been found to be highly expressed on the surface of cells within the otic capsule (Karosi et al., 2007), providing further evidence to support a role for measles viral infection in otosclerosis. Recently, a study of CD46 isoforms in 21 otosclerotic and 30 control stapes footplates identified different expression patterns in otosclerotic compared to control stapes using

restriction analysis and molecular weight distribution methods (Csomor et al., 2010). This was supported by a follow up study where histopathological examinations of control and affected stapes along with CD46-specific Western blot analyses were performed, identifying four novel CD46 isoforms in otosclerotic specimens (Karosi et al., 2008, Csomor et al., 2010). The research group claimed that their studies indicate the expression of a specific CD46 isoform expression pattern associated with otosclerosis. However quantitative data to indicate relative expression levels of specific CD46 isoforms in otosclerotic compared to control stapes, would be required to substantiate this claim. When all the evidence regarding viral infection is considered as a whole, there does appear to be some basis to support a role for measles viral infection as a trigger for the onset of otosclerosis; however this remains to be proven.

The argument that measles infection may be an otosclerosis susceptibility factor was further supported by a recent study that showed a reduction in incidence of otosclerosis in Germany since the introduction of the measles immunisation vaccine (Arnold et al., 2007). This research group proposed that their observations could explain the correlation between increasing measles vaccination uptake rates since the 1970s and the reported decline in incidence of otosclerosis during recent years. However, since vaccination programmes were only introduced into Germany in 1974, individuals who received vaccination in childhood are only now reaching the typical age of onset for otosclerosis. This is therefore unlikely to be a valid explanation for the decreasing incidence figures. Moreover, since otosclerosis is more common in developed countries where measles vaccination is common, than in less developed countries where vaccination is rare, it is unlikely that this theory can be substantiated. Although this theory cannot be ruled out, more evidence is required to determine whether immunisation against measles protects against the development of otosclerosis.

1.7.1.2 Involvement of endocrine factors in otosclerosis

Endocrine factors such as progesterone and oestrogen have also been suggested to be involved in otosclerosis. This is due in part due to the discrepancy in prevalence of the disease between males and females and also due to the long-held belief that otosclerosis often manifests after or during pregnancy. Oestrogen is known to stimulate osteocyte activity (Arnold et al., 1996) which may account for the increase in disease development rate during pregnancy when there is a steady increase in oestrogen levels from the tenth week of gestation. In the 1980s, a large, retrospective study of 479 females with otosclerosis documented a subjective impression of deterioration of hearing during pregnancy, concluding that pregnancy does trigger deterioration of hearing in otosclerosis patients, particularly

during subsequent pregnancies (Gristwood and Venables, 1983). However, the evidence attributing otosclerosis development to hormonal changes during pregnancy has been somewhat contradictory in more recent studies (Vessey and Painter, 2001, Lippy et al., 2005), with a 2005 study showing no difference in hearing ability between 47 women with children and 47 without, all of whom had undergone stapes surgery (Vessey and Painter, 2001, Lippy et al., 2005). The lack of support for a role of hormonal changes in otosclerosis from recent studies suggests that the link between otosclerosis and pregnancy may be spurious. It is possible that this is due to the fact that the typical age of onset of otosclerosis coincides with average child-bearing age, indicating that the link be may coincidental rather than causal. In addition, it appears that the origin for the argument that pregnancy accelerates hearing loss in otosclerosis originally stemmed from a 1939 paper by the Agency of Expert Opinion of the German Reich, indicating abortion and sterilisation as treatment for otosclerosis. It is therefore possible that the initial link between otosclerosis and pregnancy was falsified by the Nazis in an attempt to prevent reproduction for eugenic purposes (for review see Lippy *et al.* , 2005).

Prolonged use of oral contraceptives and hormone replacement therapy have also been suggested as susceptibility factors in otosclerosis (Podoshin et al., 1978), however no evidence was found to support this during a study which investigated adverse effects of oral contraceptives on otosclerosis and vestibular disorders in a large population (Vessey and Painter, 2001). Overall, with the exception of one large study conducted in the 1980s, there is little evidence to support a role for hormonal factors in otosclerosis pathogenesis. However, it should not yet be ruled out as a possible susceptibility factor. An improved understanding of the molecular mechanism involved in otosclerosis could help to elucidate what role, if any, hormonal factors play in otosclerosis pathology.

1.7.1.3 The contribution of fluoride to otosclerosis development

Another environmental factor implicated in otosclerosis is fluoride. Sodium fluoride is found naturally in ground water, but is also added to the water supply in some countries due to known dental benefits in children. An epidemiological study in the 1960s showed that otosclerosis is more common in areas with low fluoride content in the drinking water (Daniel, 1969). However a residential history questionnaire supplied to individuals born between 1948 and 1962 in two regions in Finland, one with fluoridated water and the other without, was not able to find any difference in cases of otosclerosis between the regions (Vartiainen and Vartiainen, 1997a). The same research group were also unable to find evidence that sodium

fluoride exposure affected hearing in 294 otosclerosis individuals in the years following their stapedectomy surgery (Vartiainen and Vartiainen, 1997a).

There have been several explanations as to the role of fluoride in otosclerosis development. It has been proposed that sodium fluoride could inhibit proteolytic enzymes, which could decrease the rate of disease progression (Causse et al., 1980). Alternatively, it has been hypothesised that sodium fluoride may inhibit the activity of DTDST, a diastrophic dysplasia sulphate transporter which is involved in bone matrix synthesis (Grayeli et al., 2003). This could account for the association between otosclerosis and fluoride; however, it is difficult to quantify the role of fluoride in the disease process due to difficulties in measuring a person's lifetime exposure to this ion.

Overall, the evidence for a role of environmental factors in otosclerosis is inconclusive, largely due to limitations associated with quantifying exposure to environmental agents. Having said this, whilst there is minimal evidence to support a role for endocrine factors and fluoride in otosclerosis, the role for viral infection is better substantiated by the data. Further studies are essential to establish whether or not these factors are involved in disease pathogenesis.

1.7.2 Genetics of otosclerosis

Over the past several decades, the methods for gene identification in families exhibiting monogenic inheritance of otosclerosis have centred on traditional linkage analysis followed by the selection of candidate genes for genetic testing, a technique which has proven successful at identifying the chromosomal region in which disease causing mutations are located in many Mendelian traits. Association studies have also been used to investigate genetic variants involved in otosclerosis. Whilst family studies tend to focus on the identification of rare variants with large effect sizes, association studies attempt to identify more common variants with smaller effect sizes that work in combination with other genetic and environmental factors to result in the disease phenotype. In addition, analysis of gene expression profiles in diseased stapes tissue has also been used to investigate genetic contributors to this disease. Whilst gene expression analysis is unlikely to pinpoint a specific disease causing gene, it can be used to identify dysregulated gene pathways in the tissue of interest, which can help improve understanding of molecular processes involved in disease development. Here, the various approaches taken in attempts to identify otosclerosis causing genes will be discussed and the outcomes of these strategies evaluated.

1.7.2.1 Linkage analysis in otosclerosis families

Linkage analysis is a technique that involves pinpointing a location in the genome at which a disease causing mutation is located. Thousands of genetic markers are used to identify a locus that segregates with the disease in affected family members, in the hope of identifying the disease causing mutation within the segregating region. Although this technique has been extremely successful in the identification of genetic mutations in numerous Mendelian disorders, the success of linkage analysis in otosclerosis families has been limited. This is largely due to the fact that otosclerosis exhibits reduced penetrance, and as a result, unaffected individuals in the family may possess the disease genotype but not display symptoms, which has hampered the identification of true familial controls. This has hindered investigators from effectively honing in on the disease causing region, resulting in the identification of large linked regions in which the disease causing mutation is presumed to be located. In most cases, these regions contain large number of genes and given that the disease process involved in otosclerosis is unknown, identifying appropriate candidate genes within these large regions has proven problematic. Success of linkage analysis has also been limited by difficulty in identifying appropriate families. Since otosclerosis is an adult onset disorder, there are few large families with living affected individuals across multiple generations, which are required for the success of this technique. In addition, since otosclerosis is a heterogeneous disorder in which variants within different genes are responsible for onset of the condition in different families, a pooling strategy using multiple families is not appropriate for the identification of disease causing genes.

Despite these limitations, at least eight loci have been identified through linkage analysis, which have been designated *OTSC1* to *10* (loci *OTSC6* and *OTSC9* remain unpublished) (Table 1.1). These loci vary in size from just 12 protein coding genes identified in the overlapping segregating regions between two families in *OTSC7* (Thys et al., 2007b), to as many as 423 protein-coding genes in *OTSC3* (Chen et al., 2002, Ali et al., 2007). However, despite further investigation, no disease causing mutations have been identified within any of these regions. This is likely due to the variable penetrance issue which results in the identification of large linked regions containing numerous genes from which it is difficult to select appropriate candidates for sequencing. Alternatively the presence of phenocopies within these families, where an individual displays a phenotype difficult to distinguish from otosclerosis on clinical grounds, may also have hindered the identification of disease-causing genes. Despite this, a number of good candidate genes have been identified within each of these regions, as is discussed below.

Locus	Position	County of origin of family	Publication reference	Maximum multipoint LOD score	No. protein coding genes in linked region
OTSC1	15q25-26	India	(Tomek et al., 1998)	3.4	89
OTSC2	7q34-36	Belgium	(Van Den Bogaert et al., 2001)	3.54	93
OTSC3	6p21-22	Cyprus	(Chen et al., 2002)	3.83	423
		Tunisia	(Ali et al., 2007)	2.04	
OTSC4	16q21-23	Israel	(Brownstein et al., 2006)	3.97	115
OTSC5	3q22-24	Netherlands	(Van Den Bogaert et al., 2004)	3.46	79
OTSC7	6q13-16	Greece	(Thys et al., 2007b)	7.5	72
		Netherlands		1.96	22
OTSC8	9p13-9q21	Tunisia	(Bel Hadj Ali et al., 2008)	4.13	49
OTSC10	1q41-44	Netherlands	(Schrauwen et al., 2011)	3.3	148

Table 1.1 Table of otosclerotic loci identified by linkage analysis

Otosclerotic loci identified during linkage analysis in families exhibiting autosomal dominant inheritance of otosclerosis. Number of protein coding genes located within the linked regions based on Ensembl release 69 Assembly Grch37

The first linkage analysis study for otosclerosis was performed on a large multigenerational Indian family exhibiting autosomal dominant inheritance of otosclerosis, in which no consanguinity was recorded. Five members of this family had surgically confirmed otosclerosis and an additional nine had conductive hearing loss suggestive of otosclerosis but had not undergone stapes surgery. In this family, severity of hearing loss based on audiometric analysis increased with advancing age. Genetic linkage analysis was performed using short tandem repeat polymorphisms (STRPs) throughout the genome which indicated a 14.5cM region flanked by centromeric marker *FES* and telomeric marker *D15S657* on chromosome 15q25-26 that segregated with the disease in the family (maximum lod score=3.4) (Tomek et al., 1998).

Since the identification of this first linked region, seven additional otosclerotic loci have been published that reportedly segregate with otosclerosis in large affected families. All families investigated are reported to exhibit autosomal dominant inheritance of otosclerosis. However, two of these families, the Greek family segregating with *OTSC7* and the Tunisian family segregating with *OTSC8,* reported a history of consanguinity which could be suggestive of autosomal recessive inheritance. However analysis of these family pedigrees, suggests that this is unlikely. The age of onset of hearing loss was variable across affected family members in all of the families, with the exception of the Greek family in which the *OTSC7* locus was identified, where disease onset commonly occurs during childhood at an average of age 10.

In total, six of the eight published otosclerotic loci have been mapped in just one otosclerosis family, whilst two loci, *OTSC3* and *OTSC7*, have been mapped in multiple families. The *OTSC3* locus was identified in two families, one of Cypriot origin and one of Northern Tunisian origin, and the *OTSC7* locus was identified in two families, one of Greek and one of Dutch origin. The subsequent identification of the Dutch family mapping to *OTSC7* enabled fine mapping of the region, narrowing down the candidate interval to just twelve protein coding genes (Ensembl release 69 Assembly Grch37). At the Congres Francais d'Oto-Rhino-Laryngologie et de Chirurgie de la Face et du Cou in Paris in 1999, it was reported that a Tunisian family may also map to the *OTSC1* locus (Drira, 1999), however the data from this study is not accessible for the identification of overlapping linked regions. There are also reports that two additional Belgian families may have been mapped to *OTSC2* (Ealy, 2011). However, a recent study of seven British Caucasian otosclerosis families found no linkage to the *OTSC*2 locus (Alzoubi et al., 2007), though this was explained by a lack of power associated with the small number of families recruited for analysis. This explanation is likely to be accurate as due to the highly heterogeneous nature of the disease, confirmed by the numerous otosclerotic loci identified through linkage analysis, it appears that a different genetic factor or factors are responsible for the disease in most otosclerosis families identified.

In all eight otosclerotic loci identified, a significant LOD score greater than 3.3 was identified in each of the first families mapped to the locus. This threshold of 3.3 is necessary when genome-wide linkage analysis is performed as multiple testing must be taken into consideration. However, when linkage analysis is performed to investigate segregation with a known linked region, a LOD score greater than 1.9 is sufficient to suggest linkage to the region (Lander and Kruglyak, 1995). Therefore LOD scores of 2.04 and 1.96 for the second families mapped to *OTSC3* and *OTSC7* respectively are sufficient to suggest linkage to these loci, as in these cases, the locus had previously been mapped. The highest multipoint LOD score of 7.5 was identified in the Greek family, in which the *OTSC7* locus was mapped, providing the strongest evidence for linkage of all of the families. A possible explanation for the high LOD score in this family is the severe phenotype, characterised by childhood onset hearing loss. This phenotype makes it easier to distinguish between affected and control individuals than in other families with adult onset symptoms. As a result, it is likely that variable penetrance was less of an issue when performing linkage analysis in this Greek family than in the other families.

The regions mapped during linkage analysis range in size from the overlapping region identified for *OTSC7,* which contains just 12 protein coding genes, to *OTSC3,* which contains 423 protein coding genes (Ensembl release 69 Assembly Grch37). In each family, possible

otosclerosis candidate genes were proposed within the linked regions and in some cases were sequenced, in the hope of identifying a disease causing mutation. In *OTSC1*, aggrecan was pinpointed as a possible otosclerosis candidate within the 14.5cM region mapped. This is a component of the extracellular matrix which is expressed in the bony labyrinth of the ear. However, no disease causing variants were identified through the sequencing of aggrecan, suggesting that it is not the causative gene in this family. In *OTSC2* Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase 3 *(PLOD3)*, tripartite motif containing 24 *(TRIM24)* and T-cell recptor beta (*TCRβ)* were suggested as potential otosclerosis candidates. *PLOD3* is involved in the synthesis of collagen, a key component of bone and has been implicated in Ehlers-Danlos syndrome type VIB, a connective tissue disorder that has been reported to occasionally occur concomitantly with otosclerosis (Miyajima et al., 2007). *TRIM24* is a transcriptional co-activator that plays a role in bone remodelling and is therefore a good otosclerosis candidate. *TCRβ* encodes a protein involved in antigen recognition in the immune system. Although sequencing of *TCRβ* found no causal genetic mutations within the family that mapped to this locus, affected members of this family have been found to have decreased expression of TCRβ mRNA and fewer circulating $TCR\alpha\beta^{+}$ T cells relative to controls (Schrauwen et al., 2010b). *OTSC3* maps to a region containing the MHC locus, which has been proposed as the most likely location of the disease-causing mutation within the family that mapped to this linked region. Within *OTSC4*, a number of zinc finger proteins were suggested by the authors who identified the linked region as the most likely candidates (Brownstein et al., 2006) and in *OTSC5,* procollagenase endopeptidase enchancer 2 *(PCOLCE2)* and carbohydrate sulphotransferase 2 *(CHST2)*, were suggested as candidates, though, neither was found to contain disease causing mutations. In *OTSC7*, although fine mapping of the overlapping segregating regions between the Greek and Belgian families enabled the identification of just twelve overlapping protein coding genes, none of these were found to contain a disease-causing mutation in either of the families during sequencing. It is therefore possible that the mutation resides within a regulatory element in this region. Alternatively it is possible that different mutations are responsible for the disease in each of these families. In *OTSC8,* genes encoding a tight junction protein, a protein involved in calcium signalling and osteoclast function, and a protein involved in craniofacial development were ruled out as good candidates by sequence analysis. In *OTSC10,* transforming growth factor beta 2 (*TGFβ2)* and angiotensinogen (*AGT*) were proposed as the most likely candidates but sequencing revealed no pathogenic variants (Schrauwen et al., 2011).

Although eight otosclerotic loci have been successfully identified in otosclerosis families using linkage analysis, no disease causing mutations have been found using this technique. This may

be partly because even though good candidates within the mapped regions have been identified, and in some cases screened, in total, only a small proportion of known genes within these eight regions have been sequenced. Since little is known about otosclerosis pathogenesis, it is not surprising that the selection of arbitrary candidates based on assumptions about the disease process, has not yielded promising results. A better understanding of the disease process would enable the selection of better candidates within the regions, which may increase the likelihood of identifying otosclerosis causing variants within these families. Another limitation of these studies is that in most cases, only the coding regions of these good candidates have been sequenced. Whilst this is a good starting point as the majority of disease causing mutations are located in coding regions of genes, deep resequencing of promoter and enhancer elements, as well as the 5' and 3' untranslated regions would also be useful, as it is possible that disease causing variants could reside within these regulatory regions. The limited success of linkage analysis as well as the identification of multiple loci in different families, confirms the highly heterogeneous nature of otosclerosis.

1.7.2.2 Candidate gene case control association studies for otosclerosis

Association studies can be used to identify genetic variation that is associated with a disease or trait. These studies use single nucleotide polymorphisms (SNPs) as genetic markers to identify genetic associations. A population-based candidate gene case control association study tests the relationship between polymorphisms within a candidate gene and the disease or trait of interest. In a population-based study, an association indicates that a specific allele, genotype or haplotype is significantly more or less common in a cohort of affected individuals than in a cohort of controls. This indicates that the associated variant increases or decreases a person's susceptibility to developing the disease phenotype. In addition to population-based studies, association studies can also be used to investigate common variants within family trios.

There are several limitations associated with association studies. These include that identification of an association with a specific variant does not always mean that it is involved in the development of the disease (Daly and Day, 2001). When association studies are performed, it is not feasible to investigate association of every single known variant within a region of interest. Therefore tagSNPs are used which are single SNPs located within a block of linkage disequilibrium (LD) which contains a number of SNPs that segregate with one another. Therefore finding an association with a SNP during an association study represents an association of an allele within the block of LD in which the tagSNP is located. The use of tagSNPs therefore helps to hone in on a region of the gene or genome on which further studies

can then be performed. Another considerable limitation of association studies is the identification of false positives in the data, where a polymorphism is incorrectly attributed an association with the disease or trait. It is important that cases and controls are age, gender and ethnicity matched as much as possible to avoid false positives as a result of population stratification (Koller et al., 2004). Due to the issue of false positives, replication of findings is required to confirm an association and in most cases is essential prior to publication of results. In addition, a cohort size with sufficient statistical power is crucial for success of association studies. The size of the cohort required to achieve significance is dependent upon prevalence of the disease or trait, the minor allele frequency of the polymorphism in question and the ratio of cases to control being tested. The greater the number of polymorphisms interrogated simultaneously, the larger the cohort required in order to account for corrections for multiple testing (Hong and Park, 2012). To date, members of three gene families have been investigated for associations with otosclerosis using this technique; the collagen gene family, the transforming growth factor beta gene family and genes encoding proteins involved in the renin-angiotensin-aldosterone pathway, which are discussed below.

1.7.2.2.1 Association studies to investigate members of the collagen gene family

Collagen is the most abundant protein in mammals and is the major component of connective tissues which exists in various forms as elongated fibrils (Bornstein, 1979). It is found within fibrous tissues and is abundant in bone, cartilage, skin, the gut, the cornea and a number of other tissues. It is known to be involved in the development of a number of connective tissue disorders affecting the bone including Osteogenesis Imperfecta (OI), a systemic disease caused by defective connective tissue which results in brittle bones and where patients often experience conductive hearing loss similar to that seen in otosclerosis. Collagen genes have long been postulated to be involved in the development of otosclerosis due to the crucial role of collagen proteins in the structure of bone and their role in OI. It is therefore not surprising that the earliest candidate gene case control association studies focused on members of the collagen gene family.

The first otosclerosis association study was conducted in the 1990s and investigated an association between variants in collagen type 1A1 (*COL1A1)* and otosclerosis in a small American population. An association was found between the condition and three polymorphic markers within the gene (McKenna et al., 1998). Subsequently, a larger case-control study was carried out in a population of 100 American cases and 108 controls by the same research group (McKenna et al., 2004). Analysis of the *COL1A1* region showed that of the three

polymorphic markers previously identified, one was located within an Sp1-binding site in the first intron of *COL1A1*, which showed a strong association with otosclerosis ($p < 0.0001$). This polymorphism is known to be associated with osteoporosis (Grant et al., 1996), a disease that causes a reduction in bone density, increasing the risk of fracture. Although the association found in this study was significant, the control group was not age and gender matched to the test population and the cohort size was small. It was therefore crucial that these findings were replicated before an association could be confirmed. In 2007, a larger study of American and German otosclerosis patients consisting of 206 cases and 385 controls identified a haplotype block containing the Sp1-binding site polymorphism which was present at a significantly greater frequency in affecteds than controls (Chen et al., 2007b), and a recent study in a small population of 28 Turkish patients and 50 age and gender matched controls found a significant association with this polymorphism and otosclerosis ($p<0.01$) (Ertugay et al., 2013). However, a replication study in a Spanish cohort of 100 cases and controls failed to confirm these findings (Rodriguez et al., 2004). The difference in outcomes between these studies may have been due to the differing sizes of the cohorts investigated as the results obtained from analysis of smaller cohorts are less likely to be reliable than those from larger cohorts. It is also possible that the discrepancy in findings could be due to the different ethnic origins of the cohorts investigated, as some polymorphisms are more common in certain populations. In 2012, a meta-analysis was performed to analyse the genetic association between *COL1A1* and otosclerosis by gathering data collected from all populations studied to date (Schrauwen et al., 2012). This study supported the evidence from previous studies for an association between *COL1A1* and otosclerosis, however it was expected that the effect sizes of the variants reported in earlier studies were likely to have been overestimated.

A role for *COL1A1* in disease pathogenesis can be further supported by a gene expression study, where *COL1A1* showed evidence for reduced expression in two out of nine otosclerotic stapes analysed, indicating that *COL1A1* mutations may be involved in a small percentage of cases of otosclerosis (McKenna et al., 2002). The lack of altered expression in the remaining seven stapes indicates that in the majority of cases, *COL1A1* is not dysregulated, suggesting that other genes are likely to be involved in disease aetiology and confirming the heterogeneous nature of this condition. A more recent gene expression study conducted by a Hungarian group in a larger population of 55 otosclerotic stapes, failed to identify dysregulation of *COL1A1* or Collagen type 1A2 (*COL1A2*), another type 1 collagen gene, in otosclerotic stapes tissue (Csomor et al., 2012). It has been hypothesised that otosclerosis may be caused by an increase in *COL1A1* expression which would lead to an increase in COL1A1 homotrimers, by affecting the ratio of COL1A1 to COL1A2, resulting in deposition of bone

within the otic capsule (Chen et al., 2007b). This theory is consistent with findings that mice with a targeted *COL1A2* deletion experience thickening of the stapes footplate and mild hearing loss. Despite, this no association of otosclerosis with *COL1A2* has been found (McKenna et al., 1998).

1.7.2.2.2 Association studies investigating the transforming growth factor beta superfamily of genes

The transforming growth factor beta super-family is a large family of related proteins involved in growth, development, homeostasis and regulation of the immune system. The name of this family is derived from the first member to be described, transforming growth factor beta 1 (TGFβ1), a secreted polypeptide that stimulates osteoblastogenesis and inhibits osteoclastogensis. The TGFβ super-family also includes the bone morphogenetic proteins, a group of growth factors with the ability to induce bone and cartilage formation. For these reasons, members of the TGFβ super-family have been repeatedly proposed as good otosclerosis candidate genes and have been investigated for associations with otosclerosis through association studies.

A case-control study of 632 Belgian-Dutch cases and controls conducted in 2007 found that a SNP in *TGFβ1* encoding an amino acid change from threonine to isoleucine at residue 263 was significantly associated with otosclerosis (p=0.0044) (Thys et al., 2007a). This finding was supported by a French replication set of 457 cases and 497 controls (p=0.00019). Following corrections for multiple testing, this association remained significant in both populations. Functional analysis of this SNP with a luciferase reporter assay concluded that the Isoluecine263 allele is more active than the Threonine263 allele. It was therefore suggested by the research group that Isoleucine263 may protect against otosclerosis by preventing bone resorption during early disease development. Three additional rare nonsynonymous mutations were also identified in *TGFβ1* during sequence analysis using 388 Belgian-Dutch cases and 380 controls and a French replication set of 367 cases and 497 controls (Thys et al., 2009). In addition, a recent study identified a *de novo* heterozygous mutation in the promoter region of *TGFβ1* in an individual with otosclerosis and also identified an increase in *TGFβ1* mRNA in diseased stapes compared to controls (Priyadarshi et al., 2013).

Bone morphogenetic proteins 2 and 4 *(BMP2* and *BMP4)*, also members of this family, have been associated with otosclerosis in the same large Belgian-Dutch and French populations investigated for association with *TGFβ1* (Schrauwen et al., 2008). Two SNPs were identified

that showed the same significant effect in both populations; SNP rs3178250 located in the 3' untranslated region of *BMP2* ($p=2.2x10^{-4}$) and SNP rs17563 in *BMP4* (p=0.002). Both genes are known to play a role in otic capsule development in the developing chick (Chang et al., 2002) and have therefore been proposed as potential otosclerosis candidate genes.

These studies investigating an association between otosclerosis and multiple members of the TGFβ super-family are fairly large and thus provide reliable evidence indicating that common variants within these genes may predispose an individual to otosclerosis. However, the effect size of the variants within these genes on the disease process remains unknown. Further functional studies are required to confirm involvement of these genes in disease aetiology and to quantify their contribution to pathogenesis.

1.7.2.2.3 Association studies to investigate the renin-angiotensin-aldosterone pathway

There is considerable interest in the role that hormonal factors involved in pregnancy play in the development of otosclerosis. This stems from the observation of a possible increased frequency of otosclerosis in females, and the fact that the average age of onset coincides with child-bearing age. For this reason, angiotensin-converting enzyme *(ACE)*, along with angiotensinogen (*AGT*), genes that are known to be up-regulated during pregnancy (Schrier and Durr, 1987), were investigated for an association with otosclerosis. This study in a French population of 186 cases and 526 controls, identified a polymorphism in *ACE* that is associated with otosclerosis ($p<0.05$) (Imauchi et al., 2008). This polymorphism is characterised by the insertion or deletion of a 287 base pair sequence in intron 16 of this gene, which is expected to increase the enzymatic activity of the encoded protein. A variant in AGT which causes a methionine to threonine amino acid change at residue 235 shown to be related to higher AGT plasma concentrations, was also found to be associated with otosclerosis ($p<0.01$) in this study (Imauchi et al., 2008). However, a replication study of 692 Belgian-Dutch cases and controls failed to confirm these results, suggesting that the initial results may have been false positives as a result of the small population investigated (Schrauwen et al., 2009b). In addition, a recent study found no evidence for the expression of *ACE* in 20 human otosclerotic stapes footplates (Liktor et al., 2013). This suggests that the polymorphism identified in *ACE* which increases its enzymatic activity is unlikely to be involved in otosclerosis as it would be expected that this gene would be expressed in human stapes bone if it were involved in the disease. Further investigation of these polymorphisms would be required before an association could be confirmed.

Despite some success in identifying associated variants using case control association studies, particularly within members of the TGFβ super-family, their contribution to disease pathogenesis remains unproven. This is partly due to a failure to replicate findings and due to the small cohort sizes of many of these studies. In addition, these studies are reliant upon a certain degree of knowledge of disease pathogenesis in order to identify appropriate candidate genes for analysis. Since the disease process involved in otosclerosis is yet to be elucidated, this technique therefore has limited applications for research into this condition. A recent genome wide approach to this form of analysis was performed in 2009 in the first otosclerosis Genome Wide Association Study (GWAS). Unlike candidate gene association studies, this technique requires no prior knowledge of disease pathogenesis as it interrogates many thousands of polymorphisms across the genome.

1.7.2.3 Genome wide association studies in otosclerosis

Genome Wide Association Studies (GWAS) examine large numbers of common variants simultaneously in large cohorts of cases and controls to identify associations. GWAS have been successful in the identification of genetic variants involved in many complex human diseases and common traits. Genome wide studies have the benefit over traditional association studies of requiring no prior knowledge of disease pathogenesis, however they also have their own inherent limitations and associated challenges, most of which can be overcome or minimised by effective quality control and study design. Selecting appropriate age, gender and ethnicity matched controls is even more vital than in a candidate gene study as the likelihood of false positives is greatly increased during multiple testing. For the same reason, it is crucial that even larger cohorts of cases and controls are recruited than for candidate gene studies in order to achieve sufficient statistical power. Since no prior knowledge of disease pathogenesis is required for use of this technique, greater confidence could be had in associations found using this genome-wide strategy than when using a candidate-gene based approach.

In 2009, a Genome Wide Association Study was conducted to identify variants associated with otosclerosis (Schrauwen et al., 2009a). In this study 550,000 tagSNPs were interrogated in pooled DNA samples from a discovery group of 302 Belgian-Dutch cases and controls. From the genome-wide data obtained, 250 SNPs of interest were selected for validation by individual genotyping. These SNPs included the 200 highest ranking SNPs from the pooled DNA samples, 28 SNPs located within either the linked otosclerotic loci identified during linkage analysis or in close proximity to otosclerosis candidate genes, and 22 SNPs that were located in close proximity to another high ranking SNP. Of these 250 SNPs, none were located in

otosclerosis candidate genes investigated during previous studies, however, three were involved in TGFβ signalling; rs12082710 in the intronic region of *TGFβR3* (p=0.002), rs3738676 in the 3' untranslated region of *BMP8A* (p=0.078) and rs2799090 in intronic *TGFβ2* (0.005). In addition, one SNP rs1879046 was located in an intronic region of collagen gene *COL22A1* (p=0.006). The 250 selected SNPs were subsequently individually genotyped in the same discovery group of 302 Belgian-Dutch cases and controls, of which 196 were found to be associated with the disease (p>0.05). Individual genotyping was then performed in a stepwise manner to identify significantly associated SNPs that were replicated in two additional independent populations consisting of a further 392 Belgian-Dutch cases and controls followed by a French replication cohort of 455 cases and 480 controls. It was found that two SNPs rs3914132 located within the reelin gene (*RELN*) on chromosome 7q22.1 (p=0.0003) and rs670358 located in an intergenic region on chromosome 11q13.1 between two regions of high linkage disequilibrium (p=0.005) were associated in all replication groups, so fine mapping was performed in the genomic region surrounding these SNPs. The most significant association signal ($p_{combined}$ =6.23x10^{-10}) was found in the reelin gene spanning 180kb from intron 1 to 4. This association reaches genome-wide significance as it survives correction for multiple testing using a Bonferonni correction; however the association on chromosome 11q13.1 does not. Despite this, the association with chromosome 11 was replicated in two independent populations, which can be considered stronger evidence of an association than a low p value (Neale and Sham, 2004). Resequencing of this region would be required to determine the causal variant.

The association with *RELN* was unexpected as there is no obvious role for the reelin protein, a large extracellular matrix protein involved in neuronal positioning during brain development (Quattrocchi et al., 2002), in the pathogenesis of otosclerosis. Despite this, the association between *RELN* and otosclerosis was further supported by a recent candidate gene association study in an additional European population (Schrauwen et al., 2010a), although it was not supported by a study within an Indian population (Priyadarshi et al., 2010). Further investigation is required to determine the role of reelin in the pathogenesis of otosclerosis.

1.7.2.4 Gene expression studies in otosclerotic stapes

Gene expression studies have also been employed in the investigation of genetic factors involved in otosclerosis. Although these studies tend not to be used for the identification of specific disease-causing genes, gene expression analysis of multiple genes in combination with pathway analysis can be extremely useful in identifying pathways dysregulated in diseased

tissue. Single gene expression studies can also be useful in validating the results of other genetic studies, by demonstrating altered expression of a gene of interest.

In otosclerosis, most gene expression studies conducted to date have been limited largely by difficulty in gaining access to human stapes tissue. Otosclerotic stapes bone is removed routinely during stapes surgery; however, since the preferred surgical technique is the stapedotomy which leaves the stapes footplate in place, it is difficult to accrue large numbers of stapes footplates for analysis. Whilst acquisition of stapes suprastructures poses less difficulty, these are less useful in investigating gene expression in otosclerosis since the otosclerotic foci, the sites most likely to display aberrant gene expression, tend to be localised to the stapes footplate. In addition, control stapes for comparison of gene expression between otosclerotic and unaffected stapes are also difficult to obtain and tend to be removed during only a small number of surgical procedures. For these reasons, few gene expression studies on otosclerotic tissue have been conducted to date. The majority of those that have been performed analyse only small numbers of stapes and employ a candidate gene based approach using either semi quantitative reverse transcription polymerase chain reaction (RT-PCR) to detect the presence of messenger RNA (mRNA) from genes of interest, or quantitative real time PCR (qRT-PCR) to analyse relative expression levels of genes in control and affected stapes. qRT-PCR assays have been used to analyse relative expression of single genes in otosclerotic stapes and also multiple genes simultaneously using an array based approach.

The most comprehensive gene expression study conducted on otosclerotic stapes to date was carried out in 2008 by a research group at the University of Iowa. A microarray-based study was performed using qRT-PCR to compare messenger RNA transcripts from the stapes tissue of nine otosclerotic patients against seven control stapes from patients undergoing surgery for vestibular schwannoma. Microarray analysis was performed using The Human Genome Survey microarray v2.0 from Applied Biosystems which annotates 29,098 genes. It was found that 110 of these genes were differentially expressed in otosclerotic stapes compared to controls, 92 up- and 18 down-regulated (Ealy et al., 2008). Of these genes, four were found to be located within the otosclerotic loci identified during previous linkage analysis, with three of these being located in *OTSC3*, the largest of the eight otosclerotic loci containing 423 protein coding genes, whilst the remaining gene was located in *OTSC5.* The three dysregulated genes identified within *OTSC3* were sequenced in the original family in which this locus was identified, but no mutations were found within the coding regions, untranslated regions or splice sites of these genes. The gene identified with the highest up-regulation in otosclerotic stapes compared to controls was platelet factor 4 (*PF4)*, which was up-regulated 12.6 fold in

otosclerotic stapes. *PF4* is thought to inhibit TGFβ1 signalling by inhibiting the binding of TGFβ1 to its receptor (Whitson et al., 1991). The most down-regulated gene was bone sialoprotein (*IBSP)*, which was down-regulated 6.22 fold in otosclerotic stapes. This is a major structural protein of the bone matrix required for bone formation and resorption (Wade-Gueye et al., 2012). Single gene qRT-PCR assays were used to validate the microarray results for seven genes of interest including cyclin dependent kinase inhibitor 1A (*CDKN1A)* and lymphotoxin beta (*LTB)* located in the *OTSC3* locus*,* the ATP-ase beta 3 sub-unit *(ATP1B3)* in the *OTSC5* locus as well as *PF4, IBSP* and two other genes that were found to be significantly up-regulated, low density lipprotein receptor *(LDLR)* and chemokine ligand 2 *(CCL2)*. Dysregulation of both *PF4* and *IBSP* in otosclerotic stapes was confirmed by the qRT-PCR assays; however this was not the case for the other five genes. To investigate the molecular pathways involved in otosclerosis, all 110 differentially expressed genes were analysed using the PANTHER classification system. Pathway analysis indicated dysregulation of molecular processes involved in interleukin signalling and inflammation in the otosclerotic stapes, which indicates an inflammatory response in otosclerosis patients. Whilst the dysregulated genes identified indicate involvement of a number of interesting biological pathways in otosclerosis, further studies would be required to confirm expression of the genes in a larger number of otosclerotic stapes.

1.7.3 Disease mechanisms for otosclerosis

It is crucial to gain a thorough understanding of the molecular basis of disease in order to establish the mechanisms involved in the disease process. This is vital for the development of new effective therapies and treatments. Although a number of environmental and genetic factors have been suggested to play a role in disease aetiology of otosclerosis, no disease mechanism is widely accepted. However, by considering the results obtained across a number of studies conducted to date, various common themes have emerged indicating possible disease mechanisms. Here, three common themes that have emerged whilst analysing the published literature are discussed; otosclerosis as an autoimmune condition, the role of TGFβ in otosclerosis pathophysiology and otosclerosis as a localised manifestation of a systemic connective tissue disorder.

1.7.3.1 Otosclerosis as an autoimmune condition

It has previously been proposed that an autoimmune mechanism may be involved in otosclerosis pathophysiology. Early research into otosclerosis hypothesised that the condition may occur as a result of an autoimmune reaction to type II collagen present in the residual embryonic cartilage of the *globuli interossei* within the otic capsule (Yoo et al., 1982), a theory which was supported by a study in which otosclerotic-like lesions were induced in rats immunised with type II collagens (Yoo et al., 1983). Furthermore, various cells of the immune system have been reported both within and in close proximity to otosclerotic lesions, indicating stimulation of the immune system in otosclerosis patients and a subsequent inflammatory tissue reaction (Altermatt et al., 1992). An autoimmune pathology for otosclerosis is also supported by a recent study that identified HLA class I polymorphisms in an otosclerosis cohort of Tunisian origin (Bel Hadj Ali et al., 2012), since HLA genes are known to be involved in various autoimmune conditions (Caillat-Zucman, 2009). A further study which demonstrated an increase in HLA antigens in otosclerosis patients (Gregoriadis et al., 1982) combined with the fact that the *OTSC3* region identified during linkage analysis (Chen et al., 2002) maps to a genomic region containing the MHC locus encoding the HLA genes, also support a role for an autoimmune mechanism. Furthermore, the results of the pathway analysis following the most comprehensive gene expression study on otosclerotic tissue conducted to date(Ealy et al., 2008) indicated dysregulation of multiple genes involved in the inflammatory response. Since autoimmune conditions are characterised by prolonged inflammation, this provides further evidence to suggest an autoimmune pathophysiology for otosclerosis.

1.7.3.2 The role of TGFβ in otosclerosis pathophysiology

The transforming growth factor beta super-family is a group of genes involved in growth, development, homeostasis and regulation of the immune system. One member of this family, transforming growth factor beta 1 (TGFβ1), is a cytokine that has been repeatedly proposed as an otosclerosis candidate gene due to its vital role in bone remodelling. During remodelling, TGFβ is involved in numerous processes that regulate the recruitment, differentiation and function of both osteoblasts and osteoclasts (Figure 1.9).

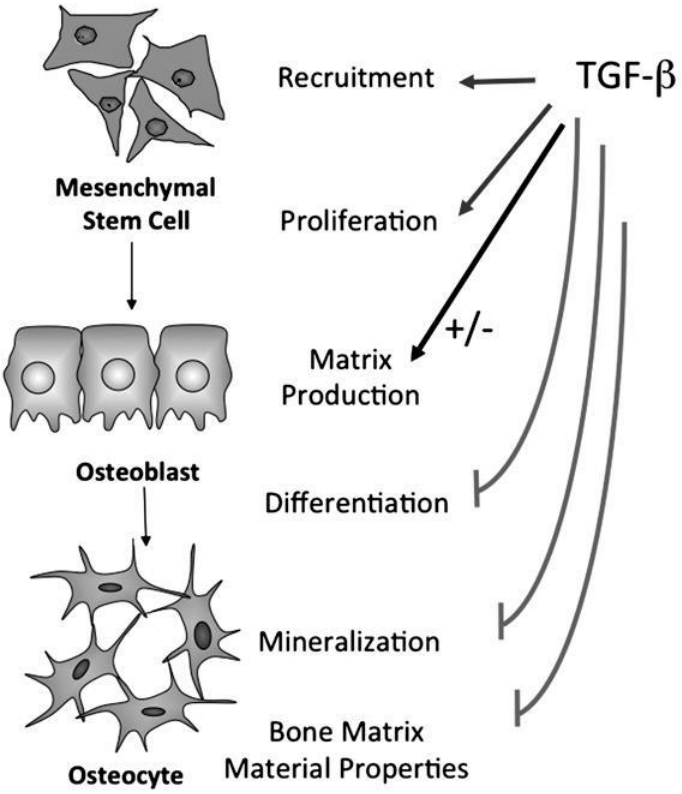


Figure 1.9 Transforming growth factor-β (TGFβ) involvement in bone remodelling

TGFβ regulates the bone remodeling cycle through the recruitment, proliferation and differentiation of osteoblasts (Tang, 2013).

For this reason, *TGFβ1*, along with genes encoding other TGFβ super-family members including bone morphogenetic proteins, have been investigated for associations with otosclerosis (Thys et al., 2007a, Schrauwen et al., 2008, Thys et al., 2009). Although there are some concerns regarding the reliability of the results obtained during otosclerosis candidate gene association studies, those associating variants within members of the TGFβ super-family and otosclerosis are amongst the most convincing. Furthermore, a study identifying dysregulation of *TGFβ1* in otosclerotic stapes provides further evidence that this gene may be involved in abnormal bone remodelling in the stapes of otosclerosis patients (Priyadarshi et al., 2013). In addition, since the most up-regulated gene identified in the Ealy et al. gene expression study (*PF4*) is known to be an inhibitor of TGFβ1 (Ealy *et al.* , 2008), it is suggestive that genes governing expression of this crucial regulator of bone remodelling may also be involved in the disease process. Given that multiple avenues of research into the genetics of otoscleorsis point to *TGFβ1*, it is a strong possibility that this gene may be involved in disease pathophysiology and is a good candidate to prioritise for future research.

1.7.3.3 Otosclerosis as a localised manifestation of a systemic connective tissue disorder

Otosclerosis has been reported to occur concomitantly in individuals with other systemic connective tissue disorders (Miyajima et al., 2007) indicating that there may be a shared genetic aetiology between these conditions and otosclerosis. Connective tissue disorders are conditions in which connective tissues are the target for disease pathology. Connective tissues are those with extensive extracellular matrices consisting of the two large structural proteins collagen and elastin. Patients with connective tissue disorders tend to have weaknesses or inflammation in their connective tissues, often accompanied by autoimmunity, another possible otosclerosis disease mechanism discussed above.

Most connective tissue disorders are caused by genetic mutations and are inherited within families. Ehlers-Danlos syndrome (EDS) is a connective tissue disorder caused by defects in collagen synthesis and has been reported to coincide with otosclerosis on rare occasions (Miyajima et al., 2007). There are many different forms of EDS which are caused by mutations in the collagen genes *COL3A1, COL5A1* and *COL5A2*, as well as in procollagen-lysines *PLOD1* and *PLOD3*. Symptoms include double-jointedness, easily damaged, bruised and stretchy skin, scarring and poor wound healing, flat feet, increased joint mobility, arthritis, joint pain and dislocation, premature rupture of membranes during pregnancy and gastrointestinal problems. Due to the role that mutations within collagen genes play in development of EDS and the proposed role of collagen variants in otosclerosis, it is possible that there is a shared genetic basis for otosclerosis and EDS.

Buschke-Ollendorff syndrome is a rare genetic disorder of connective tissue that has also been reported in individuals who have otosclerosis (Strosberg and Adler, 1981, Piette-Brion et al., 1984, Schnur et al., 1994,). It is caused by mutations in the *LEMD3* gene (LEM Domain-Containing Protein 3), a gene which can antagonise TGFβ signalling in human cells (Hellemans et al., 2004). The condition is characterised by multiple collagen-rich subcutaneous benign nevi or nodules on the skin and osteopoikilosis, benign sclerotic bone lesions which lead to joint pain. Since dysregulation of TGFβ signalling is known to be involved in the development of Buschke-Ollendorff syndrome and the proposed role of TGFβ in the development of otosclerosis discussed above, it is possible that dysregulation of the same pathway is involved in both conditions.

Osteogenesis Imperfecta (OI) is a systemic disease caused by defective connective tissue which results in brittle bones and where patients often experience conductive hearing loss similar to that of otosclerosis. It had been theorised that otosclerosis and OI have a shared genetic

aetiology and that otosclerosis is a localised manifestation of OI (Weber, 1930). However, others suggest that hearing loss in OI is caused by a series of micro-fractures combined with reparative vascular and fibrous tissue forming around the cochlea (Mezzedimi, 2013).There are eight known forms of OI, all of which are characterised by weak or brittle bones and the majority of which cause blue sclera, where the whites of the eyes exhibit a characteristic blue tinge. In the 1920s it was proposed that otosclerosis is associated with forms of OI with blue sclera (Cleminson, 1927). OI types I to IV are caused by mutations in the collagen type one genes; *COL1A1* or *COL1A2*, and vary in severity of symptoms with type II being the most severe and usually lethal within the perinatal period. They are inherited in an autosomal dominant pattern although the majority of cases are caused by *de novo* mutations, especially in the more severe forms of the condition where affected individuals rarely survive to child-bearing age (Cole and Dalgleish, 1995). OI type V is caused by mutations in *IFITM5* gene (interferon induced transmembrane protein 5) and exhibits autosomal dominant inheritance, while types VI, VII and VIII are caused by mutations in *SERPINF1* (serpin peptidase inhibitor, clade F member 1), *CRTAP* (cartilage associated protein) and *LEPRE1* (leucine proline-enriched proteoglycan (leprecan) 1) respectively, all of which exhibit autosomal recessive inheritance (Cundy, 2012). The co-occurrence of OI and otosclerosis has not been well documented, but it is thought that otosclerosis occurs most commonly in individuals with mild forms of OI, particularly with mutations in the *COL1A1* gene (McKenna et al., 1998). The mutations that cause OI are commonly nonsense or frameshift mutations that lead to nonsense mediated decay of the gene product or dominant-negative mutations that affect the structure and function of the resulting protein. This results in a lack of gene product or an abnormal gene product, impacting all bones and other tissues in which this gene is expressed throughout the body, including the temporal bone. It is possible that mutations in individuals with otosclerosis affect the same genes but have a less damaging effect on the resulting protein leading to a localised form of the condition which affects exclusively the temporal bone. This theory could also explain the high incidence of otosclerosis in the general population, as it is possible that the less damaging otosclerosis-causing alleles have been preferentially selected over numerous generations over alleles that lead to the development of OI.

Since otosclerosis affects approximately 0.3-0.4% of the population (Pearson et al., 1974) (Gapany-Gapanavicius, 1975, Declau et al., 2007b), there is a possibility that the co-occurrence of otosclerosis and other connective tissue disorders will occur in some individuals by chance. However, the phenotypic similarities, along with some genetic overlap, indicates that there may be some shared aetiology for otosclerosis and systemic connective tissue disorders.

1.7.4 Future promise for genetic research into otosclerosis

Despite advances in genetic research throughout the last decade which has begun to help unravel the genetic basis of many previously uncharacterised genetic disorders, the disease pathology in otosclerosis remains to be elucidated. This is despite the fact that a number of research groups have conducted investigations into the genetics of the condition, including studies that have used sufficiently large cohort sizes to generate statistically significant results and have employed study designs that would be expected to generate reliable findings. In addition, since otosclerosis is common, meaning that recruitment of affected individuals is less problematic than for rarer forms of deafness, it is surprising that progress in this field has been slow in comparison to genetic research into other forms of deafness.

The main reasons for the lack of progress are due to difficulties in gaining access to sufficient quantities of diseased tissue from the ear for analysis of gene dysregulation, and also due to issues associated with incomplete penetrance during linkage analysis studies in otosclerosis families. It is crucial that these obstacles are overcome in order to improve the understanding of genetic factors that contribute to otosclerosis in order to clarify the underlying disease mechanism. This could improve treatment options for those affected by aiding development of new therapies. This in turn could ease the impact of this disorder on those affected as well as helping to reduce the social and economic burden of this common form of hearing loss.

Recent developments in the field of genetic research provide new opportunities to help overcome these obstacles and unravel the genetic basis of otosclerosis. Next generation sequencing has become increasingly accessible and affordable for researchers investigating genetic conditions in recent years and has the major advantage over many traditional genetic analysis techniques of requiring no prior knowledge of disease pathogenesis. Such techniques are therefore very appealing for investigation into the genetic causes of otosclerosis where the disease process is unknown. In this research study, two next generation sequencing techniques; Whole Exome Sequencing and RNA-sequencing, neither of which have been previously used to investigate the genetics of otosclerosis, will be employed in combination with one another to investigate the genetic basis of this disorder. These unexplored techniques have great potential to yield promising results.

1.7.4.1 Whole Exome Sequencing to identify genetic variation in otosclerosis families

Whole Exome Sequencing is a next generation sequencing technique which when combined with a variant prioritisation strategy, can be used to identify disease-causing variants in families exhibiting monogenic inheritance of genetic disorders. This technology sequences only the exome, the coding region of the genome as a cost–effective alternative to whole genome sequencing. Since only approximately 1% of the genome is protein coding but around 85% of disease causing mutations are located within protein coding genes (Cooper, 1998), there is a good chance of identifying a pathogenic variant by focusing solely on the exome. This technique works on the principle of identifying a list of variants in affected individuals, and narrowing them down to the causative variants by selecting only variants that are common to affected but not unaffected individuals in the family. By applying a series of filters including comparing variants with databases of known non- pathogenic variants and by identifying whether any variants are present within candidate genes for the condition, it may be possible to identify the disease-causing mutation. In recent years, there have been numerous reports of novel genetic mutations identified using this technique, including the identification of genes involved in monogenic disorders that have so far eluded researchers (Bamshad et al., 2011). This technique therefore has potential to identify genetic mutations within otosclerosis families. A more detailed overview of Whole Exome Sequencing and its advantages and limitations are discussed at the start of Chapter 4.

1.7.4.2 RNA-sequencing to investigate gene expression in otosclerotic stapes

RNA-sequencing is a next generation approach to gene expression analysis, which analyses all RNA molecules expressed within a specific tissue or cell type. During application of this technique, a series of sequence reads are generated through high throughput sequencing, which are aligned to a reference transcriptome, generating a measure of expression for every transcript. It can therefore be used to both identify and quantify RNA transcripts and to provide information including localisation of transcription and alternative splicing. Statistical analysis is then used identify significantly dysregulated genes along with pathway analysis to identify trends in the data (McGettigan, 2013). A detailed discussion of RNA-sequencing and its uses and associated disadvantages can be found at the beginning of Chapter 5.

1.8 Aims and objectives of this thesis

Until the last decade, very little was known about the causes of otosclerosis, however recent research has implicated a number of genetic and environmental contributors. But despite these advances, the disease mechanism remains poorly understood. The aim of this project is to identify genetic contributors to the pathogenesis of otosclerosis. To achieve this aim, three key objectives were devised. These are summarised below:

1) To establish a recruitment process to identify individuals affected by otosclerosis and recruit them to the study. The recruitment process will involve:

- Collection of otosclerotic stapes tissue from participants undergoing stapes surgery, as well as collection of control stapes tissue
- Collection of a blood or saliva sample from each study participant
- The identification of probands with families that appear to exhibit a Mendelian inheritance pattern of otosclerosis
- Recruitment of both affected and unaffected relatives of these probands

2) To compare changes in gene expression between control and affected stapes bones. Techniques will include:

- Single gene semi-quantitative PCR assays to demonstrate the presence or absence of genetic transcripts from genes of interest in stapes tissue
- RNA-sequencing to identify previously unknown genetic contributors by analysing the entire transcriptome of affected and control stapes tissue
- Quantitative PCR assays to assess the level of expression of genes of interest in stapes tissue

3) To conduct genetic analysis to identify variation in the genomic DNA of individuals affected by otosclerosis compared with controls. Techniques will include:

- Whole exome sequencing to identify single gene variants within families with a strong history of otosclerosis
- Genotyping assays to follow up findings from Whole Exome Sequencing by testing the remainder of the cohort for presence or absence of the variants

Identification of genetic factors will improve understanding of the disease process which will lead to improvements in treatment for affected individuals.

2 Materials and Methods

Materials

2.1 Chemicals and Equipment

Unless otherwise stated, all chemicals were supplied by Amersham Biosciences, BDH, Invitrogen, Promega and Sigma-Aldrich.

Restriction endonucleases were purchased from Promega, New England BioLabs or Thermo Scientific.

Unless otherwise stated, experiments were performed using autoclaved double distilled deionised MilliQ water (ddH_20).

2.2 Abbreviations

dNTPs Dideoxynucleotide triphosphates

EDTA Ethylenediamine tetra-acetic acid

TAE Tris-acetate-EDTA electrophoresis buffer

2.3 Stock Solutions

Unless otherwise stated, stock solutions were made up using autoclaved double distilled deionised MilliQ water (ddH_20).

6x DNA loading buffer	15% Ficoll 0.25% bromophenol blue 0.25% xylene cyanol
50x TAE	2.0M Tris 1M Glacial Acetic Acid 50mM EDTA (pH 8.0)
TE buffer	10mM Tris-Cl (pH8.0) 1mM EDTA

Methods

This study was approved by the London Bloomsbury Research Ethics committee (reference: 11/LO/0489). Individuals with a diagnosis of otosclerosis were recruited via informed consent by Ear Nose and Throat surgeons at the Royal Free, Royal National Throat Nose and Ear Hospital, London, U.K. Exclusion criteria included children under 16 years of age, or those with a learning disability which would be likely to cause problems involving the consent procedure.

2.4 Tissue collection

2.4.1 Collection of otosclerotic and control stapes samples

Stapes suprastructure tissue samples for gene expression analysis were obtained from 109 unrelated individuals undergoing laser-assisted stapedotomy surgery at the Royal Free, Royal National Throat Nose and Ear Hospital, London, U.K. One whole stapes bone was collected from an individual undergoing stapedectomy surgery. A visual diagnosis of otosclerosis was confirmed during surgery in all cases.

Patients were identified in advance of surgery and sent a letter of invitation, patient information sheet, consent form and questionnaire (see Appendix), inviting them to participate in the study. Those patients who completed and returned the consent form and questionnaire prior to surgery were enrolled in the study. During surgery, the stapes suprastructure, which is routinely removed during surgery, was placed in AllProtect medium (Qiagen) to prevent RNA degradation.

For the control group, stapes suprastructures were obtained from 9 unrelated individuals who did not have a diagnosis of otosclerosis, but who were undergoing a surgical procedure during which the stapes was removed. In addition, two whole stapes were obtained from both ears of one individual. The control stapes tissue was obtained during a variety of surgical procedures including surgery due to head trauma, glomus tumour and total petrosectomy. The stapes tissue was placed in AllProtect medium (Qiagen).

All stapes samples were stored overnight at 4°C to enable the medium to penetrate the tissue, before being stored long-term at -80°C.

2.4.2 Collection of blood and saliva samples from study participants

Blood samples for genetic analysis were obtained from 110 unrelated individuals undergoing stapes surgery at the Royal Free, Royal National Throat Nose and Ear Hospital, London, U.K. Two 5ml EDTA blood tubes were filled from a cannula sited in a peripheral vein, during surgery.

An additional 24 blood and 127 saliva samples were obtained from non-surgical participants. These included patients at the Royal Free, Royal National Throat Nose and Ear Hospital, London, U.K who had received a visual diagnosis of otosclerosis during a stapes surgery conducted prior to initiation of recruitment for this study. It also included individuals who had received a diagnosis of otosclerosis from audiometric studies but who had opted for conservative management to treat their hearing loss. In addition, samples were also obtained from a self-selecting population of individuals with a diagnosis of otosclerosis who were recruited via advertisements on social networking sites or in charity newsletters.

Non-surgical patients were contacted by mail and were sent a letter of invitation, patient information sheet, consent form and questionnaire (see Appendix), inviting them to participate in the study. Those patients who completed and returned the consent form and questionnaire were enrolled in the study and were asked to visit the phlebotomist during their next visit to the Royal Free, Royal National Throat Nose and Ear Hospital in London to provide a venous blood sample which was collected in two 5ml EDTA tubes. If no appointment was scheduled and it was not possible to attend the clinic, participants were sent Oragene® DNA genotek saliva collection kits by mail.

All blood samples were refrigerated at 4°C before being transferred to 2ml eppendorf tubes and stored at -80°C. Returned saliva collection kits were stored at -20°C.

2.4.3 Collection of saliva samples from relatives

From the completed questionnaires received from both surgical and non-surgical participants, individuals with a strong family history of otosclerosis were identified. The consent forms returned by these participants were consulted to ensure that these individuals had stated that they were willing to be contacted with regards to recruitment of family members. Consenting participants were contacted by telephone and family pedigrees were produced. The contact details of both affected and unaffected family members were then taken. Family members were contacted by mail and were sent a letter of invitation, information sheet, consent form and questionnaire (see Appendix), inviting them to participate in the study. Those 42 relatives

who completed and returned the consent form were enrolled in the study and were sent Oragene® DNA genotek saliva collection kits by mail. Returned saliva collection kits were stored at -20°C.

2.4.4 *Culture and harvest of MG-63 cells*

The human osteosarcoma MG-63 cell line was obtained from Dr Vehid Salih at the Eastman Dental Hospital, University College London. MG-63 cells were cultured in Dulbecco's modified Eagle medium (Invitrogen) supplemented with glutamax, 10% foetal bovine serum and 1% penicillin/streptomycin (10,000 units ml−1 penicillin and 10,000 µg ml−1 streptomycin) and were maintained at 33°C and 5% carbon dioxide. Cells were seeded at a density of approximately 1-2 x 10^5 cells per 75 cm^2 flask (Nunc) and passaged by trypsinisation when confluency reached approximately 80%. Cells were transferred to 6 well plates before being harvested and homogenised using a QIAshredder column (Qiagen), prior to cDNA preparation.

2.5 cDNA preparation

2.5.1 RNA purification

The human stapes samples were thawed and any excess medium surrounding the stapes was removed. The stapes were viewed under a dissecting microscope and any attached connective tissue or dried blood was removed using forceps. The three whole human stapes bones (one from an otosclerosis patient undergoing stapedectomy surgery and two from each ear of a single control individual) were dissected into suprastructure and footplate components under the microscope, which were then processed separately. The suprastructures from both ears of the control individual were pooled prior to processing, as were the footplates. The stapes tissue was disrupted and homogenised using the PRO200® hand held homogeniser (Bio-Gen) in in 500µl QIAzol Lysis Reagent. RNA was purified from the stapes using the RNeasy lipid tissue mini kit (Qiagen) according to the manufacturer's protocol. An on column DNase digest was performed using RQ1 RNase-Free DNase (Promega).

RNA was extracted from the homogenised human MG-63 cells using the RNeasy mini kit (Qiagen) in accordance with the manufacturer's protocol. These samples served as positive controls during semi-quantitative and quantitative real time PCR assays to demonstrate successful amplification of the desired DNA fragment.

The RNA product was eluted in 30µl of ddH_2O and the concentration was quantified using the Nanodrop® ND-1000 spectrophotometer (Thermo Scientific). Half the RNA product was archived at -80°C while the other half was immediately reverse transcribed to cDNA.

2.5.2 Reverse Transcription of RNA to cDNA

Two alternative methods were used to convert RNA to cDNA:

Method 1: Qiagen Omniscript Reverse Transcription kit

RNA from the majority of human tissue samples was converted to cDNA using the Omniscript Reverse Transcription kit (Qiagen) with random primers according to the manufacturer's protocol. The cDNA product was stored at -20°C.

Method 2: Clontech SMARTer™ kit

RNA from 12 selected human stapes samples (see chapter 5 table 5.1.) was converted to cDNA using the SMARTer™ kit (Clontech), according to the manufacturer's protocol. The cDNA product was stored at -20°C. The SMARTer™ kit is more expensive than the Omniscript kit, but has increased sensitivity for efficient conversion of small quantities of RNA to full length CDNA. This kit was therefore used only for the 12 RNA samples that would be sent for RNA-sequencing.

	Total no. samples collected	**No. samples processed using Omniscript kit**	**No. samples processed using Clontech kit**
Otosclerotic suprastructure samples	110	103	7
Otosclerotic footplate samples	1	0	1
Control suprastructure samples	10	7	3
Control footplate samples	1	0	1

Table 2.1 Table showing numbers of human stapes samples collected

Table shows numbers of otosclerotic and control suprastructures and footplates processed using Omniscript and Clontech reverse transcription kits. The one control footplate sample and one of the control suprastructure samples consisted of structures pooled from both ears of the same individual.

2.6 Gene Expression Assays

2.6.1 RNA-sequencing

The RNA isolated from the suprastructures of 7 study participants with otosclerosis and the footplate from the stapes of one of these individuals was reverse transcribed to cDNA using the SMARTer™ kit (Clontech), according to the manufacturer's protocol. In addition, RNA purified from the suprastructures of 3 additional control participants and the footplate RNA from one of these was also reverse transcribed using the SMARTer™ kit. The concentration of the cDNA generated using the Clontech SMARTer™ kit was quantified using the Nanodrop® ND-1000 spectrophotometer (Thermo Scientific) and was purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocol. The product was eluted in 50µl PE buffer. 2µl of the purified cDNA was sent to UCL Genomics for quality control using the Bioanalyzer Agilent 2100. Quality control is a crucial early step in the bioinformatics pipeline of RNA-Sequencing to ensure high quality sequences and bases in order to assure a reliable final result. The 12 cDNA samples were then sent to Otogenetics Corp (Norcross, GA, U.S.A.) where cDNA fragmentation was performed on the samples and cDNA libraries constructed before high-throughput paired-end RNA sequencing using Illumina HiSeq 2000 was carried out. Paired-end 100 nucleotide reads were aligned to human genomic assembly hg19, which generated a read count for every transcript.

2.6.2 Semi-quantitative Reverse Transcription Polymerase Chain Reaction

For all cDNAs processed using the Omniscript kit, a GAPDH assay was performed to establish cDNA integrity. All PCR assays were conducted using MG-63 cDNA in addition to the stapes cDNA as positive controls and ddH_20 as a control for background contamination. Reactions were set up on ice and contained approximately 400ng cDNA from each stapes sample or MG-63 cell cDNA, 0.5µM each primer, 0.3mM dNTPs, an optimised concentration of $MgCl_2$, 5 µl of 5x green PCR buffer and 1.25U GoTaq polymerase (Promega) in a total volume of 20µl. Primer pairs were designed to span exon-exon junctions to avoid the amplification of genomic DNA (Table 2.2). Thermal cycling was performed under the following reaction conditions: Denaturation at 93°C for 15 seconds, annealing at optimum temperature for 30 seconds, extension at 72°C for 30 seconds, for up to 40 cycles, with an initial denaturation step of 93°C for 3 minutes. PCR reactions were carried out in capped 0.5ml thin walled Eppendorf tubes or 96 well Eppendorf plates in an Eppendorf Mastercycler Gradient machine and resulting bands analysed by agarose gel electrophoresis (2.9).

Primer name	Forward or reverse primer	Gene	Primer sequence	Amplicon
hGAPDHF	Forward	*GAPDH*	GGAAGGTGAAGGTCGGAGTCA	227 bp
hGAPDHR	Reverse	*GAPDH*	GAAGATGGTGATGGGATTTCCA	
RELN1F (A)	Forward	*RELN (001)*	ACCATGTGGAGGTCGTCCTAGTA	262 bp
RELN1R (B)*	Reverse	*RELN (001)*	CACTCGGTCTTGAGAAGGGCTTTC	
RELN2F (C)	Forward	*RELN (002)*	GCAATACAGCGTCAACAACGG	209bp
RELN2R (D)	Reverse	*RELN (002)*	GTTTGCGAGTGAGGACGACCT	
RELN2F (C)	Forward	*Truncated RELN*	GCAATACAGCGTCAACAACGG	241 bp
RELN3R (E)**	Reverse	*Truncated RELN*	CCTTCCTGAAATGGACATGG	

Table 2.2 Primers used during semi quantitative RT-PCR assays

* Reverse primer B for human RELN-001 (Csomor, et al. 2011)
** Reverse primer E for human truncated RELN variant (Ovadia and Shifman 2011)

2.6.3 Quantitative Real Time PCR (qRT-PCR)

Taqman® gene expression assays were performed on human otosclerotic and control cDNA samples that had been processed using the Omniscript kit. Real-time PCR was conducted using the ABI PRISM 7500 Sequence Detection System (Applied Biosystems) with primer/probe pairs obtained from Applied Biosystems. The reactions contained 50ng stapes or MG-63 cDNA, 5µl Taqman® Fast Universal PCR Mastermix (Applied Biosystems) in a total volume of 10µl. Thermal cycling was performed under the following reaction conditions: 95°C for 10 minutes followed by 45 cycles of 95°C for 15 seconds and 60°C for 1 minute.

A Taqman® assay was performed for the detection of *RELN*-specific mRNA (chapter 6) on cDNA extracted from the 71 otosclerotic and 5 control stapes suprastructures that had been collected at the time of the study and on cDNA extracted from MG-63 cells. Assays were carried out in triplicate on all otosclerotic samples and 6 replicates were performed for control

samples and MG-63 cell cDNA. It was ensured that the Taqman® assay chosen was located in a region that could detect all *RELN* isoforms.

In addition, Taqman® assays were performed for the detection of *ASPN, FBLIM1, FKBP14, LAIR1, LUM* and *SPARCL1*-specific mRNA, to validate the results of the RNA-sequencing study (chapter 5), on cDNA extracted from 81 otosclerotic and 5 control stapes. 10 additional stapes had been collected by the time that this study was performed since the *RELN* study. Assays on otosclerotic samples were performed once, as each otosclerotic sample was treated as a replicate. However, due to the small number of controls collected, assays on controls were carried out in triplicate.

Taqman® assays were performed for detection of *SERPINF1*-specific mRNA (chapter 4) on cDNA extracted from 81 otosclerotic and 5 control stapes. Two assays were performed, one with a target located in the upstream region of the gene and another in the downstream region, in order to make a comparison between expression levels of specific exons. These assays were performed on 81 otosclerotic and 5 control stapes with assays on the control samples being performed in triplicate. On 3 otosclerotic samples with known mutations in the *SERPINF1* gene, six replicates were performed.

