

## GENERAL ARTICLE

# Whole exome sequencing reveals pathogenic variants in *MYO3A*, *MYO15A* and *COL9A3* and differential frequencies in ancestral alleles in hearing impairment genes among individuals from Cameroon

Ambroise Wonkam<sup>1,2,\*</sup>, Noluthando Manyisa<sup>1,†</sup>, Christian D. Bope<sup>3</sup>, Collet Dandara<sup>1</sup> and Emile R. Chimusa<sup>1</sup>

<sup>1</sup>Division of Human Genetics, Department of Pathology, University of Cape Town, Cape Town 7925, South Africa, <sup>2</sup>Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town 7925, South Africa and <sup>3</sup>Department of Mathematics and Department of Computer Science, Faculty of Sciences, University of Kinshasa, Kinshasa, Democratic Republic of Congo

\*To whom correspondence should be addressed at: Division of Human Genetics, Department of Pathology and Department of Medicine and Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa. Tel: 0027 21 406 63 07; Fax: 0027 21 406 68 26; Email: ambroise.wonkam@uct.ac.za

## Abstract

There is scarcity of known gene variants of hearing impairment (HI) in African populations. This knowledge deficit is ultimately affecting the development of genetic diagnoses. We used whole exome sequencing to investigate gene variants, pathways of interactive genes and the fractions of ancestral overderived alleles for 159 HI genes among 18 Cameroonian patients with non-syndromic HI (NSHI) and 129 ethnically matched controls. Pathogenic and likely pathogenic (PLP) variants were found in *MYO3A*, *MYO15A* and *COL9A3*, with a resolution rate of 50% (9/18 patients). The study identified significant genetic differentiation in novel population-specific gene variants at *FOXD4L2*, *DHRS2L6*, *RPL3L* and *VTN* between HI patients and controls. These gene variants are found in functional/co-expressed interactive networks with other known HI-associated genes and in the same pathways with *VTN* being a hub protein, that is, focal adhesion pathway and regulation of the actin cytoskeleton ( $P$ -values  $<0.05$ ). The results suggest that these novel population-specific gene variants are possible modifiers of the HI phenotypes. We found a high proportion of ancestral allele versus derived at low HI patients-specific minor allele frequency in the range of 0.0–0.1. The results showed a relatively low pickup rate of PLP variants in known genes in this group of Cameroonian patients with NSHI. In addition, findings may signal an evolutionary enrichment of some variants of HI genes in patients, as the result of polygenic adaptation, and suggest the possibility of multigenic influence on the phenotype of congenital HI, which deserves further investigations.

<sup>†</sup>Ambroise Wonkam, <http://orcid.org/0000-0003-1420-9051>

<sup>‡</sup>These two authors have equally contributed to the manuscript.

Received: June 3, 2020. Revised: October 1, 2020. Accepted: October 12, 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

Hearing impairment (HI) is a common sensory impairment, affecting nearly 500 million people worldwide (1). The burden of HI has worsened globally and is the leading cause of disability (2), with a projected increasing trend (3). More than 80% of people affected by HI live in low- and middle-income countries. The highest incidence rate is found in the sub-Saharan Africa (SSA): up to 6 per 1000 compared with about 1 per 1000 in Europeans or North Americans (4).

HI may be due to genetic, environmental and/or unknown factors. About 50% of congenital HI cases in high-income countries are due to genetic causes (5). The overshadowing environmental factors as causes of HI in SSA has meant that searching for genetic causes has not been prioritized (6–9). For example, in Cameroon, in a total of 582 patients with HI, with limited molecular investigations, the environmental factors were estimated at 52.6%, the genetic causes at 14.8% and unknown causes, some of which could be genetic, were estimated at 32.6% (8). Non-syndromic HI (NSHI) accounts for approximately 70% of genetic HI (5) and is described as the condition where HI occurs without any additional clinical manifestations. Autosomal recessive inheritance accounts for the majority (77–80%) of NSHI (5,10,11).

Variants in more than 150 genes have been reported as possible causes of HI (12). Mutations in the connexin genes *GJB2* and *GJB6* have been shown to account for most of genetic-associated HI among European, Asian and North American populations (13–15). However, a few studies investigated the genetic cause of HI among Africans (7,9,16–22), with most of them showing that *GJB2* and *GJB6* may not always be the important contributors, with the exception of Ghana, where the *GJB2*-R143W founder mutation accounts for 25.9% of HI cases among families segregating autosomal recessive NSHI (ARNSHI) (23).

Advanced genomic technologies, particularly targeted enrichment genome analysis by next-generation sequencing (NGS), have now made it possible for efficient determination of other genetic contributors without necessarily understanding the pathways involved. However, NGS has demonstrated a consistently low pickup rate for isolated cases of HI in individuals of African ancestry (24,25). This pickup rate was slightly improved in a modest group of multiplex families segregating HI from Cameroon (26). Hence, there is a great scarcity in the representation of known pathogenic gene

variants of HI in populations of African ancestry. Indeed, a recent study of pathogenic and likely pathogenic (PLP) ARNSHI variants (selected from the ClinVar and Deafness Variation databases with their frequencies from gnomAD database) estimated the prevalence of HI due to PLP as 5.2 per 100 000 individuals for Africans, compared with a highest prevalence of 96.9 per 100 000 individuals for Ashkenazi Jews (27). The knowledge deficit of genetic causes of HI among African populations is likely hindering our current understanding of the mechanism of HI, the refinement of gene–disease pairs and clinical validity curation and is ultimately affecting the development of genetic diagnoses and counseling, therapeutics and prognosis. The efficiency of using whole exome sequencing (WES) in revealing PLP and in exploring conservative evolution in HI genes in sub-Saharan Africans has been rarely investigated.

This study used WES to first identify *in silico* PLP variants and then examine the fractions of ancestral to derived alleles for 159 known HI-associated genes (Supplementary Material, Table S1) implicated in a group of Cameroonian patients with NSHI and controls with the same ethnolinguistic background. Our study identifies PLP variants in known genes among patients and reveals that rare variants within known HI-associated genes may have evolved conservatively to play critical roles in HI phenotypic variation among African HI patients, suggesting the possibility of multigenic and polygenic adaptation influencing congenital HI.

## Results

### Patients' information and variant discovery

The demographics of patients and controls are summarized in Table 1. There was an almost equal gender distribution of patients within this study (52.6% of study participants were male). The mean age was 12 years ( $\pm 4.4$ ) in the patients' population and 26 years ( $\pm 13.1$ ) for the controls group. All the patients presented with non-syndromic congenital/prelingual onset of HI, with the majority (nearly 90%) having a profound HI (91 dB and above).

Our results from WES variants calling yielded 7 158 983 variants that passed quality control (Supplementary Material, Fig. S1) of which 80 226 were exonic single nucleotide

**Table 1.** Demographics of patients with HI and controls

Demographics	Categories	n (%) Patients	N (%) Controls
Sex	Male	10 (52.6)	53 (41.1)
	Female	8 (47.4)	73 (56.6)
	Blank		3 (2.3)
Age of onset	Congenital	17 (89.5)	
	Prelingual	2 (10.5)	
	Blank		129 (100.0)
Degree of HI left	Severe	1 (5.3)	
	Profound	17 (89.4)	
	Blank <sup>a</sup>	1 (5.3)	129 (100.0)
Degree of HI right	Severe	2 (10.5)	
	Profound	16 (84.2)	
	Blank <sup>a</sup>	1 (5.3)	129 (100.0)
Transmission	Autosomal recessive	18 (100)	
	Not applicable		129 (100.0)

<sup>a</sup>No data available.

Table 2. Candidate variants associated with HI in Cameroonian patients

No. of patients (age in years)	Family pedigree	Genotype	Gene	Region	SNP	cDNA change	Protein change	Functional effect <sup>a</sup>	ExAC_AFR MAF	ExAC_ASI MAF	Cameroonian controls MAF (N = 129)
3 (10, 10, 16)	SP, SP, MP	Homozygous	MYO3A	10p12.1	rs140301218c.C424T		p.H142Y	Loss of function of protein kinase domain	0.0026	0.0007	0
4 (12, 8, 10, 10)	MP, MP, SP, MP	Homozygous	MYO15A	17p11.2	rs138861831c.4888 T		p.R1630C	Loss of function of myosin head-like domain	0.0083	0.0003	0
2 (16, 11)	SP, MP	Homozygous	COL9A3	20q13.33	rs145035835c.C406A		p.G136S	Loss of function of the triple-helical region 3	0	0.001	0

<sup>a</sup> Exonic non-synonymous 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. Abbreviations: SP, simplex; MP, multiplex family; SNP, single nucleotide polymorphism; ExAC, exome aggregation consortium; AFR, African; AS, Asian; MAF, minor allele frequency

polymorphisms (SNPs), distributed as 3.6% stop loss, 2.4% stop gain, 4.4% synonymous and 94% non-synonymous or splice site variants.

