Whole exome sequencing to investigate genetic variants of non-syndromic hearing impairment in a population of African Ancestry

By Noluthando Manyisa

MNYNOL006

Submitted to the University of Cape Town in fulfilment of requirements of the degree MSc Med in Human Genetics

University of Cape Town, Faculty of Health Sciences, Department of Pathology, Faculty of Human Genetics

Supervisor: Professor Ambroise Wonkam

Co-Supervisors: Professor Collet Dandara and Dr Emile Chimusa

Word Count: 18 772

Date of Submission: February 2018

Table of Contents

Plagiarism declaration

I, Noluthando Manyisa, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 19 February 2018

Acknowledgments

This work could not have been completed without the assistance of my family friends, supervisors and group members.

Thank you to Professor Ambroise Wonkam, Professor Collet Dandara and Dr Emile Chimusa for their unwavering support and mentorship. This project would not have progressed in any way without your inexhaustible efforts.

Thank you to the members of the Wonkam Group, Bioinformatics Group and the Pharmacogenetics Group, both past and present, and Dr Lisa Roberts, Dr Shareefa Dalvie and Dr Gloudi Agenbag; who were always willing to assist in any way possible. Your support has been greatly appreciated.

I owe a huge debt of gratitude to my family and friends, who went above and beyond in order to support me and my aspirations. Your love and support has been invaluable throughout this time.

Thank you to the team of the University of Cape Town's High Performance Computing Facilities (http://hpc.uct.ac.za). The facilities provided were invaluable to the progression of this project.

List of Figures and Tables

Figures

Tables

List of Abbreviations

Abbreviation	Meaning
°C	degrees Celsius
µl	microliter
µM	micromolar
ARNSHL	Autosomal Recessive Non-Syndromic Hearing Loss
BWA	Burrows-Wheeler Aligner
CADD	Combined Annotation Dependent Depletion
dB	Decibels
DNA	Deoxyribonucleic acid
dNTPs	Deoxy nucleotide phosphates
ENT	Ear, nose, and throat
FATHMM	Functional Analysis Through Hidden Markov Models
HI	Hearing Impairment
HIV	Human Immunodeficiency Virus
LR	Logistic Regression
LRT	Likelihood Ratio Test
M	Molar
mM	millimolar
ng	nanogram
OtoSCOPE®	OTOlogic Sequence Capture Of Pathogenic Exons
PCR	Polymerase Chain Reaction
PhyloP	PHYLOgenetic P-values
RadialSVM	radial Support Vector Machine)
RNA	Ribonucleic acid
SIFT	Sorting Intolerant From Tolerant
SiPhy	SIte-specific PHYlogenetic analysis
SNP	Single Nucleotide Polymorphism
SSA	Sub-Saharan Africa
TE Buffer	Tris-EDTA Buffer
TGE	Targeted Genome Enrichment
U	Units
VCF	Variant call file

WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
X	Times

Abstract

Introduction

Hearing impairment occurs when a child has hearing loss greater than 30dB in their better hearing ear and an adult cannot detect sound lower than 40dB in the better hearing ear. It is a common sensory disorder that affecting approximately 360 million worldwide, with an incidence of 6 in 1000 live births in developing countries such as those in Sub-Saharan Africa. 50 % of hearing impairment, in developed countries, is due to genetic factors, with 70% of genetic hearing impairment being classified as non-syndromic hearing impairment, which occurs when the hearing impairment presents with no other clinical manifestations. Hearing impairment is associated with over 150 genes, of which two connexin genes, *GJB2* and *GJB6,* are the most prevalent genes associated with hearing impairment in European, Asian and North American of European ancestries populations. These genes have however been shown to be insignificant causes of Hearing Impairment in African populations

Aim

The aim of this study is to determine the rates for putative pathogenic variants in 172 hearing impairment associated genes, among Cameroonian patients affected by hearing impairment, and non-hearing-impaired controls.

Methods

Patients and controls

Patients were recruited from various schools of the Deaf and Ear, Nose and Throat (ENT) clinics in Cameroon. The patients were examined by qualified medical geneticists and ophthalmologist and detailed family history and medical history was obtained from the patients and their parents. 19 patients, who were negative for *GJB2* and *GJB6* mutations and presented with putative non-syndromic hearing impairment, were selected from a cohort of 582 patients for the present study.

The control population consisted of 130 ethnically matched groups without any personal or familial history of hearing impairment. The controls were recruited from Yaoundé Central Hospital and Laquintinie Hospital in Cameroon.

Whole exome sequencing

DNA was extracted from whole blood using the salting out procedure and the Puregene Blood kit®. The DNA was subjected to spectrometry and gel electrophoresis to determine the quantity and quality of the DNA samples. The samples were then subjected to whole exome sequencing on the Illumina platform using the Nextera Rapid Capture Exome Kit at an average read depth of 30X, whereby only 18 patients were successfully sequenced. The exomes were then subjected to FastQC and SolexaQC++ for quality control measures and aligned to the hg19 reference genome using GATK and VariantMetaCaller.

Bioinformatics analysis

Variant annotation was performed using Annovar and the annotated variants were filtered based in rarity and pathogenicity. Tests for genetic differentiation and principle component analysis was performed on the combined patient exomes and combine control exomes. The first principle component analysis included data from African populations from the 1000 Genomes Phase 3 as well as six control samples from the Democratic Republic of Congo; and the second principle component analysis analysed on the Cameroonian patients and control population. Population structure analysis was followed by protein-protein interaction analysis using custom python and R script and pathway enrichment analysis using Enrichr combined with a second custom R script.

The proportion of derived and ancestral alleles was computed by downloading the SNP ancestral alleles from Ensembl and verifying the presence of the SNPs in dbSNP database. The combined patient and control exomes were annotated using the VCFtools "fillOaa" script. The ancestral alleles were computed by dividing the number of times the alternative allele matched the ancestral allele with the number of copies of all the alternative alleles across all samples at the particular position. The ancestral alleles were categorised into six bins, based on their minor allele frequency, in the patient and control populations and this was used to contrast their proportions of derived and ancestral alleles. Furthermore, the proportion of ancestral and derived alleles in hearing impairment associated genes was computed at SNP based level for the Cameroonian population and contrasted with population from the Democratic Republic of Congo.

Variants validation by Sanger sequencing

Primers were designed to amplify the fragment surrounding the purported SNPs in *MYO15A, MYO3A*, and *COL9A3* as well as for the fragments surrounding the population specific SNPs

in *VTN, RPL3L and DHRS4L2.* Polymerase chain reaction was performed for the *MYO15A,* and *MYO3A* fragments. This was followed by purification of the PCR products and direct cycle Sanger sequencing of the PCR products. The sequencing products were then purified through ethanol precipitation and the fragments were suspended in HiDi Formamide and run on the capillary electrophoresis. The variants in *MYO3A, MYO15A* and *COL9A3* were viewed in Integrated Genomics Viewer using the Bam files as well.

Results

Putative deleterious variants

Single nucleotide polymorphism (SNPs) in *MYO3A, MYO15A* and *COL9A3*, were filtered out as putative causative mutations for three, four and two patients respectively. Direct Sanger Sequencing and viewing the patients BAM files did not confirm the presence of any of these putative pathogenic in the patients. Variations in *USH2A, HSD17B4* and MYO*1A* were also filtered out but these variants were not considered disease causing, after a careful genotype to phenotypes correlations.

Population genetics variants differentiations

At a population level, specific variations were identified in *FOXD4L2, DHRS2L6, RPL3L* and *VTN.* Significant genetic differentiation was shown to exist between the control population and the patients' population with regard to specific variants in *VTN* and *RPL3L;* furthermore, it was shown that these variants in *VTN* and *RPL3L* interact with other hearing impairment associated proteins with evidences that that VTN is hub protein for a hearing impairment associated pathway along with nine other genes. Conversely, this was not the case for variants described in *FOXD4L2* and *DHRS2L6*.

In known hearing Impairment genes, the proportion of ancestral alleles was lowest for the patients' population for variations with minor allele frequencies between 0.0 and 0.1. The proportion of derived and ancestral alleles was also shown to differ between the Cameroonian and the population from the Democratic Republic of Congo, indication possible regional differences in aetiology of Hearing impairment amongst multiple African populations.

Discussion

Low putative pathogenic variants in known hearing impairment genes among Africans

The low pick up rate for putative pathogenic variants in our patients follows a similar trend observed in the African American populations, with hearing impairment, as well as data from

targeted exome sequencing from South African and Nigerian populations. This result is also in agreeance with other studies that interrogated hearing impairment in African populations utilising other means besides next generation sequencing. This result also highlights the importance of validating any results obtained from next generation sequencing through traditional approaches such as Sanger Sequencing or viewing the BAM files on IGV, specifically in African population, poorly represented in Exome databases.

Bioinformatics Analysis Exhibited some Specific Variants among Cameroonian

Protein-protein interactions and enrichment analysis indicated that VTN and RPL3L, and their interacting proteins, are significantly associated with osteoclast differentiation, which is associated with hearing impairment in osteogenesis imperfecta. VTN was further shown as a hub protein of a protein subnetwork, along ATPB2. The presence of a second protein acting as a hub protein may account for why aberrations in VTN have not been associated with a disease; whereby ATPB2 may ameliorate the pathogenic phenotype that ought to be observed in the presence of null mutations in *VTN.*

Evolutionary adaptation of human hearing

Data indicates the patient population carried a higher proportion of derived alleles in known hearing impairment genes, at low minor allele frequencies; possibly indicating, the interactive modifiers capacities of multiple hearing impairment genes, or alternatively, the polygenetic nature of hearing impairment in some patients. The proportion of ancestral and derived alleles was contrasted in the Cameroonian and the population from the Democratic Republic of Congo and it indicated that the variations that may result in hearing impairment in the one population may not be the same variations that result in hearing impairment in the other population Due to this, it is necessary to determine the causative variants resulting in disease in each of these populations independently.

Conclusion and perspectives

The results support a low pick up rate of putative variants in 172 known genes in groups of Cameroonian patients with HI, underscoring the current Targeted panel sequencing for HI may not be relevant for some African populations. The result also support the need of confirmation of variants found in WES, as well careful genotype to phenotypes correlations, particularly among African, whose sequences exome is relatively low in Exomes databases, and as a result could lead to more false positive results. Population genetic analysis has provided novel insight

in the genetic architecture of HI among this group of Africans; particularly, the differential frequencies of ancestral alleles vs derived alleles in HI genes among patients vs controls underline the possibility of multigenic influence on the phenotype of Hearing Impairment that have not been well investigated, and may also signal evolutionary enrichment of some variants of HI genes in the populations as the result of natural selections, that deserve further investigation. The result supports the need of intensive familial studies in multiple African populations in order to unravel the novel genes and those variants that are relevant in clinical practice for people of African ancestry.

Chapter 1: Introduction and literature review

Chapter 1.1: Overview

The review by Wilson and colleagues, published in *The Lancet,* highlights the increasing burden of hearing impairment (HI) and provides recommendations to limit it (Wilson et al., 2017). HI has progressed from the 11th leading cause of disability globally in 2010 to the fourth in 2013 (Murray et al., 2015). More than 80% of people affected by HI live in low and middle-income countries, specifically in sub-Saharan Africa (SSA) with up to 6 per 1000 newborn affected per year (Olusanya et al., 2014). Major efforts are needed to reduce the burden of HI in SSA which is already ill-prepared to cope with more visible conditions such as HIV, malaria, cardiovascular diseases, and cancers. HI is a common sensory impairment affecting approximately 360 million people worldwide with the highest incidence of HI in developing countries. HI is considered disabling HI when a child cannot hear better than 30dB and an adult cannot hear better than 40dB (WHO Media Centre, 2017)

Hearing impairment may arise from genetic, environmental and unknown factors with the majority (50%) of HI cases, in developed countries, being the result of genetic factors (Schrijver, 2004). In developing countries, the environment contributes the majority of HI, but genetics investigations even of familial cases have been lacking and genetic causes underestimated (Wonkam et al., 2013)

Non-syndromic HI accounts for approximately 70% of genetic HI (Schrijver, 2004) and it describes the condition where the HI occurs without any additional clinical manifestations. Of the proportion of non-syndromic HI, 70 to 80% is due to autosomal recessive inheritance (Chung et al., 1959; Morton, 1991). Autosomal dominant inheritance accounts for 15 to 25 percent and sex-linked and mitochondrial inheritance accounts for the remaining proportion of non-syndromic HI (Chung et al., 1959; Morton, 1991)

Genetic HI is associated with more than 150 genes (The Molecular Otolaryngology and Renal Research Laboratories. The University of Iowa, 2016) of which the mutations in the connexin genes *GJB2* and *GJB6* have been shown to prevalent within European, Asian and North American populations (Chan and Chang, 2014; Hutchin et al., 2005; Liu et al., 2002). These genes have, however, been shown to be insignificant in African populations (Bosch et al., 2014a, 2014b; Gasmelseed et al., 2004; Kabahuma et al., 2011; Trotta et al., 2011; Wonkam et al., 2015) and further analysis and study is necessary in order to elucidate the prevalent genes associated with African HI.

The heterogeneity of HI renders traditional molecular techniques obsolete in trying to determine the prevalent genes associated with African HI. Through the use of next generation sequencing, we may be able to determine the prevalent genes associated with African HI. Next generation sequencing (NGS), in particular massively multi-parallel sequencing (MPS), is a sequencing technique that allows the simultaneous sequencing of the entire genome, the entire exome or specific gene regions (Shearer et al., 2010). The various NGS techniques (whole genome, whole exome or targeted genome enrichment) have been used to determine the cause of HI in various populations (Sloan-Heggen et al., 2016; Yan et al., 2016). Targeted Exome Sequencing may be also an effective tools in determining the cause of the HI in some familial cases in African populations (Lebeko et al., 2016). However, to the best of our knowledge, no study has investigated the use of whole exome sequencing in revealing mutation in HI among Africans.

The aim of this study is to determine the rates for putative pathogenic variants in 100 HI associated genes, among Cameroonian patients affected by HI, and non-hearing impaired controls, through the use of whole exome sequencing. The study also seeks to determine the genomic architecture and the evolutionary adaptation of hearing in an African population.

Chapter 1.2: Definition and epidemiology of hearing impairment

HI is so defined when an individual cannot detect sound lower than 25dB in the better hearing ear and is further defined as disabling when there is an inability to hear sound greater than 40dB in an adult or 30dB in a child (WHO Media Centre, 2017).

HI may affect one (unilateral) or both ears (bilateral) and is classified as indicated in Table 1 according to the grade of impairment.

Table 1: Hearing impairment classification

Grade of impairment	Corresponding audiometric ISO value	Performance
0 - No impairment	25 dB or better (better ear)	No or very slight hearing problems. Able to hear whispers.
1 - Slight impairment	26-40 dB (better ear)	Able to hear and repeat words spoken in a normal voice at 1 metre.

2 - Moderate impairment	41-60 dB (better ear)	Able to hear and repeat words spoken in raised voice at 1 metre.
3 - Severe impairment	61-80 dB (better ear)	Able to hear some words when shouted into the better hearing ear.
4 - Profound impairment including deafness	81 dB or greater (better ear)	Unable to hear and understand even a shouted voice.

(Adapted from http://www.who.int/pbd/deafness/hearing_impairment_grades/en/, 2017)

Further classification of HI includes the age of onset (congenital, acquired, prelingual or postlingual), frequencies at which the HI is most pronounced (low, middle or high frequency), type of HI (conductive, sensorineural or mixed), the aetiology of the HI (indicative of the cause of loss of hearing, either genetic factors or environmental factors) (Smith et al., 2005), whether the HI is progressive or stable, and, lastly, whether the HI is syndromic or non-syndromic (Tekin et al., 2001).

Disabling HI affects 360 million people worldwide (WHO Media Centre, 2017). The prevalence of HI, however, varies over population; with 1 affected individual in 500 births in the Middle East (Najmabadi and Kahrizi, 2014) and a 2 in 100 000 births in the Chinese population (Dai et al., 2007). HI occurs in 6 in every 1000 births in developing countries (Olusanya and Newton, 2007) with a 5.5 in 1000 births prevalence in South Africa (Swanepoel et al., 2009).

Chapter 1.3: Normal hearing

The ear is composed of an outer, middle and inner ear (Holme and Steel, 1999; Rędowicz, 1999)and a cross section of the structure of the ear is depicted in Figure 1.1. The outer ear transfers sound vibrations, through the auditory canal, to the tympanic membrane which then vibrates (Holme and Steel, 1999). These vibrations of the tympanic membrane are then amplified by the ossicles (the stapes, malleus and incus) in the middle ear (Holme and Steel, 1999; Tekin et al., 2001). This amplification by the ossicles results in the generation of waves in the perilymph of the cochlea resulting in movement of the tectorial membrane which produces a shearing motion of the hair cells (Tekin et al., 2001). This shearing movement of the hair cells then opens ion channels, resulting in the influx of potassium ions, causing the

depolarisation of the hair cells and the release of neurotransmitters, which activate the acoustic nerve (the eighth cranial nerve) (Holme and Steel, 1999; Ręd owicz, 1999; Tekin et al., 2001). Figure 1.2 indicates a cross section through the cochlea and Figure 1.3 indicates a schematic diagram of a hair cell.

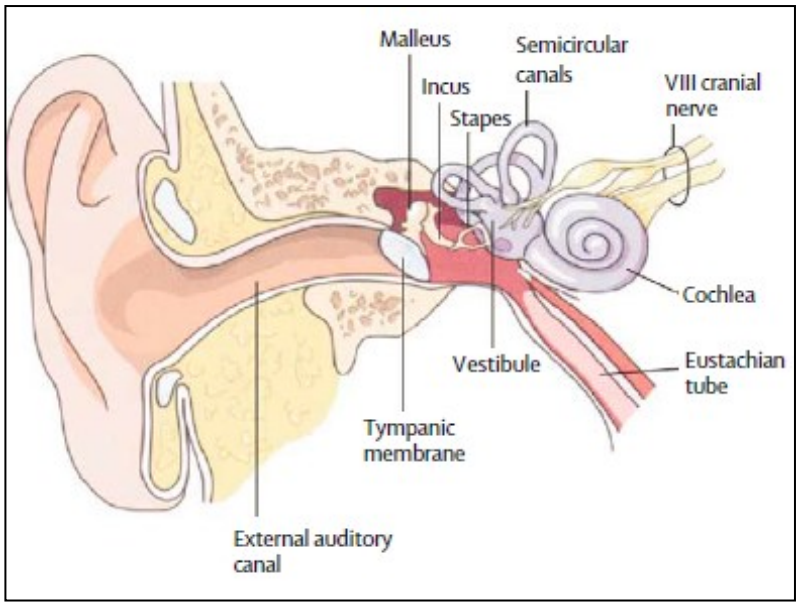


Figure 1.1: Cross section of the outer, middle and inner ear (adapted from Smith, Bale & White, 2005)**.** The outer ear is indicated by the external auditory canal and tympanic membrane. The middle ear is represented by the malleus, incus and stapes bones. The inner ear is represented by the cochlea vestibule and semicircular canals.

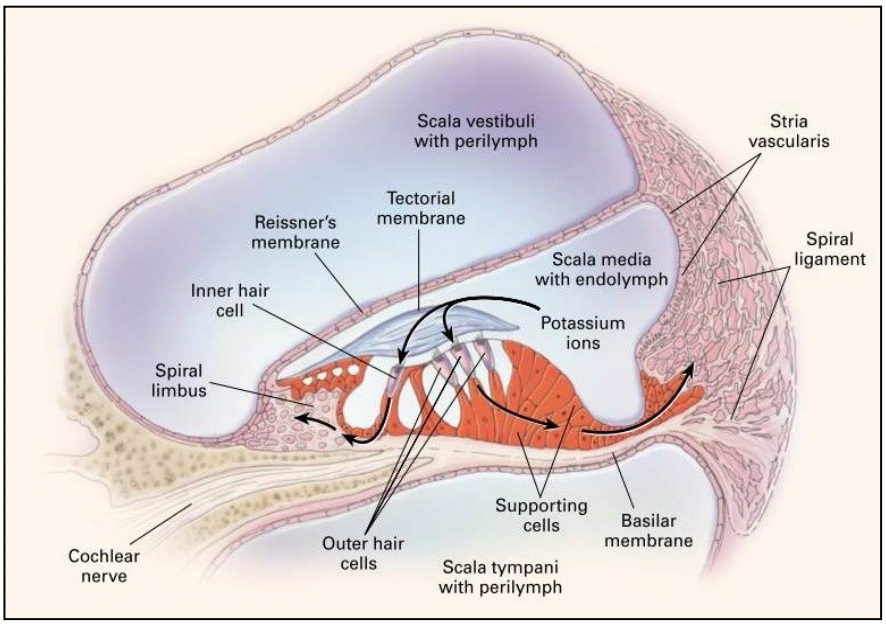


Figure 1.2: Cross section of the cochlea (adapted from Willems, 2000). The inner and outer hair cells are situated between the tectorial membrane and the basal membrane. Movement of the tectorial membrane results in the shearing motion that opens ion channels in the hair cells to depolarise the hair cells and release neurotransmitter to activate the acoustic nerve.

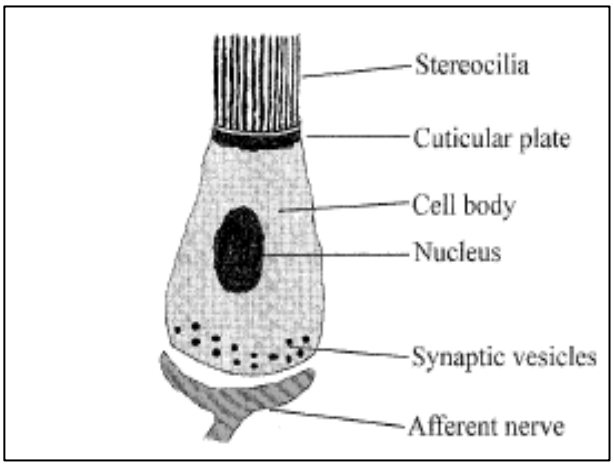


Figure 1.3: Schematic diagram of a hair cell (adapted from Ręd owicz, 1999). The diagram shows the sterocillia at the tip of the hair cell and the acoustic nerve at the base of the hair cell.

Following the depolarisation of the hair cells, potassium ions initially exit the hair cells through a channel and enter the supporting cells where they passively diffuse from cell to cell through

gap junctions (Tekin et al., 2001). The potassium ions will then be actively pumped back into the endolymph once they reach the stria vascularis and this allows the recycling of the potassium ions (Tekin et al., 2001).

Chapter 1.4: Aetiology of hearing impairment

Hearing impairment may be attributed to either genetic, environmental or unknown factors (Schrijver, 2004) and the aetiology varies from region to region, in a higher proportion of environmental and/or cases of unknown aetiology in Africa (Wonkam et al., 2013). Data from Europe and America indicates that genetic factors account for approximately 50% of congenital HI and environmental factors and HI of unknown cause each account for 25% of HI in developed countries (Schrijver, 2004).

In developing countries, the environment contributes more cases of HI, e.g. Cameroon, where putative environmental causes accounted for 52.7% of HI cases (Wonkam et al., 2013). The environmental causes of HI may be categorized as congenital causes of HI or acquired causes of HI (WHO Media Centre, 2017). The congenital causes may be due to maternal infections during pregnancy, low birth weight of the new-born, birth asphyxia of the new-born, maternal drug use during pregnancy and severe jaundice in the new-born (WHO Media Centre, 2017). The acquired causes of HI may be due to infectious disease; such as meningitis, mumps, measles and rubella (Wonkam et al., 2013); chronic ear infections, blockage of the ear canal, injury or head trauma, excessive noise and drug use or ototoxicity (WHO Media Centre, 2017). The aetiology of the environmental causes of HI is heterogeneous and similar to the genetic causes of HI in this sense.

Chapter 1.5: Genetics of hearing impairment

Genetic HI may be classified as non-syndromic if the HI occurs without other clinical manifestation, whereas syndromic HI occurs when more than one clinical manifestation is observed (Schrijver, 2004). HI is associated with approximately 152 genes (The Molecular Otolaryngology and Renal Research Laboratories. The University of Iowa, 2016) and over 1000 distinct pathogenic mutations in these genes (Shearer et al., 2011). Non-syndromic HI comprises 70% to 80% of genetic HI, of which 70% to 80% is a result of autosomal recessive inheritance (Chung et al., 1959; Morton, 1991). Autosomal dominant inheritance is responsible for 15% to 20%, X-linked inheritance is responsible for ~1% and mitochondrial inheritance is responsible for less than 1% of HI (Chung et al., 1959; Morton, 1991; Schrijver, 2004).

Chapter 1.5.1: Syndromic hearing impairment

Syndromic HI occurs when the HI is accompanied by one or more clinical features (Schrijver, 2004). This may include syndromes such as Usher Syndrome and Pendred Syndrome and these two syndromes will be discussed, in detail, below, as they may mimic non-syndromic HI, if not fully investigated.

Chapter 1.5.1.1: Usher syndrome

Usher syndrome was first reported in 1858 (Von Graefe, 1858), before been reported in 1914 by Charles Usher (Usher, 1914) who emphasised the hereditary nature of Usher Syndrome (Smith et al., 1994). Usher syndrome consists of 3 different subtypes, which have differing severity (Lévy et al., 1997). Usher syndrome type I consists of seven subtypes (Boëda et al., 2002) and it is associated with severe congenital HI, vestibular dysfunction and prepubertal onset of vision loss due to retinitis pigmentosa (Kimberling et al., 1992; Lévy et al., 1997). Usher syndrome type II is associated with a less severe form of HI as compared to Usher syndrome type I (Kimberling et al., 1992), normal vestibular function and vision loss due to retinitis pigmentosa (Weston et al., 1996). Usher syndrome type III is associated with progressive HI, as compared to congenital HI in Usher syndrome type I and type II, variable vestibular dysfunction, and vision loss due to retinitis pigmentosa (Weston et al., 1996).

Usher syndrome is the most prevalent form of combined deafness and blindness, accounting for 50% of such cases (Petit, 2001) with a prevalence of 1 in 6000 in an American population (Kimberling et al., 2010). There are 15 associated loci, of which 11 correspond to known genes and 4 correspond to unknown genes (Van Camp & Smith, accessed May 2016 and June 2017) and this is indicated in Table 2.

Table 2: Loci associated with Usher Syndrome showing known and unknown genes.

Locus	Location	Gene	Reference
USH1B	11q13.5	*MYO7A*	(Weil et al., 1995)
USH1C	11p15.1	*USH1C*	(Bitner-Glindzicz et al., 2000; Smith, 1992; Verpy et al., 2000)
USH1D	10q22.1	*CDH23*	(Bork et al., 2001; Wayne et al., 1996)
USH1E	21q21	Unknown	(Chaïb et al., 1997)
USHF	10q21-22	*PCDH15*	(Ahmed et al., 2001; Alagramam et al., 2001)
USH1G	17q24-25	*SANS*	(Mustapha et al., 2002; Weil et al., 2003)
USH1H	15q22-23	Unknown	(Ahmed et al., 2009)
USHIJ	15q23-q25.1	*CIB2*	(Riazuddin et al., 2012)
USHIK	10p11.21-q21.1	Unknown	(Jaworek et al., 2013)
USH2A	1q41	*USH2A*	(Eudy et al., 1998; Kimberling et al., 1990)
USH2B	3p23-24.2	Unknown	(Hmani et al., 1999)
USH2C	5q14.3-q21.3	*VLGR1*	(Pieke-Dahl et al., 2000; Weston et al., 2004)
USH2D	3q32	*WHRN*	(Ebermann et al., 2007)
USH3	3q21-q25	*CLRN*	(Ebermann et al., 2010)
	10q24.31	*PDZD7*	(Joensuu et al., 2001; Sankila et al., 1995)

(Adapted from http://hereditaryhearingloss.org, 2017)

Chapter 1.5.1.2: Pendred syndrome

Pendred syndrome is a syndromic HI first described in two sisters in 1896 (Pendred, 1896; Reardon et al., 1997) and the autosomal recessive inheritance pattern of Pendred syndrome was indicated in 1927 (Brain, 1927; Reardon et al., 1997). Pendred syndrome presents as enlarged vestibular aqueduct presenting with congenital, progressive HI which is associated with thyroid disease and goitre (Everett et al., 1997; Reardon et al., 1997; Sugiura et al., 2005; Yang et al., 2009).

The clinical features of Pendred syndrome, however, vary between individuals and within families (Sugiura et al., 2005). The variations may include the age of onset for the HI and goitre, as well as the presence or absence of vestibular dysfunction and fluctuations in the HI (Cremers et al., 1998). Enlargement of the aqueduct may also vary whereby some patients do not present with an enlarged aqueduct (Reardon et al., 1997).

Pendred syndrome may be the most common form of syndromic HI where it may account for up to 10% of syndromic HI (Everett et al., 1997; Fraser et al., 1960; Reardon et al., 1997). Mutations in the *SLC26A4* gene were initially identified as the causative mutation in Pendred syndrome (Everett et al., 1997). It is however now known that both digenic inheritance of mutations in *SLC26A4* and *KCNJ10* may also result in Pendred and double heterozygosity in *SLC26A4* and *FOXI1* (Yang et al., 2009, 2007).

Chapter 1.5.1.3: Other hearing impairment syndromes

Other forms of syndromic HI include Waardenburg Syndrome (Waardenburg, 1951), Jervell and Lange-Nielsen Syndrome (Jervell and Lange-Nielsen, 1957)and Branchio-oto-renal (Asherson, 1834; Přecechtěl, 1927) amongst others. The syndromes vary according to phenotype, with HI as a common feature.

Waardenburg Syndrome affects, in conjunction with the HI, the positioning of the eyes and pigmentation in the hair, skin and eyes and, periodically, resulting in abnormalities of the limbs (Hoth et al., 1993; Waardenburg, 1951) and being attributed to six known genes (*PAX3, MITG, SNAI2, EDNRB. EDN3 and SOX10)* (Attié et al., 1995; Edery et al., 1996; Hoth et al., 1993; Pingault et al., 1998; Sánchez-Martín et al., 2002; Tassabehji et al., 1994, 1992) and two unknown genes (Lalwani et al., 1994; Selicorni et al., 2002). Waardenburg syndrome type II is the most common form of syndromic HI in African populations (Wonkam et al., 2013).

Jervell and Lange-Nielsen associated HI with prolongation of the QT phase of the electrocardiogram, increased fainting episodes due to ventricular arrhythmias and an increased risk of sudden death (Jervell and Lange-Nielsen, 1957; Neyroud et al., 1997) and the syndrome

is due to mutations in the genes, *KCNQ1* (Neyroud et al., 1997)and *KCNE1* (Schulze-Bahr et al., 1997; Tyson et al., 1997).

Branchio-oto-renal syndrome is combined HI with branchial cleft cysts and renal abnormalities (Hoskins et al., 2007; Kumar et al., 2000). It has been associated with three known genes (*EYA1, SIX5* and *SIX1*) (Abdelhak et al., 1997; Ruf et al., 2004, 2003) and one unknown gene (Kumar et al., 2000). In conjunction with Usher Syndrome and Pendred Syndrome, these three syndromes indicate the phenotypic and genotypic heterogeneity of HI.

Chapter 1.6: Prevalent genes associated with Non- Syndromic hearing impairment

The connexin genes *GJB2* and *GJB6* account for the highest proportion of genetic HI in European, Asian and North American populations (Chan and Chang, 2014; Hutchin et al., 2005; Liu et al., 2002; Najmabadi and Kahrizi, 2014; Pandya et al., 2003). Mutations in *SLC26A4* have also been indicated to have a relatively high proportion amongst the Caucasian British population where mutations within this gene account for approximately 3.5% of non-syndromic HI cases (Hutchin et al., 2005). The prevalence of mutations in the connexin genes will be discussed further below with emphasis placed on their significance in African populations.

Chapter 1.6.1: Non-syndromic hearing impairment: prevalence of GJB2 and GJB6

Mutations in the connexin genes *GJB2* and *GJB6* are the most frequent cause of non-syndromic HI (del Castillo et al., 2002; Gasparini et al., 2000) with mutations in *GJB2* present in 50% of cases of non-syndromic HI in Europe, North America and Asia (Chan and Chang, 2014; Liu et al., 2002; Najmabadi and Kahrizi, 2014; Pandya et al., 2003). The distribution of the mutations in *GJB2* is indicated in Figure 1.4. The mutations in *GJB6* are present in approximately 20% of HI cases in Europe and North America (del Castillo et al., 2002; Schrijver, 2004).

