

PROGRAMA DE DOCTORADO EN BIOMEDICINA (B11.56.1)

**RARE ALLELIC VARIANTS IN MENIERE'S DISEASE
FROM FAMILIAL TO SPORADIC CASES**

VARIANTES ALÉLICAS RARAS EN LA ENFERMEDAD DE MENIERE
DE LOS CASOS FAMILIARES A LOS CASOS ESPORÁDICOS



**UNIVERSIDAD
DE GRANADA**

**INTERNATIONAL PhD THESIS
TESIS DOCTORAL CON MENCIÓN INTERNACIONAL**

ALVARO GALLEGO MARTINEZ
February 2019

GRANADA

Editor: Universidad de Granada. Tesis Doctorales
Autor: Álvaro Gallego Martínez
ISBN: 978-84-1306-154-2
URI: <http://hdl.handle.net/10481/55469>

Index

Contents

Index.....	1
Grants and funding.....	3
Abstract	5
Resumen.....	6
Abbreviations	7
Introduction.....	9
Anatomy of the inner ear	11
2. Meniere’s Disease	14
3. Epidemiology.....	17
4. Pathophysiology	18
5. Human genomics.....	19
6. Variants.....	21
7. Next-generation sequencing technologies in HL and MD.	27
8. Genetics.....	28
Hypothesis.....	31
Goals.....	35
Methods	39
1. Familial MD analysis.....	41
1. Diagnosis of cases.....	41
2. Familial samples	41
3. Whole-exome sequencing (WES)	42
4. Pipeline testing for candidate SNV prioritization	42
5. Benchmarking procedures	45
6. Statistical analysis.....	46
7. Bioinformatics tools for rare SNV selection	46
8. Control datasets to filter by Spanish population variants	47
9. Minor allelic frequency filtering.....	47
10. Prioritization.....	47
11. Linkage analysis	47
12. Validation by Sanger Sequencing	47
13. RNA extraction from cochlea and semicircular canals	48
14. Expression analysis in tissue.....	49

15.	Protein 3D modelling.....	49
16.	Variant submission	49
2.	Sporadic MD analysis.....	50
1.	Sporadic samples.....	50
2.	DNA extraction	50
3.	Selection of target genes.....	51
4.	Preparation of pools.....	51
5.	Haloplex protocol (capture, enrichment, barcoding).....	51
6.	Data generation pipelines	52
7.	Positive control SNV validation	53
8.	Selection and prioritization of pathogenic SNV.....	53
9.	Validation of candidate pathogenic SNV	54
10.	Population statistics	54
11.	Position of variants in significant enriched genes.....	55
	Results	59
1.	Prioritizing variants in exome datasets	59
2.	Comparison of prioritizing strategies with FMD exome datasets	59
3.	Benchmark in exome datasets containing variants described in AD-SNHL and CNM genes	61
4.	Families study.....	64
5.	Sporadic cases study	71
6.	Rare variants analysis	75
7.	Mitochondrial rare variants.....	78
8.	Gene burden analysis	82
	Discussion.....	92
1.	Candidate variant selection in singletons and small families.....	92
2.	Familial MD.....	94
3.	Sporadic MD.....	98
	Conclusions	106
	Bibliography.....	108
	Supplementary data	

Grants and funding

Alvaro Gallego Martinez work and this thesis was feasible thanks to the following funds:

Grant from ISCII PI13/01242

Grant from Meniere's Society, UK

Grant from Luxembourg National Research Fund INTER/Mobility/17/11772209 during three months in the Bioinformatics Core of the University of Luxembourg.

Abstract

Meniere's disease [MD; MIM 156000] is a chronic disorder characterized by attacks of vertigo associated with sensorineural hearing loss (SNHL) involving low to medium frequencies. Although its etiology remains unknown, its prevalence is about 0.5 to 1 / 1000 individuals, affecting more to familial cases than sporadic cases. MD shows a high clinical heterogeneity and incomplete phenotypic forms that complicate its diagnosis.

The goal of this thesis is to obtain a better and comprehensive image of the genetics surrounding MD disease and to characterise its sporadic and familial forms. Therefore, the first goal is to increase what we know about familial MD in two novel families with MD phenotype and variable expressivity. Second goal is to understand genetics in many diagnosed MD cases around Spain and to find genetic markers associated to sporadic MD. As a previous step, we conducted a study to improve the bioinformatic protocol for candidate variants discovery by using multiple tools.

The findings in this Thesis supports that: 1) Integration of multiple prediction and prioritization tools can be used as a protocol for discovery of candidate variants in exome datasets for families where few individuals are sequenced. 2) Validation of two novel candidate variants in SEMA3D and DPT in two families confirms genetic heterogeneity in familial MD with incomplete penetrance and variable expressivity. 3) The study of sporadic cases determined how some SNHL-related genes have an enrichment of missense variants in Spanish population. The candidate list includes well-known genes as GJB2, ESRRB, CLDN14, USH1G and SLC26A4.

Resumen

La enfermedad de Meniere [EM;MIM 156000] es un trastorno crónico caracterizado por ataques de vértigo asociados a pérdida de audición neurosensorial en bajas a medias frecuencias. Aunque su etiología permanece desconocida, su prevalencia es de alrededor de 0.5 a 1 por 1000 individuos, afectando en mayor medida a casos familiares que a esporádicos. La enfermedad de Meniere muestra una alta heterogeneidad clínica y formas con fenotipos incompletos que complican su diagnóstico.

El objetivo de esta tesis es obtener una exhaustiva imagen de la genética de la enfermedad de Meniere, y caracterizar sus formas esporádicas y familiares. Así pues, el primer objetivo es incrementar la información que tenemos sobre la enfermedad de Meniere familiar mediante el estudio de dos familias con casos de Meniere y expresividad heterogénea. El segundo objetivo consiste en entender la genética de un gran número de pacientes de Meniere españoles esporádicos y encontrar marcadores genéticos asociados. Como paso previo, este estudio nos conduce irremediablemente a una puesta a punto de los protocolos para el descubrimiento de variantes candidatas.