### In silico putative deleterious variants in known genes among HI patients

Upon filtering variants, results indicated PLP variants in 9/18 patients (Table 2).

In addition, putative PLP variants in *USH2A*, *HSD17B4* and *MYO1A* were identified in various patients (Supplementary Material, Table S2), but careful consideration of phenotype to genotype associations, in addition to the recent expert curation of the clinical validity of 164 HI gene–disease pairs by Clinical Genome (ClinGen) (28), allowed us to exclude them as potential candidates. Indeed, the variant found in *USH2A* (MIM:276901) is associated with Usher syndrome type IIA and was defined as benign; *HSD17B4* (MIM:601860) is associated with Perrault syndrome (MIM:233400) that has a clinical presentation not found in the patients; while *MYO1A* has been excluded as a HI-associated gene (28).

Considering all HI patients and controls together, we identified novel population-specific candidate variants in four genes: *FOXD4L6*, *DHRS4L2*, *RPL3L* and *VTN* (Supplementary Material, Table S3), which were not previously associated with HI or any other human conditions.

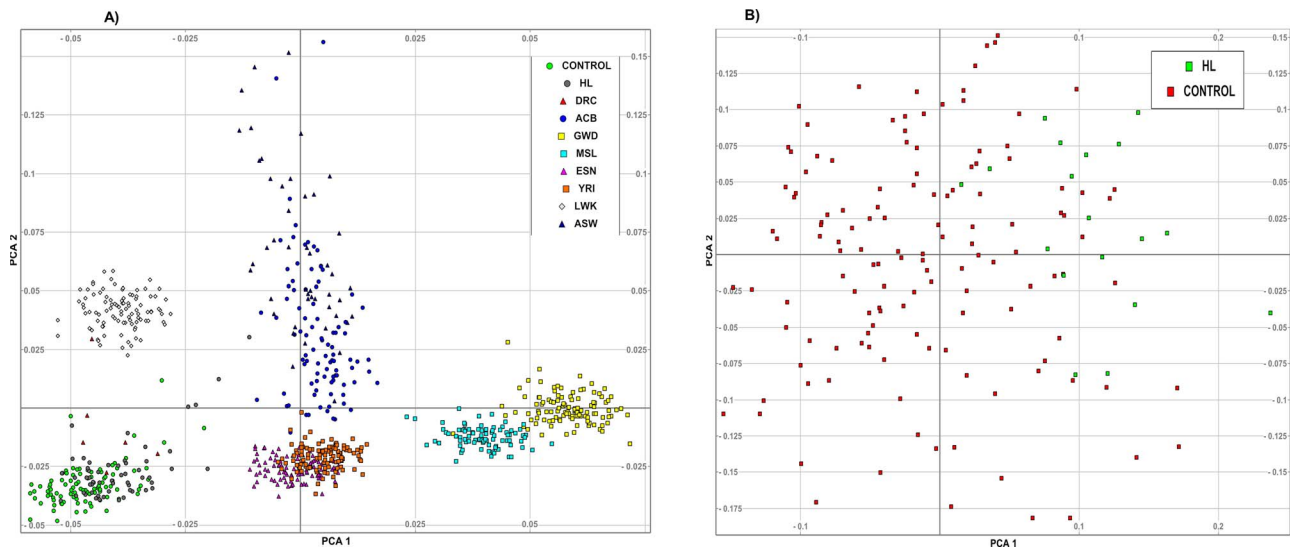
### Population structure

Principal component analysis (PCA) was performed with data from Cameroonian controls and HI patients along with those of eight other African populations (Fig. 1). Cameroonian control and patient groups cluster together in the lower left corner of the plot in Figure 1A, confirming some level of population homogeneity. Regarding the second PCA, Figure 1B represents data from Cameroonian controls and patients and indicates the inherent diversity within this population.

### Genetic differentiation between Cameroon hearing-impaired patients and controls

Following WES, aggregated SNP frequencies within 159 known HI genes were analyzed for unusual differences between the patients and controls using Fisher's combined probability (see Materials and Methods). The significance level for difference between cases and controls had a *P*-value of <0.05/159, which is  $3.1 \times 10^{-4}$ , and the permutation adjusted *P*-value was also obtained after 100 000 permutations. We investigated frequencies difference in population-specific gene variants, using an approach that is similar to gene burden analysis (29), a well-calibrated approach that accounts for small-to-moderate sample size and rare variants. The results are summarized in Table 3, where up to 26 genes were identified for the differences between patients and controls, indicating potential genetic differentiation (*P*-value < 0.05).

Importantly, two identified *in silico* candidate variants in Supplementary Material, Table S2, including *RPL3L* (*P*-value =  $3.1 \times 10^{-2}$ ) and *VTN* (*P*-value =  $6.1 \times 10^{-4}$ ), additionally indicated genetic differentiation between HI patients and controls (Table 3). These gene variants have not been previously associated with HI. The samples' quality at these gene regions was checked back in the Binary Alignment Map (BAM) files (Supplementary Material, Fig. S2).



**Figure 1.** PCA plot for nine African populations. (A) The confirmation of the origin of the control and patient groups from the same ethnolinguistic background, which cluster in the lower left corner of the plot and are represented by the green and gray circles for the controls and patients groups, respectively. The inherent diversity of the individuals within the patients and controls groups is represented in (B), where the red squares indicate the controls group and the green squares represent the HI patients group. DRC: Demographic Republic of Congo (red triangle); ABC: African Caribbean in Barbados (blue circle); GWD: Gambian in Western Division –Mandinka (yellow square); MLS: Mende in Sierra Leone (light blue square); ESN: Esan in Nigeria (purple triangle); YRI: Yoruba in Ibadan, Nigeria (orange square); LWK: Luhya in Webuye, Kenya (grey diamond); ASW: African Ancestry in Southwest United States (blue triangle).