In all assays, eukaryotic 18s rRNA was used as the endogenous control. For Applied Biosystems assay ID numbers, see Table 2.3

Gene	Target exons	Applied Biosystems Assay ID
ASPN	7-8	Hs00214395
FBLIM1	2-3	Hs00378472
FKBP14	1-2	Hs00215735
LAIR1	9-10	Hs00253790
LUM	2-3	Hs00929860
RELN	62-63	Hs01022646
SERPINF1	3-4	Hs011006934
SERPINF1	6-7	Hs011006937
SPARCL1	10-11	Hs00949881

Table 2.3 Table showing Applied Biosystems assay ID numbers for each Taqman® assay performed

2.7 Genomic DNA preparation from blood and saliva

Blood was thawed at room temperature and genomic DNA was isolated using Illustra™ DNA Extraction kit BACC2 (GE Healthcare) according to the manufacturer's protocol. Saliva samples were thawed and genomic DNA was isolated using the DNA genotek prepIT™.L2P kit according to the manufacturer's protocol.

The DNA pellet was re-suspended in between 100µl and 400µl 0.1x TE buffer according to the size of the pellet, and placed on a rotator for 2 days at 4°C. The gDNA was stored at -20°C.

2.8 Detection of genomic variation

2.8.1 Whole Exome Sequencing

Whole Exome Sequencing was performed on 10 genomic DNA samples at The Wellcome Trust Sanger Institute, Cambridge, thanks to collaboration with Professor Karen Steel.

Prior to sequencing, the concentration of the gDNA generated using the Illustra™ DNA Extraction kit BACC2 (GE Healthcare) or Oragene® DNA genotek kit was quantified using the Nanodrop® ND-1000 spectrophotometer (Thermo Scientific) and was purified using a DNeasy blood and tissue kit (Qiagen) according to the manufacturer's protocol. The product was eluted in 60µl EB buffer (Qiagen). The purified gDNA was sent to The Wellcome Trust Sanger Institute, Cambridge. After target enrichment, sequencing was performed using the Illumina HiSeq 2000 with 100-bp paired-end reads. Reads were mapped to the reference human genome (GRCh37_hs37d5) and both single-nucleotide polymorphism (SNPs) and insertions-deletions (indels) were identified using SAMtools (http://samtools.sourceforge.net/). Following prioritisation of variants for subsequent analysis, variants of interest were confirmed in the probands of each family exome sequenced via genotyping assays including restriction endonuclease digests or Sanger sequencing (2.8.2).

2.8.2 Genotyping

Two different genotyping methods were used to identify the presence of genomic variants within gDNA samples; Sanger sequencing and restriction endonuclease digests. For both techniques, the genomic region of interest, in most cases an exon containing known or predicted sequence variants, was amplified through semi-quantitative PCR prior to Sanger sequencing or restriction enzyme digests.

2.8.2.1 Semi-Quantitative PCR

For all gDNA samples, a GAPDH assay was performed to establish gDNA integrity and ddH_20 was used as a control for background contamination. Reactions were set up on ice and contained approximately 200ng gDNA from each blood or saliva sample, 0.5µM each primer, 0.3mM dNTPs, an optimised concentration of $MgCl^2$, 5 µl of 5x colourless PCR buffer and 1.25U GoTaq polymerase (Promega) in a total volume of 25µl. Primers were designed where possible in the introns surrounding the exon of interest and at least 50 bases away from the variant to be genotyped in order that they could also be used as sequencing primers (Table 2.4). In some cases, where exons were large, it was necessary to locate the primers within the exon itself as an amplicon greater than 500 bases is not desirable for effective sequencing. In the case of *TRIM17*, two pairs of overlapping amplicons were designed in order to capture the full exonic sequence. Thermal cycling was performed under the following reaction conditions: Denaturation at 93°C for 15 seconds, annealing at optimum temperature for 30 seconds, extension at 72°C for 30 seconds, for up to 40 cycles, with an initial denaturation step of 93°C for 3 minutes. PCR reactions were carried out in capped 0.5ml thin walled Eppendorf tubes or 96 well Eppendorf plates in an Eppendorf Mastercycler Gradient machine and resulting bands analysed by agarose gel electrophoresis (2.9).

Primer name	Forward or reverse primer	Gene	Primer sequence	Amplicon
COL1A2F	Forward	*COL1A2*	ACTGAGAACCAGAGTCAACCAC	506bp
COL1A2R	Forward	*COL1A2*	GCCGTTTGCTCCAGGATTAC	
CYP2D6F	Reverse	*CYP2D6*	CTCCATCACAGAAGGTGTGAC	505bp
CYP2D6R	Reverse	*CYP2D6*	CTCACACCTCCCTAGTGCAG	
DNAH5F	Forward	*DNAH5*	GTTAGGGTGGAATAAGCCTCA	469bp
DNAH5R	Reverse	*DNAH5*	AAGAATGGAAGACGTACGTGG	
EGLN3F	Forward	*EGLN3*	GTACCATGCATTTGGAGTGG	446bp
EGLN3R	Reverse	*EGLN3*	CTGAGGTTAGGTTTGATATGCT	
ERCC6F	Forward	*ERCC6*	CTGGCTGTTCATCTTCCTGAGT	443bp
ERCC6R	Reverse	*ERCC6*	CTCACAGTAAGACATCTAAGC	
FOXK1F	Forward	*FOXK1*	ACCGTCTGAATGCGGAGGACA	429bp
FOXK1R	Reverse	*FOXK1*	ATGATCACTGGCTGGGCGCTGA	
FRYLF	Forward	*FRYL*	CCACGTGGATTTGGATAAAGT	505bp
FRYLR	Reverse	*FRYL*	GTGAACTATCCTTCTCCCTTAC	
FZD2F	Forward	*FZD2*	TCCTGTGCTCCATGTACGCA	490bp
FZD2R	Reverse	*FZD2*	AGCACCGACCAGGTGAGGAT	
GNGT1F	Forward	*GNGT1*	CAAGGCTAGTGTGCATTGC	520bp
GNGT1R	Reverse	*GNGT1*	GGATTGGGCTTGGCCAGAG	
LEPRE1F	Forward	*LEPRE1*	CAGTCATCTCCCACCAGGTCC	376bp
LEPRE1R	Reverse	*LEPRE1*	GGACCGCATCCTGTTCTCTCTG	
MIA3F	Forward	*MIA3*	CTGCAGCTACAGATTCTGAG	416bp
MIA3R	Reverse	*MIA3*	TACAAGTGCATCAGCTGTTGAG	
mir183F	Forward	*mir183*	GCAAGGCCAGAAGGTCAGCT	473bp
mir183R	Reverse	*mir183*	GCCAAGCAGATGGCACTGGTG	
PDLIM5F	Forward	*PDLIM5*	TGACCTCATGATCCGCCCA	500bp
PDLIM5R	Reverse	*PDLIM5*	TGACAACTACCAATCAGTCTC	
PTK6F	Forward	*PTK6*	CTCCCTGTGCTGGCTTCAG	320bp
PTK6R	Reverse	*PTK6*	GGGCTCGAGGCCAGAGGTC	

SERPINF1Exon2F	Forward	*SERPINF1*	AGTGACTAGCCCTGCCCAAC	309bp
SERPINF1Exon2R	Reverse	*SERPINF1*	AAGCCTGGCCTGGAACCT	
SERPINF1Exon3F	Forward	*SERPINF1*	CGTGAGGAGACAGTCCCTGT	345bp
SERPINF1Exon3R	Reverse	*SERPINF1*	TCAGCCACGTTTACGCAGAGG	
SERPINF1Exon4F	Forward	*SERPINF1*	GCCTACTTGGGCTCTCAGCAGA	311bp
SERPINF1Exon4R	Reverse	*SERPINF1*	ACATGCCTCAGGCAACTTGG	
SERPINF1Exon5F	Forward	*SERPINF1*	TGCTGAGCGCTAAACCAGAAC	323bp
SERPINF1Exon5R	Reverse	*SERPINF1*	AGAATTGGAGACGCGCTCACC	
SERPINF1Exon5bF	Forward	*SERPINF1*	AGATGCTGGCTGGGAAGTCAG	440bp
SERPINF1Exon5bR	Reverse	*SERPINF1*	CTGAGATCGCACCACTGCACTC	
SERPINF1Exon6F	Forward	*SERPINF1*	CTGACAGCTAAGCTCCCTTGA	369bp
SERPINF1Exon6R	Reverse	*SERPINF1*	TGTAGACAGCTGTCGGATCTCA	
SERPINF1Exon7F	Forward	*SERPINF1*	CTGGATGAAGGACGAGACCA	420bp
SERPINF1Exon7R	Reverse	*SERPINF1*	AGCCCTTGCGTTCTGCTTAGC	
SMAP1F	Forward	*SMAP1*	ACATTGGTCTCTACCCGCTGG	451bp
SMAP1R	Reverse	*SMAP1*	GCATCTTAACCCTGTGATTCTC	
TRIM17aF	Forward	*TRIM17*	GACCCTGAATGCCCCACACCT	419bp
TRIM17aR	Reverse	*TRIM17*	CTGGGAGGCTCCATCAGCATG	
TRIM17bF	Forward	*TRIM17*	CAAGGGGACCAAGTACTTATCC	586bp
TRIM17bR	Reverse	*TRIM17*	GAGCCACAGATTCCTGTCATTC	
VPS53F	Forward	*VPS53*	AGTAGAGACCTGACCTCAGGTG	299bp
VPS53R	Reverse	*VPS53*	CAAGATGTGCCAAAAGCTGCTG	
ZNF225F	Forward	*ZNF225*	TACCCATCGTAGAGTCCACAGTGGA	447bp
ZNF225R	Reverse	*ZNF225*	TCTCTACTGTGGAGTCTCTGATGG	

Table 2.4 Primers used for RT-PCR during genotyping and Sanger sequencing

2.8.2.2 Sanger sequencing

Sanger sequencing was performed by Source BioScience Life Sciences. Approximately 100ng purified PCR product (2.10) was sequenced in each instance. Sequencing primers were in all cases either the forward or reverse primer that had been used during amplification of the exon or genomic region containing the variant to be genotyped. These primers are shown in Table 2.4.

DNA sequencing chromatograms were visualized using Finch TV version 1.4.0 Geospiza©. During segregation analysis in otosclerosis families and identification of specific known variants in a cohort of familial otosclerosis probands, chromatograms were analysed for the presence of heterozygous variants by observation of a double peak on the chromatogram at the position of the known variant, indicating the presence of two variant alleles. Individuals were also checked for homozygosity of the variant at this site, although this was expected to be unlikely due to the rarity of the variants analysed. During investigation for the presence of unknown variants within an exon, the whole chromatogram representing the full exonic region was analysed by eye for the presence of double peaks indicative of a heterozygote. Chromatograms were also checked for the presence of homozygous variants for known common alleles within the exon, as these would not be easily visualised by a double peak.

2.8.2.3 Restriction endonuclease digestion

Restriction endonuclease digests were used to detect the presence of a specific variant within a PCR product. Typically 1µg of PCR product was digested with 5 units of the appropriate enzyme in 1x of the appropriate restriction enzyme buffer in a total volume of 20µl. Reactions were typically left for 1-2 hours or overnight at 37°C unless a different optimal temperature was specified. Resulting DNA fragments were analysed by gel electrophoresis (2.9) against a digested sample from a known heterozygote which acted as a positive control. The restriction endonuclease digests performed are summarised in Table 2.5.

Gene	Variant	PCR amplicon length	Restriction Endonuclease	Fragment sizes wt	Fragment sizes het	Fragment sizes mutant
CYP2D6	c.345delC	505bp	DrdI	505bp	505+278+226bp	278+226bp
DNAH5	c.7883T>G	469bp	OliI	469bp	469+319+150bp	319+150bp
EGLN3	c.496C>T	446bp	AciI	308+138bp	446+308+138bp	446bp
mir-183	c.81G>T	473bp	AvaII	473bp	473+303+170bp	303+170bp
PTK6	c.652A>T	320bp	MlsI	207+113bp	320+207+113bp	320bp
SERPINF1	c.441G>C	440bp	TseI	314+126bp	440+314+126bp	440bp
TRIM17	c.915C>G	419bp	RsaI	334+85bp	419+334+85bp	419bp
VPS53	c.107C>G	299bp	PpuMI	182+117bp	299+182+117bp	299bp

Table 2.5 Table showing predicted fragment sizes from restriction endonuclease digests

Table shows the predicted fragment sizes for wild-type, heterozygous and homozygous mutant sequences with each restriction endonuclease.

2.9 Agarose gel Electrophoresis

DNA bands were visualised by agarose gel electrophoresis on 1.5-3% agarose gels, made using 1x TAE buffer and 0.25µg/ml ethidium bromide. 1x DNA loading buffer was added to each sample where appropriate and was electrophoresed alongside 5µl 1kb or 100bp DNA ladder (Promega) at approximately 80mA for 30 minutes. DNA bands were visualised by exposure to ultra-violet light using a 3UV™ Transilluminator GelDoc-It imaging system (Jencons-PLS).

2.10 Purification of PCR products

Specific PCR products were purified using either the QIAquick PCR purification kit (Qiagen) with a microcentrifuge, or the QIAquick 96 PCR purification kit (Qiagen) along with the QIAvac 96 vacuum manifold and a vacuum pump, according to the manufacturer's instructions. The purified PCR product was eluted in 30µl of Qiagen Buffer EB (10 mM Tris·Cl,pH 8.5).

2.10.1 Purification of PCR product from low melting point agarose gels

Specific PCR products were recovered using standard agarose gel electrophoresis (2.9) with the use of low melting point agarose. The gel was exposed to UV light under a TFX-20M UV Transilluminator (BDH) and the required bands excised using a scalpel. PCR products were purified using the QIAquick gel extraction kit (Qiagen) with a microcentrifuge according to the manufacturer's instructions. The purified PCR product was eluted in 30µl of Qiagen Buffer EB (10 mM Tris·Cl,pH 8.5).

3 Cohort Characteristics

3.1 Overview

Throughout this chapter, the data collected from the questionnaire responses of study participants will be analysed and evaluated in the context of the themes discussed in Chapter 1 of this thesis. The analysis will be focussed on three key areas; (i) the clinical characteristics of otosclerosis in the cohort members; (ii) exposure to environmental factors predicted to be involved in the disease process; and (iii) the frequency of other diseases and disorders in members of the cohort. This analysis will provide a general overview of the demographics of the cohort recruited, in order to establish whether the cohort is typical of a group of individuals with otosclerosis. It could also enable the identification of trends in the data and indicate reasons for susceptibility to this form of conductive hearing loss.

3.2 Results

3.2.1 Collection of DNA samples from study participants

Patient recruitment was opened in July 2011 after ethical approval had been received from the London Bloomsbury Research Ethics Committee. As of July 2013, 263 otosclerosis patients had been recruited to this study. Human tissue samples collected include; 110 otosclerotic stapes suprastructures; one otosclerotic footplate; 11 control stapes suprastructures; and two control footplates. In addition, 138 blood samples and 127 saliva samples were collected from otosclerosis probands, totalling 263 genomic DNA samples. Saliva samples from 42 family members of selected probands with a strong family history of the condition were also obtained (Figure 3.1).

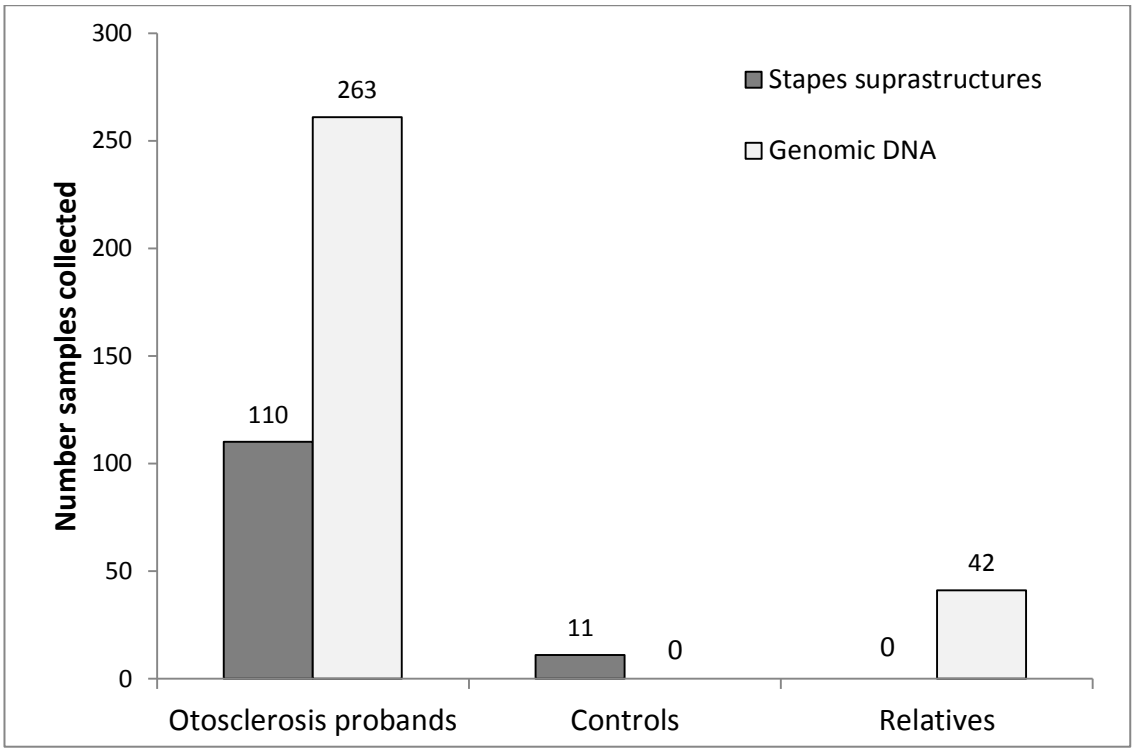


Figure 3.1 Bar graph showing numbers of stapes and genomic DNA samples collected

3.2.2 Characteristics of the otosclerosis cohort

Of the 263 study participants recruited, 186 were female and 77 male. The cohort therefore displays a female to male ratio of 2.4:1. The average age of male particpants was 55, the average of female participants was 52 and the overall average for the cohort was 53.

Each study participant was sent a questionnaire in the post (see Appendix). The questionnaire was completed and returned by 232 of the 263 participants, representing a questionnaire response rate of 88.2%. The questionnaire was designed to capture the characteristics of the cohort including their age of symptom onset, ethnicity and whether or not they have a strong family history of otosclerosis. It also collected details of participants' experience of unilateral or bilateral hearing loss, tinnitus and vertigo.

Of the 232 questionnaire respondents, 157 (67.7% of respondents) reported the first symptoms of hearing loss between the ages of 21 and 40, 11 (4.7% of respondents) were under the age of 10 at onset and just one (0.4% of respondents) was over the age of 61 (Figure 3.2).

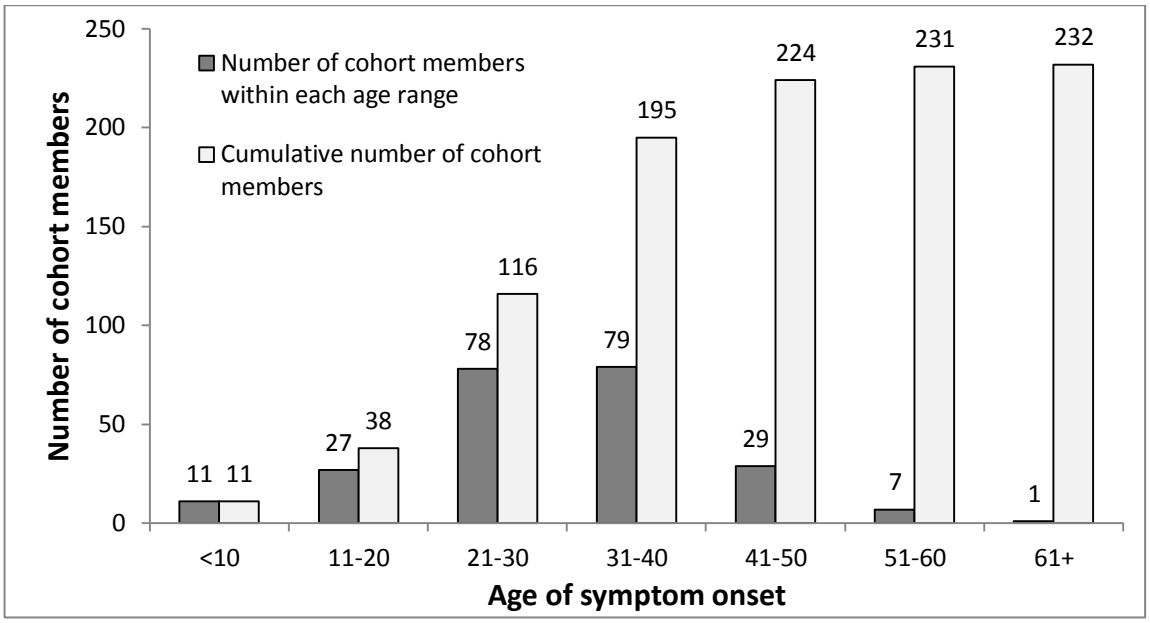


Figure 3.2 Bar graph showing age of hearing loss onset of questionnaire respondents

A comparison of the age of hearing loss onset in males and females showed that a greater percentage of females exhibited symptoms at a young age (Figure 3.3). This difference was most marked in the 21-30 age range, where 35.1% of females developed symptoms in comparison to just 29.5% of males. In contrast, males were more likely to develop hearing loss at a more advanced age, in particular in the 41-50 age range where 18.0% of males developed symptoms, in comparison to 10.5% of females. However, when a Student's two-tailed paired-end t-test was conducted to compare difference in age of onset distribution between males and females, there was no significant difference(p=0.44).

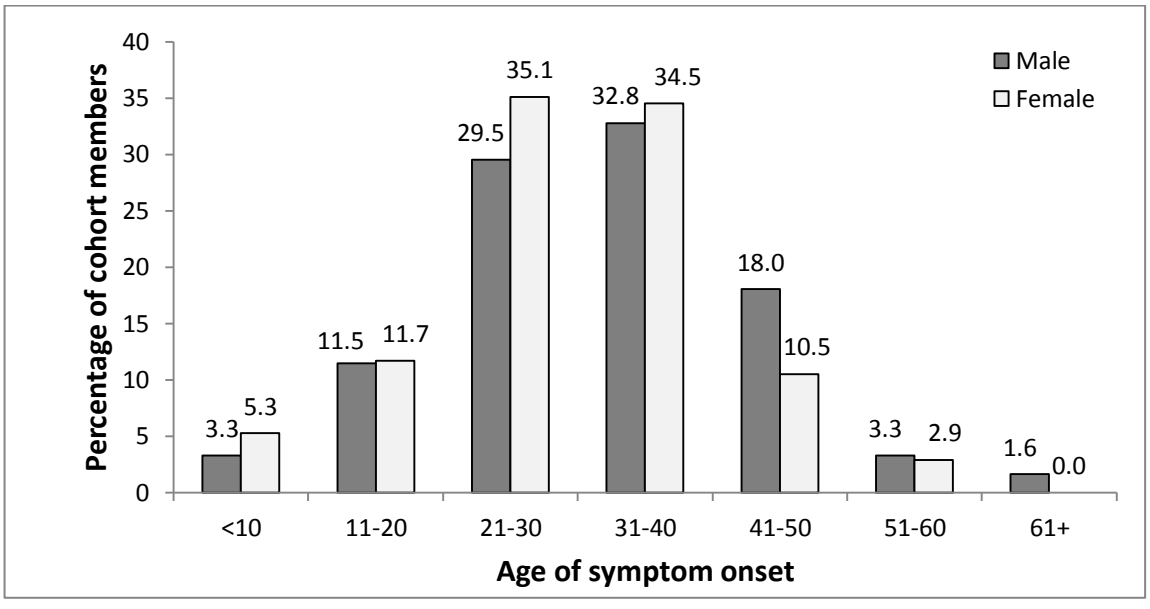


Figure 3.3 Bar graph showing age of hearing loss onset of male and female questionnaire respondents

The ethnicity of respondents was wide ranging. Of the respondents, 173 (74.6% of respondents) described themselves as British or Irish, 24 (10.3% of respondents) identified themselves as being of other European origin, 11 (4.7% of respondents) were of Asian origin, five (2.2% of respondents) of African or Caribbean origin and 19 respondents were of other ethnic origins (Figure 3.4).

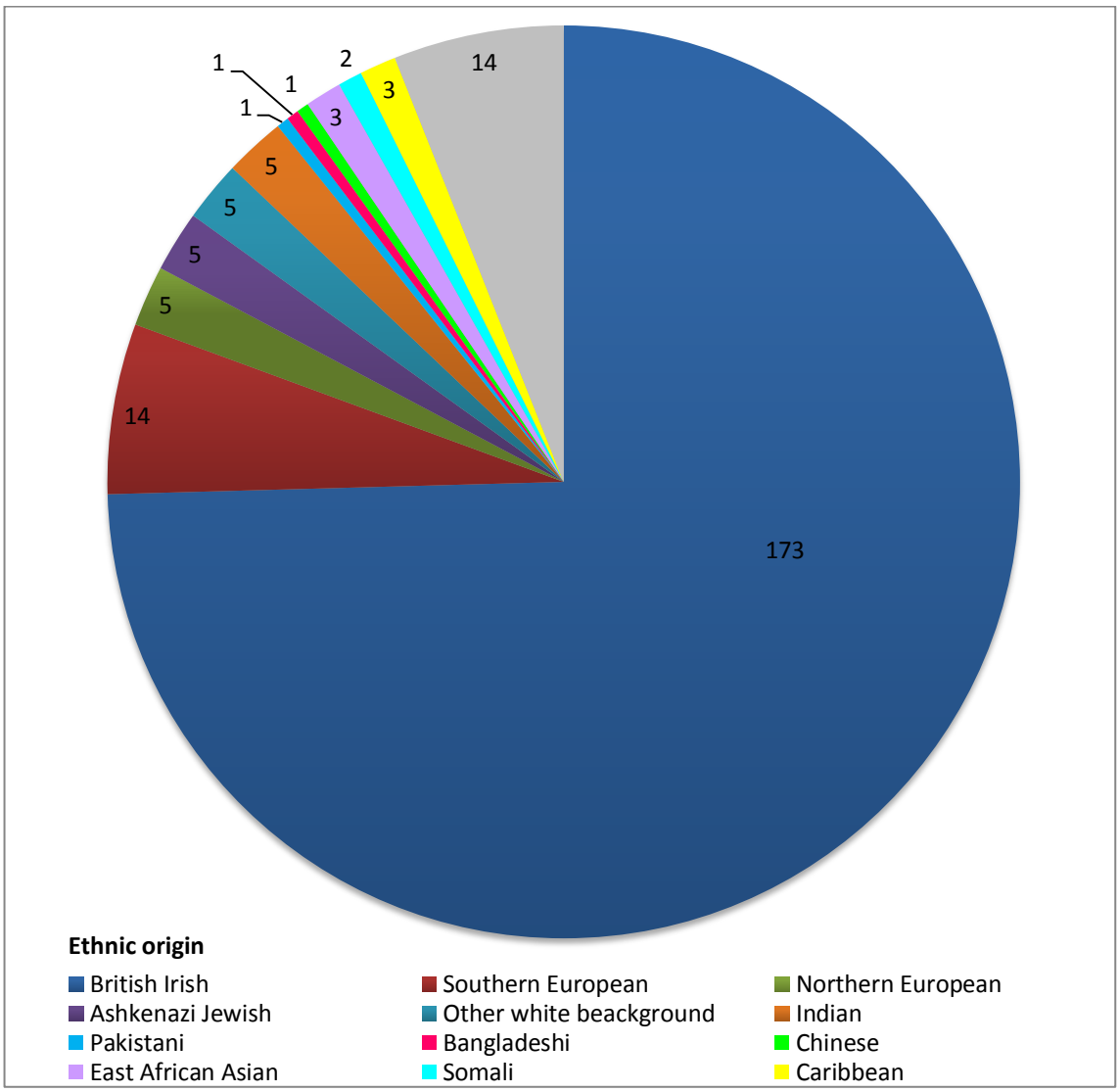


Figure 3.4 Pie chart showing ethnicity of questionnaire respondents

Of the 232 respondents, 41 (17.7% of respondents) reported a strong family history of otosclerosis with two or more affected relatives. An additional 41 (17.7% of respondents) had some family history of the condition with just one affected relative. A further 97 (41.8% of respondents) reported a family history of other forms of hearing loss, including conditions resembling otosclerosis without formal diagnosis. No family history of hearing loss was reported by 52 respondents (22.4% of respondents) (Figure 3.5).

Overall, a greater proportion of female than male respondents had a strong family history of otosclerosis. However, a greater proportion of males than females reported just one affected relative. A substantially greater proportion of females than males reported a family history of other forms of hearing loss aside from otosclerosis, whilst a greater proportion of males than females indicated that they had no family history of hearing loss. One study participant was adopted into her family and so did not know about the relevant medical history of her biological relatives (Figure 3.5).

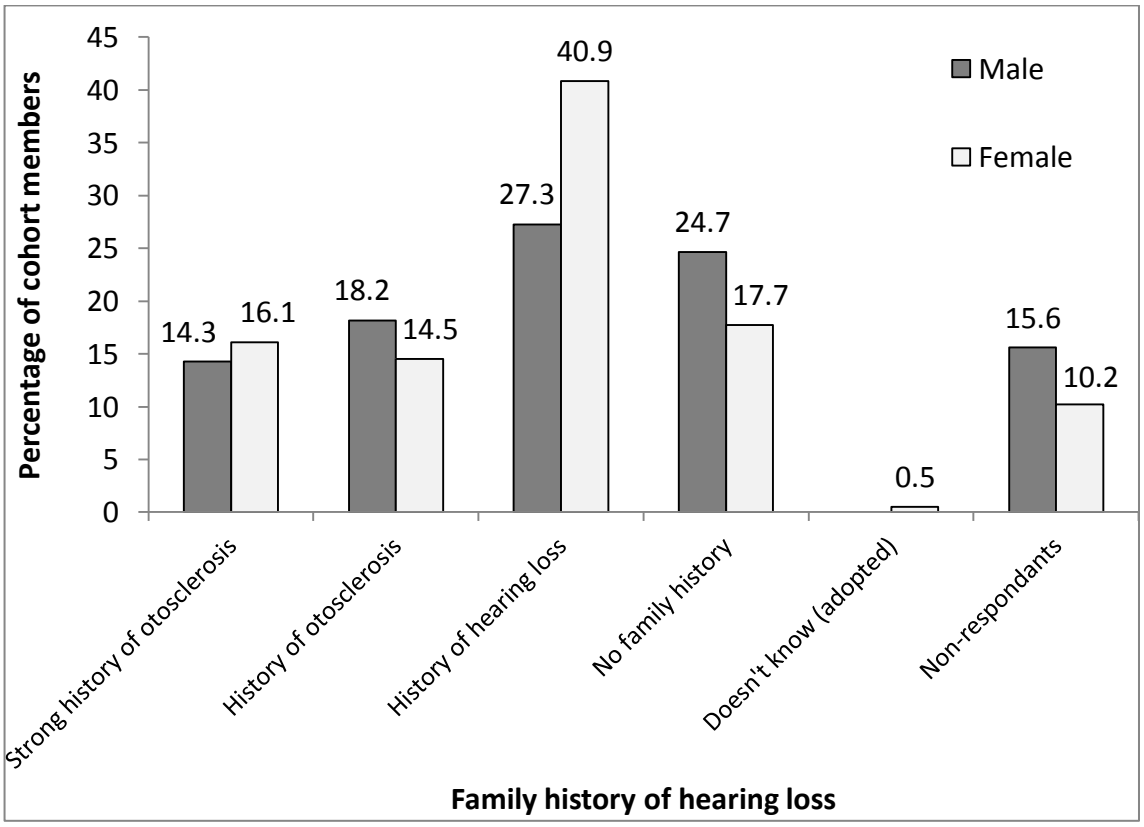


Figure 3.5 Bar graph showing family history of hearing loss in male and female questionnaire respondents

Analysis of the questionnaire responses showed that the majority of study participants experienced bilateral hearing loss, with 162 (69.8% of respondents) reporting hearing loss in both ears. Tinnitus symptoms were reported by 158 participants (68.1% of respondents). A smaller proportion of the cohort, 72 respondents (31.0% of respondents), reported symptoms of vertigo (Figure 3.6).

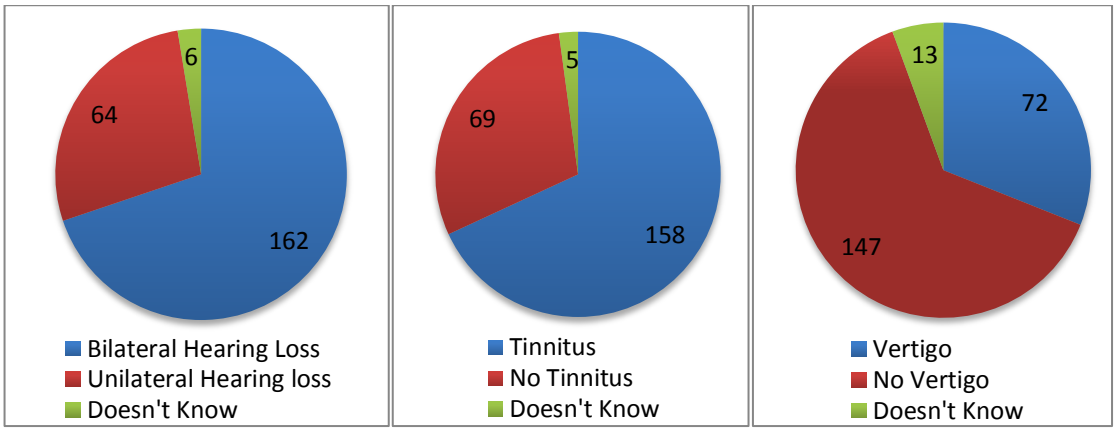


Figure 3.6 Pie charts showing proportion of respondents with bilateral hearing loss, tinnitus and vertigo

3.2.3 Environmental factors to which cohort members have been exposed

The questionnaire also collected information on environmental factors that could have influenced a participant's susceptibility to developing otosclerosis. This information included data on whether respondents had experienced measles, mumps or rubella infection and whether they had been vaccinated against these three viruses. The questionnaire also collected information in relation to periods of residence abroad, in order to identify respondents who had lived in countries with exceptionally high or low fluoride content in the drinking water. In addition, data was collected from women in relation to pregnancy, breastfeeding, use of oral contraceptives and hormone replacement therapy.

Of the 232 questionnaire respondents, 125 had been infected by measles (53.9% of respondents), whilst 91 (39.2% of respondents) had received a vaccination against the virus (Table 3.1). Trends in the data showed that with decreasing age of respondents there was an increase in the number of cohort members who had received the measles vaccination. All participants under the age of 22 had been vaccinated against measles, whilst all respondents over the age of 82 had been infected with the virus (Figure 3.7). Fewer respondents had been infected by mumps than measles, with 102 (44.0% of respondents) reporting infection. Of the respondents, 79 were vaccinated against mumps (34.1% of respondents). Rubella infection was the least common of the three viruses, with just 57 participants (24.6% of respondents) having been infected, although 104 (44.8% of respondents) were vaccinated against the virus. A large proportion of respondents did not know whether or not they had been infected by the viruses and an even larger proportion were unsure if they had received vaccinations (Table 3.1).

	Measles		Mumps		Rubella	
	Infection	**Vaccination**	**Infection**	**Vaccination**	**Infection**	**Vaccination**
Yes	125	91	102	79	57	104
No	64	79	81	82	110	64
Doesn't Know	43	62	49	71	65	64

Table 3.1 Table of the number of questionnaire respondents infected by and vaccinated for measles mumps and rubella viruses

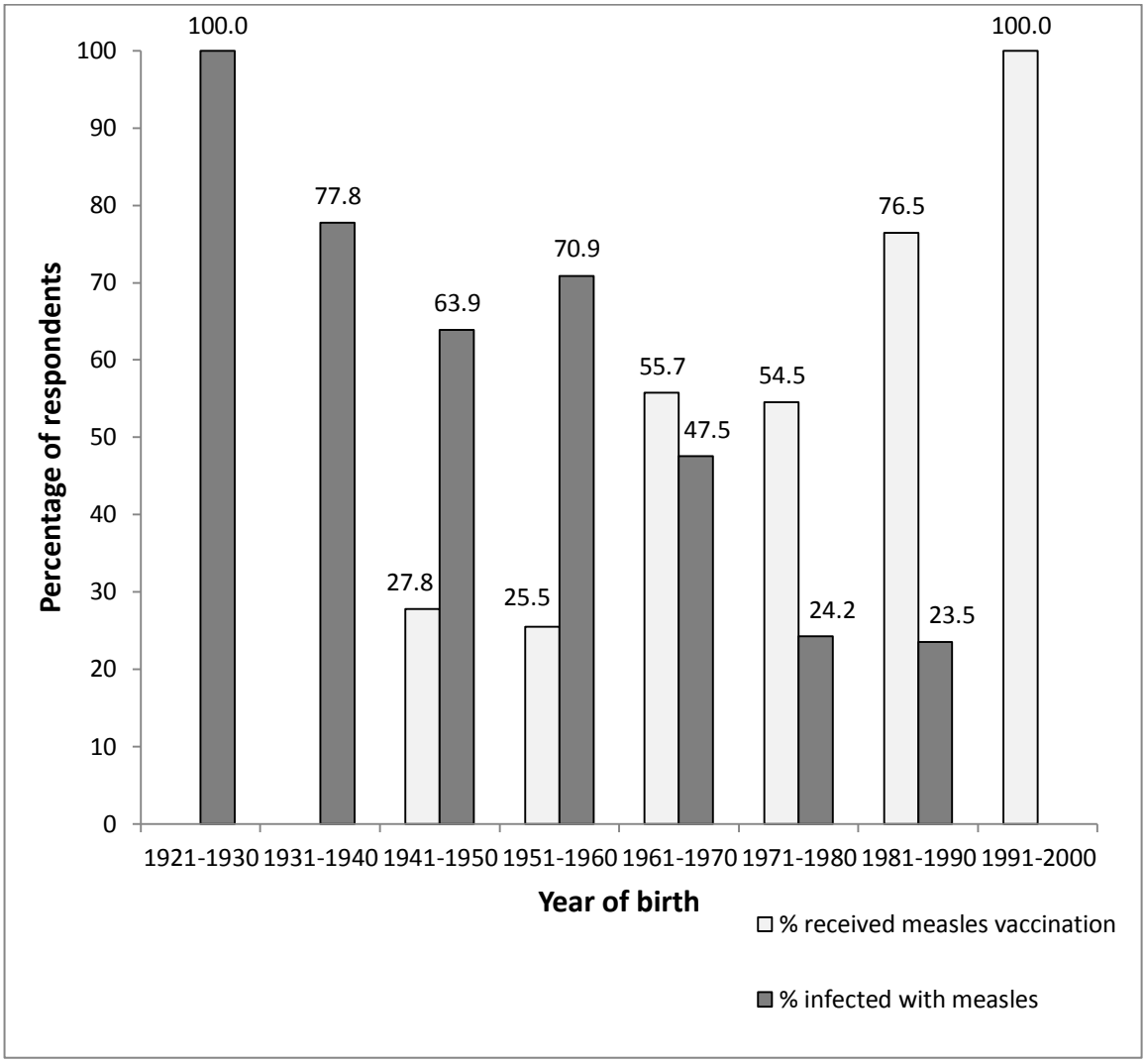


Figure 3.7 Bar graph showing percentage of respondents born in each decade vaccinated for and infected by measles

Analysis of the questionnaire responses identified the number of respondents who had lived abroad for a period of 5 years or more. This data was collected in order to identify potential links between fluoride content in drinking water and onset of otosclerosis. Table 3.2 shows that nearly 10% of the cohort have lived outside the UK within Europe during their lifetime, whilst nearly 16% have lived in non-European continents.

Continent/region	no. respondents who have lived abroad	% respondents who have lived abroad
Africa	8	3.45
Asia	6	2.59
Australia	10	4.31
Caribbean	1	0.43
Europe	22	9.48
Middle East	4	1.72
North America	7	3.02
South America	1	0.43

Table 3.2 Table of number and percentage of respondents who have lived outside the UK for over five years

All consenting females were requested to complete some additional questions regarding pregnancy, breastfeeding, contraceptives and hormone therapy to establish any links between otosclerosis and hormones, as has been hypothesised in published literature on the topic. Of the 186 female study participants, 170 responded to the questionnaire circulated to all participants. Of these respondents, 118 (69.4% of female respondents) had been pregnant, with two pregnancies the most common number. Of the 118 respondents who had been pregnant, 98 (83.1%) had breastfed and 47 (39.8%) had noticed a deterioration in hearing during or shortly after their pregnancy (Figure 3.8).

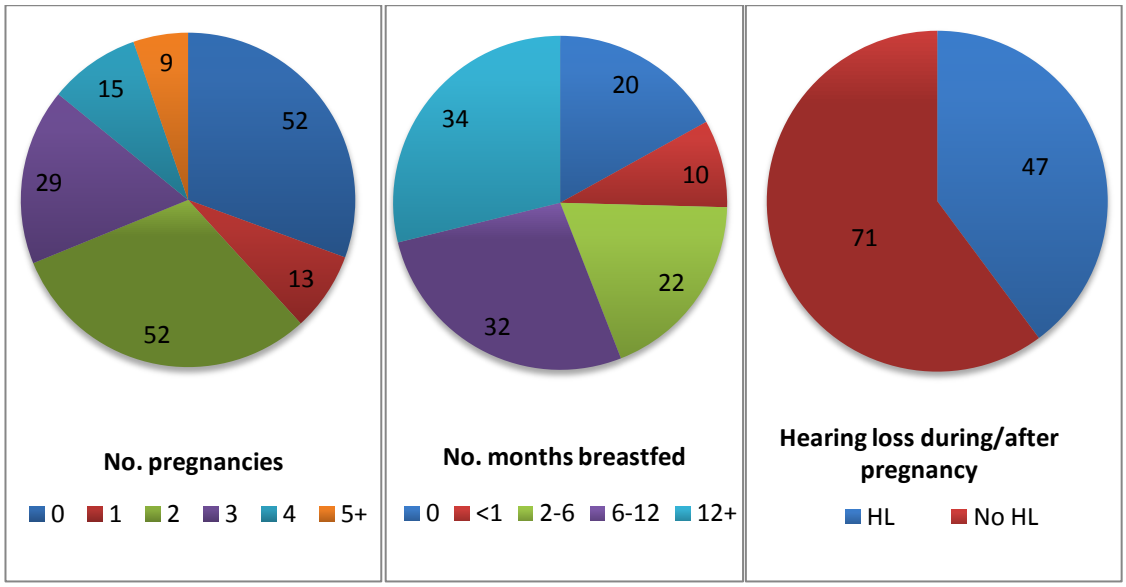


Figure 3.8 Pie charts showing number of pregnancies in female respondents, number of months for which respondents had breastfed, and proportion of respondents who developed hearing loss during or after pregnancy

Of all 170 female respondents, 126 (74.1% of respondents) had taken oral contraceptives with 40% of these 126 having taken them for longer than 5 years (Figure 3.9).

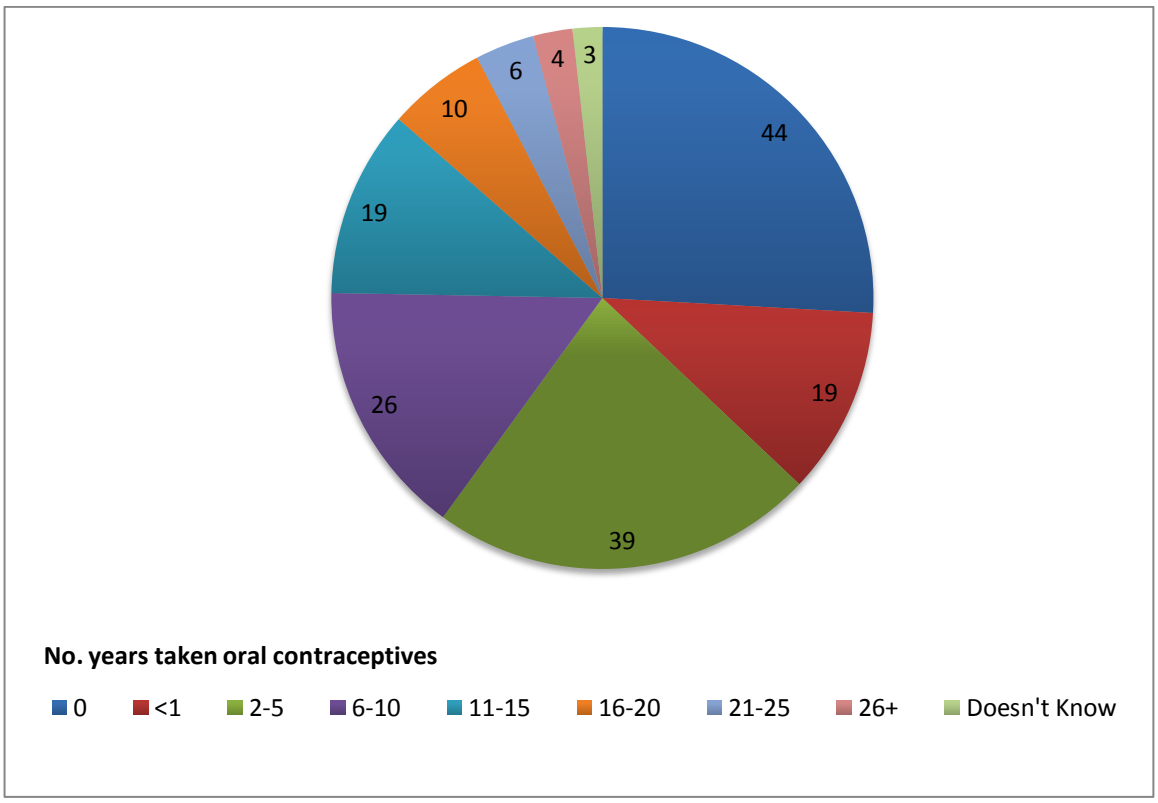


Figure 3.9 Pie chart showing number of years for which female respondents had taken oral contraceptives

Eighty individuals (47.1% of respondents) had experienced the menopause, with 25 (14.7% of respondents) having taken or currently taking hormone replacement therapy (HRT) (Table 3.3).

No. years taking HRT	No. respondents
<1	7
2-5	7
6-10	6
11+	5

Table 3.3 Table showing number of years for which questionnaire respondents had taken HRT

3.2.4 Incidence of autoimmune disease and connective tissue disorders in cohort members

The questionnaire also requested information regarding particpants' medical history including conditions that may have links with otosclerosis. Data was collected on autoimmune conditions, broken bones and connective tissue disorders. It was found that 39.2% of the 232 respondents of both genders had broken a bone at some point in their lifetime. History of an autoimmune condition was reported by 18.1% of respondents. This is approximately double the predicted prevalence of autoimmune conditions in the general population of 7.6-9.4%, a prevalence range derived from a large recent meta-analysis (Cooper and Stroehla, 2003) (Table 3.4). When a Chi squared test was performed to compare the prevalence of autoimmune conditions in the otosclerosis cohort and the general population based on the greater 9.4% prevalence statistic derived form this meta-analysis, the difference was found to be highly significant (X^2=20.6, df=1, p=5.64x10^{-6}). In addition, a further four respondents (1.7% of respondents) reported a connective tissue disorder, including two cases of Osteogenesis Imperfecta and one case each of Osteomalacia and Plantar fasciitis.

Autoimmune condition	no. respondents affected	% of questionnaire respondents affected	Prevalence in general population
Unspecified Autoimmune Condition	21	9.05%	n/a
Bullous pemphigoid	1	0.43%	Data unavailable
Graves' disease	2	0.86%	1.15%
Lichen Sclerosus	2	0.86%	Data unavailable
Multiple Sclerosis	1	0.43%	0.06%
Osteoporosis	14	6.03%	4.34%
Sjogrens disease	1	0.43%	0.01%
Total	42	18.10%	3.23%

Table 3.4 Table of number and percentage of respondents who reported history of autoimmune conditions

For data on the prevalence of the condition in the general population see (Cooper and Stroehla, 2003).

3.3 Discussion of cohort characteristics

3.3.1 Ascertainment of DNA from study participants

Over the course of the two year recruitment period from July 2011 to July 2013, 263 otosclerosis patients donated genomic DNA samples for the benefit of this research study. This is thought to be the second largest collection of genomic DNA samples from otosclerosis patients worldwide, second only to the largest collection held by the Department of Medical Genetics at the University of Antwerp in Belgium. In addition, the 110 otosclerotic stapes suprastructures collected during this study are thought to be the largest collection of non-formalin-fixed otosclerotic stapes samples collected to date, consisting of more stapes samples than the largest published collection of 85 samples held at the University Medical School of Debrecen in Hungary. The combined genomic DNA and stapes collections for this study are thought to be the largest global collection of overlapping samples from the same individuals,

with both stapes cDNA and genomic DNA held from 110 otosclerosis patients. This collection therefore offers the first opportunity to validate the results of genetic studies by conducting gene expression studies to identify the impacts of genetic variants on gene expression in specific otosclerotic stapes.

Recruitment for this study was initially slightly slower than expected, as initial predictions suggested that 100 stapes surgeries were performed per year at the Royal National Throat Nose and Ear hospital in London. In reality, this prediction had failed to account for the large number of patients with bilateral hearing loss who had both ears operated on during that time, so was an overestimate of the true number of participants who could be recruited. To overcome this, recruitment was expanded to two new UK sites at the Sunderland Royal Hospital and the Freeman Hospital in Newcastle-upon-Tyne, where as of July 2013, the first consenting participants were recruited. This research study is now the first multi-centre study in the UK recruiting otosclerosis patients from around the country.

During the recruitment phase of the study, a large number of probands reporting a strong family history of otosclerosis that also had large numbers of living affected relatives were recruited. Following successful attempts to contact both affected and unaffected family members, a total of 42 relatives were recruited from 16 unrelated otosclerosis families, in addition to the 16 probands from each of these families. This cohort is therefore extremely useful for investigations into genetic factors involved in familial otosclerosis.

One obstacle encountered during sample collection for this study was a difficulty in gaining access to large numbers of control stapes as there are very few medical scenarios during which the stapes bone is removed routinely. As of July 2013, 11 control suprastructures and 2 control footplates had been collected which came predominantly from individuals undergoing total petrosectomy surgery or those receiving surgery following head trauma. There was also difficulty in gaining access to stapes footplates. The collection of footplates is not easy in the UK since the laser assisted stapedotomy procedure, which leaves the footplate in place during surgery, is the favoured surgical technique. However, it is important to compare gene expression in footplate and suprastructure as the footplate is the predominant location of the otosclerotic foci, where it is most likely that aberrant gene expression will occur in otosclerosis. To overcome this, a collaboration has been established with the University of Athens in Greece, where the stapedectomy surgery, during which the whole stapes is removed, is routinely performed. This will enable access to be gained to large number of stapes footplates, which will be useful to validate any future gene expression studies conducted using stapes suprastructures.

3.3.2 Analysis of cohort characteristics

The gender distribution of the cohort recruited displays a female to male ratio of 2.4:1. This is roughly in line with reports from the 1930s which suggested that twice as many females as males are affected by otosclerosis (Davenport, 1933). However this fact has been widely disputed (Larsson, 1960, Morrison, 1967) with claims that the gender bias is not a true reflection of incidence but is caused by the fact that females are more likely to be seen in clinical practice, as women are more likely than men to consult their General Practitioner regarding hearing loss, resulting in more referrals to the clinic (Browning and Gatehouse, 1992). The data collected from the otosclerosis cohort suggests that more women are being referred, but cannot provide evidence to clarify whether or not this is a true reflection of disease incidence.

The data collected also suggests that the most common age of onset of hearing loss in the otosclerosis cohort is between the ages of 21 and 40. This coincides with previous published reports of the average age of onset being in the 3rd decade of life (Gapany-Gapanavicius, 1975). Eleven members of the cohort reported first experiencing hearing loss in childhood below the age of 10 and one member of the cohort was over age 61 at onset. This data therefore supports the literature, suggesting that the condition may develop as early as infancy or as late as the sixth decade in rare cases. The age of onset data also shows an interesting gender distribution with a greater percentage of females exhibiting symptoms at a younger age than males, especially in the 21-30 age range where 35.1% of females develop symptoms in comparison to just 29.5% of males. Since this age range coincides with child-bearing age in females and is also the age range at which females are most likely to be taking oral contraceptives, this difference in age distribution between the genders could provide support to previous studies that have suggested a hormonal role in disease pathogenesis (Vessey and Painter, 2001, Lippy et al., 2005, Podoshin et al., 1978). However, since there was not found to be a statistical difference in age distribution between males and females, this trend may be coincidental.

Ethnicity is predominantly British or Irish in the otosclerosis cohort with three quarters of study participants describing themselves as such. According to the 2011 UK census, 47.1% of London's population describe themselves as British or Irish. Therefore this cohort has significantly more British or Irish individuals than would be expected from a randomly selected cross-section of London's population. This may be because otosclerosis is particularly common in Caucasians, affecting approximately 0.3-0.4% of the population (Declau et al., 2007a), and less common in other ethnic groups including black African and Oriental populations (Declau et

al., 2007a). Alternatively, it may reflect the ethnic composition of the catchment area of the Royal National Throat Nose and Ear Hospital which is located in Islington in North London, a borough which has a less diverse ethnic population than some other London boroughs.

In total, 35% of questionnaire respondents reported a positive family history of otosclerosis. This is lower than the percentage that would have been expected from consulting the published literature which estimates familial cases to comprise 50% of all reported incidences of otosclerosis. However, despite the fact that 50% is the figure cited in review articles (Thys and Van Camp, 2009), the data upon which it is based originates from the 1970s and 80s with no evidence in the current literature to substantiate it. The data collected during this study could therefore provide an indication for current incidence figures for familial otosclerosis. However, it must be considered that there may have been some individuals who reported a family history of hearing loss other than otosclerosis who had a family member with hearing loss due to undiagnosed otosclerosis. It is therefore likely that the figure of 35% is an underestimate of the true incidence of familial cases of otosclerosis.

The data collected from the otosclerosis cohort regarding the proportion of individuals experiencing bilateral hearing loss, tinnitus and vertigo is slightly lower than has been presented in the published literature, with roughly 70% of the cohort experiencing bilateral loss and tinnitus, and with fewer reporting symptoms of vertigo. When considering bilateral hearing loss in the cohort, it is important to be aware that this statistic is not a true indication of the proportion of the cohort who will experience bilateral loss throughout their lifetime. Since the questionnaire is issued to patients shortly after a clinic visit, it is likely that many of these will be patients having their first visit who have only recently developed symptoms, and since otosclerosis usually develops unilateral initially, it is possible that these individuals could go on to develop bilateral loss in the future. It therefore likely that the true incidence of bilateral loss in the cohort would be slightly higher, which may bring the percentage up to a figure more similar to the predicted 70-85% as suggested by the literature (Emmett, 1993).

Since the questionnaire was not completed by a control population, statistical analyses performed on the data collected was limited. This is because is was not possible to compare the questionnaire responses of otosclerosis study participants with a population of unaffected individuals. This must be taken into consideration when evaluating the outcomes of the questionnaire analysis.

3.3.3 Environmental factors involved in otosclerosis development

The questionnaires sent to study participants included questions to collect information on exposure to measles, mumps and rubella infection and vaccination against these viruses. This data was collected because a large number of studies in the past have resulted in the publication of conflicting reports linking viral infection and the development of otosclerosis (Arnold and Friedmann, 1987), McKenna et al., 1989(Niedermeyer et al., 1994, McKenna et al., 1996). By gathering data on incidence of viral infection and vaccination, it was therefore hoped that the cohort could be analysed to identify whether it was overrepresented by those who had been infected or vaccinated against these viruses compared to the general population. Unfortunately, the data collected during this study is incomplete, as a large proportion of respondents were unsure of whether or not they had been infected or had received vaccination against these viruses in the past. This was particularly true in regards to vaccination, possibly because in the UK vaccinations are carried out at a young age so were not within memory of the cohort members. Despite this limitation in data collection, the proportion of cohort members who had been infected by the measles virus appeared to be high at 53.9%, even when those who did not know whether or not they had been infected were excluded from the analysis. However, when considering whether 53.9% is greater than the proportion of individuals that would be expected to have been infected by measles in the general population, the age distribution of the cohort must be considered. This is because older cohort members are more likely to have been exposed to measles infection than younger members, as a result of the reduction in incidence of measles following the introduction of the vaccination programme in the UK in 1970. Therefore given that 78% of the cohort were born prior to 1970, it is not surprising that 53.9% of the cohort had been infected with the measles virus.

Since the risk of developing measles fluctuates each year based on vaccination uptake rates, it is extremely difficult to measure the general population risk of being infected by measles. It is therefore not possible to compare the incidence of infection in the otosclerosis cohort, which includes individuals born over a time period spanning eight decades, with incidence figures for the general population. As a result, it is not possible to establish from this data whether the otosclerosis cohort is overrepresented for those who had been infected or vaccinated in the past. This is a difficulty that has been encountered by many research groups investigating the link between viral infection and otosclerosis and may partly explain why the role of viral infection in otosclerosis aetiology is yet to be elucidated.

In addition to gathering information on viral infection and vaccination, the questionnaires sent to study participants also attempted to collect data on the exposure of participants to fluoride ions in their drinking water. Fluoride is a naturally occurring ion in ground water which can be found at particularly high concentrations on the West coast of the USA, Argentina, some parts of Northern and Eastern Africa, Pakistan and regions of both Northern and Southern India, as well as various parts of China. In many developed countries, it is added to the water supply due to its known dental benefits in children. Fluoride is known to inhibit bone resorption and has thus been suggested to contribute to otosclerosis when present at low levels in the drinking water (Daniel, 1969) although not all studies support this theory (Vartiainen and Vartiainen, 1997a). Fluoride therapy has also been used as a treatment to slow the progression of hearing deterioration in otosclerosis patients with a recent review of the literature concluding that there is some evidence to suggest that sodium fluoride may be of benefit to preserve hearing in otosclerosis patients (Cruise et al., 2010). The questionnaire was designed in an attempt to capture information on where study participants had previously lived, in order to work out whether or not they had been exposed to low levels of fluoride during their lifetime. However, no particular country or region emerged from the data collected as being over-represented as a prior residence of cohort members. Since fluoride concentrations fluctuate within localised regions depending on the natural fluoride concentration in ground water and the varying choices made by local authorities on whether or not to fluoridate water supplies, it is extremely difficult to measure a person's lifetime exposure to fluoride. This limitation may explain why despite a promising link between otosclerosis and fluoride based upon clinical evidence suggesting that fluoride treatment is beneficial in slowing down progression of otosclerosis, no research has been published in the past decade to investigate exposure to fluoride in drinking water and development of otosclerosis.

A set of additional questions were asked specifically to female questionnaire recipients in order to gather data on pregnancy and use of both oral contraceptives and hormone replacement therapy in females with otosclerosis. These questions were included in the questionnaire due to the discrepancy in prevalence of otosclerosis between males and females and also due to the fact that otosclerosis often manifests after or during pregnancy, with many women reporting early symptoms shortly after giving birth and or during breastfeeding, although this evidence is somewhat contradictory (Vessey and Painter, 2001, Lippy et al., 2005). The data collected from the questionnaire appears to be fairly representative of a cohort of adult women, however it is noteworthy that roughly 40% of women in the cohort who had been pregnant reported to experience changes in their hearing ability during or shortly after their pregnancy. This could indicate a potential link between hearing

deterioration in females with otosclerosis and hormonal changes during pregnancy. Despite this, it must be taken into account that the average age of onset of otosclerosis coincides with the typical child-bearing age of women in the UK, which is supported by the age of onset data collected during this study where a higher proportion of females than males reported onset of symptoms between the ages of 21 and 30. Therefore it is difficult to establish whether pregnancy is involved in development of otosclerosis or if any correlation seen is coincidental.

3.3.4 Otosclerosis as an autoimmune condition

The questionnaire completed by study participants also included questions designed to gather information on autoimmune conditions and other connective tissue disorders in otosclerosis patients. Analysis of data collected indicates that the cohort is overrepresented by those experiencing autoimmune disorders in comparison to the general population, with 18.1% of questionnaire respondents reporting an autoimmune condition compared to a predicted 7.6-9.4% in the general population. When a Chi squared test was performed, it was found that there was a significant difference in prevalence figures of autoimmune diseases in the otosclerosis cohort compared to the general population. To perform this statistical test, prevalence data from the largest recent meta-analysis (Cooper and Stroehla, 2003) was used to estimate prevalence in the general population. The largest percentage in the prevalence range of 9.4% was used to ensure that calculations were conservative.

One autoimmune condition of particular interest is Sjogren's disease, which was reported by one of the questionnaire respondents, resulting in a cohort frequency of 0.43% which is substantially higher than the 0.01% reported frequency in the general population (Cooper and Stroehla, 2003). Sjogren's disease is a progressive systemic autoimmune condition that affects the exocrine glands resulting in dryness of the skin, nasal passages and genitals. It frequently occurs in individuals with other systemic connective tissue diseases (Theander and Jacobsson, 2008) including Bullous Pemphigoid (Yamamoti et al., 1998), a condition reported by another cohort member, which is characterised by production of autoantibodies against collagen XVII. Sjogren's disease is more common in women than men, with nine in ten individuals with the condition being female, and most frequently occurs roughly around the time of menopause. It is thought that it may be caused by activation of the transcription factor NFκB, which results in the upregulation of the human leukocyte antigen in the epithelial tissues of patients. Sjogren's disease has also been reported repeatedly in individuals with the common connective tissue condition Plantar Fasciitis, which affects another cohort member, and which occurs when a tight band of connective tissue develops around the arch of the foot.

In addition to Sjogren's disease and its related conditions, the cohort is also over-represented by individuals with other systemic autoimmune conditions including Multiple sclerosis which affects 0.43% of questionnaire respondents compared to 0.06% of the general population, and Osteoporosis which affects 6.03% of questionnaire respondents compared to 3.34% of the general population. In addition, 21 additional cohort members reported an autoimmune condition but did not specify which one. Since multiple autoimmune conditions often occur concomitantly in a single individual, the over-representation of autoimmune conditions in the otosclerosis cohort is suggestive that otosclerosis may itself have an autoimmune pathology.

In summary, the otosclerosis cohort recruited during this study has provided the largest stapes tissue bank worldwide for gene expression studies and is the first instance of a large collection of stapes and genomic DNA samples from overlapping individuals. The cohort recruited tends to display characteristics in line with what would be expected from the published literature. However, the information gathered regarding exposure to environmental agents thought to play a role in otosclerosis was less informative than had been hoped, as it is very difficult to measure a person's exposure to their environment, an issue that has been encountered by other research groups. Despite this, a large proportion of female cohort members reported hearing deterioration during or shortly after pregnancy suggesting that the link between hormones and otosclerosis is worth future investigation. In addition, the cohort appeared to be over-represented by individuals reporting autoimmune conditions, suggesting that perhaps a mechanism of autoimmunity may be involved in development of otosclerosis.

4 Whole Exome Sequencing (WES) identifies novel genetic variants that segregate with otosclerosis in families

4.1 Introduction

Whole Exome Sequencing (WES) is a powerful Next Generation Sequencing (NGS) technique involving targeted sequencing of the exome; the protein coding portion of the human genome. When accompanied by an effective variant prioritisation strategy, WES can be used to identify genetic variants involved in monogenic inheritance of Mendelian traits (Bamshad et al., 2011). The first study to employ this technique was a proof-of concept study which successfully identified the causal mutation within the *MYH3* gene known to underlie the rare autosomal dominant disorder Freeman-Sheldon syndrome (Ng et al., 2009). Since its publication in 2009, more than 100 additional genes involved in various monogenic disorders have been characterised through use of this technique (Rabbani et al., 2012). WES is based on the principle that that the exome constitutes approximately 1% of the genome but accounts for about 85% of disease causing mutations in Mendelian disease (Cooper, 1998). Therefore focusing solely on the exome rather than the entire genome is a quick and cost-effective approach for identification of pathogenic variants.

During WES, a library of short fragments of DNA are prepared from genomic DNA, which are sequenced in millions of parallel reactions before being mapped to a reference genome based on the sequence alignment. This enables variant calling of coding variants. These variants are then subjected to a stepwise prioritisation process (Figure 4.1) which ensures that only those shared by affected family members but absent in unaffected family members are retained. Rare variants and those predicted to have a damaging effect on the resulting protein are prioritised, with the aim of identification of disease causing mutations within candidate genes.

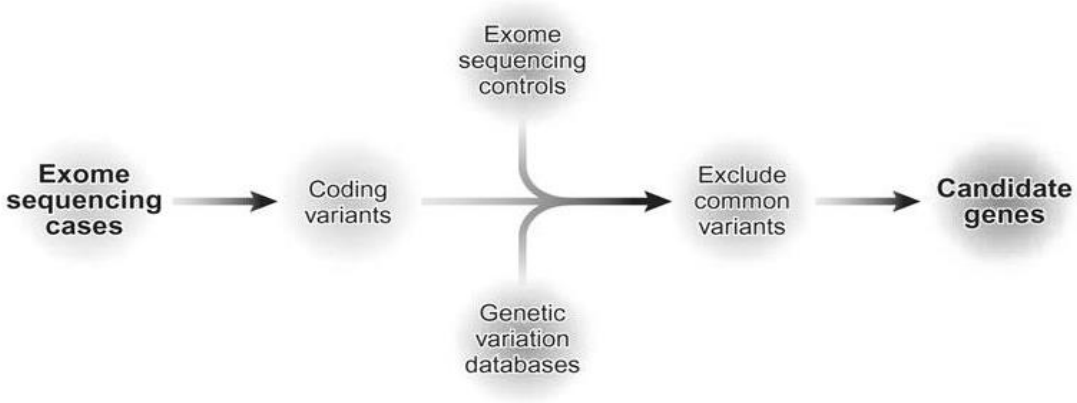


Figure 4.1 Variant prioritisation following WES

Flow chart showing variant prioritisation to identify disease causing variants within candidate genes (from (Biesecker, 2010)).

Prior to the introduction of WES, the traditional approach to gene identification for Mendelian traits was through use of linkage analysis studies. However, the success of these studies was limited due to the issue of variable penetrance and the requirement to recruit large numbers of affected family members (see Chapter 1 section 1.7.2.1). In addition, the NGS methods available, such as Whole Genome Sequencing (WGS,) were too costly for the majority of research groups and were both slow and labour-intensive, generating vast quantities of data, which led to a wealth of bioinformatics obstacles (Wang et al., 2013). Although these techniques have many useful applications, the introduction of WES has offered an attractive alternative option for identification of pathogenic variants, which is increasing in popularity amongst many research groups. With WES, it is possible to identify disease causing variants within small affected families with as few as two affected members, giving this NGS technique wider applications than traditional linkage analysis. It is also a more cost-effective option than WGS, at approximately one sixth of the cost. This reduced cost means that it is possible to perform WES on larger numbers of family members than would be possible for WGS, which increases statistical power of the study, making the identification of a disease causing variant more likely. In addition, WES is less labour-intensive than WGS meaning that results can be made available rapidly with fewer bioinformatics obstacles (Majewski et al., 2011).