Nuestros descubrimientos apoyan: 1) La integración de múltiples herramientas de predicción y priorización puede ser usada como protocolo para el descubrimiento de variantes candidatas en muestras de exoma de familias con pocos individuos secuenciados. 2) La validación de dos variantes candidatas en los genes SEMA3D y DPT en dos familias confirman la heterogeneidad genética en casos familiares con penetrancia incompleta y expresividad variable. 3) El estudio de casos esporádicos ha determinado que ciertos genes tienen un enriquecimiento en variantes exónicas sin-sentido en la población española. Esta lista incluye genes como GJB2, CLDN14, USH1G, ESRRB y SLC26A4.

Abbreviations

AAO-HNS	American Academy of Otolaryngology-Head and Neck Surgery
AD	Autoimmune Disease
AIED	Autoimmune Inner Ear Disease
BMD	Bilateral Meniere Disease
CNM	CentroNuclear Myopathy
CSVS	Collaborative Spanish Variant Server
DH	Delayed Hydrops / Delayed MD
EAONO	European Academy of Otology and Neurotology
EH	Endolymphatic Hydrops
ES	Endolymphatic Sac
EVS	Exome Variant Server
ExAC	Exome Aggregation Consortium
FMD	Familial Meniere Disease
HC	Hair Cells
HGMD	Human Gene Mutation Database
HL	Hearing Loss
HPO	Human Phenotype Ontology
IHC	Inner Hair Cells
INDEL	Insertion/Deletion
MAF	Minor Allele Frequency
MD	Meniere Disease
NGS	Next Generation Sequencing
OHC	Outer Hair Cells
PCR	Polymerase Chain Reaction
RL	Reticular Lamina
SC	Supporting Cells
SD	Standard Deviation
SGN	Spiral Ganglion Neurons
SL	Spiral Ligament
SMD	Sporadic Meniere Disease
SNHL	Sensorineural Hearing Loss (AD; Autosomal Dominant – AR; Autosomal Recessive)
SNV	Single Nucleotid Variant
SV	Structure Variant
UMD	Unilateral Meniere Disease
VCF	Variant Call Format
VM	Vestibular Migraine
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Introduction

Anatomy of the inner ear

Hearing and balance sense is particularly addressed by the ear, a sensory organ located in the temporal bone in both sides of the head in humans. The ear is composed by three large pieces or parts (Figure 1):

- External ear, which includes the pinna and the external auditory canal and it is separated from the middle ear by the tympanic membrane.
- Middle ear, containing the ossicular chain with 3 bones: malleus, incus and stapes
- Inner ear or labyrinth, which is divided in anterior labyrinth or cochlea that contains the hearing organ (organ of Corti), and the posterior labyrinth, which includes the semicircular canals and the vestibular end organs.

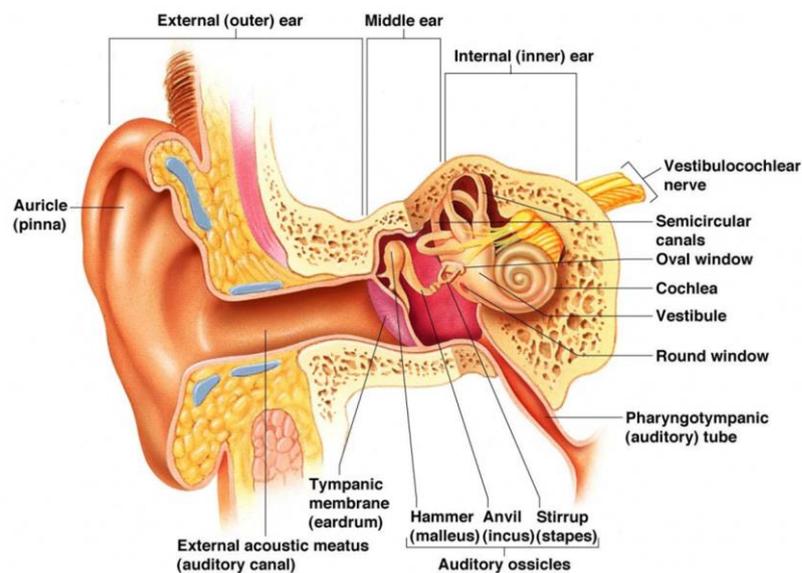


Figure 1. Anatomy of the ear. (Extracted from *Human Anatomy and Physiology*, Marieb et al. 2009¹)

a. Anatomy of the cochlea

The human cochlea is a membrane conduct of 3-3.5 cm long that rolls itself around two and a half turns following a central axis called modiolus, where the cochlear nerve is located. The cochlea is enveloped by a bone layer called the otic capsule within the temporal bone. The interior of the cochlea is filled by two fluids, the perilymph and the endolymph which remain separated in different compartments. So, the cochlea is divided in three parts or compartments:

- The vestibular duct (scala vestibuli), which lies superior to the cochlear duct and the oval window and contains perilymph.
- The tympanic duct (scala tympani), which lies inferior to the cochlear duct and also contains perilymph.

- The cochlear duct (scala media), which lies between both previous ducts and hosts the organ of Corti and contains endolymph.

The vestibular and tympanic ducts join on the apex of the cochlea, which is named helicotrema. Both ducts communicate with the middle ear through two discontinuities on the bone layer: the oval window, connecting the middle ear and the vestibular duct and where the stapes is located; and the round window, which communicates with the middle ear through the tympanic duct. The organ of Corti is limited by the Reissner membrane and the basilar membrane.

So, the vestibular duct and the tympanic duct content is an ionic fluid, the perilymph, similar to the cerebrospinal liquid. However, the cochlear duct contains the endolymph, with a major concentration of K^+ and Na^+ ions. The differences in ions concentration generates an electrochemical gradient between both fluids called endocochlear potential (around +80Mv), which is essential for the functioning of sensory cells in the organ of Corti².