**Table 3.** Genetic difference between gene variants in the hearing-impaired patients and controls

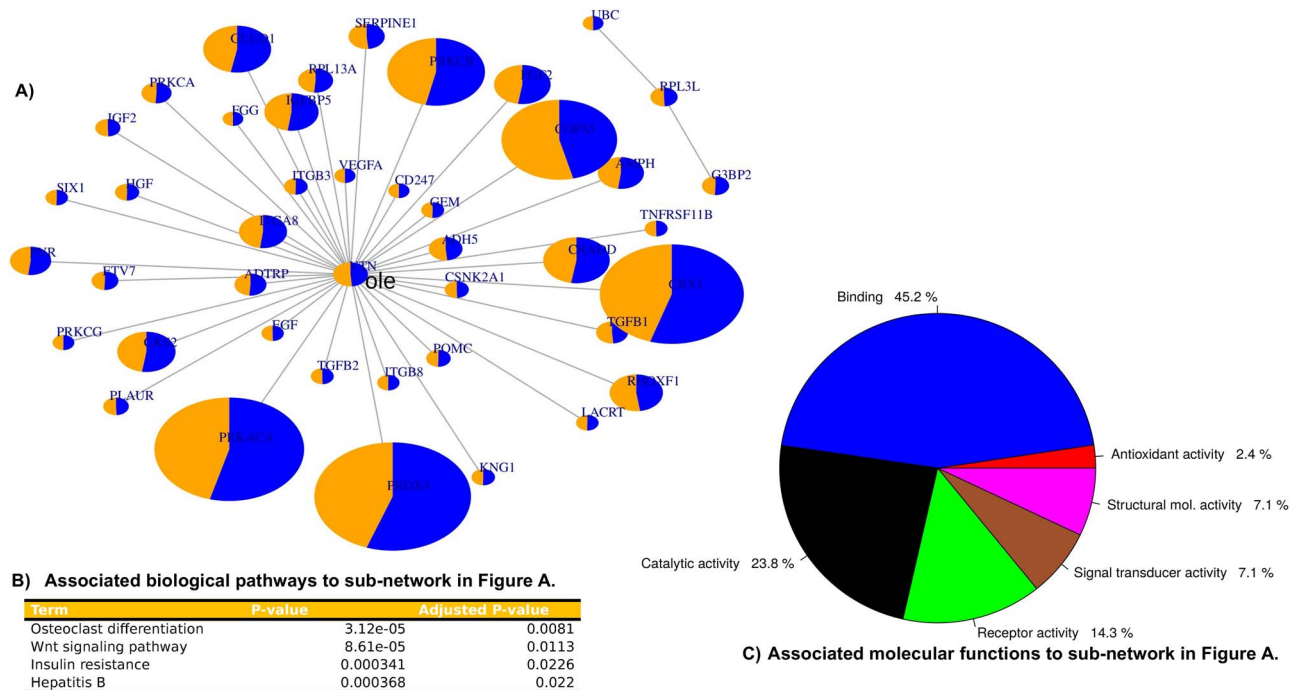
Gene	Gene-specific frequency HI	Gene-specific frequency CONTROL	P-values	Adjusted P-value
PRDX4	0.2109	0.2618	2.65e-31	1.17e-08
PRKACA	0.2716	0.3211	9.71e-30	4.34e-08
CBX1	0.2242	0.2726	1.83e-28	9.07e-08
COPSS5	0.2908	0.2488	4.35e-22	1.07e-07
PRKCB	0.2540	0.2916	4.52e-18	2.47e-07
GLUD1	0.2251	0.2538	2.36e-11	3.77e-06
CRADD	0.2571	0.2853	4.66e-11	2.67e-06
ITGB6	0.2426	0.2687	1.11e-09	3.47e-04
CKS2	0.2500	0.2753	3.40e-09	1.03e-04
FGF2	0.2376	0.2623	8.16e-09	1.11e-04
IGFBP5	0.2561	0.2799	2.68e-08	8.17e-04
RHOXF1	0.2655	0.2422	4.51e-08	7.27e-04
ITGA8	0.2559	0.2770	6.81e-07	4.47e-04
AMPH	0.2404	0.2606	1.91e-06	6.17e-04
PVR	0.2336	0.2515	2.43e-05	4.27e-03
SERPINE1	0.2513	0.2360	3.01e-04	2.17e-03
RPL13A	0.2585	0.2730	5.66e-04	1.07e-03
VTN	0.2601	0.2457	6.14e-04	1.11e-03
ADH5	0.2346	0.2209	1.16e-03	2.12e-02
ADTRP	0.2690	0.2816	2.77e-03	2.37e-02
TGFB1	0.2553	0.2428	2.84e-03	4.47e-02
PRKCA	0.2598	0.2710	7.42e-03	3.17e-02
RPL3L	0.2577	0.2487	3.13e-02	1.01e-01
G3BP2	0.2306	0.2396	3.23e-02	1.07e-01
ETV7	0.2570	0.2657	3.84e-02	1.09e-01
PLAUR	0.2668	0.2593	7.45e-02	1.17e-01

Genes that have been previously associated with HI. All genes above the red line have a significant genetic difference between patients and controls. Table indicates aggregated frequencies for SNPs in each gene. P-value was assessed using Fisher's combined probability

### Protein–protein interactions and community network enrichment analysis

The four identified genes (VTN, RPL3L, FOXD4L6 and DHRS2L6) with population-specific candidate variants were queried for

their functional or co-expressed interactive subnetwork, and therefore, the associated pathways and molecular functions of the resulting subnetwork. The protein products of VTN and RPL3L are indicated to functionally interact in 42 other protein–protein



**Figure 2.** Reconstituted HI PPI network including enrichment analysis of the associated pathways and molecular functions. Figure 2 indicates PPI 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. FOXD4L6 and DHRS4L2 have no interactions with other HI-associated proteins, as indicated in (A). The interactions were queried in the Enrichr database that indicated the major pathways associated, which includes osteoclast differentiation, as indicated in (B). (C) The associated molecular function, which includes binding and catalytic activities.

pairings (Fig. 2). The subnetwork in Figure 2A was shown to be significantly associated with four pathways including osteoclast differentiation ( $P$ -value =  $3.2 \times 10^{-5}$ ), the networks interactions revealed major associated pathways, which notably included osteoclast differentiation Figure 2B or catalytic activity Figure 2C.

Moreover, the interactions of the protein products of VTN, RPL3L, FOXD4L6 and DHRS2L6 with the other 159 HI-associated genes (Supplementary Material, Table S1) were investigated using a community network analysis. The analysis was performed by combining the candidate genes with the 159 known HI-associated genes present on the ClinVar database (30) and the Deaf Variation database (12). VTN is a hub protein for one of the HI-associated pathways along with nine other genes. The community network analysis provided eight protein subnetworks, which had 10 hub proteins, as indicated in Figure 3A. The associated pathways and molecular functions are indicated in Figures 3B and C, 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 yet been implicated in HI. The entire community network is associated with the focal adhesion pathway ( $P$ -value =  $5.3 \times 10^{-35}$ ) and regulation of the actin cytoskeleton ( $P$ -value =  $1.3 \times 10^{-42}$ ), as indicated in Figure 3B. The abundant molecular processes associated with the network are binding and catalytic activity, as indicated in Figure 3C.

#### Molecular dynamic simulations for mutated variants

Molecular dynamic (MD) simulations (Fig. 4) showed that for COL9A, substitution of the uncharged non-polar amino acid GLY136 with the uncharged polar amino acid SER136 renders the protein structure inflexible and may impact binding interactions, stability and conformation of the protein. For MYO3A,

substitution of the positively charged and hydrophilic amino acid HIS142 with the uncharged polar and hydrophobic amino acid TRP142 could impact binding interactions and the stability of structure and conformation of the protein (Fig. 5).

#### Evolutionary adaptation of human hearing: proportion of ancestral and derived alleles

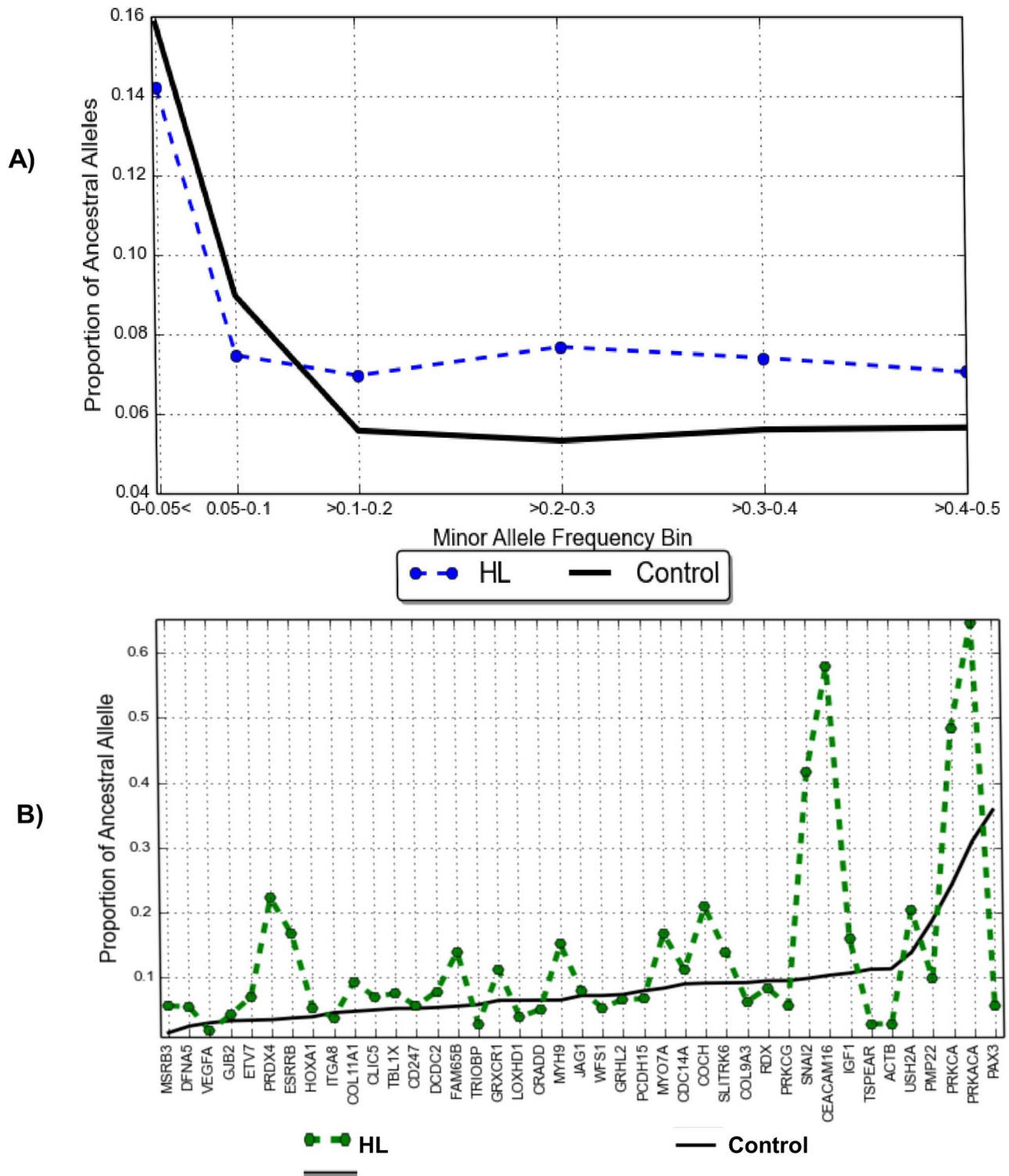
Figure 4A demonstrates that the HI patients have a lower proportion of derived alleles and therefore a higher proportion of ancestral alleles at low minor allele frequencies (0.00–0.1 in Fig. 4A), suggesting that rare variants may have evolved conservatively to play a critical role in HI variable expression. The figure further shows that the control population has a higher proportion of ancestral alleles at higher minor allele frequencies. The analysis, furthermore, indicates a low proportion of ancestral alleles in *GJB2* among patients, which translates into a high proportion of derived alleles within the gene (Fig. 4B).