However, there are also clear limitations for the use of WES for identification of disease-causing genes. Our current knowledge of the human genome limits the use of this technique, as disease causing variants located within, as yet, unidentified genes, will be overlooked during this approach. Similarly, variants located in non-coding regulatory regions of the genome including transcriptional promoter and enhancer elements, as well as those that regulate

alternative splicing may also be missed. In addition, this technique may not be ideal for investigating structural variation within genomes including translocations and inversions, as well as copy number variants and variation involved in triplet repeat disorders (Wang et al., 2013). WES cannot be used to identify mitochondrial variation as it focuses on the protein-coding region of chromosomes only. Furthermore, for conditions that exhibit variable penetrance, problems may be encountered during segregation analysis that follows WES, due to difficulties in discriminating between phenotypically unaffected family members who may or may not possess the disease causing mutation (Majewski et al., 2011).

To date, no research has been published using WES for gene discovery in otosclerosis. Here it is used to identify novel genetic variants within four families exhibiting monogenic inheritance of familial otosclerosis. The two greatest challenges faced when performing WES were the difficulty in selection of the best possible individuals for sequencing, as well as the analysis, interpretation and filtering of large numbers of variants. In this chapter, strategies for selecting appropriate samples to be sequence are discussed, followed by a description of how variants were identified, annotated and prioritised. The individuals selected for sequencing and variant prioritisation strategy employed was just one of many possible options, and was decided upon as the most likely strategy to yield the desired result.

The prioritisation steps are based on a series of assumptions that can be made about the disease causing variant, which may appear overly stringent in comparison to other research techniques, as it is possible that the true disease causing mutation could be eliminated at any stage of the filtering process. However, it is crucial to use stringent filters initially in order to narrow down the large number of variants to a small pool of those most likely to be disease causing. If no variants within good candidate genes were identified at the end of the process, reduced-stringency filters would be applied, in what is an iterative process.

This chapter is divided into two parts. Part I will describe the identification of all variants within individuals who have been sequenced, followed by prioritisation of candidate variants that segregate with familial otosclerosis, whilst Part II will involve further analysis performed on the most interesting candidates to emerge from Part I. Due to the step-wise nature of this analysis, the results and discussion sections will be combined in each part of this chapter, as it is essential to interpret the results at each stage in order to make decisions on how best to proceed. A general discussion section will summarise the findings at the end of the chapter.

Whole Exome Sequencing Part I

Identification of rare variants that segregate with familial otosclerosis

4.2 Identification of ten individuals for WES

WES was performed at the end of year one of this research project at the Wellcome Trust Sanger Institute in Cambridge. In order to obtain genomic DNA samples from appropriate otosclerosis families for sequencing, probands that reported a strong family history of otosclerosis in their questionnaire responses were prioritised for family recruitment. This ensured that genomic DNA was collected from sufficient numbers of families for WES.

At the end of year one, sixteen otosclerosis families had been recruited. This included the recruitment of thirty five individuals with confirmed otosclerosis based on confirmation of otosclerosis during surgery or audiometric analysis along with physical examination. It also included one individual with presumptive otosclerosis who was not able to confirm a diagnosis but who was presumed to be affected based on reported symptoms characteristic of otosclerosis that strongly indicated the presence of the condition in this family member. In addition, four of the individuals reported symptoms of hearing loss that were caused by other hearing conditions including presbycusis, or had undiagnosed hearing loss. For this reason it was not possible to exclude the possibility of otosclerosis in these individuals. Eighteen unaffected family members were also recruited who had not reported any prior symptoms of hearing loss.

In total ten individuals across four of these sixteen families designated A (Figure 4.2), B Figure 4.3), C (Figure 4.4) and D (Figure 4.5)were selected for WES, to give the best possible chance of identifying the disease causing variant.

These four families were prioritised for WES based on the following criteria:

- The family appeared to exhibit clear monogenic inheritance of otosclerosis
- Otosclerosis was present in multiple generations of the family
- The family included multiple living affected relatives
- It was possible to secure genomic DNA samples from multiple affected and unaffected family members
- Families from which stapes cDNA had been obtained were prioritised where possible

Families A, B, C and D are described in sections 4.2.1, 4.2.2, 4.2.3 and 4.2.4 respectively.

4.2.1 Family A

Family A is of Caucasian origin living in the UK. The mode of otosclerosis inheritance in this family is consistent with an autosomal dominant pattern (Figure 4.2). Autosomal dominant inheritance is assumed due to the presence of either confirmed or presumptive otosclerosis in three of five siblings in generation III, in addition to one confirmed case in generation IV. In this family, it appears that otosclerosis is inherited from two sisters in generation II, for which the hearing status is unknown. Two living affected individuals within this family have confirmed otosclerosis, both of whom were selected for WES (Table 4.1). These individuals are the proband (IV:3) and her maternal uncle (III:2) who would be expected to share 25% of their variants. The probands mother (III:1) is the case of presumptive otosclerosis previously mentioned, where otosclerosis is presumed but cannot be confirmed. This individual reported to have undergone bilateral surgery for hearing loss, which is strongly suggestive of otosclerosis as there are few other scenarios under which bilateral hearing loss surgery would be performed. This combined with analysis of the family pedigree which suggests that this individual is likely to have passed the disease causing variant on to her affected daughter (IV:3), indicates that otosclerosis is likely in this individual. The proband's daughter (V:1) reported recent symptoms of hearing loss but audiometric analysis was unable to provide a diagnosis. It is therefore not possible to rule out hearing loss in this individual. The age of hearing loss onset ranges between the age ranges of age 11-20 and 21-30 in this family.

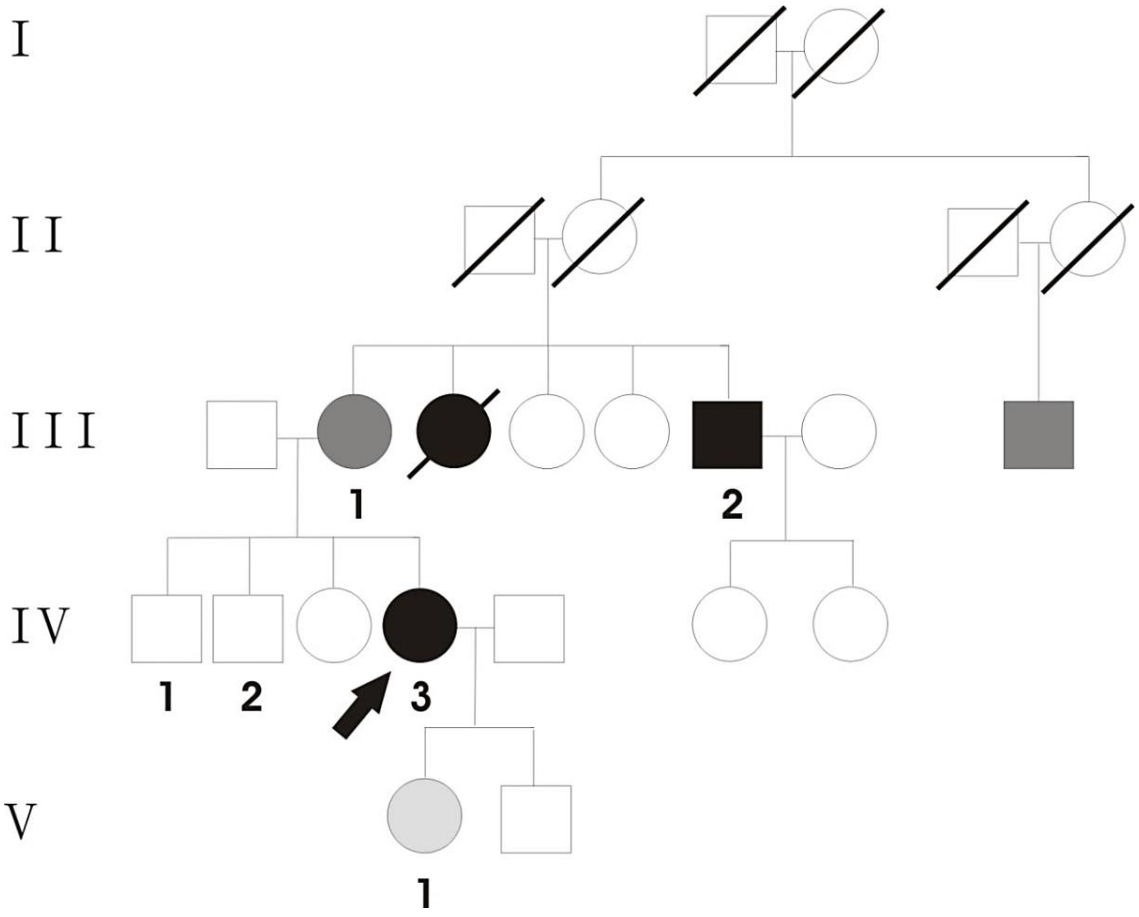


Figure 4.2 Pedigree of Family A

Square=male; circle=female; slash=deceased individual; open symbol=clinically unaffected; solid symbol=hearing impairment; black symbol=confirmed otosclerosis; dark grey symbol=presumptive otosclerosis; light grey symbol=other hearing loss; arrow=proband

Family member	Age at recruitment	Gender	Hearing status	Age range at hearing loss onset	Selected for WES
III:1	78	Female	Presumptive otosclerosis	unknown	No
III:2	64	Male	Confirmed Otosclerosis	21-30	Yes
IV:1	54	Male	Unaffected	n/a	No
IV:2	53	Male	Unaffected	n/a	No
IV:3 (Proband)	48	Female	Confirmed Otosclerosis	11-20	Yes
V:1	27	Female	Undiagnosed hearing loss	21-30	No

Table 4.1 Table showing demographics of members of family A

4.2.2 Family B

Family B is of Caribbean origin from a small island in St Vincent and the Grenadines (details with-held to protect identity of family). The proband describes herself as of black Caribbean origin with her maternal family being of French origin, whilst her paternal grandmother is of black Carib origin and paternal grandfather of Scottish origin.

The family appears to exhibit autosomal dominant inheritance of otosclerosis with variable penetrance (Figure 4.3). This assumption has been made due to presence of otosclerosis symptoms in close to 50% of the offspring of a single generation despite neither of their parents exhibiting symptoms. However, since the family comes from a small island population of approximately 4,500 people and since neither of the proband's parents experienced otosclerosis symptoms, autosomal recessive inheritance cannot be ruled out. Although DNA from only a single generation of affected individuals is available in this family, they were selected as a good family for WES because the inheritance pattern is consistent with monogenic inheritance. In addition they were prioritised because the stapes of the proband was one of those RNA sequenced (chapter 5), the results of which it was assumed could assist with the identification of the pathogenic variant in this family.

Three living affected individuals within this family have confirmed otosclerosis, two of whom were selected for WES along with one unaffected individual within the family (Table 4.2). Since the only suitable affected individuals for WES in this family are siblings who share 50% of their variants, it was decided to also select unaffected sibling III:5, in order to assist the variant prioritisation process following WES. Unaffected individual III:5 was selected in preference to individual IV:1 as this individual is younger than the age at which some of his relatives developed symptoms, so it is possible this individual could go on to develop symptoms in the future (Table 4.2). In addition, III:5 was selected in preference to II:1 as it was expected that it would be more beneficial to sequence an unaffected family member from the side of the family in which otosclerosis occurs. There was no great benefit in selecting III:5 in preference to III:6 but was chosen for the sole reason that a greater yield of genomic DNA was obtained from individual III:5. The selection of an unaffected individual raises some concerns regarding variable penetrance, especially because the father in this family appears not to have otosclerosis despite the condition seemingly being passed down the paternal line. His age at death is unknown. Symptom onset is variable in this family and ranges from age ranges 11-20 to 41-50.

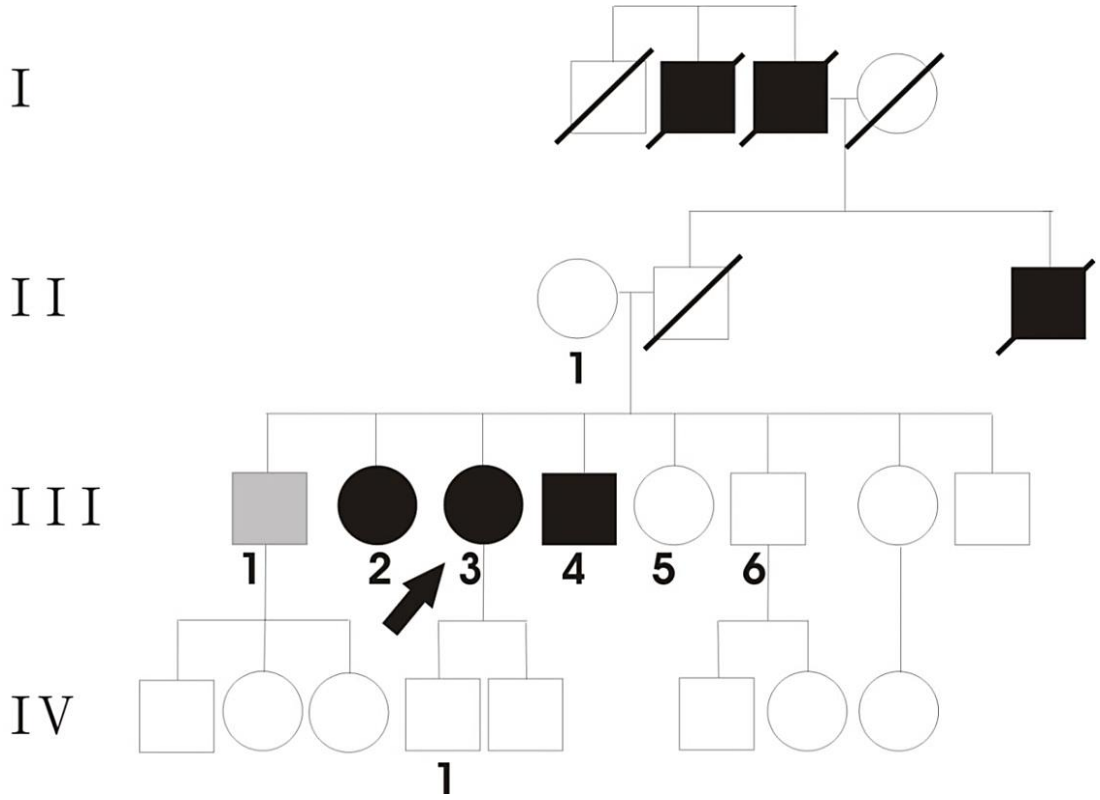


Figure 4.3 Pedigree of Family B

Square=male; circle=female; slash=deceased individual; open symbol=clinically unaffected; solid symbol=hearing impairment; black symbol= confirmed otosclerosis; light grey symbol=other hearing loss; arrow=proband

Family member	Age at recruitment	Gender	Hearing status	Age range at hearing loss onset	Selected for WES
II:1	87	Female	Unaffected	n/a	No
III:1	67	Male	Presbycusis	61+	No
III:2	65	Female	Confirmed Otosclerosis	41-50	Yes
III:3 (Proband)	63	Female	Confirmed Otosclerosis	21-30	Yes
III:4	62	Male	Confirmed Otosclerosis	11-20	No
III:5	60	Female	Unaffected	n/a	Yes
III:6	59	Male	Unaffected	n/a	No
IV:1	35	Male	Unaffected	n/a	No

Table 4.2 Table showing demographics of members of family B

4.2.3 Family C

Family C is of Caucasian origin and exhibits an otosclerosis inheritance pattern consistent with that of autosomal dominant inheritance. This can be assumed due to the presence of multiple affected family members in successive generations (Figure 4.4). The hearing status and age at death of generation I is unknown but it can be assumed that at least one of the individuals in this generation possessed the disease causing variant, as it would be very unlikely that two of their offspring in generation II would develop a spontaneous mutation leading to the disease phenotype. Three living affected individuals within this family have confirmed otosclerosis, two of which (III:1 and II:3) were selected for WES (Table 4.3). These two individuals were selected in preference to individual II:4, as they were the two most distantly related affected individuals in the family, who would be expected to share just 25% of their variants. This meant that more variants could be eliminated at the first stage of the filtering process, than if only one of these individuals was selected along with II:4. Family member II:1 has presbycusis and member III:2 has recently reported symptoms of hearing loss at age 45, which has not been diagnosed. Since symptom onset ranges from age ranges 21-30 to 31-40 in this family, it is possible that individual III:2 could be experiencing the first symptoms of otosclerosis, as although he would be the oldest reported case of symptom onset in this family, he is only 5 years older than the oldest reported case to date.

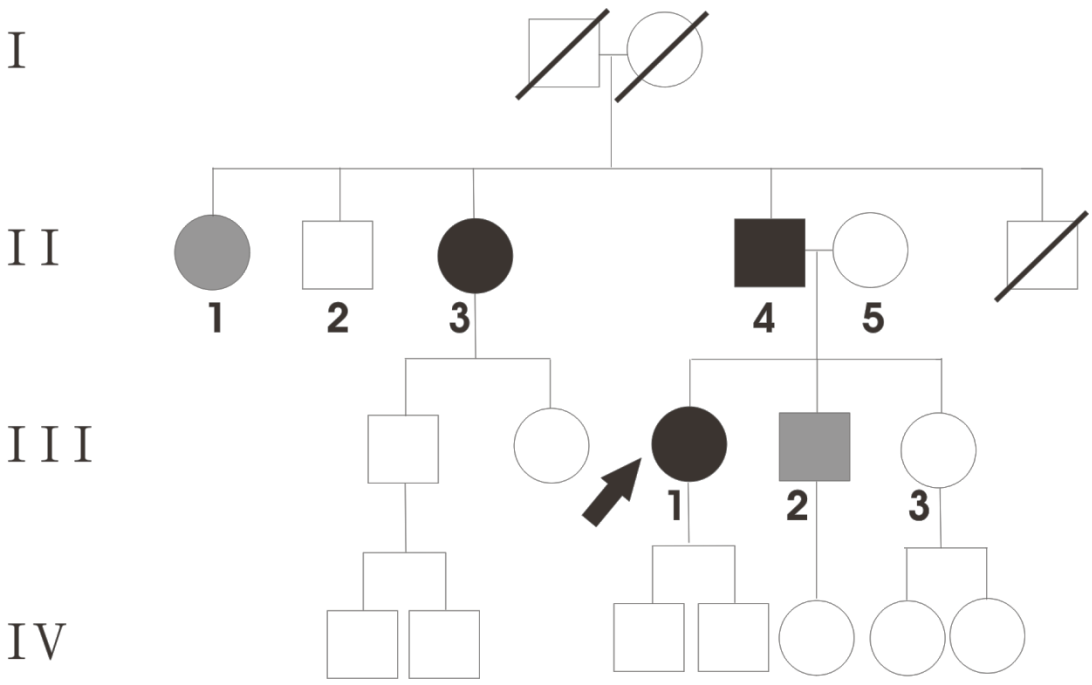


Figure 4.4 Pedigree of Family C

Square=male; circle=female; slash=deceased individual; open symbol=clinically unaffected; solid symbol=hearing impairment; black symbol=confirmed otosclerosis; light grey symbol=other hearing loss; arrow=proband

Family member	Age at recruitment	Gender	Hearing status	Age range at hearing loss onset	Selected for WES
II:1	83	Female	Presbycusis	61+	No
II:2	81	Male	Unaffected	n/a	No
II:3	78	Female	Confirmed Otosclerosis	21-30	Yes
II:4	74	Male	Confirmed Otosclerosis	31-40	No
II:5	7th decade	Female	Unaffected	n/a	No
III:1 (Proband)	46	Female	Confirmed Otosclerosis	21-30	Yes
III:2	45	Male	Undiagnosed hearing loss	41-50	No
III:3	41	Female	Unaffected	n/a	No

Table 4.3 Table showing demographics of members of family C

4.2.4 Family D

Family D is of Portuguese origin from a small rural region in Portugal (details with-held to protect identity of family). It appears that all family members are the descendants of two brothers for whom the disease status is unknown. Although this family appears to show autosomal dominant inheritance, the multiple consanguineous marriages in the family may be affecting the apparent inheritance pattern so autosomal recessive inheritance cannot be ruled out (Figure 4.5). The large numbers of individuals with confirmed or presumptive otosclerosis in this family suggests high penetrance of the disease causing allele, providing a strong rationale for monogenic inheritance in this family. Three living affected individuals (IV:2, V:1 and V:3) have confirmed otosclerosis, all of whom were successfully recruited and selected for WES (Table 4.4). Individual IV:1 who has presbycusis and unaffected individual V:2 were also recruited. Intermarriage within the family means that the siblings and their parents may share more than 50% of their variants. Therefore three affected family members were chosen to assist with the filtering process to ensure the best possible chance of identifying the pathogenic variant. The proband in this family developed symptoms under the age of 10, however age of onset data is not available for the remainder of the family.

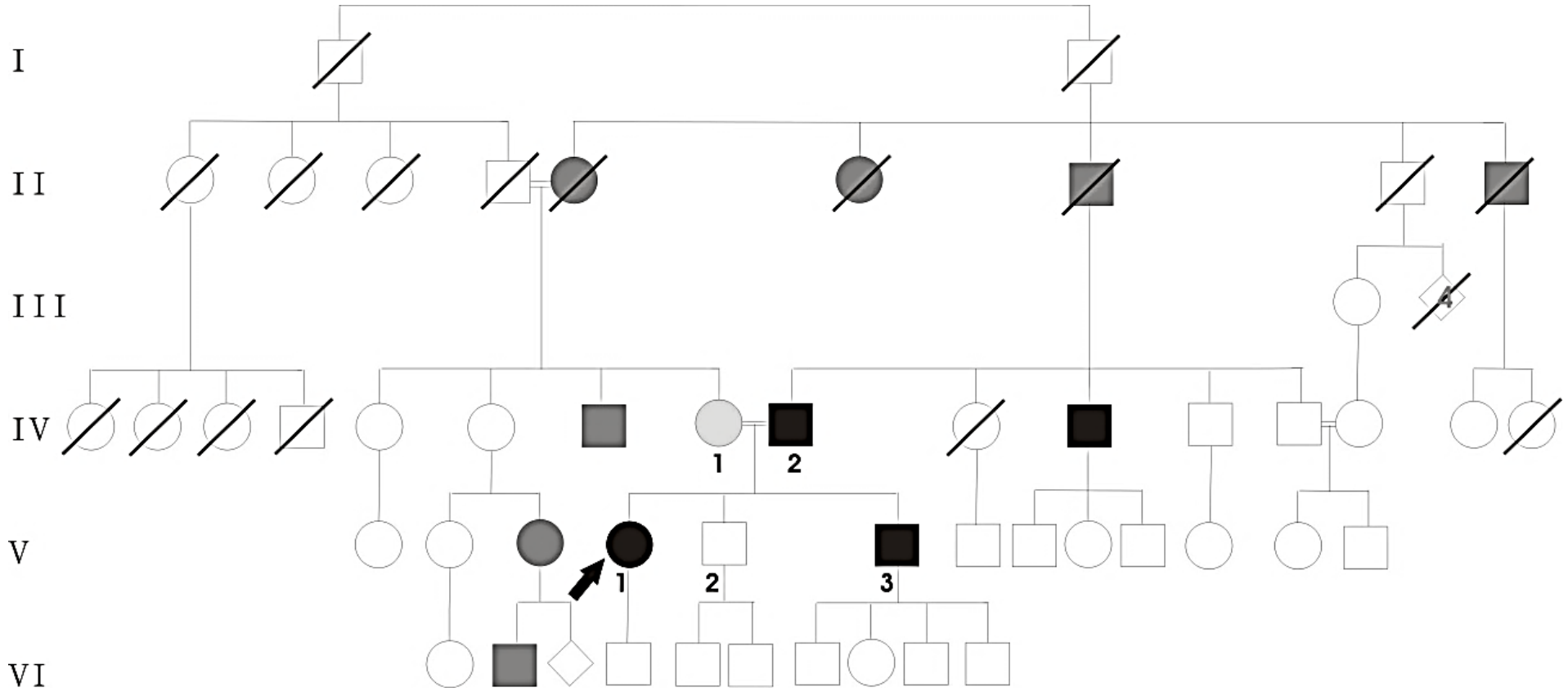


Figure 4.5 Pedigree of Family D

Square=male; circle=female; slash=deceased individual; double line=consanguineous marriage; open symbol=clinically unaffected; solid symbol=hearing impairment; black symbol=confirmed otosclerosis; dark grey symbol=presumptive otosclerosis; light grey symbol =other hearing loss; arrow=proband.

Family member	Age at recruitment	Gender	Hearing status	Selected for WES
IV:1	75	Female	Presbycusis	No
IV:2	81	Male	Otosclerosis	Yes
V:1 (Proband)	43	Female	Otosclerosis	Yes
V:2	4th decade	Male	Unaffected	Yes
V:3	46	Male	Otosclerosis	No

Table 4.4 Table showing demographics of members of family D

In this family there are two incidences, on either side of the family, in which multiple siblings have died in infancy or had severe disability. Although this could be due to an unrelated autosomal recessive genetic condition, which would not be surprising given the number of consanguineous marriages in the family, it is also possible that an allele leading to otosclerosis could be being selected for over a lethal allele. In addition, this family also has a range of other symptoms which are suggestive of a potential systemic connective tissue disorder. These symptoms include double jointedness, joint pain and muscular aches which are often triggered by illness or infection, hyperflexibility, temporomandibular joint dysfunction, weakness in the hip joints, carpal tunnel syndrome, scoliosis with old age, itchy skin after sun exposure, bad scarring, frequent miscarriages, arthritis and gastrointestinal problems. Since otosclerosis is known to co-occur with a number of systemic connective tissue disorders, it is therefore possible that this family has a connective tissue disorder that manifests coincidently with otosclerosis, or of which otosclerosis is a symptom. The literature suggests two possible genetic disorders which could explain the symptoms of this family; Ehlers-Danlos syndrome (Byers and Murray, 2012) and Buschke-Ollendorff syndrome (Schena et al., 2008). Genes that have been implicated in both conditions could assist with the variant prioritisation process following WES to help identify the pathogenic variant in family D.

4.3 WES in ten familial otosclerotic cases

WES was performed on the genomic DNA isolated from each of the ten individuals selected. After target enrichment, the whole exome DNA library from each participant was sequenced using Illumina HiSeq 2000 with 100bp paired-end reads. This technology involves random fragmentation of the genomic DNA which is then attached to an optically transparent surface prior to fragment extension and amplification to produce a flow cell which contains thousands of copies of the DNA template. These templates were then sequenced in millions of parallel reactions using sequencing-by-synthesis (SBS) technology which ensures highly accurate base calling, before being mapped to the reference human genome (GRCh37_hs37d5) based on the sequence alignment. Both single-nucleotide polymorphism (SNPs) and insertions-deletions (indels) were identified using SAMtools (http://samtools.sourceforge.net/) and GATK (http://www.broadinstitute.org/gatk/), although the data reported here refers only to SAMtools. A mean of 16.5 Gb of sequence was generated per sample with an average of 96.16% of the bases mapping to the reference genome at a coverage of at least 10x. The mean depth of coverage was 178.5x. On average 89,656 exonic variants were identified in each of the ten individuals exome sequenced. The number of variants identified was slightly higher in the family of Caribbean origin (family B) at 93,972 variants than in the three Caucasian families (families A, C and D) in which an average of 88,217 variants were mapped per individual. This may be because the reference genome against which variants are mapped is not derived from a single individual but from a collection of DNA from anonymous individuals of primarily European origin assembled in a mosaic haploid genome (Chen and Butte, 2011), so would be expected to have fewer variants in common with non-Caucasian genomes. Of all the variants mapped, 93.7% met a quality score threshold of 30. Since quality scores measure the probability that a base is called incorrectly, and the quality score of any given base (Q) can be defined by the equation $Q=-10\log_{10}(e)$ where e is the predicted probability of an incorrect base call, a quality score of 30 represents an error rate of 1 in 1000 which represents a base call accuracy of 99.9%.

The sequencing facility at the Wellcome Trust Sanger Institute supplied data files consisting of a list of all variants mapped per individual including SNPs and indels. All SNPs identified were annotated with numerous data fields which are summarised in Table 4.5. All indels identified were annotated with the same data fields with the exception of all 1000 Genome Project frequency data, 500 Exomes Project frequency data, and the predicted effect of variants based on PolyPhen and SIFT softwares.

Field name	Explanation of field name
CHROM	Chromosome number
POS	1-based position of variant. For indels, this is position preceding the indel
ID	Variant identifier, usually the dbSNP rsID where applicable
REF	Reference sequence at POS involved in variant. For a SNP, it is a single base
ALT	Comma delimited list of alternative sequence(s)
QUAL	Phred-scaled probability of all samples being homozygous reference
FILTER	Semicolon delimited list of filters that the variant fails to pass
INFO-1000G	Allele presence in the 1000 Genomes Project
INFO-1000G_AF	Allele frequency in the 1000 Genomes Project
INFO-500e1012	Allele presence in the 500 Exomes Project
INFO-AC1	EM estimate of the site allele count of the strongest non-reference allele
INFO-AF1	EM estimate of the site allele frequency of the strongest non-reference allele
INFO-AF_AFR	Allele frequency in African population of 1000 genomes project
INFO-AF_AMR	Allele frequency in Ad Mixed American population of 1000 genomes project
INFO-AF_ASN	Allele frequency in East Asian population of 1000 genomes project
INFO-AF_EUR	Allele frequency in European population of 1000 genomes project
INFO-AF_MAX	Maximum allele frequency in the 1000 genomes project populations
INFO-Condel	Consensus deleteriousness score of missense mutations predicts the allele to be deleterious or neutral
INFO-DB	Allele presence in dbSNP
INFO-DP	The number of reads covering or bridging POS
INFO-DP4	Number of 1) forward ref alleles; 2) reverse ref; 3) forward non-ref; 4) reverse non-ref alleles, used in variant calling
INFO-FQ	Consensus quality. If positive, FQ equals the phred-scaled probability of there being two or more different alleles. If negative, FQ equals the minus phred-scaled probability of all chromosomes being identical.
INFO-INDEL	Indicating the variant is an INDEL
INFO-MQ	Mapping quality score based on Phred-scaled liklihood
INFO-PV4	P-values for 1) strand bias; 2) baseQ bias; 3) mapQ bias; 4) tail distance bias
INFO-PolyPhen	PolyPhen software prediction that the allele is damaging or benign
INFO-SIFT	SIFT software prediction that the allele is deleterious or tolerated
INFO-VAA	The resulting amino acid change
INFO-VCQ	The variant is synonymous, nonsynonymous, intronic or located within a splice site

Table 4.5 Table of annotated data fields for WES variants

Table shows annotations supplied in documentation from the Wellcome Trust Sanger Institute for each of the variants identified during WES

4.4 Systematic step-wise prioritisation of variants

A series of sequential prioritisation steps were devised to identify a shortlist of genes containing those variants, which were most likely to be involved in the disease process in each family. These steps are illustrated in Figure 4.6 and are each described in detail in the text below.

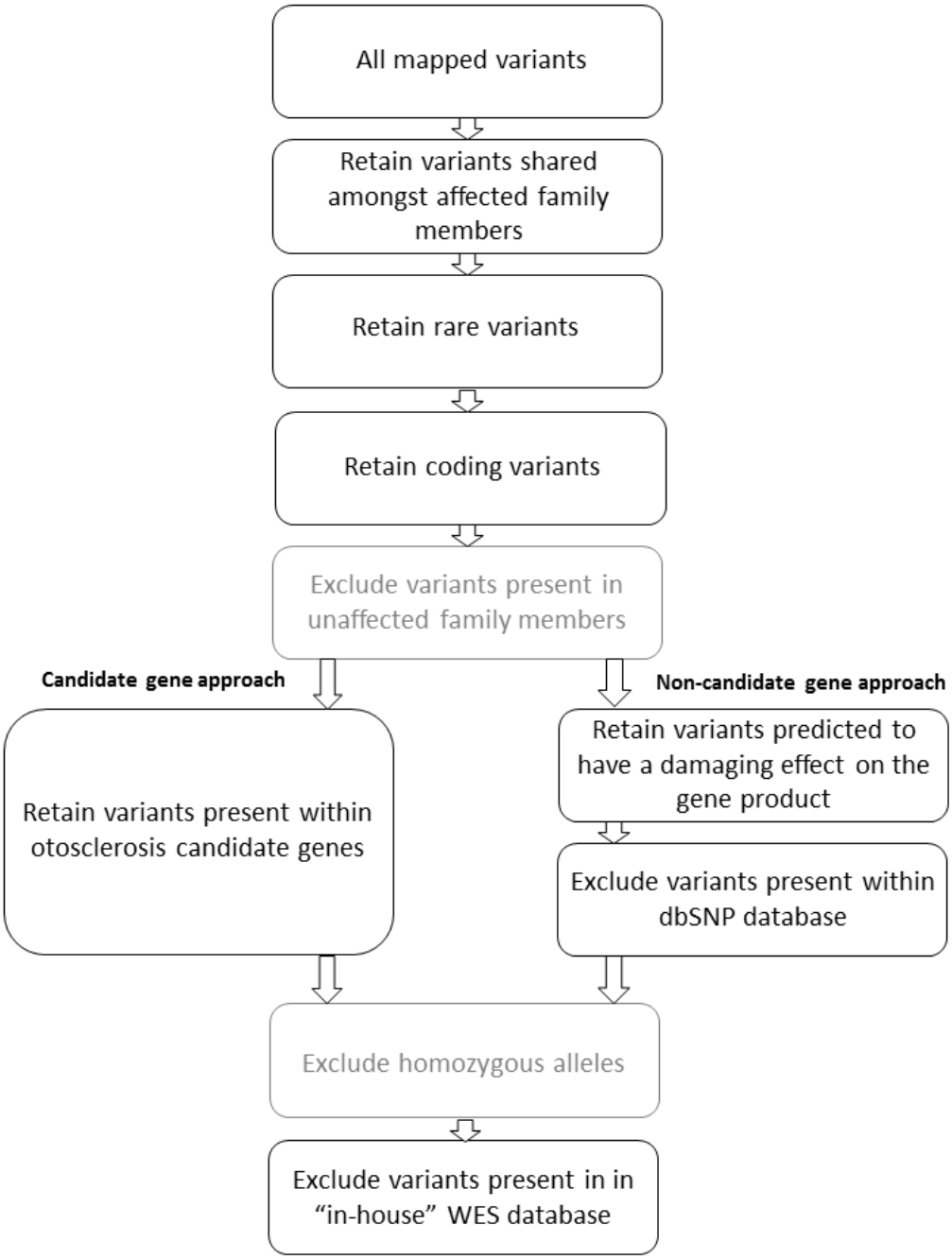


Figure 4.6 Schematic showing the step-wise variant prioritisation process employed following WES.

Steps highlighted in grey were applicable only to specific families, which is discussed in detail in the accompanying text below.

4.4.1 Retain variants shared amongst affected family members

The first stage of the filtering process was based on the assumption that the disease causing variant would be shared by all affected members of each family. Therefore only SNPs and indels which were present in all affected sequenced persons in each family were retained, leaving on average 59,511 retained variants per family (Table 4.6).

4.4.2 Retain rare variants

The second stage of filters involved the assumption that a disease causing variant would likely be present at a low frequency in the general population. Since the incidence of otosclerosis is approximately 0.003-0.004 in the Caucasian population (Declau *et al.*, 2007) and approximately half of all cases occur within families, together with the fact that most cases of familial otosclerosis are inherited in an autosomal dominant pattern suggesting that variants are likely to be heterozygous in affected family members, it can be predicted that the allele frequency of a familial otosclerosis variant in the general population would be approximately 0.0015-0.002. Since linkage analysis has identified multiple otosclerotic loci, suggesting that familial otosclerosis is highly heterogeneous, it is likely that the frequency of a specific familial variant would be even less common in the general population in reality. Based upon this assumption, the frequency of the shared SNPs in each family were compared with frequency data from the 1000 Genomes Project (http://www.1000genome.org September 2011 release). The 1000 Genomes Project is a database of variants collected from sequencing the genomes of large numbers of healthy individuals of different ethnic origins. For the three Caucasian families (families A, C and D) all SNPs present in the 379 European genomes in the 1000 Genomes Project with an allele frequency greater than 0.02 were eliminated. A cut off frequency of 0.02 was selected as this is at least 10x greater than the frequency that would be expected of a familial otosclerosis causing variant, and it is essential to be conservative when eliminating variants in order to minimise the chances of eliminating the disease causing variant. In the Caribbean family (family B) all SNPs present in the 246 African genomes in the 1000 Genomes Project, the closest ethnically matched cohort, at the same allele frequency of 0.02 were eliminated.

The datasets from the three Caucasian families were then subject to a second round of filtering by eliminating all SNPs present in the 500 Exomes Project database (http://www.sanger.ac.uk/research/projects/metabolicdisease/). This is a European cohort so was not suitable for filtering of the family B dataset.

Since frequency data on insertions and deletions was not yet available in either of these two databases, it was not possible to filter frequencies for indels. Following these filters, an average of 5,044 variants remained per family (Table 4.6).

4.4.3 Retain coding variants

The next stage of filtering was based on the assumption that a disease causing variant would likely have a direct impact on the protein it encodes. It was therefore assumed that any synonymous variants that do not cause an amino acid change could be excluded from the analysis. Similarly, variants located within non-coding genes or transcripts are unlikely to be pathogenic so were also excluded. However, it is possible that pathogenic variants may reside within non-coding transcripts, particularly in the case of microRNAs which are non-coding RNAs that play a role in transcriptional and post-transcriptional regulation. For this reason, an exception was made for variants located within microRNAs. Following the elimination of all synonymous and non-coding variants, with the exception of microRNAs, each dataset was left with an average of 429 variants (Table 4.6).

4.4.4 Exclude variants present in unaffected family members (Family B only)

In family B, since WES was also performed in an unaffected family member (Table 4.6), an additional level of filtering was then possible as it can be assumed that the unaffected family member will not possess the disease causing variant. As a result, remaining variants in this family that were present in individual III:5 were eliminated from the dataset. This resulted in 314 retained variants in family B. Caution was taken when choosing to filter against the variants present in individual III:5 because this family exhibits variable penetrance. However, since individual III:5 is 60 years old, a decade older than the oldest reported age of onset in this family, it was decided that it was unlikely that she would develop otosclerosis in the future. Therefore filtering out variants present in individual III:5 was performed, with a view to return to this stage in the filtering process if no suitable candidates survived the remaining filters (Table 4.6).

Filters	Number of retained variants following application of filter								Average number retained variants per family
	Family A (2 affecteds)		Family B (2 affecteds)		Family C (2 affecteds)		Family D (3 affecteds)		
Average number of variants mapped per affected family member	88,295		93,972		88,373		87,983		89,656
Quality control filters (Quality score>30)	82,689		87,135		82,983		83,058		83,966
Number variants shared by affected individuals	58,538		65,275		58,775		55,454		59,511
Exclude variants >0.02 freq in 1000 Genomes	5,924		8,339		5,844		5,865		6,493
Exclude variants in 500 Exome Project [NB/Filter not applied to family B]	4,001		8,339		3,910		3,925		5,044
Exclude synonymous variants	729		1,540		695		757		930
Exclude variants in noncoding genes	278		873		279		285		429
Exclude variants present in unaffected individuals [NB/Filter not applied to families A, C and D]	278		314		279		285		289
Candidate Filters	**Candidate**	**Non-candidate**	**Candidate**	**Non-candidate**	**Candidate**	**Non-candidate**	**Candidate**	**Non-candidate**	**Non-candidate filters**
Retain only variants in candidate gene list	26	112	26	158	27	94	29	112	Retain only deleterious changes (Polyphen & SIFT prediction)
		33		50		31		37	Exclude variants present in dbSNP
Exclude homozygous alleles [Filter not applied to families B and D]	18	28	26	50	17	29	29	37	
Exclude variants present in an "in house" WES database	8	22	12	44	3	21	8	27	
Total	**30**		**56**		**23 (1 overlap)**		**35**		

Table 4.6 Table illustrating the number of variants retained per family at each stage of the filtering process

The table highlights which filters were applied at each stage and the number of variants retained in families A, B (non-Caucasian), C and D, as well as the average number of variants prioritised for the next stage of filtering. Both Candidate and non-candidate gene filtering strategies are illustrated.

4.4.5 Retain variants present within otosclerosis candidate genes

At this stage, two different filtering techniques were employed in parallel on the remaining variants; a candidate gene filtering process and a non-candidate gene filtering process. The purpose of the candidate gene filters was to ensure that any variants located within a list of 494 candidate genes that had previously been implicated in otosclerosis through other studies were prioritised. These included all genes identified or investigated in the published literature through association studies, gene expression studies and within the eight loci identified through linkage analysis, as well as a list of genes known to be involved in other connective tissue disorders. It also included the 176 genes found to be dysregulated in the RNA sequencing study (chapter 5). A complete list of these 176 dysregulated genes can be found in the Appendix. On average 27 retained variants per family were located within genes in the candidate gene list (Table 4.6).

4.4.6 Retain variants predicted to have a damaging effect on the gene product

The non-candidate gene filtering process was carried out in parallel to ensure that other variants of interest, which had not been previously implicated in otosclerosis through other studies, were not overlooked. This was crucial to avoid any bias that would be introduced by using solely a candidate gene approach. The non-candidate filtering process was carried out in two stages. First it involved the use of two prediction tools Polyphen II and SIFT, to predict how likely the variant was to have a damaging effect on the resulting protein. This software uses an algorithm to predict whether the variant in question is likely to be damaging or benign. Therefore only SNPs predicted to be damaging were retained.

4.4.7 Exclude variants present within the dbSNP database

The second stage of the non-candidate gene filtering process involved retaining only those variants that were absent in the database dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). dbSNP is a public database containing information on all known human SNPs. Unlike the 1000 Genomes Project and 500 Exomes Project, dbSNP contains not only data from healthy individuals, but also from people with disease. There is therefore a small possibility that this database could contain variants from individuals with otosclerosis, so caution should be taken when applying this filter. Despite this risk, dbSNP was used as a final robust filter, with a view

to return to this stage in the filtering process if no suitable candidates survived the remaining filters. On average, 38 variants survived the non-candidate gene filtering process (Table 4.6).

4.4.8 Exclude homozygous alleles (Families A and C only)

The retained variants from both filtering strategies were then subjected to two further filters. For families A and C where an autosomal dominant inheritance pattern could be assumed with confidence, all homozygous alleles were excluded. It was possible to make this assumption as in an autosomal dominant disorder, only one copy of the disease causing variant is required to result in the disease phenotype. Due to the heterogeneous nature of familial otosclerosis, a specific familial otosclerosis variant would be expected to be rare and therefore unlikely to be homozygous in the affected family members. It can therefore be expected that all affected individuals in both families would be heterozygous for the disease causing variant. This was not a suitable filter for families B and D due to the small possibility of the condition exhibiting an autosomal recessive inheritance pattern in these families. It is therefore possible that the disease causing variant could be homozygous in affected members of these families.

4.4.9 Exclude variants present in an "in-house" WES database

Finally all variants were subject to filtering against an "in-house" WES database. This database consisted of 20 individuals with presbycusis who had undergone the same WES process at the Wellcome Trust Sanger Institute in Cambridge and in whom otosclerosis could be ruled out from analysis of audiometric data. Eliminating any variants present in this cohort removed common sequencing artefacts from the data.

Following the filtering process, the number of potential variants had been narrowed down to just 144 variants; 30 in family A , 56 in family B, 23 in family C and 35 in family D (Table 4.6). Just one of these variants, a c.915C>G SNP within a gene called *TRIM17* in family C, survived both the candidate and non-candidate gene filtering strategies.

4.5 Variant filtering for segregation analysis

An additional variant filtering strategy was then performed to identify the best possible candidates for segregation analysis in each of the four families. In order to ensure that the variants most likely to be involved in the disease process were selected, all variants were annotated by gathering further information from five online databases. These are illustrated in Table 4.7. These annotations revealed that the 144 prioritised variants were enriched for genes known to be involved in bone maintenance and development, with 8% of variants located in genes involved in these functions. This was also true of genes known to be involved in deafness with 5% of all variants being located in genes implicated in a form of hearing loss. In addition, the dataset was also found to be highly enriched for variants predicted to be located within peroxisome proliferator-activated receptor (PPAR) binding sites within families B and D. PPARs are a subfamily of nuclear receptors with a variety of functions including control of bone turnover and regulation in bone cell differentiation in mesenchymal and hematopoietic cell lineages. Of the 56 retained variants in family B, two were predicted to be located within PPARγ and one within PPARβ binding sites, whilst in family D, two of the 35 variants were located within PPARγ binding sites.

Website name	Website address	Annotation derived
Gene Cards	www.genecards.org	Function of gene
JAX lab (Jackson Lab - Mouse Genome Informatics)	www.informatics.jax.org	Mouse models of mutations within gene
OMIM (Online Mendelian Inheritance in Man)	www.ncbi.nlm.nih.gov/omim	Role of gene in human disease
SNP NEXUS	www.snp-nexus.org	Location of variant within predicted transcription factor binding sites
NHLBI (U.S. National Heart Lung and Blood Institute database)	www.nhlbi.nih.gov	Frequency of variant in NHLBI cohort

Table 4.7 Table of websites used during annotation of variants and genes

In order to prioritise variants for segregation analysis, each variant was ranked according to a list of set criteria that were based upon these annotations. Variants were given high rankings if located within a gene with a known biological role involved in bone development or maintenance, or known to be involved in conductive hearing loss, based on the annotations derived from the Gene Cards and JAX lab websites. This ensured that variants located within genes most likely to have an impact on bone growth or conductive hearing loss were included in further analysis as it would be expected that these genes would be the most likely to play a role in otosclerosis pathogenesis. Genes known to be involved in systemic connective tissue disorders when mutated according to the OMIM website were also ranked highly, as it was expected that a gene involved in another connective tissue disorder would be more likely than other genes to be involved in otosclerosis. Genes involved in connective tissue disorders were given especially high rankings in family D, as this family appears to have a range of symptoms suggestive of a wider connective tissue disorder. Other prioritisation criteria included variants that were located within a PPAR binding site as predicted by SNP NEXUS, or variants that were absent or rare in the NHLBI cohort which contains SNPs identified through a number of studies in individuals with heart, lung or blood conditions.

In addition to these annotations, variants were also given a high ranking if they were known to be located within a highly conserved region of a gene or if they had been found to be significantly dysregulated in otosclerotic stapes in the RNA sequencing study (chapter 5). Conservation was measured using the online software GERP (http://mendel.stanford.edu/SidowLab/downloads/gerp/), which estimates conservation by identifying constrained elements in multiple species alignments. A GERP value greater than 2.5 indicates conservation of the variant. Highly conserved variants were prioritised as it is more likely that a variant conserved between species is crucial for proper gene function, since a variant compromising the function of a gene would act as a selection pressure during evolution, resulting in its elimination. The location of variants within genes dysregulated in otosclerotic stapes based on the RNA sequencing study data was also taken into account as it is likely that an otosclerosis causing variant will have an impact on the structure or function of the gene it encodes within the stapes. Of the 144 variants identified during the prioritisation process, 7% were located within genes significantly up-regulated and 4% within genes significantly down-regulated in otosclerotic stapes compared to control stapes. The RNA sequencing data was particularly useful for ranking the variants in family B, as the stapes from the proband in this family was one of the otosclerotic stapes sequenced. Therefore variants in genes that were significantly dysregulated in this particular stapes compared to controls were strongly prioritised. In addition, variants encoding premature stop codons and frameshift

mutations, as well as those predicted have a strong damaging effect on the resulting protein based on predictions from Polyphen and SIFT, were prioritised, as it would be expected that these would have the greatest effect on the encoded protein product.

Due to concerns regarding the subjectivity of candidate variant selection at this stage, ranking of variants was performed independently by both Joanna Ziff and Sally Dawson, before a list of candidates was agreed upon. Based on the ranking criteria, it was decided that the 23 best candidate variants would be prioritised for segregation analysis in the first instance, with a second round of variants selected if no good segregating causal genes were identified. The variants selected for segregation analysis in each family are summarised in Table 4.8 and Table 4.9 and are further detailed at the start of each family section in Table 4.10, Table 4.12, Table 4.14 and Table 4.16.

All 23 variants were successfully validated by Sanger sequencing or restriction endonuclease digests of the proband of each family, indicating the high quality of the variant calling method. However variant c.1300G>A in hedgehog interacting protein-like 2 (*HHIPL2)* in family B was found to be homozygous in the proband through RFLP genotyping, whilst the WES data suggested that it was heterozygous. This variant is further discussed in section 4.8.

Gene	Gene name	Variant	Family	Main reason(s) why variant selected
DNAH5	Dynein Axonemal Heavy Chain 5	ENST00000265104:c:7883T>G	A	Previously unreported variant located within highly conserved region of gene that encodes critical component of the osteoclastic bone resorptive machinery.
FRYL	Furry Homolog-like (Drosophila)	ENST00000358350:c:227G>T	A	Previously unreported variant located within highly conserved region of gene that was found to be downregulated in otosclerotic stapes during RNA sequencing.
PTK6	Protein Tyrosine Kinase 6	ENST00000217185:c:652A>T	A	Previously unreported variant located in a gene that is activated by osteopontin which plays important role in bone remodelling and which was found to be upregulated in otosclerotic stapes during RNA sequencing.
SMAP1	Stromal Membrane-associated Protein 1	ENST00000370452:c:1224_1226dupCAT	A	Located within linked region OTSC7 which contains just 22 protein coding genes.
ACE	Angiotensin I Converting Enzyme	ENST00000290866:c:3304C>A	B	Located within PPARγ binding site and within a gene found to be associated with ototsclerosis in a candidate gene case control association study.
ANKS1A	Ankyrin Repeat and SterileAalpha Motif 1A	ENST00000373990:c:1291C>A	B	Rare variant located with PPARγ binding site.
CYP2D6	Cytochrome P450	ENST00000360608:c:345delC	B	Deletion leading to framsehift located in a gene expressed in spongy bone after fracture that has been linked with bone resoprtion and turnover in post-menopause women.
HHIPL2	Hedgehog Interacting Protein-like 2	ENST00000343410:c:1300G>A	B	Previously unreportedhighly conserved variant located within PPARβ binding site.
LEPRE1	Leucine Proline-enriched Proteoglycan 1	ENST00000296388:c:1626G>A	B	Previously unreported variant located in gene found to be downregulated in stapes of proband that is required for collagen synthesis, and which can cause Osteogeneis Imperfecta Type VIII when mutated in humans and delayed ossification when mutated in mice.
MIA3	Melanoma Inhibitory ActivityFamily 3	ENST00000344922:c:707G>A	B	Previously unreported highly conserved SNP located in gene required for collagen secretion and which impairs collagen secretion and arrest of chondrocyte maturation when mutated in mice.
SERPINF1	Serpin Peptidase Inhibitor, Clade F	ENST00000254722:c:601G>A	B	Rare variant in gene which can cause Osteogenesis Imperfecta Type VI when mutated.
VPS53	Vacuolar Protein Sorting-53 Homolog	ENST00000437048:c:107C>G	B	Previously unreported variant found to be upregulated in otosclerotic stapes duruing RNA sequencing.

Table 4.8 Table showing variants selected for segregation analysis in families A and B

The trranscript ID in the variant column is taken from the Ensembl database. Details of variants selected in families C and D continue over page

Gene	Gene name	Variant	Family	Main reason(s) why variant selected
EGLN3	EGL Nine Homolog 3	ENST00000250457:c:496C>T	C	Previously unreported variant located in gene involved in Osteosarcoma.
ERCC6	ATP-dependent Helicase ERCC6	ENST00000355832:c:2913G>C	C	Previously unreported variant located in gene involved in Ossifying Fibroma.
PDLIM5	PDZ and LIM Domain 5	ENST00000317968:c:842G>A	C	Previously unreported highly conserved variant involved in BMP signalling.
TRIM17	Tripartite Motif Containing 17	ENST000003666980:c:915C>G	C	Previously unreported variant that introduces premature stop codon.
COL1A2	Collagen Type I Alpha 2	ENST00000297268:c:808G>A	D	Rare variant located within PPARγ binding site in gene encoding the pro-alpha2 chain of type I collagen which can cause Ehlers Danlos Syndrome and Osteogenesis Imperfecta when mutated in humans and bone deformities when mutated in mice.
FOXK1	Forkhead Box K1	ENST00000328914:c:1384G>A	D	Previously unreported variant in gene that mediates TGFβ signalling in drosophila.
FZD2	Frizzled Homolog 2 (Drosophila)	ENST00000315323:c:655C>T	D	Rare variant in gene involved in wnt signalling which leads to abnormal cochlear hair cells in approximately 50% of mice homozygous for a reporter allele.
GNGT1	Guanine Nucleotide Binding Protein	ENST00000248572:c:82C>A	D	Previously unreported variant encoding gene involved in BMP signalling and mutations in which can cause Osteoporosis,Ehlers-Danlos syndrome, type VIIB and Osteogenesis Imperfecta Type II.
mir183	MicroRNA 183	ENST00000384958:n:81G>T	D	Rare highly conserved variant encoding gene in family of microRNAs that regulate chloride intracellular channel 5 expression in inner hair cells.
SHANK2	SH3 and Multiple Ankyrin Repeat Domains 2	ENST00000294018:c:2171C>T	D	Rare variant located with PPARγ binding site and in gene in which mutations can cause autosomal recessive deafness.
ZNF225	Zinc Finger Protein 225	ENST00000262894:c:1698:A>T	D	Rare variant in gene which activates the *COL1A2* promoter, which mediates TGFβ activation and which results in increased osteoclast cell number, abnormal trabecular bone morphology and decreased bone mineral density when mutated in mice.

Table 4.9 Table showing variants selected for segregation analysis in families C and D

The trranscript ID in the variant column is taken from the Ensembl database.

4.6 Rationale for segregation analysis and confirmation of a causal role

Segregation analysis was performed for the 23 top priority variants in the remaining family members, to identify which of these variants segregated with the disease in each relevant family. Since it could be assumed that all affected family members would share the disease causing variant, segregation analysis was used to eliminate any variants that were absent in additional affected family members from further analysis. Similarly, variants present in unaffected family members could also be eliminated from further analysis. However, due to variable penetrance of otosclerosis, it is possible that unaffected family members may possess the disease genotype without displaying otosclerosis symptoms. For this reason it was decided that if just one unaffected family member possessed the variant, it would not be sufficient to eliminate it from further analysis. However, if more than one unaffected individual possessed the variant, it was decided that it was appropriate to assume that it was unlikely to be disease causing and thus eliminate it from further analysis. In addition, when performing segregation analysis, the age of the family members and age of symptom onset in the family had to be considered. Since otosclerosis is an adult onset condition, it is possible that some family members with the disease genotype but no symptoms may go on to develop otosclerosis in the future. For this reason, if a family member younger the maximum age of disease onset in the family possessed the variant, it was not eliminated from further analysis, unless two additional unaffected family members also possessed the variant.

Identifying a rare nonsynonymous variant that segregates with a disease in a family exhibiting monogenic inheritance of a Mendelian trait, may indicate a possible role for the variant in the disease process. However, this evidence alone is not sufficient to indicate that the variant is disease causing, even if it is located within a gene with a predicted biological role relevant to the phenotype. Further evidence is essential in order to implicate the variant in disease pathogenesis. The best evidence to support this would be to identify the same rare variant within an unrelated family also exhibiting monogenic inheritance of the disease, or identifying alternative pathogenic variants within the same gene. For this reason, those variants that survived segregation analysis were genotyped in 53 additional unrelated probands, in addition to the 4 familial probands, who had been recruited to the study, each of whom had reported a strong family history of otosclerosis in their questionnaire responses.

4.7 Genotyping in Family A

In family A, four variants were prioritised for segregation analysis. These are illustrated in Table 4.10.

Gene	Gene name	Chromosomal Location	DNA Change	Amino Acid Change	Minor Allele Frequency	GERP statistic	Polyphen Prediction
DNAH5	Dynein Axonemal Heavy Chain 5	5p15.2	ENST00000265104:c:7883T>G	F/C Nonsynonymous coding	Previously unreported	5.87	Probably Damaging(0.973)
FRYL	Furry Homolog-like (Drosophila)	4p11	ENST00000358350:c:227G>T	R/L Nonsynonymous coding	Previously unreported	5.91	Probably Damaging(0.999)
PTK6	Protein Tyrosine Kinase 6	20q13.3	ENST00000217185:c:652A>T	I/F Nonsynonymous coding	Previously unreported	2.85	Possibly Damaging (0.762)
SMAP1	Stromal Membrane-associated Protein 1	6q13	ENST00000370452:c:1224_1226dupCAT	H Amino acid duplication	No information on indels	INDEL	No information on indels

Gene	In OTSC Linked Region	RNA Sequencing Study Data	Biological Role	Mouse model
DNAH5	n/a	n/a	Part of a microtubule-associated motor protein complex with ATP-ase activity. Constitutes critical components of the osteoclastic bone resorptive machinery.	Mice homozygous for a disruption in this gene display postnatal lethality, hydrocephalus, and respiratory infections
FRYL	n/a	Downregulated in otosclerotic stapes	Plays a key role in maintaining the integrity of polarized cell extensions during morphogenesis, regulates the actin cytoskeleton and may function as a transcriptional activator.	n/a
PTK6	n/a	Upregulated in otosclerotic stapes	Cytoplasmic nonreceptor protein kinase which may function as an intracellular signal transducer in epithelial tissues. It is activated by osteopontin (OPN) aka bone sialoprotein which plays important role in bone remodelling.	Mice homozygous for a null allele display increased villus length in jejunum and ileum
SMAP1	OTSC7 Dutch family (region has 22 protein coding genes)	n/a	GTPase activating protein that acts on ARF6. Plays a role in clathrin-dependent endocytosis. May play a role in erythropoiesis.	Mice homozygous for a knock-out allele exhibit perturbed receptor trafficking and myelodysplasia

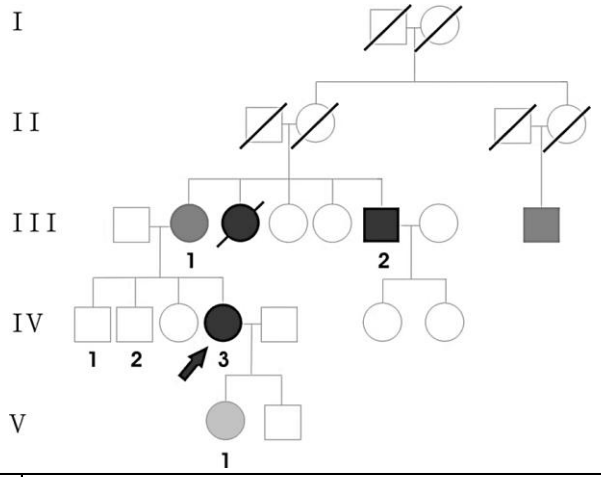
Table 4.10 Table illustrating the 4 variants prioritised for segregation analysis in family A

The DNA change column includes the transcript ID from the Ensembl database (The Ensembl ID stated is in most cases the major transcript ID, with the exception of *SMAP1* where the variant is not located within the major transcript). The minor allele frequency data is stated from the 1000 Genomes project European population and the NHLBI cohort American European population. The Polyphen prediction score is a rating from 0-1 where 0 indicates variants predicted to be benign and 1 indicates variants that are very likely to be damaging.

Genotyping was performed for each of these variants in all members of family A through Sanger sequencing. All four of these variants were detected in both confirmed cases of otosclerosis in individuals III:2 and IV:3, as well as in the presumptive case of otosclerosis in individual III:1 (Table 4.11). However, due to the presence of each variant in at least one unaffected family member, none of the four variants fully segregated with the disease. Unaffected individual IV:1 was heterozygous for SNP c.227G>T in *FRYL* and duplication c.1224_1226dupCAT in *SMAP1*, while unaffected individual IV:2 was heterozygous for SNP c.7883G>T in *DNAH5* and c.652A>T in *PTK6*. Neither individual IV:1 or IV:2 has otosclerosis symptoms and at ages 54 and 53 respectively, are substantially older than the oldest reported age of onset in this family, which is at age 30. However due to variable penetrance, none of these genes were ruled out on the basis of just one unaffected individual possessing the variant. Family member V:1, who has symptoms of hearing loss but whose audiogram is not suggestive of otosclerosis, is heterozygous for the variants in *FRYL* and PTK6, however is uninformative for the purpose of segregation analysis as at age 27, she is younger than the maximum age of symptom onset in this family.

4.7.1 Genotyping variants that segregate with otosclerosis in family A in an otosclerosis cohort

All four variants were genotyped in a cohort of 56 additional unrelated individuals, each reporting a family history of otosclerosis. Variant c.7883T>G in *DNAH5* and c.652A>T in *PTK6* were genotyped through restriction endonuclease digests using OliI and MlsI respectively (see Appendix), but neither variant was detected in any additional probands within the cohort, resulting in cohort allele frequencies of 0.0088 (Table 4.11). SNP c.227G>T in *FRYL* and the 1224_1226dupCAT duplication in *SMAP1* were genotyped through Sanger sequencing due to no appropriate restriction endonuclease for RFLP genotyping. The variant in *FRYL* was not identified in any additional probands and no other variants were identified in the cohort within *FRYL* exon 6, the exon in which the variant in family A was located. However the three base CAT duplication found in the affected members of family A (Figure 4.7A) and which results in a duplication of amino acid 408, was found in two other members of the cohort in a heterozygous state and in one other member of the cohort in a homozygous state resulting in a cohort allele frequency of 0.044 (Figure 4.7B and Table 4.11).



Pedigree of Family A

square=male; circle=female

slash=deceased individual

open symbol=clinically unaffected

solid symbol=hearing impairment

black symbol=confirmed otosclerosis

dark grey symbol=presumptive otosclerosis

light grey symbol=other hearing loss

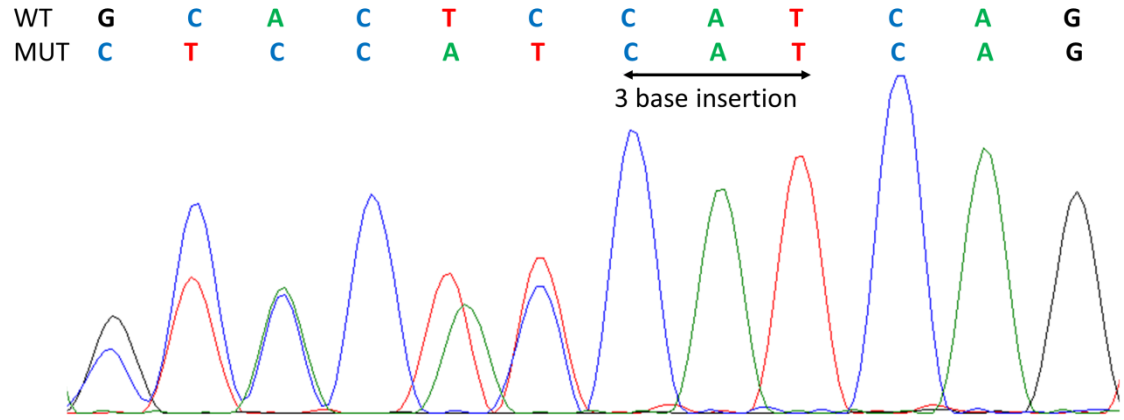
arrow=proband.

Gene	Variant	Individual						Segregates in family A	Genotyping method	Frequency of allele in otosclerosis cohort	Frequency of allele in EUR population
		III:1	III:2	IV:1	IV:2	IV: 3	V:1				
		Presumptive Otosclerosis	Confirmed Otosclerosis	Unaffected	Unaffected	Confirmed Otosclerosis	Undiagnosed hearing loss				
		Age 78	Age 64	Age 54	Age 53	Age 48	Age 27				
DNAH5	**7883T>G**	Het (T/G)	Het (T/G)	Wt (T/T)	Het (T/G)	Het (T/G)	Wt (T/T)	Yes with exception of one unaffected	RFLP - OliI	0.0088	Previously unreported
FRYL	**227G>T**	Het (G/T)	Het (G/T)	Het (G/T)	Wt (G/G)	Het (G/T)	Het (G/T)	Yes with exception of one unaffected	Sanger Sequencing	0.0088	Previously unreported
PTK6	**652A>T**	Het (A/T)	Het (A/T)	Wt (A/A)	Het (A/T)	Het (A/T)	Het (A/T)	Yes with exception of one unaffected	RFLP - MlsI	0.0088	Previously unreported
SMAP1	**1224_1226dupCAT**	Het (C/CCAT)	Het (C/CCAT)	Het (C/CCAT)	Wt (C/)	Het (C/CCAT)	Wt (C/)	Yes with exception of one unaffected	Sanger Sequencing	0.044	Data not avilable for indels

Table 4.11 Table showing genotypes of all members of family A

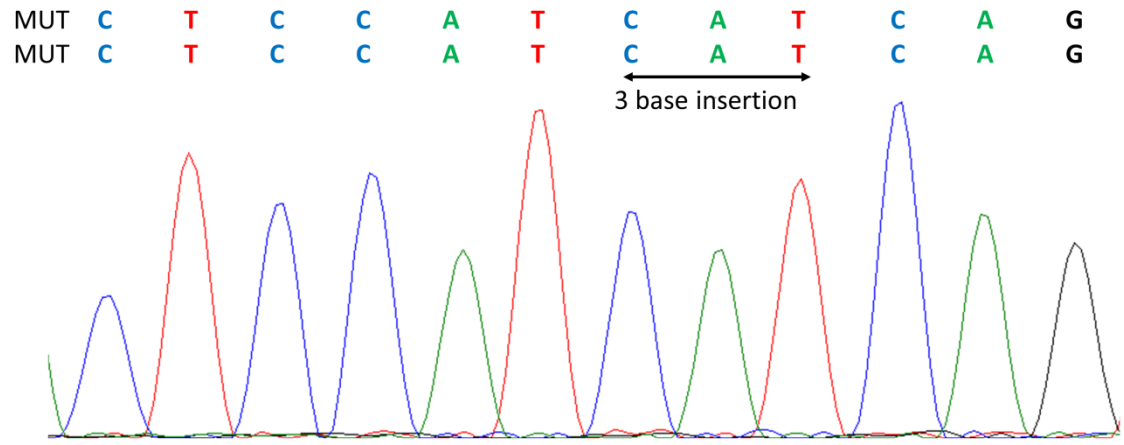
The frequency in the general population is based on 1000 Genomes Project data.

A.