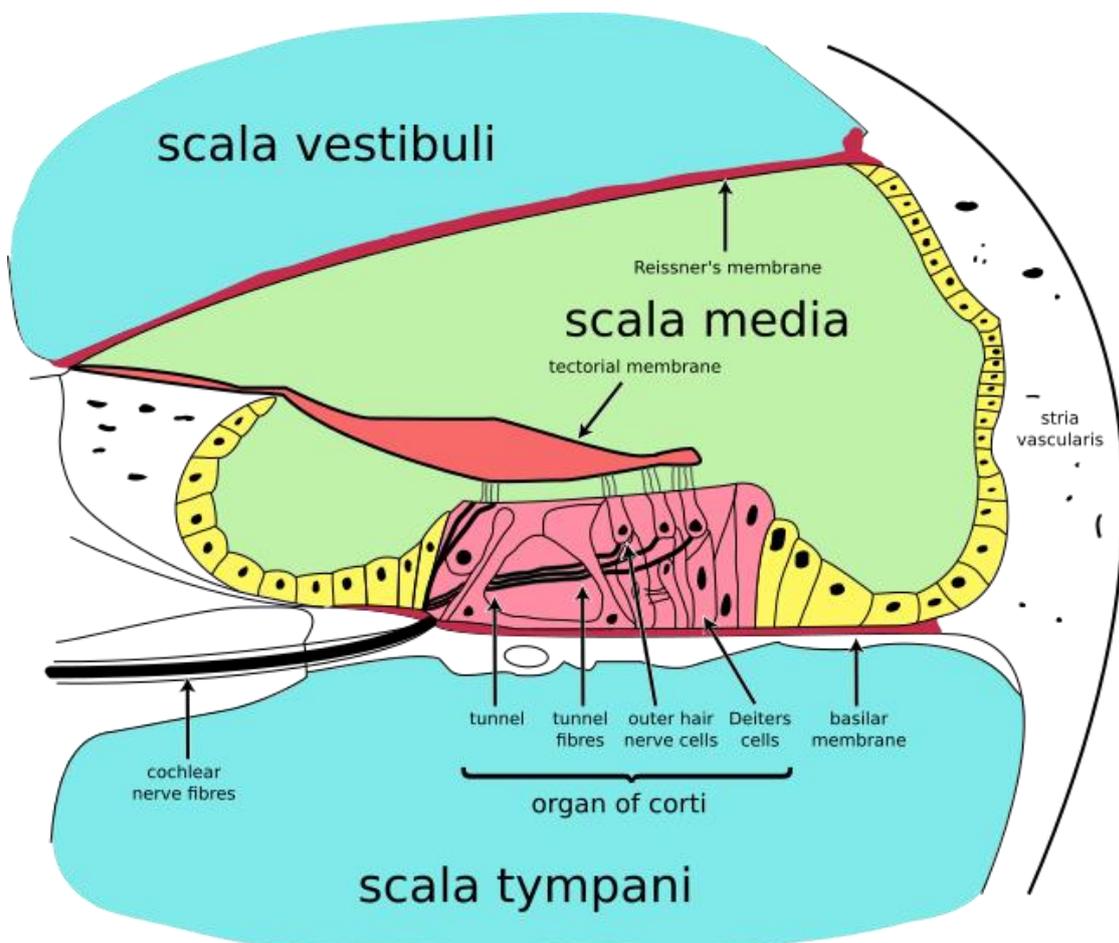


Figure 2. Organ of Corti cross-section. (Extracted from Wikipedia Commons under free license).

The organ of Corti is a spiral structure located in the cochlear duct. This organ is formed by dozens of highly specialized cell types (Figure 2). According to their function, these cells are grouped in two major cell types: hair cells and supporting cells^{3,4}. These hair cells are separated in two types: inner hair cells and outer hair cells. These cells are sensorineural and are innervated to receive and send neurotransmitters to code the auditory information⁵. Deflection of the hair cell stereocilia activates ionic channels positively which depolarizes the cell, resulting in a receptor potential. Calcium ions trigger the release of neurotransmitters along the space between the hair cell and the nerve terminal, converting the depolarization in an electrical nerve signal⁶.

Supporting cells are non-sensory cells anchoring between hair cells⁴. Supporting cells cover the entire epithelium, from basal lamina to lumen. At the surface, they join the hair cells through tight and adherens junctions to the reticular lamina⁷. Communication between cells is made through gap junctions. Although they are not sensory, their role as structural integrators of the entire matrix is suspected to be more necessary than previously thought⁸.

b. Anatomy of the vestibular labyrinth

Balance is the results of the integration of the sensory information that is provided by the vestibular, visual and proprioceptive systems in the brainstem and the cerebellum. The vestibular system is the posterior part of the labyrinth inside the otic capsule in the temporal bone. The vestibular system is composed by five different organs sensitive to angular acceleration (semicircular canals) and linear acceleration (otolith organs). The perilymph fluid appears through those canals, connecting them to the endolymphatic sac. The membranous labyrinth is situated inside the bony labyrinth comprising the otolithic organs.

The vestibule can be detailed in three different parts: the saccule, the utricle-macula and the semicircular canals.

The saccule is related to balance and gravity perception. It is covered by sensory cells and on the top, they present an extracellular structure, the otolithic membrane which contains hundred of otoconia, whose capability to detect linear accelerations and head tilts allows to vertical comprehension. When the head moves in the vertical plane, the deflection of the stereocilia in the hair cells opens the mechanotransduction channels with a fast increase of K^+ and induces the depolarization of the cells, which transmits signals to the vestibular nuclei in the brainstem.

The utricle macula is horizontally situated. Hair cells of the utricle distinguish different degrees of tilting of the head. Stereocilia and kinocilium are the organs in charge of sending

signals about the tilting in the horizontal plane. Movements of those organs send depolarizing or hyperpolarizing signals according to the direction of the movement. The afferent nerve fibers send this signal to the vestibular nuclei which integrate it with the visual information and generate a vestibule-ocular reflex to stabilize the eyes in the visual field.

Semicircular canals differ from the utricle and contain an ampulla with sensory cells. This part of the ampulla is called crista ampullaris and collects information of angular acceleration and deceleration of the head.

2. Meniere's Disease

Meniere's disease (MD) is an inner ear syndrome characterized by episodes of spontaneous vertigo (that can last from some minutes to hours) and is often associated with low to medium frequency sensorineural hearing loss (SNHL), aural fullness and tinnitus⁹. Vertigo usually disappears at the latest stages of the disease. Hearing loss develops after the first vertigo attacks, developing into total deafness at the final phase of the disease.