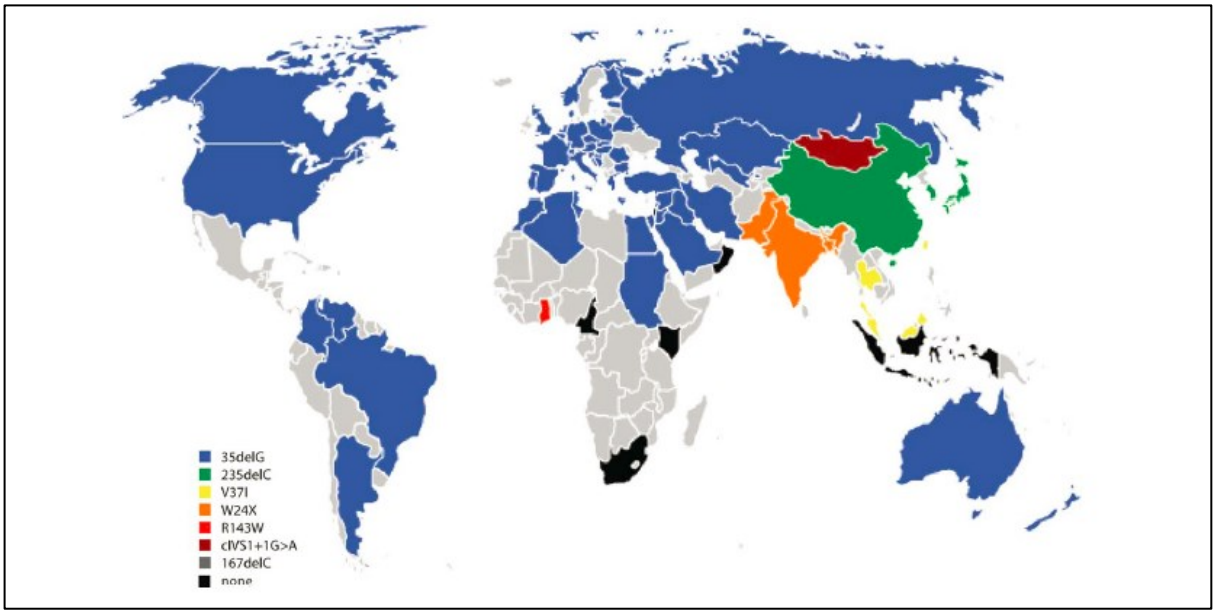


Figure 1.4: Distribution of *GJB2* mutations across the world (Adapted from Chan and Chang, 2014)The different colours indicate different *GJB2* mutations, however areas in black have no *GJB2* mutations and areas in grey have yet to be studied.

Chapter 1.6.2: *GJB2* and *GJB6* mutations are rare among Africans

Several studies examining the prevalence of mutations in the connexin genes *GJB2* and *GJB6*, in African populations, have taken place with varying sample sizes. The *GJB2* and *GJB6* mutations are, however, not prevalent in the Sub-Saharan African populations (Bosch et al., 2014a; Kabahuma et al., 2011; Wonkam et al., 2015) and this is summarised in Table 3.

Table 3 lists nine studies that investigated the prevalence of *GJB2* and/or *GJB6* mutations within Africa populations. Four studies found that *GJB6* mutations had no prevalence within the African context whereas only two studies found the relevance of specific *GJB2* mutations in the Ghanaian and Moroccan populations. The other 6 studies found little significance of *GJB2* mutations within their studied populations. The insignificance of *GJB2* and *GJB6* mutations highlights the necessity of further investigation in African HI in order to elucidate the prevalent genes among African populations and this may be achieved through next generation sequencing tools.

Table 3: Frequency and type of pathogenic *GJB2* and *GJB6* mutations in African populations

Population	Sample Size	*GJB2* mutations (allele frequency)	*GJB6* mutations (frequency)	Study
South African	182	Exon 2 (0%)	D13S1830 (0%)	(Kabahuma et al., 2011)
Kenyan	406	Whole gene (0%)	-	(Gasmelseed et al., 2004)
Sudanese	183	Whole gene (2.73%)	-	(Gasmelseed et al., 2004)
Cameroonian	180	c.424_426delTTC (0.28%) c.499G >A (0.28%)	-	(Bosch et al., 2014b)
Cameroonian	100	-	D13S1830 (0%)	(Bosch et al., 2014a)
Northern Cameroonian	78	c.269insT (0.64%)	-	(Trotta et al., 2011)
Ugandan	115	c.208C>G (0.43%)	Exon 3(0%) D13S1830 (0%)	(Javidnia et al., 2014)
Ghanaian	365	c.427C>T (15.07%)) c.608_609TC>AA (0.14%) c.236T>C (0.14%) c.641T>C (0.14%) c.35insG (0.14%) c.533T>C (0.55%) c.551G>A (0.14%) c.589G>T (0.14%)	-	(Hamelmann et al., 2001))
Moroccan	116	g.35delG (21.55%)	342-kb deletion (0%)	(Gazzaz et al., 2005)

Chapter 1.7: Next generation sequencing and hearing impairment

Next generation sequencing is a technique which allows low-cost simultaneous sequencing of multiple genes or regions of an organism's genome (Shearer et al., 2011; Tekin et al., 2017). Whole genome sequencing is the sequencing of the entire genome of an organism (Shearer et

al., 2010) which results in $3.2*10^{10}$ bps for analysis when working with humans (Gao and Dai, 2014). Whole genome sequencing has recently being used to identify causative mutations in 30, unrelated, consanguineous Pakistani families (Shafique et al., 2014).

The entire exome of the individual's genome, which accounts for 1% of the genome of an organism (Teer and Mullikin, 2010), may be screened instead of the entire genome. This is because approximately 85% of pathogenic mutations are within the protein coding regions of the DNA and/or the splice sites of the genes (Shearer et al., 2011) and it provides a cost-effective alternative to whole genome sequencing with regard to monogenic disease. Whole exome sequencing, in conjunction with homozygosity mapping, has been used to elucidate the causative mutation which resulted in HI in a consanguineous Pakistani family (Walsh et al., 2010). The mutation in the *GPSM2* was identified by the authors, when focussing on exonic data from the *DFNB82*. The notation *DFNB82* indicates that the locus is inherited in an autosomal recessive manner and that this is the 82nd autosomal recessive HI associated locus identified. This region spans 3.1Mb of chromosome 1p13.1 and it partially overlaps with *DFNB32* (Walsh et al., 2010). This led the authors to identify a mutation in the gene that results in a nonsense mutation at p.R127X (Walsh et al., 2010). Whole exome sequencing will, however, result in data from all the genes of the individual and this may not necessarily be a cost effective diagnostics approach, if the genes associated with the phenotype are already known, as the entire exome is targeted for sequencing and analysis (Shearer et al., 2011, 2010). Targeted gene enrichment analysis allows for the identification of mutations in specific genes (Shearer et al., 2011; Tekin et al., 2017) which have been shown to be statistically relevant to the biological question. This is done by designing probes that are complementary to the genes of interest, which is more cost effective and time effective, when compared to whole genome or whole exome sequencing as only the relevant genes are sequenced (Shearer et al., 2011, 2010), provided the genes associated with the phenotype are known. OtoSeq® and OtoSCOPE®, MiamiOtoGenes are three tools which use targeted gene enrichment to capture genes associated with HI for sequencing (Shearer et al., 2011; Sivakumaran et al., 2013; Tekin et al., 2017).

OtoSeq® is a targeted exome tool developed by the Cincinnati Children's Hospital Medical Center (Sivakumaran et al., 2013). The tool scans 24 sensorineural HI genes, and an additional 68 genes associated with genetic disorders that are not related to sensorineural HI (Sivakumaran et al., 2013). OtoSeq® was used in 2013 to elucidate the causative mutations in 34 consanguineous Pakistani HI families (Shahzad et al., 2013). 24 mutations in 28 families were identified, including 11 novel mutations (Shahzad et al., 2013).

OtoSCOPE® or otologic sequence capture of pathogenic exons uses a target gene approach that use solution based target enrichment to capture targeted exons (Shearer et al., 2011, 2010). OtoSCOPE® forms a diagnostic panel for non-syndromic HI that targets exons of all the 57 genes associated with HI for direct sequencing (Shearer et al., 2010). The panel currently screens for mutations in 152 genes and microRNAs associated with, both non-syndromic and syndromic, HI (https://morl.lab.uiowa.edu/deafness, January 2017). OtoSCOPE® was used in conjunction with whole exome sequencing to determine the causative mutation in a multigenerational family of European descent (Azaiez et al., 2014). The OtoSCOPE® panel was initially used to screen for mutations in an affected individual in the family but the panel did not return any variations that could be associated with the HI phenotype in the family and the authors then moved to whole exome sequencing for 4 affected and 4 unaffected family members (Azaiez et al., 2014). A novel c.533C>T mutation was discovered in the *TBC1D24* gene which encodes a protein that may be involved in oxidative stress resistance and catalytic activity (Azaiez et al., 2014).

MiamiOtoGenes uses solution based custom capture to target 146 genes, their 5' and 3' untranslated regions, and 25bp intronic flanking regions, for targeted genome enrichment, with the targeted regions sequenced on the Illumina HiSeq2000 platform (Tekin et al., 2017). The genes targeted for sequencing were selected from candidate genes discovered by the researchers during their research and genes available on various databases, including Hereditary Hearing Loss Homepage (http://hereditaryhearingloss.org), Ensembl (http://www.ensembl.org), and OMIM (http://www.ncbi.nlm.nih.gov/omim/) (Tekin et al., 2017). Annotation of the variants proceeded through the use of SeattlSeq137 (Ng et al., 2010), Online Mendelian Inheritance in Man database (OMIM) and Human Gene Mutation Database (HGMD) (Stenson et al., 2009) and the platform resolved 10 of the 14 patients (71%) (Tekin et al., 2017). The platform was further used to resolve HI in a multi-ethnic cohort of 342 hearing impaired probands from South Africa, Nigeria, Tunisia, Turkey, Iran, India, Guatemala and the United States of America, with a resolution rate of 15% (53/342) (Yan et al., 2016). The platform identified 57 unique variants, accounting for compound heterozygosity in 6 individuals; including 31 novel variants (Yan et al., 2016). The resolution rate was highest in the Iranian and Turkish population (57% and 42% respectively) and lowest in the Guatemalan, South African and Nigerian populations (0%, 4% and 4% respectively) (Yan et al., 2016).

Chapter 1.7.1: Resolution of hearing impairment in African patients using targeted exome sequencing

In order to resolve ARNSHL amongst patients from SSA, OtoSCOPE® was used, recently, to screen African HI patients from 10 Cameroonian families, who were negative for mutations in *GJB2, GJB6* and *GJA1* using direct cycle sequencing (Bosch et al., 2014a, 2014b). This resulted in 12 mutations in 7 families (Lebeko et al., 2016). One family presented with insufficient DNA quantities for OtoSCOPE® screening and two families presented with no mutations in any of the 57 HI associated genes present on the OtoSCOPE® panel.

The failure of the panel to resolve the cause of HI within all the probands is similar to other studies in literature that have had varying success in using Targeted Genome Enrichment to resolve HI within their cohorts (Azaiez et al., 2014; Brownstein et al., 2011; Choi et al., 2013; Moteki et al., 2016) and this is summarised in Table 4.

Table 4: Resolution rate of hearing impairment using targeted genome enrichment.

Population	Sample size n	Resolution Rate n (%)	Gene	Study
Middle Eastern	11	6 (54.55%)	*TMC1* *CDH23* *MYO15A* *WFS1* *TECTA*	(Brownstein et al., 2011)
Asian	20	13 (65%)	*WFS1* *COCH* *EYA4* *MYO6* *GJB3* *COL11A2* *OTOR*	(Choi et al., 2013)
Asian	194	52 (26.80%)	*KCNQ4* *WFS1* *MYO6* *TECTA* *MYO5A*	(Moteki et al., 2016)

			CDH23 *SLC26A4* *STRC*	
Caucasian	1	0 (0%)	N/A	(Azaiez et al., 2014)

Brownstein et al. (2011) improved the efficiency of TGE by increasing the number of genes screened against. They screened against 246 genes which included 80 human HI associated genes and 162 genes associated with HI in the mouse (Brownstein et al., 2011). Choi et al. (2013) combined TGE and Sanger Sequencing within their 32 patient cohort. They initially screened the cohort for mutations in *SLC26A4*, *POU3F4*, mtDNA and *GJB2*. This was followed by TGE, against 80 HI associated genes, of 19 probands, negative for mutations in the above mentioned genes and mtDNA, as well as 1 proband who was positive for *GJB2* causative mutations (Choi et al., 2013).

The use of the TGE excluded the possibility of discovering mutations in genes that have not been previously associated with HI. For this reason, samples from one of the two families, who returned no causative mutations following OtoSCOPE® screening (Lebeko et al., 2016), were subjected to whole exome sequencing. Whole exome sequencing has been shown to have a higher resolution rate, as indicated in Table 5, and it has the potential to discover causative mutations in genes that have not been previously associated with HI.

Table 5: Resolution of Hearing Impairment using whole exome sequencing

Population	Sample size n	Resolution rate n (%)	Genes	Study
Caucasian	7	7 (100%)	*RAI1* *OTOF* *SLC26A4*	(Churbanov et al., 2016)
Middle Eastern	30	17 (56.67%)	*TRIOBP* *TMC1* *LOXHD1* *TMPRSS3* *MYO15A* *GIPC3* *ILDR1* *MYO7A* *TECTA*	(Diaz-Horta et al., 2012)
Asian	2	2 (100%)	*TMC1*	(Gao et al., 2013)
Caucasian	3	3 (100%)	*TBC1D24*	(Azaiez et al., 2014)
South Asian	1	1 (100%)	*GRXCR2*	(Imtiaz et al., 2014)

The use of WES was capable of detecting a novel variant in the *GRXCR2* gene that had only been associated with HI in 3 affected individuals from a Pakistani family (Imtiaz et al., 2014). This result showed the clinical utility in detecting causative mutations in genes that had not been previously associated with HI. This result combined with the insignificance of the prevalent and known HI genes within our population makes WES a favourable approach to resolving HI within the African population.

Chapter 1.8: Rationale of the study

1. The most common genes in non-syndromic HI (*GJB2* and *GJB6*) among Europeans and Asians affected by HI are uncommon among people of African ancestry

2. The use of targeted exome sequencing capturing known HI genes has a low pick up rate among African patients. For example, following screening using the targeted exome sequencing platform OtoSCOPE®, in 2/9 families the approach did not uncover pathogenic mutations, which potentially indicates that the potential of discovering novel genes associated with HI among people of African descent.
3. Whole exome sequencing that explores the sequencing results for all coding genes which may allow the identification of novel HI associated genes in African population samples.
4. The use of WES data from control and HI patient populations, where there are no known prevalent HI genes, could establish the minor allele frequencies of pathogenic variants within African populations and therefore the list of genes that are more relevant in clinical practice to these populations.

Chapter 1.10: Thesis outline

Chapter 1:

In chapter 1, we review extant literature regarding HI and the study of genetic HI in African populations. The chapter includes consideration of the prevalent genes associated with HI and their impact on African populations. The chapter also considers the utility of next generation sequencing and its impact on the resolution of the genetic causes of HI in populations that have utilised next generation sequencing.

Chapter 2:

Chapter 2 communicates the methodology used in the study. The chapter indicates the study population recruitment and inclusion. It also indicates the bioinformatics pipeline that was followed in order to analyse the whole exome sequencing data. The chapter concludes with the validation techniques utilised to validate the whole exome sequencing data.

Chapter 3:

The results from the various methodologies are indicated in this chapter. The chapter addresses the demographics of the study population as well as the population structure. The chapter indicates the results from filtering the variations determined by the whole exome sequencing and bioinformatics analysis. It further goes on to indicate the downstream analysis of the interaction of novel genes with known HI associated genes and maps their protein products in

the greater HI protein network. The chapter further the proportion of ancestral alleles in known genes among the hearing impaired and the control population as well as in the Cameroonian and the Democratic Republic of Congo population. The chapter concludes by indicating the outcome of validating the whole exome sequencing results through Sanger Sequencing.

Chapter 4

The results obtained are discussed in this chapter. The relevance of the results obtained, in terms of the novel variants discovered and the disparity between the filtering outcome and the analysis from Sanger Sequencing is taken into consideration. Practical implications of the protein-protein interactions, the protein network and the proportion of ancestral and derived alleles is considered and the limitations of the study and future perspective and indicated.

Chapter 5:

The conclusions drawn from the study are stated in this chapter. These conclusions are drawn from the key points in the discussion and build on the perspectives and limitations of the study.

Chapter 1.11: Aim and objectives

The aim of this project is to determine the minor allele frequency and/or the carrier rates for variations in 172 HI associated genes in Cameroonian HI patients and controls; and coincidently resolve the cause of HI within the hearing-impaired patients. The project also aims to compare the minor allele frequencies and carrier rates between and to determine the proportion of ancestral and derived alleles in the patients and controls

Objective 1

The first objective is the bioinformatics analysis of whole exome sequencing data from the patient and control populations in order to discover variations in known HI genes. This analysis will be compared to other African populations available in the 1000 Genome Project.

Objective 2

The second objective is to determine the genomic architecture of HI and the evolution of hearing, in the population.

Objective 3

The third objective is to determine the effect of the candidate variants on protein structure and investigate the potential that the variants are founder effects.

Chapter 2: Patients and methods

Chapter 2.1: Ethical consideration

Ethical approval for this study was granted by the University of Cape Town's Human Research Ethics Committee (HREC 621/2017) and the Cameroon National Ethics Committee of the Ministry of Public Health (N°123/CNE/SE/2010). Informed written and signed consent was provided by study participants 18 years old or older. Informed verbal assent was provided by minors (younger than 18 years old) and informed written and signed consent was provided by the parents or guardians of the minors.

Chapter 2.2: Patient recruitment and inclusion

Chapter 2.2.1: Patient recruitment

The patient population was recruited from various Schools of the Deaf and ENT outpatient clinics in Cameroon as reported in Wonkam et al. (2013). The patients were recruited from seven of the ten regions of Cameroon, with the intention to that this would provide a diverse study population and had to have HI that presented itself before the age of 15 years (Wonkam et al., 2013).

The patients were examined by a medical geneticist and an ophthalmologist (Lebeko et al., 2017), to determine the nature of the HI, and the detailed family and medical history was obtained from the patients, their parents and their medical records (Wonkam et al., 2013). A total of 582 patients were recruited into the original study (Wonkam et al., 2013), but only 18 patients with non-syndromic HI, presenting with putative genetic cause, following an autosomal recessive inheritance pattern were included in this present study.

Chapter 2.2.2: Inclusion and exclusion criteria

Inclusion criteria:

1. Patients of black-African ancestry presenting with putative non-syndromic HI following an autosomal recessive inheritance pattern were included in this study. The HI needed to be of putative genetic origin or of unknown aetiology. In addition, all the patients included didn't have any causative mutations in *GJB2*, *GJB6* and *GJA1,* as exclude with Sanger sequencing of the entire coding regions.

Eighteen (18) patients, who met the above criteria, were chosen at random for inclusion in this study.

Exclusion criteria:

1. Patients presenting with putative syndromic HI or HI due to environmental factors
2. Presented with causative mutations in *GJB2*, *GJB6* and *GJA1*

Chapter 2.2.3: Control population

The control population consisted of 130 ethnically matched individuals who do not have HI and do not have a family history of HI. The control population was recruited from Yaoundé Central Hospital and Laquintinie Hospital in Douala, both hospitals are in Cameroon.

Chapter 2.3: DNA extractions and quality checks

Patient DNA was, previously, extracted from whole blood using the modified version of the salting out procedure (Bosch, 2013; Miller et al., 1988) and control DNA was extracted using the Puregene Blood Kit (Qiagen, Alameda, CA, USA), in accordance to the manufacturer's instructions (Geard, 2016); both methods are available in the appendix. The concentration of the extracted DNA and the quality of the extraction was determined using spectrophotometry, which was done using the Nanodrop ND-1000 (Thermo Fischer Scientific, Waltham, USA) where TE buffer was used as the blank. The spectrophotometric readings provided the 260/280 and 260/230 readings, which were used to determine if the presence of protein and RNA contamination, respectively. The spectrophotometric readings also provided the concentration of the DNA which would be used to make working DNA dilutions.

The integrity of the DNA was assessed using a 2% agarose gel and visualised using Sybr Safe (Invitrogen by Thermo Fisher Scientific, California, USA) under UV light produced by the UV Doc (UVItec Ltd). This gave an indication of the integrity of the DNA by indicating whether the DNA was fragmented, and the degree of fragmentation, or if it was intact, which has consequences for downstream analysis as fragmented DNA may hamper the polymerase chain reaction.

Chapter 2.4: Whole exome sequencing

The patient and control population were subjected to whole exome sequencing. This is supported by previous data that evidenced the low pick up rates of mutation in known genes in populations of African ancestry with HI (Bosch et al., 2014a, 2014b; Gasmelseed et al., 2004; Gazzaz et al., 2005; Javidnia et al., 2014; Kabahuma et al., 2011; Lebeko et al., 2017, 2016; Trotta et al., 2011; Wonkam et al., 2015) and the large genetic heterogeneity of HI,

whereby over 1000 distinct mutations (Shearer et al., 2011; Sloan-Heggen et al., 2016; Tekin et al., 2017) in over 100 genes have been associated with HI (The Molecular Otolaryngology and Renal Research Laboratories. The University of Iowa, 2016). The use of whole exome sequencing may allow for the detection of novel mutations in genes that have not been previously associated with HI before.

Whole exome sequencing of Patient and Control DNA was done by Omega Bioservices (Norcross, GA, USA). The DNA concentration was first quantified using the QuantiFlour dsDNA System on a Quantas Flourometer (Promega, Madison, WI, USA). The DNA library was prepared, using the Illumina Nextera Rapid Capture Exome Kit (Illumina, San Diego, CA, USA), by tagmenting 50ng of genomic DNA using the Nextera transposomes. The libraries were then hybridized with a 37M probe pool to enrich the sequences and were subjected to whole exome sequencing using the Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) using a 100bp run format, with an average read depth of 30X. 18 of the 19 patients were successfully sequenced.

Chapter 2.5: Bioinformatics analysis

Chapter 2.5.1: Reads alignment and post-alignment quality checks

Following sequencing, the quality of the whole exome sequencing needs to be verified prior to any downstream analysis. The quality control measures give an indication of how trustworthy each variant call may. The whole exome sequencing data was subjected to quality control measures using FastQC (Van Der Auwera et al., 2014) and SolexaQC++ (Cox et al., 2010).

The sequencing read was aligned to a reference sequence, following quality control measures, to determine the genetic sequences of each individual. The reads were aligned to hg19, build37, downloaded from the UCSC database (Rosenbloom et al., 2015) using BWA (Li et al., 2008; Li and Durbin, 2009) with local realignment performed using the Genome Analysis Toolkit (GATK) software (McKenna et al., 2010).

Chapter 2.5.2: Variant discovery

The ensemble variant approach was used for variant calling approaches from the GATK3.0 (haplotypeCaller) and SamTools (Garrison and Marth, 2012) variant callers. The final variant call set was obtained by combining the call sets from the two mentioned variant callers using VariantMetaCaller (Gézsi et al., 2015), and this indicated which variants in the exomes differed from the reference.

Chapter 2.5.3: Variant annotation and mutation prioritization

Variant annotation was performed on the resulting VCF files that consisted of the individual patient VCF files, a combined file of all 18 patients and a combined file of all 105 controls and, lastly, a combined VCF file consisting of all the patients and controls. The variants were annotated in order to provide information on the presence of single nucleotide polymorphisms (SNPs) and if the SNPs resulted in amino acid changes and, if so, the effects of the amino acid change on the protein.

ANNOVAR (Wang et al., 2010) was then used to independently perform gene based annotation for each of the above mentioned files in order to catalogue whether the SNPs caused protein coding changes and to identify the affected amino acids. ANNOVAR was used on the "2015Dec18", where the population frequency for each variant was obtained from the 1000 Genomes Project exome and targeted exon datasets (The 1000 Genomes Project Consortium, 2015) and COSMIC (Forbes et al., 2015). Genetic function information was obtained from RefGene (O'Leary et al., 2016; Pruitt and Maglott, 2001) and different functional predictions were obtained from ANNOVAR's ljb_all. The variants were then filtered for exonic variants that were considered deleterious or damaging according to 10 different functional scores; including SIFT (Sorting Intolerant From Tolerant) (Ng and Henikoff, 2006, 2003, 2002), LRT (likelihood ratio test) (Fujita et al., 2011), MutationTaster (Lubeck et al., 2012), MutationAssessor (Reva et al., 2011, 2007), FATHMM (Functional Analysis through Hidden Markov Models) (Shihab et al., 2014), RadialSVM (radial Support Vector Machine)and LR (Logistic Regression) (Dong et al., 2015), CADD (Combined Annotation Dependent Depletion) (Kircher, 2014), GERP (Cooper et al., 2005; Davydov et al., 2010), PolyPhen2 (Adzhubei et al., 2010), PhyloP (phylogenetic P-values) (Pollard et al., 2010) and SiPhy (SIte-specific PHYlogenetic analysis) (Garber et al., 2009). We additionally included conserved and segmental duplication sites, dbSNP code (Sherry et al., 2001) and clinical relevance reported in dbSNP 138 (National Center for Biotechnology Information, n.d.; Sherry et al., 2001).

Once variant annotation was completed, possible disease-causing variants were isolated with the data files and this was done through filtering/mutation prioritization. The annotated datasets were filtered for rare variants, exonic variants, non-synonymous variants, stop codons, and variants with predicted functional significance, natural selection on codon, and deleteriousness (Li et al., 2013, 2012). We retained variants that were predicted as "Deleterious" (D), "Probably Damaging" (D), "Disease Causing Automatic" (A) or "Disease Causing" (D), according to 10 functional score approaches.

Following annotation and variant prioritization, downstream analysis was performed on the combined patient VCF file and the combined control file. This allowed for the detection of genetic differentiation and population structure.

Chapter 2.5.4: Genetic differentiation and population structure

The population structure and genetic difference for the hearing-impaired and the control population was analysed. Population structure analysis was done to indicate whether the control and hearing-impaired population originated from the same population. The presence or absence has implications for the downstream bioinformatics analysis on either account population stratification in the case of the presence substantial sub-structure in the data. Analysis of the genetic difference was done to determine if there was an inherent difference in particular genomic region, which showed unusual differences between the control and hearing-impaired group. Population structure analysis was performed based on principle component analysis (PCA) using smartpca (Patterson et al., 2006; Price et al., 2006). The data included 186 samples from Yoruba (YRI) in Nigeria, 173 from Esan (ESN) in Nigeria, 280 from Gambia (GWD) in Western Divisions in the Gambia, 116 Luhya (LWK) in Webuye, Kenya, 112 from African Ancestry in Southwest US (ASW), 128 from Mende (MSL) in Sierra Leone and 123 from African Caribbean in Barbados (ACB) from the 1000 genomes Phase3 (The 1000 Genomes Project Consortium, 2015). The data also included 6 samples from the Democratic Republic of Congo (DRC) (unpublished data). The PCA was initially performed on the patient and control Cameroonian cohort. It was then performed on the merged data set of the eight African populations and the control and patient Cameroonian cohort. The first two components, in both PCAs, were plotted against each other using Genesis2.

The genetic difference between the HI patient and control groups was investigated by performing a statistic test of difference to detect unusual genetic difference between the HI patient group and the control group from the aggregate SNP frequency at gene level (Chimusa et al., 2015). SNP frequency was assigned to a given gene, if the SNP was located within the gene or within the gene's downstream or upstream region using dbSNP138 database. The SNP allele frequencies were aggregated using Fisher's Combined Probability (Fisher, 1958).

Chapter 2.5.5: Pathway enrichment analysis and protein-protein interactions

A comprehensive human Protein-Protein Interaction (PPI) network (Chimusa et al., 2015; Wu et al., 2009) was used to analyse how each of the variant genes are layered in a biological network to allow us to extract a sub-network. The association between the variant genes in the

subnetwork with human phenotypes and their potential biological pathways, processes, and molecular functions, were examined. This was done using custom scripts in python and R (R Core Team, 2016) and the enrichment analysis was performed using Enrichr (Chen et al., 2013; Kuleshov et al., 2016) and Panther (Mi et al., 2013; Thomas et al., 2003), based on the identified variant genes.

The list of candidate genes was combined with the list of 159 known HI associated genes obtained from Deafness Variation Database (The Molecular Otolaryngology and Renal Research Laboratories. The University of Iowa, 2016) and ClinVar (Landrum et al., 2016), 163 genes in total, to perform a community network analysis. A clustering script in R's igraph package (Csárdi and Nepusz, 2006) was used to produce a network plot that would allow for the identification of hub proteins in the sub-networks of HI genes.

Chapter 2.6: Evolutionary adaptation of human hearing

SNP ancestral alleles were downloaded from the Ensembl, a 59 comparative 32 species alignment (Paten et al., 2008), and further determined the presence of the SNPs in the dbSNP database. Our final patient and control VCF files were annotated using the VCFtools 'fillOaa' script (Danecek et al., 2011), with the ancestral allele recorded using the 'AA' INFO tag ("The 1000 Genome Project Consortium Ancestral Allele Reference Sequence," 2012).

The ancestral allele for each SNP was computed (independently in the patient and control datasets), by dividing the number of times the alternative allele matched the ancestral allele with the number of copies of all alternative alleles across all samples at the particular SNP. The fraction of the derived allele is equivalent to 1 minus the fraction of the ancestral allele.

It has been previously shown that the derived alleles are more often minor alleles (allele frequency less than 50%) and are more often associated with risk than ancestral alleles (Gorlova et al., 2012). We, therefore, investigated the relationship between the fraction of the derived allele and the minor allele frequency of each phenotype. To this end, the alternative alleles were categorised in 6 bins (0-0.05, >0.05-0.1, >0.1-0.2, >0.2-0.3, >0.3-0.4, >0.4-0.5) with respect to the patient and control group frequencies and the fractions of the derived alleles were independently computed for each bin. Furthermore, the fraction of ancestral/derived alleles for all known and candidate HI genes was computed. To this end, the fraction of ancestral/derived alleles were computed at the SNP-based level to the gene, considering all the SNPs located within the gene and downstream and upstream region (Chimusa et al., 2015) and

the Bioinformatics pipeline, including analysis of the fraction of derived and ancestral alleles is indicated in Figure 2.6.1 below.