Chromatogram of a heterozygote for the CAT duplication c.1224_1226dupCAT in *SMAP1*

B.



Chromatogram of a homozygote for the CAT duplication c.1224_1226dupCAT in *SMAP1*

Figure 4.7 Chromatograms showing c.1224_1226dupCAT in *SMAP1*

(A) Chromatogram of the genotype of the proband in family A showing the heterozygous three base CAT duplication in exon 10 of the SMAP1 gene. (B) Chromatogram of the genotype of the additional cohort member who is homozygous for the three base CAT duplication in exon 10 of the SMAP1 gene (Reverse complement of original sequence with reverse primer shown for clarity).

Each of the three additional study participants in whom the c.1224_1226dupCAT duplication had been identified were contacted, in an attempt to recruit additional affected family members for genotyping. Two of these individuals had no living affected relatives with the condition, but one of the heterozygous probands had an affected son with a confirmed otosclerosis diagnosis who was successfully recruited. Sanger sequencing showed that he was also heterozygous for the three base duplication.

Sanger sequencing also revealed an additional variant within exon 10 of *SMAP1*, the exon in which the variant in family A is located. This nonsynonymous c.1247T>C SNP is present in the cohort at a frequency of 0.510 and causes amino acid change from methionine to threonine. However this SNP is very common and is present in the European cohort of the 1000 Genomes project database at a frequency of 0.505. It was therefore assumed that it was too common to be a possible disease causing variant and therefore eliminated from further analysis

4.7.2 Discussion of variants segregating with otosclerosis in family A

Following segregation analysis of the four variants prioritised in family A within *DNAH5, FRYL, PTK6* and *SMAP1*, all were found to segregate with otosclerosis in most family members with the exception of a single heterozygous unaffected individual in each case. Due to the issue of variable penetrance, none of these genes were ruled out as possible causative variants. It is possible that any of these four variants may contribute to disease pathogenesis in this family. However, only one of these variants, the CAT duplication.1224_1226dupCAT in *SMAP1,* was identified in other members of the otosclerosis cohort

The identification of the duplication c.1224_1226dupCAT in *SMAP1* in additional unrelated individuals with otosclerosis suggested that this was the most likely of the four candidates to be involved in otosclerosis. This three base pair duplication results in duplication of amino acid 408 in a specific *SMAP1* isoform known as *SMAP1*-003 (ENST00000370452), which is located on chromosome 6q13. It is located in exon 10 of 11 of *SMAP1-003*, an exon which is unique to this specific *SMAP1* protein coding transcript. *SMAP1* is a GTPase activating protein that plays a role in endocytosis. It encodes a protein similar to the mouse stromal membrane-associated protein-1, suggesting that it may be involved in the stimulatory activity of stromal cells, the connective tissue precursor cells in the bone marrow that go on to differentiate into osteoblasts. It is therefore possible that aberrant expression of this *SMAP1* isoform as a result of a mutation, could interfere with the bone remodelling process, suggesting a potential role for *SMAP1* in the aetiology of otosclerosis.

SMAP1 is located within the OTSC7 region identified by Guy VanCamp's research group at the University of Antwerp within a Dutch family through linkage analysis (Thys et al., 2007b). The OTSC7 region mapped in this family is the smallest of the eight published otosclerotic loci, containing just 22 protein coding genes. This suggests that this Dutch family has a mutation in the immediate region surrounding *SMAP1* that segregates with otosclerosis. It is therefore possible that this family also has a variant within *SMAP1.* It would be useful to genotype this

Dutch family to identify whether or not they possess this variant or another variant within the gene, which could indicate that variants in *SMAP1* are over-represented in families with otosclerosis. Having said this, a second otosclerosis family of Greek origin was also mapped to this region through linkage analysis, but the overlap between the critical regions of these two families did not contain the *SMAP1* gene. Whilst it is possible that different genes were responsible for the disease in each of these families due to the heterogeneous nature of otosclerosis, it is likely that the disease causing gene in these families is located within the overlapping region, and therefore a gene other than *SMAP1* is responsible for pathogenesis.

In total, in the cohort of 57 unrelated probands, which includes the probands from all four families plus an additional 53 probands, duplication c.1224_1226dupCAT was found to be homozygous in one and heterozygous in three unrelated individuals with a strong family history of otosclerosis. The minor allele frequency within the cohort was therefore 0.044. However, following genotyping it was found that this duplication is more common in the general population than was initially expected, with a minor allele frequency of 0.03 based on data from the European cohort within the 1000 Genomes Project (www.1000genome.org). This was missed during the initial variant prioritisation process as at that time, frequency data for insertions and deletions was not available. Had this been known, this duplication would not have survived the 0.02 frequency cut off threshold required to qualify the variant for segregation analysis. In addition, analysis of the RNA sequencing data (chapter 5) shows that *SMAP1* exon 10, the exon in which duplication c.1224_1226dupCAT is located, is not detected in any of the otosclerotic or control stapes samples. Since this exon is unique to transcript *SMAP1-003*, it indicates that this transcript is not expressed in the adult human stapes. This suggests that it is unlikely that the protein encoded by this transcript is involved in the development of otosclerosis, as it would be expected that any transcript involved in the disease process would be likely to be expressed in human stapes bones.

Despite this, the variant should not be ruled out as having possible involvement in otosclerosis. It is present in the cohort at a frequency 1.5x greater than that in the general population and was found in a homozygous state in one individual in the cohort, the odds of which are just 0.0009 in the general population. In addition, it was also found to be present in the affected son of one of the additional heterozygous individuals in the cohort. Although it is unlikely that this duplication is causative, it is possible that it could be a disease susceptibility factor which, when combined with other variants, could result in the otosclerosis phenotype.

For future analysis of family A, it would be necessary to decide whether to further follow up any of the four segregating variants in *DNAH5, FRYL, PTK6* or *SMAP1*, or whether to identify

additional variants that were not initially prioritised for segregation analysis. To follow up the segregating variants, it would be necessary to recruit a larger cohort of familial probands in the hope of identifying one of these variants within an additional unrelated family, which would support the evidence for a role of the variant in disease pathology. Alternatively, further Sanger sequencing could be performed on the remainder of the genes to identify if any additional variants within each gene were present within the cohort. This could provide supporting evidence for a role of these genes in otosclerosis pathogenesis.

Since none of these genes fully segregated with the disease in family A, in addition to the fact that the variants in *DNAH5, FRYL* and *PTK6* were not identified in any additional cohort members, combined with the lack of evidence to implicate *SMAP1* in disease pathogenesis, suggests that it would be desirable to prioritise additional variants for segregation analysis and follow up genotyping in the remainder of the otosclerosis cohort, in the hope of identifying the disease causing variant in family A.

4.8 Genotyping in Family B

In family B, a large number of good candidate variants were identified during step-wise prioritisation and filtering, of which eight were prioritised for segregation analysis (Table 4.12).

Gene	Gene name	Chromosomal Location	DNA Change	Amino Acid Change	Minor Allele Frequency	GERP statistic	Polyphen Prediction
ACE	Angiotensin I Converting Enzyme	17q23.3	ENST00000290866:c.3304C>A	P/T Nonsynonymous coding	0.008 (1000 Genomes) 0.004 (NHLBI)	4.76	Probably Damaging(0.997)
ANKS1A	Ankyrin Repeat and Sterile Aalpha Motif 1A	6p21.31	ENST00000373990:c.1291C>A	P/T Nonsynonymous coding	0.002 (1000 Genomes)	4.56	Probably Damaging(0.98)
CYP2D6	Cytochrome P450	22q13.2	ENST00000360608:c.345delC	Frameshift coding	No information on indels	INDEL	No information on indels
HHIPL2	Hedgehog Interacting Protein-like 2	1q41	ENST00000343410:c.1300G>A	R/C Nonsynonymous coding	Previosuly unreported	5.52	Benign(0.404)
LEPRE1	Leucine Proline-enriched Proteoglycan 1	1p34.2	ENST00000296388:c.1626G>A	G/R Nonsynonymous coding	Previosuly unreported	-10.3	Unknown(0)
MIA3	Melanoma Inhibitory Activity Family 3	1q41	ENST00000344922:c.707G>A	D/G Nonsynonymous coding	Previosuly unreported	5.36	Probably Damaging(0.999)
SERPINF1	Serpin Peptidase Inhibitor, Clade F	17p13.3	ENST00000254722:c.601G>A	D/N Nonsynonymous coding	0.004 (1000 Genomes) 0.003 (NHLBI)	2.16	Benign(0)
VPS53	Vacuolar Protein Sorting-53 Homolog	17p13.3	ENST00000437048:c.107C>G	P/R Nonsynonymous coding	Previosuly unreported	4.73	Probably Damaging(1)

Gene	In OTSC Linked Region	RNA Sequencing Study Data	Biological Role	Mouse model
ACE	n/a	n/a	Role in the renin-angiotensin system.Associated with ototsclerosis in association study. SNP is located within PPARγ binding site.	Mice homozygous for targeted mutations show variable phenotypes.
ANKS1A	OTSC3 (region has 423 protein coding genes)	n/a	Regulates EPHA8 receptor tyrosine kinase signaling to control cell migration and neurite retraction. SNP is located with PPARγ binding site	Mice homozygous for a gene trapped allele exhibit a normal phenotype.
CYP2D6	n/a	n/a	Role in drug metabolism. Linked with bone resoprtion and bone turnover in post-menopause women. Expressed in spongy bone after fracture.	Gene absent in mice
HHIPL2	OTSC10 (region has 148 protein coding genes)	n/a	Involved in Hedgehog signalling cascade. SNP is located within PPARβ binding site.	n/a
LEPRE1	n/a	Downregulated in stapes of proband compared to controls	Member of the collagen family. Mutations cause Osteogeneis Imperfecta Type VIII	KO Mice exhibit decreased bone density, delayed ossification.
MIA3	OTSC10 (region has 148 protein coding genes)	n/a	Required for collagen secretion.	Mice homozygous null display impaired collagen secretion.
SERPINF1	n/a	n/a	Inhibitor of angiogenesis. Induces extensive neuronal differentiation in retinoblastoma cells. Mutations cause Osteogenesis Imperfecta Type VI	LOF results in increased microvasculature
VPS53	n/a	Upregulated in otosclerotic stapes	May be involved in transport of endosomes to the Golgi.Necessary for proper lysosomal sorting of acid hydrolases.	KO mice embryonic lethal prior to E12.5

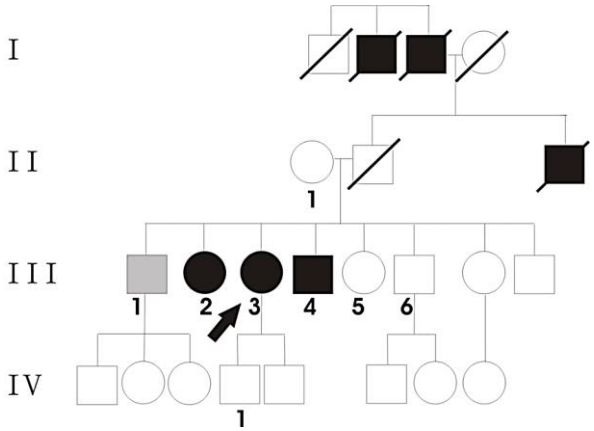
Table 4.12 Table illustrating the 8 variants prioritised for segregation analysis in family B

The DNA change column includes the transcript ID from the Ensembl database (The Ensembl ID stated is in most cases the major transcript ID, with the exception of *ANKS1A* where the variant is located within a transcript predicted to be present in Ensembl release 67 Assembly Grch37 but is no longer predicted to be present in the current Ensembl database). The minor allele frequency data is stated from the 1000 Genomes project African population and the NHLBI cohort African-American population. The Polyphen prediction score is a rating from 0-1 where 0 indicates variants predicted to be benign and 1 indicates variants that are very likely to be damaging.

In family B, all variants with the exception of SNP c.3304C>A in *ACE* were identified in all affected family members; III:2, III:3 and III:4 during Sanger sequencing (Table 4.13). The *ACE* variant did not segregate with otosclerosis in the family, as affected individual III:4 was wild-type for the variant. In addition, variant c.1626G>A within *LEPRE1* and c.707A>G in *MIA3* were ruled out of further analysis, as both were heterozygous in individual II:1, the 87 year old unaffected mother of the three affected siblings, who it was assumed could not possess the disease causing variant as the trait appears to have been passed down the paternal line. The variants in both of these genes were also present in at least two of the proband's unaffected siblings, suggesting that they are unlikely to be causative variants within this family. SNP c.1300G>A in *HHIPL2* was also ruled out of further analysis as the G/G genotype of individual III:5 appears to be inaccurate given the A/A genotype of her mother, individual II:1, as it would be expected that individual III:5 would possess at least one A allele inherited from her mother (Table 4.13). Due to the large number of other variants analysed in this family, it would be very unlikely that the maternity of this individual would need to be questioned or that there had been a mix up in the samples when genotyping. In addition, since the WES data did not match with the RFLP genotyping data for this variant, it would appear that this this SNP is located within a region prone to sequencing or mapping artefacts. Deletion c.345delC in *CYP2D6* was found to be heterozygous in unaffected individual III:1 who has presbycusis, but since otosclerosis exhibits variable penetrance in this family, the deletion in *CYP2D6* was not ruled out of further analysis. SNP c.601G>A in *SERPINF1*, c.107C>G in *VPS53* and variant c.1291C>A in *ANKS1A* fully segregated with the family with the exception of unaffected member IV:1, who at age 35 is younger than the oldest reported age of onset in this family at age 50 (Table 4.13).

4.8.1 Genotyping variants that segregate with otosclerosis in family B in an otosclerosis cohort

The variants in *SERPINF1*, *VPS53*, *ANKS1A* and *CYP2D6* were genotyped in a cohort of 56 additional unrelated individuals with familial otosclerosis. The variants in *ANKS1A*, *CYP2D6* and *VPS53* were genotyped using TseI, DrdI and PpuMI respectively through restriction endonuclease digests (see Appendix). However, the variants were not detected in any additional probands, resulting in a cohort allele frequency of 0.0088 in each case (Table 4.13). The c.601G>A nonsynonymous SNP in *SERPINF1* was genotyped through Sanger sequencing (Figure 4.8), due to no appropriate restriction enzyme for RFLP genotyping, but was also not identified in any additional probands.



Pedigree of Family B

square=male; circle=female

slash=deceased individual

open symbol=clinically unaffected

solid symbol=hearing impairment

black symbol=confirmed otosclerosis

dark grey symbol=presumptive otosclerosis

light grey symbol=other hearing loss

arrow=proband

Gene	Variant	Individual								Segregates in family B	Genotyping method	Frequency of allele in otosclerosis cohort	Frequency of allele in EUR population
		II:1	III:1	III:2	III:3	III: 4	III:5	III:6	IV:1				
		Unaffected	Presbycusis	Confirmed Otosclerosis	Confirmed Otosclerosis	Confirmed Otosclerosis	Unaffected	Unaffected	Unaffected				
		Age 87	Age 67	Age 65	Age 63	Age 62	Age 60	Age 59	Age 35				
ACE*	**3304C>A**	Wt (C/C)	Het (C/A)	Het (C/A)	Het (C/A)	Wt (C/C)	Wt (C/C)	Wt (C/C)	Wt (C/C)	No	Sanger sequencing	Genotyping not performed	0.008
ANKS1A*	**1291C>A**	Wt (C/C)	Wt (C/C)	Het (C/A)	Het (C/A)	Het (C/A)	Wt (C/C)	Wt (C/C)	Het (C/A)	Yes	RFLP- TseI	0.0088	0.002
CYP2D6	**345delC**	Wt (C/C)	Het (C/-)	Het (C/-)	Het (C/-)	Het (C/-)	Wt (C/C)	Wt (C/C)	Het (C/-)	Yes with exception of one unaffected	RFLP - DrdI	0.0088	Previously unreported
HHIPL2 **	**1300G>A**	Mut (A/A)	Het (G/A)	Het (G/A)	Mut (A/A)	Mut (A/A)	Wt(G/G)	Het (G/A)	Het (G/A)	No	Sanger sequencing	Genotyping not performed	Previously unreported
LEPRE1	**1626G>A**	Het (G/A)	Mut (A/A)	Het (G/A)	Mut (A/A)	Het (G/A)	Wt (G/G)	Mut (A/A)	Het (G/A)	No	Sanger sequencing	Genotyping not performed	Previously unreported
MIA3	**707A>G**	Het (A/G)	Het (A/G)	Het (A/G)	Het (A/G)	Het (A/G)	Wt (A/A)	Wt (A/A)	Het (A/G)	No	Sanger sequencing	Genotyping not performed	Previously unreported
SERPINF1	**601G>A**	Wt (G/G)	Wt (G/G)	Het (G/A)	Het (G/A)	Het (G/A)	Wt (G/G)	Wt (G/G)	Het (G/A)	Yes	Sanger sequencing	0.0088	0.0041
VPS53	**107C>G**	Wt (C/C)	Wt (C/C)	Het (C/G)	Het (C/G)	Het (C/G)	Wt (C/C)	Wt (C/C)	Het (C/G)	Yes	RFLP- PpmuI	0.0088	Previously unreported

Table 4.13 Table showing genotypes of all members of family B

The frequency in the general population is based on 1000 Genomes Project data.

*Genotyping of *ACE* and *ANKS1A* was performed by Rimpal Nata **Genotyping of *HHIPL2* was performed by Kara Forsyth

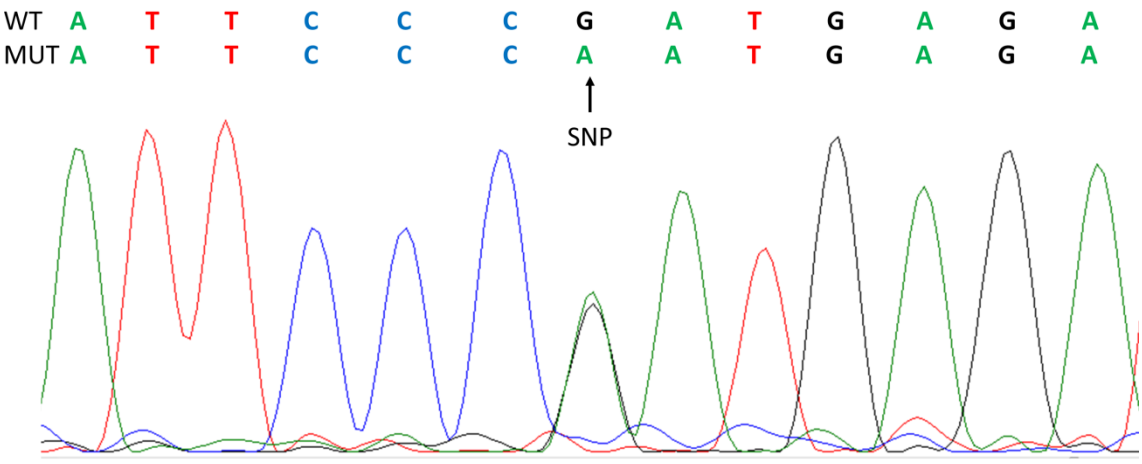


Figure 4.8 Chromatogram showing heterozygote for variant c.601G>A in *SERPINF1*

Chromatogram of the genotype of the proband in family B showing the heterozygous G>A SNP in exon 5 of the *SERPINF1* gene.

4.8.2 Discussion of variants segregating with otosclerosis in family B

It is possible that any of the four segregating variants in *SERPINF1*, *VPS53*, *ANKS1A* and *CYP2D6* may contribute to disease pathogenesis in this family, however, none were found in any additional members of the cohort. Despite this, the role of the protein encoded by *SERPINF1* as well as the location of the c.601G>A SNP within this gene suggests that it is potentially an excellent candidate variant within this family.

SERPINF1 encodes PEDF (Pigment epithelium-derived growth factor) also known as Caspin, a 50kDa protein which is 418 amino acids in length. PEDF is a neutrotrophic protein that is known to induce neurodifferentiation in retinoblastoma cells. It is a potent inhibitor of angiogenesis and lacks the serine protease inhibitory activity that is characteristic of most other members of the *SERPIN* gene family. PEDF is known to be highly expressed in collagen rich tissues including bone, cornea and cartilage. 16 unique variants located within this gene have been reported within individuals with the connective tissue disorder Osteogenesis Imperfecta (OI) type VI (www.oi.gene.le.ac.uk). Patients with OI frequently experience conductive hearing loss similar to that of otosclerosis, which has led researchers to postulate that otosclerosis may be a localised manifestation of OI (Weber, 1930). PEDF is not detected in the serum of individuals with OI type VI but is in those with other forms of the condition (Rauch et al., 2012), suggesting that mutations in this gene may affect its expression in certain tissues. If this is the case, then it is possible that the family B mutation could affect the expression of PEDF in the stapes, since bone is a collagen rich tissue in which PEDF is normally highly expressed. There is therefore a possible role for the *SERPINF1* gene in the disease process in otosclerosis.

Further evidence of a role for *SERPINF1* in otosclerosis pathogenesis is the location of the SNP within the gene. SNP c.601G>A results in an amino acid change from aspartic acid 201 to asparagine and is located within a region that encodes an acidic part of the PEDF protein, where one aspartic acid residue and two glutamic acid residues are located within a five amino acid region. Acidic negatively charged regions of this protein are known to interact directly with positively charged regions on collagen genes (Meyer et al., 2002). Since the SNP in family B, causes a change from an acidic residue to a neutral one, it is possible that the variant could inhibit interactions between PEDF and collagen, which could result in effects on the bone matrix, of which collagen is the predominant component. In addition, the variant in family B is also located within a predicted NFκB binding site. The receptor activator of NFκB (RANK) also plays an important role in the rate of bone turnover as it is expressed on the surface of pre-osteoclasts which compete with osteoprotegerin (OPG), a protein secreted from osteoblasts, for binding to the RANK ligand on the surface of pre-osteoblasts. It is therefore possible that a variant within an NFκB binding site could have an impact on the NFκB signalling pathway, which could result in an imbalance in osteoblastogenesis relative to osteoclastogenesis, suggesting another potential way in which the family B mutation could lead to disease. Furthermore, there is also evidence that this c.601G>A SNP may affect expression of various PEDF isoforms (see Chapter 4 Part II). For these reasons, this SNP is an excellent otosclerosis candidate variant in family B.

In family B, it is not possible to rule out any of the four segregating variants for a role in otosclerosis pathophysiology. However, since the 601G>A SNP in *SERPINF1* is such a good candidate, it was decided that its role in otosclerosis aetiology would be further investigated as the top priority for this family, in preference to any of the other variants. Further analysis of *SERPINF1* is discussed in more detail in Chapter 4 Part II.

4.9 Genotyping in Family C

In Family C, step-wise prioritisation of variants and filtering generated the smallest number of good otosclerosis candidate variants of all the four families. Four variants were prioritised for segregation analysis (Table 4.14), including the variant in *TRIM17*, which was the only variant to survive both candidate and non-candidate gene filtering strategies.

Gene	Gene name	Chromosomal Location	DNA Change	Amino Acid Change	Minor Allele Frequency	GERP statistic	Polyphen Prediction
EGLN3	EGL Nine Homolog 3	14q13.1	ENST00000250457:c.496C>T	R/W Nonsynonymous coding	Previosuly unreported	2.47	Probably Damaging(0.997)
ERCC6	ATP-dependent Helicase ERCC6	10q11.23	ENST00000355832:c.2913G>C	K/N Nonsynonymous coding	Previosuly unreported	4.64	Probably Damaging(0.988)
PDLIM5	PDZ and LIM Domain 5	4q22.3	ENST00000317968:c.842G>A	R/H Nonsynonymous coding	Previosuly unreported	5.31	Probably Damaging(1)
TRIM17	Tripartite Motif Containing 17	1q42.13	ENST00000366980:c.915C>G	Y/* Stop codon gained	Previosuly unreported	STOP	Stop Codon

Gene	In OTSC Linked Region	RNA Sequencing Study Data	Biological Role	Mouse model
EGLN3	n/a	n/a	Cellular oxygen sensor that catalyzes, under normoxic conditions, the post-translational formation of 4-hydroxyproline in HIF alpha proteins. Involved in Osteosarcoma	Homozygous null mice display decreased apoptosis in SCG neurons
ERCC6	n/a	n/a	Essential factor involved in transcription-coupled nucleotide excision repair which allows RNA polymerase II-blocking lesions to be removed from active genes. Involved in Ossifying Fibroma	Homozygous mutant mice exhibit UV sensitivity and inactivation of transcription-coupled repair
PDLIM5	n/a	n/a	May play a role in the regulation of cardiomyocyte expansion and contributes to the regulation of dendritic spine morphogenesis in neurons. Involved in BMP signalling.	Mice homozygous for a KO allele exhibit impaired cardiac muscle contractility
TRIM17	OTSC10 (region has 148 protein coding genes)	n/a	Negatively regulates cell proliferation and interacts with mitotic checkpoint.	n/a

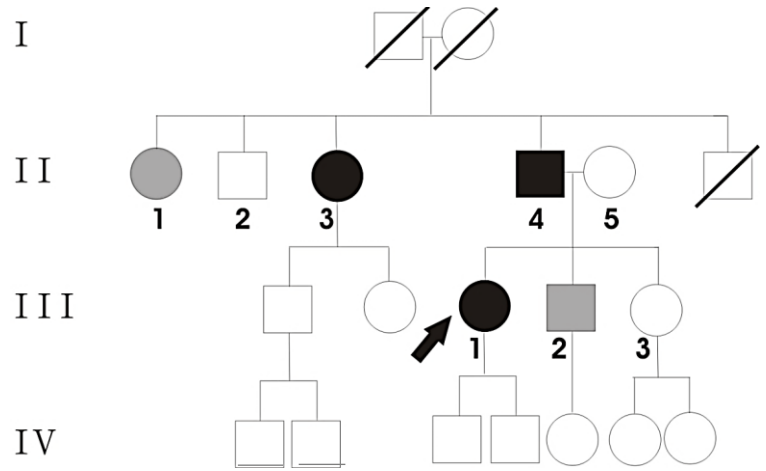
Table 4.14 Table illustrating the 4 variants prioritised for segregation analysis in family C

The DNA change column includes the transcript ID from the Ensembl database (The Ensembl ID stated is in all cases the major transcript ID). The minor allele frequency data stated is based on data from the 1000 Genomes project European population and the NHLBI cohort European American population. The Polyphen prediction score is a rating from 0-1 where 0 indicates variants predicted to be benign and 1 indicates variants that are very likely to be damaging.

In Family C, all four of these variants were identified in the three affected family members; II:3, II:4 and III:1 through Sanger sequencing. However, variant c.2913G>C in *ERCC6* and c.915C>G in *PDLIM5* were ruled out as each variant was also present in two or more unaffected family members. Since otosclerosis exhibits variable penetrance, the presence of a variant in more than one affected relative is suggestive that this variant is not involved in disease pathogenesis. Variant c.496C>T in *EGLN3* was found to be heterozygous in just one unaffected family member and c.915C>G in *TRIM17* segregated fully with the disease (Table 4.15). In this family, individual III:2 who reported early symptoms of hearing loss, is uninformative for the purpose of segregation analysis. This is because it is not known whether this individual has otosclerosis or another hearing condition.

4.9.1 Genotyping variants that segregate with otosclerosis in family C in an otosclerosis cohort

Genotyping was performed in a cohort of 56 additional unrelated individuals with familial otosclerosis for the segregating variants in *EGLN3* and *TRIM17*. Both variants were genotyped using restriction endonuclease digests (see Appendix). Enzyme AciI was used to detect the presence or absence of the variant in *EGLN3* and RsaI for the variant in *TRIM17*. However, neither variant was identified in any of the additional cohort members, resulting in a cohort allele frequency of 0.0088 for both variants (Table 4.15).



Pedigree of Family C

square=male; circle=female

slash=deceased individual

open symbol=clinically unaffected

solid symbol=hearing impairment

black symbol=confirmed otosclerosis

dark grey symbol=presumptive otosclerosis

light grey symbol=other hearing loss

arrow=proband

Gene	Variant	Individual								Segregates in family C	Genotyping method	Frequency of allele in otosclerosis cohort	Frequency of allele in EUR population
		II:1	II:2	II:3	II:4	II:5	III:1	III:2	III:3				
		Presbycusis	Unaffected	Confirmed Otosclerosis	Confirmed Otosclerosis	Unaffected	Confirmed Otosclerosis	Undiagnosed hearing loss	Unaffected				
		Age 83	Age 81	Age 78	Age 74	7th decade	Age 46	Age 45	Age 41				
EGLN3	**496C>T**	Wt (C/C)	Wt (C/C)	Het (C/T)	Het (C/T)	Wt (C/C)	Het (C/T)	Het (C/T)	Het (C/T)	Yes with exception of one unaffected	RFLP - AciI	0.0088	Previously unreported
ERCC6	**2913G>C**	Het (G/C)	Het (G/C)	Het (G/C)	Het (G/C)	Wt (G/G)	Het (G/C)	Het (G/C)	Het (G/C)	No	Sanger sequencing	Genotyping not performed	Previously unreported
PDLIM5	**842G>A**	Het (G/A)	Het (G/A)	Het (G/A)	Het (G/A)	Wt (G/G)	Het (G/A)	Het (G/A)	Wt (G/G)	No	Sanger sequencing	Genotyping not performed	Previously unreported
TRIM17	**915C>G**	Wt (C/C)	Wt (C/C)	Het (C/G)	Het (C/G)	Wt (C/C)	Het (C/G)	Het (C/G)	Wt (C/C)	Yes	RFLP - RsaI	0.0088	Previously unreported

Table 4.15 Table showing genotypes of all members of family C

The frequency in the general population is based on 1000 Genomes Project data.

4.9.2 Discussion of variants segregating with otosclerosis in family C

Following segregation analysis of variants identified in family C, the variant in *TRIM17* was found to fully segregate with otosclerosis in all family members and *EGLN3* with all but one heterozygous unaffected individual, which due to the issue of variable penetrance, was not ruled out as a possible causative variant. It is possible that either of these variants may contribute to disease pathogenesis in this family, however, neither was identified in additional members of the cohort. Despite this, the c.915C>G variant within *TRIM17* remains a very good candidate in this family.

Variant c.915C>G has not previously been reported in the literature and results in the introduction of a premature stop codon at amino acid 305 of the TRIM17 protein. This variant was the only one of the 144 surviving the variant prioritisation process across all four families to meet the criteria of both the candidate gene and non-candidate gene filtering strategy. It is located within the OTSC10 region, which was identified by Guy VanCamp's research group at the University of Antwerp (Schrauwen et al., 2011). It is therefore possible that this family could also possess a variant within *TRIM17,* however, since OTSC10 is one of the larger of the eight published otosclerotic loci, containing 148 protein coding genes, this is not necessarily the case. Despite this, since this SNP was the only prioritised variant found to segregate fully in family C, it is a strong candidate for this family.

TRIM17 is a member of the Tripartite Motif family. These genes tend to play a role in protein degradation and are related to one another through their structures which contain a number of similar functional domains. TRIM proteins have E3 ubiquitin ligase activity and play key roles in regulating signalling cascades involved in immune responses. TRIM17 contains three zinc-binding domains, a RING type zinc finger at residues 16-65, a B-box type zinc finger at residues 94-135, a coiled-coil region at 196-221 and an SPRY domain at residues 348-477 (Ogawa et al., 1998) . Since the *TRIM17* transcript is prematurely truncated at amino acid 305 in individuals possessing the SNP, affected members of family C are therefore missing the SPRY domain, a protein interaction module that recognises specific protein partners rather than a consensus sequence (Woo et al., 2006). Little is currently known about SPRY domains but they are known to be highly conserved with an ancient evolutionary origin (Rhodes et al., 2005). They are believed to have functions in immune system related processes and are found in negative regulators of cytokine activity (Nicholson and Hilton, 1998). Approximately 150 human proteins contain SPRY domains, including members of the TRIM family. It has been reported that the SPRY domain of the TRIM5α protein, binds to the viral capsid of some retroviruses including HIV, targeting HIV-1 restriction (Yap et al., 2005). This is of interest in the context of

otosclerosis given the link between the condition and a number of viral infections including measles. In addition, another family member of *TRIM17* known as *TRIM24* is known to be located within the OTSC2 locus and is a transcriptional co-activator that plays a role in bone remodelling.

It is also possible that the premature stop codon in *TRIM17* would lead to degradation of the entire protein, although this may be unlikely given that the variant is located in the final protein coding exon of the gene. TRIM17 is known to initiate neuronal apoptosis and is controlled by the PI3K/Akt/GSK3 pathway in cerebellar granule neurons (Lassot et al., 2010). It can negatively regulate cell proliferation through interaction with ZW10, a mitotic checkpoint protein (Endo et al., 2012). In addition, the variant in family C is predicted to be located within a GATA-1 binding site. These are genomic sequences which bind GATA-1 transcription factors involved in regulating erythroid development including the switch in production from foetal to adult haemoglobin. A mutation in this gene, such as the c.915C>G variant in family C could therefore inhibit any of the gene's functions, many of which could play a role in disease pathology.

Future research into the disease causing variant in family C, could focus either on follow up analysis of the segregating variants in *EGLN3* or *TRIM17*, or alternatively could focus on identification of additional variants that were not initially prioritised for segregation analysis. In this family, the choice of which variants to select for segregation analysis was easier than in the other three families as the four variants selected stood out as far better candidates than any of those remaining. It was therefore decided that it would be beneficial to prioritise the segregating variants for further analysis, rather than to select less desirable candidates. Although it was not possible to rule out the variant in *EGLN3*, since variant c.915C>G in *TRIM17* was such a good candidate, its role in otosclerosis aetiology was further investigated as the top priority for this family. Analysis of *TRIM17* is discussed in more detail in Chapter 4 Part II.

4.10 Genotyping in Family D

In family D, many good candidate variants emerged from the step-wise prioritisation and filtering processes, of which seven were prioritised for segregation analysis (Table 4.16).

Gene	Gene name	Chromosomal Location	DNA Change	Amino Acid Change	Minor Allele Frequency	GERP statistic	Polyphen Prediction
COL1A2	Collagen Type I Alpha 2	7q21.3	ENST00000297268:c:808G>A	V/I Nonsynonymous coding	0.0010 (1000 Genomes) 0.0028 (NHLBI)	4.86	Benign(0)
FOXK1	Forkhead Box K1	7p22.1	ENST00000328914:c:1384G>A	E/K Nonsynonymous coding	Previosuly unreported	4.94	Possibly Damaging(0.932)
FZD2	Frizzled Homolog 2 (Drosophila)	17q21.31	ENST00000315323:c:655C>T	R/C Nonsynonymous coding	0.0001 (NHLBI)	3.44	Possibly Damaging(0.732)
GNGT1	Guanine Nucleotide Binding Protein	7q21.3	ENST00000248572:c:82C>A	L/M Nonsynonymous coding	Previosuly unreported	2.95	Unknown(0)
mir183	MicroRNA 183	7q32.2	ENST00000384958:n:81G>T	P/L Nonsynonymous coding	0.0002 (NHLBI)	5.86	Possibly Damaging(0.799)
SHANK2	SH3 and Multiple Ankyrin Repeat Domains 2	11q13.4	ENST00000294018:c:2171C>T	P/L Nonsynonymous coding	0.0002 (NHLBI)	4.88	Possibly Damaging(0.799)
ZNF225	Zinc Finger Protein 225	19q13.31	ENST00000262894:c:1698:A>T	R/S Nonsynonymous coding	0.0009 (NHLBI)	1.67	Possibly Damaging(0.926)

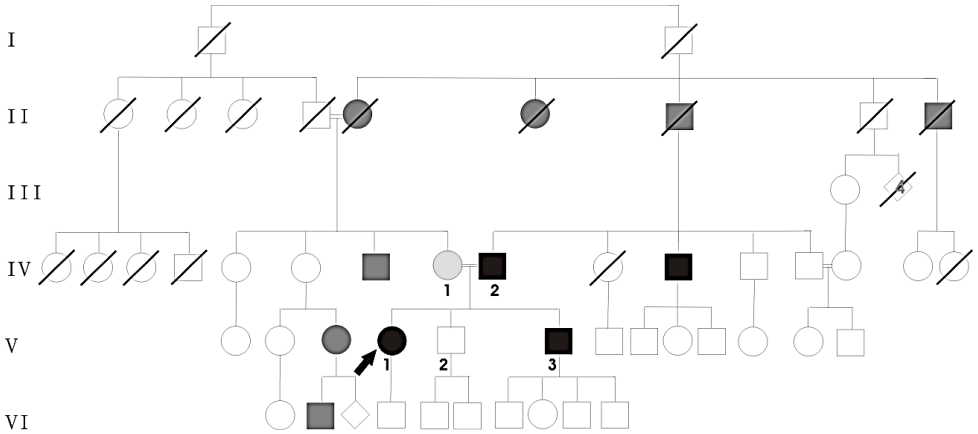
Gene	In OTSC Linked Region	RNA Sequencing Study Data	Biological Role	Mouse model
COL1A2	n/a	n/a	Encodes the pro-alpha2 chain of type I collagen. Mutations cause Ehlers Danlos Syndrome and Osteogenesis Imperfecta. SNP located within PPARγ binding site.	Mutant mice exhibit bone deformities.
FOXK1	n/a	Upregulated in otosclerotic stapes	Role in myogenic differentiation and in remodeling processes of adult muscles. Mediates TGFβ signalling in drosophila.	KO mice are runted and exhibit reduced myogenic progenitor cells
FZD2	n/a	n/a	Involved in wnt signalling.	Mice homozygous for reporter allele have abnormal hair cells
GNGT1	n/a	n/a	Involved in BMP signalling. Mutations causes Osteoporosis,Ehlers-Danlos syndrome, type VIIB, Osteogenesis Imperfecta Type II	Homozygous null mice display impaired light adaptation
mir183	n/a	n/a	Inhibits metastasis of osteosarcoma. mir183 family members regulate chloride intracellular channel 5 expression in inner hair cells	n/a
SHANK2	n/a	n/a	Adapter protein in the postsynaptic densityof excitatory synapses. Mutations cause Deafness, autosomal recessive 63. SNP located with PPARγ binding site	Mice homozygous for null mutations display hyperactivity
ZNF225	n/a	n/a	Activates COL1A2 promoter, mediates TGFβ activation, Involved in wnt signalling.	KO mousedisplay increased osteoclast cell number

Table 4.16 Table illustrating the 7 variants prioritised for segregation analysis in family D

The DNA change column includes the transcript ID from the Ensembl database (The Ensembl ID stated is in all cases the major transcript ID). The minor allele frequency data stated is based on data from the 1000 Genomes project European population and the NHLBI cohort European American population. The Polyphen prediction score is a rating from 0-1 where 0 indicates variants predicted to be benign and 1 indicates variants that are very likely to be damaging.

In this family, all seven variants were identified in all three affected family members; IV:2, V:1 and V:III through Sanger sequencing. However, the c.2171C>T variant in *SHANK2* was found to be homozygous in unaffected family member V:2 and heterozygous in IV:1, so was eliminated from further analysis. In addition, variant c.1384G>A in *FOXK1* was found to be heterozygous in just one unaffected family member V:2. Although due to variable penetrance, the presence of a variant in just one unaffected family member would not usually be sufficient to rule out the variant as being disease causing, an exception has been made in this case. This is because due to the fact that this family has numerous consanguineous marriages, it can be assumed that family members would be more likely to share a greater proportion of their variants than families with no consanguinity. This would likely mean that any variants within additional modifier genes that may be involved in influencing disease penetrance would be more likely to be shared amongst family members than in other families. Therefore incomplete penetrance is less likely to be a big concern than in non-consanguineous families. For this reason, the variant in *FOXK1* was not prioritised for genotyping in the remainder of the cohort, despite its presence in just one unaffected family member (Table 4.17).

Since genomic DNA was only ascertained for five members of family D, segregation analysis was less informative than in the other three families. For this reason and due to time and cost limitations, only the two most interesting variants were prioritised in this family for genotyping the remainder of the cohort, with a view of returning to this stage if no confirmatory mutations emerged from the analysis at a later date. The choice of variants to follow up is discussed in section 4.10.1. below.



Pedigree of Family D

square=male; circle=female

slash=deceased individual

double line=consanguineous marriage

open symbol=clinically unaffected

solid symbol=hearing impairment

black symbol=confirmed otosclerosis

dark grey symbol=presumptive otosclerosis

light grey symbol=other hearing loss

arrow=proband

Gene	Variant	Individual					Segregates in family D	Genotyping method	Frequency of allele in otosclerosis cohort	Frequency of allele in EUR population
		IV:1	IV:2	V:1	V:2	V:3				
		Presbycusis	Confirmed Otosclerosis	Confirmed Otosclerosis	Unaffected	Confirmed Otosclerosis				
		Age 75	Age 81	Age 43	4th decade	Age 46				
COL1A2	**808G>A**	Wt (G/G)	Het (G/A)	Het (G/A)	Wt (G/G)	Het (G/A)	Yes	Sanger sequencing	0.0088	0.003
FOXK1	**1384G>A**	Wt (G/G)	Het (G/A)	Het (G/A)	Het (G/A)	Het (G/A)	Yes with exception of one unaffected	Sanger sequencing	Genotyping not performed	Previously unreported
FZD2	**655C>T**	Wt (C/C)	Het (C/T)	Het (C/T)	Wt (C/C)	Het (C/T)	Yes	Sanger sequencing	Genotyping not performed	0.0001
GNGT1	**82C>A**	Wt (C/C)	Het (C/A)	Het (C/A)	Wt (C/C)	Het (C/A)	Yes	Sanger sequencing	Genotyping not performed	Previously unreported
mir-183	**81G>T**	Wt (G/G)	Het (G/T)	Het (G/T)	Wt (G/T)	Het (G/T)	Yes	RFLP- AvaII	0.0088	0.0002
SHANK2*	**2171C>T**	Het (C/T)	Het (C/T)	Mut (T/T)	Mut (T/T)	Het (C/T)	No	Sanger sequencing	Genotyping not performed	0.0002
ZNF225	**1698A>T**	Wt (A/A)	Het (A/T)	Het (A/T)	Wt (A/A)	Het (A/T)	Yes	Sanger sequencing	Genotyping not performed	0.0009

Table 4.17 Table showing genotypes of all members of family D

The frequency in the general population is based on 1000 Genomes Project data.

*Genotyping of *SHANK2* was performed by Kara Forsyth

4.10.1 Discussion of which segregating variants in family D to prioritise for follow up analysis

Although it is possible that any of the segregating variants in family D may contribute to disease pathogenesis, two variants stood out as being the most likely candidates in this family. These are c.81G>T in *mir-183* and c.808G>A in *COL1A2*, which were selected for follow up analysis as both are located in particularly good otosclerosis candidate genes.

COL1A2 is a very large gene, 5411 bases in length and coding for a 1366 amino acid protein 129kDA in size. SNP c.808G>A in exon 17 of the *COL1A2* gene on chromosome 7q21 encodes a valine to isoleucine amino acid change at residue 270. It is a fibril-forming collagen that is abundant in bone and has been suggested as an otosclerosis candidate gene in the past. Association studies have been performed in an attempt to associate SNPs within this gene with otosclerosis but none have been identified. 474 unique variants have been reported in this gene, 44 of which are located in exon 17 in which the family D SNP was identified. Many of these variants have been implicated in various connective tissue disorders including both Osteogenesis Imperfecta and Ehlers Danlos syndrome. Since family D appears to exhibit symptoms of a systemic connective tissue disorder, *COL1A2* is an excellent candidate in this family. In addition to the variant in *COL1A2*, the three affected members of Family D also have a nonsynonymous variant in *COL21A1*, a member of the FACIT (Fibril Associated Collagens with Interrupted Triple helices) family of collagens. This SNP c.1748G>C is fairly uncommon, with a minor allele frequency (MAF) in the European population (1000 Genomes Project cohort) of 0.024. However this variant did not quite meet the criteria required to prioritise it for segregation analysis as variants with a MAF greater than 0.02 were not prioritised. FACIT collagens localise to tissues containing type 1 collagens and are known to be found in extracellular matrix component of blood vessel walls (Chou and Li, 2002). There are no *COL21A1* orthologues in rodents (Fitzgerald and Bateman, 2004), possibly due to its location near to evolutionary breakpoints. It is possible that this SNP in combination with SNP c.808G>A in family D could result in the otosclerosis phenotype. In addition, SNP c.808G>A is located within a PPARγ binding site. Since PPARγ plays a role in the regulation of bone turnover, it is possible that a SNP located within a binding site for this gene could interfere with the pathway involved in bone growth and maintenance. This suggests a potential mechanism by which SNP c.808G>A could cause aberrant bone growth around the stapes, leading to otosclerosis.

mir-183 is a non-coding microRNA, a short molecule with a role in transcriptional and post-transcriptional regulation of gene expression. *mir-183* is a member of a family of microRNAs including *mir-96* and *mir-182* which are known to regulate chloride intracellular channel 5

expression in inner ear hair cells (Gu et al., 2013). Overexpression of members of this gene family induce duplicated otocysts and extra hair cells. In contrast, knockdown of *mir-183* causes reduced numbers of hair cells in the inner ear and defects in the semi-circular canals (Li et al., 2010). Family member *mir-96* has been implicated in familial sensorineural deafness (Solda et al., 2012) and mutations in this gene in mice are associated with progressive hearing loss (Lewis et al., 2009). This suggests a potential link between this family of microRNAs and hearing loss which could explain a role for SNP c.81G>T in the disease process. Having said this, *mir-183* and its family members appear to be involved in sensorineural rather than conductive hearing loss when mutated. Although otosclerosis causes conductive hearing loss, approximately 10% of clinical cases also have a sensorineural component to their deafness (Browning and Gatehouse, 1984), it is therefore possible that the SNP in *mir-183* plays a role in this. Since the proband of family D was recruited as a non-surgical participant, access was not available to her audiometric data to investigate whether or not she had mixed or solely conductive hearing loss. If sensorineural hearing loss is occurring in this family, it is possible that SNP c.81G>T in *mir-183* may be involved.

The biological roles of *COL1A2* and *mir-183* with respect to bone development and hearing loss respectively, suggest that variants within these genes are the most promising candidate variants identified within this family. These variants were therefore genotyped in a cohort of 56 additional probands. The variant in *mir-183* was genotyped using RFLP genotyping with AvaII (see Appendix), however, this variant was not found to be present in any additional members of the cohort, resulting in a cohort frequency of 0.008 (Table 4.17). The variant in *COL1A2* was genotyped using Sanger sequencing, as no appropriate restriction endonuclease for RFLP genotyping was available, but again none of the 56 probands possessed this nonsynonymous G>A SNP (Figure 4.9 and Table 4.17).

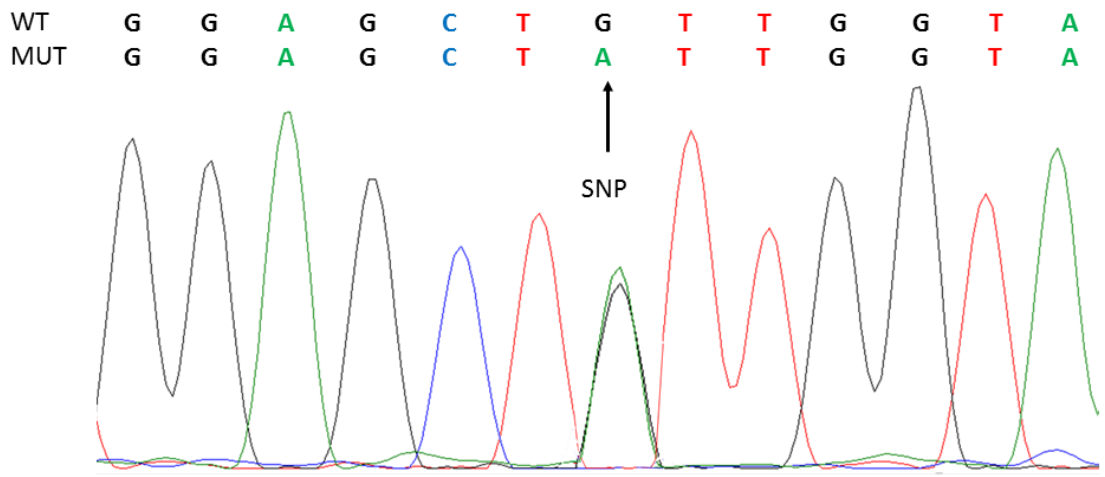


Figure 4.9 Chromatogram of a heterozygote for variant c.808G>A in COL1A2

Chromatogram of the genotype of the proband in family D showing the heterozygous G>A SNP in exon 17 of the COL1A2 gene.

4.10.2 Future work for family D

In family D, a large number of good candidate genes emerged from the step-wise prioritisation and filtering processes, but since the number of family members recruited was small, segregation analysis was not particularly informative. The reason that this family was selected for WES despite the small number of family members recruited, was because it was thought that this family had the potential to yield promising results due to its large size and due to the additional connective tissue disorder symptoms experienced by multiple family members. It was expected that larger numbers of family members could be recruited in the long term than in the other families and it was hoped that by prioritising variants in genes known to be involved in other connective tissue disorders, that it would assist in the identification of a disease causing variant. However, acquiring these additional DNA samples has proven more difficult in reality. In the future, it would be extremely useful to continue to attempt to recruit other family members, to help prioritise which of the segregating variants to follow up. This could provide further evidence to indicate whether it would be beneficial to continue to pursue the segregating variants in *mir-183* and *COL1A2*, or whether to return to the three variants in *FZD2, GNGT1* and *ZNF225* that had also segregated with the disease but were not prioritised for genotyping in the remainder of the cohort. Due to previously published investigations suggesting that *mir-183* is involved in hearing loss and that collagen type I genes are associated with otosclerosis, both the c.81G>T SNP in *mir*-183 and the c.808G>A SNP in *COL1A2* were further investigated as the top priority candidates for this family. Analysis of *mir*-183 and *COL1A2* are discussed in more detail in Chapter 4 Part II.

Whole Exome Sequencing Part II

Analysis of prioritised variants that segregate with familial otosclerosis

During WES followed by a stepwise variant prioritisation process, four variants within strong otosclerosis candidate genes were identified in three of the four families analysed. In family B SNP c.601G>A was identified in *SERPINF1,* in family C variant c.915C>G was found in *TRIM17* and in family D variant c.81G>T was identified in *mir-183* as well as SNP c.808G>A in *COL1A2*. In Part II of this chapter, further analyses which was performed on each of these four candidate variants will be discussed in turn.

4.11 Evaluation of *TRIM17* as an otosclerosis candidate gene in family C

In family C, variant c.915C>G in *TRIM17* was found to segregate with otosclerosis. This variant results in the introduction of a stop codon at amino acid 305 of *TRIM17,* suggesting that the *TRIM17* transcript in affected family members may be truncated upstream of its SPRY domain, a protein interaction module which functions in immune system related processes. This is the only variant to have survived both candidate gene and non-candidate gene filtering strategies following WES. There was therefore a strong rationale for genotyping the whole of exon 6, in which the c.915C>G variant was located, in a familial otosclerosis cohort.

4.11.1 Genotyping *TRIM17* exon 6 in a familial otosclerosis cohort

Sanger sequencing was performed on the remainder of exon 6 of *TRIM17* in a familial otosclerosis cohort of 56 unrelated probands. In addition to variant c.915C>G, which was confirmed in the proband of family C using Sanger sequencing (Figure 4.10), an additional rare variant was identified in one member of the cohort. This c.1233C>T SNP is synonymous, encoding amino acid leucine 411 and is present in the European cohort at a very low minor allele frequency estimated from the NHLBI cohort (http://www.nhlbi.nih.gov/) to be approximately 0.001. Although it is rare, the fact that it does not cause an amino acid change and since it is not located at a splice site, suggests that it is unlikely to be a variant that

contributes to disease pathogenesis, as it would be expected that a disease causing variant would affect the structure and as a result the function of the protein. This SNP was therefore considered unlikely to be a disease-causing variant.

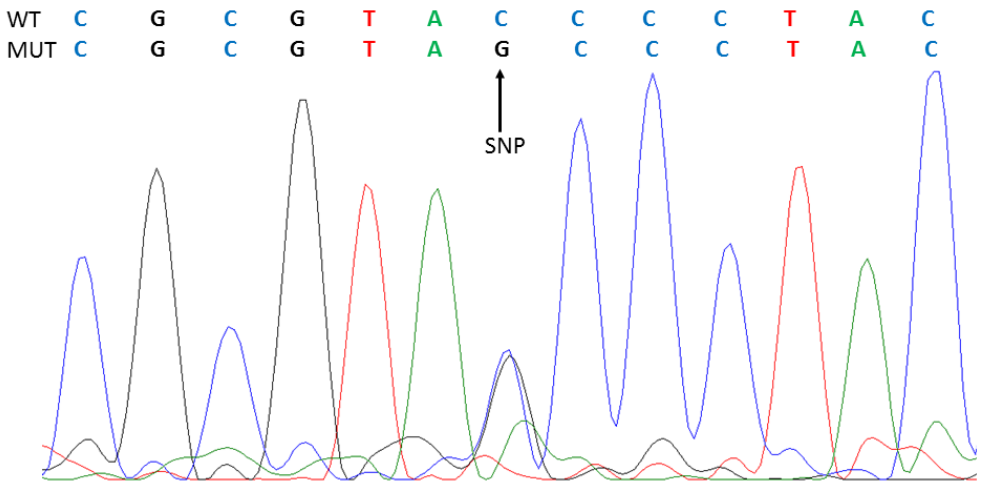


Figure 4.10 Chromatogram of a heterozygote for variant c.915C>G in *TRIM17*

Chromatogram of the genotype of the proband in family C illustrating the heterozygous C>G SNP in exon 6 of the *TRIM17* gene.

4.11.2 Analysis of RNA-sequencing data to investigate *TRIM17* expression in stapes

Analysis of the RNA sequencing data (chapter 5) suggests that *TRIM17* is expressed within human stapes but at fairly low levels with an average RPKM of 0.017 across all suprastructures. This is substantially lower than the average RPKM value of 5.4 across all genes sequenced in all stapes. Despite the low RPKM value, a difference in average RPKM was identified between the control and affected stapes, suggesting that the expression of *TRIM17* is lower in otosclerotic stapes compared to controls, although this doesn't quite reach significance thresholds at p=0.06. However, the fold difference in RPKM value between control and affected stapes is fairly large, with *TRIM17* being 2.9 times down-regulated in otosclerotic stapes compared to controls.

4.11.3 Discussion of *TRIM17* as an otosclerosis candidate gene

*T*he c.915C>G variant in *TRIM17* appears to be the most promising candidate from family C. Not only is this a previously unreported variant that introduces a premature stop codon, but it was also the only gene across all four families to survive both candidate gene and non-candidate gene filtering processes following WES. The fact that no other good candidate genes emerged from family C also gives further support to suggest that this variant may be involved in otosclerosis pathogenesis. Despite this, more evidence would be required to indicate its involvement in the pathophysiology of otosclerosis. For this, the identification of additional rare nonsynonymous variants in *TRIM17* within otosclerotic families would be essential.

The fact that *TRIM17* is expressed at low levels within the human stapes raises questions as to its likely involvement in otosclerosis, as it is likely that any gene involved in otosclerosis will be expressed in human stapes tissue. Despite this, the level of expression of a gene will depend on its role. For example, a structural protein such as collagen would be expected to be highly expressed in bone tissue, whilst a transcription factor would be likely to be expressed at a lower level. Since the exact function of *TRIM17* is unknown, it is not possible to predict whether it would be expected that this gene would be highly expressed in stapes or not. Therefore the low level of expression of this gene cannot be considered sufficient evidence to rule out the c.915C>G SNP as a possible disease causing variant in family C. Furthermore, since *TRIM17* appears to be 2.9 times down-regulated in affected stapes compared to controls, although this does not quite reach statistical significance, a difference in expression in otosclerotic stapes could indicate involvement of the gene in disease pathology.

To further investigate a role for *TRIM17* in otosclerosis, the remaining exons of the gene could be sequenced in the otosclerosis cohort. Since *TRIM17* is an excellent candidate gene and since no other promising segregating candidates emerged in family C, it would be recommended to focus efforts on this gene rather than to select other lower priority genes for segregation analysis. Since *TRIM17* contains five additional protein coding exons, Sanger sequencing would be fairly costly, but it would be crucial in order to identify if any other rare nonsynonymous variants were located within the gene. In the absence of identification of mutations within multiple families, there is yet insufficient evidence of a role for this gene in disease pathogenesis.

4.12 Evaluation of *mir-183* and *COL1A2* as candidate genes in family D

In family D, variant c.81G>T in *mir-183* and SNP c.808G>A in *COL1A2* were found to segregate with otosclerosis. *mir-183* is a small non-coding micro RNA that is a member of a gene family thought to play a role in sensorineural hearing loss. *COL1A2* encodes a fibril-forming type I collagen that has previously been investigated for associations with otosclerosis, although no association between this gene and the condition has been confirmed. Mutations in *COL1A2* are known to be involved in connective tissue disorders, meaning that this variant is an excellent candidate in members of family D, who have a range of symptoms suggestive of a connective tissue disorder. The biological roles of both these candidate genes provides a strong argument for performing further analyses on the variants found within these genes in family D. However, due to time and cost limitations, it was not possible to perform Sanger sequencing in the otosclerosis cohort for both genes. Due to previous association studies which have implicated type I collagen genes in otosclerosis, *COL1A2* was prioritised for genotyping over *mir-183*, whilst computational analysis of the secondary RNA structure of *mir-183* was also performed.

4.12.1 Genotyping *COL1A2* exon 17 in a familial otosclerosis cohort

The sequencing data from exon 17 of the *COL1A2* gene, which was obtained during Sanger sequencing in the 56 unrelated familial probands, was analysed for additional variants within this exon. An additional coding change was identified in one individual in the cohort. This was a previously unreported 9 base pair duplication c.817_825dupGCTGGTCCT in exon 17 of *COL1A2*, (Figure 4.11) just 9 bases downstream of the c.808G>A SNP found in family D. This duplication encodes the insertion of three additional amino acids: alanine 273 –glycine 274– proline 275. The study participant possessing this duplication was contacted in an attempt to recruit additional family members for segregation analysis. Genomic DNA samples were obtained from her mother who also had a diagnosis of otosclerosis and her father who had no family history of hearing loss. Sanger sequencing showed that her mother was also heterozygous for this duplication and that her father was wild type.

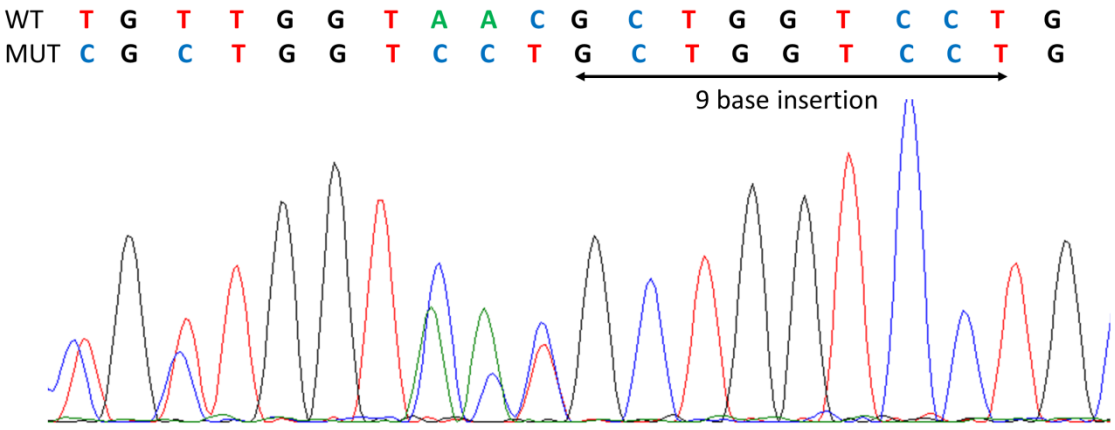


Figure 4.11 Chromatogram of a heterozygote for variant c.817_825dupGCTGGTCCT in *COL1A2*

Chromatogram of the genotype of the additional cohort member who is heterozygous for a nine base GCTGGTCCT duplication in exon 17 of the COL1A2 gene (Reverse complement of original sequence with reverse primer shown for clarity).

4.12.2 Predicting the effect of SNP c.81G>T on the secondary structure of *mir-183*

Variant c.81G>T in *mir-183* in family D is located at nucleotide 81 of the 150 bases that code for the pre-microRNA hairpin (*mir-183*). The functional portion of this non-coding RNA is the highly conserved mature-microRNA sequence (*miR-183*), which is encoded by nucleotides 27-48 of the hairpin structure and includes the seed sequence, the short region at its 5' end, which is thought to play a role in target specificity. Variant c.81G>T is not located within this mature RNA sequence but is located within the conserved miRNA* region, which is the complementary region on the opposite arm of the hairpin that binds to the mature RNA sequence during hair pin folding. It is thought that in some miRNAs, the mature sequence can be derived from either arm of the hairpin. If this were the case in *mir-183* and the mature RNA was derived from the miRNA* region, it is possible that variant c.81G>T could affect the structure of the mature RNA produced, which in turn could have an impact on the regulation of *mir-183* target genes.

Alternatively, due to the precise secondary structure of microRNAs, it is possible that this variant could result in changes to the hairpin structure which could affect normal processing of the mature RNA. RNAfold software (http://bibiserv.techfak.uni-bielefeld.de/rnafold/) was therefore used to predict the effect of the c.81G>T variant on folding of *mir-183* hairpin according to two models; the conventionally used Minimum Free Energy (MFE) model which predicts the structure of the RNA molecule using an algorithm based on the assumption that

the molecule will fold into the least energy state, and the more recently developed Centroid model, which is thought to more accurately predict secondary structure by taking into account known reference structures (Hajiaghayi et al., 2012). The figure below shows the predicted secondary structure of the pre-microRNA *mir-183* both with and without variant c.81G>T, according to the MFE and Centroid models (Figure 4.12).

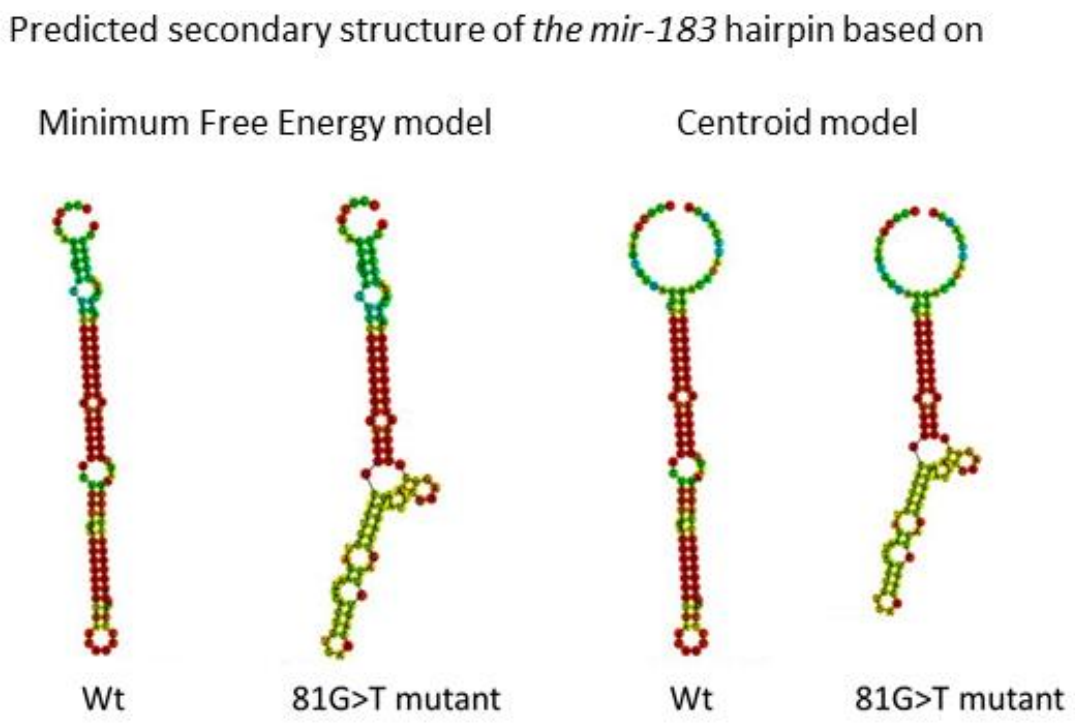


Figure 4.12 Secondary structure predictions for pre-microRNA *mir-183*

Images of the predicted secondary structure of the pre-microRNA *mir-183* both with and without variant c.81G>T according to the MFE and Centroid models using RNAfold software (http://bibiserv.techfak.uni-bielefeld.de/rnafold/).

4.12.3 Discussion of *mir-183* and *COL1A2* as candidate genes in family D

Both *mir-183* and in *COL1A2* are good candidate genes in family D. Computational predictions that suggest that variant c.81G>T in *mir-183* may affect the secondary structure of the pre-microRNA, indicates that this variant could have an impact on the processing of this pre-microRNA into the mature micro-RNA. Given that mutations within mature micro-RNAs in this gene family are known to play a role in hearing loss, it is possible that abnormal processing of the mature micro-RNA could result in a similar effect on target genes, leading to deafness. Whilst SNP c.81G>T *in mir-183* appears to be a promising candidate variant in family D, more evidence would be required to suggest its involvement in the pathogenesis of otosclerosis including identification of additional variants throughout the gene within otosclerotic families.

In *COL1A2*, an additional rare variant was identified in exon 17, the same exon in which the family D c.808G>A SNP was identified. This c.817_825dupGCTGGTCCT insertion, which was identified in an additional cohort member and her mother, is a novel duplication that has not been previously reported in the literature. However a rare variant, c.821G>A/T, has been reported within this 9 base pair region resulting in a p.glycine>aspartic acid/valine change at residue 274 in individuals with Osteogenesis Imperfecta (Ward et al., 2001, Shaheen et al., 2012, Marini et al., 2007). In addition, another rare 9 base pair duplication c.825_833dupGCTGGTCCC has been reported immediately downstream of c.817_825dupGCTGGTCCT, resulting in the 3 amino acid insertion alanine- glycine – proline in a further patient with Osteogenesis Imperfecta (Gentile et al., 2012). Due to the repetitive nature of the *COL1A2* gene, this nine base duplication c.825_833dupGCTGGTCCC results in the exact same amino acid sequence for *COL1A2* as the duplication c.817_825dupGCTGGTCCT found within the member of the otosclerosis cohort.