#### Discussion

To the best of our knowledge, this is a rare study that utilizes WES to investigate HI in a SSA population. We identified very rare PLP ARNSHI variants in three known genes (*MYO3A*, *MYO15A* and *COL9A3*) (Table 2). *MYO3A* was first associated with progressive ARNSHI in a multigenerational Israeli family (31) and has since been associated with progressive HI in mice (32). A recessive form of HI with pathologic variants in *MYO3A* has also been reported in a Chinese family (33). On the other side, variants in *MYO3A* have been associated with autosomal dominant progressive HI in an African American family (34) and among individuals



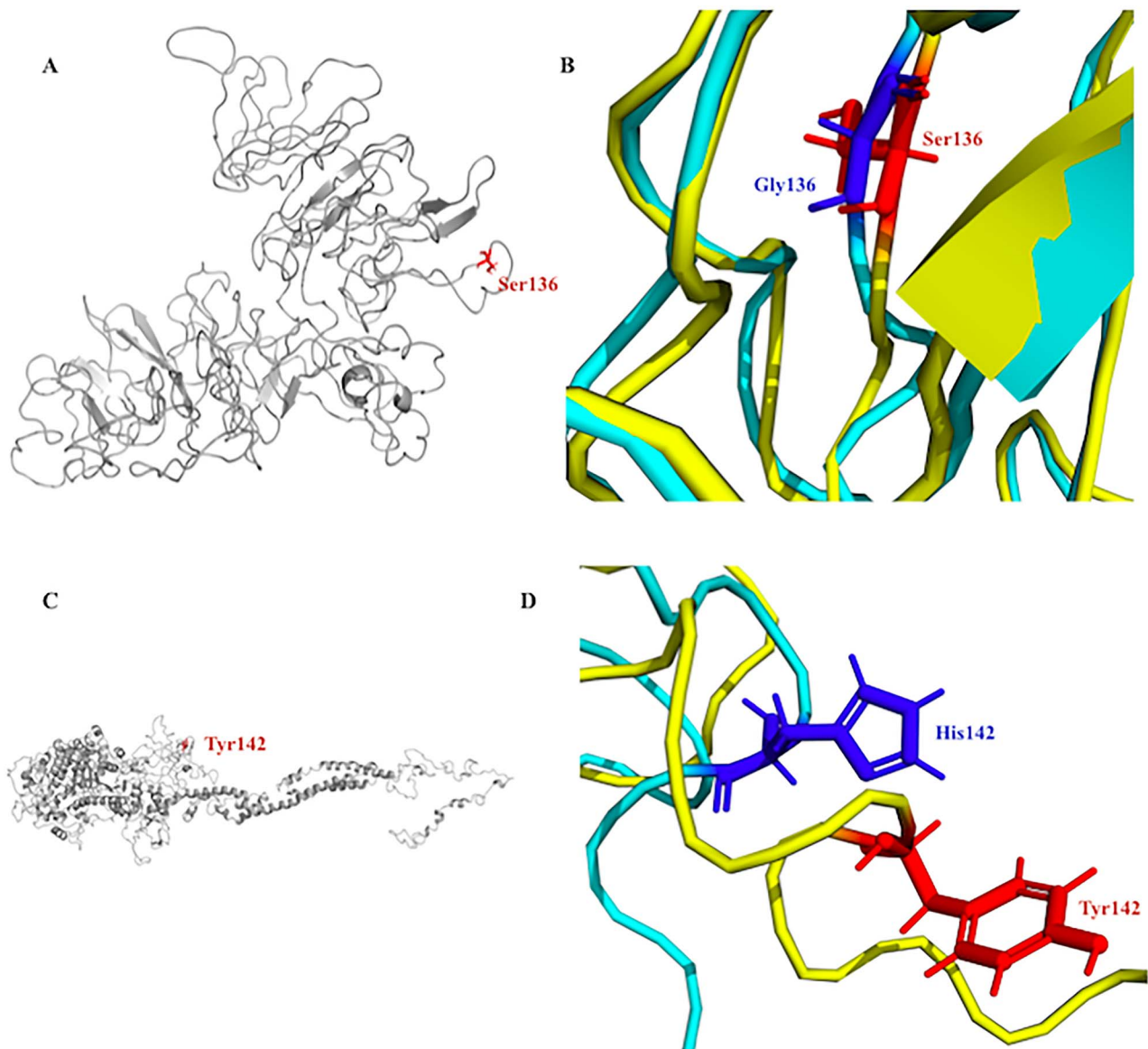




**Figure 4.** Proportion of ancestral alleles in the Cameroonian patients with HI and controls group. (A) Analysis indicates a higher proportion of ancestral alleles for rare variants (MAF of 0.00-0.1) in the controls groups as compared with the patients group. (B) Distribution of ancestral alleles proportion based on known HI genes. HL: Hearing Loss.

alleles in known HI genes at low minor frequencies [the 0.00-0.10 minor allele frequency (MAF) bins]. To our knowledge, this analysis is the first to explore such an approach in HI research. This result suggests that rare variants may have evolved conservatively from the evolutionary lineage and supports the possibility of multigenic and polygenic adaptation

influencing congenital HI. A point of interest is that the proportion of derived alleles in *GJB2* contrasts the generally low implication of this gene in HI in most African populations (9,17-19,21,47) and particularly in Cameroonian families segregating ARNSHI (20). This finding, that is, the high proportion of *GJB2*-derived alleles among HI patients could be due to



**Figure 5.** Molecular modeling scheme of the COL9A p.Gly136Ser and MYO3A p.His142Tyr mutants. (A) COL9A protein with mutated residue (SER136) colored red. The non-conservative substitution of uncharged non-polar GLY136 amino acid with uncharged polar amino acid SER136 renders the protein structure inflexible and may impact the binding interactions, stability and conformation of the protein. (B) Enlargement of the COL9A mutation site comparing the configuration of the wild type (colored cyan) and mutant (colored yellow) protein and illustrating the flexibility of the structures. (C) MYO3A protein product showing the mutated residue (TYR142) colored red. The substitution of the positively charged and hydrophilic amino acid HIS142 with the uncharged polar and hydrophobic amino acid TRP142 could impact the binding interactions and the stability of structure and conformation of the protein. (D) Enlargement of the MYO3A mutation site comparing the configuration of the wild type (colored cyan) and mutant (colored yellow) protein and illustrating the flexibility of the structures.

*GJB2* being a hypervariable gene, suggests possible mutation episodes throughout its evolution. Alternatively, in some cases of congenital HI, variants in *GJB2* could be modifiers in the HI phenotypes, possibly in combination with other HI gene variants in a multigenic model. Therefore, this study supports future exploration of the possible polygenic nature of some cases of congenital HI. The possibility of polygenic HI has been previously shown using GWAS that revealed numerous suggestive loci in the most common form of deafness: adult-onset progressive HI (48–51). Future studies in a larger cohort of HI patients and controls from multiple settings in Africa are needed to validate our findings.