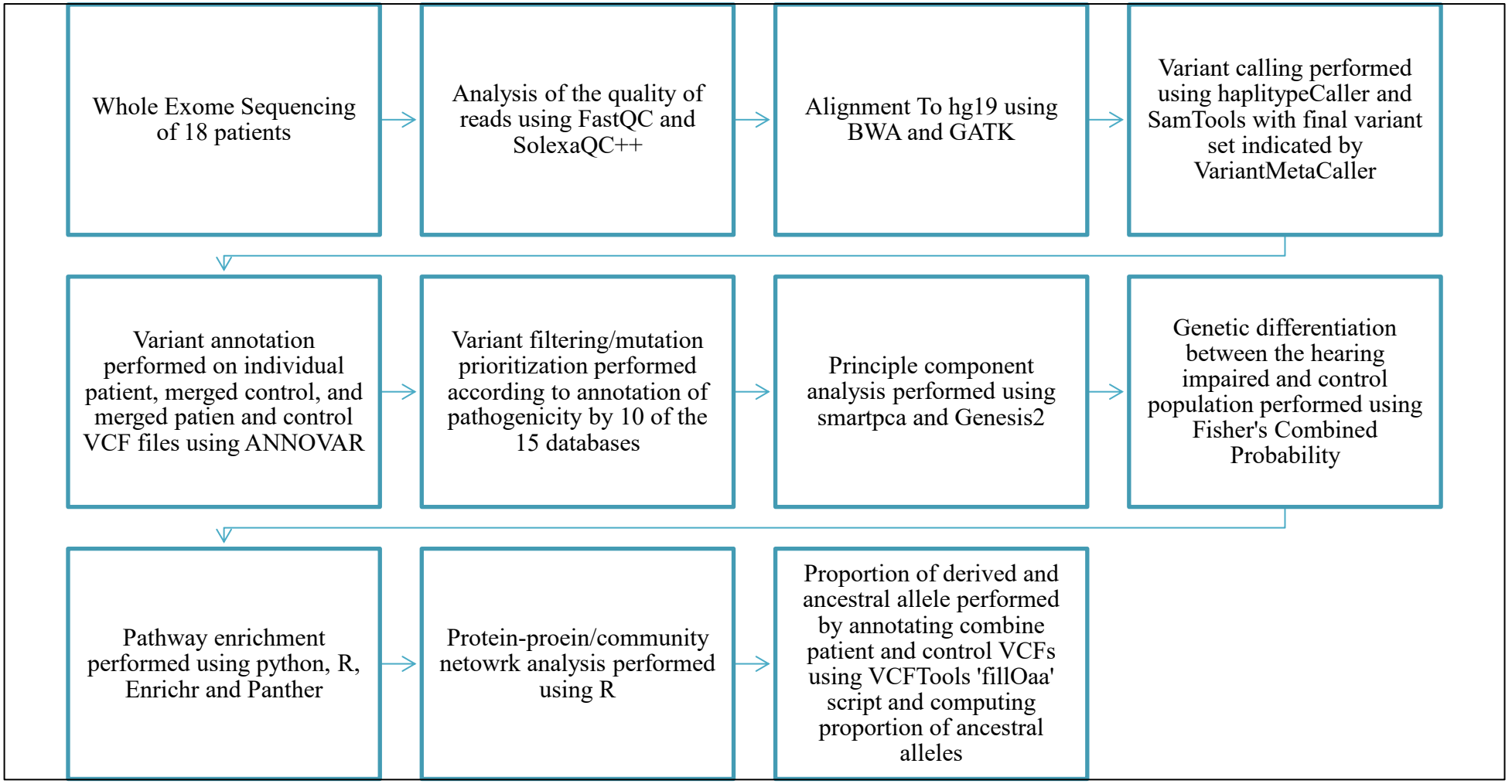


Figure 2.6.1: Bioinformatics pipeline used to analyse whole exome sequencing data. Pipeline indicates the steps taken in the analysis of the whole exome sequencing in order to generate the results

Chapter 2.7: Validation by Sanger Sequencing

Chapter 2.7.1: Primer Design

Primers were designed to amplify the variants identified in *VTN, RPL3L, DHRS4L2, MYO3A* and *MYO15A* during filtering. The primers are described below in Tables 6. These primers are necessary for amplification of the target regions and are designed to decrease non-specific binding and amplification.

Table 6: Properties of the primers amplifying fragments surrounding target SNPs

Gene	Direction	SNP	Sequence	Length (bp)	GC Content (%)	Melting Temperature (°C)	Molecular Weight (g/mol)	Expected Fragment Size (bases)
VTN	Forward	rs370266094	CCA GGG CAA GAC TGG AGA T	19	57.9	56.8	5 886.9	406
	Reverse		TGG GAA CGG CTG CCT TA	17	58.8	56.7	5 226.4	
RPL3L	Forward	rs143345186	AAG CCC TGC CTG TGC TC	17	64.7	58.4	5 122.4	420
	Reverse		TTC TGG GAC CAT CCA GCT CTA	21	52.4	57.4	6 357.2	
DHRS4L2	Forward	rs145720846	TAG TAG GCA CAC GTA GGA GTT A	22	45.5	54.5	6 823.5	390
	Reverse		TCT AGG AAA GCA GGG AAG GA	20	50.0	55.0	6 264.1	

MYO3A	Forward	rs140301218	TAC TTT CAT TGG ACC AAC AGG AT	23	39.1	54.1	7 022.6	643
	Reverse		GCC AAA TTA TCT CTG CAC TTC TTA C	25	40.0	54.2	7 526.9	
MYO15A	Forward	rs138861831	GGG TCC TTC TTC TCC TCT ACT	21	52.4	55.0	6 290.1	554
	Reverse		ACT GGT GTA GCC TCT CTT CT	20	50.0	54.9	6 050.0	

The primers were designed by downloading the genomic sequence for each gene from Ensembl (Aken et al., 2016) and locating the region of interest surrounding the SNP. The region of interest was then copied into the PrimerQuest Tool, available from Integrated DNA Technologies website, and specifying design parameters of the primer pairs generated.

For the primer pairs indicated in Table 6, the primers needed to amplify a region between 350 and 650bp, with a minimum of 100bp upstream and downstream of the target SNP, which allows the target region to be long enough for satisfactory resolution by the capillary electrophoresis. The primers needed to amplify between 54°C and 60°C and contain a GC content between 30% and65%, which allows the primers to bind selectively to the region of interest. The last parameter was that the primers needed to be between 16bp and 30bp long, which is the optimal primer length.

The primers were then analysed for specificity to ensure that minimal non-specific binding occurs which could interfere with efficient amplification of the target region or with downstream analysis of the PCR products. This analysis was done by imputing the generated primers into Primer-BLAST (Ye et al., 2012)

The primers for *VTN, RPL3L* and *DHRS4L2* were redesigned due to poor amplification. The primers were redesigned to specification listed above but a GC-clamp was incorporated into the 3' end of the primer to assist in securing the primer to the template strand and improving amplification. The primers generated were check for specificity in Primer-BLAST and In-Silico PCR (Kent, 2002), based on the GRCh38/hg38 reference genome assembly build, available from the UCSC Genome Browser. The new primers for *VTN, RPL3L* and *DHRS4L2* are indicated in Table7.

Table 7: Properties of the redesigned primers amplifying fragments surrounding target SNPs

Gene	Direction	SNP	Sequence	Length (bp)	GC Content (%)	Melting Temperature (°C)	Molecular Weight (g/mol)	Expected Fragment Size (bp)
VTN	Forward	rs370266094	TACTCCCAGTACTGTTTCCC	20	50.0	55.9		451
	Reverse		TGCCTTAGGGCAAGGG	16	62.5	55.2		
RPL3L	Forward	rs143345186	TAGCTCCTGCGCTTACCG	18	61.1	59.2		719
	Reverse		CCTTGGCTGCAGTATGGG	18	61.1	57.8		
DHRS4L2	Forward	rs145720846	GAAAGTGATGTCAGTTCTTGGGAACTAAGG	30	43.3	63.8		570
	Reverse		CTGACTGAGATGGGATGATTCTCTGACG	28	50.0	64.5		

Chapter 2.7.2: Primer Optimization

The PCR experiments, with the above specified primers, were optimised for the best conditions for the polymerase chain reactions to occur. This can be done through temperature and concentration gradients for the various reagents that are necessary for the reactions to occur as well as adjusting reaction parameters in order to optimise amplification.

The various primer pairs were initially subjected to a temperature gradient using control DNA, within 5°C above and below the predicted melting temperature. A second temperature gradient would be run with the gradient 1°C above and below the temperature that produces the brightest band in the initial temperature gradient. The temperature gradients allowed the primer efficacy to be observed at 8 different temperature points in order to assess the best temperature for primer annealing in the PCR reaction. The primers amplifying the variants in *MYO15A* and *MYO3A* had sufficient amplification for there to be no need to run a concentration gradient to further optimise the primer pairs. The primer pairs amplifying the fragments in *VTN*, *DHRS4L2* and *RPL3L3* had poor amplification coupled with non-specificity and were thus redesigned and the new primer pairs provided satisfactory amplification of the target fragments with no non-specific binding.

Chapter 2.7.3: Polymerase Chain Reaction and product detection

The fragments of interest were amplified using polymerase chain reaction (PCR) in order to provide enough copies of the fragment of interest for sequencing. The reaction mixture included 16.9µl of distilled, DNase free, water (Sabax water for injections, Adcock Ingram, Johannesburg, South Africa); 5µl of GoTaq 5X Buffer (Promega, Madison, Wisconsin, USA); 1µl of 5mM dNTPs (Solis Biodyne, Tartu, Estonia); 0.5µl of 20µM of the forward primer and of the reverse primer (Whitehead Scientific, Johannesburg, South Africa); 0.1µl of 5U/µl of GoTaq Polymerase (Promega, Madison, Wisconsin, USA); and 1µl of 100ng DNA.

This PCR was performed by initially denaturing the genomic DNA into single stranded DNA by heating it to 95°C for 3 minutes. This was followed by 35 cycles of denaturing the DNA by heating it to 94°C for 30 seconds, followed by annealing the primers to the DNA by lowering the temperature to 61°C for 40 seconds. The primers were then elongated, and the DNA transcribed, by increasing the temperature to 72°C for 1 minute. Following the cycles of amplification, the fragment length for all the fragments was ensured by allowing the fragments to be elongated at 72°C for 10 minutes.

Chapter 2.7.4: PCR clean-up

The product of the PCR reactions may be cleaned prior to sequencing to eliminate unincorporated dNTPs and to degrade the primer pairs and single stranded DNA. This was done by incubating 8.9µl of the PCR product with 1µl of Thermosensitive FastAP (Thermo Fischer Scientific, Waltham, USA) and 0.1µl of Exo1 endonuclease (Thermo Fischer Scientific, Waltham, USA) at 37°C for 1 hour. This was followed by inactivating the enzymes by raising the temperature of the reaction to 75°C for 15 minutes.

The Thermosensitive FastAP will degrade the excess dNTPs in the PCR product. The Exo1 endonuclease will degrade single stranded DNA and the primers.

Chapter 2.7.5: Direct Cycle Sequencing and Capillary Electrophoresis

The cleaned PCR products were subjected to direct cycle sequencing. This allows the DNA sequence to be determined for each fragment determined through the use of chain termination sequencing and capillary termination as developed by Sanger, Nicklen & Coulson (1977).

Sequencing was performed using the BigDye ® Direct Cycle Sequencing Kit (Thermo Fischer Scientific, Waltham, USA). 2µl of the PCR cleaned product were incubated with 1µl BigDye® Terminator Mix, 2µl BigDye® Dilution Buffer, 0.5µl of the forward or reverse primer and 4.5µl dH_2O. The reaction was run with an initial denaturation step at 96°C for 5 minutes. This was followed by 30 cycles of denaturation at 96°C for 30 seconds, primer annealing at 50°C for 15 seconds and strand elongation at 60°Cfor 4 minutes.

The sequencing product was cleaned up by precipitating the sequencing product, in 50µl 100% (absolute) ethanol and 2µl sodium acetate, with vortexing to mix, overnight at -20°C. The DNA was pelleted out of the solution by centrifuging at 10 000g for 10 minutes. The supernatant was discarded whilst carefully ensuring not to disturb the pellet of DNA. 50µl of 70% ethanol was then used to resuspend the DNA pellet using vortexing, whereby these steps allow the DNA to precipitate out of solution and to be concentrated in preparation for capillary electrophoresis. The DNA was once again pelleted out by centrifuging at 10 000g for 10 minutes. The supernatant was once again discarded, and the pellet was allowed to airdry for 2 hours before it was resuspended in 10µl HiDi Formamide and the run on the 3130xl Genetic Analyzer ABI Prism (Applied Biosystems) capillary electrophoresis.

Chapter 2.7.7: Sequencing Results Analysis

The sequencing results from capillary electrophoresis were processed using the *Sequencing Analysis Software* (Applied Biosystems). The software performed the base calling from the

capillary electrophoresis data and, thus providing the DNA sequence, for each of the samples subjected to capillary electrophoresis. The DNA sequences were then analysed/visualised using the Unipro UGENE (Golosova et al., 2014; Okonechnikov et al., 2012).

Chapter 2.7.8: BAM file analysis

The sequencing results obtained by Direct Sanger Sequencing presented were discordant with the result from bioinformatics analysis of the whole exome sequencing with regard to the putative causative mutations in the patients. Due to this discordance, the BAM files for each of the putatively resolved was viewed using Integrative Genomics Viewer (IGV) software (Robinson et al., 2011; Thorvaldsdottir et al., 2018).The SNP position for rs140301218,in *MYO3A*, rs138861831, in *MYO15A*, and rs14503583, in *COL9A3*, was determined using Ensembl GRCh37 (Flicek et al., 2014; Zerbino et al., 2018) and the SNPs were compared to a patient, or patients, who were indicated to not carry the variations by bioinformatics analysis as well as the hg19 reference sequence.

Summary of bioinformatics methodology

Ethics approval was granted for this study by the Human Research Ethics Committee if the University of Cape Town. Nineteen hearing impaired patients and 130 ethnically matched controls from the Cameroonian population were included in this study. Patients were previously recruited from Schools of the Deaf and ENT out-patient clinics if they presented with HI that presented before 15 years of age. Patients included in this study had to present with putative non-syndromic HI that was not a result of causative mutations in *GJB3*, *GJB6* and *GJA1*.

DNA samples from the patient and control populations were subjected to whole exome sequencing, following successful integrity checks. The whole exome sequences were then subjected to quality checks and aligned to the reference hg19 genome assembly. Variant calling was performed, and the variants were annotated and filtered to determine putative causative mutations resulting in HI in the patients and variations that may putatively modify the HI phenotype that were present in the entire Cameroonian population.

The population structure of the Cameroonian population was analysed through the use of principal component analysis by comparing the patient and control population to other African populations and by comparing the patient and control population to one another. The patient

and control population were also analysed to identify the presence of significant genetic difference between them.

The variations discovered in the novel genes were included into a list of known HI associated genes and these genes were used to determine protein-protein interactions, enrichment analysis and pathway analysis. This would generate an interaction plot and a network plot which incorporates the newly associated genes, and this may provide possible insight into the manner by which they affect HI.

Haplotype analysis was performed on the patients segregating the three novel, putative causative mutations, which could cumulatively resolve the HI in 50% of the patient population. The analysis would allow for insight into whether the variations are founder mutations in the Cameroonian population.

Analysis of the proportion of derived and ancestral alleles concluded the bioinformatics methodology. This analysis was done to determine the proportion of ancestral, and conversely derived, alleles in the patient and control populations, as well as in the Cameroonian population and select samples from the Democratic Republic of Congo. The analysis contrasted the proportion of ancestral alleles within the patient and control population at various minor allele frequencies, as well as indicating the proportion of ancestral alleles in genes that may be significant in HI in the Cameroonian and Congolese populations.

Summary of molecular methodology

The variants discovered in the whole exome sequencing and subsequent bioinformatics analysis were validated through direct cycle sequencing. This was done by designing primers specific to the genomic sequence surround the variants of interest. The five primer pairs were optimised to their optimal experimental conditions through a temperature gradient and these conditions were utilised to amplify the target regions surrounding the five variants.

The PCR products were cleaned up in preparation for sequencing by degrading unincorporated dNTPs and the primers from the PCR reaction. This was followed by direct cycle sequencing of the PCR products using the chain termination technique. The sequencing product was then up by precipitating with 70% ethanol and sodium acetate overnight which was followed with by centrifugation of the mixture to pellet out the DNA. The supernatant was discarded, and the DNA was re-suspended in 100% ethanol and subjected to centrifugation once again, to allow the concentration of the DNA prior to electrophoresis. The supernatant was discarded and the

pellet was air dried before been re-suspended in HiDi Formamide and subjected to capillary electrophoresis.

Chapter 3: Genomics architecture of non-syndromic hearing impairment results

Chapter 3.1: Patients clinical information

The patient population consisted of an almost equal distribution of male and female participants. The mean age of the population was 13 years old (± 4.48 years). The youngest individual in the cohort was 6 years old and the oldest individual was 27 years old. The majority of the HI presented as congenital (88.9%) and profound HI (88.9% in right ear and 83.3% in left ear) and the demographics of the population are summarised in Table 6

Table 8: Demographics of the Patient Population

Demographic		Summary statistics	Percentage (%)
Gender	Male	10	55,56
	Female	8	44,44
Age of Onset	Congenital/Prelingual	16	88,89
	Prelingual	2	11,11
Degree of hearing impairment – Left	Severe	1	5,56
	Profound	16	88,89
	No Data	1	5,56
Degree of hearing impairment – Right	Severe	2	11,11
	Profound	15	83,33
	No Data	1	5,56
Transmission	Familial	12	66,67
	Sporadic	6	33,33
Mechanism	Sensorineural	15	83,33
	Not Determined	3	16,67
Consanguinity	Yes	0	0,00
	No	18	100,00
Endogamous Population	Yes	7	38,89
	No	11	61,11

Table 6 also indicates that the majority of the HI was sensorineural (83.3%) and familial HI (66.7%). None of the patients were from consanguineous families, though 7 (39%) patients were from endogamous populations.

Chapter 3.2: Variant filtering

Variant filtering was performed on both the individual datasets and the combined patient dataset. Filtering variants in the combined patient dataset resulted in population specific variants, which were present in the merged control and patient population, which were classified as Damaging/Deleterious by 10 of the 15 annotation databases. The variants discovered are indicated below in Table 9, population specific variants, and Table 10, individual/clinical variants.

Table 9: Candidate variants discovered at the population level of the Cameroonian merged patient and control population

Gene	Region	snp138	A1/A2	Variant Type	cDNA Change	Protein change
FOXD4L6	9p11.2	rs71246017	G/C	Non-synonymous	-*	-*
DHRS4L2	14q11.2	rs145720846	G/A	Non-synonymous	c.134G>A	p.Gly45Asp
RPL3L	16p13.	rs143345186	C/T	Non-synonymous	c.151G>A	p.Ala51Thr
VTN	17q11.2	rs370266094	G/T	Non-synonymous	c.719C>A	p.Pro240His

Data for the SNP rs71246017 in *FOXD4L6* is not available. All the variants, including the rs71246017, occur in the exonic region of the genes and all are classified as non-synonymous. All four genes have not been previously associated with HI and each individual in the merged dataset needed to have the variant for it to filter out.

Variant filtering indicated putative pathogenic variants, indicated in Table 10, in *MYO3A, MYO15A* and *COL9A3* genes which may result in non-syndromic, autosomal recessive phenotype observed in the patients. The variants observed in the genes are most probably pathogenic, however, rs138861831 in *MYO15A* has conflicting clinical annotation in ClinVar whereby it has been annotated as likely benign (1 study) and with uncertain significance (2 studies) in three studies. *COL9A3* is typically associated with Stickler Syndrome, however, there is a study associating a recessive mutation in *COL9A3* to non-syndromic HI (Asamura et al., 2005). The most likely variation associated with HI in the patient population is the variation in *MYO3A* whereby the gene has been previously associated with HI in an African-American family (Grati et al., 2016).

Variant filtering indicated variations in *USH2A, MYO1A* and *HSD17B4,* as shown in Table 11. The variations in these genes are unlikely to be causative mutations in the patients as *MYO1A* has been excluded as a HI associated gene in several prior studies (Eisenberger et al., 2014). The variation in *USH2A* has been indicated to be benign by ClinVar and *USH2A* is associated with Usher Syndrome Type 2A (Eudy et al., 1998; Kimberling et al., 1990). Finally, the *HSD17B4* is associated with Perrault syndrome (Jenkinson et al., 2013; Perrault et al., 1950; Pierce et al., 2010)and D-bifunctional protein deficiency (Grønborg et al., 2010; Suzuki et al., 1994; Watkins et al., 1989), both syndromic conditions that present with HI, therefore the variations in *HSD17B4* are most likely not causative mutations in the patients as the phenotype observed does not match with the phenotype of either Perrault syndrome or D-bifunctional protein deficiency.

Table 10: Candidate variants found in 9/18 Cameroonian patients affected by Non-Syndromic Hearing impairment, with whole exome sequencing and analysis of 172 known genes[£]

Number of Patients (age in years)	**Families Pedigrees**	**Genotypes**	**Gene**	**Region**	**SNP**	**cDNA Change**	**Protein Change**	**ExAC_AFR MAF**	**ExAC_AS MAF**	**ExAC_EUR MAF**	**Cameroonian Controls MAF (*N* =161)**
3 (10, 10, 16)	SP, SP, MP	Homozygous	*MYO3A*	10p12.1	rs140301218	c.C424T	p.H142Y	0.0026	0.0007	0	0
4 (12, 8, 10, 10)	MP, MP, SP, MP	Homozygous	*MYO15A*	17p11.2	rs138861831	c.C4888T	p.R1630C	0.0083	0.0003	0	0
2(16, 11)	SP, MP	Homozygous	*COL9A3*	20q13.33	rs145035835	c.G406A	p.G136S	0	0.001	0	0

[£]Exonic. nonsynonymous variants that were considered deleterious/damaging according to 14 different functional scores of annotation databases, including SIFT, Polyphen2_HDIV, Polyphen2_HVAR, Likelihood ratio test, MutationTaster, MutationAssessor, FATHMM, fathmm-MKL, RadialSVM, LR, CADD, PROVEAN, MetaSVM and MetaLR, as previously reported.[8]

Abbreviations Families' Pedigrees, SP: simplex; MP: multiplex family; SNP: Single Nucleotide Polymorphism; ExAC: Exome Aggregation Consortium; AFR: African; EUR: European; AS: Asian; MAF: minor allele frequency;

Table 11: Candidate variants discovered in individual patients in Genes that are unlikely to explain the HI phenotype in our patients

Gene	Region	snp138	A1/A2	Genotype (Het/Hom)	cDNA change	Protein change	Patients	Pedigree (Simplex/Multiplex)	Camerooni an control MAF	ExAC AFR	ExAC AS	ExAC EUR
USH2A	1q41	rs111033481	C/T	Hom	c.G2546A	p.C849Y	CAM104	Simplex	0	0.0153	0.0006	0
							CAM051	Multiplex				
							CAM056	Multiplex				
							CAM076	Multiplex				
MYO1A	12q13.3	rs149803771	C/T	Hom	c.G304A	p.E102K	CAM103	Simplex	0	0.0002	8.67E-0.005	0
							CAM056	Multiplex				
							CAM075	Multiplex				
							Cam076	Multiplex				

							Cam078	Multiplex				
HSD17B4	5q23.1	rs35281104	G/A	Hom	c.G1631A	p.R544H	Cam119	Simplex	0	0.0009	0	0
							Cam051	Multiplex				
							Cam078	Multiplex				

Chapter 3.3: Population structure

The principal component analysis (PCA) was performed on the Cameroonian population, consisting of both patients and controls, with other African populations and on the Cameroonian patients and control population solely. The PCA, in Figure 3.3.1, indicates that the Cameroonian population clusters together in the bottom left corner. The populations represented in the PCA, from Phase 3 of the 1000 genomes project, are the Yoruba (YRI) and Esan (ESN) from Nigeria, individuals from the West Divisions of Gambia (GWD), the Luhya from Webuye in Kenya (LWK), individuals of African Ancestry from the Southwest of the USA (ASW), the Mende in Sierra Leone (MSL) and individuals of African Caribbean from Barbados (ACB). Six individuals from the Democratic Republic of Congo (DRC) were also included in the PCA.

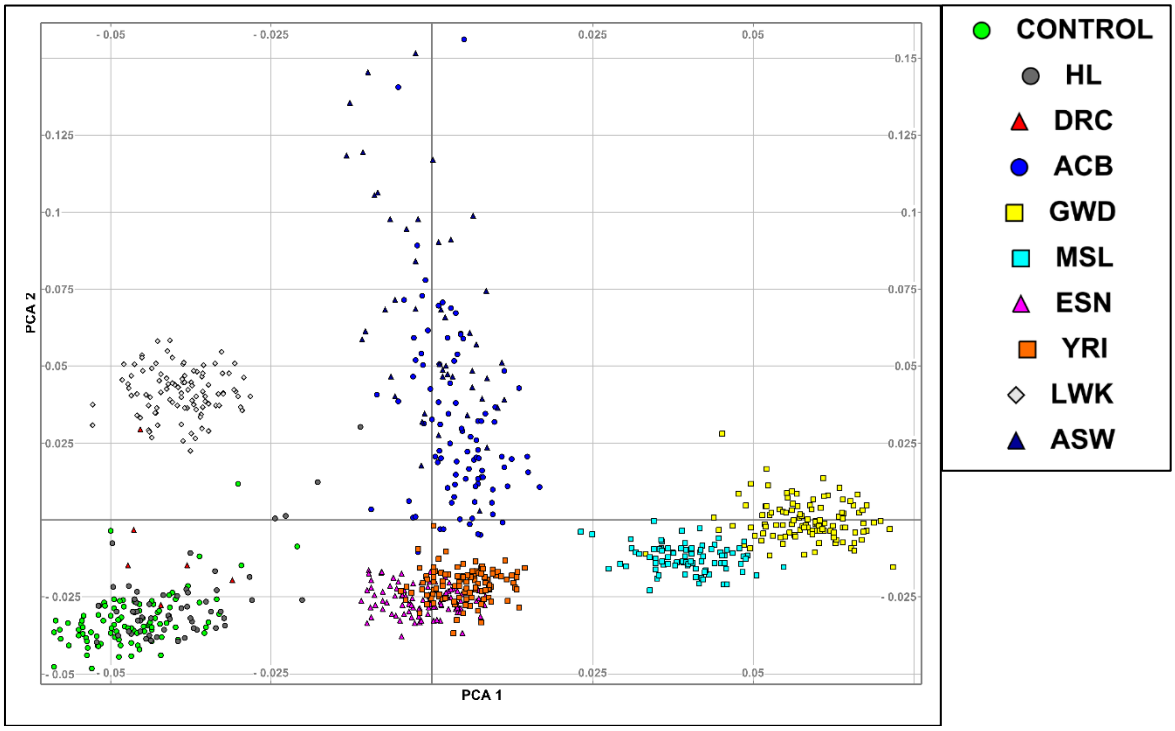


Figure 3.3.1: PCA comparing Cameroonian population with other African populations with key. The plot indicates the distribution of the Cameroonian population when compared to individuals of African ancestry in populations from Democratic Republic of Congo, Nigeria, the Gambia, Kenya, Barbados and the United States of America. The Cameroonian population clusters in the lower left corner of the plot with the Congolese population interspersed within the cluster. The key alongside indicates which symbol correspond to which population in the PCA.

The PCA in Figure 3.3.2 indicates the structure of the Cameroonian population, which has a distinct cluster in the lower left corner of the PCA. The five of the six individuals from the DRC population are interspersed within the Cameroonian population with only one individual clustering with the Kenyan (LWK) population.

A second PCA plot compares the patient and control Cameroonian population to one another. The PCA, Figure 3.3.2, shows the two populations dispersed in the plot. Neither the control nor the patient population forms a distinct cluster.

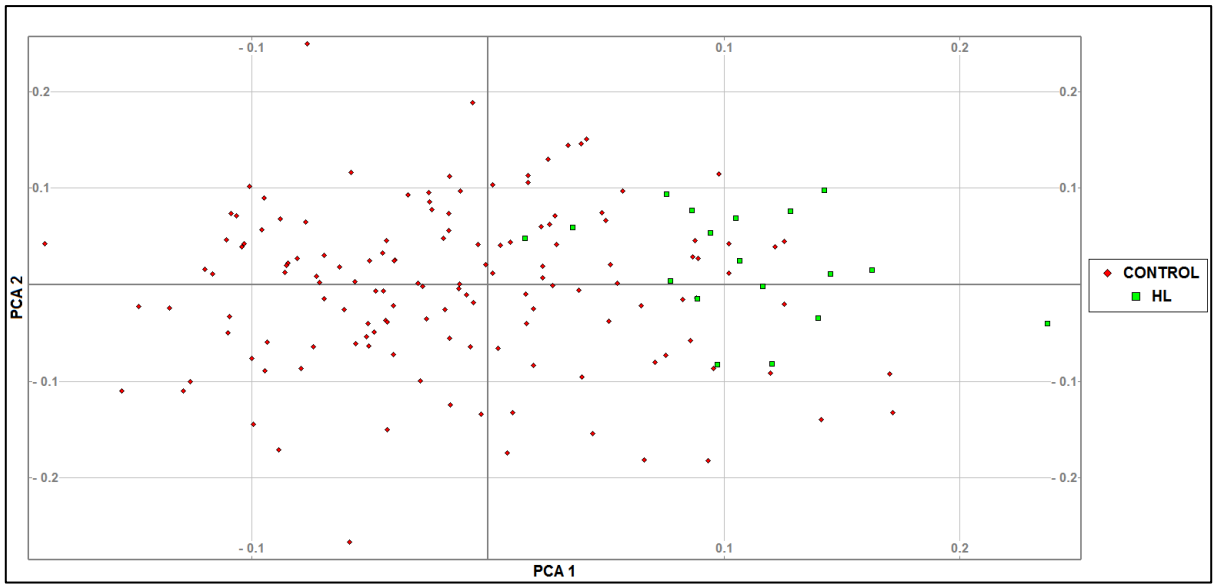


Figure 3.3.2: PCA plot of the Cameroonian patient and control populations with key.

The plot indicates the diversity of the Cameroonian patient and control population which fail to form distinct clusters in the PCA.

Chapter 3.4: Genetic differentiation

Genetic differentiation between the control and patient populations was determined by analysing the unusual difference of the patient and control populations' aggregated SNP frequencies. This was done using Fisher's combined probability and the results are summarised in Table 12.

Table 12: Identified genetic difference between the Patient and Control populations.

Gene	HI SNP frequencies	CONTROL SNP frequencies	p-value
PRDX4	0.2109	0.2618	2.65E-31
PRKACA	0.2716	0.3211	9.71E-30
CBX1	0.2242	0.2726	1.83E-28
COPS5	0.2908	0.2488	4.35E-22
PRKCB	0.2540	0.2916	4.52E-18
GLUD1	0.2251	0.2538	2.36E-11
CRADD	0.2571	0.2853	4.66E-11
ITGB6	0.2426	0.2687	1.11E-09
CKS2	0.2500	0.2753	3.40E-09
FGF2	0.2376	0.2623	8.16E-09
IGFBP5	0.2561	0.2799	2.68E-08
RHOXF1	0.2655	0.2422	4.51E-08
ITGA8	0.2559	0.2770	6.81E-07
AMPH	0.2404	0.2606	1.91E-06
PVR	0.2336	0.2515	2.43E-05
SERPINE1	0.2513	0.2360	3.01E-04
RPL13A	0.2585	0.2730	5.66E-04
VTN	0.2601	0.2457	6.14E-04
ADH5	0.2346	0.2209	1.16E-03
ADTRP	0.2690	0.2816	2.77E-03
TGFB1	0.2553	0.2428	2.84E-03
PRKCA	0.2598	0.2710	7.42E-03
RPL3L	0.2577	0.2487	3.13E-02
G3BP2	0.2306	0.2396	3.23E-02
ETV7	0.2570	0.2657	3.84E-02
PLAUR	0.2668	0.2593	7.45E-02
IGF2	0.2707	0.2636	9.14E-02
POMC	0.2874	0.2937	1.31E-01

HGF	0.2595	0.2656	1.49E-01
CSNK2A1	0.2723	0.2663	1.54E-01
ITGB3	0.2738	0.2793	1.96E-01
KNG1	0.2458	0.2511	2.19E-01
TGFB2	0.2628	0.2583	2.93E-01
GEM	0.2536	0.2581	2.97E-01
TNFRSF11B	0.2372	0.2329	3.05E-01
ITGB8	0.2525	0.2563	3.75E-01
LACRT	0.2732	0.2770	3.75E-01
**SIX1*	0.3040	0.3078	3.77E-01
EGF	0.2376	0.2340	3.96E-01
PRKCG	0.2451	0.2477	5.53E-01
CD247	0.2644	0.2661	7.09E-01
VEGFA	0.2628	0.2614	7.47E-01
UBC	0.2681	0.2676	9.20E-01
FGG	0.2181	0.2185	9.41E-01

All genes above the red line have significant genetic difference between the patient and control population. Genes with an asterisk (*) have been previously associated with HI.

The aggregated SNP frequencies was analysed in 44 genes for significant genetic difference in the patient and control population. 25 genes indicate significant genetic difference between the two populations (at $p<0.05$). One gene, *SIX1* (MIM:601205), has been previously associated with HI, but it is shown that there is no significant difference between the patient and control population.