The proband possessing duplication c.817_825dupGCTGGTCCT is a 30 year old female with bilateral hearing loss which begun in the second decade of life. She does not have a diagnosis of Osteogenesis Imperfecta or any other connective tissue disorder besides otosclerosis, but has reported to have broken three bones in her life time; her collar bone at age three and a finger and toe in her late twenties. This may suggest that the duplication in *COL1A2* may be having an effect on her bones, causing her to be more susceptible to bone damage and breakages. Her 60 year old mother also has bilateral hearing loss that begun in the second decade. She has never broken a bone but has reported chronic neutropenia, an autoimmune condition associated with various collagen vascular disorders. Given that c.817_825dupGCTGGTCCT and c.825_833dupGCTGGTCCC result in the same amino acid sequence in the encoded protein, it is surprising that the family with otosclerosis possessing the c.817_825dupGCTGGTCCT duplication do not have Osteogenesis Imperfecta, whereas the individual with duplication c.825_833dupGCTGGTCCC does. One explanation for this is that perhaps the individual with the c.825_833dupGCTGGTCCC duplication identified by Gentile et al. also possesses an additional Osteogenesis Imperfecta causing variant within the *COL1A2* gene or within another Osteogenesis Imperfecta gene that was not identified. Another explanation could be that the otosclerosis family with the c.817_825dupGCTGGTCCT duplication have a mild form of Osteogenesis Imperfecta with associated stapes fixation which is undiagnosed. If this is the case, it is suggestive that there is a shared genetic aetiology between otosclerosis and Osteogenesis Imperfecta. To investigate the link between these two conditions further, the Italian research group reporting duplication c.825_833dupGCTGGTCCC was contacted in an attempt to gather phenotypic information in regards to hearing loss from

the individual with this duplication; however, the research group were not able to provide us with this information. If this individual did have conductive hearing loss, it could support the theory that otosclerosis is a localised manifestation of Osteogenesis Imperfecta, as has previously been suggested in the literature (Weber, 1930). Although the subsequent identification of an additional variant within *COL1A2* in an unrelated otosclerosis family, is consistent with a role for *COL1A2* in the disease process, there is still insufficient evidence to implicate this gene as causative in family D. Recruitment of additional members of family D would be required to provide more convincing data from segregation analysis.

Analysis of the RNA-sequencing data (chapter 5) suggests that *COL1A2* may be a more promising candidate gene in family D than *mir-183*. Analysis of this data showed that that *mir-183* was not expressed in human stapes, whilst *COL1A2* was found to be highly expressed with an average RPKM of 6.65 across all stapes suprastructures. Since it is likely that any gene involved in otosclerosis will be expressed in human stapes tissue, this calls into question the role that *mir-183* could play in the disease process. However, given that microRNAs tend to be expressed at very low levels, it is possible that this gene is expressed in the stapes but at such low levels that it was not detected during RNA-sequencing. In contrast, since *COL1A2* encodes a structural protein, it would be expected to be highly expressed in bone tissue whether or not it was involved in the disease process. In addition, no significant difference in expression was found in *COL1A2* between otosclerotic and control samples, which might be expected of a gene involved in the disease process. Since no stapes were available from members of family D for gene expression analysis, it was not possible to investigate changes in expression of *mir-183* or *COL1A2* within the stapes of individuals who possessed these variants. This would have been useful in order to determine if these variants had an impact on expression levels of these genes, which could help to prioritise one of the candidates in this family.

For further investigations into causative variants in family D, Sanger sequencing could be performed in the remainder of the otosclerosis cohort of 56 probands on *mir-183*. Since this is a short gene, just 150 bases in length, this would be relatively inexpensive. This would be crucial in order to identify if any other rare nonsynonymous variants were located within the gene, as in the absence of identification of additional mutations within this gene in other families, it is difficult to support a role for *mir-183* in disease pathogenesis at this stage. For further analysis of *COL1A2*, it would not be possible to perform Sanger sequencing in the remainder of the gene in the otosclerosis cohort as it is an extremely large gene consisting of 52 exons so would not be financially viable. Targeted re-sequencing of specific regions of *COL1A2* may be a better option, however, with a lack of knowledge regarding disease

pathogenesis, it would be difficult to identify the regions of the gene that are most likely to contain mutations involved in otosclerosis. For this reason, it may be most beneficial to wait until a larger number of probands with familial otosclerosis have been recruited in order to perform sequencing on exon 17 in these individuals. If no additional mutations within this exon in other families are identified, it is difficult to provide enough evidence that strongly supports a role for *COL1A2* in disease pathogenesis.

4.13 Evaluation of *SERPINF1* as a family B candidate gene

In family B, SNP c.601G>A in *SERPINF1* was found to segregate with otosclerosis. This SNP results in an aspartic acid to asparagine amino acid change in exon 5 of the *SERPINF1* gene, a gene involved in the connective tissue disorder Osteogenesis Imperfecta type VI which is known to occasionally occur concomitantly with otosclerosis. The biological role of the gene, which is highly expressed in bone and cartilage, provides a strong justification to prioritise *SERPINF*1 for follow up analysis in a cohort of 56 familial otosclerosis probands.

4.13.1 Genotyping SERPINF1 exon 5 in a familial otosclerosis cohort

When the sequencing data from exon 5 of the *SERPINF1* gene, which was obtained during Sanger sequencing in the 56 unrelated familial probands, was analysed for additional variants within the exon, an additional SNP c.441G>C was identified at the second base of exon 5 in two unrelated probands in the cohort (Figure 4.13).

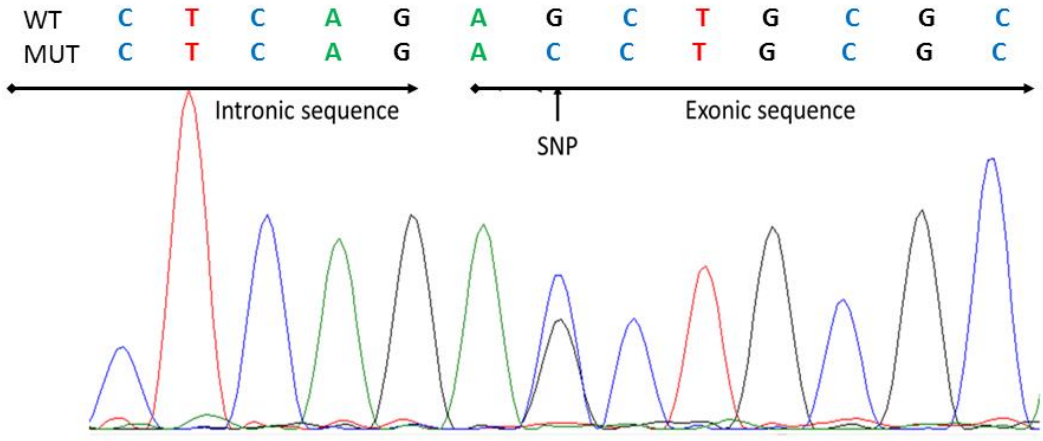


Figure 4.13 Chromatogram of a heterozygote for variant c.441G>C in *SERPINF1*

Chromatogram of the genotype of one of the two additional cohort members with the heterozygous G>C SNP at the second base of exon 5 in the *SERPINF1* gene.

Both probands possessing this rare variant were 57 year old British females with a strong family history of otosclerosis. Neither proband had any living relatives with otosclerosis so segregation analysis was not possible in either of these families. However, gDNA from an unaffected elderly maternal aunt was secured from one of these probands who had a strong history of otosclerosis in her maternal line. The variant was not found in this unaffected aunt who had a mother, sister and niece all with early onset hearing loss characteristic of otosclerosis. It was not possible to secure genomic DNA from any family members of the other proband with this variant. However this participant was recruited via the surgical recruitment path and has received stapedotomy surgery on both ears within the last three years at the Royal National Throat Nose and Ear Hospital in London. As a result, access was available to both her stapes bones for gene expression analysis (4.13.6).

SNP c.441G>C causes an amino acid change from lysine 147 to asparagine and is present in the general population at a very low frequency, estimated from data from the NHLBI cohort of 4,300 European American individuals (http://www.nhlbi.nih.gov/) to be approximately 0.0003, which is much lower than the frequency in the familial otosclerosis cohort of 0.0175. This variant is therefore 58 times more common in the otosclerosis cohort than in the general population. In addition, this variant was not identified when a cohort of 190 controls were genotyped by restriction endonuclease digest using TseI, confirming that this SNP is rare in the general population.

4.13.2 Genotyping the remainder of *SERPINF1* in a familial otosclerosis cohort

Following the identification of additional rare SNP c.441G>C in *SERPINF1* in two unrelated probands in the cohort, the remainder of the coding region (exons 2-8) of the *SERPINF1* gene was sequenced in 56 familial otosclerosis probands. In total 8 additional variants were identified in the cohort. A nonsynonymous SNP c.167C>G was identified in *SERPINF1* exon 3 within one of the 56 probands (Figure 4.15 A). This SNP causes an alanine to glycine change at amino acid 56 and is very rare in the general population with an allele frequency of just 0.0001 based on the European American data from the NHLBI cohort. Two other nonsynonymous SNPs were also identified in exon 3. These were c.202G>C and c.215C>T, both of which occurred in the otosclerosis cohort at a frequency similar to that of the general population based on data from the NHLBI cohort (http://www.nhlbi.nih.gov/). In addition, four variants were identified in exon 4 of *SERPINF1*. One of these was a previously unreported c.331G>A variant (Figure 4.15 B) which causes an amino acid change from aspartic acid 111 to

asparagine. The other three SNPs within exon 4 are located within a highly polymorphic region spanning nine bases that contains six known SNPs. One of these is a synonymous SNP c.389T>C which is very common in the general population according to data from the NHLBI cohort. Another is a nonsynonymous SNP c.392C>A (Figure 4.15 C) which causes an amino acid change from alanine 131 to aspartic acid and which is has an allele frequency of just 0.0008 in the NHLBI cohort. The third is a common nonsynonymous SNP c.395C>G which occurred in the otosclerosis cohort at a frequency similar to that of the general population as predicted by NHLBI. The final SNP was a common synonymous SNP c.963T>C within exon 7 which again occurred in the otosclerosis cohort at a frequency similar to general population frequencies. The location and frequency information on each of these SNPs in addition to the c.601G>A SNP in family B and the c.441G>C SNP also identified in the cohort are summarised in Figure 4.14 and Table 4.18.

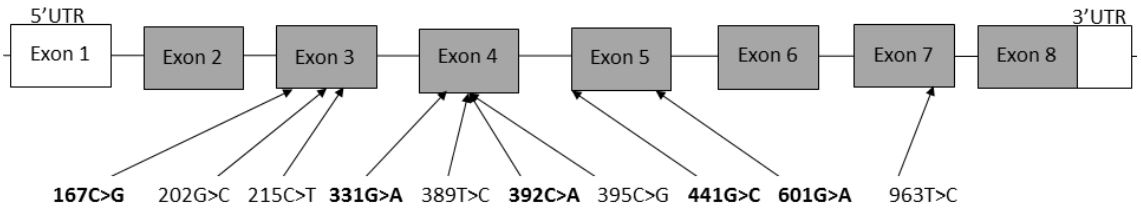


Figure 4.14 Illustration of variant locations in *SERPINF1*

Image illustrating the location of the 10 SNPs identified in the familial otosclerosis cohort in the *SERPINF1* gene. Rare SNPs in the general population are highlighted in bold.

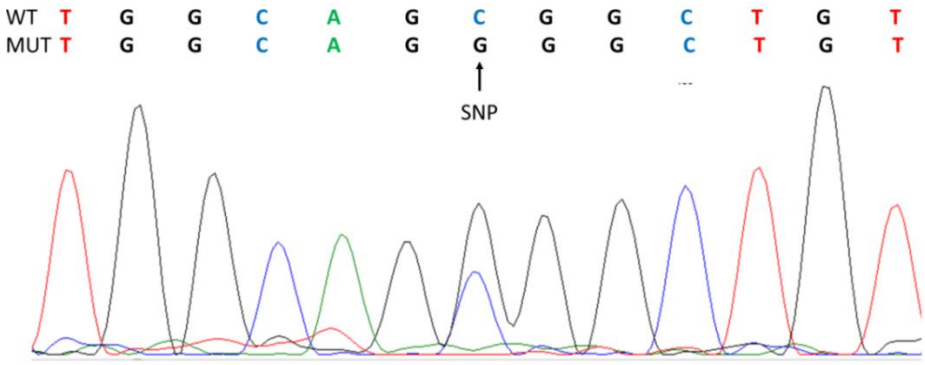
SNP	Exon	Frequency of allele in otosclerosis cohort	Frequency of allele in NHLBI cohort	Amino Acid change	Variant Prediction software
c.167C>G	3	0.0088	0.0001	56 A/G	Possibly damaging(0.428)
c.202G>C	3	0.0088	0.0036	68 V/L	Benign (0)
c.215C>T	3	0.6404	0.6519	72 T/M	Benign (0)
c.331G>A	4	0.0088	Previously unreported	111 D/N	Probably damaging(0.915)
c.389T>C	4	0.2807	0.2942	Synonymous	Synonymous
c.392C>A	4	0.0088	0.0008	131 A/D	Probably damaging(0.993)
c.395C>G	4	0.0526	0.0327	132 P/H	Probably damaging(0.929)
c.441G>C	5	0.0175	0.0003	147 K/N	Benign(0.129)
c.601G>A	5	0.0088	Previously unreported	201 D/N	Benign(0)
c.963T>C	7	0.2895	0.2921	Synonymous	Synonymous

Table 4.18 Table showing the 10 variants identified in *SERPINF1*in the otosclerosis cohort

Table shows all variants identified in *SERPINF1* ENST00000254722 *in* a cohort of 57 unrelated probands with familial otosclerosis. Rare SNPs in the general population are highlighted in bold. The minor allele frequency of the variant is based on data from the European American population in the NHLBI cohort. The Polyphen prediction score is a rating from 0-1 where 0 indicates variants predicted to be benign and 1 indicates variants that are very likely to be damaging.

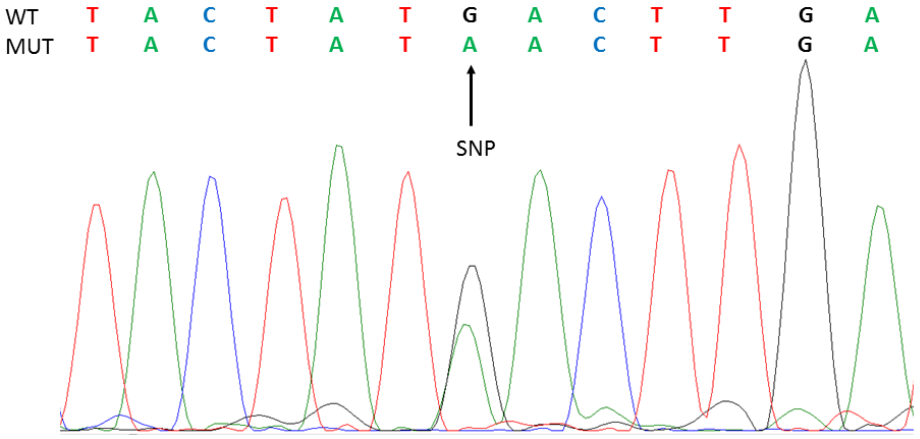
Please note that although the original *SERPINF1* SNP c.601G>A was identified in family B of Caribbean origin; here the frequencies of a European cohort are stated as this best represents the ethnicity of the otosclerosis cohort recruited.

A.



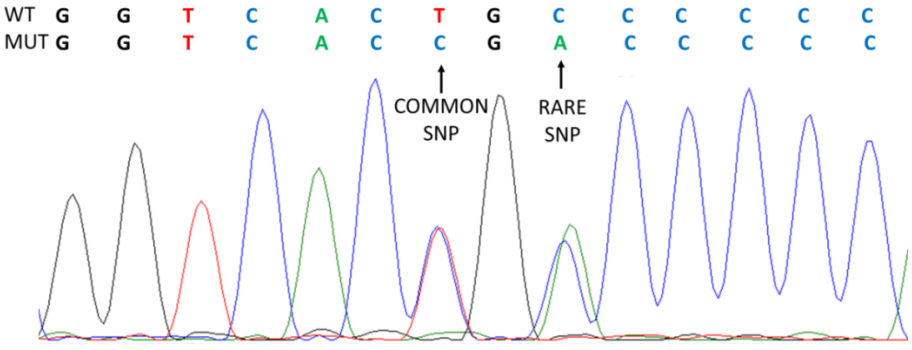
Chromatogram of a heterozygote for variant c.167C>G in *SERPINF1*

B.



Chromatogram of a heterozygote for variant c.331G>A in *SERPINF1*

C.



Chromatogram of a heterozygote for variant c.392C>A in SERPINF1

Figure 4.15 Chromatograms showing variants in *SERPINF1*

(A) Chromatogram of the genotype of the additional cohort member who has the heterozygous c.167C>G SNP in exon 3 of the *SERPINF1* gene. (B) Chromatogram of the genotype of the additional cohort member who has the heterozygous c.331G>A SNP in exon 4 of the SERPINF1 gene. (C) Chromatogram of the genotype of the additional cohort member who has the heterozygous c.392C>A SNP in exon 4 of the SERPINF1 gene. This individual also has common synonymous c.389T>C SNP in this region.

Of the eight additional SNPs identified, two were synonymous and six nonsynonymous. Since synonymous SNPs do not cause an amino acid change, it was decided that these were unlikely to be disease causing, so were eliminated from further analysis. Of the remaining six variants, three occurred at a frequency in the cohort similar to the frequency that would be expected in the general population based on data from the NHLBI cohort. It was therefore decided that these were also unlikely to be involved in the disease process as it would be expected that a disease causing variant would be over-represented in the otosclerosis cohort. Therefore three additional variants of interest were identified in the cohort; c.167C>G in exon 3, c.331G>A in exon 4 and c.392C>A also in exon 4. These variants were also predicted to be damaging according to Polyphen prediction software. It was not possible to recruit family members from any of the probands possessing these three SNPs for segregation analysis. However, two of these individuals, one with SNP c.167C>G in exon 3 in the other with SNP c.331G>A in exon 4 were recruited via the surgical recruitment path. As a result, access was available to their stapes bones for gene expression analysis (4.13.6). In total, including the two previously identified SNPs, five rare nonsynonymous variants of interest were identified within the coding region of *SERPINF1.*

4.13.3 Analysis of *SERPINF1* isoforms

SERPINF1 has 5 known protein coding, 3 putative protein coding and 4 non-protein coding transcripts (Figure 4.16). The nonsynonymous c.441G>C variant, which was identified in two unrelated otosclerosis probands, is located at the second nucleotide of exon 5 of the major protein-coding transcript *SERPINF1-001*. The location of this SNP suggests that it may play a role in altered expression of specific *SERPINF1* transcripts, as the 5' end of an internal exon typically acts as splice acceptor site, indicating that this SNP may be involved in splicing. The locations of the additional *SERPINF1* variants were also investigated, to identify if any were located at a splice sites within any of the 12 *SERPINF1* transcripts. Whilst three were not found to be located at a splice site, the Family B SNP c.601G>A was located at the final base of truncated exon 5 in non-protein coding transcript *SERPINF1-008* (Figure 4.16). This suggests that both exon 5 SNPs (c.441G>C and c.601G>A) could have an effect on splicing, and could thus impact the expression of certain isoforms. Furthermore, whilst the family B SNP c.601G>A is the final base in the terminal exon of non-protein coding transcript *SERPINF1-008*, SNP c.441G>C is located 160 bases upstream at the second base of this same exon. This suggests that the sites of variants c.441G>C and c.601G>A are likely to be significant for splicing of *SERPINF1*.

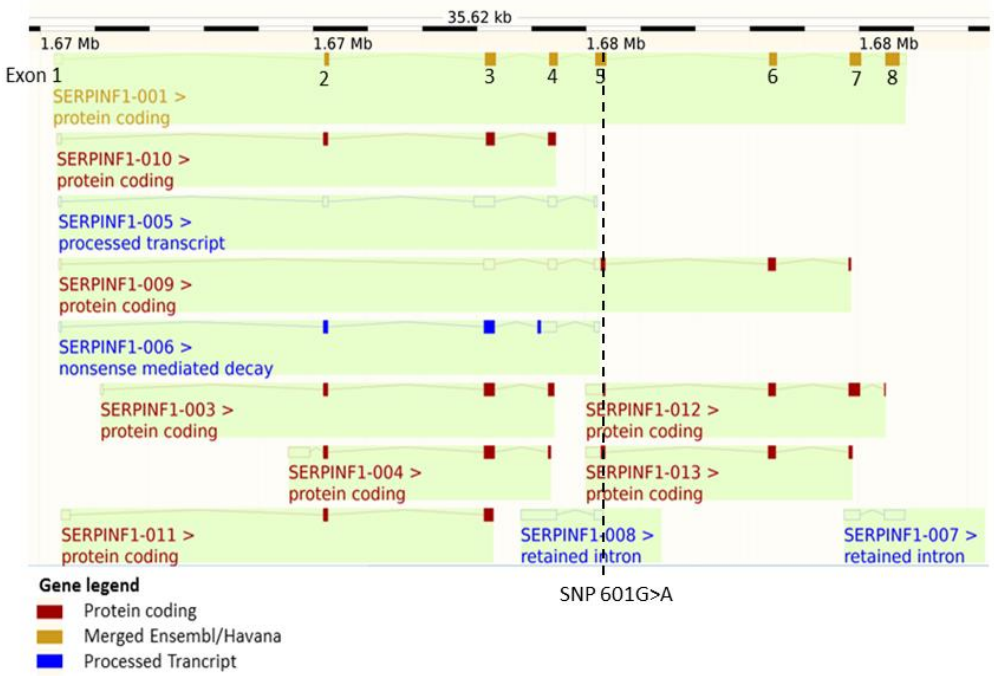


Figure 4.16 Image illustrating all 12 *SERPINF1* transcripts.

Transcripts 1, 3, 4, 10 and 11 are known to be protein coding, transcripts 9, 12 and 13 are putative, transcripts 5, 6, 7 and 8 are predicted to be non-protein coding. Exon numbers are labeled on SERPINF1 transcript 001. Location of family B SNP c.601G>A is shown (Image from www.ensembl.org)

There is also evidence that suggests that Family B variant c.601G>A could affect translation of putative transcript *SERPINF1-012*. This transcript is 729 bases in length and codes for a 138 amino acid protein spanning from exon 5 to exon 8. It encodes an untranslated region (UTR) at the 5' end of exon 5 where SNP c.601G>A is the final base of the 5'UTR, meaning that the SNP is located at the -1 position (one base upstream of the AUG start codon) in *SERPINF1-012*. It therefore lies within the Kozak sequence, a highly conserved element at the 5' end of a messenger RNA transcript which spans the translational initiation site. This sequence is essential for binding of the 43s ribosome complex at the initiation of translation. There are no other initiation signals within this exon, suggesting that if the variant in family B resulted in inhibition of ribosomal binding at this site, translation of the entire transcript could be compromised.

4.13.4 Analysis of RNA-sequencing data to investigate *SERPINF1* expression in human stapes

To investigate whether there were any differences in the expression of specific *SERPINF1* isoforms in the stapes tissue of the proband in family B who possessed SNP c.601G>A, RNA-sequencing data (chapter 5) was analysed in this individual. During RNA-sequencing, reads were mapped against a specialist reference transcriptome known as the spliceome, a sequence database consisting of sequences of all known splice isoforms, which enabled detection and relative quantification of specific isoforms within the samples. Since this tool generates a read count for each exon within each sample analysed, it is useful for comparing differential expression of specific isoforms between affected and control tissue. It was found that the read count in *SERPINF1* was greater for exons downstream of exon 5 than for those upstream. The average read count for exons 1 to 4 was just 18.67 compared to 78.58 for exons 5 to 8 in this individual. This finding was unexpected as it would be expected that if SNP c.601G>A did impair ribosome binding in the Kozak sequence of transcript *SERPINF1-012*, that it would result in reduced expression of this transcript and thus a lower read count for exons downstream of exon 5. However, when the read count for each exon of this gene was analysed in all 8 otosclerotic and 4 control samples sequenced, it was found that this pattern was consistent throughout all samples. On average the read count was greater for exons downstream of exon 5 than for upstream exons (Table 4.19).

Exon	Average read count		Read Count in Stapes from Family B Proband	p value
	Control stapes	Otosclerotic stapes		
exon 1	7.90	12.75	12.17	0.52
exon 2	11.65	17.33	14.17	0.55
exon 3	40.89	40.77	23.17	1.00
exon 4	39.66	40.28	25.16	0.98
exon 5	134.84	52.78	49.16	0.02
exon 6	135.33	42.16	45.16	0.02
exon 7	228.17	74.02	107.99	0.03
exon 8	293.77	64.75	112.00	0.01

Table 4.19 Table showing read count data for *SERPINF1* exons

Table shows the average read count of *SERPINF1* exons in the 8 otosclerotic and 4 control stapes and from the stapes in the proband of family B. p-values calculated using a Student's two-tailed paired-end t-test illustrate a significant difference in expression between affected and control stapes of exons 5-8 but not 1-4.

Although the read count for exons 1-4 was fairly consistent across all control and otosclerotic stapes, there was a marked reduction in the read count of downstream exons in otosclerotic stapes compared to controls. Moreover, when the expression levels both up and downstream of exon 5 in the cohort of the 8 otosclerotic and 4 control stapes samples were compared, it was found that there was a significant difference ($p<0.05$) in expression of *SERPINF1* exons 5 to 8, but this was not the case for exons 1 to 4 (Table 4.19).

Further analysis of the RNA-sequencing data found that the 5'UTRs unique to downstream transcripts *SERPINF1-012* and *SERPINF1-013* were detected in both otosclerotic and control human stapes. However, the level of expression of these UTRs was lower than for the coding region of exon 5. (Figure 4.17). In the stapes suprastructure of the proband of family B who possessed SNP c.601G>A, the difference in read count between the UTR and the coding region of exon 5 is larger than in the three control suprastructures (Figure 4.17), particularly when compared to control stapes 1 and 2.

4.13.5 Investigating alternative splicing in *SERPINF1*

Due to the evidence suggesting that a mechanism of alternative splicing in *SERPINF1* may be involved in otosclerosis, an *in silico* splicing assay was carried out on all rare nonsynonymous variants identified within *SERPINF1*. Spliceport (www.spliceport.cs.umd.edu) analysis was performed on both wild-type and mutant sequences surrounding each variant to calculate the probability of a splice donor or acceptor site within the sequence. Splice donor and acceptor sites are highly conserved sequences at which the spliceosome, a complex of ribosomal RNA and associated proteins, assembles during splicing. The 5' donor splice site is located at the upstream region of the intron being excised, whilst the 3' acceptor splice site is located at the downstream region of the intron (Figure 4.18). Using a feature generation algorithm (FGA) framework, the Spliceport software classifies all AG dinucleotides into acceptor and non-acceptor sites and all GT dinucleotides into donor and non-donor sites. When applied to the five variants identified in *SERPINF1,* four were not found to be located in the immediate region surrounding an AG or GT dinucleotide, so were not assigned a Spliceport computational score. However SNP c.441G>C results in a G to C change at the G of an AG dinucleotide immediately downstream of the donor site. A computational score of -1.05 was assigned to the wild-type G allele, which is greater than the threshold score of -10 implemented, above which any score is considered significant (sensitivity value=100%, AG false positive rate = 85.21%). However, when the analysis was run to include the C allele, no computational score was generated. This suggests that c.441G>C may affect the splice acceptor site at the 5' end of exon 5.

Figure 4.17 Graphical representation of read count in *SERPINF1* from RNA-seq data in human stapes

Read count data shown from the family B proband as well as three control stapes suprastructures, to show the relative levels of expression of the 5' UTR region of exon 5 that is unique to *SERPINF1* transcripts 12 and 13 compared to the protein coding region of exon 5. Screenshot from DNAnexus

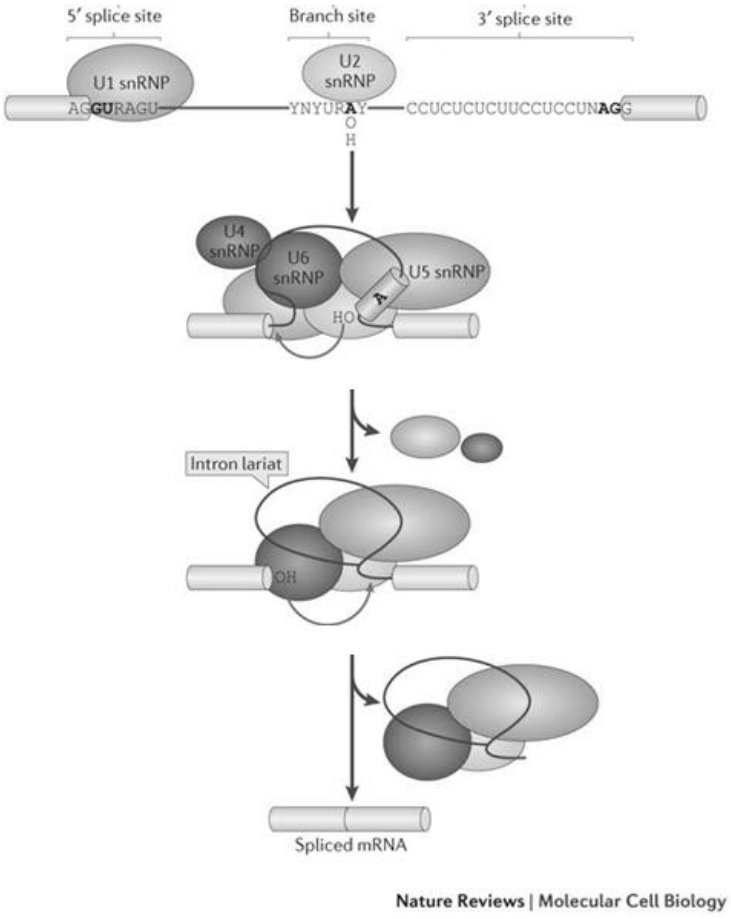


Figure 4.18 Image illustrating the process of splicing

Image shows binding of the U1 and U2 snRNPs to the 5' splice site and branch site respectively, assembly of the mature spliceosome, nucleophilic attack and lariat excision. (Adapted from Kornblihtt et al., 2013)

ESE finder (www.rulai.cshl.edu/tools/ESE) is a second splice prediction software that was used to predict whether the variants were located within an additional consensus sequence known as the branch site. This sequence is less strongly conserved than the donor and acceptor sites but contains an invariant adenosine residue approximately 18 nucleotides upstream of the acceptor site, followed by a polypyrimidine tract (Figure 4.18). This adenosine residue is recognised by a specialist component of the splicesome complex called the U2 snRNP (small nuclear ribonuclear protein) which plays a crucial role in lariat excision during the splicing together of adjacent exons (Kornblihtt et al., 2013).

This software predicted that the c.392C>A SNP in exon 4 would result in the introduction of a branch point adenosine residue within the conserved branch point sequence (Table 4.20). ESE finder also predicted that variant c.331G>A in exon 4, which had not been reported in previous literature and SNP c.601G>A in Family B may reduce efficiency of binding of the U2 snRNP to the branch point consensus sequence. However, there was a smaller difference in the branch

site threshold score generated by ESE finder between the wild type sequence and sequence containing the variant for both of these variants, than was the case for the 392 C>A SNP (Table 4.20). Neither SNP c.167C>G in exon 3 nor the c.441G>C SNP at the second base of exon 5 were predicted to be located within a branch point consensus sequence.

SNP	Exon	ESE Finder Branch site score	
		Wild-type sequence	Sequence containing variant
c.167 C>G	3	0	0
c.331 G>A	4	3.20	2.79
c.392 C>A	4	1.19	7.49
c.441G>C	5	0	0
c.601G>A	5	3.71	3.31

Table 4.20 Table showing predicted impact on splicing of *SERPINF1* variants

Table shows the 5 variants identified in *SERPINF1* and the predicted impact of each variant on splicing according to ESE finder software. Scores reflect the extent of matching to a degenerate consensus and are adjusted for background nucleotide composition. The default threshold setting used was 0. The scoring system for ESEfinder ranges from the default threshold signifying no effect, to positive values which predict an increasing probability that a variant is damaging to a branch site.

4.13.6 Investigating the relative expression of alternative *SERPINF1* isoforms

Since the RNA-sequencing data, as well as the results of genotyping followed by an *in silico* splicing assay, suggested that alternative splicing of *SERPINF1* may be occurring in otosclerotic stapes, quantitative PCR was performed in a large number of stapes in order to analyse relative expression levels of specific *SERPINF1* isoforms. Two Taqman® gene expression assays were performed to analyse relative expression levels of specific *SERPINF1* transcripts. One of these assays only amplified transcripts containing exons upstream of exon 5, whilst the other only amplified transcripts containing exons downstream of exon 5. The probe of the upstream assay (Hs011006934-Applied Biosystems ®) binds to the exon-exon junction between exons 3 and 4, so would be predicted to amplify *SERPINF1* transcripts 1, 3, 4, 5, 9 and 10. The probe of the downstream assay (Hs011006937-Applied Biosystems ®) binds to the exon-exon junction between exons 6 and 7 so would be predicted to amplify *SERPINF1* transcripts 1, 9, 12 and 13 (Figure 4.19).

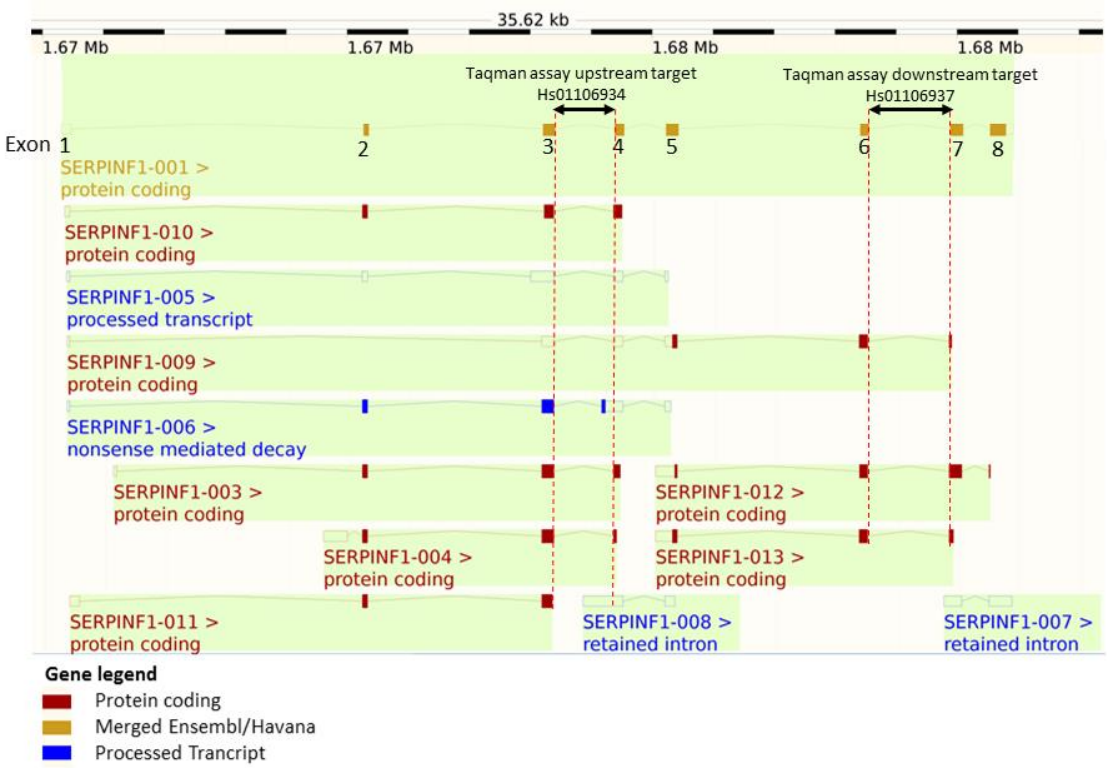


Figure 4.19 Image illustrating the location of the upstream and downstream Taqman assay probes for all *SERPINF1* transcripts.

The upstream probe Hs011006934 binds at the junction between exons 3 and 4. It can therefore detect transcripts 1, 3, 4, 5, 9 and 10. The downstream probe Hs011006937 binds at the junction between exons 6 and 7. It can therefore detect transcripts 1, 9, 12 and 13 (Image from www.ensembl.org)

Both upstream and downstream assays were performed on cDNA extracted from 81 otosclerotic and 5 control stapes suprastructures. Of the 81 otosclerotic stapes, three were from individuals known to possess variants within the *SERPINF1* gene that had been predicted to have an effect on alternative splicing. Of these individuals, one had the c.167C>G SNP in exon 3, one had the c.331G>A SNP in exon 4 and the other had the c.441G>C SNP in exon 5. The Taqman® assays were carried out in triplicate in each of the five controls and six replicates were performed for the three samples with known mutations in *SERPINF1*, to provide a better estimate of the true expression levels in these stapes, whilst the remaining 78 otosclerotic samples were performed once and analysed as one sample. *SERPINF1* expression levels were compared between otosclerotic and control stapes and in those individual patients with known variants. The design of the assays also allowed a comparison between levels of downstream and upstream exons in the same group to indicate whether there were differences in the transcripts expressed.

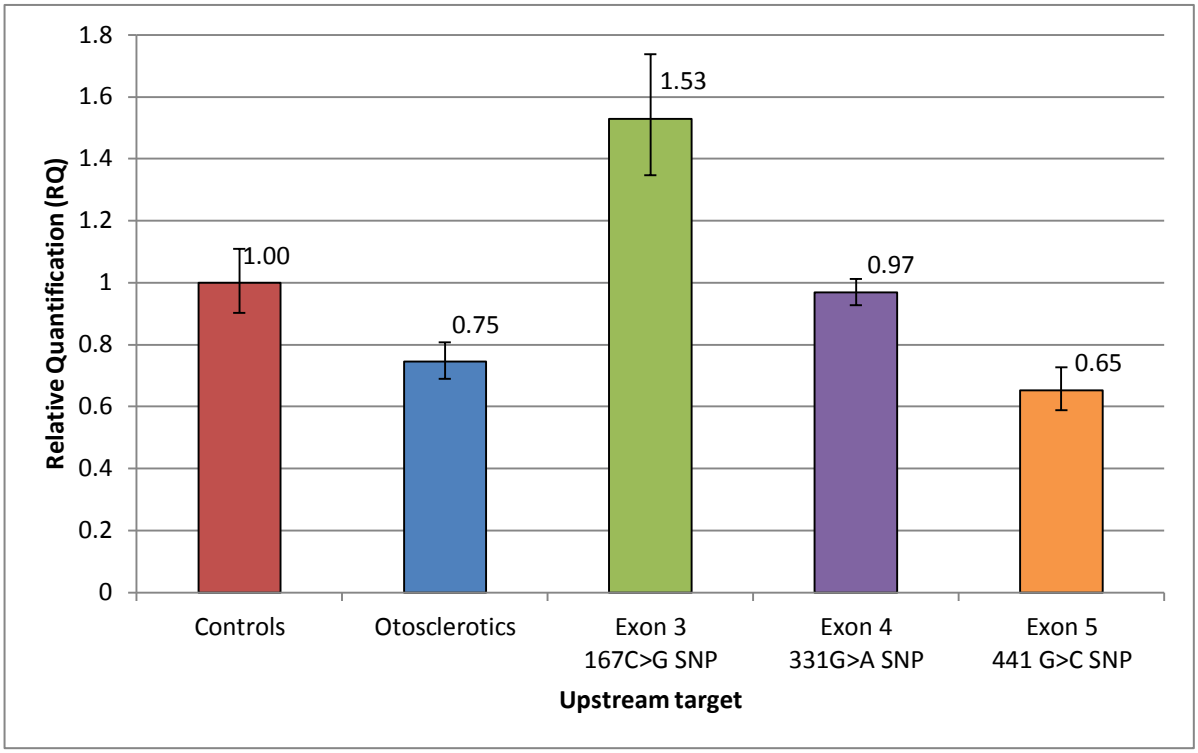
The results from the Taqman® assays are shown in Figure 4.20. These results show that *SERPINF1* expression is reduced in otosclerotic samples compared to controls, for both upstream and downstream exons. The degree of this reduction is significantly greater for the downstream assay (RQ=0.52) than for the upstream the upstream assay (RQ=0.75). The downstream assay indicates expression of *SERPINF1* in otosclerotic stapes is only half that of control stapes, whereas the upstream assay suggests expression in otosclerotic stapes at three quarters that of control stapes.

In the stapes of the individual with variant c.167C>G in exon 3, *SERPINF1* appeared to be expressed at higher levels than in the control stapes for both assays, although this only reached significance for the upstream assay. In the stapes of the individual with SNP c.331G>A in exon 4, *SERPINF1* appeared to be expressed at slightly lower levels than in the control stapes, although this did not reach significance for either assay. The stapes from individuals with both these variants appeared to have reduced expression of downstream compared to upstream exons relative to controls, however again this did not reach significance.

In the individual with SNP c.441G>C in exon 5, the SNP which was found in two cohort members and is known to be located near a splice acceptor site, *SERPINF1* was expressed at significantly lower levels than in the control stapes for both assays and was expressed at even lower levels than the otosclerotic stapes cohort as a whole. In addition, the downstream exons were expressed at lower levels than the upstream exons when compared to controls. This indicates that this variant may be affecting splicing of *SERPINF1* in this individual, causing a reduction in expression of *SERPINF1* and particularly of exons downstream of exon 5.

Since relative quantification rather than absolute quantification assays were performed, it was considered unsuitable to make a direct comparison between expression levels of the upstream and downstream assays within a sample category due to possible differences in assay efficiencies.

A



B

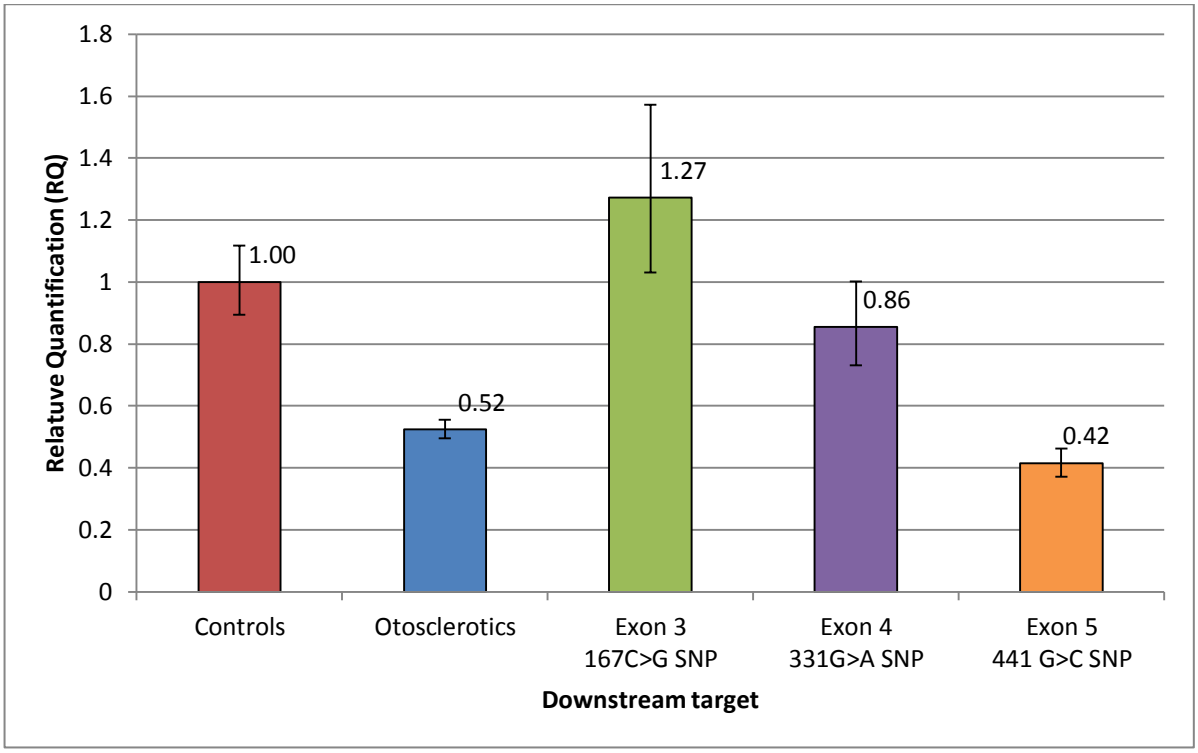


Figure 4.20 Graphs showing the relative expression of *SERPINF1* transcripts in otosclerotic and control stapes suprastructures

Assays were located at two targets within the gene, one at the exon 3-4 boundary (A) and one at the exon 6-7 boundary (B). 18s ribosomal RNA served as endogenous control. Error bars represent 95% confidence intervals.

4.13.7 A possible mechanism of alternative splicing in *SERPINF1* in otosclerosis pathophysiology

The data obtained from both RNA-sequencing assays indicate that there is increased expression of downstream compared to upstream exons in all human stapes. This suggests that alternatively spliced *SERPINF1* transcripts *SERPINF1-012* and/or *SERPINF1-013*, which encode the downstream region of the protein, are expressed in human stapes in addition to the major full length transcript *SERPINF1-001.* Further evidence to support this is the detection of the 5'UTR of transcripts *SERPINF1-012* and *SERPINF1-013* in both otosclerotic and control stapes during RNA-sequencing. Since this UTR sequence is unique to these transcripts, it is suggestive that transcripts *SERPINF1-012* and *SERPINF1-013* are expressed in the human stapes. In addition, since the level of expression of this 5'UTR region is reduced compared to that of the remainder of exon 5 in the human stapes, it indicates that other transcripts containing exon 5 that do not possess the UTR, such as major full length transcript *SERPINF1-001* are also expressed in the stapes. Due to the overlapping nature of transcripts *SERPINF1-012* and *SERPINF1-013*, it has not been possible to confirm which of these are expressed in human stapes, but the data indicates that it is extremely likely that either one or both of these transcripts are expressed.

RNA-sequencing and quantitative real time PCR assays also suggests that expression of *SERPINF1* is reduced in otosclerotic stapes compared to controls, in particular for exons downstream of exon 5. The most plausible explanation for this is that in individuals with otosclerosis, there is a reduction in expression of downstream transcripts *SERPINF1-012* and/or *SERPINF1-013* compared to controls. There are other possible theories, for example that there is an increase in expression of transcripts encoding upstream exons in otosclerotic stapes, which would support the relative reduction in expression of downstream exons within these stapes, but would not account for the difference in expression between control and affected stapes. Therefore a reduction in expression of transcripts encoding downstream exons seems to be the most likely explanation.

This theory is corroborated by the results of the gene expression assays in the individuals in whom rare nonsynonymous variants in *SERPINF1* were identified, as well as the results of the *in silico* splicing assay. In the stapes of the family B proband possessing the c.601G>A SNP, there is a reduction in expression of downstream exons compared to controls based on the results from the RNA-sequencing data. There is also a greater difference in read count between the 5'UTR of *SERPINF1-012* and *SERPINF1-013* and the coding region of exon 5 in the stapes of the family B proband than there is in control stapes, indicating a reduction in

expression of downstream transcripts. These observations could be explained by the fact that this SNP is located at the -1 position of the Kozak sequence of transcript *SERPINF1-012,* a sequence that is required for translation initiation. The Kozak consensus sequence of *SERPINF1-012* is weak, as it possesses a pyrimidine base at the -3 position. The presence of a pyrimidine rather than a purine at this position causes translation to become more sensitive to mutations located at the -1 position (Kozak, 1986). Therefore the G>A variant at -1 of *SERPINF1-012,* resulting in a change from a well conserved G base to a less well conserved A base, could lead to reduced binding affinity of the translational machinery and thus a reduction in expression of this transcript. Alternatively, based upon the results of the *in silico* splicing assay, it is possible that this SNP in family B could have a weak effect on binding of the U2 snRNP to the branch point consensus sequence, which would be predicted to inhibit recruitment of the splicesome at this site. Since the SNP is located at that start of *SERPINF1-012* and only 40 bases downstream of the *SERPINF1-013* start site, it is possible that it could reduce splicing at this site and thus lead to a reduction expression of downstream transcripts. This could therefore provide an alternative explanation as to how the G>A SNP in family B could result in the reduction in expression of downstream exons observed.

In the stapes from the individual with the c.167C>G variant in exon 3, the quantitative PCR assays found that expression of *SERPINF1* is greater than in control stapes for both up and downstream exons and that there is no significant difference in expression of downstream than upstream exons. This is consistent with the results of the *in silico* splicing assay which do not predict this SNP to be located within a splice acceptor, donor or branch point consensus site and is therefore not predicted to affect splicesome recruitment. Since this variant is located in exon 3, it would not be expected to affect splicing between exons 4 and 5 and is therefore unlikely to affect expression of *SERPINF1- 012* and *SERPINF1-013.* It is possible that this variant may have another effect on *SERPINF1* that is involved in the otosclerosis disease process, or may be non-casual. However, it must also be considered that there was greater variability in the expression of *SERPINF1* between replicates in this stapes compared to the other samples analysed, resulting in larger error bars than for other samples. More repeats of the assay could be performed to find out if there is any true difference in expression levels of this gene between this individual and control stapes.

In the stapes of the individual with variant c.331G>A in exon 4, there was reduced expression of downstream compared to upstream exons when compared to controls, based on findings from the quantitative PCR assays, although this was not significant. This trend is consistent with the results of the *in silico* splicing assay suggesting that like the variant found in family B,

variant c.331G>A may have a weak effect on binding of the U2 snRNP to the branch point consensus sequence, which could impact on expression of downstream exons. Since this effect is predicted to be weak, it is consistent with the lack of significance in the expression data. It would have been interesting to have had access to the stapes from the individual with variant c.392C>A in exon 4, as the *in silico* splicing assay predicted this variant to have a larger effect on splicing. However, since this individual was recruited through the non-surgical arm of the study, this was not possible.

Of greatest interest, are the results from the individual with variant c.441G>C in exon 5, a variant which was also found in a non-surgical study participant and which is predicted to be located near a splice acceptor site. The results from the quantitative PCR assay show that the reduction in expression of downstream exons is greater in this stapes than in both control stapes and in other otosclerotic stapes. There is therefore convincing evidence that the presence of this SNP directly affects splicing of *SERPINF1* and thus causes a reduction in expression of downstream exons.

It is surprising that the reduction in expression of downstream relative to upstream exons is not only seen in individuals with known variants in *SERPINF1*, but also in the larger cohort of 78 otosclerotic stapes when compared to controls. Whilst 46% of these otosclerotic stapes had reduced expression of upstream exons compared to controls, 88% had reduced expression of downstream exons compared to controls. Although it is possible that some of these 78 individuals possess rare nonsynonymous variants in *SERPINF1,* Sanger sequencing was performed on the genomic DNA of eight of these individuals, in whom no variants were identified. This therefore suggests that a reduction in expression of downstream exons also occurs in individuals who do not possess variants in *SERPINF1*. There is no obvious explanation for this finding, but it may be due to an interacting pathway of molecular factors that result in a pathogenic shift in expression away from the downstream isoforms. A hypothetical model of *SERPINF1* isoform expression in normal human stapes and otosclerotic human stapes both in the presence and absence of known *SERPINF1* variants is illustrated in Figure 4.21.

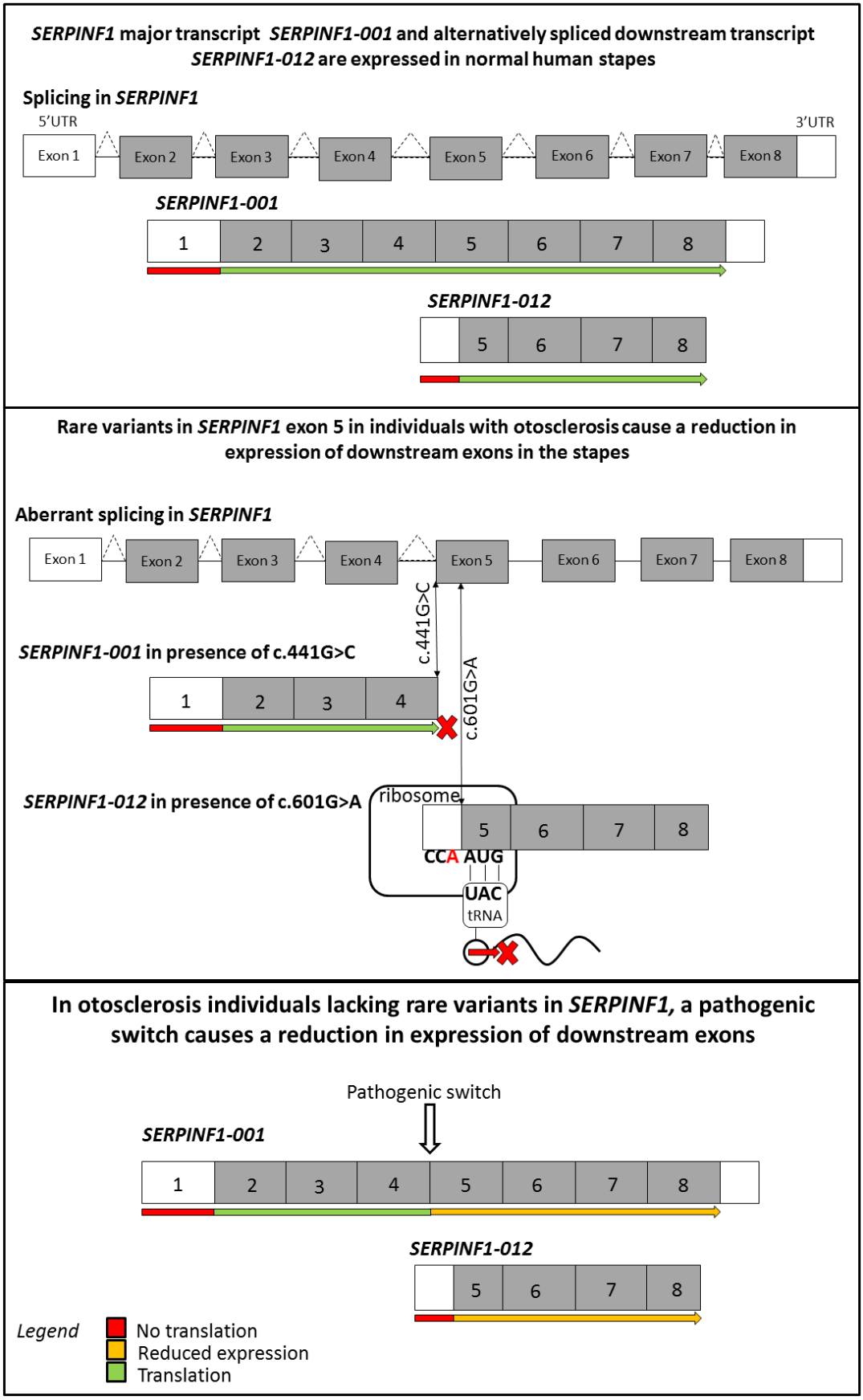


Figure 4.21 Hypothetical model of expression of *SERPINF1* in normal and otosclerotic human stapes in the presence and absence of known variants in *SERPINF1*.

4.13.8 Discussion of *SERPINF1* as an otosclerosis candidate gene

In *SERPINF1*, five rare nonsynonymous variants were identified in six unrelated probands from a cohort of 57 individuals with a history of familial otosclerosis. Therefore in total, 10.5% of the otosclerosis cohort recruited possessed a rare nonsynonymous variant within exons 3-5 of this gene, which together is a far greater percentage than would be expected in the general population. Several lines of evidence provide support for mutations in *SERPINF1* resulting in differential expression of various PEDF isoforms. *In silico* splicing assays suggest that variants within the gene may influence efficiency of alternative splicing of *SERPINF1*, and the location of SNP c.601G>A in family B indicates that it may affect efficiency of translation of certain *SERPINF1* transcripts. This together with data collected from analysis of the RNA sequencing study and quantitative PCR analysis, suggests that there may be differential expression of specific *SERPINF1* transcripts within otosclerotic stapes.

Mutations within this gene have been linked to the connective tissue disorder Osteogenesis Imperfecta, a condition in which conductive hearing loss similar to that of otosclerosis is often experienced. To date 16 unique sequence variants in *SERPINF1* have been identified in Osteogenesis Imperfecta VI. These include nonsense and frameshift mutations within all exons with the exception of exon 5, each of which causes nonsense mediated decay of mRNA transcripts, leading to loss of PEDF expression (Becker et al., 2011). Since none of the mutations identified in the otosclerosis cohort are nonsense or frameshift mutations, it is unlikely that they would lead to nonsense mediated decay. However, it is possible that the variants identified have a less extreme impact on the protein they encode and thus result in a less extreme phenotype. This would support previously published theories that otosclerosis is a localised manifestation of Osteogenesis Imperfecta (Weber, 1930). Although it is expected that a loss of PEDF results in production of undermineralised bone, the mechanism by which this occurs is unknown. In bone, since PEDF binds to collagen with high affinity and is actively expressed in osteoblastic regions of active bone formation, it is possible that PEDF plays an important functional role in bone matrix remodelling (Tombran-Tink and Barnstable, 2004). *SERPINF1* therefore represents an extremely promising otosclerosis candidate gene both for family B and for other otosclerosis families.

Furthermore, an investigation of known and predicted functional domains located with PEDF reveals that variants identified within the otosclerosis cohort could have an effect on protein structure and function. The c.441G>C SNP near the splice acceptor site of exon 5 encoding a lysine to asparagine amino acid change at residue 147 is located within a highly conserved region spanning residues 141-151 which has been proposed as a putative nuclear localisation

signal (Tombran-Tink et al., 2005). It is therefore possible that this SNP could affect nuclear localisation of the PEDF protein. In addition, PEDF contains a reactive centre loop (RCL), a proteinase recognition site that is thought to be a critical component of serpin function located at the C terminus of the protein. Although the role of the RCL in PEDF is unknown, it has been suggested to be involved in secretion of PEDF (Shao et al., 2003). It is thus a possibility that a reduction in expression of downstream exons encoding the RCL as a result of aberrant splicing in otosclerotic stapes, could have an impact on secretion of PEDF.

To further investigate alternative splicing of *SERPINF1* in human otosclerotic stapes, an in vitro translation assay could be performed to identify whether the family B SNP c.601G>A has an effect on translation of transcript *SERPINF1-012*. If this SNP was found to affect the resulting gene product, it may indicate that the variant in family B is affecting the translation of specific *SERPINF1* transcripts in human stapes. This would suggest that these transcripts are essential in the human stapes and would lead towards a better understanding of the otosclerosis disease process.

An alternative option for future work would be the construction of a minigene system to characterise alternative splicing elements. This could be used to identify cis-acting elements that define the usage of splice sites and to identify trans-activating factors that regulate alternative splicing. Through the amplification of the gene region containing the known variant(s), insertion of this region into an appropriate plasmid, followed by transfection into a cell line, it may be possible to identify if these variants are having an effect on the transcripts being transcribed. This could indicate which *SERPINF1* transcripts are affected by the variant and thus also indicate which transcripts are essential for stapes function. As a result, this technique could also bring investigators closer to elucidating the otosclerosis disease mechanism.

4.14 Discussion of WES

4.14.1 WES was successful in identification of segregating variants in four otosclerosis families

The results from this study indicate that WES is a useful tool for investigating disease causing genes in families exhibiting autosomal dominant inheritance of otosclerosis. The high sequencing depth and read quality obtained during WES has resulted in the identification of large numbers of variants with high confidence. The step-wise variant prioritisation process has been effective at narrowing down the large number of potential disease causing variants from an average of 89,656 variants per family to a small pool of those that are most likely to be involved in the disease process. Although the variant selection process may be considered somewhat stringent, it was essential in order to reduce the long lists of variants to manageable levels. To minimise the risk of eliminating the true disease causing variant at each stage in the prioritisation process, the benefits of using the filter versus the limitations of omitting it were weighed up and appropriate decisions were made. Each family was considered in turn at every stage to ensure that the most appropriate prioritisation steps were implemented.

Following the step-wise prioritisation process, a further filtering strategy was implemented to select those variants of greatest interest for segregation analysis within the otosclerosis families. This involved annotation of the genes in which variants were located, with respect to their biological role. From these annotations, along with various other factors, variants were ranked by those perceived to be of greatest interest, and top ranking variants were selected for segregation analysis. This variant selection process is therefore unavoidably subjective, but was crucial, in order to hone in on the most interesting candidates. The greatest limitation associated with this filtering strategy is that there is a limit to human knowledge as to the function and role of genes, so it is possible that good candidates were missed if the role of the genes in which they are located is currently unknown. Despite this, 23 variants within good candidate genes were prioritised for segregation analysis in the four families.

Segregation analysis has proven particularly useful in families B and C where large numbers of family members were recruited. This resulted in the elimination of a substantial proportion of variants in each family, enabling the remaining variants to be prioritised further for the next stages of analysis. However, segregation analysis proved less informative in families A and D, where it had not been possible to secure genomic DNA samples from as many relatives as had initially been anticipated when the families were selected for WES. This meant that it was not

possible to definitively eliminate variants in these families. In family A, since only four genes had been prioritised for segregation analysis, it was possible to follow up all four in the subsequent genotyping stage; however in family D, due to time and cost constraints, it was necessary to select just the two highest priority candidates for follow up. In hindsight, families A and D may not have been the optimal choice for WES out of the sixteen otosclerosis families recruited. However, at the end of year one of this research project at which time the families for WES were selected, they appeared to be amongst those most likely to yield promising results.

4.14.2 Genotyping segregating variants in a familial otosclerosis cohort enabled prioritisation of variants of interest

Genotyping 56 additional probands for each of the variants surviving segregation analysis was successful in the identification of one of these variants in unrelated cases of familial otosclerosis. The c.1224_1226dupCAT duplication in *SMAP1* in family A was identified in a heterozygous state in two of the 56 unrelated probands and in a homozygous state in one of the 56 probands. Despite this, its relatively high frequency in the general population calls into question the likelihood of its involvement in disease pathogenesis. Although at least one strong otosclerosis candidate variant emerged from each of the other three otosclerosis families for further analysis; the c.601G>A variant in *SERPINF1* in family B, c.915C>G variant in *TRIM17* in family C and the c.81G>T variant in *mir-183* and c.808G>A variant in *COL1A2* in family D, none were identified in unrelated cohort members. However, this neither confirms nor refutes a role for these variants in disease pathogenesis. Due to the heterogeneous nature of otosclerosis, it would be expected that any disease causing variant found within an otosclerosis family would be very rare. As a result it is possible that although a variant may be disease causing, it may not be found in any additional cohort members. On-going recruitment could help to tackle this, as the more individuals that are recruited with a strong family history of otosclerosis, the greater the likelihood of identifying an unrelated family who shares the variant. Since recruitment for this study is on-going, it may be possible to perform these genotyping assays on newly recruited participants reporting a strong family history of otosclerosis at a later date.

4.14.3 Further analysis was performed on four variants of interest

Of the four prioritised variants, those in *SERPINF1* and *COL1A2* stand out as the most promising candidates identified in this study, due to the identification of additional rare variants within the otosclerosis cohort. Although the c.915C>G variant in *TRIM17* and the c.81G>T variant in *mir-183* remain good candidates in their respective families, without the identification of any additional variants in the remainder of the cohort, their role as disease causing variants cannot be confirmed. Although there is greater evidence to support a role for *COL1A2* in the disease process, due to the small number of genomic DNA samples secured from family D, this is insufficient to confirm a role for this gene in otosclerosis. Additional research is required in order to suggest possible mechanisms by which these genes may be involved in the disease process.

However, for *SERPINF1* there is greater evidence for a role in disease pathophysiology. Rare nonsynonymous variants were identified in 10.5% of the familial otosclerosis cohort, and since familial otosclerosis is highly heterogeneous, it would be expected that any disease causing variant identified within an otosclerosis family would be rare. Since the otosclerosis cohort have a substantially greater than expected frequency of mutations within this gene, there is convincing evidence that *SERPINF1* may be involved in the disease process. Moreover, the location of the variants identified in the cohort, within sites predicted to be involved in alternative splicing or aberrant expression of specific transcripts, further supports a role for *SERPINF1* in otosclerosis. This combined with the gene expression data obtained from both RNA-sequencing and quantitative PCR assays, which suggest that transcripts coding for the downstream exons of *SERPINF1* are less highly expressed in otosclerotic than control stapes, implicates abnormal expression of specific *SERPINF1* transcripts in the disease process. There is therefore highly persuasive evidence that mutations in *SERPINF1* result in the production of alternatively spliced *SERPINF1* transcripts in the stapes of individuals affected by otosclerosis. In future research, it would be necessary to elucidate the exact role that *SERPINF1* transcripts play in bone formation in the stapes in order to determine if variants within this gene are causing monogenic inheritance of otosclerosis.

It is necessary to evaluate the choice of the four variants for follow up analysis, as by focussing on these variants, other possible disease causing variants may have been overlooked. For this reason, it must be taken into consideration that of the four prioritised variants; the c.601G>A variant in *SERPINF1* and the c.808G>A variant in *COL1A2*, were initially genotyped in the remainder of the otosclerosis cohort using Sanger sequencing rather than restriction endonuclease digests, as was the case for the majority of other variants identified. This was

because no suitable restriction endonuclease cut at the site of these variants. Since the Sanger sequencing revealed additional variants of interest in each of these genes, which would not have been possible if restriction endonuclease digests had been performed, it is possible that an element of bias was introduced into the study, as following identification of additional variants, these genes were prioritised over others for further analysis. Had appropriate restriction enzymes been available, no additional variants would have been identified and therefore it is possible that these two good candidates would not have been prioritised for follow up. Having said this, given the biological role of these two genes, both of which are known to be involved in systemic connective tissue disorders, they would have been excellent candidates for follow up even if no additional variants had been identified within these genes.

In summary, through WES followed by a step-wise prioritisation process, a number of good candidate variants have been identified which may be involved in disease pathology. At least one strong candidate gene has emerged from three of the four otosclerosis families: *SERPINF1* in family B, *TRIM17* in family C and *mir-183* and *COL1A2* in family D. In family A, it would be necessary to return to an earlier stage of the filtering process to identify additional good candidate variants for segregation analysis. For the four prioritised variants identified within the other three families, analysis of RNA sequencing data as well as further genotyping on these genes revealed that all are good candidates for involvement in disease pathology. In particular, strong evidence suggests that alternative splicing of *SERPINF1* in the human stapes may be involved in the disease process leading to otosclerosis. It is hoped that this research will contribute to the current literature to help improve understanding of the pathways involved in pathogenesis of otosclerosis, with the hope of improving treatment options for otosclerosis patients in the future.