The present study has a few limitations: first, the modest sample size of the HI individuals. Indeed, with the extreme

genetic heterogeneity, larger sample sizes are needed to capture the genetic profile that is representative of the population of Cameroon. Second, the study was performed on isolated HI cases and without the possibility of variant segregation studies within the family or in trio. Lastly, purposely designed future functional analysis should interrogate the effects of PLP on protein production and functions as well as explore the hypothesis of multigenic causes in congenital HI, using appropriate cellular or animal models. We used multiple tools instead of direct Sanger sequencing to validate variants, which is possibly another strength of this study. Indeed, studies have shown that a single round of Sanger sequencing is more likely to incorrectly refute a true-positive variant from NGS than to correctly identify a false-positive variant from NGS (52,53). Therefore, our



preliminary data from an understudied African population have provided novel insights into the genomic architecture of HI.

## Conclusion

This WES investigation identified PLP variants associated with HI in 9/18 patients in three known genes: *MYO3A*, *MYO15A* and *COL9A3*. We additionally reported novel population-specific variants in four genes (*FOXD4L6*, *DHRS4L2*, *RPL3L* and *VTN*) of which *RPL3L* and *VTN* interacted with other HI genes, with *VTN* being a hub protein, suggesting a possible modifier action in the HI phenotypes. Lastly, we reported for the first time, differential frequencies of ancestral alleles versus derived alleles in known HI genes among patients versus controls. These findings may signal an evolutionary enrichment of some variants of HI genes in patients as the result of polygenic adaptation and suggest the possibility of multigenic influence on the phenotype of congenital HI, which deserves further investigations.

## Materials and Methods

### Ethical approval

The study was performed in accordance with the Declaration of Helsinki. It was granted ethics approval by the Cameroon National Ethics Committee (ethics approval N°123/CNE/SE/2010 and N°033/CNE/DNM/07) as well as the University of Cape Town Human Research Ethics Committee (ethics approval HREC REF: 455/2014, and HREC REF: 132/2010). Written informed consent was obtained from patients 18 years or older or from parents/guardians for minors, accompanied by verbal assent from the minor.

### Patients and controls

Patients were recruited from various schools for the deaf, ear, nose and throat (ENT) clinics in Cameroon. The patients were examined by qualified medical geneticists and ENT specialists. Detailed family and medical histories were obtained from the patients and their parents. In this study, 18 Cameroonian patients presenting with NSHI of either putative genetic origin (revealed by one or more affected family members or consanguinity) or of unknown origin were selected from a cohort of 582 patients (8). All the patients were tested negative for *GJB2* and *GJB6* mutations as previously reported (16,17).

The control group consisted of 129 ethnically matched individuals without personal or familial history of HI and was recruited at Yaoundé Central and Douala Laquintinie Hospitals in Cameroon.

### Genomic characterization

WES was performed at Omega Bioservices (Norcross, GA, USA) on the DNA of 18 Cameroonian patients living with HI and 129 controls from the same ethnolinguistic background. DNA concentration was quantified using the QuantiFlour dsDNA System on a Quantas Fluorometer (Promega, Madison, WI, USA); 50 ng of genomic DNA was used for library preparation following an Illumina Nextera Rapid Capture Exome kit (Illumina, San Diego, CA, USA) that uses Nextera transposomes. The libraries were then hybridized with a 37 M probe pool to enrich the sequences and were then sent for WES using the Illumina HiSeq 2500 (Illumina), 100 bp run format, with an average read depth of 30X.

### Variant calling, annotation and prioritization

The WES data was subjected to quality control using FastQC (54) and SolexaQC++ (55). The sequencing reads were aligned to hg19, build37, (56) using Burrows-Wheeler Aligner (BWA) (57,58), and local realignment was performed using Genome Analysis Toolkit (GATK) (59). Variant calling was performed using an ensemble approach, whereby GATK3.0 (HaplotypeCaller) and SamTools (60) variant callers were utilized. The final variant call set was obtained by consensus using VariantMetaCaller (61). In addition, we excluded the deletion/insertion variants with breakpoints outside the coding region.

After high confidence variants were obtained with the VariantMetaCaller, the resulting Variant Call Format (VCF) file was split into two groups: 18 HI patients and 129 controls. ANNOtate VARiation (ANNOVAR) (62) was used to independently perform gene-based annotation in the two groups in order to catalog whether SNPs caused protein-coding changes and to identify the affected amino acids. ANNOVAR was used with the '2015Dec18' setting, where the population frequency for each variant was obtained on data extracted from the 1000 Genomes Project exomes and targeted exon datasets (63) and Catalog of Somatic Mutations in Cancer (COSMIC) (64). Genetic function was determined from RefGene (65,66) and different functional predictions were made from ANNOVAR's ljb\_all function. The variants set was filtered for exonic variants that were considered deleterious/damaging according to 10 different functional scores, including Sorting Intolerant From Tolerant (SIFT) (67–69), likelihood ratio test (LRT) (70), MutationTaster (71), MutationAssessor (72,73), FATHMM (74), Radial support vector machine (RadialSVM) and Likelihood Ratio (LR) (75), Combined Annotation Dependent Depletion (CADD) (76), GERP (77,78), PolyPhen2 (Poly morphism Phen otyping) (79), PhyloP and SiPhy (80). Additionally, conserved and segmental duplication sites, the dbSNP code (81) and clinical relevance reported in dbSNP138 (National Center for Biotechnology Information; (81)) were included. Thus, in the resulting annotated dataset, the variants were filtered based on whether they were: (1) rare, (2) exonic, (3) resulting in non-synonymous change, (4) resulting in a stop codon, (5) variants of known functional significance, (6) naturally selected and (7) likely deleterious (82,83). Variants that had a functional prediction status of either 'Deleterious' (D), 'probably damaging' (D), 'disease\_causing\_automatic' (A) or 'disease\_causing'(D) by all 10 functional score approaches were retained.

### Variant quality control and confirmation

We double-checked the presence and quality of variants discovered against all sample BAM files by using FastQC (54) to ensure that the reads of which the call was made or variants discovered from BAM files were of good quality. We ran FastQC on all final BAM files prior to the variant calling, then we aggregated the result from FastQC into a single report by using MultiQC (84). The summary evaluations of FastQC have resulted in six modules (Supplementary Material, Fig. S1). Furthermore, we have specifically rechecked the quality of reads in genomic regions of all our identified genes in Table 2 and generated the correlation matrix and principal components plot of all sample reads at those specific genomic regions (Supplementary Material, Fig. S2).

### Genetic differentiation and population structure

A statistical test of difference was performed to detect the possibility of unusual genetic difference between the hearing-

impaired patient and control groups. The aggregated SNP frequency for all SNPs in a given gene was computed. In this regard, SNPs were mapped to their associated genes using dbSNP database and differences in the aggregate SNP frequencies were determined using Fisher's combined probability (85).

Population structure was analyzed based on PCA using smartpca (86,87). The data from the patient and control groups were combined with data from 186 individuals of Yoruba origin (YRI) in Nigeria; 173 of Esan (ESN) in Nigeria; 280 from Western Divisions in Gambia (GWD); 116 Luhya (LWK) from Webuye, Kenya; 112 from African Ancestry in Southwest USA (ASW); 128 from Mende (MSL) in Sierra Leone and 123 of African Caribbean origin in Barbados (ACB) from the 1000 Genomes Phase3 (63). We performed a PCA the merged set of our cohort and that of seven African groups, and in a separate analysis on the merged data sets from patients and control. The first two components were plotted against each other using Genesis2.

### Fractions of ancestral and derived alleles

SNP ancestral alleles were downloaded from Ensembl and included 59 comparative alignments of 32 species (88,89). The SNPs were then checked for those present in the dbSNP database. The final patients and controls VCF files were annotated using the VCFtools 'fillOaa' script (90), with the ancestral allele recorded using the 'AA' INFO tag (89). The fraction of ancestral alleles for each SNP was computed by dividing the number of times the alternative alleles matched with ancestral alleles independently in patient and control datasets by the total number of copies of all alternative alleles across all samples for the particular SNP. The abovementioned fraction of ancestral alleles per SNP was aggregated at gene level. The fraction of derived alleles is equivalent to 1 minus the fraction of ancestral alleles. Previous studies have shown that derived alleles are more often minor alleles (<50% allele frequency) and are more often associated with risk as compared with the ancestral alleles (91).

The relationship between the fraction of derived allele and minor allele frequency in each gene was investigated. To do so, the alternative alleles were categorized into 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 derived alleles in each bin were independently computed. The fraction of ancestral/derived alleles for all known and candidate HI genes was further computed. This was done by aggregating the fraction of ancestral/derived alleles at SNP-based level to gene, taking into consideration all SNPs located within the gene's downstream or upstream region (92).