Chapter 3.5: PPI analysis and enrichment analysis

Filtering variants resulted in population specific variants listed in Table 9. These variants have not been previously associated with HI, and as such, were queried in a protein-protein interaction network to identify the interactions of their protein products with other HI associated protein.

Figure 3.5.1 indicates the protein-protein interactions associated with *VTN, RPL3L, FOXD4L6* and *DHRS4L2*, as well as the associated pathways and molecular functions. The protein products of *VTN* and *RPL3L* are associated with 42 pairings/interactions, with *FOXD4L6* and

DHRS4L2, indicating no interactions with other HI associated proteins, as indicated in Figure 3.5.1A. The interactions were queried in the Enrichr database which indicated the major pathways associated, which includes osteoclast differentiation, as indicated in Figure 3.5.1B. Figure 3.5.1C indicates that the associated molecular function, which include binding and catalytic activity.

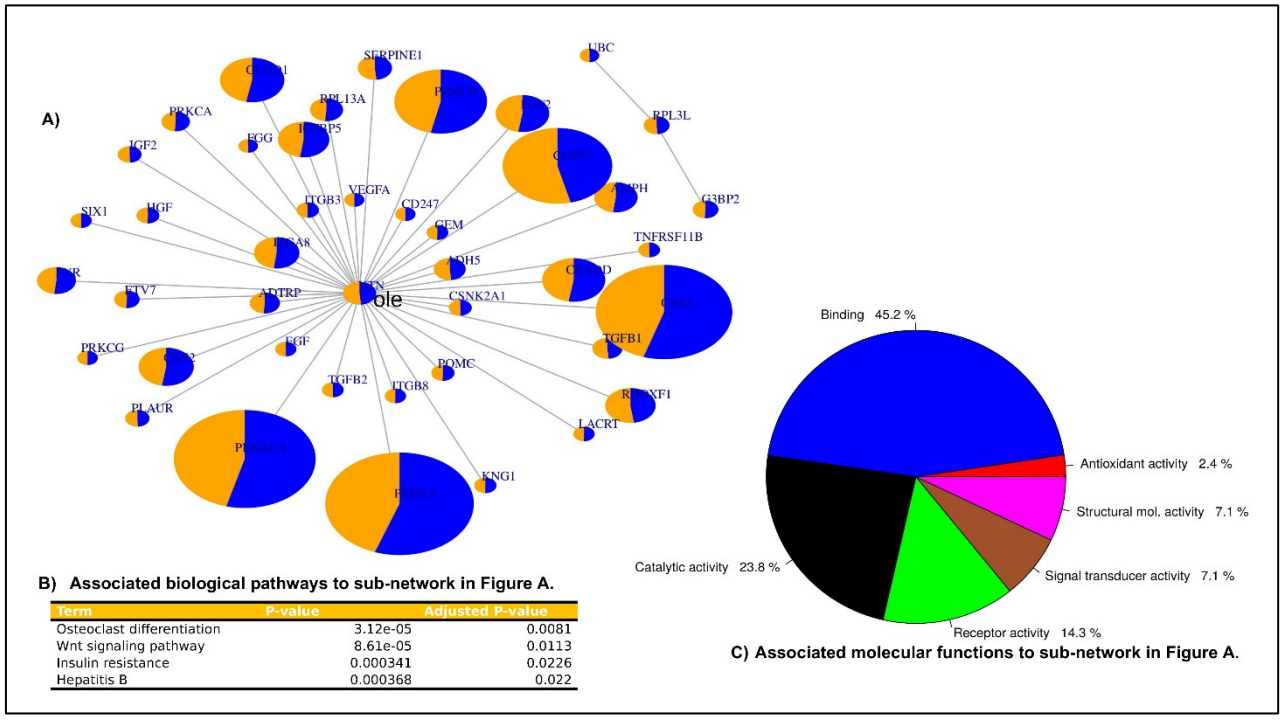


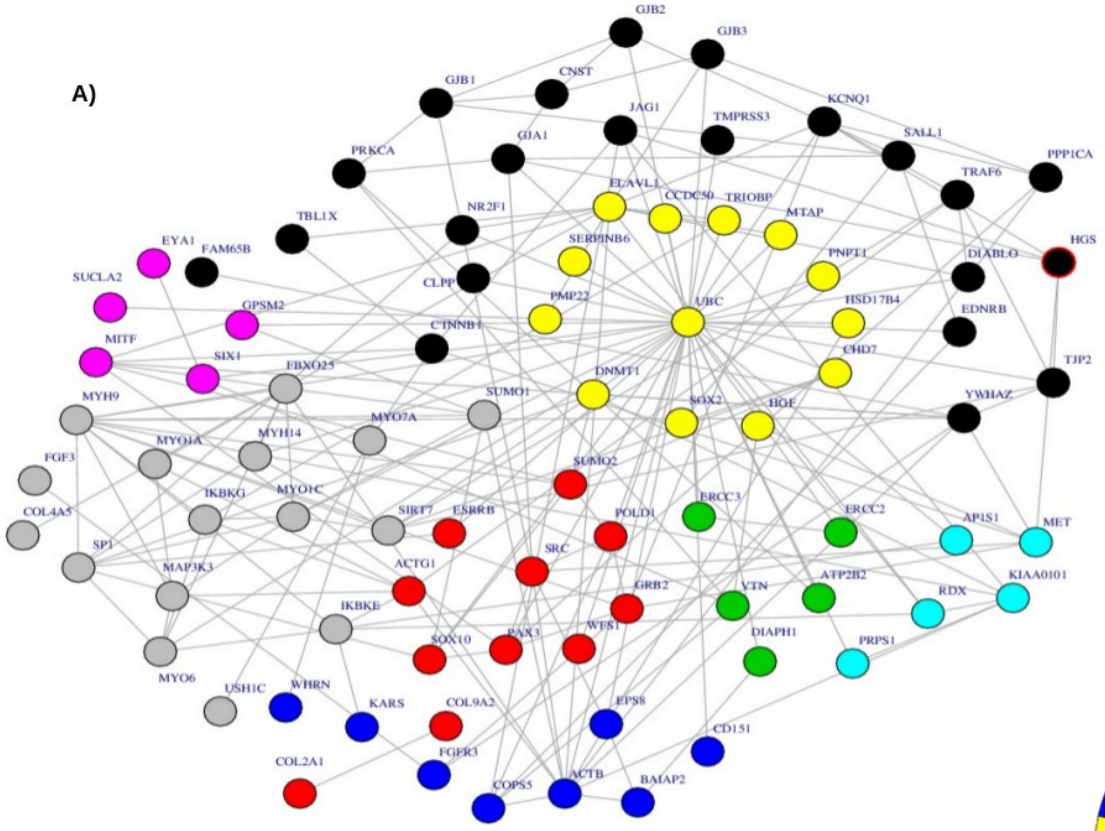
Figure 3.5.1: Reconstituted hearing impairment PPI network including enrichment analysis of the associated pathways and molecular functions. Figure 3.5.1A indicates the reconstituted PPI network including *VTN* and *RPL3L3* and other hearing impairment proteins. Figure 3.5.1B indicates the associated biological pathways and Figure 3.5.1C indicates the molecular functions associated with the network.

Chapter 3.6: Community network analysis

The candidate population genes, in Table 8, were combined with 159 known HI associated genes, present on the ClinVar database (Landrum et al., 2016) and the Deaf Variation Database (The Molecular Otolaryngology and Renal Research Laboratories. The University of Iowa, 2016) to perform the community network analysis. The community network analysis provided eight protein subnetworks which had 10 hub proteins as indicated in Figure 3.5.2A and the associated pathways and molecular functions are indicated in Figure 3.5.2B and Figure 3.5.2C respectively.

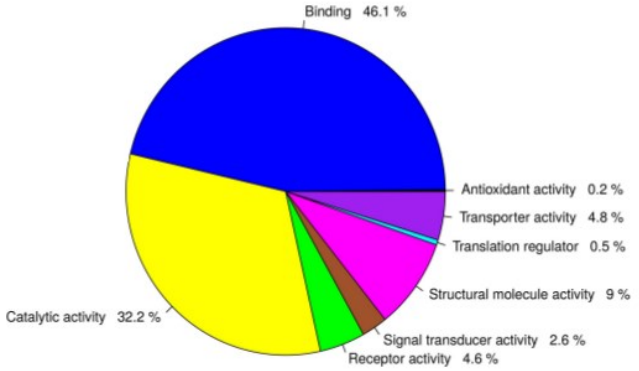
Five of the 10 hub proteins, ACTB, MITF, SSCTG1, MYH9, and ATP2B2, are known HI associated proteins, whereas the other hub proteins, UBC, SRC, YWHAZ, KIAA010, and VTN, have not. The entire network is centred around the UBC protein which is connected to the proteins in the yellow subnetwork as well as all the other networks.

The entire community network is associated with focal adhesion pathway as well as regulation of the actin cytoskeleton, indicated in Figure 3.5.2B. The abundant molecular processes associated with the network are binding and catalytic activity, as indicated in the pie graph in Figure 3.5.2C



B) Associated biological pathways to sub-network in Figure A).

Term	P-value	Adjusted P-value
Pathways in cancer	8.68e-71	2.25765e-68
Regulation of actin cytoskeleton	1.31e-42	1.70006e-40
Focal adhesion	5.26e-35	4.56157e-33



C) Associated molecule functions to sub-network in Figure A).

Figure 3.5.2: Community network analysis of the combined hearing impairment associated proteins and the candidate proteins identified. The network in Figure 3.5.2A indicates eight subnetworks, centred around UBC, with 10 hub proteins including UBC. The network is associated with regulation of the actin cytoskeleton and focal adhesion pathways, Figure 3.5.2B, and with binding and catalytic activity molecular functions, Figure 3.5.2C.

Chapter 3.7: Evolutionary adaptation of human hearing

The proportion of ancestral and derived alleles were computed for the HI associated genes using custom scripts. This proportion was analysed within the patient and control population, indicated in Figure 3.7.1, and in the control Cameroonian population and the DRC Sickle Cell Disease (SCD) population, that did not present with HI, indicated in Figure 10. In both figures, the proportion of derived alleles is equivalent to $Proportion\ of\ derived\ allele = 1 - Proportion\ of\ ancestral\ allele$.

It is shown in Figure 3.7.1 that the patient population has a decreased proportion of ancestral alleles at low minor allele frequencies (0.0 to 0.1), as compared to the control population. This proportion of ancestral alleles increases at higher minor allele frequencies, such that the control population has a decreased proportion of ancestral alleles at the higher minor allele frequencies.

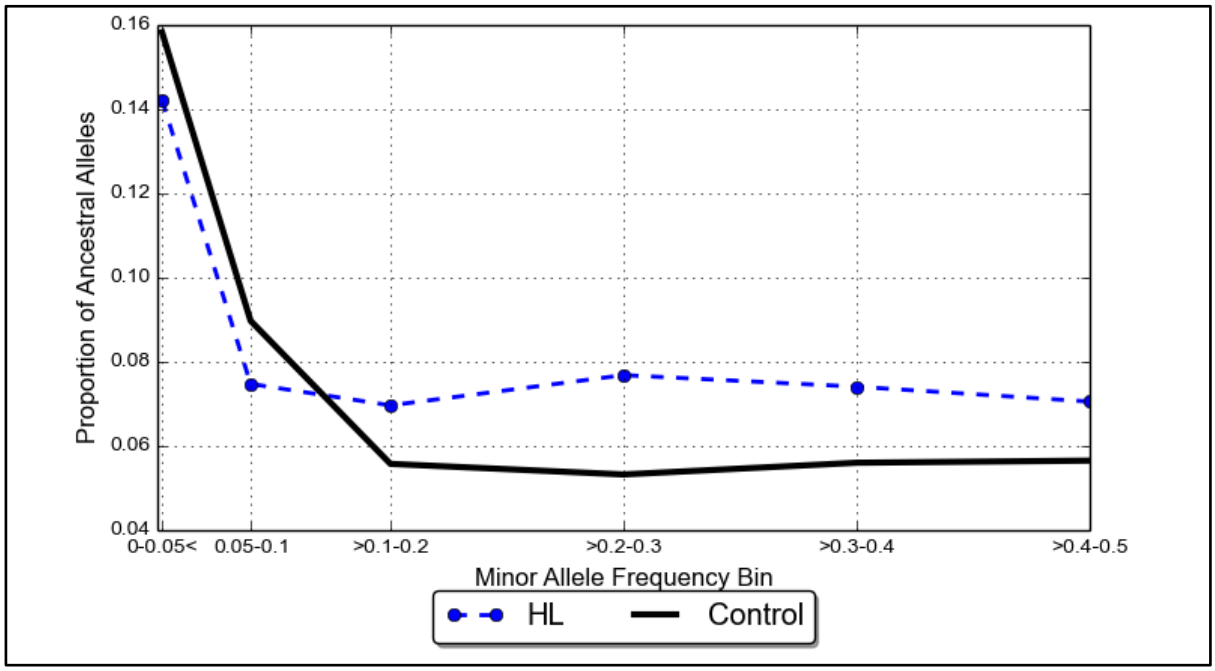


Figure 3.7.1: Proportion of ancestral alleles in the Cameroonian patient and control population. The patient population, denoted HL in the figure, has a lower proportion of ancestral alleles, and conversely a higher proportion of derived alleles, at low minor allele frequencies (between 0.0 and 0.1) as compared to the control population. The proportion of ancestral alleles in the patient population is greater at minor allele frequencies about 0.15, when compared to the control population.

The proportion of ancestral alleles in the Cameroonian control population and the SCD population from the DRC is indicated in Figure 3.7.2. The proportions of ancestral alleles vary

for various genes, but notably, both populations indicate a low proportion of ancestral alleles for *GJB2,* which translates into a high proportion of derived alleles within both populations.

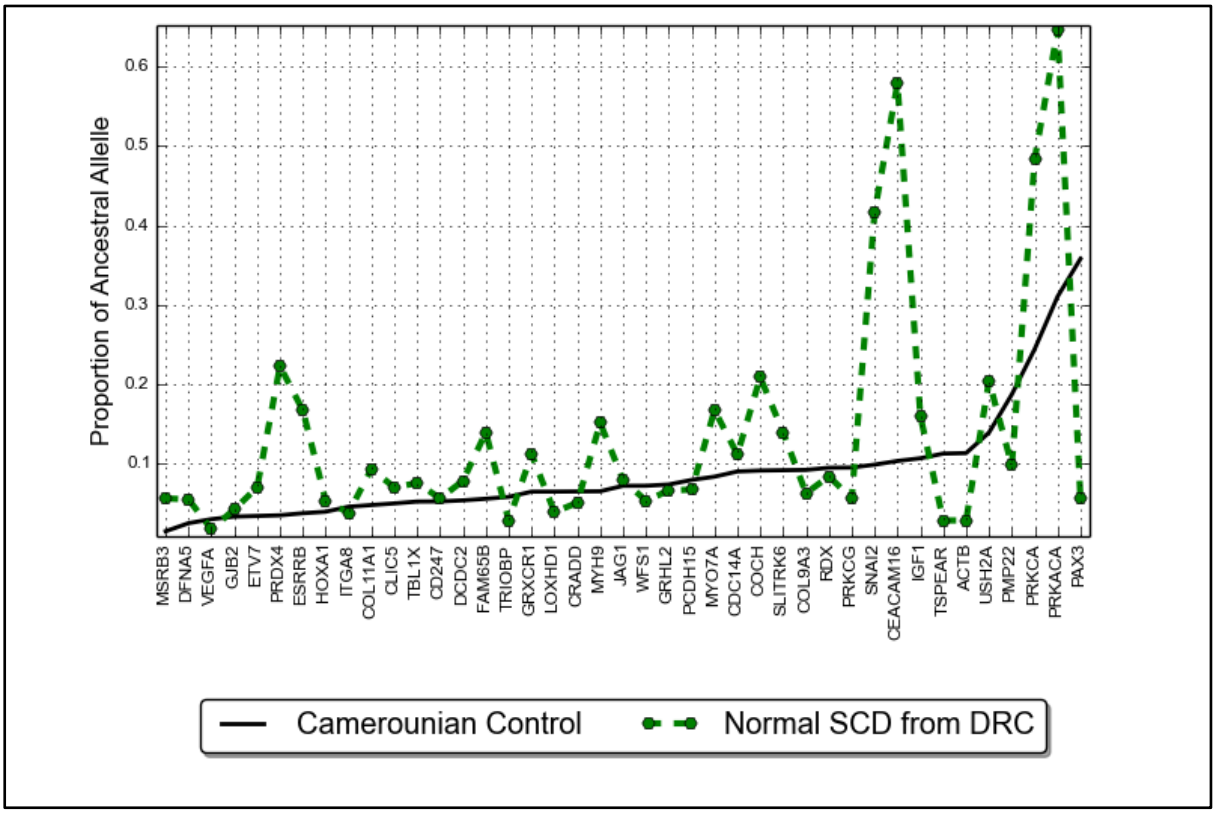


Figure 3.7.2: Proportion of ancestral and derived alleles in hearing impairment associated genes in the Cameroonian control population and a Democratic Republic of Congo Sickle Cell Disease population. The DRC SCD population does not present with hearing impairment and the figure illustrates the variation in the proportion of ancestral alleles in between two African populations.

Chapter 3.8: Validation

The candidate variants were validated through the use of Direct Cycle, or Sanger, Sequencing of each patient. Sequencing analysis of the patients indicated that the three individuals purported to have the rs140301218, in *MYO3A*, do not carry the variation and as such, the variation is most likely a false positive in this individual. A similar occurrence was observed for rs138861831, in *MYO15A*, in the four individuals purported to carry the variation, whereby Sanger sequencing indicated that the individual did not carry the variation observed in the whole exome sequencing and subsequent bioinformatics analysis. The electropherograms for individual CAM119 and an unrelated individual negative for the variation are indicated in

Figure 3.8.1, for comparison. Similarly, the electropherograms for individual CAM076 and an unrelated individual are indicated in Figure 3.8.2, for comparison.

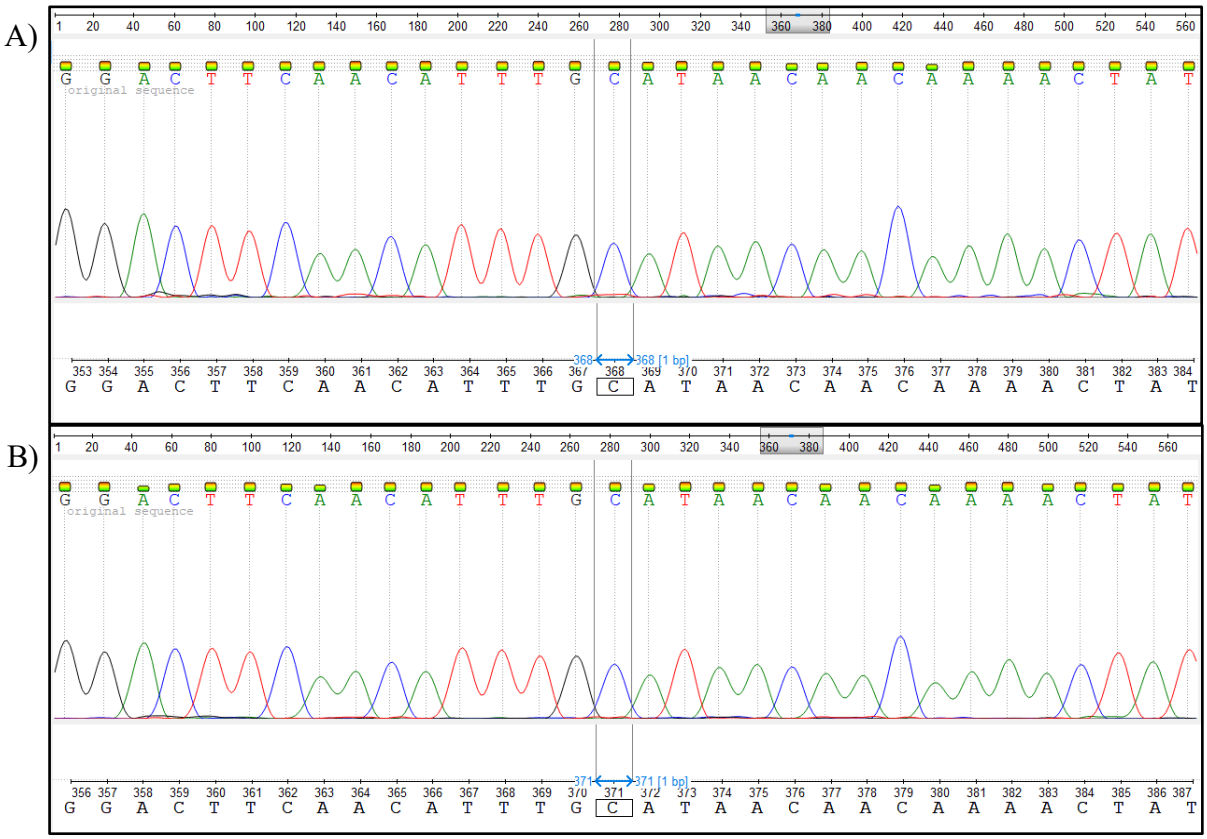
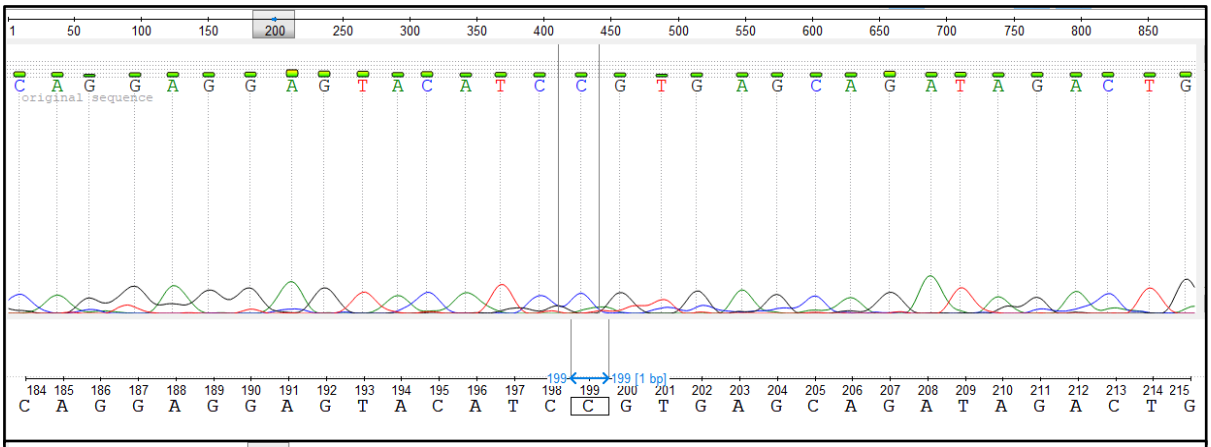


Figure 3.8.1: Electropherogram for CAM119 and an unrelated individual.
Electropherogram A) indicates the DNA sequence for individual CAM119, with the SNP position highlighted. CAM119 is homozygous for the CC allele and does not carry the T allele. Electropherogram B) indicates the DNA sequence for an unrelated individual, also with the SNP position highlighted. The individual also carries the wildtype C allele.

A)



B)

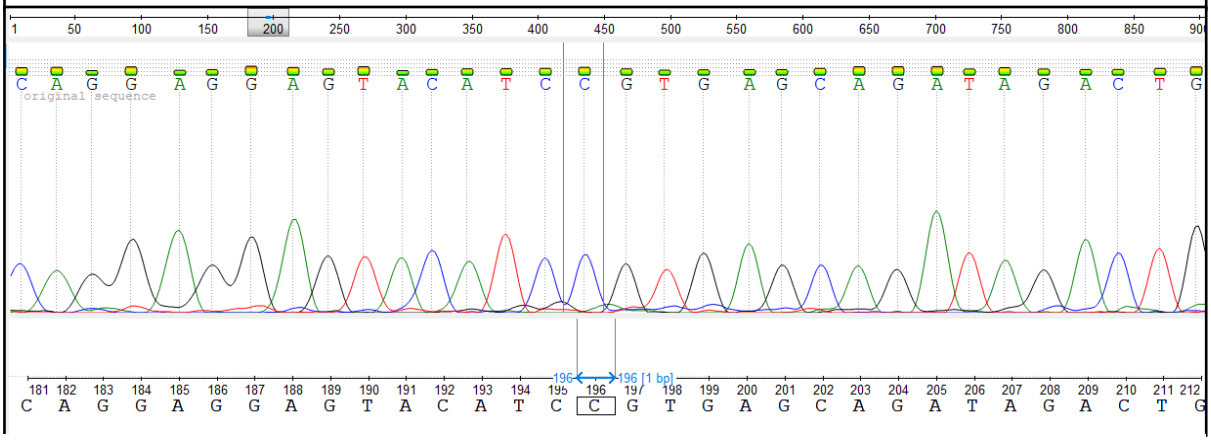


Figure 3.8.2: Electropherogram for CAM076 and an unrelated individual. A) Electropherogram A) indicates the DNA sequence for CAM076, with the SNP position highlighted. CAM076 is homozygous for the CC allele and does not carry the T allele. Electropherogram B) indicates the DNA sequence for an unrelated individual, with the SNP position highlighted. The individual carries the wildtype C allele, in homozygosity, at the SNP position.

CAM103, oddly enough, presented with two differing genotypes. The individual was shown to carry the wildtype C allele in an initial round of sequencing, which was followed by the individual presenting with an A allele in the second round of sequencing. This discrepancy is indicated in Figure 3.8.3.

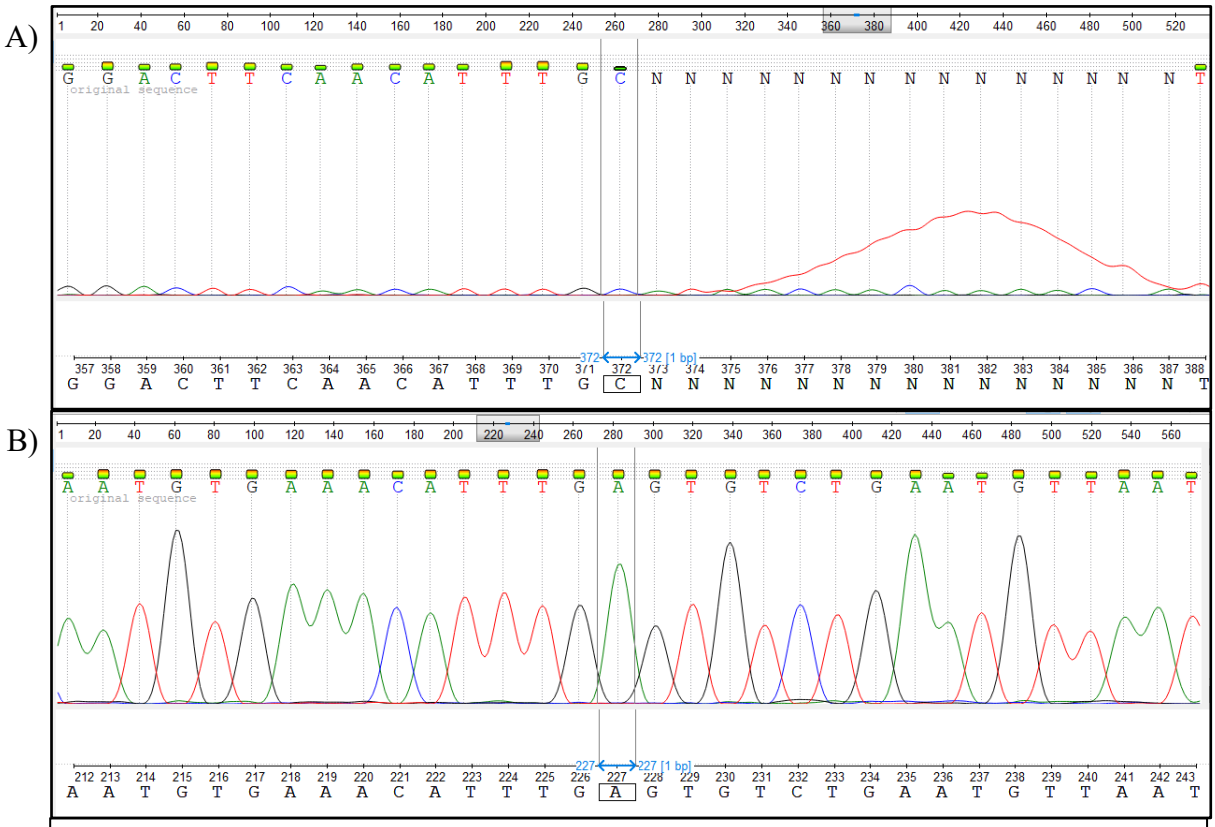


Figure 4.8.3: Sanger Sequencing discrepancy. Sequencing of the same variant in the same patient sample resulted in a discrepancy in the bases called for rs140301218. The base is called as a C allele in electropherogram A) and as an A allele in electropherogram B).

The results from the Sanger Sequencing were confirmed by viewing the respective SNPs in the individual patient BAM files, in IGV. This was additionally done for the two patients harbouring the rs14503583, in *COL9A3*. The analysis indicates the absence of all the identified clinical SNPs in all 9 patients that were purported to harbour them and this is indicated in Figures 3.8.4, Figure 3.8.5 and Figure 3.8.6 below. Reads for each of the SNPs for each person for *MYO3A* and *MY015A* were above 30X, with the highest read depth being 58X. Read depth for *COL9A3* for each of the patients was above 10X but below 20X.

The analysis of the BAM files, furthermore, addressed the discrepancy in the sequencing results for CAM103. The analysis was in agreeance with the sequencing results indicated in Figure 4.8.3B.