Episodic vertigo is one of the most common features of the syndrome. This symptom is usually experienced as a sensation of rotation of the head in which the patient perceives his environment or himself turning around. After repetitive episodes, patient can develop chronic imbalance, driving them to difficulties into staying on foot and an increased risk of falls. The episodes can last from minutes to hours, but their duration and severity are usually unpredictable and have a significant impact in the health-related quality of life between patients.

The etiology of vertigo is unknown yet. Most hypothesis suggest changes in endolymph pressure, the rupture of the membranous labyrinth or sudden changes of the ionic composition of the endolymph, driving to a dysfunction of the mechanotransduction channels located in the tip-links of the stereocilia in the hair cells epithelia of the saccule or canals, and generating an abnormal nerve discharge^{10,11}.

Sensorineural hearing loss (SNHL) is the most common feature in this syndrome, because it usually worsens over time. SNHL can start developing in one ear (unilateral SNHL)¹² or both simultaneously (bilateral SNHL)¹³. The progression of SNHL can be fast (around weeks or months) or very slow (years). The most common progression observed in patients is slow, but extreme MD phenotypes can develop a moderate to severe SNHL at early stages of life¹⁴. Hearing loss can progress to profound deafness at the latest stages of age.

SNHL is diagnosed by pure-tone audiometry. Audiograms are required to differentiate MD from other inner ear diseases and to observe disease progression¹⁵.

Tinnitus is another symptom in MD patients, but this presence could be vastly delayed along the age of the patient. Tinnitus is the perception of sounds in the ear when no actual sound is present¹⁶.

In most cases, tinnitus' loudness usually increases during vertigo attacks or after hearing starts to decrease. Tinnitus is often a very disabling symptom due to the difficulty of speech comprehension while hearing an annoying noise. Tinnitus may appear in some patients without relation to the attacks, it may become permanent and it is usually associated with anxiety¹⁷.

Aural fullness is the perception of blocking, fullness and ear pressure. This sensation could be constant during the disease and his intensity can worsen after several vertigo attacks.

Diagnosis of MD

MD diagnosis is often difficult due to its heterogeneous nature. MD symptoms overlap with some other conditions as vestibular migraine, otosclerosis and others¹⁸⁻²¹.

MD previous definition followed the guidelines established by American Academy of Otolaryngology-Head and Neck surgery (AAO-HNS) in 1995²², but since 2015 a new diagnostic criteria has been established by the Classification Committee of the Bárány Society, the Japan Society for Equilibrium Research, the European Academy of Otology and Neurology (EAONO), the Equilibrium Committee of the American Academy of Otolaryngology-Head and Neck Surgery (AAO-HNS) and the Korean Balance Society⁹ (Table 1).

Symptoms	Definite MD	Probable MD
Hearing loss (HL)	Audiometrically documented low-to-medium frequency SNHL on an affected ear during or after one vertigo episode.	
Spontaneous vertigo	2 or more episodes of vertigo during 20 min to 12 hours	2 or more episodes of vertigo during 20 min to 24 hours
Tinnitus / aural fullness	Fluctuating aural symptoms	Fluctuating aural symptoms
Other	Excluded	Excluded

Table 1. Diagnostic criteria for MD, according to the Classification Committee of the Barany Society (2015).

Pure-tone audiometry is required for all patients. Due to the clinical heterogeneity and the overlapping symptoms with vestibular migraine, some other differential diagnosis is needed along with physical examination and additional tests as speech audiometry, auditory evoked potentials and vestibular and imaging tests among others (Figure 3). The differential diagnosis includes some common and rare diseases such as:

1. Autoimmune inner ear disease (AIED): as an inflammatory disease of the inner ear, this disease incurs in some recurring episodes of sudden to progressive bilateral SNHL²⁰.

2. Vestibular migraine (VM): this disease is the most common cause of vertigo in the world, after benign paroxysmal positional vertigo¹⁸.

3. Delayed Hydrops (DH) or delayed MD: DH is usually found in patients that have suffered longstanding unilateral profound SNHL. There can be several years between the appearance of SNHL and vertigo¹⁴.

4. Monogenic SNHL disorders with vestibular dysfunction such as DFNA9, caused by mutations in COCH gene²³⁻²⁵.

5. Transient ischemic attacks involving the internal auditory artery with a similar phenotype of fluctuating hearing loss, tinnitus and episodic vertigo. This etiology should be suspected in older individuals (> 60 years old) with cardiovascular risk factors^{26,27}.

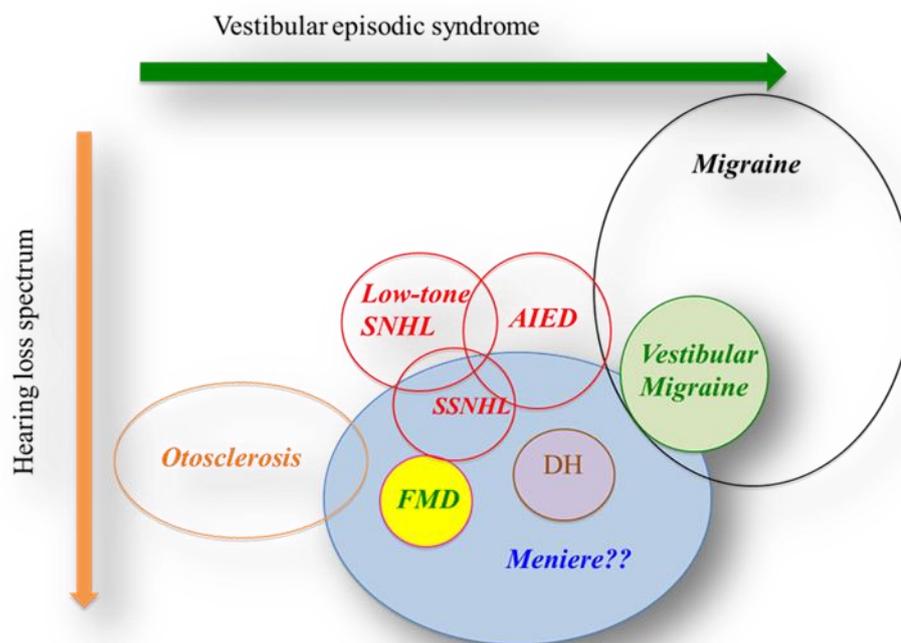


Figure 3. Inner ear disorders with overlapping symptoms with Meniere disease.