5 Whole transcriptome expression analysis identifies dysregulated genes in otosclerotic stapes

5.1 Introduction

Whole transcriptome expression analysis by Next Generation Sequencing (NGS) is acquiring increasing importance in investigating molecular mechanisms involved in complex human disease. RNA-sequencing (RNA-seq) is one such NGS technique, which provides a high throughput transcriptome–wide approach to gene expression analysis, allowing comprehensive evaluation of all RNA molecules expressed within a tissue type (Ozsolak and Milos, 2011). Since its initial application, RNA-seq has made significant progress in quantification and characterisation of transcriptomes, particularly in the fields of cancer research and neurodegenerative disease (Costa et al., 2013). It is now increasingly acquiring an important role in elucidating molecular pathways involved in the development of other complex human diseases and traits.

To date, RNA-seq has not been used to analyse the gene expression profiles of otosclerotic tissue, however previous attempts have been made to characterise gene expression in the stapes. Most of these gene expression studies have focused on a small number of candidate genes, with only one employing a transcriptome-wide approach. In this study, a whole genome hybridisation array was used to investigate gene expression in nine otosclerotic and seven control stapes (Ealy et al., 2008). This provided a comprehensive stapes gene expression dataset, enabling the identification of 110 dysregulated genes. However, of seven dysregulated genes prioritised for validation using quantitative reverse transcription PCR (qRT-PCR) in the same stapes samples, dysregulation of only two of these genes was confirmed. This may have been due to the small number of stapes analysed during this study. As a whole transcriptome approach, RNA-seq has great promise for identifying novel genetic pathways involved in otosclerosis, through the provision of a comprehensive stapes gene expression dataset and has the benefit over more traditional techniques of being able to detect and quantify previously uncharacterised transcripts, as well as transcripts expressed at low levels (Ozsolak and Milos, 2011).

During RNA-seq, millions of short RNA fragments are generated from a messenger RNA sample. These are converted into a library of cDNA fragments by a fragmentation process

followed by the addition of sequencing adaptors to each cDNA fragment. High throughput sequencing generates a series of sequence reads which are aligned to a reference transcriptome. Reads that align to exons, genes and splice junctions within the transcriptome are counted, which enables a read count for each transcript to be generated. From this read count, transcript-level expression is calculated. This data is then subject to statistical analysis as well as visualisation on an appropriate data analysis platform. Following this, pathway analysis along with various other analyses such as functional annotation may be performed to identify trends in the data (Wang et al., 2009) (Figure 5.1).

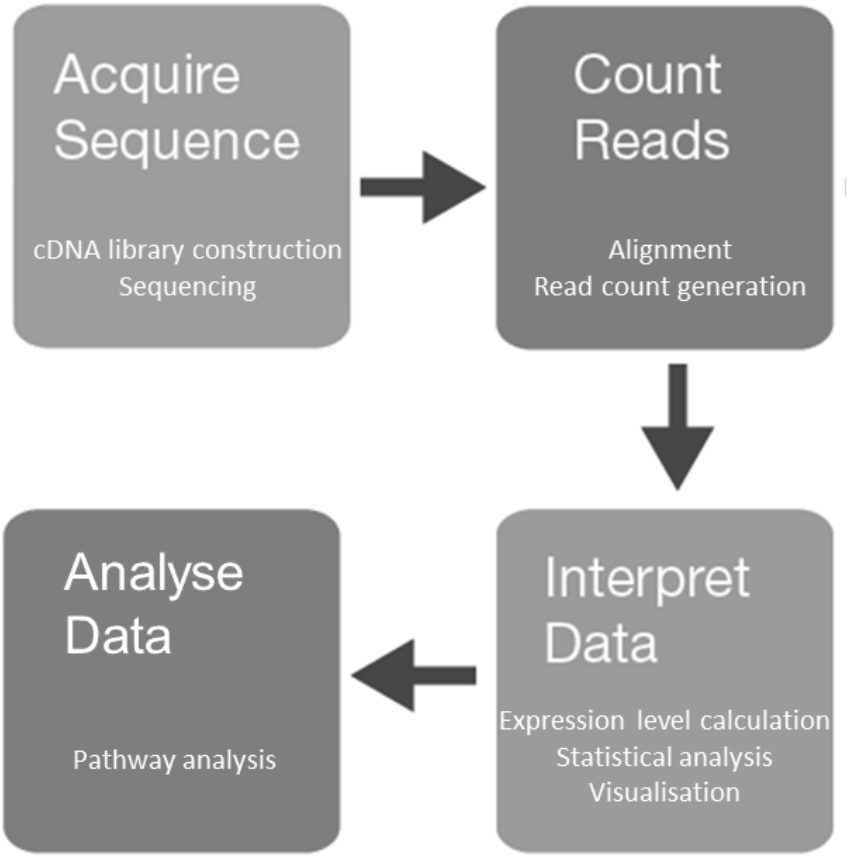


Figure 5.1 Flow chart showing the four stages involved in RNA-seq

Stage include acquisition of the sequence, read counting, data interpretation and data analysis

Although this next generation sequencing technique has many benefits over traditional analyses, there are some obstacles which have been encountered in previous otosclerosis gene expression studies that are likely to remain issues when employing an RNA-seq approach. These include difficulties in gaining access to appropriate human tissue. This is particularly problematic in regards to acquisition of large numbers of control stapes, as there are few clinical scenarios during which healthy stapes are removed. It is also likely to be true of

ascertainment of otosclerotic stapes footplates, as the stapedotomy procedure where the footplate is not routinely removed is standard practice in the UK.

Furthermore, the RNA-seq technique also has a number of additional limitations. The fragmentation process performed during library construction may introduce experimental bias by leading to an under-representation of RNAs with a stable secondary structure, including non-coding RNAs and those with a high GC content (Risso et al., 2011). In addition, during RNA-seq, since more reads will map to long than short transcripts of similar expression levels, short transcripts are at an inherent statistical disadvantage over longer transcripts (Oshlack and Wakefield, 2009). A similar issue is true of genes expressed at low levels such as transcription factors, when comparing expression of genes between experimental groups, as when absolute differences in expression levels of a specific transcript are compared, those expressed at low levels will be less likely to be significantly different between groups than more highly expressed transcripts. To overcome this obstacle, it is essential to consider relative expression levels of transcripts between experimental groups in addition to absolute expression levels. Furthermore, false positive results may occur during RNA-seq due to misalignment of reads, and are particularly common in gene families with paralogous members or genes with highly repetitive regions. RNA-seq also leads to issues with downstream computational analysis as a result of the large volumes of data generated which must be analysed, validated and interpreted. Despite attempts to streamline analysis of such data between research groups, there is currently no consensus as to the best method of data analysis. This introduces an element of variability in analysis techniques between research groups and subjectivity in interpretation of findings (Ozsolak and Milos, 2011). Moreover, this technique is more expensive than traditional microarray based techniques, costing roughly $1,000 U.S. to perform per sample. Despite these limitations, the great potential benefits associated with whole transcriptome profiling combined with a steady reduction in cost, has meant that RNA-seq has emerged as a popular technique of choice for researchers performing gene expression analysis (Qi et al., 2011).

Here, a small number of otosclerotic and unaffected stapes samples were subjected to RNA-seq in order to provide a transcriptome-wide gene expression dataset. Differentially expressed genes of interest within the otosclerotic stapes identified within this small discovery group, were further investigated using qRT-PCR in a larger cohort of stapes. In this chapter, the choice of stapes selected for RNA-seq will be discussed, along with statistical analysis of the data obtained, in order to provide a stapes gene expression dataset and to identify genes significantly dysregulated in otosclerotic stapes compared to unaffected stapes. Functional

annotation and pathway analysis performed on datasets of these dysregulated genes will then be outlined and trends identified using this technique will be summarised. Following this, the choice of which dysregulated candidate genes to select for follow up single gene expression studies using qRT-PCR will be discussed, as well as the results from these single gene assays. The results and discussion sections will be combined in this chapter, as it is essential to interpret the findings following RNA-seq in order to select the best possible candidates for follow up analysis. An overall general discussion will summarise the findings at the end of the chapter in section 5.3.

5.2 Results and Discussion

5.2.1 Selection of stapes for RNA-seq

Throughout the course of this research project, 110 otosclerosis patients were recruited to the surgical arm of the study. Of these study participants, 109 received stapedotomy surgery, during which their stapes suprastructure was collected, whilst one participant received a stapedectomy surgery, where their whole stapes was acquired. A further 11 stapes suprastructures and two footplates were collected from unaffected individuals receiving middle-ear surgery.

Of these samples, twelve were selected for RNA-seq including four unaffected and eight affected stapes samples (Table 5.1). Within the unaffected group, which is referred to hereon as the control group, two whole stapes, one from each ear from the same individual (ConStap1L and ConStap1R) were dissected into footplate and suprastructure components and pooled. In addition, the whole otosclerotic stapes (OtoStap1) which had been collected during stapedectomy surgery, was also dissected into footplate and suprastructure components, which were analysed separately.

5.2.1.1 Rationale for choice of stapes for RNA-seq

During selection of stapes for RNA-seq, four control stapes samples were chosen in addition to eight otosclerotic samples. This enabled a comparison to be made between gene expression datasets in diseased and healthy tissue. The lack of control samples available was a major limitation when selecting stapes for sequencing. For this reason, RNA from all control stapes that had been collected at the time of sequencing were included in the RNA-seq study, and all that were subsequently collected were used in follow up analysis.

Sample ID	Footplate/ Suprastructure	Unaffected/ Confirmed otosclerosis	Age at recruitment	Gender	Disease and Hearing status	
ConStap1L&RFP	Footplate	Unaffected	11	m	Child with meningitis, whole stapes removed	
ConStap1L&RSS	Suprastructure	Unaffected	11	m		
ConStap2SS	Suprastructure	Unaffected	51	m	Stapes removed following head trauma	
ConStap3SS	Suprastructure	Unaffected	34	f	Exploratory surgery following injury to ear	
OtoStap1FP	Footplate	Confirmed otosclerosis	50	m	Peripheral fixation	Conductive Hearing Loss
OtoStap1SS	Suprastructure	Confirmed otosclerosis	50	m		
OtoStap2SS	Suprastructure	Confirmed otosclerosis	75	m	Peripheral fixation	Mixed Hearing Loss
OtoStap3SS*	Suprastructure	Confirmed otosclerosis	63	f	Peripheral fixation	Conductive Hearing Loss
OtoStap4SS	Suprastructure	Confirmed otosclerosis	36	f	Peripheral fixation	Conductive Hearing Loss
OtoStap5SS	Suprastructure	Confirmed otosclerosis	47	f	Peripheral fixation	Conductive Hearing Loss
OtoStap6SS	Suprastructure	Confirmed otosclerosis	56	f	Peripheral Fixation	Conductive Hearing Loss
OtoStap7SS	Suprastructure	Confirmed otosclerosis	37	m	Peripheral fixation	Conductive Hearing Loss

Table 5.1 A table illustrating the twelve stapes selected for RNA sequencing

* OtoStap3SS collected from individual III:3 in family B (Chapter 4) for whom WES was performed

Since the stapes footplate is the predominant location of histologic changes during otosclerosis, it is important that a comparison can be made between expression patterns in the stapes footplate and suprastructure. Therefore all three stapes footplates collected, two of which were from the same unaffected individual and one of which was from an otosclerosis patient, were selected from the cohort for analysis. This will help to clarify any doubts concerning differences in gene expression between footplates and suprastructures.

Stapes suprastructure OtoStap3SS was selected for sequencing because this stapes came from an individual who had a very strong family history of otosclerosis and was a member of one of the families selected for Whole Exome Sequencing (WES) (see chapter 4). It was hoped that the results from the RNA-seq on this particular stapes would assist in the candidate gene identification process following WES in this family.

The remaining otosclerotic stapes were selected from individuals of age range from 36-75, with peripheral fixation of the stapes which is indicative of early or mid-stage disease, who had either conductive or mixed hearing loss. By selecting stapes with early or mid-stage stage disease, the samples were standardised as much as possible, in order to increase the likelihood of finding any differentially expressed genes, as it would be expected that individuals with similar disease characteristics would also share similar gene expression profiles in their stapes. Approximately equal numbers of individuals of each gender were selected (3 male and 4 female stapes suprastructures) to account for any gender-specific changes in gene expression. It is important to account for this due to the reported variation in incidence of otosclerosis between the genders.

Due to the high cost of RNA-seq, only small numbers of stapes were included in the analysis. This small cohort of stapes was therefore considered a discovery group. A larger cohort of stapes was later used to further investigate the results obtained.

5.2.2 Overview of RNA-seq data

RNA-seq was performed on twelve individual stapes samples by the U.S.A based sequencing company Otogenetics Corp (Norcross, GA). Before the samples were sent to Otogenetics, a quality control procedure was performed on the cDNA isolated from each of the stapes samples using the Agilent Bioanalyzer 2100. The results obtained were sent to Otogenetics which confirmed that the sample integrity and quantities of cDNA were sufficient for RNA-seq. The samples were then sent to Otogenetics which performed cDNA fragmentation on the samples and constructed cDNA libraries before carrying out high-throughput paired-end RNA sequencing using Illumina HiSeq 2000. Paired-end 100 nucleotide reads were aligned to human genomic assembly hg19, which generated a read count for every transcript. An average of 2.3 billion bases and 23 million reads were generated per sample (Table 5.2). This data was analysed using the platform provided by DNAnexus Inc (Mountainview, CA), a cloud-based data analysis platform for sequencing data.

Sample ID	Total bases	Total reads
ConStap1L&RFP	2,788,077,000	27,880,770
ConStap1L&RSS	2,899,915,000	28,999,150
ConStap2SS	3,036,257,600	30,362,576
ConStap3SS	2,903,808,400	29,038,084
OtoStap1FP	3,307,667,000	33,076,670
OtoStap1SS	2,436,615,000	24,366,150
OtoStap2SS	843,953,600	8,439,536
OtoStap3SS*	2,782,302,200	27,823,022
OtoStap4SS	3,272,769,800	32,727,698
OtoStap5SS	2,167,111,800	21,671,118
OtoStap6SS	903,777,800	9,037,778
OtoStap7SS	2,269,353,400	22,693,534

Table 5.2 Table showing a summary of RNA-seq statistics for each sample

* Stapes from individual III:3 in family B (Chapter 4) for whom WES was performed

Through DNAnexus the expression level of each transcript was quantified through two tools for expression profiling of annotated transcripts; RNA-Seq and 3SEQ/transcriptome. Using these tools, the expression levels of specific transcripts were quantified by mapping reads against a reference transcriptome, generating a measure of RNA relative abundance for 21,563 annotated genes. This data was represented in two formats; RPKM (reads per kilobase of transcript per million mapped reads) values and z scores. The statistical analysis performed on the RPKM values and z scores generated for each gene are discussed in sections 5.2.3.1 and 5.2.3.2 respectively.

5.2.3 Identification of a gene expression profile in human stapes suprastructures

To compare gene expression datasets for affected and control stapes, two statistical analyses were performed in parallel, using the RPKM data and z score data independently. This analysis was performed using only the data from stapes suprastructures, in order to avoid introducing inconsistency and bias through use of the footplate data.

5.2.3.1 RPKM analysis

The RPKM value is a measure of the level of expression of a transcript which is calculated by normalising the read count for the transcript against the transcript length across samples of varying coverage. This enables expression levels of transcripts to be compared without bias of transcript length. To identify differential expression in control and affected human stapes suprastructures from the RPKM data, a two tailed paired Student's *t*-test was conducted on all 21,563 annotated genes, to identify those genes with a significant difference in expression between the control and affected groups. The fold difference in expression between the mean RPKM for the control and affected groups was also calculated because *t*-tests frequently result in arbitrarily small fold-changes in expression being considered statistically significant. It is generally considered that genes are differentially expressed if they show a fold-change of at least 1.5 and also satisfy p value threshold of 0.05 (Raouf et al., 2008). In total 2,220 genes met both these criteria with 1,673 genes found to be significantly down-regulated in human stapes suprastructures compared to controls and 547 found to be significantly up-regulated (Table 5.3).

During statistical analysis of large gene expression datasets such as these, where multiple comparisons have been performed, the probability of incorrectly assigning significance to a gene is in excess of 0.05. Therefore corrections for multiple testing were performed on the RPKM dataset. A Bonferroni correction was carried out in order to control for the probability of incorrectly assigning significance; however none of the differentially expressed genes in stapes suprastructures reached significance after Bonferroni correction. Therefore the False Discovery Rate (FDR) was also calculated. This exerts less stringent controls over false discovery than the Bonferroni correction, however again, none of the transcripts reached significance (Benjamini et al., 2001). Although this raises concerns regarding false positive results in the data, it has been postulated that Bonferroni and FDR corrections are too stringent for transcriptome-wide expression data and are not routinely applied to whole transcriptome analysis.

5.2.3.2 z score analysis

The z score is a measure of the level of expression of a transcript, a statistic which is generated from just one read per transcript, calculated from the read count. It is normalised by the total number of reads in the sample, thus allowing a comparison of data between samples. From the z score data, z ratios were calculated by dividing the difference in mean z score between the affected and control stapes by the standard deviation of difference in mean z score across all 21,563 annotated genes. This normalises the dataset in order to reduce bias across the samples. A z ratio that is greater than ±1.96 is considered to be significant at 95% confidence intervals (Cheadle et al., 2003). A positive z ratio represents genes that are up-regulated in otosclerotic stapes compared to controls whilst negative z ratios represent down-regulated genes. In total of the 21,563 annotated genes, 176 were found to be dysregulated; 97 were found to be up- regulated and 79 down-regulated in otosclerotic stapes (Table 5.3).

5.2.3.3 A comparison of RPKM and z score gene expression datasets

Genes identified as significantly differentially expressed through analyses of RPKM values, were compared with those identified through the z score analysis, and only those that reached significance in both analyses were prioritised for further analysis. In total, 101 genes were dysregulated in both datasets; 51 genes that were down-regulated and 50 that were up-regulated (Table 5.3). The full list of 101 dysregulated genes can be found in the Appendix.

	RPKM [p value threshold = 0.05 & Fold difference threshold = 1.5]	**z ratio** [Threshold = +/- 1.96]	**Overlap**
Up-regulated	547	97	50
Down-regulated	1673	79	51
Total	2220	176	101

Table 5.3 Table showing the number of up and down-regulated genes identified during RNA-seq

Table shows numbers of genes that met statistical analysis thresholds for the RPKM analysis, z ratio analysis and the total number of genes that met threshold criteria in both analyses

5.2.4 Identification of a gene expression profile in human stapes footplates

Statistical analysis was also performed on the stapes footplate data in order to identify if there was any overlap in genes found to be dysregulated in otosclerotic suprastructures and footplates. Since only one control and one affected stapes footplate had undergone RNA-seq, a Student's *t*-test was not possible as this statistical test relies on there being multiple data points within a single group. For this reason, RPKM data was not analysed for the footplate samples. z ratios were calculated by dividing the difference in z score between the affected and control footplate by the standard deviation of difference in z score across all 21,563 annotated genes. In total, of the 21,563 annotated genes, 97 were found to be dysregulated in the footplate samples; 6 were found to be up-regulated and 91 down-regulated. These 97 genes were then compared with those found to be dysregulated in the suprastructure samples based on z score analysis. In total, 63 genes were dysregulated in both tissue types; 4 genes that were up-regulated and 59 that were down-regulated (Figure 5.2).

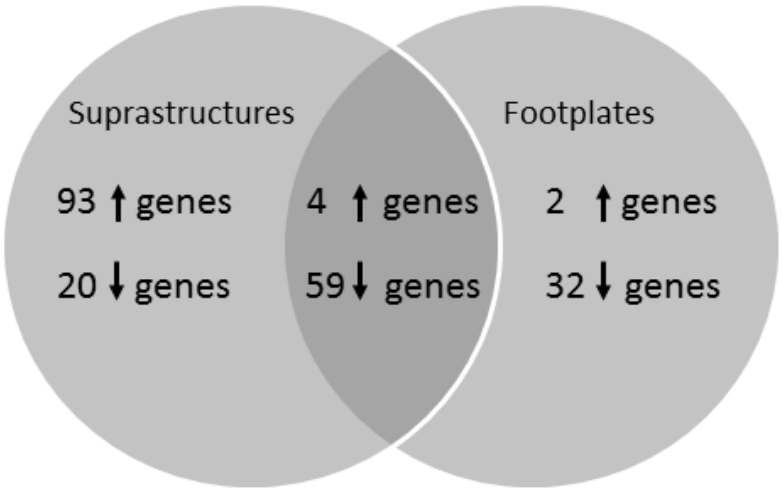


Figure 5.2 Venn diagram illustrating overlapping dysregulated genes from stapes suprastructure and footplate analyses

Number of genes identified is based on z score data analysis

The predominant focus of this chapter is on the genes found to be dysregulated in stapes suprastructures rather than footplates. This is because since only one control and one affected footplate sample was available for analysis; confidence in the results from the footplate analysis is not as strong as for the suprastructure analysis. Having said this, genes that were found to be dysregulated in stapes footplates in addition to suprastructures were considered as good potential candidates for follow up gene expression analysis in the remainder of the cohort (see section 5.2.6).

5.2.5 Identifying molecular pathways in the stapes gene expression transcriptome

The stapes gene expression dataset identified during z score analysis was analysed using two tools; Ingenuity Pathway analysis (IPA) (http://www.ingenuity.com), and the DAVID functional classification tool (http://david.abcc.ncifcrf.gov). Each of these tools will be discussed in turn before pathways identified from use of both tools are summarised.

5.2.5.1 Ingenuity Pathway Analysis

Gene network analysis was performed using Ingenuity Pathway Analysis (IPA) on all 176 significantly dsyregulated genes identified in stapes suprastructures based on the z score analysis. This was chosen, as opposed to the gene list identified through RPKM analysis, because z score data is normalised by the total number of reads in each sample, meaning that greater confidence can be had in genes identified through this technique than RPKM analysis. A network score was generated by IPA for each network of genes identified. This score takes into account the likelihood of a group of genes being identified in a single gene network purely by chance. (Letwin et al., 2006). The figure below illustrates the five highest scoring networks identified using IPA (Figure 5.3).

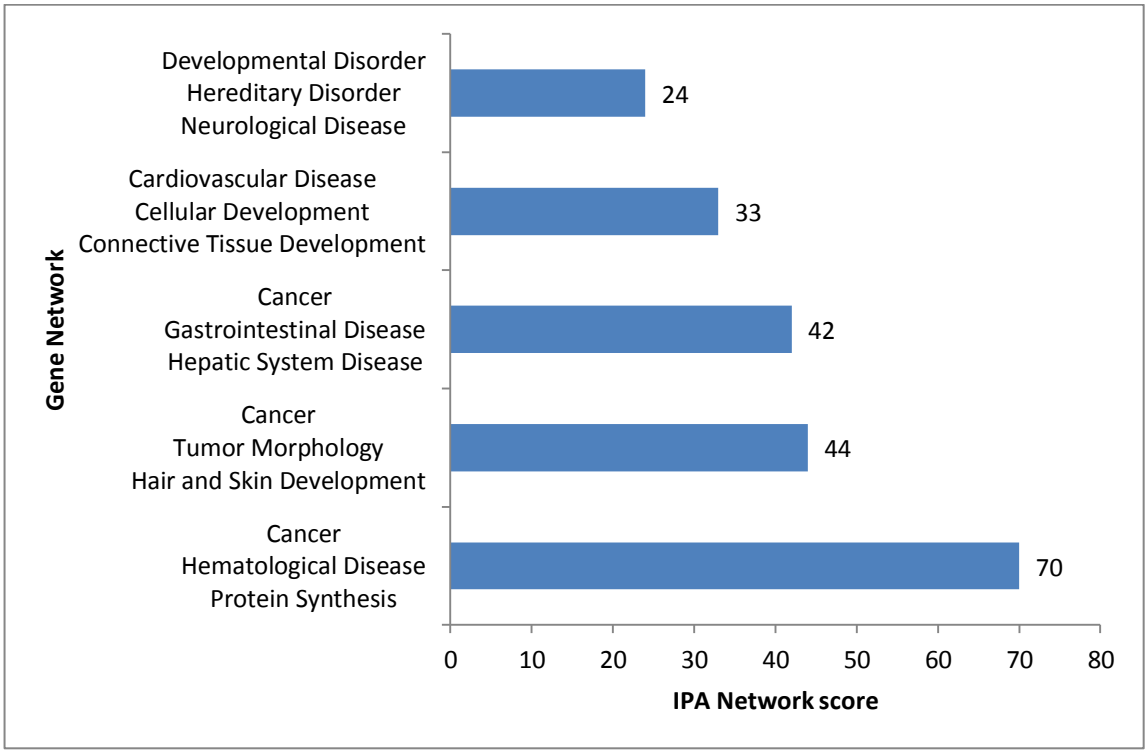


Figure 5.3 Bar graph showing five highest scoring networks from IPA analysis and their respective network scores.

Networks with a score greater than 2 are considered to be statistically significant with confidence of 99% (Letwin et al., 2006)

IPA was also used to identify trends in biological functions amongst the dysregulated genes. This analysis considers the number of genes in the dataset that are involved in three biological function categories; diseases and disorders, molecular and cellular functions, and physiological system development and function. The graph below shows the five top biological functions for each of these three criteria including the number of genes dysregulated within each category (Figure 5.4).

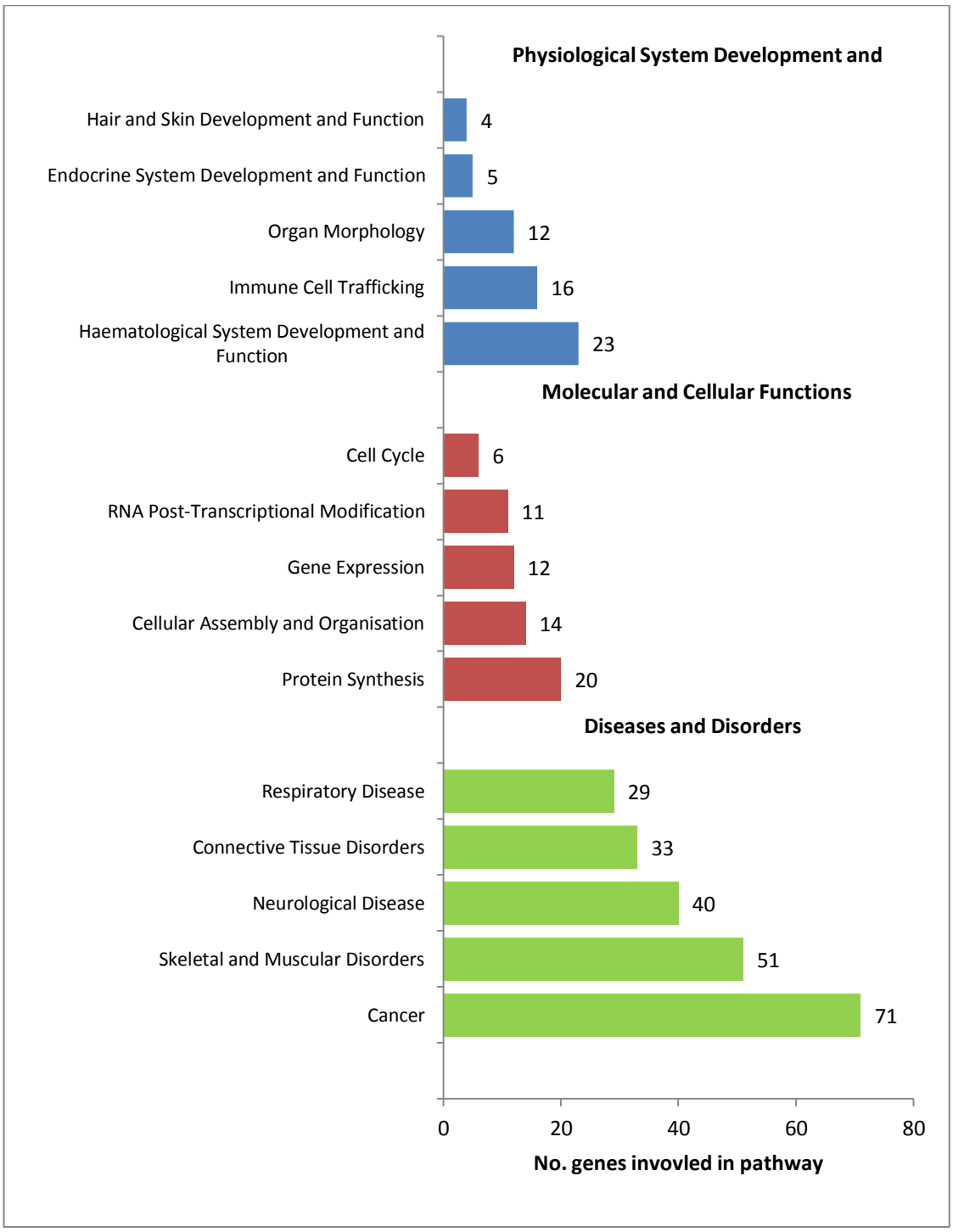


Figure 5.4 Bar graphs showing top biological functions identified during IPA analysis

5.2.5.2 DAVID Functional classification analysis

The 176 significantly dsyregulated genes identified in stapes suprastructures based on the z score analysis were also analysed using the gene functional classification tool DAVID Bioinformatics resource v6.7, in order to identify functionally related gene groups within the list of dysregulated genes (http://david.abcc.ncifcrf.gov). This was performed by analysing the Gene Ontology (GO) terms attributed to each gene. These are standardised terms that represent gene and gene product attributes across species and databases (http://www.geneontology.org/), thus allowing cross-comparisons to be made between gene functions in order to group together those genes with shared or similar functions. Genes that were both significantly up and down-regulated were analysed using the DAVID functional classification tool. During analysis, clusters of genes were assigned an enrichment score based on agreement between GO terms, with a higher score representing greater similarity between terms. The three highest scoring clusters for up-regulated genes are genes encoding zinc finger proteins, transcription factors and purine nucleotide binding proteins, whilst the three highest scoring clusters for down-regulated genes are ribosomal genes, genes encoding leucine-rich repeat proteins and myosin proteins (Table 5.4).

Gene Cluster	Up or Down-regulated	No. genes	Top 3 scoring genes in cluster	Enrichment score
Zinc Finger Proteins	Up-regulated	8	*ZBTB8A* *ZNF793* *ZNF461*	1.17
Transcription Factors	Up-regulated	4	*SIX4* *FOXK1* *TAF8*	0.39
Purine Nucleotide Binding Proteins	Up-regulated	6	*HSP90AB2P* *CAM1KD* *PTK6*	0.35
Ribosomal Genes	Down-regulated	36	*RPL23A* *RPL15* *RPS25*	40.87
Leucine-rich Repeat Proteins	Down-regulated	6	*LUM* *OMD* *ASPN*	1.32
Myosin Proteins	Down-regulated	3	*MYL6* *MYL12A* *MYL12B*	2.70

Table 5.4 Table of the three top scoring gene clusters for both up and down-regulated genes from the RNA-seq dataset

Clusters identified from functional gene classification using DAVID (http://david.abcc.ncifcrf.gov).

5.2.5.3 Discussion of pathway and functional classification analysis

Both the IPA gene network and biological function analysis suggest that the dataset is highly enriched for genes involved in cancer. Cancer-related gene networks constitute the three top scoring networks, and 71 out of the 176 dysregulated genes are known to be involved in cancer, based on biological function analysis. Since both cancer and otosclerosis are characterised by cellular hyperproliferation, it is of interest that these conditions share molecular pathways. It is therefore more likely that dysregulated genes involved in cell proliferation are markers of disease tissue rather than genes involved in disease pathogenesis. However, caution must be taken when interpreting this as due to an overwhelming body of research on the genetics of cancer, an extremely large number of genes have been implicated in the various forms of this disease through numerous different genetic studies. This makes it more likely that genes with a known involvement in cancer will be present in any randomly selected group of genes in comparison to other disorders which have not been investigated in so much depth.

The IPA gene network analysis and biological function analysis also suggest that the dataset is highly enriched for genes involved in connective tissue disorders with 33 out of the 176 dysregulated genes known to be involved in connective tissue disorder pathways. In addition, the DAVID functional classification analysis suggests that the dataset is over-represented for genes containing leucine- rich repeats, a motif which is frequently found in connective tissue proteins (Bengtsson et al., 1995). Since it has been suggested that otosclerosis occurs concomitantly with other connective tissue disorders, it is possible that genes involved in such disorders may also be involved in the otosclerosis disease process. These 33 dysregulated genes are known to be involved in numerous connective tissue disorders, three of which have been linked to at least three of the dysregulated genes in the otosclerosis dataset. 29 dysregulated genes are involved in arthritis, and three in each of osteosarcoma and pigmented villonodular synovitis, a disease characterised by inflammation and overgrowth of the joint lining (Figure 5.5). One of these genes, *SPP1* (Osteopontin), which is involved in the attachment of osteoclasts to the mineralised bone matrix, is involved in all three of these disorders. Genes involved in these disorders are therefore excellent candidates for involvement in disease pathogenesis. Having said this, it must also be taken into consideration that since RNA-seq was performed on RNA extracted from bone tissue, genes encoding connective tissue proteins that form key constituents of bone are likely to be dysregulated in otosclerotic stapes. It is therefore possible that these genes could be markers for disease tissue rather than being involved in disease pathogenesis.

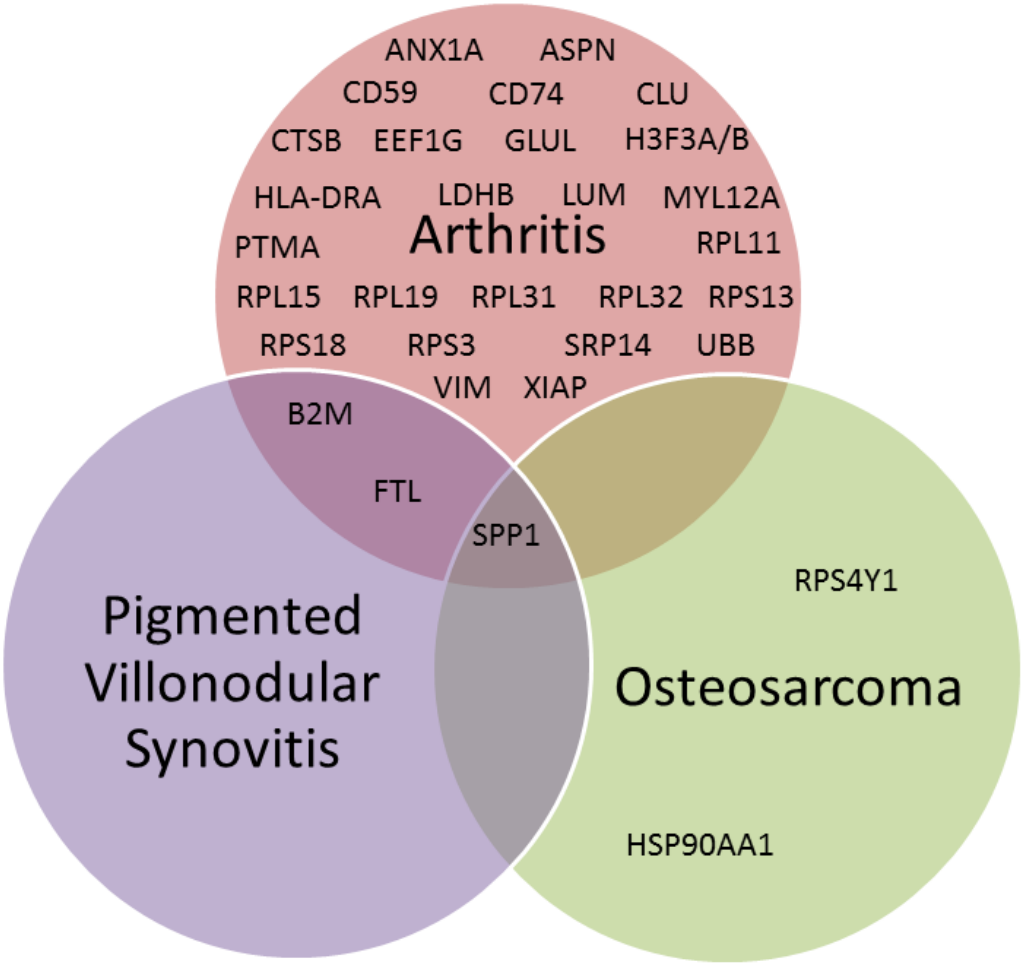


Figure 5.5 Venn diagram showing the dysregulated genes known to be involved in arthritis, osteosarcoma and pigmented villonodular synovitis

The IPA biological function analysis of molecular and cellular functions also suggests that the dysregulated gene list is enriched for genes involved in functions involved in translational processes including protein synthesis, gene expression and RNA post-transcriptional modification. This coincides with the outcome from the DAVID analysis where 36 dysregulated ribosomal genes were identified in the dataset, with an extremely high enrichment score of 40.87. Since ribosomes are involved in the translational process, dysregulation of large numbers of these genes are likely to have an impact on protein synthesis and gene expression. It is therefore possible that local dysregulation of ribosomal genes is common in this type of gene expression analysis where disease tissue is being compared with healthy tissue. These findings may therefore suggest that aberrant expression of ribosomal genes is a result of the disease process rather than a causative factor.

The IPA biological function analysis of physiological system development and function suggests that genes involved in immune cell trafficking and endocrine system development and

function are enriched in the list of dysregulated genes. This coincides with suggestions in the literature that viral infection and hormones may be involved in otosclerosis. It is likely that if these factors are involved in otosclerosis pathogenesis, that genes encoding proteins involved in these functions would be dysregulated in the diseased stapes. Since the evidence for viral infection and a hormonal role in disease pathology from previous research studies is somewhat conflicting, the outcome of this analysis could provide some supporting evidence for the previous studies that made a positive link between these factors and otosclerosis.

The DAVID functional classification analysis also suggests that both zinc finger proteins and transcription factors are over-represented in the list of dysregulated genes. Since both these groups encode proteins involved in regulating gene expression, it is possible that dysregulation of these genes could be involved in regulation of a large numbers of other genes and gene pathways. Given that both zinc finger proteins and transcription factors have previously been suggested by various research groups as possible contributors to otosclerosis pathophysiology and that both have been implicated in numerous other genetic diseases, it is possible that dysregulation of these could contribute to development of otosclerosis.

Pathway and functional classification analysis can be useful in the identification of interacting genes and those with shared biological functions. However, since genes interact with one another as well as environmental factors in numerous interacting pathways, of which few have been well characterised, the identification of a dysregulated gene or group of related genes may indicate involvement in a disease process but cannot be used to identify a causative gene. It is also possible that these dysregulated genes are markers for the disease rather than being involved in the pathologic switch that leads to otosclerosis in the stapes. For this reason, any findings obtained during pathway or functional classification analysis must be considered in the relevant context and their impact should not be over-stated. Having said this, such tools are extremely useful for helping to prioritise genes for follow up analysis after RNA-seq, as is discussed in the following section.

5.2.6 Selection of candidate genes for follow up gene expression studies

A small number of genes dysregulated in the discovery group were selected as interesting candidates for follow up gene expression analysis in the remainder of the cohort. All were selected from the list of 101 genes that were found to be significantly dysregulated through both z score and RPKM analysis. By prioritising overlapping genes from both analyses, greater confidence can be had that the each gene is significantly dysregulated in the otosclerotic

stapes than by using the results of just one analysis alone. In order to ensure that the genes most likely to be involved in the disease process were selected for follow up, all 101 genes were annotated by gathering further information from online databases. These annotations combined with the pathway analysis and functional annotation analysis, revealed that eight of the 101 genes had roles in bone growth and maintenance and two were involved in hearing loss, according to data from the Gene Cards website (Table 5.5). Mutations within four of these genes led to bone and growth defects when mutated in mice whilst mutations in one of these genes caused a hearing loss phenotype in mice, based on information from the JAX lab website. One of these genes has been implicated in the connective tissue disorder Ehlers Danlos Syndrome, according to the OMIM website. These annotations prioritised genes most likely to have an impact on bone regulation or conductive hearing loss, or to be involved in another connective tissue disorder, as genes involved in these processes are good otosclerosis candidates.

Website	Bone abnormalities	Hearing anomalies	Connective Tissue Disorders
Gene Cards	8	2	
JAX lab (Jackson Lab - Mouse Genome Informatics)	4	1	
OMIM (Online Mendelian Inheritance in Man)			1

Table 5.5 Table showing the number of genes known to be involved in bone abnormalities, hearing anomalies and connective tissue disorders

Number of genes identified from data obtained from three online databases.

In addition to these annotations, when selecting the most promising dysregulated genes for further investigation studies in larger numbers of stapes, it was also taken into account if genes were located within one of the eight otosclerotic loci identified during linkage analysis, found to be dysregulated in the 2008 gene expression analysis of otosclerotic stapes (Ealy et al., 2008), or contained variants that had been associated with otosclerosis during the 2009 Genome Wide Association Study (Schrauwen et al., 2009a). However, none of these genes were found to be dysregulated in the Ealy et al. study, nor were any associated SNPs identified within these genes in the otosclerosis Genome Wide Association Study. Although four of the genes were located in linked regions, all of these were located in the two largest linked

regions, OTSC3 and OTSC10, which contain 423 and 148 protein coding genes respectively. Since these genes were located within linked regions containing such a large number of possible genes, it was decided that these genes would not be prioritised solely on this basis. Genes were also given greater priority for follow up analysis if they were found to be differentially expressed in stapes footplates as well as stapes suprastructures, since the stapes footplate is the predominant location of histologic changes during disease.

5.2.6.1 Six candidate genes of interest were selected for further investigation

Six genes of interest were selected for follow up studies using qRT-PCR, including three that were up-regulated and three that were down-regulated in otosclerotic tissue. Based on the annotations derived from the three online databases, as well as taking into account gene expression in footplate samples, the following six genes illustrated in Table 5.6 were selected.

Gene	Up or down regulated	z ratio	RPKM		Dysregulated in footplates
			p value	Fold difference	
ASPN (Asporin)	Down	-3.02	0.0004	5.3	No
FBLIM1 (Filamin binding LIM protein 1)	Up	2.6	0.0009	2.0	No
FKBP14 (FK506 binding protein 14)	Up	4.0	0.03	1.5	No
LAIR1 (leukocyte-associated immunoglobulin-like receptor 1)	Up	2.6	0.01	1.7	No
LUM (Lumican)	Down	-16.3	0.006	2.7	Yes (z=-13.9)
SPARCL1 (SPARC-like 1)	Down	-3.9	0.002	3.4	Yes (z= -3.0)

Table 5.6 Table illustrating the 6 genes prioritised for follow up analysis in the remainder of the cohort

The table highlights the 6 genes selected for follow up analysis. The table identifies the gene name, whether it was found to be up or down-regulated during the RNA-seq study, the z ratio, p value and fold difference calculated during statistical analysis for each gene, the role of the gene according to annotations and whether or not the gene was also found to dysregulated in stapes footplates

5.2.6.2 Justification for selection of six candidate genes for further analysis

The gene which was found to be the most significantly down-regulated in otosclerotic stapes compared to controls based upon RPKM values is Asporin (*ASPN*). *ASPN* is one of the six leucine-rich repeat proteins identified during the DAVID functional classification analysis and is known to be highly expressed within cartilage. It is thought to regulate chondrogenesis by inhibiting TGFβ-1 induced gene expression. Since members of the TGFβ superfamily have been associated with otosclerosis during association studies, *ASPN* is an excellent otosclerosis candidate. It is also known to bind both collagen and calcium and is thought to induce collagen mineralisation. Polymorphisms within this gene have been associated with osteoarthritis, suggesting that this gene may be involved in disorders characterised by bone defects.

ASPN is located in a cluster of four linked genes on chromosome 9q22. The other genes within this cluster are Osteomodulin (*OMD*), Osteoglycin (*OGN*) and Extracellular matrix protein 2 (*ECM2*). All four of these genes are extracellular matrix proteins that are members of the small leucine-rich proteoglycan family and are all bone or cartilage extracellular matrix proteins. Like *ASPN, ECM2* is significantly downregulated in the otosclerotic suprastructures based on RPKM calculations, although this is not the case for z score calculations (Figure 5.6). In contrast, *OMD* is significantly downregulated based on z score calculations but not RPKM values. *OGN* does not reach significance in either statistical analysis but regulation appears to be in the same downward direction in otosclerotic stapes. Given that these genes are linked, share similar functions and show a general trend of down-regulation in the otosclerotic stapes, it raises the possibility that they may be involved in the development of disease. In addition, since these four genes are also linked in the mouse, it suggests that this cluster is evolutionarily conserved, indicating that these genes may share various crucial gene functions. For this reason, *ASPN*, the most significantly dysregulated of these four genes and the only one significantly down-regulated based in both RPKM and z score analysis, is an excellent candidate for follow up gene expression studies.

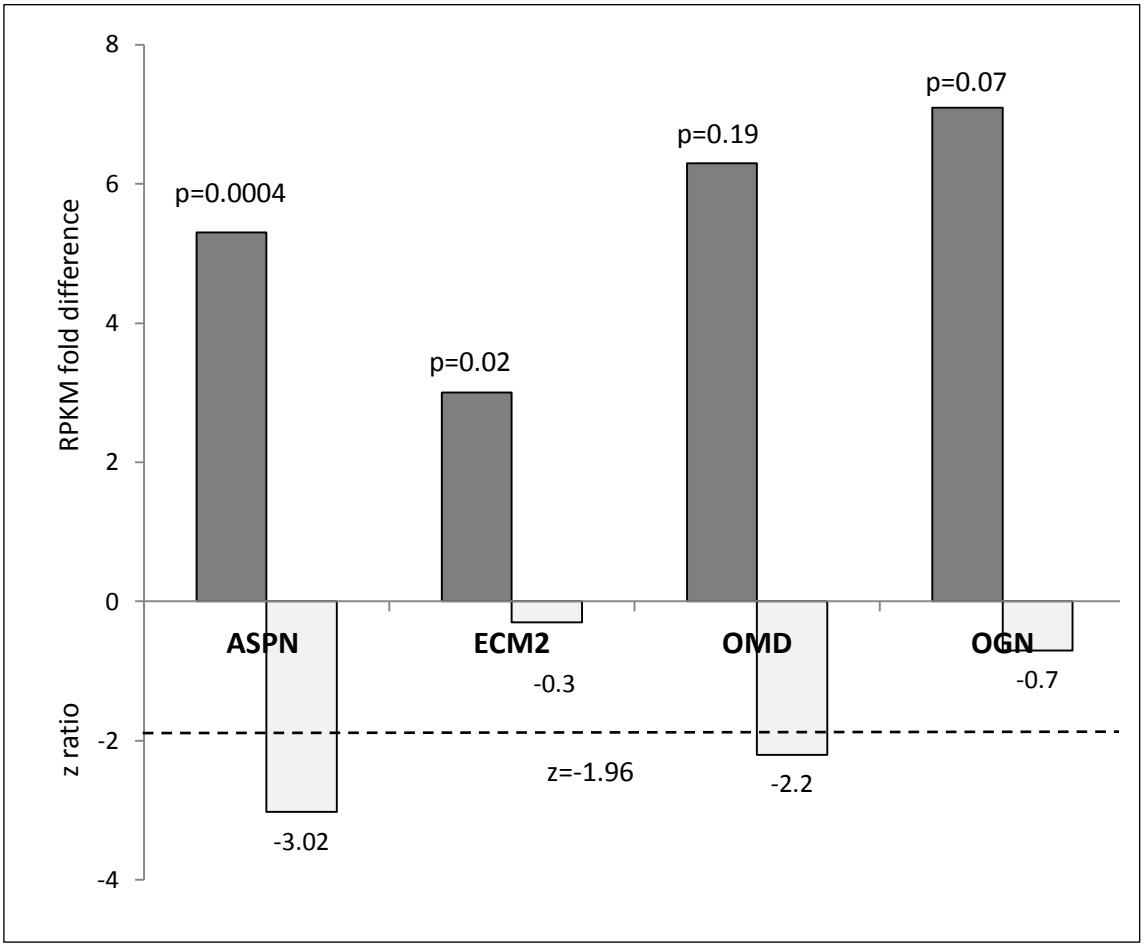


Figure 5.6 Bar graph showing the RPKM fold difference and p values as well as the z ratios calculated for each of the four linked genes located within the cluster on chromosome 9q22

In addition to ASPN, two other leucine-rich repeat extracellular matrix proteins identified during DAVID functional classification analysis, Lumican (*LUM*) and SPARC-like 1 (*SPARCL1*) were selected for follow up. *LUM* is a gene known to regulate collagen fibril organisation and to be secreted by osteoblasts. It is therefore possible that reduced expression of *LUM* in otosclerotic tissue could be caused by abnormal osteoblast function. LUM also has many similarities to another small leucine-rich proteoglycan called Decorin (*DCN*), a component of connective tissue which binds to type I collagen fibrils and plays a role in bone matrix assembly, which was also found to be downregulated in otosclerotic suprastructures based on z score analysis. SPARCL1 is a non-collagenous bone constituent known to function in morphogenesis of the inner ear, thus making it an excellent otosclerosis candidate gene. Both *LUM* and *SPARCL1* were also found to be dysregulated in otosclerotic footplates, giving further evidence to suggest that these genes may be involved in the otosclerosis disease process.

Among the up-regulated genes, FK506 binding protein 14 (*FKBP14*) was identified as a promising candidate gene for follow up qRT-PCR. This gene is known to be involved in accelerated protein folding and has been identified as a causative gene in a specific form of the connective tissue disorder Ehlers Danlos syndrome that is associated with sensorineural hearing loss (Baumann et al., 2012). This is therefore an excellent otosclerosis candidate gene and so was selected for follow up analysis.

Leukocyte-associated immunoglobulin-like receptor 1 (*LAIR1*) was also prioritised for follow up. This encodes a receptor found on the surface of mononuclear cells, a type of cell that differentiates into osteoclasts during bone remodelling. This receptor is known to interact with extracellular matrix collagens in regulation of the immune system. Given that it has been suggested that otosclerosis may be an autoimmune condition and that otosclerosis is characterised by abnormal bone remodelling within the otic capsule, *LAIR1* is also a promising candidate for follow up studies.

Filamin binding LIM protein 1 (*FBLIM1*) was also selected. This gene is involved in actin filament assembly and stabilisation in the cell. Although it doesn't have a role that suggests obvious involvement in otosclerosis pathology, it was selected for follow up due to the extremely low p value calculated during RPKM analysis. Unlike the down-regulated gene list where there were numerous good candidate genes to select from, there were fewer good candidates in the set of up-regulated genes. For this reason *FBLIM1* was selected for follow up.

5.2.7 qRT-PCR assays on the six genes selected for follow up analysis

The expression levels of the three genes, *ASPN, LUM* and *SPARCL1,* that were found to be down-regulated through RNA-seq in 7 otosclerotic and 3 control stapes suprastructures, were analysed in a larger cohort of 81 additional otosclerotic and 5 additional control stapes using qRT-PCR Taqman® assays (Figure **5.7**). The results of these assays showed that *ASPN* was expressed at significantly lower levels in the otosclerotic stapes relative to control samples with the relative quantification (RQ) value for *ASPN* being 0.35, representing an approximately three fold reduction in expression. However, the results showed that *LUM* and *SPARCL1* were expressed at significantly higher levels in otosclerotic compared to control samples in the larger stapes cohort. This finding was unexpected as it showed an opposite trend in expression of *LUM* and *SPARCL1* to that shown in the RNA-seq study.

It was noticed that the cycle threshold (CT) values of all three replicates for one of the control assays for *LUM* were substantially higher than for the other controls, with an average ΔCT

value of 17.98 compared to 12.30 for the remainder of the control replicates. This represents a fold difference of $2^{5.68}$ which is equivalent to a 51.27 fold difference in expression for this sample. Therefore, this outlier was excluded from the analysis (Figure **5.7**). This is discussed in more detail in section 5.2.7.1.

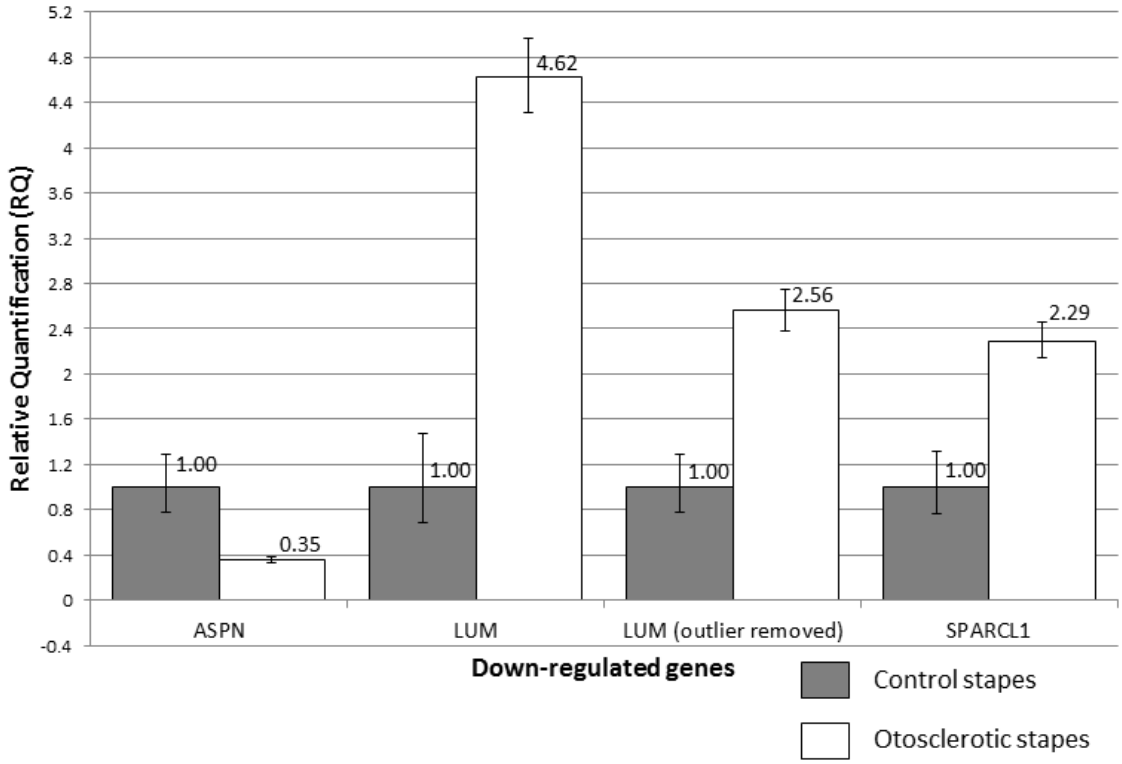


Figure 5.7 Graph showing relative expression of 3 down-regulated genes in control and otosclerotic stapes

ASPN, LUM, LUM with outlier removed and SPARCL1 relative expression in 5 control stapes (grey) compared to 81 otosclerotic stapes (white). 18s ribosomal RNA served as an endogenous control. Error bars represent 95% confidence intervals.

The expression levels of the three genes, *FBLIM1, FKBP14* and *LAIR1,* that were found to be up-regulated through RNA-seq of 7 otosclerotic and 3 control stapes suprastructures, were also analysed in the larger cohort of 81 additional otosclerotic and 5 additional control stapes using qRT-PCR Taqman® assays. However, the *FBLIM1* assay was unable to detect the presence of *FBLIM1* mRNA in the majority of stapes samples (data not presented). This is discussed in detail in section 5.2.7.1. The results of the Taqman® assays for the additional two up-regulated genes showed that *FKBP14* and *LAIR1* were both expressed at significantly lower levels in the otosclerotic stapes relative to control samples (Figure 5.**8**). This trend is in the opposite direction to the trend shown in the data from the RNA-seq study.

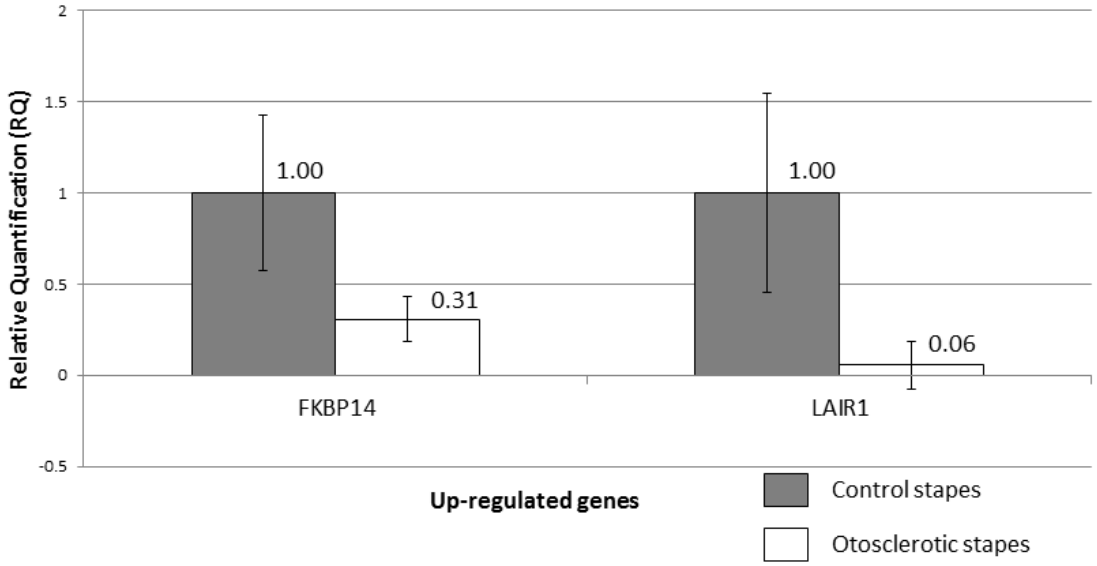


Figure 5.8 Graph showing relative expression of 2 up-regulated genes in control and otosclerotic stapes

FKBP14 and LAIR1 relative expression in 5 control stapes (grey) compared to 81 otosclerotic stapes (white). 18s ribosomal RNA served as an endogenous control. Error bars represent 95% confidence intervals.

5.2.7.1 Discussion of discrepancies between RNA-seq and qRT-PCR data

A likely explanation for the discrepancy in findings between the RNA-seq study and the qRT-PCR analysis is the small number of otosclerotic and control stapes analysed in the RNA-seq study, in addition to the small number of control stapes available for follow up qRT-PCR assays. These small sample sizes reduce the reliability of the findings, particularly in the case of transcripts with variable expression between samples. Since otosclerosis is thought to be a multigenic complex disease, caused by mutations in a variety of different genes as well as a number of environmental contributors, it is possible that different genetic factors are responsible for the condition in different individuals, which could lead to varying patterns of gene expression between stapes. Therefore selecting just a small subset of the stapes for whole transcriptome analysis, may be misleading if these samples are not representative of diseased stapes in general. However, due to cost constraints, it was not possible to analyse large numbers of samples, a common problem encountered when using NGS techniques. This concern was raised at an early stage in the project, for which reason the otosclerotic stapes selected for sequencing were intentionally chosen to be representative of the cohort as a whole. However variability between samples is still likely to be an issue in such a small cohort. Due to the small size of the discovery group, particularly in regards to the control stapes, variability between samples is the most likely explanation for the differences in findings between the two gene expression assays.

This explanation is particularly convincing when explaining the difference in findings between RNA-seq and qRT-PCR assays for *LUM,* as this gene is extremely highly expressed in bone tissue. *LUM* has a mean RPKM value across all stapes suprastructures of 508.5 (Table 5.7) and therefore falls within the top 20 most highly expressed of all 21,563 genes annotated in the RNA-seq study. There is likely to be greater variability in absolute expression levels of more highly expressed genes between samples than lowly expressed genes, it is possible that dramatically different expression of a highly expressed gene in just one sample could skew the data. This was not a concern when performing the qRT-PCR assay on the otosclerotic stapes, as since a large number of samples were analysed, any anomalies would likely be averaged out. However, since only five control stapes were available for analysis, just one control sample could skew the data if it had a different level of expression to the majority of other controls. This theory is supported by the identification of the one control stapes with an exceptionally high CT value for *LUM,* which demonstrates that expression of this gene is variable between samples.

An alternative possible explanation for the discrepancy in results between the RNA-seq study and Taqman® assays in regards to *SPARCL1* is that this gene has numerous splice isoforms. The major transcript *SPARCL1-001* (ENST00000282470) consists of 11 exons, of which the target for the Taqman® assay is located between exons 10 and 11. This means that this major isoform will be detected by the Taqman® assay. However, a number of other known *SPARCL1* isoforms contain an additional exon located between exons 10 and 11. If these isoforms are expressed in stapes tissue, they will therefore not be detected using the Taqman® assay, although would have been using RNA-seq. This could provide a potential alternative explanation for the difference in outcome of the two studies.

An explanation for the lack of detection of *FBLIM1* in the Taqman® assays despite detection during RNA-seq, is that RNA-seq is a more sensitive technique and can therefore detect transcripts expressed at lower levels. This explanation is supported by the fact that *FBLIM1* was detected in just two of the five control and six of the 81 otosclerotic stapes (Table 5.7) during qRT-PCR, suggesting that this gene is expressed at low levels in the human stapes. Furthermore, in the two control samples in which *FBLIM1* was detected, expression was not consistent across all replicates, as it was detected in just one of the three replicates in each case. This indicates that there may not have been sufficient cDNA for accurate detection of this transcript. Furthermore, the average CT value was 38.2 for affected stapes and 38.1 for control stapes further indicating low level expression of this transcript in the human stapes. In addition, *FBLIM1* had the lowest normalised read count of all six genes prioritised for follow up

gene expression assays at just 0.11 across all suprastructures, when the normalised read count data from the RNA-seq study was analysed (Table 5.7). The low level of expression of this gene had been overlooked at earlier stages of analysis during the candidate gene selection process, as the focus had been on the RPKM and z score data rather than the normalised read count data. Since RPKM normalises the read count against the length of the transcript and z score normalises by the total number of reads, these statistics are useful when considering relative expression levels, but not as effective as normalised read count data when considering absolute levels of expression.

Gene	Up or down regulated	Mean normalised read count across all suprastructures	Mean RPKM value across all suprastructures	No. (%) control suprastructures	No. (%) otosclerotic suprastructures
ASPN	Down	1.65	50.8	5 (100%)	78 (96%)
FBLIM1	Up	0.11	14.8	2 (40%)	6 (7%)
FKBP14	Up	0.18	19.4	5 (100%)	64 (79%)
LAIR1	Up	0.12	10.1	5 (100%)	70 (86%)
LUM	Down	14.29	508.5	5 (100%)	81 (100%)
SPARCL1	Down	2.97	88.2	5 (100%)	79 (98%)

Table 5.7 Table showing the normalised read count data from RNA-seq analysis for each of the three up and down-regulated genes selected for qRT-PCR analysis and number and percentage of control and affected stapes samples in which each transcript was detected.

By observing the normalised read count data for all six genes selected for follow up analysis, it can be seen that the other two up-regulated genes *FKBP14* and *LAIR1* also have low mean read counts, especially compared to the three down-regulated genes. In addition, they were also found to be detected in fewer stapes during Taqman® assays than the three down-regulated genes. Whilst *ASPN, LUM* and *SPARCL1* were detected in almost all of the 81 affected stapes, with the exception of two otosclerotic stapes for *ASPN* and one for *SPARCL1*, *FKBP14* was detected in only 64 otosclerotic samples and *LAIR1* in just 70 (Table 5.7). There is an issue associated with non-detection of transcripts during qRT-PCR, as the 7500 System SDS software used to analyse the results of the Taqman® assays assigns a CT value of 0 to samples that do

not meet threshold detection levels. These samples are therefore excluded from the analysis, meaning that only the small number of samples in which the mRNA is detected are analysed. This means that only those samples in which the gene is most highly expressed are included in the analysis, compromising the reliability of the results obtained. In order to tackle this issue, both *FKBP14* and *LAIR1* were re-analysed by assigning a maximum CT value of 46 to all samples in which the transcript was not detected, and the relative quantification values were calculated manually. This ensured that the lowly expressed transcripts, which may have required more rounds of cycling before the florescence signal crossed the detection threshold, were not omitted from the analysis. For this reason, all non-detected transcripts were assigned a CT value of 46, making the assumption that these transcripts may have been detected, had more than 45 rounds of cycling been performed. This enabled the findings to be re-assessed in order to increase the reliability of the results by ensuring inclusion of a greater number of samples. However, the re-analysis of the data did not cause any substantial changes to the RQ value with a decrease from 0.15 to 0.11 for *FKBP14* and an increase from 0.06 to 0.07 for *LAIR1*.

Since it is likely that any gene involved in the otosclerosis disease process, will be expressed in human stapes, genes that are not expressed or are expressed at very low levels in this tissue are not likely to be good pathologic candidates. In hindsight, it therefore seems that the three up-regulated genes selected for follow up assays were perhaps not the best choice, and that normalised read count data and absolute difference in expression levels between otosclerotic and control samples, in addition to fold difference, should have been considered when making the candidate gene selection.

5.3 Discussion of RNA-seq

In this study, the entire transcriptomes of 8 otosclerotic and 4 control stapes samples were sequenced using Illumina HiSeq 2000 paired end RNA-seq. Statistical analysis was performed to distinguish significant changes in gene expression between the diseased and control stapes suprastructures, providing stapes gene expression datasets for both otosclerotic and healthy stapes. This is the first study to perform transcriptome-wide NGS on human stapes and thus provides a unique dataset upon which to base future studies. The number of otosclerotic stapes analysed during this RNA-seq study are comparable to the most comprehensive published otosclerosis gene expression study conducted to date (Ealy et al., 2008) which employed a whole genome hybridisation array technique, and has the additional benefit over this previous study of being able to detect and quantify previously uncharacterised transcripts

and those expressed at low levels. Furthermore, the follow up qRT-PCR analysis performed in 81 additional otosclerotic stapes following this RNA-seq study was much larger than the nine otosclerotic stapes used for validation of the array-based study.

Pathway and functional analysis using IPA and DAVID showed that the dataset was enriched for genes involved in pathways including immune cell trafficking and endocrine system development. These findings support the somewhat contradictory evidence in the published literature which suggests that viral infection and hormones such as oestrogen may be involved in otosclerosis pathogenesis. The analysis also showed that the dataset was enriched for genes encoding connective tissue proteins. Since otosclerosis is a connective tissue disorder and such disorders often occur concomitantly with other connective tissue disorders, these genes represent good pathologic candidates in otosclerosis. The dataset was also enriched for genes encoding regulatory proteins such as transcription factors and zinc fingers. Since regulatory proteins are known to be involved numerous genetic disorders, these findings could indicate a role for regulatory proteins in the disease process. However, the results from pathway analysis and functional annotation should not be over-stated due to limited knowledge on the complex nature of interacting gene networks. Despite this, the results of the pathway and functional annotation analysis were taken into consideration during the selection of candidate genes of interest for follow up analysis.