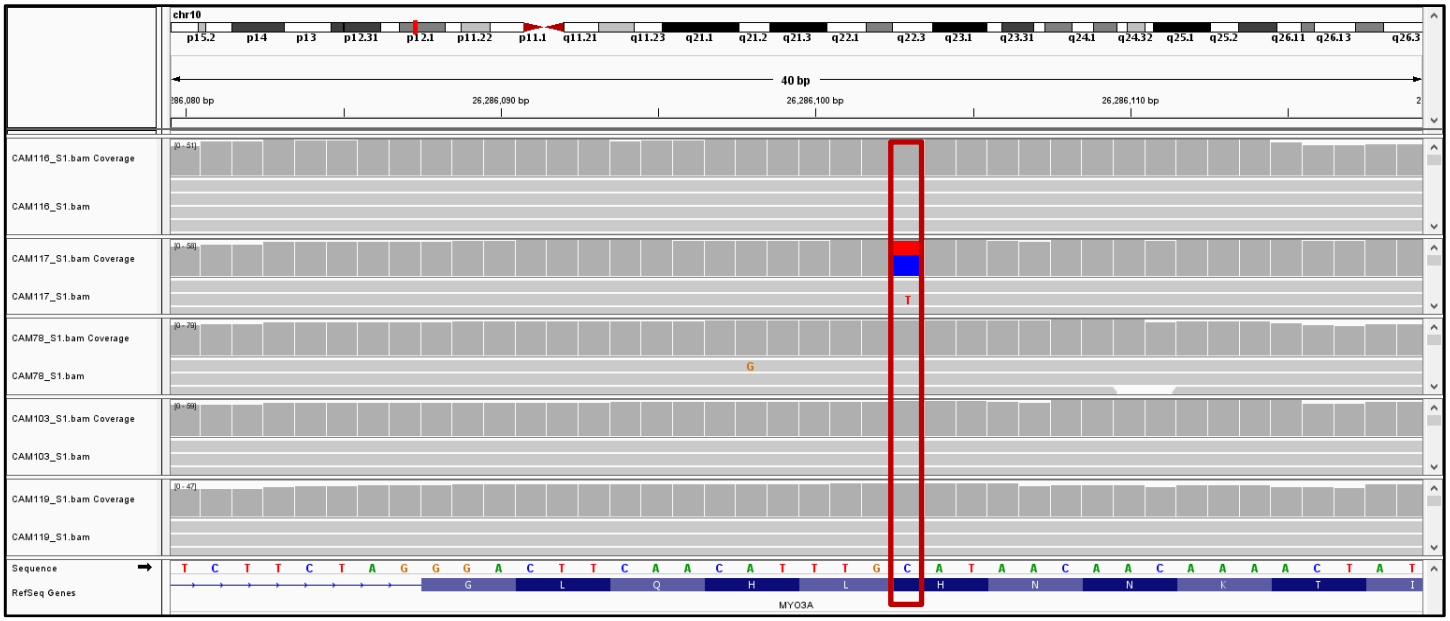


Figure 3.8.4: BAM file view for*MYO3A* sequences. The BAM file indicates the 3 supposed *MYO3A* patients, CAM78, CAM103 and CAM119, as well as two control patients, CAM116 and CAM117. The SNP of interest is shaded in grey, in the patients as well as one control individual, as are other nucleotides in the BAM files indicating that the SNP in the patients and the particular control individual is homozygous wildtype. The SNP is shaded in red and blue in a single control individual who is heterozygous for the mutant T allele

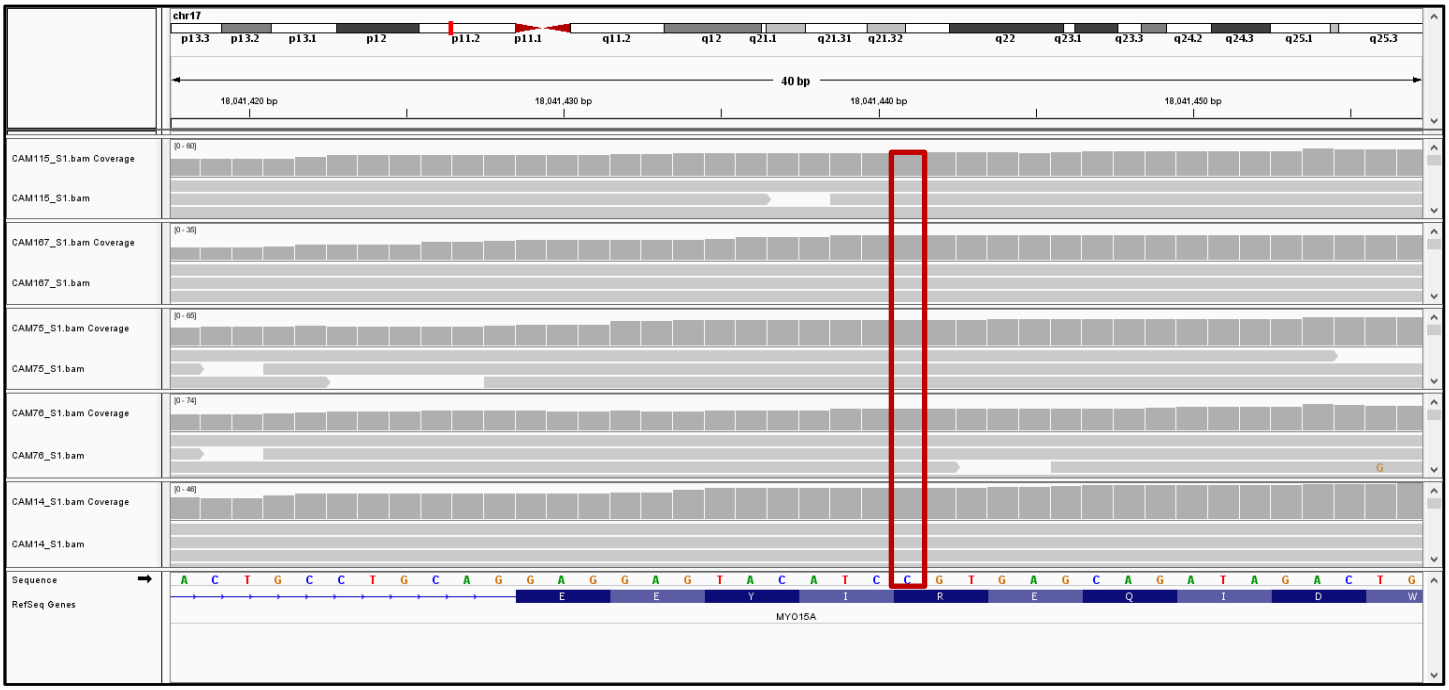


Figure 3.4.5: BAM file view of *MYO15A* sequences. The BAM files indicate the four purported *MYO15A* patients, at the top, CAM115, CAM167, CAM075 and CAM76, and one control individual, CAM014. The SNP of interest for that the patients were indicated to carry is shaded in grey, similarly to the control individual. This shading indicates that the nucleotide at that position is the same as the reference sequence

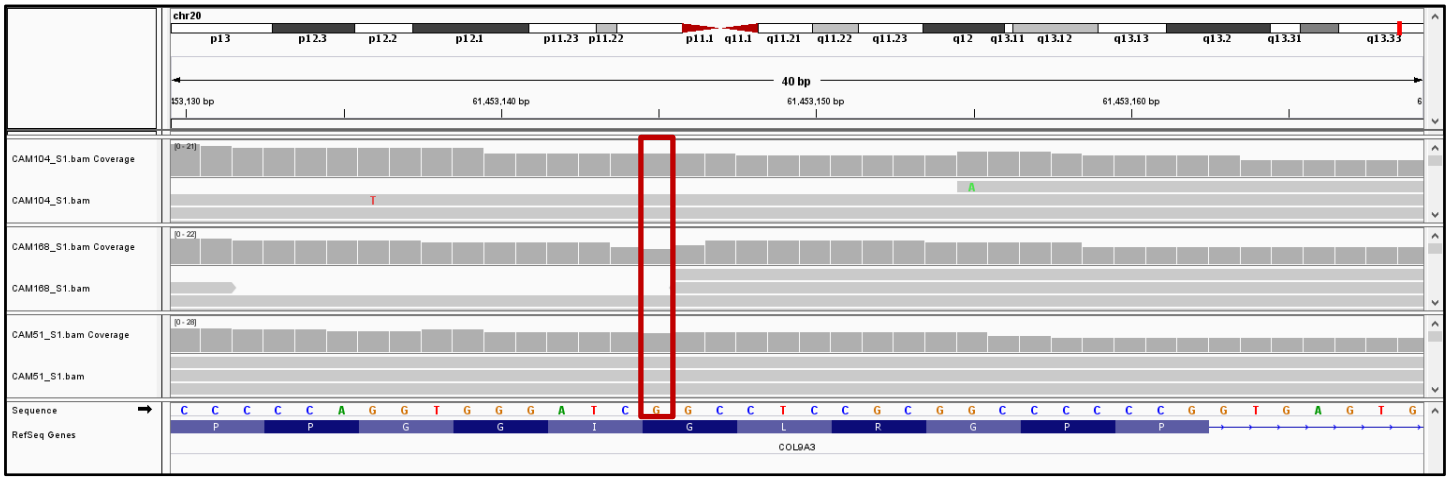


Figure 3.8.6: BAM file view of *COL9A3* sequences. The BAM files indicate the two purported *COL9A3* patients, CAM104 and CAM168, and a control individual, CAM51. The SNP of interest is shaded in grey in the two patients and the control patient, which indicates that the nucleotide at that position is the same as in the reference sequence.

The results provide indicate that the causative mutations resulting in HI were resolved for 0% of the patient population which is extremely low.

Summary of results

The patient population consisted of 18 hearing impaired individuals, of which the mean age of the population was 13 years old. 10 individuals of the population were male, and the majority of individuals were profoundly deaf in both the left and the right ear. The HI was familial in 12 of the individuals, with sensorineural HI contributing to the HI in the majority of the patients. None of the patients were from consanguineous families but seven of the patients were from endogamous populations.

Putative protein and function changing variants were discovered at a population level in *VTN, RPL3L3, FOXD4L6* and *DHRS4L2* in the merged patient and control population. These population specific genes have never been associated with HI before. Variants putatively associated to the HI were found in 11 of the 18 patients were identified in *MYO3A, MYO15A, USH2A, COL9A3, MYO1A* and *HS17B4*, with some individuals possessing more than one of the putative causative variants; these variants are all novel, but unconfirmed by Sanger sequencing and BAM files analysis.

Population structure was analysed using principal component analysis that revealed some homogeneity, as the PCA indicated that the patient and control populations originate from one population group when compared to other African populations and each other. Further analysis determined the presence of genetic difference in 44 genes; among them, 25 genes, including *VTN* and *RPL3L,* displayed significant genetic difference between the patients and control populations.

The newly identified population specific genes variants were incorporated into a list of known HI associated genes and analysis of the interaction in their protein products was performed. *VTN* and *RPL3L* were shown to interact with 42 other proteins, with RPL3l interacting with 2 proteins and VTN interacting with 40. The interactions were shown to be significantly associated with osteoclast differentiation and the bulk of the molecular activity included binding and catalytic activity.

The protein network was reconstituted to incorporate the candidate population specific variants, and this indicated that *VTN* was a hub protein of one of the eight protein networks, where this role is shared with *ATP2B2*. The protein network is centred around *UBC* and it is significantly associated with focal adhesion and regulation of the actin cytoskeleton. The bulk

of molecular activity associated with the network is binding activity and catalytic activity as well.

Chapter 4: Discussion and perspectives

Chapter 4.1: Originality of study and general comments

HI is a common sensory disorder affecting 360 million people worldwide, of which the highest incidence of HI is within developing countries such as the countries in Africa (WHO Media Centre, 2017). Despite the high incidence of HI within African countries, few studies have been performed on African populations and the prevalent genes associated HI within African populations still needs to be determined (Lebeko et al., 2017, 2016; Trotta et al., 2011). Furthermore, the majority of studies focus on the prevalence of the connexin genes, *GJB2* and *GJB6*, which have been repeatedly shown to be of negligible significance within the African context (Gasmelseed et al., 2004; Javidnia et al., 2014; Kabahuma et al., 2011; Lasisi et al., 2014; Trotta et al., 2011).

This study aspired to remedy the deficit by attempting to determine the prevalent genes associated with HI in a Cameroonian population. To the best of our knowledge, this is the first study to use whole exome sequencing to study HI within an African population. The study has also identified specific population protein changing variants in *VTN, DHRS4L2, FOXD4L8,* and *RPL3L*, which may have possible modifier action in the HI phenotype, owing the critical hub roles in their protein-protein in interactions with multiple known HI genes. These findings are novel and deserve to be explored in other African populations. The association between *VTN, RPL3, FOXD4L6,* and *DHRS4L2* and HI was, furthermore, studied in terms of its impact on the HI network and what role each of the protein products may play in the network. This information and the interaction of the protein products with other HI associated genes is a step forward in determining the action of these proteins in HI.

Exceptionally, in the present study, further analysis allowed the determination of the proportion of derived and ancestral alleles within the Cameroonian population. This analysis is the first report to our knowledge, to explore such approach in hearing loss. Data indicates the hearing-impaired patient population carried a higher proportion of derived alleles in know HI genes, at low minor allele frequencies; possibly indicating, the interactive modifiers capacities of multiple HI genes, or alternatively, the polygenetic nature of HI in some patients. The proportion of ancestral and derived alleles was contrasted in the Cameroonian and the population from the Democratic Republic of Congo and it indicated that the variations that may result in HI in the one population may not be the same variations that result in HI in the other population. Due to this, it is necessary to determine the causative variants resulting in disease

in multiple populations independently, to capture the full picture of genetic heterogeneity of hearing loss among Africans.

Variants discovered in *COL9A3, MYO3A* and *MYO15A* have yet to be described in literature, and, cumulatively, putatively could have account for HI in 9 of the 18 patients, but were not confirmed by Sanger sequencing, nor the BAMs file analysis. This may reflect the current difficulty of inferring variant deleteriousness in African population, whose exomes are not yet well represented in Exome databases; a call for corrective action to global scientific community to remedy this.

.

Chapter 4.2: Candidate clinical variants

Putative causative mutations were discovered in *MYO3A, MYO15A, USH2A, COL9A3, MYO1A* and *HSD17B4*. The six genes have been previously associated with HI, with the *USH2A*, *COL9A3* and *HSD17B4* been associated with syndromic HI.

USH2A (MIM:276901), encoded at position 1q41, is associated with Usher Syndrome Type IIA, which is congenital, mild HI and prepubertal onset of vision loss due to retinitis pigmentosa (Eudy et al., 1998; Kimberling et al., 1990), as indicated in Chapter 1.5.1.1. The phenotype associated with this gene suggests that the gene may not be the causative mutation within the patients that carry the variation, as the patients have non-syndromic HI rather than syndromic HI. Furthermore, the variation, rs111033481, is indicated to be benign in the ClinVar database (Landrum et al., 2016). For these particular patients, further analysis will be necessary to identify the variation that results in the HI.

HSD17B4 (MIM:601860), encoded at position 5q23.1, is associated with Perrault syndrome (MIM:233400) and D-bifunctional protein deficiency (MIM:261515). Perrault syndrome was initially described as Turner Syndrome presenting with deaf-mutism by Perrault, Klotz & Housset (1950). The syndrome presents as sensorineural HI in both males and females, accompanied by premature ovarian failure in females (Jenkinson et al., 2013; Pierce et al., 2010). Additional clinical features may include cerebellar ataxia, peripheral neuropathy and learning difficulties (Jenkinson et al., 2013) though the syndrome is clinically and genetically heterogeneous as it may result from variations in several genes (Chatzispyrou et al., 2017; Jenkinson et al., 2013; Pierce et al., 2013, 2010). Conversely, the gene is associated with D-bifunctional protein deficiency (MIM:261515) which presents with a severe phenotype, that

includes early onset hypotonia and seizures and facial abnormalities (Grønborg et al., 2010; Suzuki et al., 1994; Watkins et al., 1989). The severe phenotype associated with pathogenic variations in this gene is not demonstrated within the Cameroonian patient population, whereby all patients were evaluated by a qualified medical geneticist and they presented with no additional clinical features above the HI, and as such the possibility of this gene being the causative mutation for the HI within the patients who carry the variation is highly unlikely and further analysis will be necessary to determine the causative mutations in these patients.

MYO1A, encoded on 12q13.3, has been previously associated autosomal dominant HI, indicated as *DFNA48*. The gene was associated with autosomal dominant HI in an Italian cohort (D'Adamo et al., 2003; Donaudy et al., 2003). The role of *MYO1A* in HI was challenged by Eisenberger et al. (2014) who indicated that other variations in other genes are responsible for HI in purported *DFNA48* cases of HI. As such, the variation in *MYO1A* was not considered to be the cause of HI in the patients who carry the variation and further investigation is required to resolve the cause of HI in those patients.

The variations in *MYO1A, USH2A* and *HSD17B4* may be explained away as the cause of HI in the patients due to abovementioned literature which indicates the implausibility of these variants as causative mutations in the patients. This is especially of note in *HSD17B4* and *USH2A,* granting that the benign nature of the variation in *USH2A* is not considered, which are associated with syndromic HI. The patient population was specifically selected for non-syndromic HI which means it is highly improbable that these genes are the causative mutations in the implicated patients. The variations in *COL9A3*, *MYO3A*, and *MYO15A* require further consideration.

Chapter 4.2.1: Collagen type IX alpha 3 chain.

COL9A3, encoded at position 20q13.33, is associated with Stickler Syndrome, which is mixed HI associated with myopia, cataracts, and retinal detachment, as well as facial abnormalities and precocious arthritis among other clinical manifestations (Robin et al., 2017). The syndrome is currently associated with six collagen genes, *COL2A1, COL11A1, COL11A2, COL9A1, COL9A2,* and *COL9A3,* though it is possible that other genes may also be responsible for the syndrome (Robin et al., 2017). The syndrome is inherited in an autosomal dominant pattern when the causative mutations are in the *COL2A1, COL11A1* or *COL11A2* genes. Conversely, the syndrome is inherited in a autosomal recessive pattern if the causative mutations are in the *COL9A1, COL9A2* or *COL9A3* genes (Robin et al., 2017).

COL9A3 forms one of the three alpha chains of Type IX Collagen and it is expressed in the inner ear and localised on the tectorial membrane (Abe et al., 2003). The HI associated with *COL9A3* is due to *COL9A3* and the Type IX being pertinent to the shape of the tectorial membrane and essential for the 3D structure of Type II collagen (Asamura et al., 2005), which is also associated with HI.

The variation identified in *COL9A3* has yet to be described in literature, the absence in clinical information accompanying the variant on rs145035835 on ClinVar and its rarity in both the Cameroonian patient and control population would have made it an attractive variant if it were validated by Sanger Sequencing.

Chapter 4.2.2: *Myosin 3A*

MYO3A, encoded at position 10p12.1, and was first reported to be associated with progressive, autosomal recessive, non-syndromic HI, previously identified as DFNB30 locus (MIM:60711 and MIM:606808), by Walsh et al. (2002) in a multi-generational, Israeli family, originating from an endogamous population, indicated the Figure 4.3.1. The family was shown to carry three recessive *MYO3A* mutations that segregated with the family and resulted in HI in homozygotes and compound heterozygotes, where carriers were unaffected (Walsh et al., 2002). Studies have indicated that *MYO3A* is responsible for the elongation of stereocilia by transporting *espin 1* to the plus ends actin filaments in hair cells in cell culture (Salles et al., 2009). Furthermore, that aberration in *MYO3A,* in mice models, results in fewer outer hair cells that indicate degeneration by the tenth month of life, which explains the progressive nature of the HI (Walsh et al., 2011).

MYO3A has been associated with autosomal dominant, progressive HI in an African American family (Grati et al., 2016); the first study to do so. The study indicates that the mutation results in an aberration of the protein's ATPase activity which the authors postulate may affect the motor-walking capabilities of the protein (Grati et al., 2016). This finding is also one of the first documented incidences where *MYO3A* was associated with a family of African descent and this variant is present in the heterozygous state in nine out of 121 064 alleles (Grati et al., 2016), of which seven were in individuals of African ancestry and no homozygotes were observed. Though the *MYO3A* variant was not confirmed by Sanger Sequencing in the Cameroonian population, the study by Grati et al. (2016) may indicate that *MYO3A* may be an important gene in African non-syndromic HI and future studies ought to make allowances to study it.

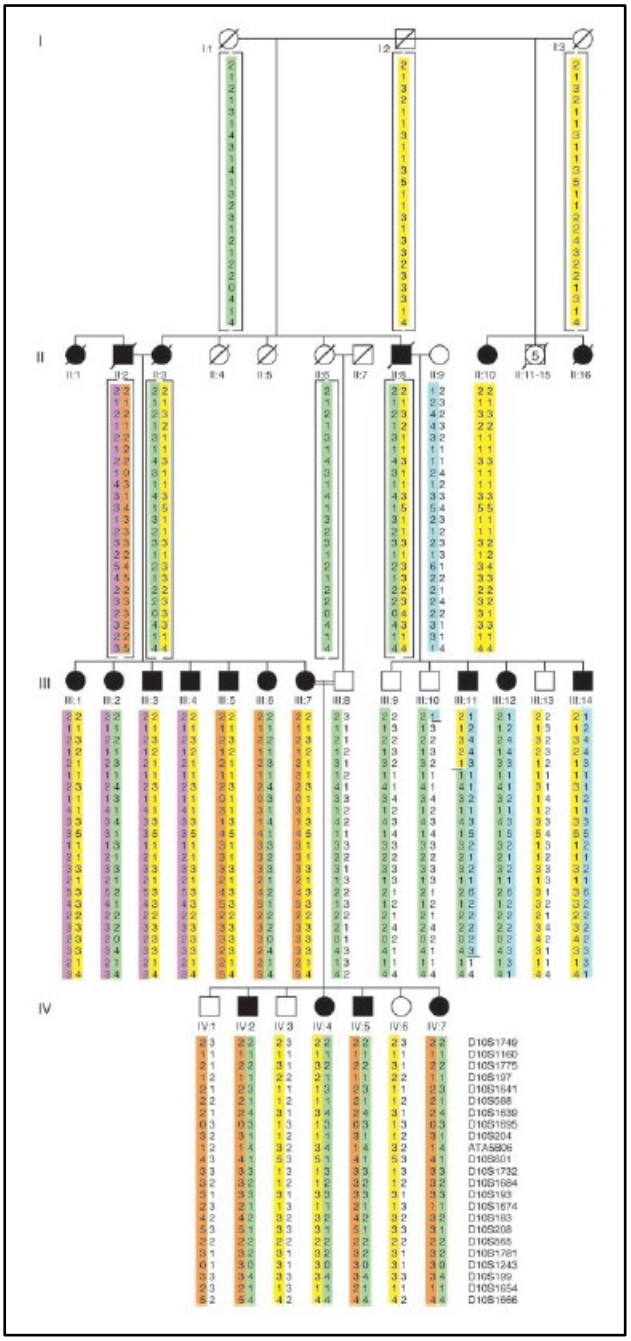


Figure 4.2.2.1: Multi-generational Israeli family segregating autosomal-recessive non-syndromic hearing impairment due to *MYO3A* mutations (adapted from Walsh et al., 2002). Proband family in which *MYO3A* was identified as the gene responsible for DFNB30. Family segregates three *MYO3A* mutations that result in hearing impairment in homozygotes and compound heterozygotes carrying two different mutations. Female members of the family are indicated by circles and male members of the family are indicated by squares. Affected individuals are indicated by filled shapes and unaffected individuals are blank.

Chapter 4.2.3: *Myosin 15A*

MYO15A, encoded at position 17p11.2, was initially associated with the *DFNB3* HI locus in the individuals from the Bengkala and , neighbouring, Bali villages in Indonesia (Friedman et al., 1995; Winata et al., 1995), with the Bengkala families indicated in Figure 4.2.2.1. *MYO15A* was identified as the DFNB3 locus in 1998 through fine mapping (Liang et al., 1998; Probst et al., 1998; Wang et al., 1998). Probst et al. (1998) indicated that the HI was as a result of homozygous pathogenic mutations shortening the stereocilia bundles and disrupting the arrangement of the cell cytoskeleton in mice models. Further analysis identified that *MYO15A*

is expressed as 2 isoforms in hair cells (Anderson et al., 2000; Belyantseva et al., 2005; Liang et al., 1999), whereby isoform 1 includes exon 2 of the *MYO15A*. gene which encodes a longer N-terminal domain and isoform 2 excludes exon 2 and the subsequent elongated N-terminal domain (Belyantseva et al., 2003; Liang et al., 1999), this is indicated in Figure 4.2.2.2.

Isoform 2 is necessary for the early development of the stereocilia bundles ,including stereocilia elongation and the arrangement of the stereocilia bundles (Fang et al., 2015). Mutations in this isoform shortened stereocilia result in shortened stereocilia and slight perturbation of the stereocilia architecture being observed in mice models (Anderson et al., 2000; Probst et al., 1998). The longer isoform of MYO15A is necessary for maintenance of the bundles (Belyantseva et al., 2005) and mutations that allow for the normal isoform 2, whilst disrupting isoform 1, result in normal stereocilia length and architecture at birth in mice models (Fang et al., 2015). The stereocilia, however, degenerate following 14 weeks after birth which coincides with when isoform 1 expression increases and isoform 2 expression decreases in mice models (Fang et al., 2015).

The pathophysiology of these mutations differs in the sense that mutations that selectively target the elongated N-terminal domain present in isoform 1 will result in a progressive HI due to postnatal degeneration of the stereocilia, as in the case of a Pakistani individual (Nal et al., 2007), on whom the pathogenic variation used in the isoform 2 mice models were based (Fang et al., 2015). Pathogenic variations that occur along the length of *MYO15A*, after the elongate N-terminal domain, will result in congenital HI due to the aberration in early development of the stereocilia bundles, as in the individuals from the Bengkala village (Friedman et al., 1995; Winata et al., 1995).

Studies out of Morocco have implicated *MYO15A* in profound congenital HI in the population (Salime et al., 2017) but no studies have linked *MYO15A* to HI in Sub-Saharan African patients.

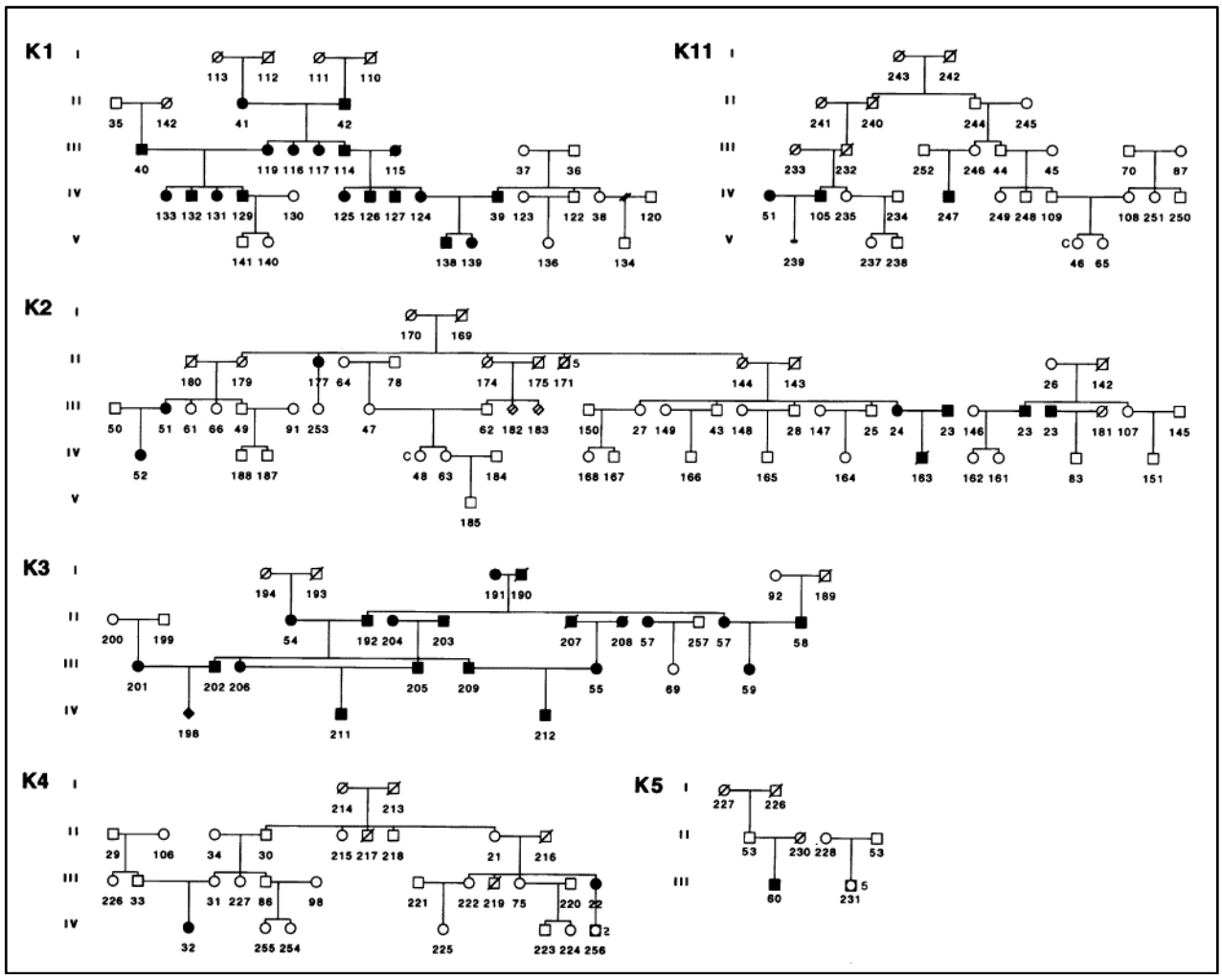


Figure 4.2.3.1: Pedigree of Bengkala families segregating *MYO15A* mutations responsible for DFNB3 non-syndromic hearing impairment (adapted from Winata et al., 1995). Pedigrees of six families segregating non-syndromic hearing impairment in the Bengkala village in Indonesia. The filled shapes indicate hearing impaired individuals, whereas unaffected individuals remain blank. Females are indicated by circles, males are indicated by squares and individuals of unspecified sex are indicated by triangles.

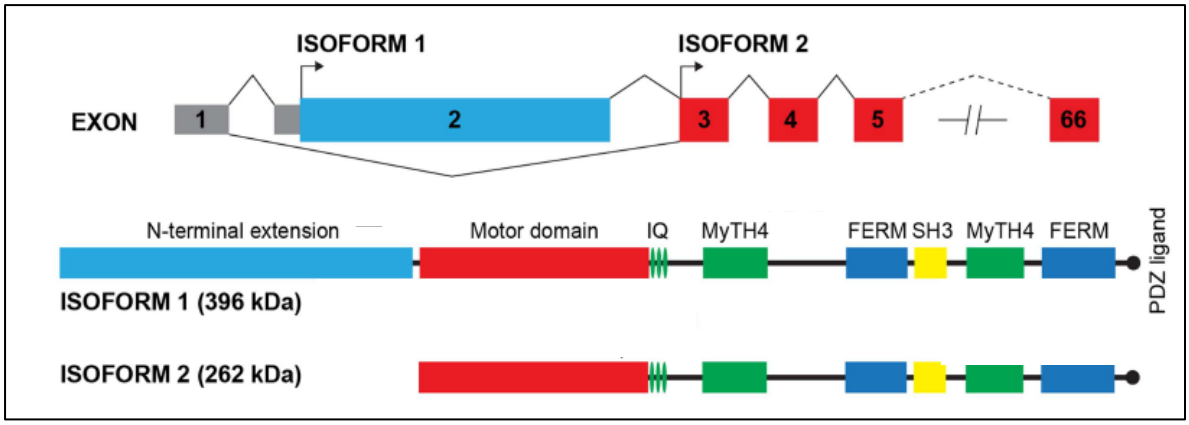


Figure 4.2.3.2: Diagram of *MYO15A* indicating Isoform 1 and Isoform 2 (adapted from Fang et al., 2015). Isoform 1 and 2 of MYO15A are alternative splice variants of *MYO15A* which are expressed at different stages of stereocilia development. Isoform 2 is initially expressed and results in early formation of stereocilia bundles. Isoform 2 is expressed later in stereocilia development and is necessary for elongation and maintenance of the stereocilia bundles.

Chapter 4.2.4: Resolution of hearing impairment

This study was capable of segregating putative causative mutations in 9 out of the 18 patients. If confirmed this could have been an improved diagnostic rate as compared to other studies utilising whole exome sequencing to study HI in African populations (Sloan-Heggen et al., 2016); the study rather confirmed the low pick-up rate of causative mutations in known genes of HI. Furthermore, it increased the need to explore more familial cases in African populations so as to facilitate the discovery of novel genes and variants that are relevant to African HI, for use in clinical practice. Indeed, there is limited literature available regarding HI in African populations that doesn't specifically focus on the connexin genes. Several studies have investigated the prevalence of the concern genes in African populations (Gasmelseed et al., 2004; Javidnia et al., 2014; Kabahuma et al., 2011; Lasisi et al., 2014; Trotta et al., 2011), but there has been limited research into the other genetic causes of HI in African populations.

Chapter 4.3: Candidate population specific variants

Filtering variants from the WES data of entire Cameroonian population resulted in four variants being identified in four different genes that were population specific, as indicated in Table 7. These variants are not causative mutations of HI in the population and were discovered within the entire Cameroonian cohort, patients and controls, whereby the variants would only be

filtered if they were present in each individual in the merged data file. The variations also needed to be deemed pathogenic by 10 of the 15 filtering algorithms in order to be considered candidate variants, which makes these four protein changing variants filtered highly attractive variants of interest.

Chapter 4.3.1: *Forkhead Box D4 Like*

FOXD4L6, encoded at position 9p11.2, is one of six duplications of the *FOXD4* gene (Jackson et al., 2010). The *FOX* gene family encodes transcription factors (Jackson et al., 2010), whereby the first *FOX* gene was discovered in *Drosophila melanogaster* as *fkh* (Weigel et al., 1989). *fkh* was shown to be restricted in the nucleus (Weigel et al., 1989) and it contains an approximately 100 to 110 amino acid DNA binding domain conserved within members of this family (Fan et al., 2002; Jackson et al., 2010; Lai et al., 1991).The gene family has been shown to be implicated in embryonic development (Ruiz et al., 1993) and they have been shown to be oncogenes (Li and Vogt, 1993) and they have been implicated in various cancers (Katoh et al., 2013; Zhang et al., 2014). Despite the suggested importance of the *FOX* genes, the functional roles of many of these genes is unknown (Jackson et al., 2010; Zhang et al., 2014) and this notion extends to *FOXD4L6.*

The segregation of the *FOXD4L6* variant with the Cameroonian population makes it of interest, as it may be associated with the HI phenotype in the population, though it does not actively cause HI in and of itself. The lack of literature/information regarding its exact function, however, renders the justification of the possible association rather complex as the possible action of *FOXD4L6* is completely unknown. *FOXD4L6* will need to be extensively studied within the population and its expression and possible function in the inner ear will need to be determined. The possibility does exist that it interacts with other HI associated genes, as it is a transcription factor, but to what extent would need to be determined.