3. Epidemiology

The prevalence of MD is highly variable in the world due to the difficulties in diagnosis. However, some population studies carried out points to biased assumptions in prevalence. MD seems to be more prevalent in industrialized European countries rather than Asian and African populations, with ranges around 225 cases per 100000 individuals²⁸. A study conducted in US estimated a MD prevalence of 190 cases per 100000 individuals²⁹, pretty similar to a study done in UK where prevalence of 157 cases per 100000 were detailed³⁰. A larger study considering only Caucasian ascendant people showed a prevalence quite higher than the rest of the studies (1-2/1000)³¹, while Finnish and Japanese studies detail quite lower prevalence (43 and 36 cases per 100000 individuals respectively)^{32,33}. A last study made in Cantabria situated Spanish MD prevalence in 75 cases per 100000 individuals³⁴, half way between Europeans and Asians.

Ethnic differences or geographical issues could result in this variation of prevalence in MD, as it is found in other similar complex diseases. However, a methodological bias in the estimation of prevalence cannot be rule out. Before 1995, the lack of standardization on the diagnostic criteria could result in this biased diagnosis among the different regions. Also, most

of MD patients do not develop all the symptoms at early stages of the disease, so this could mask and make more difficult diagnosis.

4. Pathophysiology

MD etiology remains vastly unknown, but major hypothesis considers MD as a multifactorial complex disease. Genetic, autoimmune, autoinflammation and allergy remain as major hypothesis nowadays. All these mechanisms would lead to the accumulation of endolymph in the cochlear duct and explain the finding of dilatation and rupture of cochlear duct in temporal bone necropsies of patients with MD.

4.1. *Hydrops*

Some studies point endolymphatic hydrops (EH) as a very common finding in histopathological studies of the temporal bone of MD patients³⁵. EH is a dilatation of the membranous labyrinth by an increase in volume of endolymph^{36,37}. Cellular stress is supposed to produce an ionic imbalance that can cause a cochlear-vestibular dysfunction.

Spiral ligament of the cochlea could be the trigger of MD. Then, endolymph produced in the stria vascularis in the cochlear duct should be reabsorbed in the endolymphatic sac. Any kind of interruption in on the flow could lead to an accumulation of endolymph, developing EH. A decrease in the absorption of endolymph in the ES could be caused by ionic imbalance, genetic mutations, viral infections, dietary factors, autoimmune diseases, allergic responses or vascular irregularities. The endolymph in the area pressures the cochlear duct and produces the dilation of the Reissner's membrane³⁸.

4.2. *Cellular and molecular alterations caused by EH*

Mastoid of an affected ear is usually smaller in size and the vestibular aqueduct is shorter, with a narrow periaqueductal and external opening. Moreover, the oval window membranes are found to be thicker in MD patients than controls³⁹. At advanced ages, patients also lose hair cells and suffer atrophy of supporting and epithelial cells on the organ of Corti and the tectorial membrane.

A decrease in spiral ganglion neurons (SGN) has been found in temporal bone in MD patients, before any damage in cochlear hair cells was found. Besides that, a significant decrease of type I afferent nerve endings number and synapses at the base of the inner and outer hair cells are found in some patients⁴⁰. So, SGN loss could occur in the initial steps of the degenerative cascade of the cochlea, before the loss of hair cells.

These findings do not correlate too well with the presence of EH. So, the lack of hair cells loss compared to the damaged nerve endings raise concern about the role of EH in the molecular origin of MD. Therefore, current data support the hypothesis that EH is a epiphenomenon associated with hearing loss in a variety of inner ear disorders, rather than MD itself⁴¹.

5. Human genomics

Elemental and basic genetic information is saved in living species as desoxyribonucleic acid (known as DNA). This molecule was discovered back in 1869 by Friedrich Miescher⁴². This acid is compound by a combination of four nucleotides, an organic molecule form by the covalent union of a pentose (five-carbon sugar), a nitrogenous base and a phosphate group. The four nucleotides can be divided according to contain either a purine base (guanine or G and adenine or A) or pyrimidine base (cytosine or C and thymine or T). The DNA molecule shows a singular structure discovered by Watson and Crick⁴³ call double helix. The basal structure appears after the bonding between two pairs of nucleotides (A – T and C – G).

Human DNA is a complex repetitive molecule form with near 3 billion of nucleotides in different combinations. This molecule represents the entire genome of a human. Each cell in the human body present the same DNA generally (avoiding mosaicism), and it appears in an extreme coiled form known as chromosome. Human cells have 23 pairs of chromosomes: 22 of them are called autosomes while the last 23rd pair is called sexual chromosome, as it refers to the determination of the sex in an individual. This last pair is different in males and females. In males, both chromosomes in the pair are different, only collocating in similarity in a region known as pseudoautosomal region. Autosomes are exactly the same between both sexes. Changes in chromosome number usually result in known chromosomal aberrations as Down syndrome and Patau syndrome.

Inside chromosomes, information is vastly inaccessible until it is decoiled by specialized proteins. However, most of the DNA doesn't code biological information and it is called non-coding regions. Exons of genes, that are the regions coding for proteins, are called coding regions. Most of the region between genes doesn't code for any protein. Intergenic regions, however, have been described as potentially regulating zones^{44,45}.

As a project to study the vast human genome, in 2001 appeared The Human Genome Project (HGP)⁴⁶ as a collaborative work between a lot of genomic groups interested in a standard knowledge database of the human genome. This standard genomic database allowed groups to compare genomes of individuals with different diseases or features and look for

similarities and differences between them. The HGP was used as a key tool to understand conserved regions along the genome, regions with special interest in biology. This project took nearly 13 years and its cost was astronomic (around \$2.7 billions). Since then, the interest in genomic data started to grow and the price for sequencing human genomes started to diminish while technology was experiencing a high improvement (Figure 4, Figure 5).