A difficulty encountered during the candidate gene selection process was distinguishing between dysregulated genes that were likely to be involved in disease pathology, and genes that were dysregulated as a result of the disease process. It is likely that in disease tissue, there will be local dysregulation of genes that act as a pathogenic switch in initiating the change that leads to onset of symptoms, as well as genes whose expression is altered as a result of pathogenicity. The gene annotation stage for dysregulated genes was designed in an attempt to overcome this obstacle with the aim of prioritising those genes most likely to be involved in otosclerosis pathophysiology. However, it is possible that due to this, genes that are disease markers rather than causative factors were among those selected for follow up studies.

A further limitation of the candidate gene selection process was the strategy involved in the selection of these genes for follow up analysis. Following annotation of the 51 down-regulated genes, a large number of good candidates emerged based on the biological role of the gene combined with statistical analysis of both suprastructure and footplate samples, enabling selection of three promising otosclerosis candidates for follow up. However, when the 50 up-regulated genes were annotated, there were fewer good candidates to select for follow up. As a result, the three best possible choices were selected, however following analysis, when it

emerged that these genes selected were all expressed at very low levels, it became apparent that these had probably not been the best choice. In hindsight, it may have been more beneficial to have selected the six top candidates from the dysregulated gene set as a whole, not taking into consideration if genes were up or down-regulated. Alternatively, perhaps genes should have been selected based solely on statistical analysis findings and the gene functions should not have been considered. This would have resulted in a less biased approach to gene selection. In addition, it would have been beneficial to take into account the normalised read count data in addition to the RPKM data to identify genes that were expressed at high enough levels to be efficiently detected using qRT-PCR. Absolute difference in expression in addition to fold difference could also have been considered during the selection process. Whilst absolute difference was initially dismissed as a useful parameter due to concerns that it could discriminate against lowly expressed genes that may have been of interest such as transcription factors, in hindsight, it may have been useful to consider, as excluding it from the prioritisation process meant that three very lowly expressed genes were selected for follow up. If further genes were to be selected from the RNA-seq dataset for follow up in the future, it would be useful to take all factors into account to ensure that a balance is obtained between absolute and fold difference in expression.

The relative expression levels of the three up and three down-regulated genes selected from the dysregulated gene set were analysed using qRT-PCR in a follow up cohort of 81 otosclerotic and 5 control stapes suprastructures collected at the Royal National Throat Nose and Ear Hospital in London. The relative expression levels of one of the three down-regulated genes, *ASPN,* was reduced in the otosclerotic stapes compared to controls in the larger stapes cohort, supporting the results from the RNA-seq study and suggesting a possible role for this gene in disease pathogenesis. However, the results from the RNA-seq study were not replicated in any of the other five genes selected for follow up analysis. The most likely explanation for this is the small size of the discovery group used for RNA-seq and the lack of available control stapes for analysis. This is a problem that was also encountered in the whole genome hybridisation array study which analysed gene expression in nine otosclerotic and seven control stapes (Ealy et al., 2008). During follow up analysis for this study, only two of seven dysregulated genes prioritised for validation were found to be dysregulated. The issue of non-validation during follow up analysis is therefore not unique to this RNA-seq study and has been encountered by other research groups investigating gene expression in otosclerotic stapes. It is likely that this is due to the small cohort sizes of the discovery groups used in both these studies, meaning that a number of possible dysregulated genes of interest may not have reached statistical

significance. Genes of interest may therefore have been overlooked in the datasets, leading to the selection of less desirable candidates for follow up analysis.

Variability in expression levels of genes between samples is another crucial contributing factor to the lack of consistency between the results of the RNA-seq study and the qRT-PCR assays. It would have been desirable to perform Taqman® assays on the same stapes samples that had undergone RNA-seq to ensure that results from both assays were consistent. However, this was not possible because the cDNA samples sent for RNA-seq had been reverse transcribed using the highly sensitive SMARTer™ kit (Clontech) rather than using the Qiagen Omniscript Reverse Transcription kit, which was used for the samples analysed using Taqman® assays. cDNA processed using the Clontech kit is not suitable for Taqman® assays because 18s ribosomal RNA is not polyadenylated so is not reverse transcribed during the Clontech reverse transcription process. It is therefore not suitable for use as an endogenous control during the Taqman® assay. If it were possible to perform Taqman® assays on the same samples as the RNA-seq, it would be possible to evaluate whether or not the discrepancy in findings was due to differing patterns of expression between stapes.

One of the greatest advantages of the data obtained during the RNA-seq study for this research project as a whole has been its value in regards to the WES study (chapter 4). The identification of dysregulated genes in the otosclerotic relative to control stapes during RNA-seq has provided an invaluable tool to assist with prioritisation of candidate variants following WES. After WES, rare nonsynonymous variants identified in families exhibiting monogenic inheritance of otosclerosis were prioritised for segregation analysis if located within a gene found to be dysregulated in otosclerotic stapes during the RNA-seq study. This was extremely useful in highlighting good pathologic candidates in these families, in particular in family B, in which the stapes of the proband was one of those sequenced.

In summary, this study has shown that RNA-seq can be used to facilitate the identification of genetic factors involved in complex disorders such as otosclerosis. However, this technique is limited by the high costs involved, which limits the number of samples that can be analysed and thus reduces the reliability of findings. Care must therefore be taken when selecting appropriate genes for follow up analysis, particularly in the case of lowly expressed genes and any findings must be followed up in a larger cohort of stapes using more cost-effective approaches. Despite the limitations associated with this technique, RNA-seq is an effective tool that can provide valuable information to elucidate the disease pathways involved in otosclerosis and in particular to help prioritise genes of interest for future research.

6 Investigating reelin (RELN) expression in otosclerotic stapes

6.1 Introduction

Reelin (RELN) is a an extracellular matrix protein involved in neuronal positioning during brain development (Quattrocchi et al., 2002), which was found to be associated with otosclerosis during a genome wide association study (Schrauwen et al., 2009a). Analysis of variants within 694 Belgian-Dutch cases and controls, as well as a replication cohort of 455 French cases and 480 controls, revealed a significant association signal that reached genome wide significance in the *reelin* (*RELN)* gene, spanning 180kb from intron 1 to 4 of this 65 exon gene. The association of this gene with otosclerosis was confirmed in a study of four additional European populations (Schrauwen et al., 2010a) as well as a study in a Tunisian population, which also suggested an interaction with sex (Khalfallah et al., 2010). However, the association was not replicated in an Indian population (Priyadarshi et al., 2010). The association with *RELN* was unexpected as there is no obvious role for the reelin protein, a large 388kDa protein 3,460 amino acids in length, in the pathogenesis of otosclerosis.

Subsequent expression studies have produced conflicting evidence regarding presence of the *RELN* transcript in human stapes tissue. Schrauwen et al., in their genome wide association study, successfully detected *RELN* transcripts in 1 of 2 human stapes footplate samples through semi-quantitative RT-PCR assays (Schrauwen et al., 2009a). However, a study published in 2011 by Csomor et al. did not detect *RELN* mRNA in 47 stapedial footplates using the same technique (Csomor et al., 2011). Despite absence of *RELN* mRNA, Csomor et al. did detect the reelin protein in stapes footplates using immunohistochemistry (Csomor et al., 2011). Together with the fact that *RELN* is an unlikely candidate for an otosclerosis gene, evidence of a lack of expression in disease tissue raised concerns that the genetic link to *RELN* may be spurious.

In this chapter, gene expression of *RELN* within human and mouse stapes bones has been investigated using a combination of Quantitative real-time PCR (qRT-PCR), Semi-quantitative Reverse Transcription PCR and RNA-seq, in order to clarify this discrepancy in the literature.

6.2 Results

6.2.1 Selection of stapes for gene expression analysis

The expression of *RELN* transcripts was investigated in stapes suprastructures from 78 otosclerosis patients and 8 unaffected individuals, in order to clarify the conflicting reports on its expression in the published literature, and to investigate whether there was any difference in the expression of *RELN* in the stapes between individuals with otosclerosis and unaffecteds. The stapes cDNA used in this study included 71 otosclerotic and 5 unaffected control suprastructure samples for PCR assays and a further 7 otosclerotic and 3 control suprastructure samples as well as 1 otosclerotic and 1 control footplate sample for RNA-seq. Expression was investigated in both suprastructure and footplate samples to investigate whether there was any difference in expression of *RELN* between these two distinct parts of the stapes bone, which are believed to have different embryonic origins. See Table 6.1 for a summary of stapes analysed in this study.

Species	Stapes Samples	Total no. Samples	Gene Expression Assay	No. Samples Used in Assay
Human	Control Suprastructures	8	PCR assays	5
			RNA-seq	3
	Control Footplates	1	PCR assays	0
			RNA-seq	1
	Otosclerotic Suprastructures	78	PCR assays	71
			RNA-seq	7
	Otosclerotic footplates	1	PCR assays	0
			RNA-seq	1

Table 6.1 Table showing the total number of human stapes suprastructure and footplate samples used for PCR assays and RNA-seq

6.2.2 An alternatively spliced variant of *RELN* is detected in human stapes

RELN is a large gene located on chromosome 7q21 which is composed of 65 exons. Two known protein coding isoforms exist in humans including the major full length *RELN* transcript ENST00000428762 (*RELN-001*) and an alternatively spliced *RELN* transcript ENST00000343529 (*RELN-002*), which lacks the penultimate 6 nucleotide micro exon, exon 64. In addition, it has been proposed that a prematurely truncated *RELN* transcript exists in human tissue (Lambert de Rouvroit et al., 1999). In order to identify which of these isoforms were expressed in human stapes suprastructures, sequence-specific primers were designed to amplify each of two confirmed human *RELN* isoforms independently of one another using Semi-quantitative Reverse Transcription PCR. To amplify the major *RELN* transcript (*RELN-001*), the forward primer (A) was designed to span exons 63 and 64 with the 3' end binding to exon 64 which is unique to the full-length transcript, whilst the reverse primer (B) was designed to bind to exon 65 (Figure 6.1). To amplify the alternatively spliced *RELN* transcript (*RELN-002*), the forward primer (C) was designed to bind to exon 62 while the 5' end of the reverse primer (D) was designed to bind to exon 65 with the 3' end binding to exon 63, thus skipping the excised exon 64 (Figure 6.1). Primers were also designed to amplify the prematurely truncated *RELN* transcript. Primer C was used as the forward primer, while the reverse primer (E) was designed to bind only to alternative exon 63a, which exists only in this isoform (Figure 6.1). *GAPDH* primers were also designed to confirm integrity of the cDNA samples. The primer sequences can be found in Chapter 2 Materials and Methods table 2.2.

Subsequently, qRT-PCR was performed on the human stapes suprastructures in order to identify if there was any difference in the level of gene expression between the affected and control stapes. A Taqman® gene expression assay (Hs01022646-Applied Biosystems®) which bound to the exon-exon junction between exons 62 and 63 was performed, as the location of the target ensured that all three *RELN* isoforms would be detected if present (Figure 6.1). The Taqman® assay was carried out in triplicate for each otosclerotic sample, whilst six replicates were performed on cDNA from the five control samples.

All PCR assays were performed on cDNA extracted from 71 otosclerotic and 5 control stapes suprastructures as well as on cDNA extracted from MG-63 cells, a human osteosarcoma cell line which served as a positive control for each PCR assay to demonstrate successful amplification of the desired DNA fragment.

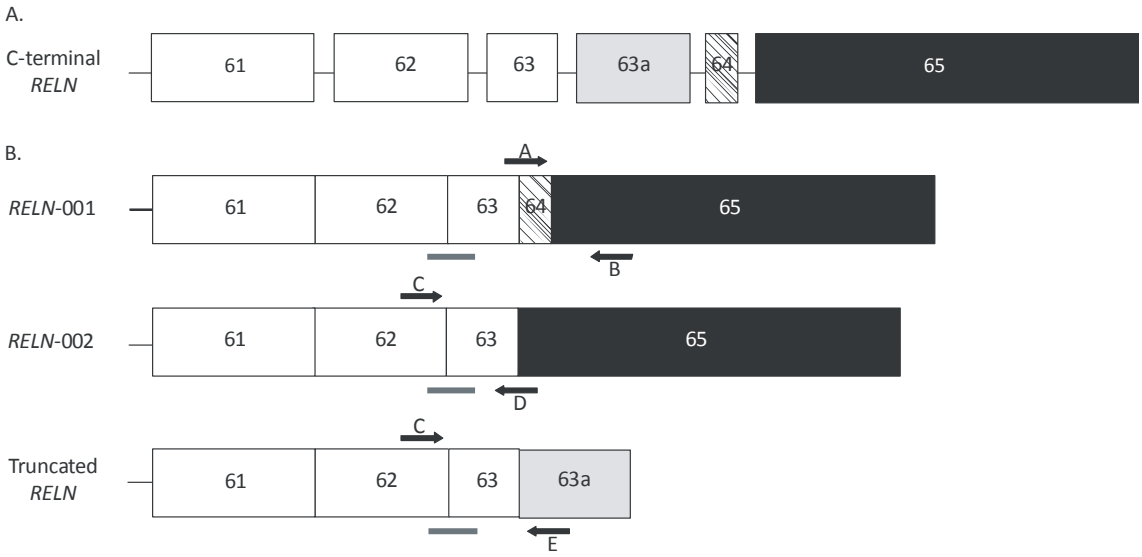


Figure 6.1 Structure of the C-terminal region of the *RELN* gene and its three mRNA isoforms illustrating the positions of the Taqman® gene expression assay amplicon and isoform-specific primers

A) Genomic exon structure of the C-terminus of the *RELN* gene (not to scale). Isoform-specific exons 63a and 64 are indicated by grey and hatched shading respectively. Exon 65 which is absent from the truncated isoform is indicated by black shading.

B) Structure of the three RELN mRNA isoforms; the major isoform *RELN-001*, an alternatively spliced isoform *RELN-002* from which exon 64 is absent, and a prematurely truncated transcript containing an isoform specific terminal exon 63a. Location of the Taqman® assay amplicons are illustrated in grey below the transcript. Arrows indicate forward and reverse isoform-specific RT-PCR primers. A and B indicate the forward and reverse primers which can only amplify *RELN-001*, C and D indicate the forward and reverse primers which amplify *RELN-002*, E indicates the reverse primer which along with forward primer C can amplify only the truncated variant.

Through semi-quantitative RT-PCR assays, the alternatively spliced *RELN* isoform, lacking the 6 nucleotide micro exon (*RELN-002*) was detected in 43 of the 71 otosclerosis and 4 of the 5 control samples (see Figure 6.2 for a representative sample). Of the otosclerosis samples, *RELN-00*2 was detected in 32 out of 52 female samples and 11 out of 19 male samples. Although both *RELN* isoforms were detected in the MG-63 cells, the major full length RELN transcript (*RELN-001*) was not detected in any of the 71 otosclerosis or 5 control human stapes suprastructures. This indicates that whilst both transcripts are expressed in the osteosarcoma cell line, only the alternatively spliced transcript, *RELN-002*, is expressed in human stapes. There does not appear to be any difference in expression of this isoform between the genders. In addition, the prematurely truncated *RELN* transcript was not detected in any of the human stapes samples or MG-63 cells, indicating that this predicted transcript may not be expressed in bone tissue. Human *GAPDH* cDNA was detected in all human stapes suprastructures and MG-63 cells, confirming successful RNA purification and reverse transcription in all samples.

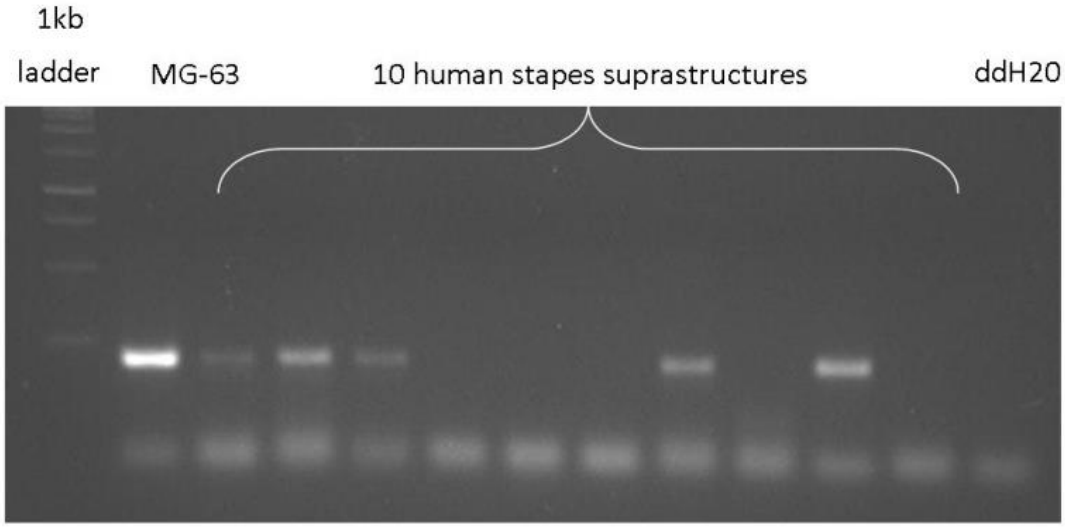


Figure 6.2 A 2% agarose gel showing *RELN* transcript variant 2 RT-PCR amplicons in human stapes cDNA.

RELN-002 was detected in MG-63 cells and 5 out of 10 human stapes suprastructures (a representative cohort sample), illustrated by a 209 bp fragment.

Although the *RELN-002* PCR product was detected in a large number of the stapes suprastructures, the faint bands visible on the gel relative to the brighter band detected in the MG-63 sample suggests that this transcript is expressed at low levels in human stapes. This indicates that the semi-quantitative PCR assay may lack the sensitivity to detect this transcript efficiently. The results from the qRT-PCR assay confirm that *RELN* is expressed at significantly lower levels in the stapes relative to MG-63 cells (Figure 6.3) and also suggests that expression of *RELN* is significantly reduced in otosclerotic stapes compared to controls with detection of *RELN* in 39 out of the 71 otosclerotic stapes compared to 4 out of the 5 controls (Figure 6.3). The CT values in human stapes were high in comparison to those in MG-63 cells, with an average CT of 31.2 for MG-63 cDNA, 36.6 for otosclerotic stapes cDNA and 36.7 for control stapes cDNA. Overall, the proportion of stapes in which *RELN* was detected in qRT-PCR assays is comparable to the proportion of stapes in which *RELN-002* was detected through the semi-quantitative PCR assays. This suggests that it is likely that the qRT-PCR assay is detecting solely *RELN-002*. The outcome of all PCR assays is summarised in Table 6.2.

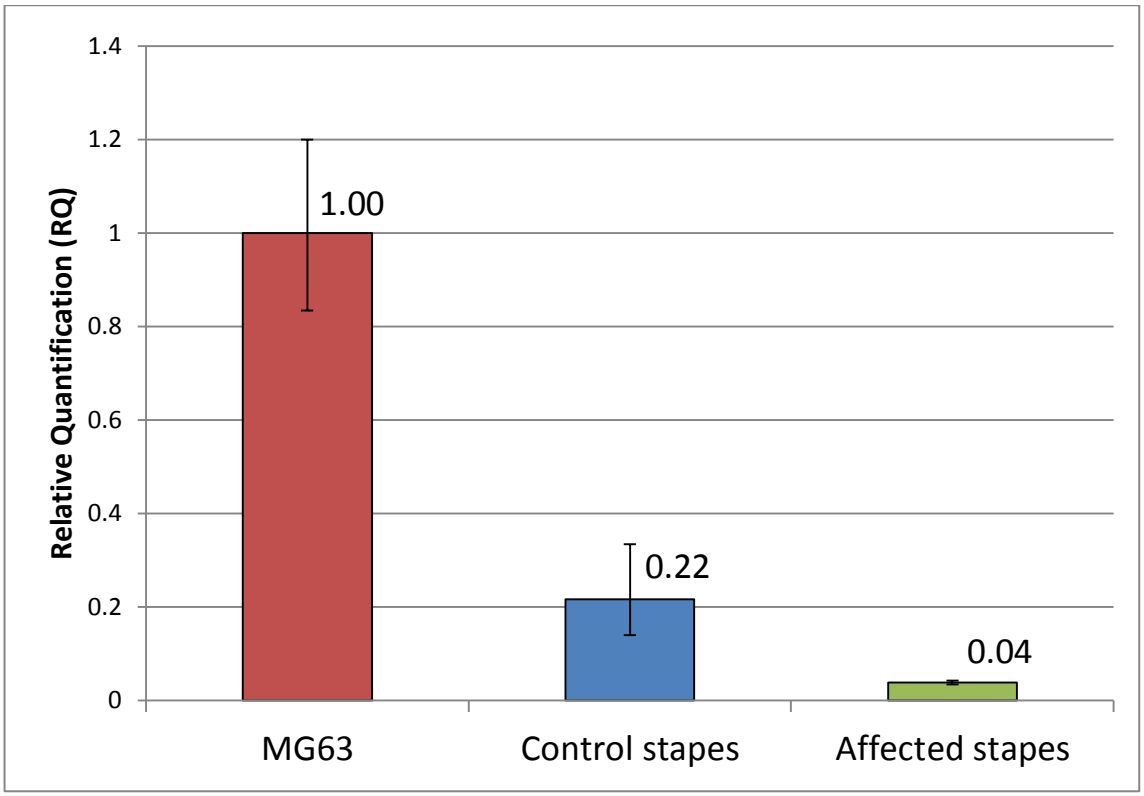


Figure 6.3 *RELN* relative quantification graph.

Relative expression of *RELN* in MG-63 cells (red), 5 control stapes (blue) and 71 otosclerotic stapes (green) shows that *RELN* is more highly expressed in MG-63 cells compared to stapes and is more highly expressed in control than otosclerotic stapes. 18s ribosomal RNA served as an endogenous control. Error bars represent 95% confidence intervals.

	No. (%) of samples in which transcript is detected				
	Semi-quantitative RT-PCR				**Quantitative Real Time PCR**
Sample	***GAPDH***	***RELN-001***	***RELN-002***	***Truncated RELN***	
MG-63 cells (1)	1 (100%)	1 (100%)	1 (100%)	0 (0%)	1 (100%)
Female stapes suprastructures (52)	52 (100%)	0 (0%)	32 (62%)	0 (0%)	29 (56%)
Male stapes suprastructures (19)	19 (100%)	0 (0%)	11 (58%)	0 (0%)	10 (53%)
Control suprastructures (5)	5 (100%)	0 (0%)	4 (80%)	0 (0%)	4 (80%)
Total stapes suprastructures (76)	76 (100%)	0 (0%)	47 (62%)	0 (0%)	43 (57%)

Table 6.2 Table showing *RELN* transcript detection in human stapes

Table showing the number of stapes suprastructure samples and MG-63 cells in which each of the three *RELN* isoforms was detected using semi-quantitative and qRT-PCR assays and the results from a *GAPDH* assay confirming cDNA integrity.

Detection of *RELN* was not consistent across the three replicates performed for each stapes sample. Whereas *RELN* was detected in all replicates in MG-63 cells, it was detected in all replicates in only 8 of the 43 suprastructures that were positive for *RELN* expression in qRT-PCR assays (Table 6.3).

Samples in which *RELN* was detected in qRT-PCR assays	**No. replicates in which *RELN* was detected**		
	1 replicate	**2 replicates**	**All replicates**
MG-63 cells (1)	0	0	1
Female stapes suprastructures (29)	8	14	7
Male stapes suprastructures (10)	8	2	0
Control suprastructures (4)	1	2	1
Total stapes suprastructures (43)	17	18	8

Table 6.3 Table showing the number of replicates in which *RELN* was detected in qRT-PCR assays in human stapes suprastructures as well as MG-63 cells

6.2.3 Expression of *RELN* in human stapes is confirmed by RNA-seq

RELN mRNA was detected in all 12 human stapes samples through RNA-seq (chapter 5), confirming the expression of this gene in human stapes. However, the expression of *RELN* transcripts was very low, with a mean RPKM (reads per kilobase of transcript per million mapped reads) value of just 0.058 for control stapes and 0.060 for otosclerotic stapes (Table 6.4). This is approximately 100 times fewer reads than the mean RPKM value of 5.421 across all 21,563 annotated genes. RNA-seq was unable to detect the presence of the major *RELN-001* isoform in human stapes tissue, as exon 64 which is unique to this transcript was not detected in any of the stapes samples. Although it was difficult to detect the presence of specific isoforms due to the low read count for *RELN* in the RNA-seq data, both exons 63 and 65 were detected in the control stapes footplate with reasonably high read counts of 1.999 and 0.999 respectively.

Control human stapes								
Stapes ID	ConStap1L&R FP	ConStap1L&R SS	ConStap2 SS	ConStap3 SS				
RPKM	0.04032	0.07157	0.05982	0.05957				
Mean RPKM	0.058							
Otosclerotic human stapes								
Stapes ID	OtoStap1 FP	OtoStap1 SS	OtoStap2 SS	OtoStap3 SS	OtoStap4 SS	OtoStap5 SS	OtoStap6 SS	OtoStap7 SS
RPKM	0.04968	0.06389	0.05276	0.04973	0.09777	0.05587	0.05327	0.04573
Mean RPKM	0.060							

Table 6.4 Table showing RPKM values from RNA-seq for *RELN* in four control and eight otosclerotic stapes

6.3 Discussion of expression of *RELN* in human stapes

6.3.1 Resolving controversies in *RELN* expression in otosclerosis

Reelin is a large extracellular matrix protein that plays an important role in brain development and function. Despite no obvious role for this protein in the development of conductive hearing loss due to otosclerosis, strong evidence for an association between variants in *RELN* and otosclerosis have been reported (Schrauwen et al., 2009a). The *RELN* gene reportedly exists in three alternative isoforms which are conserved across species, which suggests functional importance (Lambert de Rouvroit et al., 1999). In this study, the expression of *RELN* isoforms in human stapes were analysed to investigate a discrepancy reported regarding detection of *RELN* transcripts in stapes tissue. Whilst Schrauwen et al., reported successful detection of *RELN* transcripts in 1 of 2 human stapes footplates in their 2009 research (Schrauwen et al., 2009a), a 2011 study by Csomor et al. failed to detect the *RELN* transcript in a much larger study (Csomor et al., 2011). In designing transcript specific primers for this study, it was observed that PCR primers designed by Schrauwen et al and Csomor et al. would be predicted to have different transcript specificities. Primers used by Csomor et al. would exclusively amplify the major *RELN* transcript (*RELN-001*), as the 3' end of the oligonucleotide binds to exon 64 which is unique to this isoform. In contrast, the primers used by Schrauwen et al. would detect all *RELN* transcripts if present in the tissue (Figure 6.4).

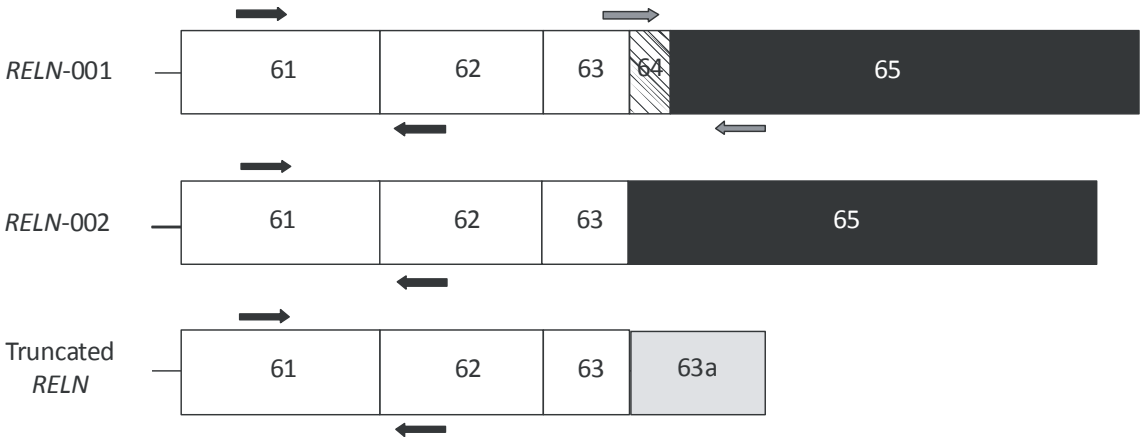


Figure 6.4 Structure of the three *RELN* mRNA isoforms, illustrating the positions of primers designed by Schrauwen et al. and Csomor et al.

Structure of the three RELN mRNA isoforms; the major isoform *RELN-001*, an alternatively spliced isoform *RELN-002* from which exon 64 is absent, and a prematurely truncated transcript containing an isoform specific terminal exon 63a. Isoform-specific exons 63a and 64 are indicated by grey and hatched shading respectively. Exon 65 which is absent from the truncated isoform is indicated by black shading. Black arrows indicate the forward and reverse primers designed by Schrauwen et al. that can amplify all three *RELN* isoforms. Grey arrows indicate the forward and reverse primers designed by Csomor et al. which can only amplify the major transcript *RELN-001*.

Therefore a possible explanation for the discrepancy in findings between the two research groups is that only the alternatively spliced *RELN*-002 transcript or the truncated *RELN* transcript is present in human stapes samples. Given the primer positions selected by each group, only the assay used by Schrauwen et al. would be able to detect *RELN-002* and the truncated *RELN* isoform in their RT-PCR assays. This would also explain why Csomor et al. were able to detect the reelin protein in human stapes in the absence of *RELN* mRNA detection. The results presented here from much larger numbers of human stapes suprastructures support the theory that only the alternatively spliced *RELN-002* isoform is expressed in human stapes.

Transcript specific RT-PCR assays and RNA-seq failed to detect the presence of the major *RELN*-001 isoform and the truncated isoform in human stapes but were successful in the detection of the alternatively spliced isoform *RELN*-002 lacking the penultimate 6 nucleotide micro exon 64 encoding the amino acids Val-Ser. Whilst little is known about the impact of this two amino acid sequence on protein function, it is though that the C-terminal splicing events in *RELN* may affect downstream signalling (Nakano et al., 2007). Hence, it can be concluded that isoform specificity is the most likely explanation for the discrepancy between previous studies. However, it must also be taken into account that Csomor et al. used stapes footplates from otosclerosis patients while Schrauwen et al. used control stapes footplates from individuals undergoing translabyrinthine surgery for vestibular schwannoma, which could result in differences in gene expression due to the differing disease status of the tissue. Alternatively, it

is possible that the discrepancy is due to the variability in expression of *RELN* between individual stapes, an explanation which has been previously proposed by other research groups and which is supported by the results of this study.

It must also be considered that whilst the RT-PCR assays conducted by Schrauwen et al. and Csomor et al. used human stapes footplates, in this study the stapes analysed by RT-PCR originated from suprastructure tissue. Laser-assisted stapedotomy is the preferred surgical technique in the UK for treatment of conductive otosclerosis, which results in removal of only the suprastructure and not footplate, meaning that in the UK it is difficult to gain access to stapes footplate samples for analysis. Since the primary origin of otosclerotic lesions is postulated to be the stapes footplate rather than suprastructure, differences in gene expression could exist between the suprastructure and footplate. For this reason, analysis of RNA-seq data from the one otosclerotic and one control footplate collected was performed. *RELN* was detected in both footplate samples, and although expression levels in the otosclerotic sample were too low to definitively detect the presence of specific isoforms, in the control footplate, the alternatively spliced transcript was detected, whilst the full-length transcript was not. This indicates that the alternatively spliced transcript is expressed in human stapes footplates as well as suprastructures. Whilst the absence of the full length major transcript in just one control footplate is not sufficient to rule out the possibility that this transcript may be expressed in stapes footplates, it indicates that there is some concordance of expression of *RELN* between footplate and suprastructures, although more footplates would need to be collected to confirm this. For this reason, a recent collaboration has been established between the research groups of Dr Sally Dawson at UCL Ear Institute in London and Dr Athanasios Bibas at the National and Kapodistrian University of Athens to initiate recruitment of Greek otosclerosis patients, as in Greece, the stapedectomy is the preferred surgical procedure for treatment of otosclerosis, where the entire stapes, including the footplate is removed. By gaining access to larger numbers of stapes footplates, the expression patterns between footplates and suprastructures can be analysed and the suitability of using suprastructures for this type of analysis can be determined.

6.3.2 A critical analysis of *RELN* expression data

Through the use of three different gene expression assays; Quantitative real-time PCR, Semi-quantitative Reverse Transcription PCR and RNA-seq, *RELN* mRNA was successfully detected in human stapes suprastructures. In both types of PCR assay, there appeared to a reduction in expression of *RELN* in otosclerotic stapes compared to controls. However, since only 5 control stapes were available for analysis, larger number of control stapes would need to be collected in order to support this. The results from all three methods also suggested that this gene is expressed at very low levels within human stapes and that there is considerable variation between stapes samples and between replicates for a single stapes sample. The high CT values in the human stapes, which are at the limit of detection, may account for the inconsistency between replicates, since transcripts expressed at low levels are less likely to meet the threshold for detection. However, it is also possible that it may be due to contamination of samples with *RELN* cDNA, or due to pipetting error which may have led to variability in the amount of cDNA pipetted into each well and thus a difference in detection of *RELN* between the replicates. Whilst these are both plausible explanations, the lack of variability between replicates in the MG-63 assay, which had a lower CT value than human stapes, suggests that this is not the case. Evidence suggesting that the variability is due to low level expression can be seen when the results from the semi-quantitative and quantitative PCR assays are compared. This is because there is consistency between the two assays as to which stapes samples *RELN* was detected in, indicating that the results from both assays support one another. In total, *RELN* was detected in the qRT-PCR assay, as well as in the semi-quantitative PCR assay, in 34 out of 76 suprastructures. *RELN* was detected solely in the qRT-PCR assay in 9 stapes samples and only in the semi-quantitative PCR assay in 12 samples. *RELN* was not detected in 21 of the stapes samples in either of the assays. Therefore, the results of both assays concur in 55 out of the 76 samples, representing consistency between the assays of 72% (Table 6.5).

Assay in which *RELN* was detected	Female stapes suprastructures (52)	Male stapes suprastructures (19)	Control suprastructures (5)	Total stapes suprastructures (76)
RELN detected in qRT-PCR and semi-quantitative PCR assay	22	8	4	34
RELN detected in only qRT-PCR assay	7	2	0	9
RELN detected in only semi-quantitative PCR assay	9	3	0	12
RELN detected in neither qRT-PCR or semi- quantitative PCR assay	14	6	1	21
Expression of *RELN* consistent between qRT-PCR and semi-quantitative PCR assay	36 (69%)	14 (74%)	5 (100%)	55 (72%)

Table 6.5 Table showing the number of samples in which detection of *RELN* was consistent in both qRT-PCR assays and semi-quantitative PCR assays in human stapes suprastructures

The argument that low expression of *RELN* is causing the inconsistency between replicates is further supported by the fact that there is a strong correlation between the samples in which *RELN* is detected using qRT-PCR and the quantity of cDNA present in each stapes sample as measured by the 18s endogenous control assay. The average 18s cycle threshold (CT) value for the samples in which *RELN* was detected in one or more replicates was 21.87, whereas for samples in which it was not detected in any replicates, the CT value was 23.01. This difference is highly significant (p=0.00005), suggesting that the reason for the inconsistency in replicates is due to the lack of sensitivity of the assay as a result of the low expression of the *RELN* transcript in human stapes. If it is considered that due to low levels of expression, detection of *RELN* in at least one of the PCR assays is sufficient evidence to suggest that it is expressed in the stapes, it can be determined that *RELN* is expressed in 73% of female otosclerotic stapes, 68% of male otosclerotic stapes and 80% of control stapes. This suggests that *RELN* is expressed in the majority of stapes but that expression is variable between both affected and control individuals. This is consistent with the findings from previous research groups, where conflicting reports have been published as to whether or not *RELN* is expressed in human stapes, and where detection of *RELN* mRNA varies between stapes samples.

6.3.3 Investigating a link between *RELN* expression and gender

Variation in isoform specific expression of *RELN* between males and females was also investigated, as there is a reported interaction with gender in the prevalence of otosclerosis and in the genetic effect of *RELN* (Khalfallah et al., 2010). In addition, variation in the *RELN* gene has also been linked to risk of schizophrenia and a recent study suggested that women of specific risk genotype (SNP rs7341475 *GG*), experience a higher proportion of micro exon skipping in brain tissue than those with alternative genotypes, while men showed the opposite trend (Ovadia and Shifman, 2011). SNP rs7341475 is located within intron 4 of *RELN* and is in strong linkage disequilibrium with a number of SNPs which showed significant association with otosclerosis in the genome wide association study (Schrauwen et al., 2009a)(Figure 6.5). However, no evidence was found in this study to support a difference in expression of this alternatively spliced isoform between the genders.

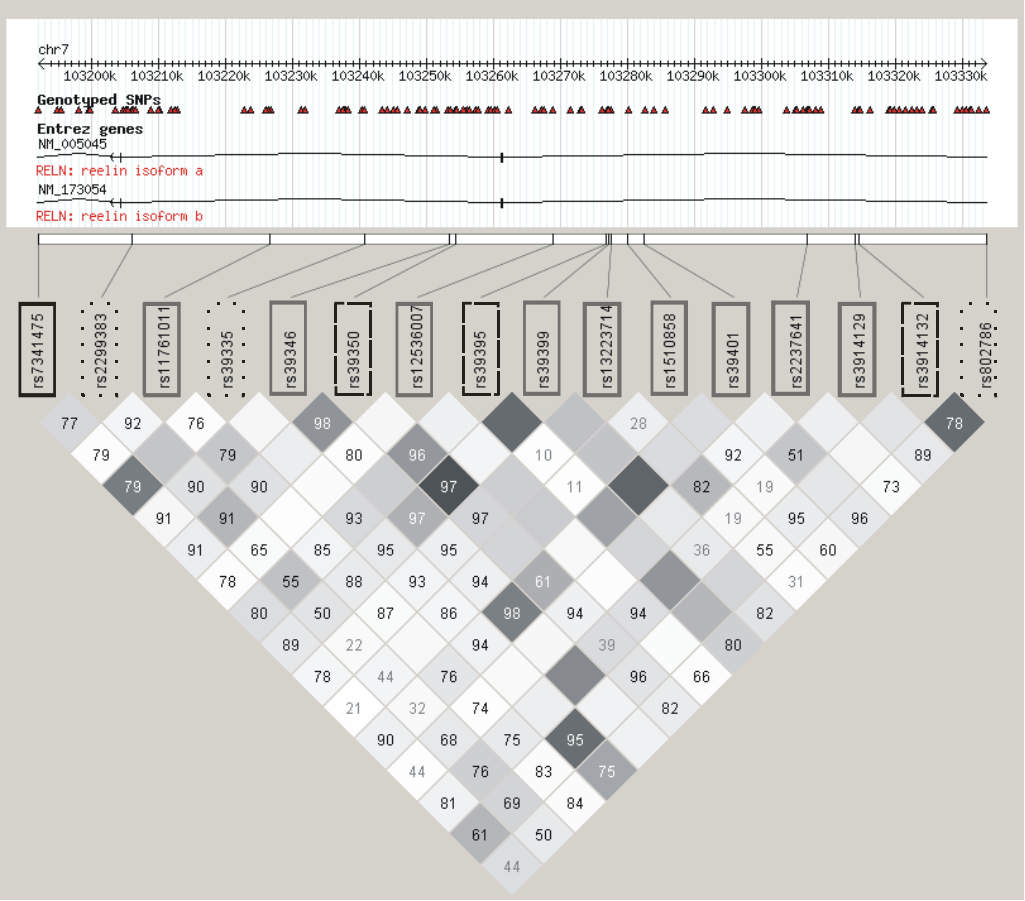


Figure 6.5 Haploview plot depicting linkage disequilibrium between the rs7341475 SNP associated with microexon skipping and 15 SNPs associated with otosclerosis

The strength of LD (D' value) is shown in shades of grey. SNP rs7341475 which was associated with micro exon skipping in the Ovadia et al. study is highlighted in black (Ovadia and Shifman, 2011). All other SNPs were found to be associated with otosclerosis in the discovery and/or one of the replication groups in the otosclerosis GWAS (Schrauwen et al., 2009a). The nine SNPs highlighted in grey were associated with otosclerosis in either the discovery group or a replication cohort. The three SNPs highlighted by dotted lines were associated with otosclerosis in the discovery group and a replication cohort. The three SNPs highlighted by dashed lines were associated with otosclerosis in the discovery group and two replication cohorts. (HapMap release 21, phase III).

In summary, *RELN*-specific mRNA is present in human stapes footplates and suprastructures in both individuals affected by otosclerosis and controls, although at low levels. However, only the *RELN*-002 isoform which undergoes alternative splicing was detected in the human stapes. This provides an explanation for the discrepancy in results published by Schrauwen et al. who were able to detect the *RELN*-002 isoform in their PCR assay and Csomor et al. who were not. This study therefore resolves the controversies regarding *RELN* expression in the stapes. Further studies are required to determine the exact role of the *RELN* isoforms and their possible functions in the pathogenesis of otosclerosis.

7 General Discussion

Throughout the course of this research project, two next generation sequencing technologies; RNA-seq and WES, have been used in combination with one another to identify new otosclerosis candidate genes, in an attempt to shed light on the disease process involved in otosclerosis. WES has proven to be an extremely useful tool in identifying a number of promising pathologic candidate variants within genes of interest in families exhibiting monogenic inheritance of otosclerosis. This technique has great value for genetic analysis of otosclerosis families to identify segregating candidate variants for follow up analysis in a larger cohort of otosclerosis individuals. RNA-seq also has been useful in the provision of an otosclerosis stapes transcriptome, although the reliability of this data is limited as a result of the small cohort size of the discovery group and lack of available control stapes for follow up analyses. Despite these limitations, by using the results of the RNA-seq study to inform variant prioritisation following WES and by making use of the alternative splicing data produced during RNA-seq, one very promising novel otosclerosis candidate gene, *SERPINF1*, has emerged during this study. Thus, by combining genotyping data with gene expression data, progress has been made in beginning to understand the disease process involved in otosclerosis. A cross-comparison of the results obtained across the studies conducted during this research project have indicated a number of possible themes which could explain how *SERPINF1,* as well as other genes of interest identified during this project, could be potential triggers for otosclerosis pathogenesis. These common themes and possible otosclerosis disease mechanisms are discussed below followed by a summary of the conclusions drawn from this PhD thesis.

7.1 Could a shift in isoform expression of otosclerosis candidate genes be involved in disease pathogenesis?

The identification of possible splice site variants in the *SERPINF1* gene provides evidence to suggest a possible alternative splicing disease mechanism for otosclerosis. Alternative splicing is the process by which multiple messenger RNA products are produced from a single gene. It plays a crucial role in the regulation of eukaryotic gene expression and increases the coding capacity of the human genome. Approximately 95% of all mammalian genes undergo alternative splicing (Pan et al., 2008). An increasing number of examples are emerging that show that alternative splicing can lead to human disease (Tazi et al., 2009). Various hereditary diseases of the skeletal system are known to be caused by splice site mutations, including 195

known splice site mutations within the *COL1A1* gene that lead to the brittle bone disease Osteogenesis Imperfecta (Fan and Tang, 2013).

Through WES followed by Sanger sequencing, five rare nonsynonymous mutations were identified in six unrelated otosclerosis families within the *SERPINF1* gene. The location of these variants, within sites predicted to be involved in either alternative splicing or reduced expression of specific *SERPINF1* transcripts based on an *in silico* splicing assay, suggest that an alternative splicing mechanism of this gene may be involved in otosclerosis. This, combined with gene expression data obtained from both RNA-seq and qRT-PCR assays, suggests that downstream *SERPINF1* transcripts may be less highly expressed in otosclerotic than control stapes in both familial and non-familial cases, implicating a role for alternative splicing of *SERPINF1* in the development of otosclerosis. A more detailed description of this possible disease mechanism can be found in Chapter 4 section 4.13.7 including figure 4.21.

Furthermore, the results from the *RELN* expression study in human stapes also indicate involvement of a potential alternative splicing mechanism in otosclerosis. Semi-quantitative RT-PCR assays along with data from the RNA-seq study indicate that the major *RELN* transcript is not expressed in human stapes but that an alternatively spliced variant, lacking the penultimate six nucleotide microexon is expressed. This, combined with qRT-PCR assays which indicate that expression of *RELN* may be reduced in otosclerotic compared to control stapes, suggests that there may be a reduction in expression of this alternatively spliced *RELN* transcript in otosclerotic stapes. Having said this, due to the small number of control stapes on which the quantitative real time PCR assays were performed, these results should be treated with caution, as a greater number of control stapes would need to be collected and analysed to confirm these findings.

Further research into the characterisation of splice site mutations within otosclerosis candidate genes could help improve understanding of the disease mechanism, which could potentially provide a basis for diagnosis as well as the foundation for development of new effective therapies and treatments for otosclerosis.

7.2 A possible link between heparan sulphate modulation of TGFβ signalling and the otosclerosis disease process?

PEDF, the protein encoded by *SERPINF1*, is known to bind to heparan sulphate, a glycosaminoglycan which regulates transforming growth factor beta (TGFβ) signalling by modulating the assembly of latent TGFβ1 (Chen et al., 2007a). The TGFβ super-family is a group of genes involved in growth, development, homeostasis and regulation of the immune system. *TGFβ1* is a polypeptide abundant in the bone matrix that is known to be a potent stimulator of osteoblastogenesis and inhibitor of osteoclastogensis, for which reason it has been repeatedly proposed as an otosclerosis candidate gene. Various members of this family including *TGFβ1* and bone morphogenetic proteins *BMP2* and *BMP4* have been associated with otosclerosis during association studies, however a precise role for this gene family in development of otosclerosis has not been defined. In this study, a possible link between heparan sulphate and modulation of TGFβ signalling in otosclerosis has emerged as a common theme across multiple lines of research.

Based on homology modelling, there is a putative binding site for heparan sulphate on PEDF located between residues 134 and 214 of the protein. This region is densely populated with lysine residues that are able to interact with glycosaminoglycans such as heparan sulphate (Alberdi et al., 1998). This putative binding site corresponds to a region spanning *SERPINF1* exons 4 and 5, including the region in which both exon 5 variants identified in the otosclerosis cohort are located. Since the c.441G>C variant identified in two unrelated members of the otosclerosis cohort causes a lysine to asparagine change at residue 147, it is possible that this variant could disrupt a heparan sulphate binding site. This would lead to a reduction in binding of heparan sulphate to PEDF, which would result in accumulation of heparan sulphate and thus have an effect on TGFβ signalling (Figure 7.1). Alternatively, it is possible that the presence of variants that lead to alternative splicing or reduced expression of downstream *SERPINF1* transcripts that cause a reduction in expression of exon 5, could reduce the number of heparan binding sites available, thus also leading to heparan sulphate accumulation and altered TGFβ signalling.

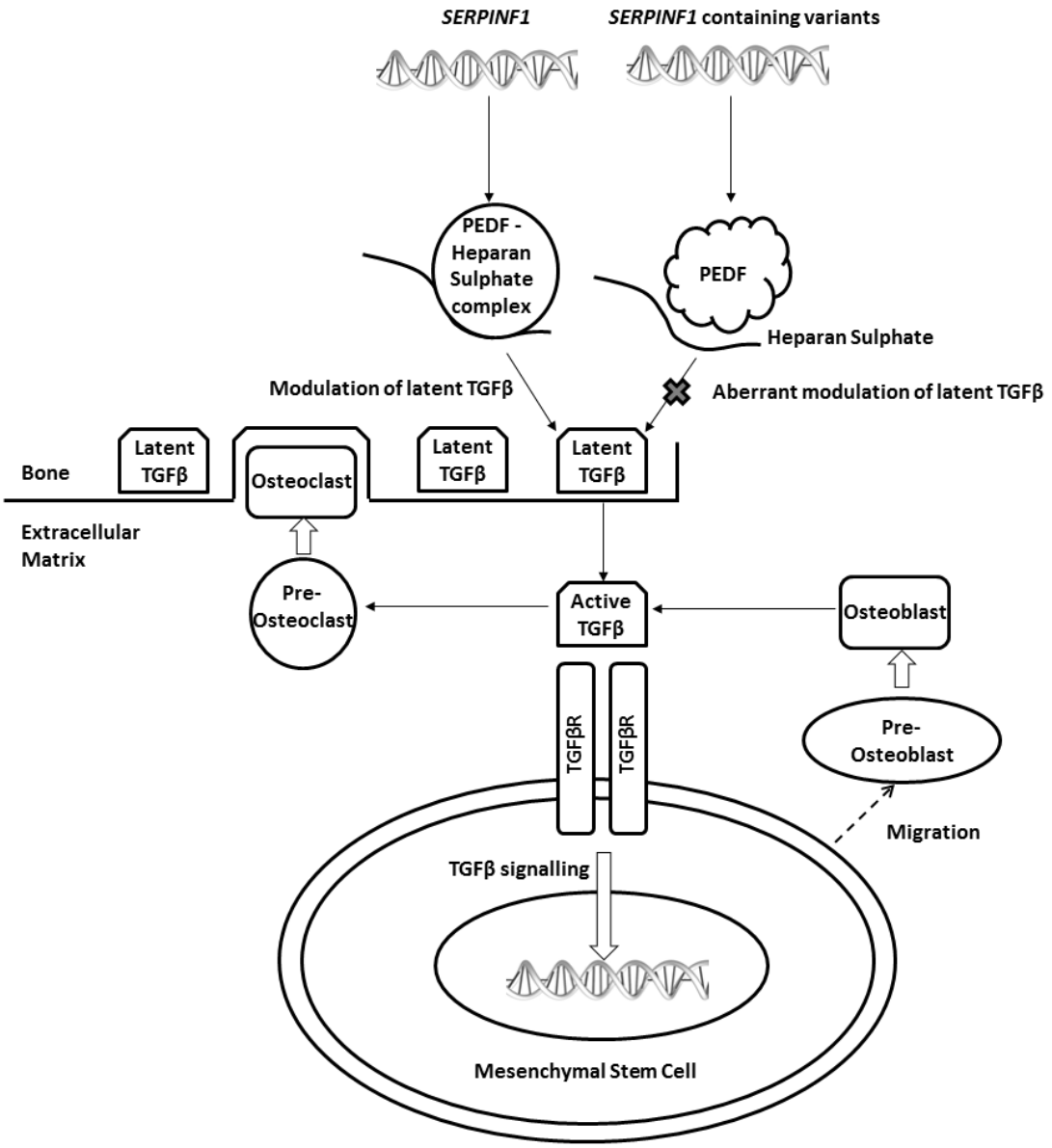


Figure 7.1 Schematic showing the possible effect of *SERPINF1* variants on TGFβ signaling via heparan sulphate mediated modulation of latent TGFβ.

In addition, PEDF is known to compete with heparan sulphate for binding sites located within collagen genes. Both PEDF and heparan sulphate bind to similar recognition sites located within the triple helix region of collagen genes (Sekiya et al., 2011). The triple helix region of *COL1A2* is the region in which the two variants identified within this gene in the otosclerotic cohort are located, although it doesn't appear that either of these variants are located directly within the PEDF or heparan sulphate binding sites themselves. Reduced expression of downstream *SERPINF1* transcripts as a result of alternative splicing could therefore reduce

competition for binding by heparan sulphate, which could have an impact on both TGFβ signalling and the collagen matrix within bone tissue.

Further evidence from this study to support a possible role for heparan sulphate modulation of TGFβ signalling in otosclerosis can be found in the pathway and functional analyses performed following RNA-seq. These studies revealed that genes dysregulated in otosclerotic stapes were enriched for those encoding a number of small leucine-rich proteoglycans (SLRPs). Following quantitative real-time PCR to validate these findings, the SLRP *ASPN* (Asporin) was confirmed to be significantly downregulated in otosclerotic compared to control stapes. Asporin regulates chondrogenesis by modulating the TGFβ1 signalling pathway. It does this through inhibition of the interaction between TGFβ1 and its type II receptor in the presence of heparan sulphate (Kou et al., 2010). This could implicate a potential role for heparan sulphate in combination with *ASPN* in the regulation of the TGFβ signalling pathway in the stapes.

Further indications that heparan sulphate may play a role in otosclerosis disease pathology, stems from the observation of otosclerosis in children with Hunter's syndrome (Zechner and Moser, 1987). This rare X-linked lysosomal storage disorder is caused by deficiency or absence of the enzyme iduronate-2-sulphatase which results in accumulation of heparan sulphate. Due to the rarity of childhood otosclerosis, it has been proposed there is a causal connection between Hunter's syndrome and otosclerosis.

This possible link between heparan sulphate and otosclerosis also provides a plausible explanation for the persistent reports of a role for viral infection in the disease. Heparan sulphate proteoglycans possess large polysaccharide chains which serve as receptors for viruses. Binding of a viral protein to heparan sulphate is frequently the first step in a cascade of interactions leading to viral entry into a cell and the initiation of infection (Shukla and Spear, 2001). It is therefore possible that otosclerosis is a xenogenetic disease where genetic and environmental factors are interlinked in predisposing an individual to otosclerosis.

Furthermore, it has also been proposed that reelin, the large extracellular matrix protein found to be associated with otosclerosis during the genome wide association study, may serve a similar function to proteoglycan molecules such as heparan sulphate. During quantitative real-time PCR assays conducted during this study, *RELN* was found to be expressed in otosclerotic stapes at less than a fifth of the level at which it is expressed in control stapes, indicating that this gene, like various members of the SLRP family may be downregulated in otosclerotic stapes. Although Reelin is itself not a proteoglycan, its size and structure indicates that it may have a similar role to proteoglycans, such as by providing binding sites and concentrating

growth factors (Tissir and Goffinet, 2003). This is supported by a recent study where reelin was found to be an essential negative regulator of TGFβ1-induced cell migration in an oesophageal carcinoma cell line (Yuan et al., 2012). This publication represents the first time that reelin has been postulated to be involved in a genetic pathway involving other otosclerosis candidate genes and offers a plausible link between this gene and a possible disease mechanism involving TGFβ.

Through analysis of data collected from a combination of next generation sequencing and single gene assays, a trend has emerged implicating proteoglycans and in particular heparan sulphate modulated TGFβ signalling in otosclerosis pathogenesis. A number of proteoglycans including the SLRPs found to be dysregulated in otosclerotic stapes through RNA-seq, as well as heparan sulphate, not only interact directly with members of the TGFβ superfamily to modulate signalling, but also interact via genes in which variants were identified during sequencing in otosclerosis families. Further experimentation would be required to investigate this theory.

7.3 Osteogenesis Imperfecta and Otosclerosis: A shared genetic aetiology?

The identification of multiple variants within two genes, *COL1A2* and *SERRPINF1* during this research project, both of which are known to be involved in the systemic connective tissue disorder Osteogenesis Imperfecta (OI), suggests a possible shared genetic aetiology for this condition and otosclerosis. There are numerous similarities between the clinical features of otosclerosis and the hearing loss associated with OI. These similarities have previously led researchers to postulate that otosclerosis may be a localised manifestation of OI (Weber, 1930). Like in otosclerosis, the temporal bones of OI patients frequently exhibit delayed ossification of the auditory ossicles, and OI-associated conductive hearing loss, which usually develops in early adulthood shows similarities to that of otosclerosis patients. This, in combination with an association found between variants in the OI-causing gene *COL1A1* and otosclerosis, suggests that some cases of clinical otosclerosis may be caused by mutations similar to those found in mild forms of OI (McKenna et al., 1998).

The results from the WES study conducted during this research project support the concept of a shared genetic aetiology between otosclerosis and OI. Within two of the four families sequenced, rare nonsynonymous segregating variants were identified within two of the six known OI genes; *SERPINF1* and *COL1A2*. Although these variants were not identified in any

other cohort members from a population of 53 additional probands reporting a strong family history of otosclerosis, four further rare nonsynonymous variants within *SERPINF1* were found in five unrelated probands through Sanger sequencing of the remainder of the gene. In addition, an unknown 9 base insertion was identified within the *COL1A2* gene in one further proband and her affected mother. Due to the large size of *COL1A2*, which consists of 52 exons, only the exon in which the familial mutation had been found was sequenced in the remainder of the cohort. It is therefore possible that additional *COL1A2* mutations could reside within other exons in additional members of the cohort. Therefore in total, rare nonsynonymous variants were identified within known OI genes in eight of the 57 otosclerosis probands, constituting approximately 14% of the cohort. In reality, this percentage could be far higher since only 6% of the exons present within the six known OI genes were sequenced during this study. This provides strong evidence to suggest that genes known to be involved in OI may also contribute to otosclerosis pathogenesis.

Furthermore, the Ingenuity Pathway Analysis performed on the dysregulated gene set identified during RNA-seq of otosclerotic and control stapes suggested that 33 of the 176 dysregulated genes were involved in connective tissue disorders. In addition, the functional classification analysis performed on the same dysregulated gene set using DAVID suggested that the dataset is over-represented for genes that encode leucine- rich repeat motifs, which are common in connective tissue proteins (Bengtsson et al., 1995). This evidence is consistent with the concept of a shared genetic aetiology for otosclerosis and other connective tissue disorders.

In individuals with OI VI, 16 unique sequence variants have been identified within *SERPINF1*. These variants include both nonsense and frameshift mutations which lead to nonsense mediated decay of *SERPINF1* mRNA, and thus loss of expression of PEDF (Becker et al., 2011). Since it appears that the *SERPINF1* variants identified in the otosclerosis cohort do not cause nonsense mediated decay but rather are more likely to affect alternative splicing of this gene, it is possible that these variants lead to loss of expression of specific *SERPINF1* isoforms in the stapes rather than complete loss of expression of this gene in all tissues of the body. This could explain the phenotypic differences seen between OI VI and otosclerosis and would support the theory that otosclerosis is a localised manifestation of OI.

The findings of this research have important implications for the treatment of both otosclerosis and the otopathology associated with OI. An improved understanding of the genes involved in the development of conductive hearing loss in these conditions may help in the elucidation of the disease processes. This in turn could assist the development of specific

targeted therapies tailored to an individual's needs based on their genetic composition. Since personalised medicine is becoming a realistic option for treatment of disease through improvements in pharmacogenomics, tailored drug therapy for these otopathologies could be an alternative to surgical options in the not too distant future.

7.4 Conclusions

Throughout the course of this PhD research project, a UK-wide recruitment strategy has been successfully implemented to establish large collections of both genomic DNA and cDNA from the stapes of otosclerosis patients. This valuable resource has enabled the implementation of a range of genetic analyses including both traditional approaches and NGS techniques to investigate the genetics of otosclerosis, and has opened up a range of opportunities for future research.

Single gene PCR assays conducted during this research project have helped to resolve a discrepancy in the published literature regarding the expression of the otosclerosis candidate gene *RELN* in human stapes.

WES followed by segregation analysis in otosclerosis families has enabled the identification of a number of good otosclerosis pathologic candidate variants within *SERPINF1*, *TRIM17*, *mir-183* and *COL1A2* genes. Subsequent genotyping assays in a larger cohort of familial otosclerosis probands identified additional variants within two of these candidate genes, *SERPINF1* and *COL1A2*. Both these genes are excellent otosclerosis candidates due to their known role in the systemic connective tissue disorder Osteogenesis Imperfecta, in which patients frequently display an otopathology with a similar phenotype to that seen in otosclerosis.

RNA-seq has enabled the provision of an otosclerosis stapes transcriptome which has assisted the variant prioritisation process following WES and will provide a useful gene expression dataset for future genetic research into otosclerosis. It has also been valuable in evaluating expression of the candidate genes prioritised during WES, suggesting a possible shift in expression of specific *SERPINF1* isoforms in the stapes, a theory which was supported by subsequent *in silico* splicing assays and qRT-PCR in a larger stapes cohort. Thus, the two NGS technologies employed during the project; WES and RNA-seq, have enabled the identification

of a promising new otosclerosis candidate gene; *SERPINF1,* for which there is strong evidence for a possible role in disease aetiology.

It is hoped that this new knowledge will help to unravel the molecular mechanisms involved in disease pathology. Improved knowledge of the genetic pathways involved in otosclerosis will pave the way for future research in order to help pinpoint genes and gene families in which to search for both rare variants that contribute to the familial form of the condition, and the multiple common variants which are expected to be involved in the complex form of the disease. It will be crucial to maintain strong links between genetic research into otosclerosis and the clinical treatment for otosclerosis patients, as not only will such relationships enable the provision of otosclerotic patient material on which to perform genetic analyses, but will also assist in translation of any laboratory findings into possible treatments for patients. This in turn could provide future therapeutic options to halt the progression of hearing loss for individuals affected by otosclerosis.

Appendix

Study participant questionnaire

UCL EAR INSTITUTE &
CENTRE FOR AUDITORY RESEARCH
GRAY'S INN ROAD
LONDON, WC1X 8EE

Telephone 020 7679 8935
FAX 020 7679 8990

ROYAL NATIONAL THROAT
NOSE & EAR HOSPITAL
GRAY'S INN ROAD
LONDON, WC1X 8DA

Telephone 020 7915 1300
FAX 020 7833 5518

Royal National Throat, Nose and Ear Hospital

Investigating the causes of otosclerosis (PhD Research project)

Participant Questionnaire: 1.7	Participant number:

Thank you for helping with this project, which will help us to understand the causes of otosclerosis. Please see the attached **Participant Information Sheet** for more details. We are looking for people of any age who have developed otosclerosis. Your contribution will involve answering the few questions below, and giving a saliva sample for genetic testing.

Please read through the questions and answer all those that apply.

1 . What is your age? ☐

2 . Are you: ☐ Male
☐ Female

3 . QUESTIONS ABOUT HEARING LOSS

a Does anyone else in your family have otosclerosis?

- ☐ **Yes, other member/s of my family have otosclerosis**
- ☐ **No, there are no other members of my family with otosclerosis**
- ☐ **Don't know**

If you answered yes, please provide us with details, including how this family member(s) is related to you.

_____

_____

_____

b Does anyone else in your family have hearing loss who has not been diagnosed with otosclerosis?

- ☐ **Yes**
- ☐ **No**

If you answered yes, please provide us with details, including how this family member(s) is related to you and if the cause of their deafness is known, for example, childhood deafness, surgery, trauma.

_____

_____

_____

c At what age did you first experience symptoms of hearing loss?

- ☐ **less than 10 years**
- ☐ **age 11-20**
- ☐ **age 21-30**
- ☐ **age 31-40**
- ☐ **age 41-50**
- ☐ **age 51-60**
- ☐ **age 61 or older**
- ☐ **Don't know**

d Have you ever experienced significant problems with the following?

	Bilateral hearing loss (hearing loss in both ears)	**Tinnitus** (ringing in the ears)	**Vertigo** (dizziness /poor balance)
Yes	☐	☐	☐
No	☐	☐	☐

4 . QUESTIONS ABOUT YOUR HEALTH

a Have you ever had any of the following conditions? Please tick all that apply.

☐ **Osteoporosis**

☐ **Rheumatoid Arthritis or another Autoimmune condition**

If you answered yes, please provide us with brief details about your condition.

_____

_____

_____

b Have you ever broken a bone/s?

☐ **Yes**

☐ **No**

If you answered yes, please provide us with brief details.

_____

_____

_____

5 . QUESTIONS ABOUT MEASLES, MUMPS AND RUBELLA

a Have you ever had any of the following **vaccinations**?

	Measles	Mumps	Rubella
Yes	☐	☐	☐
No	☐	☐	☐
Don't know	☐	☐	☐

b Have you ever had any of the following **infections**?

	Measles	Mumps	Rubella
Yes	☐	☐	☐
No	☐	☐	☐
Don't know	☐	☐	☐

6 . DETAILS ABOUT WHERE YOU LIVE

It has been suggested that the flouride content in our drinking water may play a role in the development of otosclerosis.
We would therefore like to ask you for some information about where you live and where you have lived in the past. This information will help us to investigate whether there is a link between otosclerosis and the flouride content of the water supplied by your local authority.

Please be assured that this information is strictly confidential and will be used for the purposes of this study only.

a What is your postcode area (ie. The first part of your postcode)
eg. W1, NW2, SE16

☐

b For how long have you lived in this postcode area?

- ☐ **less than 1 year**
- ☐ **1-5 years**
- ☐ **6-10 years**
- ☐ **11 years or more**

c If you have previously lived in any other postcode area for **longer than 5 years**, please give details.

_____
_____
_____

d If you have lived abroad for **longer than 5 years**, please give details.

_____
_____
_____

QUESTIONS ABOUT YOUR ETHNICITY

We would like to ask you for some information about your **ethnicity**. This information will help us to investigate whether there is a link between otosclerosis and your genetic background.

Please be assured that this information is strictly confidential and will be used for the purposes of this study only.

7. Please indicate which of the following ethnic groups best describes you. This may not necessarily be your country of birth or your nationality, but where your ancestors (ie. parents and grandparents) are originally from. If you are of mixed ethnicity, please tick all that apply.

White

- ☐ British/Irish
- ☐ Northern European
- ☐ Southern European
- ☐ Of Ashkenazi Jewish descent*
- ☐ Of Sephardi Jewish descent*
- ☐ Any other white background

Asian or Asian British

- ☐ Indian
- ☐ Pakistani
- ☐ Bangladeshi
- ☐ East African Asian
- ☐ Any other Asian background

Black or Black British

- ☐ Caribbean
- ☐ African
- ☐ Somali
- ☐ Any other black background

Other

- ☐ Chinese
- ☐ Japanese
- ☐ Arab/Iranian/North African
- ☐ Any other background

If you answered other to any of the questions above, please give details.

_____

* We ask these questions because Jewish people share genetic similarities, which may differ from other European populations.

QUESTIONS FOR FEMALES ONLY

Otosclerosis may be more common in females than in males. It has been suggested that female reporductive hormones may play a role in the development of otosclerosis.
We would therefore like to ask you for some information about pregnancy, contraceptives and hormone therapy. This information will help us to investigate whether there is a link between otosclerosis and female reproductive hormones.

Please be assured that this information is strictly confidential and will be used for the purposes of this study only.

8 .a Have you ever been pregnant?

☐ **Yes**

☐ **No** Please skip to question 9

b How many pregnancies have you had? ☐

c Have you ever breast fed?

☐ **Yes**

☐ **No** Please skip to question 8e

d For approximately how many years **in total** have you breast fed?

☐ **less than 1 month** ☐ **6-12 months**

☐ **2-5 months** ☐ **1 year or more**

e Did you notice any changes in your hearing ability during your pregnancy/pregnancies

☐ **Yes**

☐ **No**

If you answered yes, please provide us with details about these changes.

_____

_____

_____

9. a Are you currently taking, or have you ever taken oral contraceptives ('the pill')?

- ☐ **Yes, currently taking**
- ☐ **Yes ,previously taken**
- ☐ **No, never taken** please skip to question 10

b For approximately how many years **in total** have you taken oral contraceptives ('the pill')?