### Pathways enrichment analysis and protein–protein interactions

A comprehensive human protein–protein Interaction (PPI) network (46,92,93) was used to analyze how each of the gene variants are layered in a biological network, allowing us to extract a subnetwork. The association between gene variants in the subnetwork with human phenotypes and their potential biological pathways, processes and molecular functions was examined. This was performed using custom scripts in R (94), and the enrichment analysis was performed using Enrichr (95,96) and Panther (97,98) based on the identified gene variants.

To enable a community network analysis, the list of candidate gene variants was combined with those of the 159 known HI-associated genes (Supplementary Material, Table S1) obtained

from the Deafness Variation database (12) and ClinVar (30). A clustering script in R's igraph package (99) was used to produce a network plot, which would allow for identification of hub proteins in the subnetworks of HI genes.

### 3D protein structure prediction for functional characterization of novel variants

MD simulations were conducted to assess the effect of novel variants on protein functions. Amino acid sequences were obtained from UniProt (COL9A has UniProt protein identifier Q14050 and MYO3A has UniProt protein identifier Q8NEV4). The web server generated 1436 amino acids out of 1610 amino acids, with the missing structure of 174 amino acids on the tail end of the protein structure. The tertiary structures of the COL9A- and MYO3A-encoded proteins were generated using the I-tasser homology web server (100). All MD simulations were conducted with the GROMACS package, version 4.6.5 (101) using Amber (AMBER99SB-ILDN) force field ((102)). The system was solvated in a dodecahedron box of water. The temperature and pressure were maintained at 300 K using the Parrinello–Donadio–Bussi V-rescale thermostat (103) and a pressure of 1 kPa using the Berendsen barostat (104). The short-range non-bonded interactions were modeled using Lennard Jones potentials. The long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) algorithm (105,106). The LINC algorithm was used to constrain hydrogen bond lengths (107). Then, the velocities were assigned according to the Maxwell–Boltzmann distribution at 300 K. The equilibration of the structure NVT (constant number of particles, volume and temperature) and NPT (constant number of particles, pressure and temperature) was recorded for 10 ns each. Subsequently, a 10 ns NPT ensemble MD was conducted.

### Supplementary Material

Supplementary Material is available at HMG online.

### Acknowledgements

The authors thank members of the clinical unit of the Division of Human Genetics, University of Cape Town for contributing to data generations. Finally, the authors are grateful to patients and families who participated in this study.

*Conflict of Interest statement:* We declare that no competing interest exists.

### Funding

South African Medical Research Council's (SAMRC's) Self-Initiated Research (SIR); the National Health Laboratory Services (NHLS) start-up grant, South Africa; the Wellcome Trust/AAS Ref: H3A/18/001 to A.W.; the National Institutes of Health, USA (grant number U01HG009716 to A.W.).

### Authors' Contributions

N.M., E.C., C.B., C.D. and A.W. contributed to the conception and design of the study; N.M. and A.W. collected the data; N.M., E.C., C.B., C.D. and A.W. performed molecular analysis and interpretation of data; A.W. and N.M. wrote the first draft of the manuscript and E.C. wrote a section of the manuscript. All

authors contributed to the revisions and approved the submitted version.

## References

- WHO Media Centre. (2018). Deafness and hearing loss. Available at: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> [Accessed January 15, 2019].
- 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. et al. (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.
- Wilson, B.S., Tucci, D.L., Merson, M.H. and Donoghue, G.M.O. (2017) Review global hearing health care: new findings and perspectives. *Lancet*, **6736**, 31073–31075.
- Olusanya, B.O., Neumann, J. and Saunders, J.E. (2014) The global burden of disabling hearing impairment : a call to action. *Bull. World Health Organ.*, **92**, 367–373.
- Schrijver, I. (2004) Hereditary non-syndromic sensorineural hearing loss: transforming silence to sound. *J. Mol. Diagn.*, **6**, 275–284.
- Adadey, S.M., Awandare, G., Amedofu, G.K. and Wonkam, A. (2017) Public health burden of hearing impairment and the promise of genomics and environmental research: a case study in Ghana, Africa. *OMICS*, **21**, 638–646.
- Rudman, J.R., Kabahuma, R.I., Bressler, S.E., Feng, Y., Blanton, S.H., Yan, D. and Liu, X.Z. (2017) The genetic basis of deafness in populations of African descent. *J. Genet. Genomics*, **44**, 285–294.
- Wonkam, A., Noubiap, J.J.N., Djomou, F., Fieggen, K., Njock, R. and Toure, G.B. (2013) Aetiology of childhood hearing loss in Cameroon (sub-Saharan Africa). *Eur. J. Med. Genet.*, **56**, 20–25.
- Wonkam Tingang, E., Noubiap, J.J., Fokouo, J.V., Oluwole, O.G., Nguéfack, S., Chimusa, E.R. and Wonkam, A. (2020) Hearing impairment overview in Africa: the case of Cameroon. *Genes (Basel)*, **11**, 233.
- Chung, C.S., Robison, O.W. and Morton, N.E. (1959) A note on deaf mutism. *Ann. Hum. Genet.*, **23**, 357–366.
- Morton, N.E. (1991) Genetic epidemiology of hearing impairment. *Ann. N. Y. Acad. Sci.*, **630**, 16–31.
- The Molecular Otolaryngology and Renal Research Laboratories. The University of Iowa. (2016) Deafness Variation database. Available at: <http://deafnessvariationdatabase.org/> [Accessed June 2, 2017].
- Chan, D.K. and Chang, K.W. (2014) GJB2-associated hearing loss: systematic review of worldwide prevalence, genotype, and auditory phenotype. *Laryngoscope*, **124**, 34–53.
- Hutchin, T., Coy, N.N., Conlon, H., Telford, E., Bromelow, K., Blyndon, D., Taylor, G., Coghill, E., Brown, S., Trembath, R. et al. (2005) Assessment of the genetic causes of recessive childhood non-syndromic deafness in the UK-implications for genetic testing. *Clin. Genet.*, **68**, 506–512.
- 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. et al. (2002) The prevalence of connexin 26 (GJB2) mutations in the Chinese population. *Hum. Genet.*, **111**, 394–397.
- Bosch, J., Lebeko, K., Noubiap Nziale, J.J., Dandara, C., Makubalo, N. and Wonkam, A. (2014) 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.
- Bosch, J., Noubiap, J.J.N., Dandara, C., Makubalo, N., Wright, G., Domelevo Entfellner, J.-B., Tiffin, N. and Wonkam, A. (2014) Sequencing of GJB2 in Cameroonians and black South Africans and comparison to 1000 Genomes Project data support need to revise strategy for discovery of non-syndromic deafness genes in Africans. *OMICS*, **18**, 1–6.
- Gasmelseed, N.M.A., Schmidt, M., Magzoub, M.M.A., Macharia, M., Elmustafa, O.M., Ototo, B., Winkler, E., Ruge, G., Horstmann, R.D. and Meyer, C.G. (2004) Low frequency of deafness-associated GJB2 variants in Kenya and Sudan and novel GJB2 variants. *Hum. Mutat.*, **23**, 206–207.
- Kabahuma, R.I., Ouyang, X., Du, L.L., Yan, D., Hutchin, T., Ramsay, M., Penn, C. and 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.
- Wonkam, E.T., Chimusa, E., Noubiap, J.J., Adadey, S., M., Fokouo, J., V., F., Wonkam, A. (2019) GJB2 and GJB6 mutations in hereditary recessive non-syndromic hearing impairment in Cameroon. *Genes (Basel)*, **10**, 844.
- Trotta, L., Iacona, E., Primignani, P., Castorina, P., Radaelli, C., Bo, L.D., Coviello, D. and Ambrosetti, U. (2011) GJB2 and MTRNR1 contributions in children with hearing impairment from northern Cameroon. *Int. J. Audiol.*, **50**, 133–138.
- Wonkam, A., Bosch, J., Noubiap, J.J.N., Lebeko, K., Makubalo, N. and 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.
- Adadey S., M., Manyisa, N., Mnika, K., de Kock, C., Nembaware, V., Quaye, O., Amedofu G. K., Awandare, G., Wonkam, A. (2019) GJB2 and GJB6 mutations in non-syndromic childhood hearing impairment in Ghana. *Front. Genet.* **10**, 841.
- 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. et al. (2016) Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum. Genet.*, **135**, 441–450.
- Yan, D., Tekin, D., Bademci, G., Foster, J., Cengiz, F.B., Kannan-Sundhari, A., Guo, S., Mittal, R., Zou, B., Grati, M. et al. (2016) Spectrum of DNA variants for non-syndromic deafness in a large cohort from multiple continents. *Hum. Genet.*, **135**, 953–961.
- 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. et al. (2016) Targeted genomic enrichment and massively parallel sequencing identifies novel nonsyndromic hearing impairment pathogenic variants in Cameroonian families. *Clin. Genet.*, **90**, 288–290. [10.1111/cge.12799](https://doi.org/10.1111/cge.12799).
- Chakchouk, I., Zhang, D., Zhang, Z., Francioli, L.C., Santos-Cortez, R.L.P., Schrauwen, I. and Leal, S.M. (2019) Disparities in discovery of pathogenic variants for autosomal recessive non-syndromic hearing impairment by ancestry. *Eur. J. Hum. Genet.*, **27**, 1456–1465.
- DiStefano, M.T., Hemphill, S.E., Oza, A.M., Siegert, R.K., Grant, A.R., Hughes, M.Y., Cushman, B.J., 311 Azaiez, H., Booth, K.T., Chapin, A. et al. (2019) Clin gen expert clinical validity curation of 164 hearing loss 312 gene-disease pairs. *Genet. Med.*, **21**, 2239–2247.