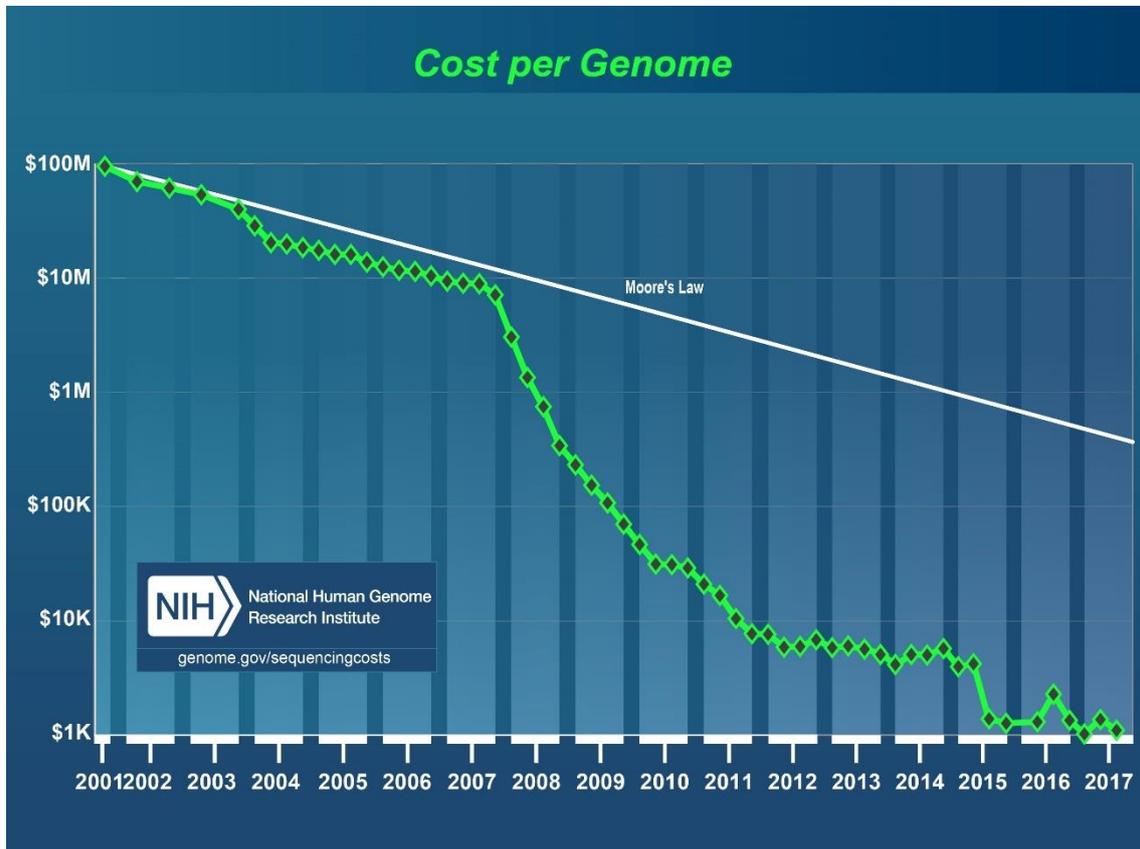


Figure 4. Cost per genome registered since first genome sequenced by NIH research institute (extracted from NIH Research Institute genome.gov/sequencingcosts).

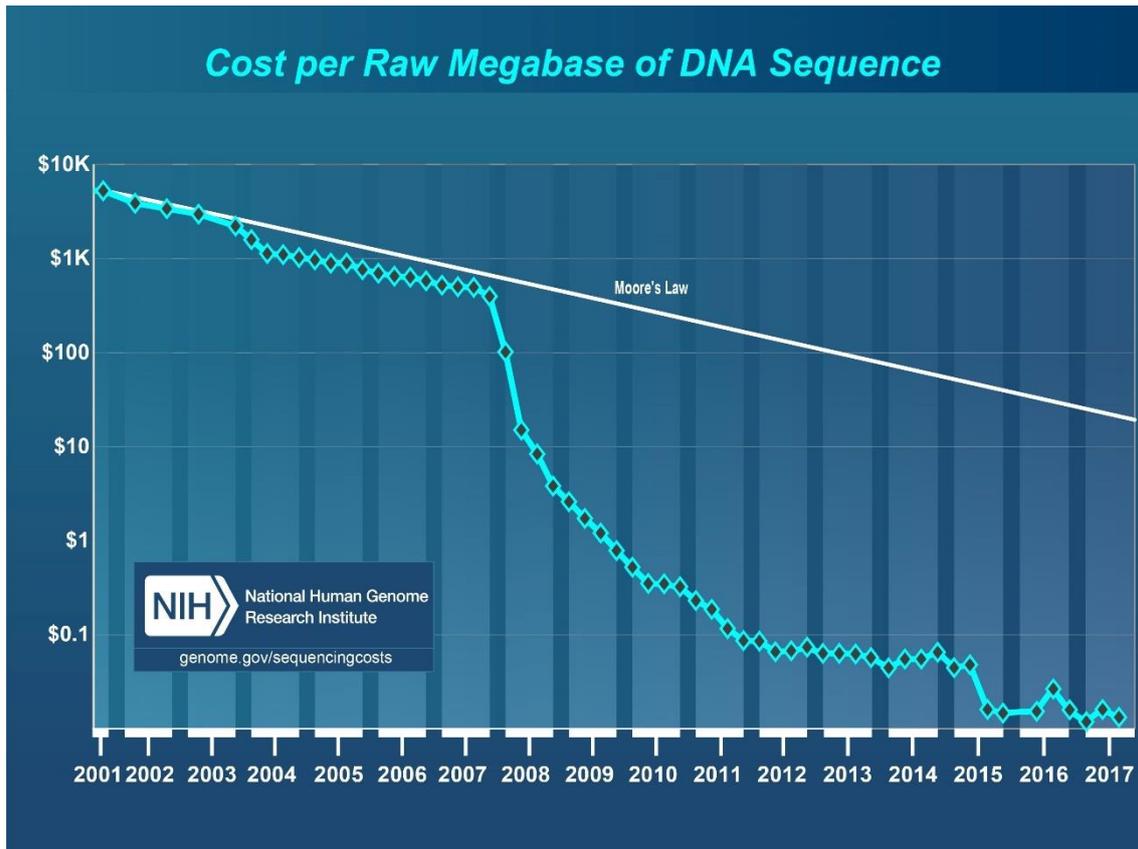


Figure 5. Cost per raw megabase of DNA sequence since the first sequence genome registered by the NIH research institute (extracted from NIH Research Institute genome.gov/sequencingcosts).