Chapter 4.3.2: *Short chain dehydrogenase/reductase member 4 like 2*

DHRS4L2, encoded at position 14q11.2, as part of the *DHRS4* gene cluster (Gabrielli and Tofanelli, 2012). The gene arose from duplication of *DHRS4* and forms the gene cluster with *DHRS4* and *DHRS4L1* (Gabrielli and Tofanelli, 2012). *DHRS4L2* has an alternative transcription start site in an exon prior to its exon 1 and the gene may be alternatively splice to produce alternative two transcripts (Zhang et al., 2009); denoted exon Ea in Figure 4.3.2.1a. The first alternative transcript of *DHRS4L2* does not contain exons 1, 4, 5 and 6 and the second alternative transcript of *DHRS4L2* is missing the above mentioned exons as well as exon 3

(Zhang et al., 2009), as indicated in Figure 4.3.2.1b. *DHRS4L2* has been shown to be expressed only in neuroblastoma cells and that expression is limited to transcription of the DNA into RNA (Zhang et al., 2009), whereby the RNA product is stable and may be involved in gene expression regulation (Su et al., 2010), though this needs to be verified through molecular experiments.

The absence of a known DHRS4L2 protein means that the function of the protein is unknown which makes its association with the Cameroonian population difficult to conceptualise. The possibility exists that *DHRS4L2* is involved in regulating gene expression as indicated by Su et al. (2010) but further experimentation will be necessary to verify this. Alternatively, *DHRS4L2* may possibly encode a protein that has yet to be identified similarly to *DHRS4L1* which was identified through a novel bioinformatics approach coupled with mass spectrometry analysis (Carapito et al., 2015). Regardless of the existence of a protein product for *DHRS4L2*, analysis of its involvement within the Cameroonian population may possibly determine whether it may regulate gene expression, though it will be necessary to first establish if it is expressed in the inner ear first.

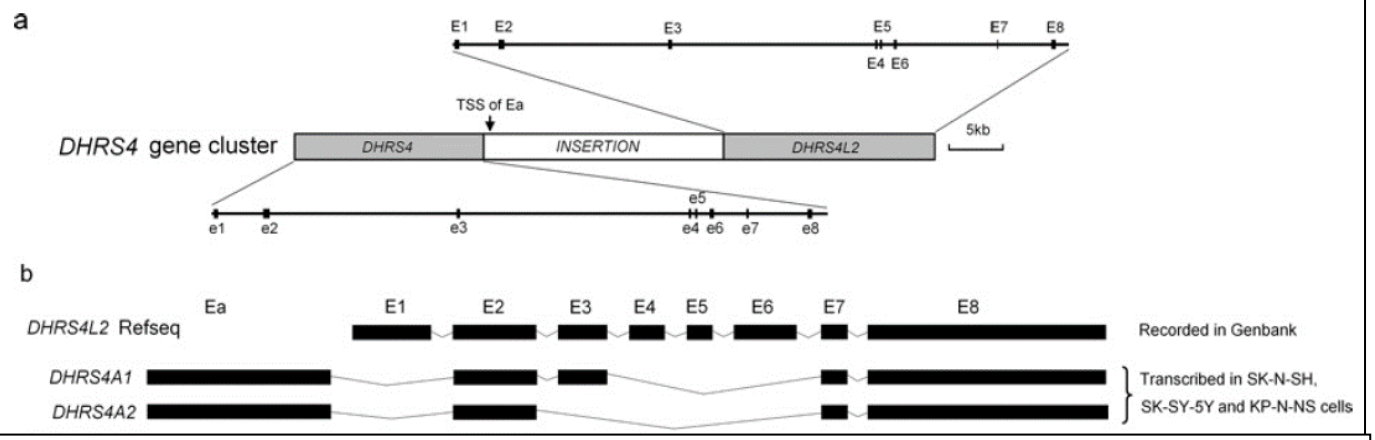


Figure 4.3.2.1: *DHRS4* gene cluster structure consisting of *DHRS4* and *DHRS4L2* (adapted from Zhang et al., 2009). Figure 3.5.2.1a indicates the structure of the DHRS4 gene cluster, including the position of the exon Ea which also possesses a transcription start site for *DHRS4L2*. Figure 5.3.2.1b indicates the alternatively spliced RNAs for DHRS4L2, which consist of exon Ea and exon exclude exons 1, 4, 5, and 6 for the long transcript and exons 1, 3, 4, 5 and 6 for the short transcript.

Chapter 4.3.3: *Ribosomal Protein L3 Like*

The significance and worth, with regards to HI, of *RPL3L*, which is encoded at position 16p13, may be disputed. RPL3L, is a structural component of large ribosomal subunit, and it has been

previously shown to be expressed at high levels in the heart, skeletal muscle and at low expression in the pancreases, brain, placenta, lung liver and kidneys (Van Raay et al., 1996). The expression pattern is contrasted by Thorrez et al. (2008) and Gupta and Warner (2014) who indicate that *RPL3L* is only expressed in muscle tissue. Instead of *RPL3L,* it is *RPL3* which is expressed in non-muscle tissue according to the 2008 study (Thorrez et al., 2008). This information results in a great deal of ambiguity regarding the plausibility of this gene playing any role in HI, including a modifier role.

A further point of contention to be noted is that *RPL3L* was part of a gene-rich interval deleted in an individual with tuberous sclerosis and acrofacial dystosis, presenting with HI (Dauwerse et al., 2002). No follow-up study has been performed, to the best of our knowledge, to refine the region and determine the gene responsible for the phenotype in this individual, and as such, further ambiguity exists regarding the role of *RPL3L* in the disease status of the individual presented by Dauwerse et al. (2002)). This ambiguity extends to the role the RPL3L may play within the Cameroonian population and future work would need to address 1) the possibility that *RPL3L* is expressed in the inner ear instead of *RPL3* and 2) the role that *RPL3L*, if any exists, plays in HI and how it functions to achieve the role.

Chapter4.3.4: *Vitronectin*

VTN, encoded at position 17q11.2, was initially sequenced by Jenne and Stanley (1985) and it is a secreted protein that exists and acts in the extracellular matrix where it is involved in cell attachment and spreading (Schvartz et al., 1999). The protein functions in the complement system where it interferes with complement-mediated cell lysis (Bhakdi and Roth, 1981; Kolb and Muller-Eberhard, 1975; Podack et al., 1977) and protects thrombin from inactivation by antithrombin III (Podack and Müller-Eberhard, 1979).

VTN was been shown to be expressed in the inner ear by Aarnisalo *et al*. (2010), which renders it a more attractive candidate than *RPL3L* in having any role in HI within the Cameroonian population. The exact site, in the inner ear, where VTN is expressed was not indicated nor was the role of VTN in the inner ear elucidated in the study. Interestingly, *VTN* deposits have been shown to be associated with fibrosis and necrosis in a variety of diseases (Schvartz et al., 1999), though mice models, with null mutations in either one or both copies of *VTN*, have been shown to be healthy (Zheng et al., 1995). This may indicate that mutations in *VTN* in and of themselves may not result in disease, but they may affect the phenotype associated with the disease. The effect on the disease phenotype may be limited to the presence of *VTN* deposits and this would need to be further investigated to draw any meaningful conclusions.

Furthermore, taking this information into account, the presence of deleterious mutations in *VTN* in the Cameroonian population will need to be explored further. *VTN* may not necessarily be pathogenic, but it could possibly be a modifier of the HI phenotype observed within the Cameroonian population. If this be the case, the study of *VTN* with respect to HI may assist in understanding the variability sometimes observed within individuals who carry the same HI mutations.

Chapter 4.3.5: Bias

The population specific variants and the clinical variants discovered were due to mapping and comparison with the hg19 reference genome (Meyer et al., 2013) available from the UCSC Genome Browser. Though the variants were annotated and filtered according to various tools that predict pathogenicity, the possibility of false positives may not be ignored. The proportion of ethnic minorities whose genomes have contributed to developing the reference genome is not stated and no data is available on this percentage. There is, however, a persistent bias in genome wide association studies whereby the majority of studies are performed in European populations (Popejoy and Fullerton, 2016) and African populations are not well represented in exome databases (Lek et al., 2016; Manrai et al., 2016) and this may possibly extend to the genome assembly. This would result in polymorphisms been labelled as pathogenic simply due to their difference from the established normal. As such, increase in the African Exome proportion in datasets should improve inference of variants deleteriousness (Lek et al., 2016; Manrai et al., 2016); as well as subsequent functional analysis of the variants discovered is imperative to establish their pathogenicity.

Chapter 4.4: Population structure and genetic differentiation

Population structure analysis was performed through the use of two principle component analyses. Figure 4.3.1 indicates that both the patient and control population originate from one population confirming that the control population is sufficient to contrast any discoveries made within the patient population against. The figure indicates that the DRC population clusters with the Cameroonian population. This may simply be due to the small sample size used for the DRC population whereby only samples were used in the PCA from the DRC. A second PCA will need to be run with a greater sample population representation from the DRC in order to verify this hypothesis.

Figure 4.3.2 is the second PCA which compares the Cameroonian hearing impaired population to the unaffected control population. Neither population indicates any clustering and the samples are dispersed freely within the PCA indicating that no population stratification that may need to be corrected for. This PCA also verifies that the two sample populations originate from one population and that the control population is an adequate control for the downstream analysis.

A novel approach analysed the genetic difference between hearing impaired population and control population. It aggregated the SNP frequencies in each gene and compared the aggregate between the populations. This approach allowed for the evaluation of possible difference between the populations which was summarised in Table 9. This data indicates a difference in aggregate frequencies between the hearing-impaired population and control population with regard to *VTN* and *RPL3L*; which both show unusual genetic differentiation. The novelty of this approach means there is no literature to compare to, but this approach adds evidence in the hypothesis that *VTN* and *RPL3L* are noteworthy genes regarding HI in the Cameroonian population. It thus supports the notion to further investigate the possible involvement of *VTN* and *RPL3L* within the HI context. Interesting, no data was produced for *FOXD4L6* and *DHRS4L2*.

Chapter 4.4: Protein-protein interactions, enrichment analysis and community network analysis

Protein-protein interactions were analysed to determine what proteins VTN, RPL3L, FOXD4L6 and DHRS4L2 interact with as they have never been associated with HI before. VTN and RPL3L were shown to interact with 42 other proteins; with FOXD4L6 and DHRS4L2 shown to interact with no other proteins; and the interactions were illustrated in Figure 4.4.1. The interactions show significant association with the osteoclast differentiation pathway. Aberrations in osteoclast differentiation has been shown to be associated with HI in mice (Akil et al., 2014; Kanzaki et al., 2006) and HI in osteogenesis imperfecta (Kuurila et al., 2002; Paterson et al., 2001; Pedersen, 1984; Pillion et al., 2014).

The protein products of *VTN, RPL3, DHRS4L2* and *FOXD4L2* were combined to produce a network plot. This network plot indicated VTN and ATPB2 as the hub proteins of one of the 8 subnetworks generated. This involvement of VTN as hub protein supports its notion as having a contribution in HI within our population. ATPB2, as a second hub protein in the subnetwork, may rescue the subnetwork should VTN possess deleterious variations and this may explain

why mice with null mutations in VTN are healthy (Zheng et al., 1995). This will need to be explored further as a similar study by Lebeko et al. (2017) indicate that the hub protein of this subnetwork was DIAPH1, which is shown in this study to be a subsidiary protein. The difference in designated hub protein may be due to the genes used whereby this study compared 159 genes whereas Lebeko et al. (2017) utilised 116 genes.

This study and Lebeko et al. (2017), do indicate differences in the hub proteins identified; but a simplistic conclusion could be that all the hub proteins identified are of importance in HI and should be studied further in order to elucidate how they fit into the HI network. This study will be greatly assisted if further genes associated with HI are discovered and incorporated into the network such that a fuller picture of the HI network is established.

The absence of *DHRS4L2* and *FOXD4L6* protein products, in Figure 4.4.1, may be due to the non-existence of the protein in the case of the former only or due to there being no annotation that indicates which proteins these particular proteins interact with. Studies will need to be performed which may allow for the exclusion of both genes as associated with HI, even in a modifier role, in the population, but the proposed gene expression regulatory role of the RNA products of DHRS4L2 and the transcription factor role of *FOXD4L6* make both genes and the products interesting subjects for future studies.

Chapter 4.6: Validation

Cycle sequencing indicated the absence of the *MYO3A* and the *MYO15A* variations in all the patients purported to carry the variation. This absence of the variation could simply be that variations are absent in all the three *MYO3A* patients and in all the four *MYO15A* patients, but this seems highly unlikely. The possibility exists that the overall negative results may be due to errors in the sequencing; which may have arisen due to incorrect primers being utilised and would be concordant with results observed by Beck et al., (2016). The discrepancy in the next generation sequencing results vs the Sanger Sequencing results was resolved for 17 of the 19 SNPs (Beck et al., 2016).

A valid approach to assess the discrepancy between Sanger Sequencing and the variant call by next generation sequencing would be to perform restriction fragment length polymorphisms for the rs140301218, in *MYO3A*, and rs138861831, in *MYO15A*. This approach is, however, not viable due to the absence of restriction sites surrounding the variants as indicated by Webcutter 2.0 (http://rna.lundberg.gu.se/cutter2/).

A more viable approach was to view the BAM files of the patients using the Integrative Genomics Viewer, or IGV (Robinson et al., 2011; Thorvaldsdóttir et al., 2013). This approach allows the researcher to visualise each variant call and the associated read depth for the variant. As such, viewing the BAM files using IGV supports the results from Sanger Sequencing for *MYO3A,* indicating that the variants observed in Bioinformatics analysis were incorrect and the patients with the purported to carry the causative mutations. This approach further indicated the absence of the variant in *COL9A3* in the patients that were purported to carry the variant and as such, the resolution rate regarding putative causative mutations in this population is 0%. The approach furthermore resolved the discrepancy in the Sanger Sequencing results for patient CAM103 by providing supporting evidence for the sequencing result indicated in 4.8.3B.

Chapter 4.7: Evolutionary adaptation of human hearing

Analysis of the fraction of ancestral alleles in the population, indicated in Figure 4.7.1, yielded expected results with regards to the proportion of ancestral and derived alleles in the patient and control populations. The hearing-impaired patient population presented with a higher proportion of derived alleles at low population frequencies (the 0.00 to 0.10 MAF bins). The proportion of the derived alleles decreased with increasing population frequencies, such that there was a higher proportion of ancestral alleles at population frequencies indicative of common alleles. This result was similar to the results of a study by Gorlova et al. (2012), which found a higher proportion of derived alleles at low population frequencies and this proportion (of derived alleles) decreases with an increase in the population frequencies, and they further showed that derived alleles are associated with disease risk (Gorlova et al., 2012).

Analysis further contrasted the proportion of ancestral alleles in the Cameroonian Population and a Population from the Democratic Republic of Congo, indicated in Figure 4.7.2. Analysis showed variations in the two populations with regards to the proportion of ancestral alleles in various HI associated genes in the two populations. This may indicate that the subset of genes responsible for HI within one population cannot be juxtaposed on the other population and that independent research into the genetic causes of HI must be done within each population. This analysis should, however, be taken with a grain of salt due to the small sample size of the DRC population, which may explain the large spikes observed in the graph.

A point of interest is that the proportion of derived alleles in *GJB2* indicates that it may be of importance to HI within both the Cameroonian population and the Democratic Republic of Congo population and, maybe, possibly in a greater African context, as indicated in Figure

4.7.2. This contrasts published literature that has shown that *GJB2* mutations are not prevalent within studied African populations (Bosch et al., 2014b; Gasmelseed et al., 2004; Javidnia et al., 2014; Kabahuma et al., 2011; Trotta et al., 2011), excepting in the Ghanaian (Hamelmann et al., 2001) and Moroccan (Gazzaz et al., 2005) populations, and the probable cause of this (the high proportion of derived alleles) may be due to *GJB2* being a hypervariable gene.

The novelty of this analysis within the African context and within African HI means there is very little literature to compare these findings to; despite this, the results from this analysis may be compared to the candidate variants discovered whereby *COL9A3* was indicated to have a low proportion of ancestral alleles, and therefore a high proportion of derived alleles, in Figure 4.7.2. This may possibly affirm its association as a putative causative mutation in HI in the individuals that carry this variation. A similar statement cannot necessarily be made about *USH2A*, which also indicates a low proportion of ancestral alleles, and thus a high proportion of derived alleles, simply because *USH2A* is solely associated with syndromic HI and none of the individuals who harboured the *USH2A* variations segregated syndromic HI.

The result, in Figure 4.7.2, may also indicate genes of interest in Cameroonian population. The worth of this information will need to be evaluated by determining the minor allele frequencies where these variations occur and establishing whether the gene, similar to *GJB2*, is hypervariable or not.

Chapter 4.8: Limitations of study

The limitations of this study were the small sample size for the hearing-impaired population whereby the whole exome sequences from 18 individuals were analysed. This means that though cumulatively, if confirmed, the variations in *COL9A3, MYO3A* and *MYO15A* could have been putative causes of HI in 50% of the studied hearing-impaired population, the sample size was too small for the results to have any statistical power. This could have been remedied by screening the variations in a greater cohort of Cameroonian hearing-impaired patients, which would aid in establishing the significance of the variations discovered.

A further limitation was that no segregation analysis was performed. This would have indicated if the variations discovered segregated with the families and would have ameliorated the lack of statistical power of the study. This could have been remedied by including the parents and of affected individuals and performing trio-studies, which would indicate the segregation pattern of the variations.

A final limitation of this study, which was beyond the scope of the study, was that no functional studies or protein modelling were performed in order to validate the pathogenicity of the variations discovered. The possible manner by which these variations cause HI would need to be studied through molecular analysis in order to conclude that the variations are indeed the causative mutations within the those who carry the variations in homozygosity.

Chapter 4.9: Practical implications of the study

The results support published literature indicating the low-resolution rate of the causative genes resulting in HI in African populations. The results further indicate that novel strategies may be necessary to discover causative variants resulting in HI in African populations, specifically the studies of multiple families segregating HI, from various settings and African populations, to discovery novel HI genes and novel variants in Known genes, that will be useful in clinical practice.

Chapter 4.10: Perspectives and research recommendations

Future work should focus on targeting familial cases of HI in multiple African populations in order to determine the prevalent genes associated with HI within African populations; as by determining these genes, it may be possible to create a diagnostic panel that may be used in the clinical settings to determine the cause of HI in patients with readily available results. This must however be augmented with whole genome sequencing and transcriptomic analysis; as the causative mutations may not necessarily be resulting in an aberration in the exonic or protein sequences, but in the expression of the genes. The errors that result in HI may be due to gene silencing which should be further explored.

A second avenue of future work should also consider mutations that affect the genomic architecture that may result in HI. This should include copy number variations which has not necessarily been explored in African populations.

Finally, a larger patient population is needed for further studies. The increased population size may result in better resolution of causative mutations resulting in HI in the population as well as lending statistical power to any results obtained. The increased population size may be recruited by collaborating with other researchers on the African continent in order to recruit patients from other African populations.

Chapter 5: Conclusions and future directions

The study aimed to determine known genes that may be prevalently associated with HI in Sub-Saharan, and specifically Cameroonian, patients. This study demonstrates the low pick up rate of putative pathogenic variant in known genes in this group of patients. The study also showed the efficacy novel approaches, in the bioinformatics analysis of whole exome sequencing data, in annotating genes variants according to their putative protein-protein interactions and network involvement in HI. This provides a novel insight into possible functions that the genes may adapt with regard to HI. The value of bioinformatics research is demonstrated in its use in indicating proteins of note which may have indirect effects on HI in the form of the hub proteins and the protein products of genes indicating unusual genetic difference between the hearing impaired and control populations. The proteins may not directly result in HI but perturbations within them may have downstream effects on the observed phenotype and this needs to be taken into account. Advancement of bioinformatics tools and methods allow for the in-silico prediction of protein interactions and incorporation of novel proteins into a protein product, but this must be substantiated with molecular analysis. The bioinformatics analysis provides an invaluable insight into the direction the molecular research ought to follow, but these are still predictions. That being said, the advancement of bioinformatics tools and the simultaneous incorporation of several tools in analysis is instrumental in generating new knowledge, especially in a heterogenous field such as genetic HI.

The study, further, indicated that whole exome sequencing results should be viewed cautiously; perhaps more so when studying African exomes/genomes, as some putative causative variants identified in whole exome sequencing will not confirmed by other techniques such as Sanger Sequencing. Whole exome sequencing resulting should be validated prior in order to ensure that the filtered variants are present in the patients. Furthermore, logical annotation of all candidate variants should be done whilst taking into account published literature regarding the involvement of the gene products in disease. This may however be complicated if there is a dearth in published literature regarding specific genes and/or their protein products.

Lastly the study has confirmed the diagnostic limitation of using WES for known HI genes among this group of Africans, and therefore emphasizes the urgent need to study large numbers of African families with HI in order to determine the relevant novel genes and variants that will be useful in clinical practice.

References

1. Aarnisalo, A.A., Green, K.M., O'Malley, J., Makary, C., Adams, J., Merchant, S.N., Evans, J.E., 2010. A method for MSE differential proteomic analysis of archival formalin-fixed celloidin-embedded human inner ear tissue. Hear. Res. 270, 15–20.
2. Abdelhak, S., Kalatzis, V., Heilig, R., Compain, S., Samson, D., Vincent, C., Weil, D., Cruaud, C., Sahly, I., Leibovici, M., Bitner-Glindzicz, M., Francis, M., Lacombe, D., Vigneron, J., Charachon, R., Boven, K., Bedbeder, P., Regemorter, N. Van, Weissenbach, J., Petit, C., 1997. A human homologue of the Drosophila eyes absent gene underlies Branchio-Oto-Renal (BOR) syndrome and identifies a novel gene family. Nat Genet 15, 157–164.
3. Abe, S., Katagiri, T., Saito-Hisaminato, A., Usami, S., Inoue, Y., Tsunoda, T., Nakamura, Y., 2003. Identification of CRYM as a Candidate Responsible for Nonsyndromic Deafness , through cDNA Microarray Analysis of Human Cochlear and Vestibular Tissues. Am J Hum Genet 72, 73–82.
4. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249.
5. Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez-Banet, J., Billis, K., Garcia-Giron, C., Hourlier, T., Howe, K.L., Kahari, A.K., Kokocinski, F., Martin, F.J., Murphy, D.N., Nag, R., Ruffier, M., Schuster, M., Tang, Y.A., Vogel, J.-H., White, S., Zadissa, A., Flicek, P., Searle, S.M.J., 2016. The Ensembl Gene Annotation System. Database (Oxford). 2016, baw093.
6. Akil, O., Hall-Glenn, F., Chang, J., Li, A., Chang, W., Lustig, L.R., Alliston, T., Hsiao, E.C., 2014. Disrupted bone remodeling leads to cochlear overgrowth and hearing loss in a mouse model of fibrous dysplasia. PLoS One 9.
7. Anderson, D.W., Probst, F.J., Belyantseva, I. a, Fridell, R. a, Beyer, L., Martin, D.M., Wu, D., Kachar, B., Friedman, T.B., Raphael, Y., Camper, S. a, 2000. The motor and tail regions of myosin XV are critical for normal structure and function of auditory and vestibular hair cells. Hum. Mol. Genet. 9, 1729–1738.
8. Asamura, K., Abe, S., Imamura, Y., Aszodi, A., Suzuki, N., Hashimoto, S., Takumi, Y., Hayashi, T., Fässler, R., Nakamura, Y., Usami, S., 2005. Type IX collagen is crucial for normal hearing. Neuroscience 132, 493–500.
9. Asherson, F.M., 1834. De fistulis collii congenitalis. Phobus Preuss ver. Zeitung,

Berlin.

10. Attié, T., Till, M., Pelet, A., Amiel, J., Edery, P., Boutrand, L., Munnich, A., Lyonnet, S., 1995. Mutation of the endothelin-receptor B gene in Waardenburg-Hirschsprung disease. Hum. Mol. Genet. 4, 2407–2409.
11. Azaiez, H., Booth, K.T., Bu, F., Huygen, P., Shibata, S.B., Shearer, A.E., Kolbe, D., Meyer, N., Black-Ziegelbein, E.A., Smith, R.J.H., 2014. TBC1D24 Mutation Causes Autosomal-Dominant Nonsyndromic Hearing Loss. Hum. Mutat. 35, 819–823.
12. Beck, T.F., Mullikin, J.C., Biesecker, L.G., 2016. Systematic evaluation of sanger validation of next-generation sequencing variants. Clin. Chem. 62, 647–654.
13. Belyantseva, I.A., Boger, E.T., Naz, S., Frolenkov, G.I., Sellers, J.R., Ahmed, Z.M., Griffith, A.J., Friedman, T.B., 2005. Myosin-XVa is required for tip localization of whirlin and differential elongation of hair-cell stereocilia. Nat. Cell Biol. 7, 148–156.
14. Belyantseva, I. a, Boger, E.T., Friedman, T.B., 2003. Myosin XVa localizes to the tips of inner ear sensory cell stereocilia and is essential for staircase formation of the hair bundle. Proc. Natl. Acad. Sci. U. S. A. 100, 13958–63.
15. Bhakdi, S., Roth, M., 1981. Fluid-phase SC5b-8 complex of human complement: generation and isolation from serum. J. Immunol. 127, 576–570.
16. Boëda, B., El-Amraoui, A., Bahloul, A., Goodyear, R., Daviet, L., Blanchard, S., Perfettini, I., Fath, K.R., Shorte, S., Reiners, J., Houdusse, A., Legrain, P., Wolfrum, U., Richardson, G., Petit, C., Blanchard, Â., Daviet, L., Perfettini, I., Fath, K.R., Shorte, S., Reiners, J., Houdusse, A., Legrain, P., Wolfrum, U., Richardson, G., 2002. Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle. EMBO J. 21, 6689–6699.
17. Bosch, J., 2013. Signifcance of connexin genes in non-syndromc deafness in Africans. University of Cape Town.
18. Bosch, J., Lebeko, K., Noubiap Nziale, J.J., Dandara, C., Makubalo, N., Wonkam, A., 2014a. In Search of Genetic Markers for Nonsyndromic Deafness in Africa: A Study in Cameroonians and Black South Africans with the GJB6 and GJA1 Candidate Genes. OMICS 18, 1–5.
19. Bosch, J., Noubiap, J.J.N., Dandara, C., Makubalo, N., Wright, G., Domelevo Entfellner, J.-B., Tiffin, N., Wonkam, A., Entfellner, J.-B.D., Tiffin, N., Wonkam, A., 2014b. Sequencing of GJB2 in Cameroonians and Black South Africans and comparison to 1000 Genomes Project Data Support Need to Revise Strategy for Discovery of Nonsyndromic Deafness Genes in Africans. OMICS 18, 1–6.

20. Brain, W.R., 1927. Heredity in simple goitre. QJM 303–319.
21. Brownstein, Z., Friedman, L.M., Shahin, H., Oron-Karni, V., Kol, N., Abu Rayyan, A., Parzefall, T., Lev, D., Shalev, S., Frydman, M., Davidov, B., Shohat, M., Rahile, M., Lieberman, S., Levy-Lahad, E., Lee, M.K., Shomron, N., King, M.-C., Walsh, T., Kanaan, M., Avraham, K.B., 2011. Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in middle eastern families. Genome Biol. 12, R89.
22. Carapito, C., Lane, L., Benama, M., Opsomer, A., Mouton-Barbosa, E., Garrigues, L., Gonzalez De Peredo, A., Burel, A., Bruley, C., Gateau, A., Bouyssié, D., Jaquinod, M., Cianferani, S., Burlet-Schiltz, O., Van Dorsselaer, A., Garin, J., Vandenbrouck, Y., 2015. Computational and mass-spectrometry-based workflow for the discovery and validation of missing human proteins: Application to chromosomes 2 and 14. J. Proteome Res. 14, 3621–3634.
23. Chan, D.K., Chang, K.W., 2014. GJB2-associated hearing loss: Systematic review of worldwide prevalence, genotype, and auditory phenotype. Laryngoscope 124, 34–53.
24. Chatzispyrou, I.A., Alders, M., Guerrero-Castillo, S., Zapata Perez, R., Haagmans, M.A., Mouchiroud, L., Koster, J., Ofman, R., Baas, F., Waterham, H.R., Spelbrink, J.N., Auwerx, J., Mannens, M.M., Houtkooper, R.H., Plomp, A.S., 2017. A homozygous missense mutation in ERAL1, encoding a mitochondrial rRNA chaperone, causes Perrault syndrome. Hum. Mol. Genet. 26, 2541–2550.
25. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.
26. Chimusa, E.R., Mbiyavanga, M., Mazandu, G.K., Mulder, N.J., 2015. AncGWAS: A post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. Bioinformatics 32, 549–556.
27. Choi, B.Y., Park, G., Gim, J., Kim, A.R., Kim, B.J., Kim, H.S., Park, J.H., Park, T., Oh, S.H., Han, K.H., Park, W.Y., 2013. Diagnostic Application of Targeted Resequencing for Familial Nonsyndromic Hearing Loss. PLoS One 8.
28. Chung, C.S., Robison, O.W., Morton, N.E., 1959. A note on deaf mutism. Ann. Hum. Genet. 23, 357–366.
29. Churbanov, A.Y., Karafet, T.M., Morozov, I. V., Mikhalskaia, V.Y., Zytsar, M. V., Bondar, A.A., Posukh, O.L., 2016. Whole Exome Sequencing Reveals Homozygous Mutations in RAI1, OTOF, and SLC26A4 Genes Associated with Nonsyndromic

Hearing Loss in Altaian Families (South Siberia). PLoS One 11, e0153841.

30. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., Sidow, A., 2005. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 15, 901–913.
31. Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485.
32. Cremers, C.W.R.J., Admiraal, R.J.C., Huygen, P.L.M., Bolder, C., Everett, L.A., Joosten, F.B.M., Green, E.D., Van Camp, G., Otten, B.J., 1998. Progressive hearing loss, hypoplasia of the cochlea and widened vestibular aqueducts are very common features in Pendred's syndrome. Int. J. Pediatr. Otorhinolaryngol. 45, 113–123.
33. Csárdi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal Complex Syst. 1695, 1–9.
34. D'Adamo, P., Pinna, M., Capobianco, S., Cesarani, A., D'Eustacchio, A., Fogu, P., Carella, M., Seri, M., Gasparini, P., 2003. A novel autosomal dominant non-syndromic deafness locus (DFNA48) maps to 12q13-q14 in a large Italian family. Hum. Genet. 112, 319–20.
35. Dai, P., Yu, F., Han, B., Yuan, Y., Li, Q., Wang, G., Liu, X., He, J., Huang, D., Kang, D., Zhang, X., Yuan, H., Schmitt, E., Han, D., Wong, L.-J., 2007. The prevalence of the 235delC GJB2 mutation in a Chinese deaf population. Genet. Med. 9, 283–289.
36. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.
37. Dauwerse, J.G., Bouman, K., va Essen, A.J., van der Hout, A.H., Koisters, G., Breuning, M.H., 2002. Acrofacial dysostosis in a patient with the TSC2-PKD1 contiguous gene syndrome. J. Med. Genet. 39, 136–141.
38. Davydov, E. V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., Batzoglou, S., 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. 6.
39. del Castillo, I., Villamar, M., Moreno-Pelayo, M., del Castillo, F., Álvarez, A., Tellería, D., Menéndez, I., Moreno, F., 2002. A deletion involving the connexin 30 gene in nonsyndromic hearing impairment 346, 243–249.
40. Diaz-Horta, O., Duman, D., Foster, J., Sirmaci, A., Gonzalez, M., Mahdieh, N., Fotouhi, N., Bonyadi, M., Cengiz, F.B., Menendez, I., Ulloa, R.H., Edwards, Y.J.K.,

Züchner, S., Blanton, S., Tekin, M., 2012. Whole-Exome Sequencing Efficiently Detects Rare Mutations in Autosomal Recessive Nonsyndromic Hearing Loss. PLoS One 7.