29. Barnett, I.J., Lee, S. and Lin, X. (2013) Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol.*, **37**, 142–151.
30. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S. et al. (2016) Clin Var: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, **44**, D862–D868.
31. 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. USA*, **99**, 7518–7523.
32. Li, P., Wen, Z., Zhang, G., Zhang, A., Fu, X. and Gao, J. (2018) Knock-in mice with Myo 3a Y137C mutation displayed progressive hearing loss and hair cell degeneration in the inner ear. *Neural Plast.*, **2018**, 4372913.
33. Qu, R., Sang, Q., Xu, Y., Feng, R., Jin, L., He, L., Wang, L. (2016) Identification of a novel homozygous mutation in MYO3A in a Chinese family with DFNB30 non-syndromic hearing impairment. *Int J Pediatr Otorhinolaryngol.*, **84**, 43–37.
34. Grati, M., Yan, D., Raval, M.H., Walsh, T., Ma, Q., Chakchouk, I., Kannan-Sundhari, A., Mittal, R., Masmoudi, S., Blanton, S.H. et al. (2016) MYO3A causes human dominant deafness and interacts with Protocadherin 15-CD2 isoform. *Hum. Mutat.*, **37**, 481–487.
35. Dantas, V.G.L., Raval, M.H., Ballesteros, A., Cui, R., Gunther, L.K., Yamamoto, G.L., Alves, L.U., Bueno, A.S., Lezirovitz, K., Pirana, S. et al. (2018) Characterization of a novel MYO3A missense mutation associated with a dominant form of late onset hearing loss. *Soc. Rep.*, **8**, 8706.
36. Salime, S., Charif, M., Bousfiha, A., Elrharchi, S., Bakhchane, A., Charoute, H., Kabine, M., Snoussi, K., Lenaers, G., Barakat, A. et al. (2017) Homozygous mutations in PJK and MYO15A genes associated with non-syndromic hearing loss in Moroccan families. *Int. J. Pediatr. Otorhinolaryngol.*, **101**, 25–29.
37. Belguith, H., Aifa-Hmani, M., Dhoub, H., Said, M.B., Mosrati, M.A., Lahmar, I., Moalla, J., Charfeddine, I., Driss, N., Arab, S.B. et al. (2009) Screening of the DFNB3 locus: identification of three novel mutations of MYO15A associated with hearing loss and further suggestion for two distinctive genes on this locus. *Genet. Test. Mol. Biomarkers*, **13**, 147–151.
38. Budde, B.S., Aly, M.A., Mohamed, M.R., Breß, A., Altmüller, J., Motameny, S., Kawalia, A., Thiele, H., Konrad, K., Becker, C. et al. (2020) Comprehensive molecular analysis of 61 Egyptian families with hereditary nonsyndromic hearing loss. *Clin. Genet.*, **98**, 32–42.
39. Rehman, A.U., Bird, J.E., Faridi, R., Shahzad, M., Shah, S., Lee, K., Khan, S.N., Imtiaz, A., Ahmed, Z.M., Riazuddin, S. et al. (2016) Mutational Spectrum of MYO15A and the molecular mechanisms of DFNB3 human deafness. *Hum. Mutat.*, **37**, 991–1003.
40. Zhang, J., Guan, J., Wang, H., Yin, L., Wang, D., Zhao, L., Zhou, H. and Wang, Q. (2019) Genotype-phenotype correlation analysis of MYO15A variants in autosomal recessive non-syndromic hearing loss. *BMC Med. Genet.*, **20**, 60.
41. Asamura, K., Abe, S., Fukuoka, H., Nakamura, Y., Usami, S. (2005) Mutation analysis of COL9A3, a gene highly expressed in the cochlea, in hearing loss patients. *Auris Nasus Larynx*, **32**, 113–117.
42. Robin, N.H., Moran, R.T., Ala-Kokko, L. (2000) [updated 2017 Mar 16]. Stickler Syndrome. In: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Stephens, K., Amemiya, A. editors. *GeneReviews*® [Internet], Seattle (WA): University of Washington, Seattle; 1993–2020.
43. Hanson-Kahn, A., Li, B., Cohn, D.H., Nickerson, D.A., Bamshad, M.J., University of Washington Center for Mendelian Genomics, Hudgins, L. et al. (2018) Autosomal recessive stickler syndrome resulting from a COL9A3 mutation. *Am. J. Med. Genet. A*, **176**, 2887–2891.
44. Nixon, T.R.W., Alexander, P., Richards, A., McNinch, A., Bearcroft, P.W.P., Cobben, J. and Snead, M.P. (2019) Homozygous type IX collagen variants (COL9A1, COL9A2, and COL9A3) causing recessive Stickler syndrome-expanding the phenotype. *Am. J. Med. Genet. A*, **179**, 1498–1506.
45. Lebeko, K., Manyisa, N., Chimusa, E.R., Mulder, N., Dandara, C. and Wonkam, A. (2017) A genomic and protein-protein interaction analyses of nonsyndromic hearing impairment in Cameroon using targeted genomic enrichment and massively parallel sequencing. *OMICS*, **21**, 90–99.
46. Javidnia, H., Carson, N., Awubwa, M., Byaruhanga, R., Mack, D. and Vaccani, J.-P. (2014) Connexin gene mutations among ugandan patients with nonsyndromic sensorineural hearing loss. *Laryngoscope*, **124**, E373–E376.
47. Friedman, R.A., Van Laer, L., Huentelman, M.J., Sheth, S.S., Van Eyken, E., Corneveaux, J.J., Tembe, W.D., Halperin, R.F., Thorburn, A.Q., Thys, S. et al. (2009) GRM7 variants confer susceptibility to age-related hearing impairment. *Hum. Mol. Genet.*, **18**, 785–796.
48. Hoffmann, T.J., Keats, B.J., Yoshikawa, N., Schaefer, C., Risch, N. and Lustig, L.R. (2016) A large genome-wide association study of age-related hearing impairment using electronic health records. *PLoS Genet.*, **12**, 1–20.
49. Lewis, M.A., Nolan, L.S., Cadge, B.A., Matthews, L.J., Schulte, B.A., Dubno, J.R. et al. (2018) Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes. *BMC Med. Genet.*, **11**, 1–12.
50. Vuckovic, D., Dawson, S., Scheffer, D.I., Rantanen, T., Morgan, A., Di Stazio, M., Vozzi, D., Natile, T., Concas, M.P., Biino, G. et al. (2015) Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss. *Hum. Mol. Genet.*, **24**, 5655–5664.
51. Beck, T.F., Mullikin, J.C., NISC Comparative Sequencing Program and Biesecker, L.G. (2016) Systematic evaluation of Sanger validation of next-generation sequencing variants. *Clin. Chem.*, **62**, 647–654.
52. Mallawaarachchi, A.C., Hort, Y., Cowley, M.J., McCabe, M.J., Minoche, A., Dinger, M.E., Shine, J. and Furlong, T.J. (2016) Whole-genome sequencing overcomes pseudogene homology to diagnose autosomal dominant polycystic kidney disease. *Eur. J. H. Genet.*, **24**, 1584.
53. 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. et al. (2014). From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33.
54. Cox, M.P., Peterson, D.A. and Biggs, P.J. (2010) Solexa QA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform.*, **11**, 485.
55. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Gurdoo, L., Haeussler, M. et al. (2015) The UCSC genome