6. Variants

Information in the DNA strands is codified by a set of rules called genetic code. This genetic code is common to most species but differs in some bacterial and fungus species. The information to obtain a protein from the DNA has two known steps where different proteins and regulators act: transcription step, where DNA is transcribed to a mRNA (messenger RNA), and translation step, where this mRNA is read and translated to protein. Both steps have their regulators, that allows a certain grade of control proofing to avoid mistakes⁴⁷.

The mRNA can be read in 6 different ways, according to the sense and the reading frame. Reading frame represent how the translation is going to divide the information of the DNA according to set of three consecutive and non-overlapping nucleotides called codons. Codons are permutations of three nucleotides to a total of 64 possible codons translating to the different aminoacid that forms a protein. Besides of the large pool of possibilities, there is redundancy in this code, meaning some codons translate the same aminoacid. However, this is an important principle of the genetic code, as it is degenerated⁴⁸, and there is not ambiguity as they always code for the same aminoacid of the 20 possible ones.

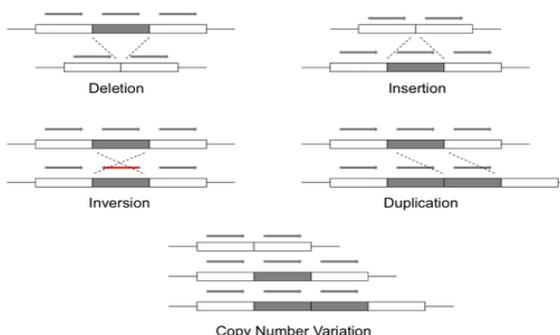
The most important codons include the unique AUG codon, or start codon, and the three STOP codons. These codons mark the beginning and the end of the translation. While the start codon codes for a methionine, the three stop codons don't code for any amino acid. Instead, they are the signal for the translation machinery to disengage.

6.1. Genetic variation

DNA replication is not perfect. Genetic variation may occur as part of the possible errors during DNA replication phase. However, this is an important feature in evolution, as genetic variation is inherited from one generation to the other, it ports new possibilities to the descendants. Mutations are irreversible and are one of the major sources of genetic variation. However, natural selection tries to maintain neutral effect mutation instead of deleterious mutation, aspect that can be measured through minor allele frequencies in population genetics: rare variants used to be related to deleterious traits and its conservation is more difficult in a healthy population⁴⁹. Although this assumption is potentially true, not all the rare mutations are directly related to diseases. Therefore, the term "variant" is used for this type of mutation instead of the former.

There are various types of genetic variants, which are all summarized in Table 2.

Table 2. Type of variants and descriptions.

<p>Reference ACTGACGCATGCATCATGCATGC</p> <p>SNP ACTGACGCATGCATCATTCATGC</p>	Nonsynonymous (no conservative)	TTC > TGC == LYS > THR
	Nonsynonymous (conservative)	TTC > TCC == LYS > ARG
	Synonymous or silent	TTC > TTT == LYS > LYS
	Nonsense	TTC > ATC == LYS > STOP
<p>Reference ACTGACGCATGCATCATGCATGC</p> <p>Insertion ACTGACGCATGGTACATCATGCATGC</p> <p>Deletion ACTGACG--TGCATCATGCATGC</p> <p style="text-align: right;">} Indel</p>	Non Frameshift	Indel length divisible by three. Do not change reading frame.
	Frameshift	Indel length not divisible by three. Change reading frame.
<p>Structural Variation</p>  <p>The diagrams show: 1. Deletion: a segment of a chromosome is missing. 2. Insertion: a segment of a chromosome is added. 3. Inversion: a segment of a chromosome is flipped 180 degrees. 4. Duplication: a segment of a chromosome is repeated. 5. Copy Number Variation: the number of copies of a chromosome or segment is altered.</p>	<ul style="list-style-type: none"> • Large deletion • Large insertion • Inversion • Duplication • Copy Number Variation 	

6.2. Single nucleotide variants (SNVs)

SNVs are single nucleotide changes in the DNA when compared with the reference genome. They are considered common if they surpass 1% of the population, in which case they are called single nucleotide polymorphism (SNPs). Attending to their functional effect, they can be divided in different categories:

Coding variants. These variants affect protein-coding regions of the genome. As a single nucleotide change, they can result in different changes in the aminoacid when translated.

Synonymous variant. The change in the codon doesn't change the translated aminoacid. The genetic code can translate for the same aminoacid by multiple codons. Most of the codons for the same aminoacid changes only on the third nucleotide. If one variant affects the third nucleotide, it is very probable that the final output could be the same. These variants are considered as functionally neutral, but certain synonymous variants can also be disease causing based on their functions, as some synonymous variants are involved in splicing⁵⁰.

Nonsynonymous variants. If the resulting aminoacid from the affected codon derives in a different aminoacid, the variant is termed as nonsynonymous. If the change derives in a stop codon or in the loss of a stop codon, the variant is called as missense.

Non-coding variants. The rest of SNVs appearing at the non-coding regions of the genome are non-coding variants. Their importance is unknown but they are a focus of study nowadays in genomic research.

Insertions and deletions (INDELs). Insertions and deletions (INDELs) are small regions of the human genome that are missing or appear somewhere other than the normal region in the reference genome. INDELs range around 1 and 10.000 bp. Based in their effect on the reading frame, they are known as frameshift indels (the resulting change affects the reading frame during translation) and non-frameshift indels (the resulting change doesn't affect the reading frame). Both types of indels can suppose different problems to the translation of the entire protein, altering his integrity and structure^{51,52}.

Structural variants (SVs). Structural variants are usually genomic alterations larger than 1000pb. They are poorly studied; however they are considered as important as indels to the developing of a disease.

Copy Number Variants (CNVs). CNVs are the most common SVs studied so far. They are large insertions, deletions or duplication occurring in the genome. They are supposed to represent a large proportion of the phenotypic variation between individuals⁵³.