- ☐ **less than 1 year**
- ☐ **2-5 years**
- ☐ **6-10 years**
- ☐ **11-15 years**
- ☐ **16-20 years**
- ☐ **21-25 years**
- ☐ **26 years or more**
- ☐ **Don't Know**

10 .a Have you gone through /are going through the menopause?

- ☐ **Yes**
- ☐ **No** please skip to question 11

b If you answered yes, approximately how old were you when you first experienced symptoms of the menopause?

☐

11 .a Are you currently taking, or have you ever taken Hormone Replacement Therpay (HRT)?

- ☐ **Yes, currently taking**
- ☐ **Yes ,previously taken**
- ☐ **No, never taken**

b If you answered yes, for approximately how many years **in total** have you taken Hormone Replacement Therapy (HRT)?

- ☐ **less than 1 year**
- ☐ **2-5 years**
- ☐ **6-10 years**
- ☐ **11-15 years**
- ☐ **16-20 years**
- ☐ **21-25 years**
- ☐ **26 years or more**
- ☐ **Don't Know**

176 RNA-seq dysregulated genes

Downregulated genes			
Gene name	Z ratio	Gene name	Z ratio
EEF1A1	-64.07	RPS15A	-3.00
HBB	-44.36	TSPAN8	-2.98
FTL	-42.74	RPS13	-2.96
B2M	-20.13	RPL31	-2.94
SLPI	-18.03	UBC	-2.93
LUM	-16.35	RPS4Y1	-2.92
RPL7	-14.42	CFH	-2.90
RPL41	-11.95	DIO2	-2.90
RPS6	-10.20	MYL12A	-2.78
RPL10	-8.90	RPL27	-2.75
RPS3A	-8.18	RPL17	-2.73
RPLP0	-7.50	EIF4A2	-2.65
MYL6	-7.35	RPL24	-2.60
DCN	-7.10	EEF1G	-2.59
RPL23A	-7.05	PPIA	-2.44
HLA-DRA	-6.75	RPL23	-2.41
RPL32	-6.42	LDHB	-2.35
RPS27A	-5.58	SRP14	-2.28
ANXA2	-5.55	RPL35A	-2.26
PFDN5	-5.28	CD59	-2.23
COCH	-5.26	OMD	-2.23
SPP1	-5.25	VIM	-2.18
MGP	-5.13	RPS25	-2.18
RPS27	-4.69	RPS29	-2.14
RPL9	-4.43	SDHD	-2.13
RPL5	-4.32	RPS12	-2.10
RPL15	-4.15	MYL12B	-2.09
RPS3	-4.12	GLUL	-2.09
RPL30	-4.10	HSP90AA1	-2.02
RPS11	-3.99	RPS7	-1.97
RPL14	-3.96		
SPARCL1	-3.84		
SKP1	-3.82		
ARL6IP5	-3.79		
RPS18	-3.78		
MT1X	-3.73		
ANXA1	-3.69		
UBB	-3.67		
RPL34	-3.65		
RPS14	-3.47		
SAT1	-3.43		
RPL19	-3.40		
CALM2	-3.38		
CTSK	-3.36		
HIST1H4C	-3.27		
RPL4	-3.19		
RPS23	-3.05		
RPL11	-3.05		
ASPN	-3.02		

Upregulated genes			
Gene name	Z ratio	Gene name	Z ratio
MALAT1	84.46	SCAI	2.74
HSP90AB2P	35.68	KCNJ5	2.73
C1orf161	19.34	LOC219347	2.71
CLU	13.27	ZNF793	2.71
APOD	12.46	PDE6A	2.70
LOC646214	10.05	ZNF347	2.70
SHISA9	8.06	DCUN1D2	2.65
LOC643406	7.69	NLRP12	2.63
FAM119A	6.91	MAP1LC3C	2.63
ACTB	6.01	LOC100128288	2.60
LOC284232	5.53	FBLIM1	2.60
PLUNC	5.43	CCDC142	2.59
LPP	4.59	LAIR1	2.57
CD74	4.56	KBTBD12	2.47
IFITM1	4.26	ZNF805	2.45
OPHN1	4.17	ZNF493	2.41
ZMAT3	4.17	PTK6	2.40
FKBP14	4.01	IL17RD	2.37
TLCD2	3.83	CYP4V2	2.34
LOC442293	3.76	CRX	2.30
ABCC9	3.75	ENTPD4	2.29
IL28RA	3.63	C11orf58	2.28
PPP1R2P1	3.60	C10orf72	2.26
CAMK1D	3.56	RBMS2	2.24
IGJ	3.53	TAF8	2.23
TMEM212	3.38	L2HGDH	2.18
KIAA1486	3.29	GEMIN8	2.16
LRRC58	3.23	ZBTB8A	2.16
ZNF461	3.23	GNE	2.14
PDP2	3.22	SLC9A4	2.14
H3F3B	3.21	LOC100272146	2.12
SLC31A1	3.19	S100A11	2.11
PHACTR4	3.13	PCDH11Y	2.10
PIGX	3.13	SKA1	2.09
TATDN3	3.08	SIX4	2.06
VPS53	3.06	C8orf79	2.05
IFITM3	3.01	RBM34	2.05
ZNF264	3.01	ZNF490	2.05
PTMA	2.97	SMG1	2.03
FOXK1	2.96	NOL9	2.03
UGGT1	2.96	CHP	2.02
TMEM41B	2.92	SPRED1	2.02
C21orf62	2.89	FMN1	2.01
HBA2	2.89	LOC90834	1.99
GOSR1	2.87	PNPT1	1.99
LOC283267	2.85	LOC100190938	1.98
GREB1	2.83	GTF2E1	1.97
XIAP	2.83	SIRPB2	1.97
CTSB	2.76		

Table showing restriction endonuclease digests

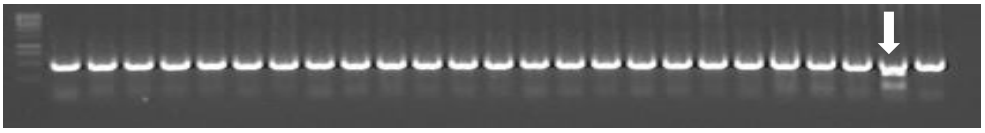
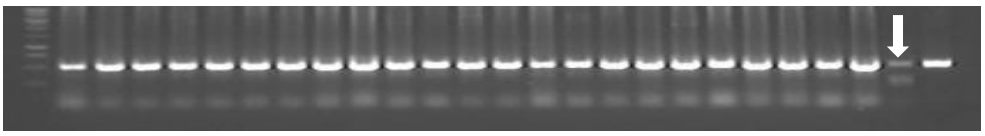
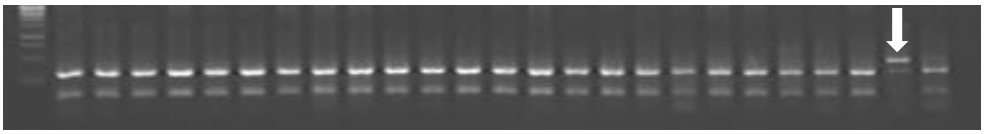
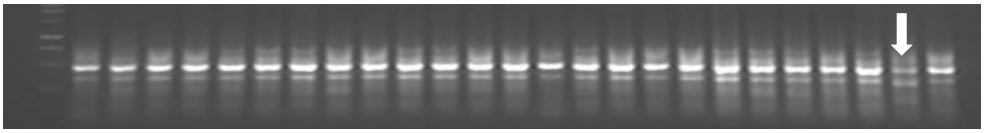
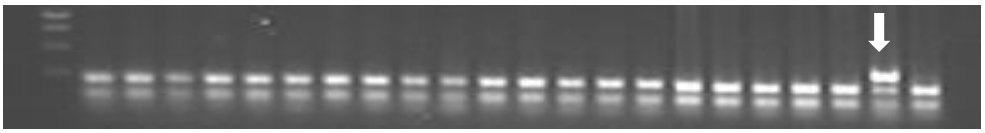
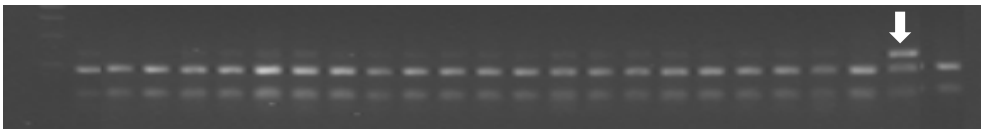
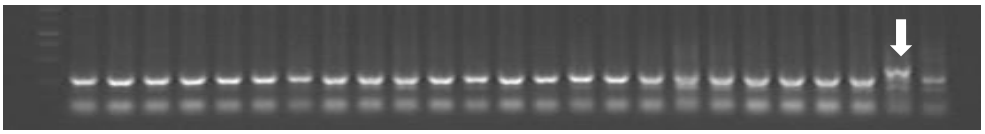
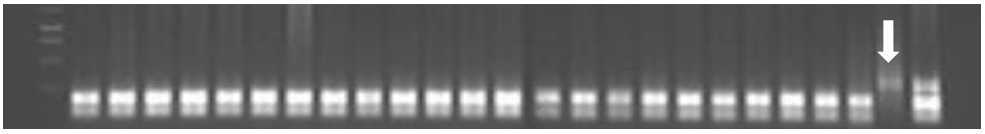
Gene	Variant	PCR amplicon length	Restriction Endonuclease	Fragment sizes wt	Fragment sizes het	Fragment sizes mutant
CYP2D6	c.345delC	505bp	DrdI	505bps	505+278+226bps	278+226bp
						
DNAH5	c.7883T>G	469bp	OliI	469bp	469+319+150bp	319+150bp
						
EGLN3	c.496C>T	446bp	AciI	308+138bp	446+308+138bp	446bp
						
mir-183	c.81G>T	473bp	AvaII	473bp	473+303+170bp	303+170bp
						
PTK6	c.652A>T	320bp	MlsI	207+113bp	320+207+113bp	320bp
						
SERPINF1	c.441G>C	440bp	TseI	314+126bp	440+314+126bp	440bp
						
TRIM17	c.915C>G	419bp	RsaI	334+85bp	419+334+85bp	419bp
						
VPS53	c.107C>G	299bp	PpuMI	182+117bp	299+182+117bp	299bp
						

Table shows the predicted fragment size and an agarose gel electrophoresis image showing 1kb DNA ladder in lane 1, digest fragments for 23 cohort members in lanes 2-24, a known heterozygote in lane 25 (indicated by arrows) which acted as a positive control, and a known wild type in lane 26.

RNA-seq dysregulated genes–z score and RPKM overlap

Downregulated genes			
gene name	Z ratio	p-value	Fold diff
ASPN	-3.02	0.00	5.28
HIST1H4C	-3.27	0.01	4.73
PFDN5	-5.28	0.01	4.19
RPS15A	-3.00	0.00	3.98
RPL32	-6.42	0.00	3.89
RPL30	-4.10	0.00	3.83
RPL23	-2.41	0.00	3.74
RPL24	-2.60	0.00	3.72
RPL27	-2.75	0.02	3.72
RPL23A	-7.05	0.01	3.71
MYL6	-7.35	0.03	3.69
RPS29	-2.14	0.04	3.68
RPL17	-2.73	0.00	3.64
RPL11	-3.05	0.00	3.64
RPS7	-1.97	0.00	3.57
RPS13	-2.96	0.01	3.55
RPS3A	-8.18	0.03	3.55
SDHD	-2.13	0.01	3.51
SLPI	-18.03	0.01	3.50
RPS23	-3.05	0.00	3.49
RPL34	-3.65	0.01	3.41
SPARCL1	-3.84	0.00	3.41
RPS6	-10.20	0.01	3.40
RPL7	-14.42	0.00	3.40
RPL9	-4.43	0.01	3.39
RPL14	-3.96	0.02	3.36
RPL5	-4.32	0.01	3.31
RPL35A	-2.26	0.01	3.31
PPIA	-2.44	0.01	3.29
ARL6IP5	-3.79	0.02	3.28
RPS27	-4.69	0.01	3.26
RPS27A	-5.58	0.00	3.15
RPL41	-11.95	0.00	3.14
SRP14	-2.28	0.03	3.11
RPS25	-2.18	0.00	3.07
RPL31	-2.94	0.01	3.06
TSPAN8	-2.98	0.01	3.04
LDHB	-2.35	0.01	3.01
MYL12A	-2.78	0.04	2.97
RPL4	-3.19	0.03	2.95
ANXA2	-5.55	0.03	2.89
VIM	-2.18	0.01	2.80
SKP1	-3.82	0.02	2.78
LUM	-16.35	0.01	2.77
RPS14	-3.47	0.01	2.73
ANXA1	-3.69	0.01	2.65
RPS18	-3.78	0.01	2.58
CALM2	-3.38	0.01	2.54
HSP90AA1	-2.02	0.03	2.33
RPL15	-4.15	0.03	2.30
MYL12B	-2.09	0.03	2.26

Upregulated genes			
gene name	Z ratio	p-value	Fold diff
ZMAT3	4.17	0.01	1.50
CAMK1D	3.56	0.01	2.12
KIAA1486	3.29	0.00	2.12
LOC219347	2.71	0.01	2.07
PTK6	2.40	0.00	2.06
LOC100128288	2.60	0.01	2.06
PPP1R2P1	3.60	0.00	2.03
LOC90834	1.99	0.00	2.00
PCDH11Y	2.10	0.01	1.98
FBLIM1	2.60	0.00	1.97
KCNJ5	2.73	0.01	1.96
IL17RD	2.37	0.00	1.95
PIGX	3.13	0.01	1.93
SIRPB2	1.97	0.01	1.93
GREB1	2.83	0.00	1.87
LOC100190938	1.98	0.02	1.86
NLRP12	2.63	0.02	1.85
TLCD2	3.83	0.01	1.85
SKA1	2.09	0.00	1.85
C21orf62	2.89	0.01	1.84
FOXK1	2.96	0.00	1.83
LOC100272146	2.12	0.01	1.82
ENTPD4	2.29	0.00	1.82
PDP2	3.22	0.01	1.81
PDE6A	2.70	0.03	1.81
KBTBD12	2.47	0.03	1.78
C8orf79	2.05	0.02	1.75
FMN1	2.01	0.01	1.75
LOC283267	2.85	0.01	1.72
C10orf72	2.26	0.00	1.70
CRX	2.30	0.02	1.69
DCUN1D2	2.65	0.02	1.68
LAIR1	2.57	0.01	1.67
ZNF264	3.01	0.01	1.67
ZNF490	2.05	0.01	1.66
IL28RA	3.63	0.01	1.65
PNPT1	1.99	0.03	1.64
CCDC142	2.59	0.01	1.62
SCAI	2.74	0.02	1.62
GNE	2.14	0.00	1.61
ZNF793	2.71	0.02	1.60
VPS53	3.06	0.02	1.60
GEMIN8	2.16	0.01	1.59
MAP1LC3C	2.63	0.03	1.58
ZBTB8A	2.16	0.01	1.57
ZNF805	2.45	0.01	1.54
TAF8	2.23	0.02	1.53
NOL9	2.03	0.02	1.53
SIX4	2.06	0.00	1.51

Bibliography

ALBERDI, E., HYDE, C. C. & BECERRA, S. P. 1998. Pigment epithelium-derived factor (PEDF) binds to glycosaminoglycans: analysis of the binding site. *Biochemistry,* 37**,** 10643-52.

ALBERS, A. E., WAGNER, W., STOLZEL, K., SCHONFELD, U. & JOVANOVIC, S. 2011. [Laser stapedotomy]. *HNO,* 59**,** 1093-102.

ALBRECHT, W. 1922. Über der vererbung der hereditären labyrinth-schwerkörigkeit und der otosclerose. *Arch Ohrenheilk Nas Kehlkopfheilk* 11**,** 244-251.

ALI, I. B., THYS, M., BELTAIEF, N., SCHRAUWEN, I., DIELTJENS, N., VANDERSTRAETEN, K., BESBES, G., MNIF, E., HACHICHA, S., ARAB, S. B. & CAMP, G. V. 2007. Clinical and genetic analysis of two Tunisian otosclerosis families. *Am J Med Genet A,* 143A**,** 1653-60.

ALTERMATT, H. J., GERBER, H. A., GAENG, D., MULLER, C. & ARNOLD, W. 1992. [Immunohistochemical findings in otosclerotic lesions]. *HNO,* 40**,** 476-9.

ALZOUBI, F. Q., OLLIER, W. R., RAMSDEN, R. T. & SAEED, S. R. 2007. No evidence of linkage between 7q33-36 locus (OTSC2) and otosclerosis in seven British Caucasian pedigrees. *J Laryngol Otol,* 121**,** 1140-7.

AMIN, S. & TUCKER, A. S. 2006. Joint formation in the middle ear: lessons from the mouse and guinea pig. *Dev Dyn,* 235**,** 1326-33.

ARNOLD, W. 2007. Some remarks on the histopathology of otosclerosis. *Adv Otorhinolaryngol,* 65**,** 25-30.

ARNOLD, W., BUSCH, R., ARNOLD, A., RITSCHER, B., NEISS, A. & NIEDERMEYER, H. P. 2007. The influence of measles vaccination on the incidence of otosclerosis in Germany. *Eur Arch Otorhinolaryngol,* 264**,** 741-8.

ARNOLD, W. & FRIEDMANN, I. 1987. [Detection of measles and rubella-specific antigens in the endochondral ossification zone in otosclerosis]. *Laryngol Rhinol Otol (Stuttg),* 66**,** 167-71.

ARNOLD, W., NIEDERMEYER, H. P., ALTERMATT, H. J. & NEUBERT, W. J. 1996. [Pathogenesis of otosclerosis. "State of the art"]. *HNO,* 44**,** 121-9.

BAMSHAD, M. J., NG, S. B., BIGHAM, A. W., TABOR, H. K., EMOND, M. J., NICKERSON, D. A. & SHENDURE, J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet,* 12**,** 745-55.

BAUMANN, M., GIUNTA, C., KRABICHLER, B., RUSCHENDORF, F., ZOPPI, N., COLOMBI, M., BITTNER, R. E., QUIJANO-ROY, S., MUNTONI, F., CIRAK, S., SCHREIBER, G., ZOU, Y., HU, Y., ROMERO, N. B., CARLIER, R. Y., AMBERGER, A., DEUTSCHMANN, A., STRAUB, V., ROHRBACH, M., STEINMANN, B., ROSTASY, K., KARALL, D., BONNEMANN, C. G., ZSCHOCKE, J. & FAUTH, C. 2012. Mutations in FKBP14 cause a variant of Ehlers-Danlos syndrome with progressive kyphoscoliosis, myopathy, and hearing loss. *Am J Hum Genet,* 90**,** 201-16.

BECKER, J., SEMLER, O., GILISSEN, C., LI, Y., BOLZ, H. J., GIUNTA, C., BERGMANN, C., ROHRBACH, M., KOERBER, F., ZIMMERMANN, K., DE VRIES, P., WIRTH, B., SCHOENAU, E., WOLLNIK, B., VELTMAN, J. A., HOISCHEN, A. & NETZER, C. 2011. Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet,* 88**,** 362-71.

BEL HADJ ALI, I., BEN SAIDA, A., BELTAIEF, N., NAMOUCHI, I., BESBES, G., GHAZOUENI, E. & BEN ARAB, S. 2012. HLA class I polymorphisms in Tunisian patients with otosclerosis. *Ann Hum Biol,* 39**,** 190-4.

BEL HADJ ALI, I., THYS, M., BELTAIEF, N., SCHRAUWEN, I., HILGERT, N., VANDERSTRAETEN, K., DIELTJENS, N., MNIF, E., HACHICHA, S., BESBES, G., BEN ARAB, S. & VAN CAMP, G.

2008. A new locus for otosclerosis, OTSC8, maps to the pericentromeric region of chromosome 9. *Hum Genet,* 123**,** 267-72.

BENGTSSON, E., NEAME, P. J., HEINEGARD, D. & SOMMARIN, Y. 1995. The primary structure of a basic leucine-rich repeat protein, PRELP, found in connective tissues. *J Biol Chem,* 270**,** 25639-44.

BENJAMINI, Y., DRAI, D., ELMER, G., KAFKAFI, N. & GOLANI, I. 2001. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res,* 125**,** 279-84.

BIESECKER, L. G. 2010. Exome sequencing makes medical genomics a reality. *Nat Genet,* 42**,** 13-4.

BORNSTEIN, P., TRAUB, W. 1979. *The chemistry and biology of collagen. In The Proteins,* New York, Academic.

BRETLAU, P., CAUSSE, J., CAUSSE, J. B., HANSEN, H. J., JOHNSEN, N. J. & SALOMON, G. 1985. Otospongiosis and sodium fluoride. A blind experimental and clinical evaluation of the effect of sodium fluoride treatment in patients with otospongiosis. *Ann Otol Rhinol Laryngol,* 94**,** 103-7.

BRETLAU, P. & JORGENSEN, M. B. 1968. Otosclerosis in the lateral semi-circular canal. *J Laryngol Otol,* 82**,** 65-7.

BRETLAU, P., SALOMON, G. & JOHNSEN, N. J. 1989. Otospongiosis and sodium fluoride. A clinical double-blind, placebo-controlled study on sodium fluoride treatment in otospongiosis. *Am J Otol,* 10**,** 20-2.

BROWNING, G. G. & GATEHOUSE, S. 1984. Sensorineural hearing loss in stapedial otosclerosis. *Ann Otol Rhinol Laryngol,* 93**,** 13-16.

BROWNING, G. G. & GATEHOUSE, S. 1992. The prevalence of middle ear disease in the adult British population. *Clin Otolaryngol Allied Sci,* 17**,** 317-21.

BROWNSTEIN, Z., GOLDFARB, A., LEVI, H., FRYDMAN, M. & AVRAHAM, K. B. 2006. Chromosomal mapping and phenotypic characterization of hereditary otosclerosis linked to the OTSC4 locus. *Arch Otolaryngol Head Neck Surg,* 132**,** 416-24.

BYERS, P. H. & MURRAY, M. L. 2012. Heritable collagen disorders: the paradigm of the Ehlers-Danlos syndrome. *J Invest Dermatol,* 132**,** E6-11.

CAILLAT-ZUCMAN, S. 2009. Molecular mechanisms of HLA association with autoimmune diseases. *Tissue Antigens,* 73**,** 1-8.

CARHART, R. 1962. Atypical audiometric configurations associated with otosclerosis. *Trans Am Otol Soc,* 50**,** 153-71.

CAULDWELL, E. W., ANSON, B.J. 1942. Stapes, fissula ante fenestram, and associated structure sin man. *Arch Otlaryng,* 36**,** 891-925.

CAUSSE, J. R., SHAMBAUGH, G. E., CAUSSE, B. & BRETLAU, P. 1980. Enzymology of otospongiosis and NaF therapy. *Am J Otol,* 1**,** 206-14.

CAWTHORNE, T. 1955. Otosclerosis. *J Laryngol Otol,* 69**,** 437-56.

CHANG, W., TEN DIJKE, P. & WU, D. K. 2002. BMP pathways are involved in otic capsule formation and epithelial-mesenchymal signaling in the developing chicken inner ear. *Dev Biol,* 251**,** 380-94.

CHAPMAN, S. C. 2011. Can you hear me now? Understanding vertebrate middle ear development. *Front Biosci,* 16**,** 1675-92.

CHEADLE, C., VAWTER, M. P., FREED, W. J. & BECKER, K. G. 2003. Analysis of microarray data using Z score transformation. *J Mol Diagn,* 5**,** 73-81.

CHEN, Q., SIVAKUMAR, P., BARLEY, C., PETERS, D. M., GOMES, R. R., FARACH-CARSON, M. C. & DALLAS, S. L. 2007a. Potential role for heparan sulfate proteoglycans in regulation of transforming growth factor-beta (TGF-beta) by modulating assembly of latent TGF-beta-binding protein-1. *J Biol Chem,* 282**,** 26418-30.

CHEN, R. & BUTTE, A. J. 2011. The reference human genome demonstrates high risk of type 1 diabetes and other disorders. *Pac Symp Biocomput***,** 231-42.

CHEN, W., CAMPBELL, C. A., GREEN, G. E., VAN DEN BOGAERT, K., KOMODIKIS, C., MANOLIDIS, L. S., ACONOMOU, E., KYAMIDES, Y., CHRISTODOULOU, K., FAGHEL, C., GIGUERE, C. M.,

ALFORD, R. L., MANOLIDIS, S., VAN CAMP, G. & SMITH, R. J. 2002. Linkage of otosclerosis to a third locus (OTSC3) on human chromosome 6p21.3-22.3. *J Med Genet,* 39**,** 473-7.

CHEN, W., MEYER, N. C., MCKENNA, M. J., PFISTER, M., MCBRIDE, D. J., JR., FUKUSHIMA, K., THYS, M., CAMP, G. V. & SMITH, R. J. 2007b. Single-nucleotide polymorphisms in the COL1A1 regulatory regions are associated with otosclerosis. *Clin Genet,* 71**,** 406-14.

CHOU, M. Y. & LI, H. C. 2002. Genomic organization and characterization of the human type XXI collagen (COL21A1) gene. *Genomics,* 79**,** 395-401.

CLEMINSON, F. J. 1927. Otosclerosis associated with Blue Sclerotics and Osteogenesis Imperfecta. *Proc R Soc Med,* 20**,** 471-4.

COLE, W. G. & DALGLEISH, R. 1995. Perinatal lethal osteogenesis imperfecta. *J Med Genet,* 32**,** 284-9.

COOPER, D. N., KRAWCZAK, M., ANTONARAKIS, S.E. 1998. *The nature and mechanisms of human gene mutation. Chapter 3 in The Genetic Basis of Human Cancer,* New York, Mc Graw-Hill.

COOPER, G. S. & STROEHLA, B. C. 2003. The epidemiology of autoimmune diseases. *Autoimmun Rev,* 2**,** 119-25.

COSTA, V., APRILE, M., ESPOSITO, R. & CICCODICOLA, A. 2013. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *Eur J Hum Genet,* 21**,** 134-42.

COULY, G. F., COLTEY, P. M. & LE DOUARIN, N. M. 1993. The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. *Development,* 117**,** 409-29.

CRUISE, A. S., SINGH, A. & QUINEY, R. E. 2010. Sodium fluoride in otosclerosis treatment: review. *J Laryngol Otol,* 124**,** 583-6.

CSOMOR, P., LIKTOR, B., SZIKLAI, I. & KAROSI, T. 2012. No evidence for disturbed COL1A1 and A2 expression in otosclerosis. *Eur Arch Otorhinolaryngol,* 269**,** 2043-51.

CSOMOR, P., SZALMAS, A., KONYA, J., SZIKLAI, I. & KAROSI, T. 2010. Restriction analysis of otosclerosis-associated CD46 splicing variants. *Eur Arch Otorhinolaryngol,* 267**,** 219-26.

CSOMOR, P., SZIKLAI, I. & KAROSI, T. 2011. Controversies in RELN/reelin expression in otosclerosis. *Eur Arch Otorhinolaryngol*.

CUNDY, T. 2012. Recent advances in osteogenesis imperfecta. *Calcif Tissue Int,* 90**,** 439-49.

DALLOS, P. 1996. Overview: cochlear neurophysiology. *Springer Handbook of Auditory Research: The Cochlea.* Berlin: Springer.

DALY, A. K. & DAY, C. P. 2001. Candidate gene case-control association studies: advantages and potential pitfalls. *Br J Clin Pharmacol,* 52**,** 489-99.

DANIEL, H. J., 3RD 1969. Stapedial otosclerosis and fluorine in the drinking water. *Arch Otolaryngol,* 90**,** 585-9.

DAVENPORT, C. B., LLOYD MILES, B., FRINK, L. B. 1933. The genetic factor in otosclerosis. *Archives of otolaryngology,* 17**,** 503-548.

DAVIS, A., SMITH, P. A., BOOTH, M. & MARTIN, M. 2012. Diagnosing Patients with Age-Related Hearing Loss and Tinnitus: Supporting GP Clinical Engagement through Innovation and Pathway Redesign in Audiology Services. *Int J Otolaryngol,* 2012**,** 290291.

DECLAU, F., JACOB, W., DORRINE, W., APPEL, B. & MARQUET, J. 1989. Early ossification within the human fetal otic capsule: morphological and microanalytical findings. *J Laryngol Otol,* 103**,** 1113-21.

DECLAU, F., VAN DEN BOGAERT, K., VAN DE HEYNING, P., OFFECIERS, E., GOVAERTS, P. & VAN CAMP, G. 2007a. Phenotype-genotype correlations in otosclerosis: clinical features of OTSC2. *Adv Otorhinolaryngol,* 65**,** 114-8.

DECLAU, F., VAN SPAENDONCK, M., TIMMERMANS, J. P., MICHAELS, L., LIANG, J., QIU, J. P. & VAN DE HEYNING, P. 2007b. Prevalence of histologic otosclerosis: an unbiased temporal bone study in Caucasians. *Adv Otorhinolaryngol,* 65**,** 6-16.

DRIRA, E. A. Etude genetique et localisation d'un gene de l'otospongiose chez les familles Tunisiennes. 06 Congres Francais d'Oto-Rhino-Laryngologie et de Chirurgie de la Face et du Cou., 1999 Paris.

DROR, A. A. & AVRAHAM, K. B. 2010. Hearing impairment: a panoply of genes and functions. *Neuron,* 68**,** 293-308.

EALY, M. 2011. *Otosclerosis-identifying genetic contributions to a complex heaing disorder.* University of Iowa.

EALY, M., CHEN, W., RYU, G. Y., YOON, J. G., WELLING, D. B., HANSEN, M., MADAN, A. & SMITH, R. J. 2008. Gene expression analysis of human otosclerotic stapedial footplates. *Hear Res,* 240**,** 80-6.

EALY, M. & SMITH, R. J. 2011. Otosclerosis. *Adv Otorhinolaryngol,* 70**,** 122-9.

EMMETT, J. R. 1993. Physical examination and clinical evaluation of the patient with otosclerosis. *Otolaryngol Clin North Am,* 26**,** 353-7.

ENDO, H., IKEDA, K., URANO, T., HORIE-INOUE, K. & INOUE, S. 2012. Terf/TRIM17 stimulates degradation of kinetochore protein ZWINT and regulates cell proliferation. *J Biochem,* 151**,** 139-44.

ERTUGAY, O. C., ATA, P., KALAYCIK ERTUGAY, C., KAYA, K. S., TATLIPINAR, A. & KULEKCI, S. 2013. Association of COL1A1 polymorphism in Turkish patients with otosclerosis. *Am J Otolaryngol*.

FAN, X. & TANG, L. 2013. Aberrant and alternative splicing in skeletal system disease. *Gene,* 528**,** 21-6.

FITZGERALD, J. & BATEMAN, J. F. 2004. Why mice have lost genes for COL21A1, STK17A, GPR145 and AHRI: evidence for gene deletion at evolutionary breakpoints in the rodent lineage. *Trends Genet,* 20**,** 408-12.

FRISCH, T., SORENSEN, M. S., OVERGAARD, S. & BRETLAU, P. 2000. Predilection of otosclerotic foci related to the bone turnover in the otic capsule. *Acta Otolaryngol Suppl,* 543**,** 111-3.

GAPANY-GAPANAVICIUS, B. 1975. The incidence of otosclerosis in the general population. *Isr J Med Sci,* 11**,** 465-8.

GENTILE, F. V., ZUNTINI, M., PARRA, A., BATTISTELLI, L., PANDOLFI, M., PALS, G. & SANGIORGI, L. 2012. Validation of a quantitative PCR-high-resolution melting protocol for simultaneous screening of COL1A1 and COL1A2 point mutations and large rearrangements: application for diagnosis of osteogenesis imperfecta. *Hum Mutat,* 33**,** 1697-707.

GIUDICESSI, J. R. & ACKERMAN, M. J. 2013. Determinants of incomplete penetrance and variable expressivity in heritable cardiac arrhythmia syndromes. *Transl Res,* 161**,** 1-14.

GORDON, M. A. 1989. The genetics of otosclerosis: a review. *Am J Otol,* 10**,** 426-38.

GRANT, S. F., REID, D. M., BLAKE, G., HERD, R., FOGELMAN, I. & RALSTON, S. H. 1996. Reduced bone density and osteoporosis associated with a polymorphic Sp1 binding site in the collagen type I alpha 1 gene. *Nat Genet,* 14**,** 203-5.

GRAYELI, A. B., ESCOUBET, B., BICHARA, M., JULIEN, N., SILVE, C., FRIEDLANDER, G., STERKERS, O. & FERRARY, E. 2003. Increased activity of the diastrophic dysplasia sulfate transporter in otosclerosis and its inhibition by sodium fluoride. *Otol Neurotol,* 24**,** 854-62.

GRAYELI, A. B., PALMER, P., TRAN BA HUY, P., SOUDANT, J., STERKERS, O., LEBON, P. & FERRARY, E. 2000. No evidence of measles virus in stapes samples from patients with otosclerosis. *J Clin Microbiol,* 38**,** 2655-60.

GREGORIADIS, S., ZERVAS, J., VARLETZIDIS, E., TOUBIS, M., PANTAZOPOULOS, P. & FESSAS, P. 1982. HLA antigens and otosclerosis. A possible new genetic factor. *Arch Otolaryngol,* 108**,** 769-71.

GRISTWOOD, R. E. & VENABLES, W. N. 1983. Pregnancy and otosclerosis. *Clin Otolaryngol Allied Sci,* 8**,** 205-10.

GU, C., LI, X., TAN, Q., WANG, Z., CHEN, L. & LIU, Y. 2013. MiR-183 family regulates chloride intracellular channel 5 expression in inner ear hair cells. *Toxicol In Vitro,* 27**,** 486-91.

GUINAN, J. J., JR., SALT, A. & CHEATHAM, M. A. 2012. Progress in cochlear physiology after Bekesy. *Hear Res,* 293**,** 12-20.

HADJIDAKIS, D. J. & ANDROULAKIS, II 2006. Bone remodeling. *Ann N Y Acad Sci,* 1092**,** 385-96.

HAJIAGHAYI, M., CONDON, A. & HOOS, H. H. 2012. Analysis of energy-based algorithms for RNA secondary structure prediction. *BMC Bioinformatics,* 13**,** 22.

HAMILTON, W. I., MOSSMAN, H.W. 1975. *Embriologia Humana. 4th.,* Buenos Aires, Intermedica.

HELLEMANS, J., PREOBRAZHENSKA, O., WILLAERT, A., DEBEER, P., VERDONK, P. C., COSTA, T., JANSSENS, K., MENTEN, B., VAN ROY, N., VERMEULEN, S. J., SAVARIRAYAN, R., VAN HUL, W., VANHOENACKER, F., HUYLEBROECK, D., DE PAEPE, A., NAEYAERT, J. M., VANDESOMPELE, J., SPELEMAN, F., VERSCHUEREN, K., COUCKE, P. J. & MORTIER, G. R. 2004. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nat Genet,* 36**,** 1213-8.

HENRY, K. R. & CHOLE, R. A. 1987. Genetic and functional analysis of the otosclerosis-like condition of the LP/J mouse. *Audiology,* 26**,** 44-55.

HILGERT, N., SMITH, R. J. & VAN CAMP, G. 2009. Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? *Mutat Res,* 681**,** 189-96.

HONG, E. P. & PARK, J. W. 2012. Sample size and statistical power calculation in genetic association studies. *Genomics Inform,* 10**,** 117-22.

IMAUCHI, Y., JEUNEMAITRE, X., BOUSSION, M., FERRARY, E., STERKERS, O. & GRAYELI, A. B. 2008. Relation between renin-angiotensin-aldosterone system and otosclerosis: a genetic association and in vitro study. *Otol Neurotol,* 29**,** 295-301.

JOHNSON, E. W. 1993. Hearing aids and otosclerosis. *Otolaryngol Clin North Am,* 26**,** 491-502.

KANIFF, T. E., SCHNEIDER, G., MATZ, G. J. 2001. The incisor absent rat: an animal model for the study of otosclerosis. *Otolaryngol Head Neck Surg*, 103:406-12

KAPINAS, K. & DELANY, A. M. 2011. MicroRNA biogenesis and regulation of bone remodeling. *Arthritis Res Ther,* 13**,** 220.

KAPUR, Y. P. & PATT, A. J. 1967. Hearing in Todas of South India. *Arch Otolaryngol,* 85**,** 400-6.

KAROSI, T., JOKAY, I., KONYA, J., PETKO, M., SZABO, L. Z. & SZIKLAI, I. 2007. Expression of measles virus receptors in otosclerotic, non-otosclerotic and in normal stapes footplates. *Eur Arch Otorhinolaryngol,* 264**,** 607-13.

KAROSI, T., KONYA, J., PETKO, M., SZABO, L. Z., PYTEL, J., JORI, J. & SZIKLAI, I. 2006. Antimeasles immunoglobulin g for serologic diagnosis of otosclerotic hearing loss. *Laryngoscope,* 116**,** 488-93.

KAROSI, T., SZALMAS, A., CSOMOR, P., KONYA, J., PETKO, M. & SZIKLAI, I. 2008. Disease-associated novel CD46 splicing variants and pathologic bone remodeling in otosclerosis. *Laryngoscope,* 118**,** 1669-76.

KAWAI, M., DEVLIN, M. J. & ROSEN, C. J. 2009. Fat targets for skeletal health. *Nat Rev Rheumatol,* 5**,** 365-72.

KENNEDY, D. W., HOFFER, M. E. & HOLLIDAY, M. 1993. The effects of etidronate disodium on progressive hearing loss from otosclerosis. *Otolaryngol Head Neck Surg,* 109**,** 461-7.

KHALFALLAH, A., SCHRAUWEN, I., MNAJA, M., FRANSEN, E., LAHMAR, I., EALY, M., DHOUIB, L., AYADI, H., CHARFEDINE, I., DRISS, N., GHORBEL, A., SMITH, R. J., MASMOUDI, S. & VAN CAMP, G. 2010. Genetic variants in RELN are associated with otosclerosis in a non-European population from Tunisia. *Ann Hum Genet,* 74**,** 399-405.

KOLLER, D. L., PEACOCK, M., LAI, D., FOROUD, T. & ECONS, M. J. 2004. False positive rates in association studies as a function of degree of stratification. *J Bone Miner Res,* 19**,** 1291-5.

KORNBLIHTT, A. R., SCHOR, I. E., ALLO, M., DUJARDIN, G., PETRILLO, E. & MUNOZ, M. J. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol,* 14**,** 153-65.

KOU, I., NAKAJIMA, M. & IKEGAWA, S. 2010. Binding characteristics of the osteoarthritis-associated protein asporin. *J Bone Miner Metab,* 28**,** 395-402.

KOZAK, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell,* 44**,** 283-92.

LAMBERT DE ROUVROIT, C., BERNIER, B., ROYAUX, I., DE BERGEYCK, V. & GOFFINET, A. M. 1999. Evolutionarily conserved, alternative splicing of reelin during brain development. *Exp Neurol,* 156**,** 229-38.

LANDER, E. & KRUGLYAK, L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet,* 11**,** 241-7.

LARSSON, A. 1960. Otosclerosis. A genetic and clinical study. *Acta Otolaryngol Suppl,* 154**,** 1-86.

LASSOT, I., ROBBINS, I., KRISTIANSEN, M., RAHMEH, R., JAUDON, F., MAGIERA, M. M., MORA, S., VANHILLE, L., LIPKIN, A., PETTMANN, B., HAM, J. & DESAGHER, S. 2010. Trim17, a novel E3 ubiquitin-ligase, initiates neuronal apoptosis. *Cell Death Differ,* 17**,** 1928-41.

LEE, K. J. 1999. *Essential Otolaryngology Head and Neck Surgery,* Stamford Connecticut, Appleton and Lange.

LETWIN, N. E., KAFKAFI, N., BENJAMINI, Y., MAYO, C., FRANK, B. C., LUU, T., LEE, N. H. & ELMER, G. I. 2006. Combined application of behavior genetics and microarray analysis to identify regional expression themes and gene-behavior associations. *J Neurosci,* 26**,** 5277-87.

LEWIS, M. A., QUINT, E., GLAZIER, A. M., FUCHS, H., DE ANGELIS, M. H., LANGFORD, C., VAN DONGEN, S., ABREU-GOODGER, C., PIIPARI, M., REDSHAW, N., DALMAY, T., MORENO-PELAYO, M. A., ENRIGHT, A. J. & STEEL, K. P. 2009. An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat Genet,* 41**,** 614-8.

LI, H., KLOOSTERMAN, W. & FEKETE, D. M. 2010. MicroRNA-183 family members regulate sensorineural fates in the inner ear. *J Neurosci,* 30**,** 3254-63.

LIKTOR, B., CSOMOR, P., SZASZ, C. S., SZIKLAI, I. & KAROSI, T. 2013. No evidence for the expression of renin-angiotensin-aldosterone system in otosclerotic stapes footplates. *Otol Neurotol,* 34**,** 808-15.

LIM, D. J. 1986. Functional structure of the organ of Corti: a review. *Hear Res,* 22**,** 117-46.

LINTHICUM, F. H., JR. 1993. Histopathology of otosclerosis. *Otolaryngol Clin North Am,* 26**,** 335-52.

LIPPY, W. H., BERENHOLZ, L. P., SCHURING, A. G. & BURKEY, J. M. 2005. Does pregnancy affect otosclerosis? *Laryngoscope,* 115**,** 1833-6.

LOLOV, S. R., ENCHEVA, V. I., KYURKCHIEV, S. D., EDREV, G. E. & KEHAYOV, I. R. 2001. Antimeasles immunoglobulin G in sera of patients with otosclerosis is lower than that in healthy people. *Otol Neurotol,* 22**,** 766-70.

MAJEWSKI, J., SCHWARTZENTRUBER, J., LALONDE, E., MONTPETIT, A. & JABADO, N. 2011. What can exome sequencing do for you? *J Med Genet,* 48**,** 580-9.

MARINI, J. C., FORLINO, A., CABRAL, W. A., BARNES, A. M., SAN ANTONIO, J. D., MILGROM, S., HYLAND, J. C., KORKKO, J., PROCKOP, D. J., DE PAEPE, A., COUCKE, P., SYMOENS, S., GLORIEUX, F. H., ROUGHLEY, P. J., LUND, A. M., KUURILA-SVAHN, K., HARTIKKA, H., COHN, D. H., KRAKOW, D., MOTTES, M., SCHWARZE, U., CHEN, D., YANG, K., KUSLICH, C., TROENDLE, J., DALGLEISH, R. & BYERS, P. H. 2007. Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat,* 28**,** 209-21.

MCGETTIGAN, P. A. 2013. Transcriptomics in the RNA-seq era. *Curr Opin Chem Biol,* 17**,** 4-11.

MCKENNA, M. J., KRISTIANSEN, A. G., BARTLEY, M. L., ROGUS, J. J. & HAINES, J. L. 1998. Association of COL1A1 and otosclerosis: evidence for a shared genetic etiology with mild osteogenesis imperfecta. *Am J Otol,* 19**,** 604-10.

MCKENNA, M. J., KRISTIANSEN, A. G. & HAINES, J. 1996. Polymerase chain reaction amplification of a measles virus sequence from human temporal bone sections with active otosclerosis. *Am J Otol,* 17**,** 827-30.

MCKENNA, M. J., KRISTIANSEN, A. G. & TROPITZSCH, A. S. 2002. Similar COL1A1 expression in fibroblasts from some patients with clinical otosclerosis and those with type I osteogenesis imperfecta. *Ann Otol Rhinol Laryngol,* 111**,** 184-9.

MCKENNA, M. J., NGUYEN-HUYNH, A. T. & KRISTIANSEN, A. G. 2004. Association of otosclerosis with Sp1 binding site polymorphism in COL1A1 gene: evidence for a shared genetic etiology with osteoporosis. *Otol Neurotol,* 25**,** 447-50.

MEYER, C., NOTARI, L. & BECERRA, S. P. 2002. Mapping the type I collagen-binding site on pigment epithelium-derived factor. Implications for its antiangiogenic activity. *J Biol Chem,* 277**,** 45400-7.

MEZZEDIMI, C., PASSALI, D. 2013. Van der Hoeve Syndrome and Stapes Surgery: Case Reports and a Review. *Journal of Rhinolaryngo-Otologies,* 1**,** 11-16.

MIYAJIMA, C., ISHIMOTO, S. & YAMASOBA, T. 2007. Otosclerosis associated with Ehlers-Danlos syndrome: report of a case. *Acta Otolaryngol Suppl***,** 157-9.

MORRISON, A. W. 1967. Genetic factors in otosclerosis. *Ann R Coll Surg Engl,* 41**,** 202-37.

MOUMOULIDIS, I., AXON, P., BAGULEY, D. & REID, E. 2007. A review on the genetics of otosclerosis. *Clin Otolaryngol,* 32**,** 239-47.

NAGER, G. T. 1969. Histopathology of otosclerosis. *Arch Otolaryngol,* 89**,** 341-63.

NAKANO, Y., KOHNO, T., HIBI, T., KOHNO, S., BABA, A., MIKOSHIBA, K., NAKAJIMA, K. & HATTORI, M. 2007. The extremely conserved C-terminal region of Reelin is not necessary for secretion but is required for efficient activation of downstream signaling. *J Biol Chem,* 282**,** 20544-52.

NEALE, B. M. & SHAM, P. C. 2004. The future of association studies: gene-based analysis and replication. *Am J Hum Genet,* 75**,** 353-62.

NG, S. B., TURNER, E. H., ROBERTSON, P. D., FLYGARE, S. D., BIGHAM, A. W., LEE, C., SHAFFER, T., WONG, M., BHATTACHARJEE, A., EICHLER, E. E., BAMSHAD, M., NICKERSON, D. A. & SHENDURE, J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature,* 461**,** 272-6.

NICHOLSON, S. E. & HILTON, D. J. 1998. The SOCS proteins: a new family of negative regulators of signal transduction. *J Leukoc Biol,* 63**,** 665-8.

NIEDERMEYER, H., ARNOLD, W., NEUBERT, W. J. & HOFLER, H. 1994. Evidence of measles virus RNA in otosclerotic tissue. *ORL J Otorhinolaryngol Relat Spec,* 56**,** 130-2.

NIEDERMEYER, H. P., BECKER, E. T. & ARNOLD, W. 2007. Expression of collagens in the otosclerotic bone. *Adv Otorhinolaryngol,* 65**,** 45-9.

OGAWA, S., GOTO, W., ORIMO, A., HOSOI, T., OUCHI, Y., MURAMATSU, M. & INOUE, S. 1998. Molecular cloning of a novel RING finger-B box-coiled coil (RBCC) protein, terf, expressed in the testis. *Biochem Biophys Res Commun,* 251**,** 515-9.

OSHLACK, A. & WAKEFIELD, M. J. 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct,* 4**,** 14.

OVADIA, G. & SHIFMAN, S. 2011. The genetic variation of RELN expression in schizophrenia and bipolar disorder. *PLoS One,* 6**,** e19955.

OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet,* 12**,** 87-98.

PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet,* 40**,** 1413-5.

PEARSON, R. D., KURLAND, L. T. & CODY, D. T. 1974. Incidence of diagnosed clinical otosclerosis. *Arch Otolaryngol,* 99**,** 288-91.

PICHICHERO, M. E. 2013. Otitis media. *Pediatr Clin North Am,* 60**,** 391-407.

PIETTE-BRION, B., LOWY-MOTULSKY, M., LEDOUX-CORBUSIER, M. & ACHTEN, G. 1984. [Dermatofibromas, elastomas and deafness: a new case of the Buschke-Ollendorff syndrome]. *Dermatologica,* 168**,** 255-8.

PODOSHIN, L., GERTNER, R., FRADIS, M., FEIGLIN, H., EIBSCHITZ, I., SHARF, M. & REITER, A. 1978. Oral contraceptive pills and clinical otosclerosis. *Int J Gynaecol Obstet,* 15**,** 554-5.

POLITZER, A. 1894. *English translation of Text-book of the Diseases of the Ear and Adjacent Organs:For students and Practicioners, 3rd ed,* London:Bailliere, Tindall & Cox.

PRIYADARSHI, S., PANDA, K. C., PANDA, A. K. & RAMCHANDER, P. V. 2010. Lack of association between SNP rs3914132 of the RELN gene and otosclerosis in India. *Genet Mol Res,* 9**,** 1914-20.

PRIYADARSHI, S., RAY, C. S., PANDA, K. C., DESAI, A., NAYAK, S. R., BISWAL, N. C. & RAMCHANDER, P. V. 2013. Genetic association and gene expression analysis of TGFB1 contributes towards the susceptibility to otosclerosis. *J Bone Miner Res*.

QI, Y. X., LIU, Y. B. & RONG, W. H. 2011. [RNA-Seq and its applications: a new technology for transcriptomics]. *Yi Chuan,* 33**,** 1191-202.

QUATTROCCHI, C. C., WANNENES, F., PERSICO, A. M., CIAFRE, S. A., D'ARCANGELO, G., FARACE, M. G. & KELLER, F. 2002. Reelin is a serine protease of the extracellular matrix. *J Biol Chem,* 277**,** 303-9.

QUESNEL, A. M., SETON, M., MERCHANT, S. N., HALPIN, C. & MCKENNA, M. J. 2012. Third-generation bisphosphonates for treatment of sensorineural hearing loss in otosclerosis. *Otol Neurotol,* 33**,** 1308-14.

RABBANI, B., MAHDIEH, N., HOSOMICHI, K., NAKAOKA, H. & INOUE, I. 2012. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet,* 57**,** 621-32.

RAOUF, A., ZHAO, Y., TO, K., STINGL, J., DELANEY, A., BARBARA, M., ISCOVE, N., JONES, S., MCKINNEY, S., EMERMAN, J., APARICIO, S., MARRA, M. & EAVES, C. 2008. Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell,* 3**,** 109-18.

RAUCH, F., HUSSEINI, A., ROUGHLEY, P., GLORIEUX, F. H. & MOFFATT, P. 2012. Lack of circulating pigment epithelium-derived factor is a marker of osteogenesis imperfecta type VI. *J Clin Endocrinol Metab,* 97**,** E1550-6.

RHODES, D. A., DE BONO, B. & TROWSDALE, J. 2005. Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence? *Immunology,* 116**,** 411-7.

RISSO, D., SCHWARTZ, K., SHERLOCK, G. & DUDOIT, S. 2011. GC-content normalization for RNA-Seq data. *BMC Bioinformatics,* 12**,** 480.

RODRIGUEZ-VAZQUEZ, J. F. 2005. Development of the stapes and associated structures in human embryos. *J Anat,* 207**,** 165-73.

RODRIGUEZ, L., RODRIGUEZ, S., HERMIDA, J., FRADE, C., SANDE, E., VISEDO, G., MARTIN, C. & ZAPATA, C. 2004. Proposed association between the COL1A1 and COL1A2 genes and otosclerosis is not supported by a case-control study in Spain. *Am J Med Genet A,* 128A**,** 19-22.

SAEED, S. R. 2002. *The genetics of otosclerosis.* University of Manchester.

SCHENA, D., GERMI, L., ZAMPERETTI, M. R., COLATO, C. & GIROLOMONI, G. 2008. Buschke-Ollendorff syndrome. *Int J Dermatol,* 47**,** 1159-61.

SCHNUR, R. E., GRACE, K. & HERZBERG, A. 1994. Buschke-Ollendorff syndrome, otosclerosis, and congenital spinal stenosis. *Pediatr Dermatol,* 11**,** 31-4.

SCHRAUWEN, I., EALY, M., FRANSEN, E., VANDERSTRAETEN, K., THYS, M., MEYER, N. C., COSGAREA, M., HUBER, A., MAZZOLI, M., PFISTER, M., SMITH, R. J. & VAN CAMP, G. 2010a. Genetic variants in the RELN gene are associated with otosclerosis in multiple European populations. *Hum Genet,* 127**,** 155-62.

SCHRAUWEN, I., EALY, M., HUENTELMAN, M. J., THYS, M., HOMER, N., VANDERSTRAETEN, K., FRANSEN, E., CORNEVEAUX, J. J., CRAIG, D. W., CLAUSTRES, M., CREMERS, C. W.,

DHOOGE, I., VAN DE HEYNING, P., VINCENT, R., OFFECIERS, E., SMITH, R. J. & VAN CAMP, G. 2009a. A genome-wide analysis identifies genetic variants in the RELN gene associated with otosclerosis. *Am J Hum Genet,* 84**,** 328-38.

SCHRAUWEN, I., KHALFALLAH, A., EALY, M., FRANSEN, E., CLAES, C., HUBER, A., MURILLO, L. R., MASMOUDI, S., SMITH, R. J. & VAN CAMP, G. 2012. COL1A1 association and otosclerosis: a meta-analysis. *Am J Med Genet A,* 158A**,** 1066-70.

SCHRAUWEN, I., THYS, M., VANDERSTRAETEN, K., FRANSEN, E., DIELTJENS, N., HUYGHE, J. R., EALY, M., CLAUSTRES, M., CREMERS, C. R., DHOOGE, I., DECLAU, F., VAN DE HEYNING, P., VINCENT, R., SOMERS, T., OFFECIERS, E., SMITH, R. J. & VAN CAMP, G. 2008. Association of bone morphogenetic proteins with otosclerosis. *J Bone Miner Res,* 23**,** 507-16.

SCHRAUWEN, I., THYS, M., VANDERSTRAETEN, K., FRANSEN, E., EALY, M., CREMERS, C. W., DHOOGE, I., VAN DE HEYNING, P., OFFECIERS, E., SMITH, R. J. & VAN CAMP, G. 2009b. No evidence for association between the renin-angiotensin-aldosterone system and otosclerosis in a large Belgian-Dutch population. *Otol Neurotol,* 30**,** 1079-83.

SCHRAUWEN, I. & VAN CAMP, G. 2010. The etiology of otosclerosis: a combination of genes and environment. *Laryngoscope,* 120**,** 1195-202.

SCHRAUWEN, I., VENKEN, K., VANDERSTRAETEN, K., THYS, M., HENDRICKX, J. J., FRANSEN, E., VAN LAER, L., GOVAERTS, P. J., VERSTREKEN, M., SCHATTEMAN, I., STINISSEN, P., HELLINGS, N. & VAN CAMP, G. 2010b. Involvement of T-cell receptor-beta alterations in the development of otosclerosis linked to OTSC2. *Genes Immun,* 11**,** 246-53.

SCHRAUWEN, I., WEEGERINK, N. J., FRANSEN, E., CLAES, C., PENNINGS, R. J., CREMERS, C. W., HUYGEN, P. L., KUNST, H. P. & VAN CAMP, G. 2011. A new locus for otosclerosis, OTSC10, maps to chromosome 1q41-44. *Clin Genet,* 79**,** 495-7.

SCHRIER, R. W. & DURR, J. A. 1987. Pregnancy: an overfill or underfill state. *Am J Kidney Dis,* 9**,** 284-9.

SCHUKNECHT, H. F. & GACEK, M. R. 1993. Cochlear pathology in presbycusis. *Ann Otol Rhinol Laryngol,* 102**,** 1-16.

SEKIYA, A., OKANO-KOSUGI, H., YAMAZAKI, C. M. & KOIDE, T. 2011. Pigment epithelium-derived factor (PEDF) shares binding sites in collagen with heparin/heparan sulfate proteoglycans. *J Biol Chem,* 286**,** 26364-74.

SHAHEEN, R., ALAZAMI, A. M., ALSHAMMARI, M. J., FAQEIH, E., ALHASHMI, N., MOUSA, N., ALSINANI, A., ANSARI, S., ALZAHRANI, F., AL-OWAIN, M., ALZAYED, Z. S. & ALKURAYA, F. S. 2012. Study of autosomal recessive osteogenesis imperfecta in Arabia reveals a novel locus defined by TMEM38B mutation. *J Med Genet,* 49**,** 630-5.

SHAMBAUGH, G. E., JR. & SCOTT, A. 1964. Sodium Fluoride for Arrest of Otosclerosis; Theoretical Considerations. *Arch Otolaryngol,* 80**,** 263-70.

SHAO, H., SCHVARTZ, I. & SHALTIEL, S. 2003. Secretion of pigment epithelium-derived factor. Mutagenic study. *Eur J Biochem,* 270**,** 822-31.

SHUKLA, D. & SPEAR, P. G. 2001. Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. *J Clin Invest,* 108**,** 503-10.

SMYTH, G. D. L. (ed.) 1997. *Scott Brown's Otolaryngology In:Otoslerosis,* Oxford: Butterworth-Heinmann.

SOLDA, G., ROBUSTO, M., PRIMIGNANI, P., CASTORINA, P., BENZONI, E., CESARANI, A., AMBROSETTI, U., ASSELTA, R. & DUGA, S. 2012. A novel mutation within the MIR96 gene causes non-syndromic inherited hearing loss in an Italian family by altering pre-miRNA processing. *Hum Mol Genet,* 21**,** 577-85.

STROSBERG, J. M. & ADLER, R. G. 1981. Otosclerosis associated with osteopoikilosis. *JAMA,* 246**,** 2030-1.

TAKADA, I., KOUZMENKO, A. P. & KATO, S. 2009. Wnt and PPARgamma signaling in osteoblastogenesis and adipogenesis. *Nat Rev Rheumatol,* 5**,** 442-7.

TANG, S. Y., ALLISTON, T. 2013. Regulation of postnatal bone homeostasis by TGFβ. *BoneKEy Reports* 2.

TANGE, R. A., DE BRUIJN, A. J. & GROLMAN, W. 1998. Experience with a new pure gold piston in stapedotomy for cases of otosclerosis. *Auris Nasus Larynx,* 25**,** 249-53.

TAZI, J., BAKKOUR, N. & STAMM, S. 2009. Alternative splicing and disease. *Biochim Biophys Acta,* 1792**,** 14-26.

THAMJARAYAKUL, T., SUPIYAPHUN, P. & SNIDVONGS, K. 2010. Stapes fixation surgery: stapedectomy versus stapedotomy. *Asian Biomedicine,* 4**,** 429-434.

THEANDER, E. & JACOBSSON, L. T. 2008. Relationship of Sjogren's syndrome to other connective tissue and autoimmune disorders. *Rheum Dis Clin North Am,* 34**,** 935-47, viii-ix.

THOMPSON, H., OHAZAMA, A., SHARPE, P. T. & TUCKER, A. S. 2012. The origin of the stapes and relationship to the otic capsule and oval window. *Dev Dyn,* 241**,** 1396-404.

THYS, M., SCHRAUWEN, I., VANDERSTRAETEN, K., DIELTJENS, N., FRANSEN, E., EALY, M., CREMERS, C. W., VAN DE HEYNING, P., VINCENT, R., OFFECIERS, E., SMITH, R. H. & VAN CAMP, G. 2009. Detection of rare nonsynonymous variants in TGFB1 in otosclerosis patients. *Ann Hum Genet,* 73**,** 171-5.

THYS, M., SCHRAUWEN, I., VANDERSTRAETEN, K., JANSSENS, K., DIELTJENS, N., VAN DEN BOGAERT, K., FRANSEN, E., CHEN, W., EALY, M., CLAUSTRES, M., CREMERS, C. R., DHOOGE, I., DECLAU, F., CLAES, J., VAN DE HEYNING, P., VINCENT, R., SOMERS, T., OFFECIERS, E., SMITH, R. J. & VAN CAMP, G. 2007a. The coding polymorphism T263I in TGF-beta1 is associated with otosclerosis in two independent populations. *Hum Mol Genet,* 16**,** 2021-30.

THYS, M. & VAN CAMP, G. 2009. Genetics of otosclerosis. *Otol Neurotol,* 30**,** 1021-32.

THYS, M., VAN DEN BOGAERT, K., ILIADOU, V., VANDERSTRAETEN, K., DIELTJENS, N., SCHRAUWEN, I., CHEN, W., ELEFTHERIADES, N., GRIGORIADOU, M., PAUW, R. J., CREMERS, C. R., SMITH, R. J., PETERSEN, M. B. & VAN CAMP, G. 2007b. A seventh locus for otosclerosis, OTSC7, maps to chromosome 6q13-16.1. *Eur J Hum Genet,* 15**,** 362-8.

TISSIR, F. & GOFFINET, A. M. 2003. Reelin and brain development. *Nat Rev Neurosci,* 4**,** 496-505.

TOMBRAN-TINK, J., APARICIO, S., XU, X., TINK, A. R., LARA, N., SAWANT, S., BARNSTABLE, C. J. & ZHANG, S. S. 2005. PEDF and the serpins: phylogeny, sequence conservation, and functional domains. *J Struct Biol,* 151**,** 130-50.

TOMBRAN-TINK, J. & BARNSTABLE, C. J. 2004. Osteoblasts and osteoclasts express PEDF, VEGF-A isoforms, and VEGF receptors: possible mediators of angiogenesis and matrix remodeling in the bone. *Biochem Biophys Res Commun,* 316**,** 573-9.

TOMEK, M. S., BROWN, M. R., MANI, S. R., RAMESH, A., SRISAILAPATHY, C. R., COUCKE, P., ZBAR, R. I., BELL, A. M., MCGUIRT, W. T., FUKUSHIMA, K., WILLEMS, P. J., VAN CAMP, G. & SMITH, R. J. 1998. Localization of a gene for otosclerosis to chromosome 15q25-q26. *Hum Mol Genet,* 7**,** 285-90.

UPPAL, S., BAJAJ, Y., RUSTOM, I. & COATESWORTH, A. P. 2009. Otosclerosis 1: the aetiopathogenesis of otosclerosis. *Int J Clin Pract,* 63**,** 1526-30.

VAN DEN BOGAERT, K., DE LEENHEER, E. M., CHEN, W., LEE, Y., NURNBERG, P., PENNINGS, R. J., VANDERSTRAETEN, K., THYS, M., CREMERS, C. W., SMITH, R. J. & VAN CAMP, G. 2004. A fifth locus for otosclerosis, OTSC5, maps to chromosome 3q22-24. *J Med Genet,* 41**,** 450-3.

VAN DEN BOGAERT, K., GOVAERTS, P. J., SCHATTEMAN, I., BROWN, M. R., CAETHOVEN, G., OFFECIERS, F. E., SOMERS, T., DECLAU, F., COUCKE, P., VAN DE HEYNING, P., SMITH, R. J. & VAN CAMP, G. 2001. A second gene for otosclerosis, OTSC2, maps to chromosome 7q34-36. *Am J Hum Genet,* 68**,** 495-500.

VARTIAINEN, E. & VARTIAINEN, J. 1997a. The influence of fluoridation of drinking water on the long-term hearing results of stapedectomy. *Clin Otolaryngol Allied Sci,* 22**,** 34-6.

VARTIAINEN, E. & VARTIAINEN, T. 1997b. Effect of drinking water fluoridation on the prevalence of otosclerosis. *J Laryngol Otol,* 111**,** 20-2.

VERSTREKEN, M., CLAES, J. & VAN DE HEYNING, P. H. 1996. Osteogenesis imperfecta and hearing loss. *Acta Otorhinolaryngol Belg,* 50**,** 91-8.

VESSEY, M. & PAINTER, R. 2001. Oral contraception and ear disease: findings in a large cohort study. *Contraception,* 63**,** 61-3.

WADE-GUEYE, N. M., BOUDIFFA, M., VANDEN-BOSSCHE, A., LAROCHE, N., AUBIN, J. E., VICO, L., LAFAGE-PROUST, M. H. & MALAVAL, L. 2012. Absence of bone sialoprotein (BSP) impairs primary bone formation and resorption: the marrow ablation model under PTH challenge. *Bone,* 50**,** 1064-73.

WANG, P. C., MERCHANT, S. N., MCKENNA, M. J., GLYNN, R. J. & NADOL, J. B., JR. 1999. Does otosclerosis occur only in the temporal bone? *Am J Otol,* 20**,** 162-5.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet,* 10**,** 57-63.

WANG, Z., LIU, X., YANG, B. Z. & GELERNTER, J. 2013. The Role and Challenges of Exome Sequencing in Studies of Human Diseases. *Front Genet,* 4**,** 160.

WANGEMANN, P., SCHACHT, J 1996. Cochlear Homeostasis. *The Cochlea. Handbook of Auditory Research.* New York: Springer.

WARD, L. M., LALIC, L., ROUGHLEY, P. J. & GLORIEUX, F. H. 2001. Thirty-three novel COL1A1 and COL1A2 mutations in patients with osteogenesis imperfecta types I-IV. *Hum Mutat,* 17**,** 434.

WEBER, M. 1930. *Archives of Pathology,* 9**,** 984.

WHITSON, R. H., JR., WONG, W. L. & ITAKURA, K. 1991. Platelet factor 4 selectively inhibits binding of TGF-beta 1 to the type I TGF-beta 1 receptor. *J Cell Biochem,* 47**,** 31-42.

WOO, J. S., SUH, H. Y., PARK, S. Y. & OH, B. H. 2006. Structural basis for protein recognition by B30.2/SPRY domains. *Mol Cell,* 24**,** 967-76.

YAMAMOTI, T., YOKOYAMA, A., MAMADA, A., MIYAZAKI, Y. & NISHIOKA, K. 1998. Familial occurrence of coexistence of bullous pemphigoid and Sjogren's syndrome. *Int J Dermatol,* 37**,** 475-6.

YAP, M. W., NISOLE, S. & STOYE, J. P. 2005. A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. *Curr Biol,* 15**,** 73-8.

YOO, T. J., STUART, J. M., KANG, A. H., TOWNES, A. S., TOMODA, K. & DIXIT, S. 1982. Type II collagen autoimmunity in otosclerosis and Meniere's Disease. *Science,* 217**,** 1153-5.

YOO, T. J., TOMODA, K., STUART, J. M., CREMER, M. A., TOWNES, A. S. & KANG, A. H. 1983. Type II collagen-induced autoimmune sensorineural hearing loss and vestibular dysfunction in rats. *Ann Otol Rhinol Laryngol,* 92**,** 267-71.

YUAN, Y., CHEN, H., MA, G., CAO, X. & LIU, Z. 2012. Reelin is involved in transforming growth factor-beta1-induced cell migration in esophageal carcinoma cells. *PLoS One,* 7**,** e31802.

ZECHNER, G. & MOSER, M. 1987. Otosclerosis and mucopolysaccharidosis. *Acta Otolaryngol,* 103**,** 384-6.

ZEHNDER, A. F., KRISTIANSEN, A. G., ADAMS, J. C., KUJAWA, S. G., MERCHANT, S. N. & MCKENNA, M. J. 2006. Osteoprotegrin knockout mice demonstrate abnormal remodeling of the otic capsule and progressive hearing loss. *Laryngoscope,* 116**,** 201-6.

ZEHNDER, A. F., KRISTIANSEN, A. G., ADAMS, J. C., MERCHANT, S. N. & MCKENNA, M. J. 2005. Osteoprotegerin in the inner ear may inhibit bone remodeling in the otic capsule. *Laryngoscope,* 115**,** 172-7.