41. Donaudy, F., Ferrara, A., Esposito, L., Hertzano, R., Ben-David, O., Bell, R.E., Melchionda, S., Zelante, L., Avraham, K.B., Gasparini, P., 2003. Multiple mutations of MYO1A, a cochlear-expressed gene, in sensorineural hearing loss. Am. J. Hum. Genet. 72, 1571–1577.
42. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., Liu, X., 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. 24, 2125–2137.
43. Edery, P., Attie, T., Amiel, J., Pelet, A., Eng, C., Hofstra, R.M.W., Martelli, H., Bidaud, C., Munnich, A., Lyonnet, S., 1996. Mutation of the endothelin-3 gene in the Waardenburg-Hirschsprung disease (Shah-Waardenburg syndrome). Nat Genet 12, 442–444.
44. Eisenberger, T., Di Donato, N., Baig, S.M., Neuhaus, C., Beyer, A., Decker, E., Mürbe, D., Decker, C., Bergmann, C., Bolz, H.J., 2014. Targeted and Genomewide NGS Data Disqualify Mutations in MYO1A, the "DFNA48 Gene", as a Cause of Deafness. Hum. Mutat. 35, 565–570.
45. Eudy, J.D., Weston, M.D., Yao, S., Hoover, D.M., Rehm, H.L., Ma-Edmonds, M., Yan, D., Ahmad, I., Cheng, J.J., Ayuso, C., Cremers, C., Davenport, S., Moller, C., Talmadge, C.B., Beisel, K.W., Tamayo, M., Morton, C.C., Swaroop, A., Kimberling, W.J., Sumegi, J., 1998. Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa. Science (80-.). 280, 1753–1757.
46. Everett, L.A., Glaser, B., Beck, J.C., Idol, J.R., Buchs, A., Heyman, M., Adawi, F., Hazani, E., Nassir, E., Baxevanis, A.D., 1997. Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). Nat. Genet. 17, 411–422.
47. Fan, Y., Linardopoulou, E., Friedman, C., Williams, E., Trask, B.J., 2002. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14. 1 and paralogous regions on other human chromosomes. Genome Res. 12, 1651.
48. Fang, Q., Indzhykulian, A.A., Mustapha, M., Riordan, G.P., Dolan, D.F., Friedman, T.B., Belyantseva, I.A., Frolenkov, G.I., Camper, S.A., Bird, J.E., 2015. The 133-kDa N-terminal domain enables myosin 15 to maintain mechanotransducing stereocilia and is essential for hearing. Elife 4, 1–22.
49. Fisher, R., 1958. Statistical Methods for Research Workers, 4th EDitio. ed. Oliver and

Boy, London.

50. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kähäri, A.K., Keenan, S., Kulesha, E., Martin, F.J., Maurel, T., McLaren, W.M., Murphy, D.N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H.S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S.J., Vullo, A., Wilder, S.P., Wilson, M., Zadissa, A., Aken, B.L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T.J.P., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D.R., Searle, S.M.J., 2014. Ensembl 2014. Nucleic Acids Res. 42, 749–755.
51. Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C.Y., Jia, M., De, T., Teague, J.W., Stratton, M.R., McDermott, U., Campbell, P.J., 2015. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 43, D805–D811.
52. Fraser, G.R., Morgans, M.E., Trotter, W.R., 1960. The syndrome of sporadic goitre and congenital deafness. Qj Med 29, 279–295.
53. Friedman, T.B., Liang, Y., Weber, J.L., Hinnant, J.T., Barber, T.D., Winata, S., Arhya, I.N., Asher, J.H., 1995. A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. Nat. Genet. 9, 86–91.
54. Fujita, A., Kojima, K., Patriota, A.G., Sato, J.R., Severino, P., Miyano, S., Dopazo, J., 2011. A fast and robust statistical test based on likelihood ratio with Bartlett correction to identify Granger causality between gene sets. Bioinformatics 27, 2349–2351.
55. Gabrielli, F., Tofanelli, S., 2012. Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes. Gene 511, 461–469.
56. Gao, X., Dai, P., 2014. Impact of next-generation sequencing on molecular diagnosis of inherited non-syndromic hearing loss. J. Otol. 9, 122–125.
57. Gao, X., Su, Y., Guan, L.-P., Yuan, Y.-Y., Huang, S.-S., Lu, Y., Wang, G.-J., Han, M.-Y., Yu, F., Song, Y.-S., Zhu, Q.-Y., Wu, J., Dai, P., 2013. Novel compound heterozygous TMC1 mutations associated with autosomal recessive hearing loss in a Chinese family. PLoS One 8, e63026.
58. Garber, M., Guttman, M., Clamp, M., Zody, M.C., Friedman, N., Xie, X., 2009. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics 25, 54–62.
59. Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read

sequencing. arXiv Prepr. arXiv1207.3907 [q-bio.GN] 9.

60. Gasmelseed, N.M.A., Schmidt, M., Magzoub, M.M.A., Macharia, M., Elmustafa, O.M., Ototo, B., Winkler, E., Ruge, G., Horstmann, R.D., Meyer, C.G., 2004. Low frequency of deafness-associated GJB2 variants in Kenya and Sudan and novel GJB2 variants. Hum. Mutat. 23, 206–7.
61. Gasparini, P., Rabionet, R., Barbujani, G., Melçhionda, S., Petersen, M., Brøndum-Nielsen, K., Metspalu, A., Oitmaa, E., Pisano, M., Fortina, P., Zelante, L., Estivill, X., 2000. High carrier frequency of the 35delG deafness mutation in European populations. Genetic Analysis Consortium of GJB2 35delG. Eur. J. Hum. Genet. 8, 19–23.
62. Gazzaz, B., Weil, D., Ra??s, L., Akhyat, O., Azeddoug, H., Nadifi, S., 2005. Autosomal recessive and sporadic deafness in Morocco: High frequency of the 35delG GJB2 mutation and absence of the 342-kb GJB6 variant. Hear. Res. 210, 80–84.
63. Geard, A., 2016. Association of variants in APOL1, MYH9 and HMOX1 with micro-alnuminuria among Sickle Cell patients from Cameroon. University of Cape Town.
64. Gézsi, A., Bolgár, B., Marx, P., Sarkozy, P., Szalai, C., Antal, P., 2015. VariantMetaCaller: automated fusion of variant calling pipelines for quantitative, precision-based filtering. BMC Genomics 16, 875.
65. Golosova, O., Henderson, R., Vaskin, Y., Gabrielian, A., Grekhov, G., Nagarajan, V., Oler, A.J., Quiñones, M., Hurt, D., Fursov, M., Huyen, Y., 2014. Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. PeerJ 2, e644.
66. Gorlova, O.Y., Ying, J., Amos, C.I., Spitz, M.R., Peng, B., Gorlov, I.P., 2012. Derived SNP Alleles Are Used More Frequently Than Ancestral Alleles As Risk-Associated Variants in Common Human Diseases. J. Bioinform. Comput. Biol. 10, 1241008.
67. Grati, M., Yan, D., Raval, M.H., Walsh, T., Ma, Q., Chakchouk, I., Kannan-Sundhari, A., Mittal, R., Masmoudi, S., Blanton, S.H., Tekin, M., King, M.C., Yengo, C.M., Liu, X.Z., 2016. MYO3A Causes Human Dominant Deafness and Interacts with Protocadherin 15-CD2 Isoform. Hum. Mutat. 37, 481–487.
68. Grønborg, S., Krätzner, R., Spiegler, J., Ferdinandusse, S., Wanders, R.J.A., Waterham, H.R., Gärtner, J., 2010. Typical cMRI pattern as diagnostic clue for D-bifunctional protein deficiency without apparent biochemical abnormalities in plasma. Am. J. Med. Genet. Part A 152, 2845–2849.
69. Gupta, V., Warner, J.R., 2014. Ribosome-omics of the human ribosome. RNA 20, 1004–13.

70. Hamelmann, C., Amedofu, G.K., Albrecht, K., Muntau, B., Gelhaus, a, Brobby, G.W., Horstmann, R.D., 2001. Pattern of connexin 26 (GJB2) mutations causing sensorineural hearing impairment in Ghana. Hum. Mutat. 18, 84–5.
71. Holme, R.H., Steel, K.P., 1999. Genes involved in deafness. Curr. Opin. Genet. Dev. 9, 309–314.
72. Hoskins, B.E., Cramer, C.H., Silvius, D., Zou, D., Raymond, R.M., Orten, D.J., Kimberling, W.J., Smith, R.J.H., Weil, D., Petit, C., Otto, E.A., Xu, P.-X., Hildebrandt, F., 2007. Transcription factor SIX5 is mutated in patients with branchio-oto-renal syndrome. Am. J. Hum. Genet. 80, 800–4.
73. Hoth, C.F., Milunsky, A., Lipsky, N., Sheffer, R., Clarren, S.K., Baldwin, C.T., 1993. Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). Am. J. Hum. Genet. 52, 455–62.
74. Hutchin, T., Coy, N.N., Conlon, H., Telford, E., Bromelow, K., Blaydon, D., Taylor, G., Coghill, E., Brown, S., Trembath, R., Liu, X.Z., Bitner-Glindzicz, M., Mueller, R., 2005. Assessment of the genetic causes of recessive childhood non-syndromic deafness in the UK - Implications for genetic testing. Clin. Genet. 68, 506–512.
75. Imtiaz, A., Kohrman, D.C., Naz, S., 2014. A Frameshift Mutation in GRXCR2 Causes Recessively Inherited Hearing Loss. Hum. Mutat. 35, 618–624.
76. Jackson, B.C., Nebert, D.W., Vasiliou, V., 2010. Update of human and mouse matrix metalloproteinase families. Hum. Genomics 4, 194–201.
77. Javidnia, H., Carson, N., Awubwa, M., Byaruhanga, R., Mack, D., Vaccani, J.-P., 2014. Connexin gene mutations among ugandan patients with nonsyndromic sensorineural hearing loss. Laryngoscope 124, E373–E376.
78. Jenkinson, E.M., Rehman, A.U., Walsh, T., Clayton-Smith, J., Lee, K., Morell, R.J., Drummond, M.C., Khan, S.N., Naeem, M.A., Rauf, B., Billington, N., Schultz, J.M., Urquhart, J.E., Lee, M.K., Berry, A., Hanley, N.A., Mehta, S., Cilliers, D., Clayton, P.E., Kingston, H., Smith, M.J., Warner, T.T., Black, G.C., Trump, D., Davis, J.R.E., Ahmad, W., Leal, S.M., Riazuddin, S., King, M.C., Friedman, T.B., Newman, W.G., 2013. Perrault syndrome is caused by recessive mutations in CLPP, encoding a mitochondrial ATP-dependent chambered protease. Am. J. Hum. Genet. 92, 605–613.
79. Jenne, D., Stanley, K.K., 1985. Molecular cloning of S-protein, a link between complement, coagulation and cell-substrate adhesion. EMBO J. 4, 3153–7.
80. Jervell, A., Lange-Nielsen, F., 1957. Congenital deaf-mutism, functional heart disease

with prolongation of the QT interval, and sudden death. Am. Heart J. 54, 59–68.

81. Kabahuma, R.I., Ouyang, X., Du, L.L., Yan, D., Hutchin, T., Ramsay, M., Penn, C., Liu, X.Z., 2011. Absence of GJB2 gene mutations, the GJB6 deletion (GJB6-D13S1830) and four common mitochondrial mutations in nonsyndromic genetic hearing loss in a South African population. Int. J. Pediatr. Otorhinolaryngol. 75, 611–617.
82. Kanzaki, S., Ito, M., Takada, Y., Ogawa, K., Matsuo, K., 2006. Resorption of auditory ossicles and hearing loss in mice lacking osteoprotegerin. Bone 39, 414–419.
83. Katoh, M., Igarashi, M., Fukuda, H., Nakagama, H., Katoh, M., 2013. Cancer genetics and genomics of human FOX family genes. Cancer Lett. 328, 198–206.
84. Kent, W.J., 2002. BLAT — The BLAST -Like Alignment Tool. Genome Res. 12, 656–664.
85. Kimberling, W.J., Hildebrand, M.S., Shearer, A.E., Jensen, M.L., Halder, J.A., Trzupek, K., Cohn, E.S., Weleber, R.G., Stone, E.M., Smith, R.J.H., 2010. Frequency of Usher syndrome in two pediatric populations: Implications for genetic screening of deaf and hard of hearing children. Genet. Med. 12, 512–6.
86. Kimberling, W.J., Möller, C.G., Davenport, S., Priluck, I.A., Beighton, P.H., Greenberg, J., Reardon, W., Weston, M.D., Kenyon, J.B., Grunkemeyer, J.A., Pieke Dahl, S., Overbeck, L.D., Blackwood, D.J., Brower, A.M., Hoover, D.M., Rowland, P., Smith, R.J.H., 1992. Linkage of usher syndrome type I gene (USH1B) to the long arm of chromosome 11. Genomics 14, 988–994.
87. Kimberling, W.J., Weston, M.D., Möller, C., Davenport, S.L.H., Shugart, Y.Y., Priluck, I.A., Martini, A., Milani, M., Smith, R.J., 1990. Localization of Usher syndrome type II to chromosome 1q. Genomics 7, 245–249.
88. Kircher, M., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. g 46, 310–315.
89. Kolb, W.P., Muller-Eberhard, H.J., 1975. The membrane attack mechanism of complement. Isolation and subunit composition of the C6b-9 complex. J. Exp. Med. 141, 724–735.
90. Kuleshov, M. V, Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D., Gundersen, G.W., Ma'ayan, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90-7.
91. Kumar, S., Deffenbacher, K., Marres, H.A., Cremers, C.W., Kimberling, W.J., 2000.

Genomewide search and genetic localization of a second gene associated with autosomal dominant branchio-oto-renal syndrome: clinical and genetic implications. Am. J. Hum. Genet. 66, 1715–20.

92. Kuurila, K., Kaitila, I., Johansson, R., Grénman, R., 2002. Hearing loss in Finnish adults with osteogenesis imperfecta: A nationwide survey. Ann. Otol. Rhinol. Laryngol. 111, 939–946.
93. Lai, E., Prezioso, V.R., Tao, W., Chen, W.S., Darnell, J.E., 1991. Hepatocyte nuclear factor 3a belongs to a gene family in mammamls that is hololog ous to the Drosophila homeotic gene fork head. Genes Dev. 5, 416–427.
94. Lalwani, A.K., San Agustin, T.B., Wilcox, E.R., 1994. A locus for Waardenburg syndrome type II maps to chromosome 1p13. 3-2.1. Am. J. Hum. Genet. 55.
95. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., Maglott, D.R., 2016. ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 44, D862–D868.
96. Lasisi, A.O., Bademci, G., Foster, J., Blanton, S., Tekin, M., 2014. Common genes for non-syndromic deafness are uncommon in sub-Saharan Africa: A report from Nigeria. Int. J. Pediatr. Otorhinolaryngol. 78, 1870–3.
97. Lebeko, K., Manyisa, N., Chimusa, E.R., Mulder, N., Dandara, C., Wonkam, A., 2017. A Genomic and Protein–Protein Interaction Analyses of Nonsyndromic Hearing Impairment in Cameroon Using Targeted Genomic Enrichment and Massively Parallel Sequencing. Omi. A J. Integr. Biol. 21.
98. Lebeko, K., Sloan-Heggen, C.M., Noubiap, J.J.N., Dandara, C., Kolbe, D.L., Ephraim, S.S., Booth, K.T., Azaiez, H., Santos-Cortez, R.L.P., Leal, S.M., Smith, R.J.H., Wonkam, A., 2016. Targeted genomic enrichment and massively parallel sequencing identifies novel nonsyndromic hearing impairment pathogenic variants in Cameroonian families. Clin. Genet. 90, 288–290.
99. Lek, M., Karczewski, K.J., Minikel, E. V, Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas,

M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., Consortium, E.A., Karczewsk, K.J., Minikel, E. V, Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

100. Lévy, G., Levi-Acobas, F., Blanchard, S., Gerber, S., Larget-Piet, D., Chenal, V., Liu, X.Z., Newton, V., Steel, K.P., Brown, S.D.M., Munnich, A., Kaplan, J., Petit, C., Weil, D., 1997. Myosin VIIA gene: Heterogeneity of the mutations responsible for Usher syndrome type IB. Hum. Mol. Genet. 6, 111–116.

101. Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

102. Li, H., Ruan, J., Durbin, R., Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores Mapping short DNA sequencing reads and calling variants using mapping quality scores 1851–1858.

103. Li, J., Vogt, P.K., 1993. The retroviral oncogene qin belongs to the transcription factor family that includes the homeotic gene fork head. Proc. Natl. Acad. Sci. U. S. A.

90, 4490–4494.

104. Li, M.X., Gui, H.S., Kwan, J.S.H., Bao, S.Y., Sham, P.C., 2012. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. Nucleic Acids Res. 40.

105. Li, M.X., Kwan, J.S.H., Bao, S.Y., Yang, W., Ho, S.L., Song, Y.Q., Sham, P.C., 2013. Predicting Mendelian Disease-Causing Non-Synonymous Single Nucleotide Variants in Exome Sequencing Studies. PLoS Genet. 9, 1–11.

106. Liang, Y., Wang, A., Belyantseva, I.A., Anderson, D.W., Probst, F.J., Barber, T.D., Miller, W., Touchman, J.W., Jin, L., Sullivan, S.L., Sellers, J.R., Camper, S.A., Lloyd, R. V., Kachar, B., Friedman, T.B., Fridell, R.A., 1999. Characterization of the human and mouse unconventional myosin XV genes responsible for hereditary deafness DFNB3 and shaker 2. Genomics 61, 243–58.

107. Liang, Y., Wang, A., Probst, F.J., Arhya, I.N., Barber, T.D., Chen, K.-S., Deshmukh, D., Dolan, D.F., Hinnant, J.T., Carter, L.E., Jain, P.K., Lalwani, A.K., Li, X.C., Lupski, J.R., Moeljopawiro, S., Morell, R., Negrini, C., Wilcox, E.R., Winata, S., Camper, S.A., Friedman, T.B., 1998. Genetic mapping refines DFNB3 to 17p11.2, suggests multiple alleles of DFNB3, and supports homology to the mouse model shaker-2. Am. J. Hum. Genet. 62, 904–915.

108. Liu, X.Z., Xia, X.J., Ke, X.M., Ouyang, X.M., Du, L.L., Liu, Y.H., Angeli, S., Telischi, F.F., Nance, W.E., Balkany, T., Xu, L.R., 2002. The prevalence of connexin 26 (GJB2) mutations in the Chinese population. Hum. Genet. 111, 394–397.

109. Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., Cai, L., 2012. MutationTaster2: mutation prediction for the deep-sequencing age. Nat. Methods Nat. Methods Nat. Methods 9, 743–748.

110. Manrai, A.K., Funke, B.H., Rehm, H.L., Olesen, M.S., Maron, B.A., Szolovits, P., Margulies, D.M., Loscalzo, J., Kohane, I.S., 2016. Genetic Misdiagnoses and the Potential for Health Disparities. new Engl. J. Med. Spec. 375, 655–665.

111. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo1, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

112. Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., Raney, B.J., Pohl, A., Malladi, V.S., Li, C.H., Lee, B.T., Learned, K., Kirkup, V., Hsu, F., Heitner, S., Harte, R.A.,

Haeussler, M., Guruvadoo, L., Goldman, M., Giardine, B.M., Fujita, P.A., Dreszer, T.R., Diekhans, M., Cline, M.S., Clawson, H., Barber, G.P., Haussler, D., Kent, W.J., 2013. The UCSC Genome Browser database: Extensions and updates 2013. Nucleic Acids Res. 41, 918–923.

113. Mi, H., Muruganujan, A., Thomas, P.D., 2013. PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 41, 377–386.

114. Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.

115. Morton, N.E., 1991. Genetic epidemiology of hearing impairment. Ann. N. Y. Acad. Sci. 630, 16–31.

116. Moteki, H., Azaiez, H., Booth, K.T., Shearer, A.E., Sloan, C.M., Kolbe, D.L., Nishio, S., Hattori, M., Usami, S., Smith, R.J.H., 2016. Comprehensive genetic testing with ethnic-specific filtering by allele frequency in a Japanese hearing-loss population. Clin. Genet. 89, 466–472.

117. Murray, C.J.L., Barber, R.M., Foreman, K.J., Ozgoren, A.A., Abd-Allah, F., Abera, S.F., Aboyans, V., Abraham, J.P., Abubakar, I., Abu-Raddad, L.J., 2015. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. Lancet 386, 2145–2191.

118. Najmabadi, H., Kahrizi, K., 2014. Genetics of non-syndromic hearing loss in the Middle East. Int. J. Pediatr. Otorhinolaryngol. 78, 2026–2036.

119. Nal, N., Ahmed, Z.M., Erkal, E., Alper, O.M., Luleci, G., Dinc, O., Waryah, A.M., Ain, Q., Tasneem, S., Husnain, T., Chattaraj, P., Riazuddin, S., Boger, E., Ghosh, M., Kabra, M., Riazuddin, S., Morell, R.J., Friedman, T.B., 2007. Mutational Spectrum of : The Large N-Terminal Extension of Myosin XVA MY O15A Is Required for Hearing. HUMANMUTATION 28, 1014–1019.

120. National Center for Biotechnology Information, N.L. of M., n.d. Database of Single Nucleotide Polymorphisms (dbSNP).

121. Neyroud, N., Tesson, F., Denjoy, I., Leibovici, M., Donger, C., Barhanin, J., Fauré, S., Gary, F., Coumel, P., Petit, C., Schwartz, K., Guicheney, P., 1997. A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome. Nat Genet 15, 186–189.

122. Ng, P.C., Henikoff, S., 2002. Accounting for human polymorphisms predicted

to affect protein function. Genome Res. 12, 436–446.

123. Ng, P.C., Henikoff, S., 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 31, 3812–3814.

124. Ng, P.C., Henikoff, S., 2006. Predicting the effects of amino acid substitutions on protein function. Annu. Rev. Genomics Hum. Genet. 7, 61–80.

125. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Abigail, W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Evan, E., Bamshad, M., Nickerson, D. a, Shendure, J., 2010. Targeted Capture and Massicely Parallel Sequencing of twelve human exomes. Nature 461, 272–276.

126. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D., 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733–D745.

127. Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O.G., Kandrov, D., Rasputin, K., Syabro, M., Tleukenov, T., 2012. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics 28, 1166–1167.

128. Olusanya, B.O., Neumann, J., Saunders, J.E., 2014. The global burden of disabling hearing impairment : a call to action. Bull. World Health Organ. 92, 367–373.

129. Olusanya, B.O., Newton, V.E., 2007. Global burden of childhood hearing impairment and disease control priorities for developing countries. Lancet 369, 1314–1317.

130. Pandya, A., Arnos, K.S., Xia, X.J., Welch, K.O., Blanton, S.H., Friedman, T.B., Garcia Sanchez, G., Liu, X.Z., Morell, R., Nance, W.E., 2003. Frequency and distribution of GJB2 (connexin 26) and GJB6 (connexin 30) mutations in a large North American repository of deaf probands. Genet. Med. 5, 295–303.

131. Paten, B., Herrero, J., Fitzgerald, S., Beal, K., Flicek, P., Holmes, I., Birney, E.,

2008. Genome-wide nucleotide-level mammalian ancestor reconstruction. Genome Res. 18, 1829–1843.

132. Paterson, C.R., Monk, E. a, McAllion, S.J., 2001. How common is hearing impairment in osteogenesis imperfecta? J. Laryngol. Otol. 115, 280–282.

133. Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2, 2074–2093.

134. Pedersen, U., 1984. Hearing Loss in Patients with Osteogenesis Imperfecta A Clinical and Audiological Study of 201 Patients. Scand. Audiol. 13, 67–74.

135. Pendred, V., 1896. Deaf-mutism and goitre. Lancet 148, 532.

136. Perrault, M., Klotz, B., Housset, E., 1950. Two cases of Turner syndrome with deaf-mutism in two sisters. Bull. Mem. Soc. Med. Hop. Paris 67, 79–84.

137. Petit, C., 2001. Usher Syndrome: From Genetics to Pathogenesis 0, 271–97.

138. Pierce, S.B., Gersak, K., Michaelson-Cohen, R., Walsh, T., Lee, M.K., Malach, D., Klevit, R.E., King, M.C., Levy-Lahad, E., 2013. Mutations in LARS2, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in Perrault syndrome. Am. J. Hum. Genet. 92, 614–620.

139. Pierce, S.B., Walsh, T., Chisholm, K.M., Lee, M.K., Thornton, A.M., Fiumara, A., Opitz, J.M., Levy-Lahad, E., Klevit, R.E., King, M.C., 2010. Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of perrault syndrome. Am. J. Hum. Genet. 87, 282–288.

140. Pillion, J.P., Santos, F., Vernick, D.M., Shapiro, J., 2014. Chapter 32 - Hearing Loss in Osteogenesis Imperfecta BT - Osteogenesis Imperfecta. Academic Press, San Diego, pp. 305–312.

141. Pingault, V., Bondurand, N., Kuhlbrodt, K., Goerich, D.E., Prehu, M.-O., Puliti, A., Herbarth, B., Hermans-Borgmeyer, I., Legius, E., Matthijs, G., Amiel, J., Lyonnet, S., Ceccherini, I., Romeo, G., Smith, J.C., Read, A.P., Wegner, M., Goossens, M., 1998. SOX10 mutations in patients with Waardenburg-Hirschsprung disease. Nat Genet 18, 171–173.

142. Podack, E.R., Kolb, W.P., Hans, J., 1977. The SC5b-7 Complex: Formation, Isolation, Properties, and Subunit Composition. J. Immunol. 119, 2024–2029.

143. Podack, E.R., Müller-Eberhard, H.J., 1979. Isolation of Human S-Protein , Complex of Complement * an Inhibitor of the Membrane Attack complex. J. Biol. Chem. 254, 9908–9914.

144. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., Siepel, A., 2010. Detection of

nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–121.

145. Popejoy, A.B., Fullerton, S.M., 2016. Genomics is failing on diversity. Nature 538, 161–164.

146. Přecechtěl, A., 1927. Pedigree of Anomalies in the First and the Second Branchial Cleft, Inherited According to the Laws of Mendel, and A Contribution to the Technique of the Extirpation of Congenital Lateral Fistulae Colli. Acta Otolaryngol. 11, 23–30.

147. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Gene 38, 904–909.

148. Probst, F.J., Fridell, R. a, Raphael, Y., Saunders, T.L., Wang, A., Liang, Y., Morell, R.J., Touchman, J.W., Lyons, R.H., Noben-trauth, K., Friedman, T.B., Camper, S.A., 1998. Correction of deafness in shaker-2 mice by an unconventional myosin in a BAC transgene. Science 280, 1444–1447.

149. Pruitt, K.D., Maglott, D.R., 2001. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res 29, 137–40.

150. R Core Team, 2016. R: A Language and Environment for Statistical Computing.

151. Reardon, W., Coffey, R., Phelps, P.D., Luxon, L.M., Stephens, D., Kendall-Taylor, P., Britton, K.E., Grossman, A., Trembath, R., 1997. Pendred syndrome - 100 years of underascertainment? QJM - Mon. J. Assoc. Physicians 90, 443–447.

152. Rędowicz, M.J., 1999. Myosins and deafness. J. Muscle Res. Cell Motil. 20, 241–248.

153. Reva, B., Antipin, Y., Sander, C., 2007. Determinants of protein function revealed by combinatorial entropy optimization. Genome Biol. 8, R232.

154. Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: Application to cancer genomics. Nucleic Acids Res. 39, 37–43.

155. Robin, N.H., Moran, R.T., Ala-Kokko, L., 2017. Stickler syndrome.

156. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz1, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–27.

157. Ruf, R.G., Berkman, J., Wolf, M.T.F., Nurnberg, P., Gattas, M., Ruf, E.-M., Hyland, V., Kromberg, J., Glass, I., Macmillan, J., Otto, E., Nurnberg, G., Lucke, B., Hennies, H.C., Hildebrandt, F., 2003. A gene locus for branchio-otic syndrome maps to chromosome 14q21.3-q24.3. J. Med. Genet. 40, 515–9.

158. Ruf, R.G., Xu, P.-X., Silvius, D., Otto, E.A., Beekmann, F., Muerb, U.T., Kumar, S., Neuhaus, T.J., Kemper, M.J., Raymond, R.M.J., Brophy, P.D., Berkman, J., Gattas, M., Hyland, V., Ruf, E.-M., Schwartz, C., Chang, E.H., Smith, R.J.H., Stratakis, C.A., Weil, D., Petit, C., Hildebrandt, F., 2004. SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. Proc. Natl. Acad. Sci. U. S. A. 101, 8090–8095.

159. Ruiz, A., Altaba, I., Cox, C., Jessell, T.M., Klar, A., Kandel, E.R., 1993. Ectopic neural expression of a floor plate marker in frog embryos injected with the midline transcription factor Pintallavis. Dev. Biol. 90, 8268–8272.

160. Salime, S., Charif, M., Bousfiha, A., Elrharchi, S., Bakhchane, A., Charoute, H., Kabine, M., Snoussi, K., Lenaers, G., Barakat, A., 2017. Homozygous mutations in PJVK and MYO15A genes associated with non-syndromic hearing loss in Moroccan families. Int. J. Pediatr. Otorhinolaryngol. 101, 25–29.

161. Salles, F.T., Jr, R.C.M., Manor, U., Dougherty, G.W., Sousa, D., Moore, J.E., Yengo, C.M., Dosé, A.C., 2009. Myosin IIIa boosts elongation of stereocilia by transporting espin 1 to the plus ends of actin filaments. Nat Cell Biol 11, 443–450.

162. Sánchez-Martín, M., Rodríguez-García, A., Pérez-Losada, J., Sagrera, A., Read, A.P., Sánchez-García, I., 2002. SLUG (SNAI2) deletions in patients with Waardenburg disease. Hum. Mol. Genet. 11, 3231–3236.

163. Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.

164. Schrijver, I., 2004. Hereditary non-syndromic sensorineural hearing loss: transforming silence to sound. J. Mol. Diagn. 6, 275–284.

165. Schulze-Bahr, E., Wang, Q., Wedekind, H., Haverkamp, W., Chen, Q., Sun, Y., Ruble, C., Hordt, M., Towbin, J.A., Borggrefe, M., Assmann, G., Qu, X., Somberg, J.C., Breithardt, G., Oberti, C., Funke, H., 1997. KCNE1 mutations cause Jervell and Lange-Nielsen syndrome. Nat Genet 17, 267–268.

166. Schvartz, I., Seger, D., Shaltiel, S., 1999. Vitronectin. Int. J. Biochem. Cell Biol. 31, 539–544.

167. Selicorni, A., Guerneri, S., Ratti, A., Pizzuti, A., 2002. Cytogenetic mapping of a novel locus for type II Waardenburg syndrome. Hum. Genet. 110, 64–67.

168. Shafique, S., Siddiqi, S., Schraders, M., Oostrik, J., Ayub, H., Bilal, A., Ajmal, M., Seco, C.Z., Strom, T.M., Mansoor, A., Mazhar, K., Shah, S.T.A., Hussain, A., Azam, M., Kremer, H., Qamar, R., 2014. Genetic spectrum of autosomal recessive non-

syndromic hearing loss in Pakistani families. PLoS One 9, 1–10.

169. Shahzad, M., Sivakumaran, T.A., Qaiser, T.A., Schultz, J.M., Hussain, Z., Flanagan, M., Bhinder, M.A., Hissell, D., Greinwald Jr, J.H., Khan, Shaheen N, Friedman, T.B., Zhang, K., Riazuddin, S., Riazuddin, S., Ahmed, Z.M., 2013. Genetic analysis through OtoSeq of Pakistani families segregating prelingual hearing loss. Otolaryngol. -- Head Neck Surg. 149, 478–487.