- browser database: 2015 update. *Nucleic Acids Res.*, **43**, D670–D681.
56. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760.
  57. Li, H., Ruan, J., Durbin, R., Li, H., Ruan, J. and 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. *Genome Res.*, **18**, 1851–1858.
  58. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. et al. (2010) The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
  59. Garrison, E. and Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. arXiv Prepr. arXiv, 1207.3907 [q-bio.GN], 9. doi:arXiv:1207.3907 [q-bio.GN].
  60. Gézsi, A., Bolgár, B., Marx, P., Sarkozy, P., Szalai, C. and Antal, P. (2015) Variant meta caller: automated fusion of variant calling pipelines for quantitative, precision-based filtering. *BMC Genom.*, **16**, 875.
  61. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
  62. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
  63. Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S. et al. (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.
  64. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. et al. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
  65. Pruitt, K.D. and Maglott, D.R. (2001) RefSeq and locus link: NCBI gene-centered resources. *Nucleic Acids Res.*, **29**, 137–140.
  66. Ng, P.C. and Henikoff, S. (2002) Accounting for human polymorphisms predicted to affect protein function. *Genome Res.*, **12**, 436–446.
  67. Ng, P.C. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, **31**, 3812–3814.
  68. Ng, P.C. and Henikoff, S. (2006) Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.*, **7**, 61–80.
  69. Fujita, A., Kojima, K., Patriota, A.G., Sato, J.R., Severino, P. and Miyano, S. (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.
  70. Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M. and Cai, L. (2012) Mutation Taster2: mutation prediction for the deep-sequencing age. *Nat. Methods*, **9**, 743–748.
  71. Reva, B., Antipin, Y. and Sander, C. (2007) Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.*, **8**, R232.
  72. Reva, B., Antipin, Y. and Sander, C. (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.*, **39**, 37–43.
  73. Shihab, H.A., Gough, J., Mort, M., Cooper, D.N., Day, I.N.M. and Gaunt, T.R. (2014) Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum. Genomics*, **8**, 11.
  74. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X. (2015) Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.*, **24**, 2125–2137.
  75. Kircher, M. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
  76. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S. and Sidow, A. (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.*, **15**, 901–913.
  77. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A. and Batzoglou, S. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.*, **6**. doi: 10.1371/journal.pcbi.1001025.
  78. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
  79. Garber, M., Guttman, M., Clamp, M., Zody, M.C., Friedman, N. and Xie, X. (2009) Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics*, **25**, 54–62.
  80. Sherry, S., Ward, M., Kholodov, M., Baker, J., Phan, L., Smigielski, E. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
  81. Li, M.X., Gui, H.S., Kwan, J.S.H., Bao, S.Y. and Sham, P.C. (2012) A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res.*, **40**, e53.
  82. Li, M.X., Kwan, J.S.H., Bao, S.Y., Yang, W., Ho, S.L., Song, Y.Q. and Sham, P.C. (2013) Predicting Mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet.*, **9**, 1–11.
  83. Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**, 3047–3048.
  84. Fisher, R. (1958) *Statistical Methods for Research Workers*, 4th edn. Oliver and Boyd, London.
  85. Patterson, N., Price, A.L. and Reich, D. (2006) Population structure and eigenanalysis. *PLoS Genet.*, **2**, 2074–2093.
  86. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
  87. Paten, B., Herrero, J., Fitzgerald, S., Beal, K., Flicek, P., Holmes, I. and Birney, E. (2008) Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.*, **18**, 1829–1843.
  88. The 1000 Genome Project Consortium. (2012) Ancestral allele reference sequence. Available at: [http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis\\_results/su%0AApporting/ancestral\\_alignments/%0A](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/su%0AApporting/ancestral_alignments/%0A).
  89. 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. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

90. Gorlova, O.Y., Ying, J., Amos, C.I., Spitz, M.R., Peng, B. and 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.
91. Chimusa, E.R., Mbiyavanga, M., Mazandu, G.K. and Mulder, N.J. (2015) Anc GWAS: a post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. *Bioinformatics*, **32**, 549–556.
92. Wu, J., Vallenius, T., Ovaska, K., Westermarck, J., Mäkelä, T.P. and Hautaniemi, S. (2009) Integrated network analysis platform for protein-protein interactions. *Nat. Methods*, **6**, 75–77.
93. R Core Team. (2016). R: a language and environment for statistical computing. Available at: <https://www.r-project.org>.
94. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R. and Ma'ayan, A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.*, **14**, 128.
95. 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. et al. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
96. Mi, H., Muruganujan, A. and 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.
97. Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H. and Karlak, B. (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.*, **13**, 2129–2141.
98. Csárdi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *Inter. J. Complex. Syst.*, **1695**, 1–9.
99. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinform.*, **9**, 1–8.
100. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D. et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.
101. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E. (2010) Improved side-chain torsion potentials for the amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.
102. Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101–014107.
103. Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F.V., Dinola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684.
104. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092. doi: [10.1063/1.464397](https://doi.org/10.1063/1.464397).
105. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T. and Lee, H. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
106. Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
107. D'Adamo, P., Pinna, M., Capobianco, S., Cesarani, A., D'Eustacchio, A., Fogu, P., Carella, M., Seri, M. and Gasparini, P. (2003) A novel autosomal dominant non-syndromic deafness locus (DFNA48) maps to 12q13-q14 in a large Italian family. *Hum. Genet.*, **112**, 319–320.
108. Donaudy, F., Ferrara, A., Esposito, L., Hertzano, R., Ben-David, O., Bell, R.E., Melchionda, S., Zelante, L., Avraham, K.B. and Gasparini, P. (2003) Multiple mutations of MYO1A, a cochlear-expressed gene, in sensorineural hearing loss. *Am. J. Hum. Genet.*, **72**, 1571–1577.
109. Eisenberger, T., Di Donato, N., Baig, S.M., Neuhaus, C., Beyer, A., Decker, E. et al. (2014) Targeted and Genomewide NGS data disqualify mutations in MYO1A, the 'DFNA48 gene', as a cause of deafness. *Hum. Mutat.*, **35**, 565–570.
110. 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. et al. (1998) Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa. *Science*, **280**, 1753–1757. Available at: <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0009624053>.
111. Gazzaz, B., Weil, D., Rais, L., Akhyat, O., Azeddoug, H. and 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.
112. Golosova, O., Henderson, R., Vaskin, Y., Gabrielian, A., Grekhov, G., Nagarajan, V., Oler, A.J., Quiñones, M., Hurt, D., Fursov, M. et al. (2014) Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. *Peer J.*, **2**, e644.
113. Grønberg, S., Krätzner, R., Spiegler, J., Ferdinandusse, S., Wanders, R.J.A., Waterham, H.R. and Gärtner, J. (2010) Typical cMRI pattern as diagnostic clue for D-bifunctional protein deficiency without apparent biochemical abnormalities in plasma. *Am. J. Med. Genet. A*, **152**, 2845–2849.
114. Hamelmann, C., Amedofu, G.K., Albrecht, K., Muntau, B., Gelhaus, A., Brobby, G.W. and Horstmann, R.D. (2001) Pattern of connexin 26 (GJB2) mutations causing sensorineural hearing impairment in Ghana. *Hum. Mutat.*, **18**, 84–85.
115. 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. et al. (2013) Perrault syndrome is caused by recessive mutations in CLPP, encoding a mitochondrial ATP-dependent chambered protease. *Am. J. Hum. Genet.*, **92**, 605–613.
116. Kimberling, W.J., Weston, M.D., Möller, C., Davenport, S.L.H., Shugart, Y.Y., Priluck, I.A., Martini, A., Milani, M. and Smith, R.J. (1990) Localization of Usher syndrome type II to chromosome 1q. *Genomics*, **7**, 245–249.
117. National Center for Biotechnology Information, N. L. of M. Database of single nucleotide polymorphisms (db SNP). Available at: <http://www.ncbi.nlm.nih.gov/SNP/>.
118. Okonechnikov, K., Golosova, O., Fursov, M. and UGENE, t. (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, **28**, 1166–1167.
119. Pierce, S.B., Walsh, T., Chisholm, K.M., Lee, M.K., Thornton, A.M., Fiumara, A., Oritz, J.M., Levy-Lahad, E., Klevit, R.E. and

- 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.
120. Suzuki, Y., Shimozawa, N., Yajima, S., Tomatsu, S., Kondo, N., Nakada, Y., Akaboshi, S., Lai, M., Tanabe, Y., Hashimoto, T. et al. (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 Available at: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=8279468>.
121. Thorvaldsdottir, H., Robinson, J. and Mesirov, J.P. (2018) Integrative genomics viewer IGV: high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.
122. Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.
123. 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. et al. (1989) Peroxisomal bifunctional enzyme deficiency. *J. Clin. Invest.*, **83**, 771–777.