170. Shearer, A.E., DeLuca, A.P., Hildebrand, M.S., Taylor, K.R., Gurrola, J., Scherer, S., Scheetz, T.E., Smith, R.J.H., 2010. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. Proc. Natl. Acad. Sci. U. S. A. 107, 21104–21109.

171. Shearer, A.E., Hildebrand, M.S., Sloan, C.M., Smith, R.J.H., 2011. Deafness in the genomics era. Hear. Res. 282, 1–9.

172. Sherry, S., Ward, M., Kholodov, M., Baker, J., Phan, L., Smigielski, E., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29, 308–11.

173. Shihab, H.A., Gough, J., Mort, M., Cooper, D.N., Day, I.N.M., Gaunt, T.R., 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. Hum. Genomics 8, 11.

174. Sivakumaran, T.A., Husami, A., Kissell, D., Zhang, W., Keddache, M., Black, a. P., Tinkle, B.T., Greinwald, J.H., Zhang, K., 2013. Performance Evaluation of the Next-Generation Sequencing Approach for Molecular Diagnosis of Hereditary Hearing Loss. Otolaryngol. -- Head Neck Surg. 148, 1007–1016.

175. Sloan-Heggen, C.M., Bierer, A.O., Shearer, A.E., Kolbe, D.L., Nishimura, C.J., Frees, K.L., Ephraim, S.S., Shibata, S.B., Booth, K.T., Campbell, C.A., Ranum, P.T., Weaver, A.E., Black-Ziegelbein, E.A., Wang, D., Azaiez, H., Smith, R.J.H., 2016. Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. Hum. Genet. 135, 441–450.

176. Smith, R.J.H., Bale, J.F., White, K.R., 2005. Sensorineural hearing loss in children. Lancet 365, 879–890.

177. Smith, R.J.H., Berlin, C.I., Hejtmancik, J.F., Keats, B.J.B., Kimberling, W.J., Lewis, R.A., Möller, C.G., Pelias, M.Z., Tranebjærǵ, L., 1994. Clinical diagnosis of the Usher syndromes. Am. J. Med. Genet. 50, 32–38.

178. Stenson, P.D., Mort, M., Ball, E. V., Howells, K., Phillips, A.D., Cooper, D.N., Thomas, N.S.T., 2009. The human gene mutation database: 2008 update. Genome Med.

1, 1–6.

179. Su, Z.-J., Zhang, Q.-X., Liu, G.-F., Song, X.-H., Li, Q., Wang, R.-J., Chen, H.-B., Xu, X.-Y., Sui, X.-X., Huang, D.-Y., 2010. Bioinformatic analysis of the human DHRS4 gene cluster and a proposed mechanism for its transcriptional regulation. BMC Mol. Biol. 11, 43.

180. Sugiura, M., Sato, E., Nakashima, T., Sugiura, J., Furuhashi, A., Yoshino, T., Nakayama, A., Mori, N., Murakami, H., Naganawa, S., 2005. Long-term follow-up in patients with Pendred syndrome: Vestibular, auditory and other phenotypes. Eur. Arch. Oto-Rhino-Laryngology 262, 737–743.

181. Suzuki, Y., Shimozawa, N., Yajima, S., Tomatsu, S., Kondo, N., Nakada, Y., Akaboshi, S., Lai, M., Tanabe, Y., Hashimoto, T., 1994. Novel subtype of peroxisomal acyl-CoA oxidase deficiency and bifunctional enzyme deficiency with detectable enzyme protein: identification by means of complementation analysis. Am. J. Hum. Genet. 54, 36–43.

182. Swanepoel, D., Störbeck, C., Friedland, P., 2009. Early hearing detection and intervention in South Africa. Int. J. Pediatr. Otorhinolaryngol. 73, 783–786.

183. Tassabehji, M., Newton, V.E., Read, A.P., 1994. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. Nat Genet 8, 251–255.

184. Tassabehji, M., Read, A.P., Newton, V.E., Harris, R., Balling, R., Gruss, P., Strachan, T., 1992. Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. Nature 355, 635–636.

185. Teer, J.K., Mullikin, J.C., 2010. Exome sequencing: The sweet spot before whole genomes. Hum. Mol. Genet. 19, 145–151.

186. Tekin, D., Yan, D., Bademci, G., Feng, Y., Guo, S., Foster, J., Blanton, S., Tekin, M., Liu, X., 2017. A next-generation sequencing gene panel (MiamiOtoGenes) for comprehensive analysis of deafness genes. Hear. Res. 333, 179–184.

187. Tekin, M., Arnos, K.S., Pandya, A., 2001. Advances in hereditary deafness. Lancet 358, 1082–1090.

188. The 1000 Genome Project Consortium Ancestral Allele Reference Sequence [WWW Document], 2012. URL http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/su%0Apporting/ancestral_alignments/%0A

189. The 1000 Genomes Project Consortium, 2015. A global reference for human

genetic variation. Nature 526, 68–74.

190. The Molecular Otolaryngology and Renal Research Laboratories. The University of Iowa, 2016. Deafness Variation Database [WWW Document]. URL http://deafnessvariationdatabase.org/ (accessed 6.2.17).

191. Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., 2003. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. Genome Res. 13, 2129–2141.

192. Thorrez, L., Van Deun, K., Tranchevent, L.C., Van Lommel, L., Engelen, K., Marchal, K., Moreau, Y., Van Mechelen, I., Schuit, F., 2008. Using ribosomal protein genes as reference: A tale of caution. PLoS One 3.

193. Thorvaldsdottir, H., Robinson, J., P.Mesirov, J., 2018. Integrative Genomics Viewer (IGV: high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192.

194. Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192.

195. Trotta, L., Iacona, E., Primignani, P., Castorina, P., Radaelli, C., Del Bo, L., Coviello, D., Ambrosetti, U., 2011. GJB2 and MTRNR1 contributions in children with hearing impairment from Northern Cameroon. Int. J. Audiol. 50, 133–8.

196. Tyson, J., Tranebjærg, L., Bellman, S., Wren, C., Taylor, J.F.N., Bathen, J., Aslaksen, B., Sørland, S.J., Lund, O., Malcolm, S., Pembrey, M., Bhattacharya, S., Bitner-Glindzicz, M., 1997. IsK and KvLQT1: Mutation in either of the two subunits of the slow component of the delayed rectifier potassium channel can cause Jervell and Lange-Nielsen syndrome. Hum. Mol. Genet. 6, 2179–2185.

197. Usher, C.H., 1914. On the inheritance of retinitis pigmentosa, with notes and cases.

198. Van Camp, G., Smith, R., n.d. Hereditary Hearing Loss Homepage [WWW Document]. URL http://hereditaryhearingloss.org

199. Van Der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Levy-moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V, Altshuler, D., Gabriel, S., Depristo, M.A., 2014. From FastQ data to high confidence varant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 11.

200. Van Raay, T.J., Connors, T.D., Klinger, K.W., Landes, G.M., Burn, T.C., 1996.

A Novel Ribosomal Protein L3-like Gene (RPL3L) Maps to the Autosomal Dominant Polycystic Kidney Disease Gene Region. Genomics 37, 172–176.

201. Von Graefe, A., 1858. Exceptionelles verhalten des gesichtsfeldes bei pigmententartung der netzhaut. Von Graefes Arch Ophthalmol 4, 250–253.

202. Waardenburg, P.J., 1951. A new syndrome combining developmental anomalies of the eyelids, eyebrows and nose root with pigmentary defects of the iris and head hair and with congenital deafness. Am. J. Hum. Genet. 3, 195–253.

203. Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M.K., Thornton, A.M., Roeb, W., Abu Rayyan, A., Loulus, S., Avraham, K.B., King, M.-C., Kanaan, M., 2010. Whole Exome Sequencing and Homozygosity Mapping Identify Mutation in the Cell Polarity Protein GPSM2 as the Cause of Nonsyndromic Hearing Loss DFNB82. Am. J. Hum. Genet. 87, 90–94.

204. Walsh, T., Walsh, V., Vreugde, S., Hertzano, R., Shahin, H., Haika, S., Lee, M.K., Kanaan, M., King, M.-C., Avraham, K.B., 2002. From flies' eyes to our ears: mutations in a human class III myosin cause progressive nonsyndromic hearing loss DFNB30. Proc. Natl. Acad. Sci. U. S. A. 99, 7518–7523.

205. Walsh, V.L., Raviv, D., Dror, A.A., Shahin, H., Walsh, T., Kanaan, M.N., Avraham, K.B., King, M.C., 2011. A mouse model for human hearing loss DFNB30 due to loss of function of myosin IIIA. Mamm. Genome 22, 170–177.

206. Wang, A., Liang, Y., Fridell, R.A., Probst, F.J., Wilcox, E.R., Touchman, J.W., Morton, C.C., Morell, R.J., Noben-Trauth, K., Camper, S.A., Friedman, T.B., 1998. Association of Unconventional Myosin MYO15 Mutations with Human Nonsyndromic Deafness DFNB3. Science (80-.). 280, 1447–1451.

207. Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.

208. Watkins, P.A., Chen, W.W., Harris, C.J., Hoefler, G., Hoefler, S., Blake, D.C., Balfe, A., Kelley, R.I., Moser, A.B., Beard, M.E., Moser, H.W., 1989. Peroxisomal bifunctional enzyme deficiency. J. Clin. Invest. 83, 771–777.

209. Weigel, D., Jürgens, G., Küttner, F., Seifert, E., Jäckle, H., 1989. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. Cell 57, 645–658.

210. Weston, M.D., Kelley, P.M., Overbeck, L.D., Wagenaar, M., Orten, D.J., Hasson, T., Chen, Z.Y., Corey, D., Mooseker, M., Sumegi, J., Cremers, C., Moller, C., Jacobson, S.G., Gorin, M.B., Kimberling, W.J., Sume, J., Cremers, C., Iller, C.M.,

Jacobson, S.G., Gorin, M.B., Kimberling, W.J., Sumegi, J., Cremers, C., Moller, C., Jacobson, S.G., Gorin, M.B., Kimberling, W.J., 1996. Myosin VIIA mutation screening in 189 Usher syndrome type 1 patients. Am. J. Hum. Genet. 59, 1074–1083.

211. WHO, 2017. Prevention of blindness and deafness [WWW Document]. URL http://www.who.int/pbd/deafness/hearing_impairment_grades/en/ (accessed 8.2.17).

212. WHO Media Centre, 2017. Deafness and hearing loss [WWW Document]. URL http://www.who.int/mediacentre/factsheets/fs300/en/# (accessed 10.4.15).

213. Willems, P.J., 2000. Genetic causes of hearing loss. New Engl. J. od Med. 342, 1101–1109.

214. Wilson, B.S., Tucci, D.L., Merson, M.H., Donoghue, G.M.O., 2017. Review Global hearing health care : new findings and perspectives. Lancet 6736.

215. Winata, S., Arhya, I.N., Moeljopawiro, S., Hinnant, J.T., Liang, Y., Friedman, T.B., Asher, J.H., 1995. Congenital non-syndromal autosomal recessive deafness in Bengkala, an isolated Balinese village. J. Med. Genet. 32, 336–43.

216. Wonkam, A., Bosch, J., Noubiap, J.J.N., Lebeko, K., Makubalo, N., Dandara, C., 2015. No evidence for clinical utility in investigating the connexin genes GJB2, GJB6 and GJA1 in non-syndromic hearing loss in black Africans. South African Med. J. 105, 23–26.

217. Wonkam, A., Noubiap, J.J.N., Djomou, F., Fieggen, K., Njock, R., Toure, G.B., 2013. Aetiology of childhood hearing loss in Cameroon (sub-Saharan Africa). Eur. J. Med. Genet. 56, 20–25.

218. Wu, J., Vallenius, T., Ovaska, K., Westermarck, J., Mäkelä, T.P., Hautaniemi, S., 2009. Integrated network analysis platform for protein-protein interactions. Nat. Methods 6, 75–77.

219. Yan, D., Tekin, D., Bademci, G., Foster, J., Cengiz, F.B., Kannan-Sundhari, A., Guo, S., Mittal, R., Zou, B., Grati, M., Kabahuma, R.I., Kameswaran, M., Lasisi, T.J., Adedeji, W.A., Lasisi, A.O., Menendez, I., Herrera, M., Carranza, C., Maroofian, R., Crosby, A.H., Bensaid, M., Masmoudi, S., Behnam, M., Mojarrad, M., Feng, Y., Duman, D., Mawla, A.M., Nord, A.S., Blanton, S.H., Liu, X.Z., Tekin, M., 2016. Spectrum of DNA variants for non-syndromic deafness in a large cohort from multiple continents. Hum. Genet. 135, 953–961.

220. Yang, T., Gurrola, J.G., Wu, H., Chiu, S.M., Wangemann, P., Snyder, P.M., Smith, R.J.H., 2009. Mutations of KCNJ10 Together with Mutations of SLC26A4 Cause Digenic Nonsyndromic Hearing Loss Associated with Enlarged Vestibular

Aqueduct Syndrome. Am. J. Hum. Genet. 84, 651–657.

221. Yang, T., Vidarsson, H., Rodrigo-Blomqvist, S., Rosengren, S.S., Enerback, S., Smith, R.J.H., 2007. Transcriptional control of SLC26A4 is involved in Pendred syndrome and nonsyndromic enlargement of vestibular aqueduct (DFNB4). Am. J. Hum. Genet. 80, 1055–63.

222. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134.

223. Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., Mclaren, W., Moore, B., Mudge, J., Murphy, N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A., Flicek, P., 2018. Ensembl 2018. Nucleic Acids Res. 46, 754–761.

224. Zhang, Q., Li, Y., Liu, G., Xu, X., Song, X., Liang, B., Li, R., Xie, J., Du, M., Xiao, L., Gan, X., Huang, D., 2009. Alternative transcription initiation and splicing variants of the DHRS4 gene cluster. Biosci. Rep. 29, 47–56.

225. Zhang, Y.L., Sun, F.T., Zhang, Z., Chen, X.X., Liu, A.X., Pan, J.J., Peng, F., Zhou, S., Sun, L.J., 2014. Comprehensive expression analysis suggests functional overlapping of human FOX transcription factors in cancer. Asian Pacific J. Cancer Prev. 15, 10475–10481.

226. Zheng, X., Saunders, T.L., Camper, S.A., Samuelson, L.C., Ginsburg, D., 1995. Vitronectin is not essential for normal mammalian development and fertility. Proc. Natl. Acad. Sci. U. S. A. 92, 12426–12430.

Appendix

Appendix 1: DNA extraction methods

DNA Extraction using the Gentra Puregene Kit (QIAGEN Sample and Assay Technologies, 2014)

- Add 9ml of RBC lysis buffer to 4ml of whole blood and mix by inverting 10 times
- Incubate for 5 minutes at room temperature (15°C to 25°C) and invert at least once during the incubation
- Centrifuge the sample for 2 minutes at 2 000gto pellet out the white blood cells
- Discard the supernatant carefully, leaving the pellet and 200µl residual liquid
- Vortex the sample vigorously until the pellet is completely resuspended
- Add 3ml cell lysis solution and pipette mix prior to vortexing for 10 seconds
- Add 15µl RNase to the solution and mix by inverting 25 times prior to incubating at 37°C for 15 minutes
- Add 1 ml protein precipitation solution and vortex, at high speed, for 20 seconds
- Centrifuge the solution for 5 minutes at 2 000g
- Add 3ml of isopropanol into a clean 15ml tube and carefully add the supernatant from the previous step by pouring
- Mix by inverting 50 times until the DNA is visible as threads or a clump
- Centrifuge for 3 minutes at 2 000g such that the DNA is visible as a clear white pellet
- Discard the pellet carefully and drain the tube by inverting on a clean piece of absorbent paper; taking care not disturb the pellet
- Add 3ml of 70% ethanol and invert several times to wash the DNA pellet
- Centrifuge at 2 000g for 1 minute
- Discard the pellet carefully and drain the tube by inverting on a clean piece of absorbent paper; taking care not disturb the pellet; and air dry for 5 to 10 minutes
- Add 300µl of DNA hydration solution and vortex at medium speeds for 5 seconds to mix
- Incubate at 65°C for 1 hour to dissolve the DNA
- Incubate overnight with gentle shaking.
- Samples can be transferred to storage tubes following brief centrifugation.

DNA extraction using modified salting out method

- If fresh blood has a volume of 3ml or less, transfer the blood into a 15ml conical tube, else make a buffy for blood greater than 3ml and transfer 3ml of the buffer into a 15ml conical tube
- Add 9ml of filter sterilised red blood cell lysis buffer and incubate for 30 minutes at room temperature following mixing
- Vortex the solutions and centrifuge for 10 minutes at 2 000rpm
- Decant the supernatant and add 1ml of fresh red blood cell lysis buffer and vortex to mix the solution until the pellet is resuspended.
    - Repeat this step and the previous centrifugation step to ensure that the pellet is clean
- Add 3ml of filter sterilised cell lysis buffer and vortex to mix
- Add 12.5µl of 20mg/ml Proteinase K and 100µl of 20% SDS to each tube and vortex to mix
- Incubate the samples, in a water bath, at 37°C for 1 to 3 days
    - There should be no pellet and the liquid should be clear following this step
- Add 1ml of 6M NaCl and vortex vigorously to mix
- Centrifuge at 2 000rpm for 20 minutes and transfer 5ml of the supernatant into a clean 15ml conical tube
- Add 10ml of 100% ethanol and mix gently but inverting 50X
- Centrifuge for 10 minutes at 2 000rpm
- Decant the supernatant and add 2ml of ice cold 70% ethanol to the solution
- Centrifuge or 10 minutes at 2 000rpm and carefully remove the supernatant without disturbing the DNA pellet
- Air dry at room temperature the pellet for 2 hours, or until dry
- Resuspend the pellet in 150µl DNA rehydration solution and leave at room temperature for one week to ensure complete resuspension.
- Transfer the DNA into storage tubes

Buffy coat preparation

- Centrifuge whole blood sample at 2 500rpm for 20 minutes
    - The buffy coat will contain the leucocytes and some erythrocytes and blood plasma

Appendix 2: Reagents and buffers

TE Buffer

5ml of 1M Tris-HCl (pH 8)

1ml of 0.5M EDTA

500ML distilled H_2O

Autoclave

Polymerase Chain Reaction

Sabax water for injections

Adcock In

Lot number: S8513

GoTaq® G@ DNA Polymerase

Promega

Lot number: 0000 1651 34

dNTP Set

Solis BioDyne

Lot number: DCTP3400

Primer

Whitehead Scientific

Gel Electrophoresis

Seakem® LE Agarose

Lonza

Lot number: 0000 5718 11

GeneRuler 1kb DNA Ladder

Thermo Scientific

Lot number: 0058 6155

SYBR ® Safe DNA gel stain

Invitrogen

Lot number:1906 577

10X TBE (1l)

7.5g EDTA Di-Sodium Salt

Nexgen Chemical

Batch number: NEX-90030

55g Boric acid

108g Tris

Melford Biolaboratories

Batch number: 2641

Procedure

- Components measured out in 1l Erlenmeyer flask
- Add 750ml distilled H_2O
- Heat content whilst stirring until the powder is completely dissolved
- Decant into 2l measuring cylinder and fill up to 1l mark with distilled H_2O
- Pour into a 1l labelled bottle.

PCR Clean-up

FastAP Thermosenstive Alkaline Phosphatase

Thermo Scientific

Lot number:0031 0272

Exonuclease I

Thermo Scientific

Lot number: 0024 9850

Direct Cycle Sanger Sequencing

BigDye® Terminator v3.1 Cycle Sequencing Kit

Applied Biosystems

Lot number:1507 380

Appendix 3: Equipment and machines

NanoDrop® ND-1000 Spectrophotometer

Serial number 4206

ND-1000 Software v3.8.1

SimpliAmo Thermal Cycler

Applied Biosysterms

PS500XT DC Power Supply

Hoefer Scientific Instruments

Mettler PE 3000 Scale

Mettler Toledo

Vortex Mixer

Labnet International Inc

Uvipro gold

Uvitec

UVIpro Version 12.3 for Windows

3130xl Genetic Analyzer

ABI Prism

Applied Biosystems | Hitachi

FoundationData Collection Version 3.1.1

Sequencing Analysis v5.4

Centrifuge 5415 D

Eppendorf

Centrifuge 5804

Eppendorf

Appendix 4: PCR and clean-up constituents and protocols

MYO3A and MYO15A PCR Constituents

PCR Protocol

Reagent	Stock	Required	n=1 (µl)
dH_2O	-	-	16.9
Taq Buffer	5X	1X	5
dNTPs	5mM	200µM	1
Forward Primer	20µM	0.04µM	0.5
Reverse Primer	20µM	0.04µM	0.5
Taq	5U/µL	0.02µM	0.1
DNA	-	100ng/µl	1
Total volume	-	-	25

Step	Temperature (°C)	Time
Initial Denaturation	95	3 Minutes
Denaturation	94	30 Seconds
Primer Annealing	61	40 Seconds
Elongation	72	1 Minute
Final Elongation	72	10 minutes

The denaturation, primer annealing and elongation steps are performed for 35 cycles for *MYO3A* and for 45 cycles for *MYO15A*

PCR Clean-up

Reagent	n=1 (µl)
PCR Product	17.8
FastAP	2.0
Exo1	0.2
Total volume	20

PCR Clean-up Protocol

	Temperature (°C)	Time
Incubation	37	1 hour
Deactivation	75	15 minutes

Appendix 5: Sequencing and clean-up constituents and protocols

Sodium Acetate Precipitation

Cycle sequencing

Reagent	n=1 (µl)
dH_2O	4.5
Dilution Buffer	2
Terminator Mix	1
Primer	0.5
PCR Product	2
Total Volume	10

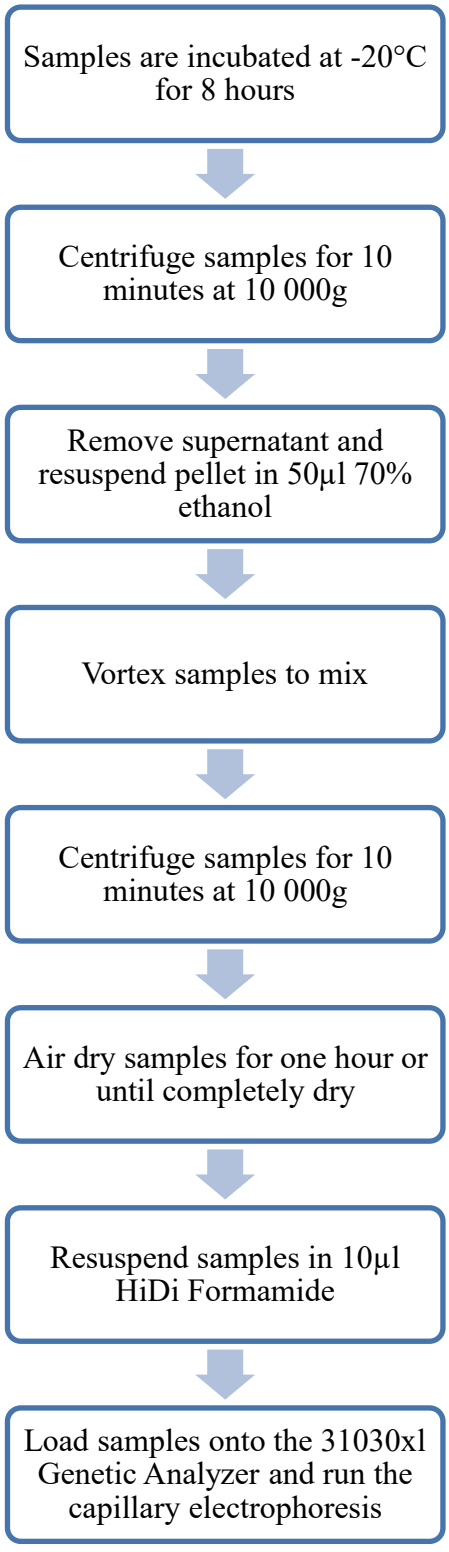
Protocol

Step	Temperature (°C)	Time
Initial Denaturation	96	5 minutes
Denaturation	96	30 seconds
Primer Annealing	50	15 seconds
Elongation	60	4minutes

Cycle sequencing proceeds with 30 cycles of the denaturation primer annealing and elongation steps.

Ethanol and Sodium Acetate Precipitation

Reagent	n=1 (µl)
100% Ethanol	50
Sodium acetate	2
Total volume	52


Samples are incubated at -20°C for 8 hours
Centrifuge samples for 10 minutes at 10 000g
Remove supernatant and resuspend pellet in 50µl 70% ethanol
Vortex samples to mix
Centrifuge samples for 10 minutes at 10 000g
Air dry samples for one hour or until completely dry
Resuspend samples in 10µl HiDi Formamide
Load samples onto the 31030xl Genetic Analyzer and run the capillary electrophoresis

EDTA Precipitation

Cycle Sequencing

Reagent	n=1 (µl)
dH_2O	8.5
Dilution Buffer	4
Terminator Mix	2
Primer	0.5
PCR Product	5
Total Volume	20

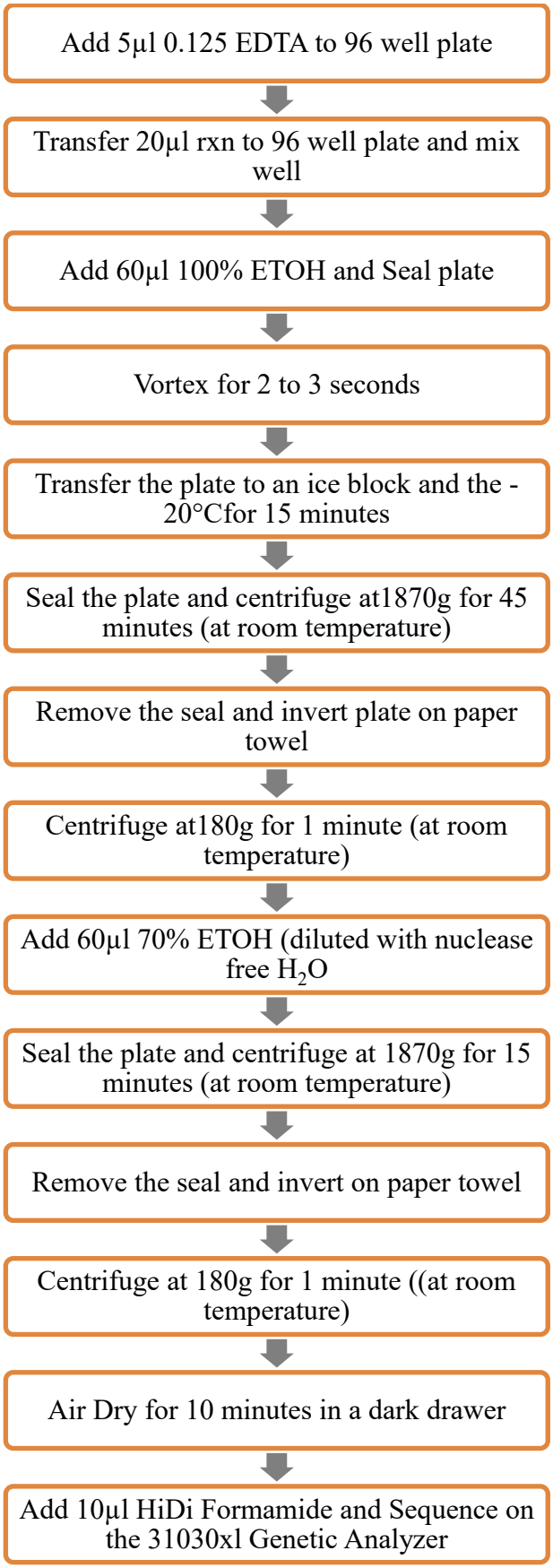
Protocol

Step	Temperature (°C)	Time
Initial Denaturation	96	5 Minutes
Denaturation	96	30 Seconds
Primer Annealing	50	15 Seconds
Elongation	60	4 Minutes

Cycle sequencing proceeds with 30 cycles of the denaturation primer annealing and elongation steps.

Ethanol and EDTA Precipitation

20µl sequencing product or 10 µl sequencing product diluted with 10µl Sabax water



x

Appendix 6: 179 hearing impairment associated genes investigated

Number	Gene name	Number	Gene name3	Number	Gene name	Number	Gene name
1	ACTG1	**51**	EPS8	**101**	MT-TK	**151**	SLC26A5
2	ADCY1	**52**	EPS8L2	**102**	MT-TL1	**152**	SLITRK6
3	ADGRV1	**53**	ESPN	**103**	MT-TL2	**153**	SMC1A
4	ADGRV1	**54**	ESRRB	**104**	MT-TS1	**154**	SMPX
5	AIFM1	**55**	EYA1	**105**	MYH14	**155**	SNAI2
6	ALMS1	**56**	EYA4	**106**	MYH9	**156**	SOX10
7	APC	**57**	F2	**107**	MYO15A	**157**	SPRY4
8	ATP1A3	**58**	FAM65B	**108**	MYO3A	**158**	STRC
9	ATP2B2	**59**	FBN1	**109**	MYO6	**159**	SYNE4
10	ATP6V1B1	**60**	FGF3	**110**	MYO7A	**160**	TBC1D24
11	BDP1	**61**	FGFR1	**111**	NARS2	**161**	TBX1
12	BSND	**62**	FGFR2	**112**	NDUFA1	**162**	TCOF1
13	C10orf2	**63**	FGFR3	**113**	NLRP3	**163**	TECTA
14	CABP2	**64**	FOXI1	**114**	OPA1	**164**	TECTB
15	CACNA1D	**65**	GATA3	**115**	OSBPL2	**165**	TIMM8A
16	CCDC50	**66**	GBE1	**116**	OTOA	**166**	TJP2
17	CD164	**67**	Gene(s)	**117**	OTOF	**167**	TMC1
18	CDC14A	**68**	GIPC3	**118**	OTOG	**168**	TMEM132E
19	CDH23	**69**	GJB2	**119**	OTOGL	**169**	TMIE
20	CDH23	**70**	GJB3	**120**	P2RX2	**170**	TMPRSS3
21	CDK5RAP2	**71**	GJB6	**121**	PAFAH1B1	**171**	TNC
22	CEACAM16	**72**	GPSM2	**122**	PAX3	**172**	TPRN
23	CIB2	**73**	GRHL2	**123**	PCDH15	**173**	TRIOBP
24	CISD2	**74**	GRXCR1	**124**	PDZD7	**174**	TSPEAR

25	CLDN14	**75**	GRXCR2	**125**	PEX1	**175**	USH1C
26	CLIC5	**76**	HARS2	**126**	PEX6	**176**	USH1G
27	CLPP	**77**	HGF	**127**	PJVK	**177**	USH2A
28	CLRN1	**78**	HOMER2	**128**	PMP22	**178**	WFS1
29	COCH	**79**	HSD17B4	**129**	PNPT1	**179**	WHRN
30	COCH	**80**	ILDR1	**130**	POLG		
31	COL11A1	**81**	KARS	**131**	POLR1C		
32	COL11A2	**82**	KCNE1	**132**	POLR1D		
33	COL2A1	**83**	KCNJ10	**133**	POU3F4		
34	COL4A3	**84**	KCNQ1	**134**	POU4F3		
35	COL4A4	**85**	KCNQ4	**135**	PRF1		
36	COL4A5	**86**	KITLG	**136**	PRPS1		
37	COL4A6	**87**	LARS2	**137**	PRRT2		
38	COL9A1	**88**	LHFPL5	**138**	PTPRQ		
39	COL9A2	**89**	LOXHD1	**139**	PTRH2		
40	CRYM	**90**	LOXL3	**140**	RAI1		
41	DCDC2	**91**	LRTOMT	**141**	RDX		
42	DFNA5	**92**	MARVELD2	**142**	ROR1		
43	DIABLO	**93**	MBTPS2	**143**	S1PR2		
44	DIAPH1	**94**	MCM2	**144**	SERPINB6		
45	DIAPH3	**95**	MET	**145**	SIX1		
46	DSPP	**96**	MIR96	**146**	SIX5		
47	DYNC1H1	**97**	MITF	**147**	SLC17A8		
48	EDN3	**98**	MSRB3	**148**	SLC22A4		
49	EDNRB	**99**	MT-CO1	**149**	SLC26A4		
50	ELMOD3	**100**	MT-RNR1	**150**	SLC26A4		