Inversions and other SVs. Inversions are regions of the DNA that appear reversed with respect to the rest of the genome. Inversions usually appear as sub products of recombination. Some diseases are caused by inversions, as Angelman syndrome, Hunter syndrome... Other SVs includes translocations and segmental uniparental disomy. Although they are rare, most of them are suspected to derive from the same steps in recombination⁵⁴.

6.3. Variant pathogenicity

Most of the variation in human genome has no functional effect and not all the detected variants can be categorized as pathogenic. However, standardization studies have agreed about which criteria is needed to determine likely-pathogenic status on a selected variant. Multi-criteria framework has been used for the last years to obtain enough evidence about the pathogenic status of known variants. One of the most relevant in the last years was published by Richards et al, 2015⁵⁵ where they determine the relevance of a multi-evidence framework to score a variant inside a benign-or-pathogenic profile. Most relevant evidence criteria are detailed in the next Table 3 (extracted from Richards et al, 2015).

Table 3. Evidence-based criteria for pathogenicity description of a variant (extracted and adapted from Richards et al, 2015⁵⁵).

<i>Evidence</i>	← BENIGN →		← PATHOGENIC →			
	<i>Strong</i>	<i>Supporting</i>	<i>Supporting</i>	<i>Moderate</i>	<i>Strong</i>	<i>Very Strong</i>
<i>Population data</i>	MAF too high for a disorder OR observation in controls inconsistent with disease penetrance			Absent in population databases	Prevalence in affected cases statistically increased over controls	
<i>Computational and predictive data</i>		Multiple lines of computational evidence suggest no impact on gene or product OR missense in gene where only truncation cause disease OR silent variant with non-predicted splice impact	Multiple lines of computational evidence support a deleterious effect on the gene or product	Novel missense change at a aminoacid where a different pathogenic missense change has been seen before OR protein length changing variant	Same aminoacid change as an established pathogenic variant	Predicted null variant in a gene where LOF is a known mechanism of disease
<i>Functional data</i>	Well-established functional studies show no deleterious effect		Missense in gene with low rate of benign missense variants and common pathogenic missense variants	Mutational hot spot or well-established functional domain without benign variation	Well-established functional studies show a deleterious effect	
<i>Segregation data</i>	Non-segregation with disease		Co-segregation with disease in multiple affected family members	Co-segregation with disease in multiple affected family members (data increase)	Co-segregation with disease in multiple affected family members (data increase)	
<i>De novo data</i>				De novo without paternity and maternity confirmed	De novo with paternity and maternity confirmed	
<i>Allelic data</i>		Observed in trans with a dominant variant OR observed in cis with a pathogenic variant		Detected in trans with a pathogenic variant in recessive disorder		
<i>Other databases</i>		Reputable source w/out shared data = benign	Reputable source where variant is pathogenic			
<i>Other data</i>		Found in a case with an alternate cause of disease	Patient's phenotype highly specific for gene			

7. Next-generation sequencing technologies in HL and MD.

Detection of variants in the DNA sample is a requirement for his validation. Since 1977 with the first sequencing method developed by Frederick Sanger⁵⁶, sequencing technology has evolved very quickly. The major goals of next generation sequencing technologies is the improvement in quality and price of sequencing. However, Sanger sequencing is still considered the gold standard for validation, although its price per sequenced base is very high. For extensive sequencing of entire exomes and genomes, Sanger sequencing has been replaced in all its forms, giving its place to different technologies specialized in different kind of studies.

7.1. Next generation sequencing

Next generation sequencing represents the next step in biomedical science. NGS technologies allow a high throughput DNA sequencing with more efficiency than older techniques as Sanger. NGS generates millions of sequences per run, optimizing in time the entire process, and allowing scientist to resequence faster than before. Today, there are a lot of platforms for NGS, making it more affordable to perform. Most known NGS platforms are Illumina, 454, Qiagen, Ion Torrect and Nanopore. All of them offers different approach to sequencing, according to the main focus of the different genomic studies.

7.2. WGS

Whole-Genome Sequencing (or WGS) is the determination of the entire DNA sequence of an individual. This technique used to be more expensive than any other in the sequencing catalog. However, with the new machines that are appearing in market nowadays, an entire genome has become more affordable.

Sequencing a genome brings a lot of advantages in the study of diseases and traits. Although the raw output of a WGS experiment is quite large, it is the entire information of the individual. Researchers uses this approach when they want to focus in non-coding regions, as it is the only method that range this large part of the genome⁵⁷. Another big advantage is the uniformity of the output which can be very useful in the discovery of CNVs and exonic variants better than in WES analysis.

7.3. WES

Whole Exome Sequencing is a method that takes only coding regions of a genome as a target for sequencing. Due to the shorter representation of coding regions against the whole genome, it is a cheaper alternative to WGS and is widely accepted for variant discovery. As it only covers coding regions, this technology lacks into other relevant information along the genome⁵⁸.

7.4. Targeted gene sequencing

Targeted gene sequencing covers a panel of chosen regions or genes of the genome. These panels can be custom or premade. They are usually made for diagnostics and variant discovery in specific known targets. Methods for targeted sequencing could be very different, from PCR enrichment to solution hybridization along interest probes. This method allows multiplexing in some platforms, making it cheaper than WES.

8. Genetics

While most of the known patients of MD are considered sporadic, some patients report relatives with SNHL or vertigo during diagnosis. Studies have determined that between 3-14% of MD patients have genetic background, and that one or most member of their families has similar symptomatology⁵⁹⁻⁶¹.

MD inheritance has been proposed differently, attending to the different families studied. Some studies show autosomal dominant patterns of inheritance⁶²⁻⁶⁴ while others shows recessive inheritance^{59,65,66}. Incomplete penetrance and variable expressivity marks differences between familial cases.

Most recent studies suggest MD could appear in two forms: familial MD or sporadic MD⁶⁰. The frequency for the former is lower (5-15%) in European population⁶¹.

To investigate the genetic contribution to MD, several methodologies to find candidate genes associated with MD have been addressed. From the pregenomic era, linkage analysis studies in few large autosomal FMD have found candidate loci at 5q14-15 in German families⁶⁵ and 12p12.3 in Swedish families⁶⁶. Besides that, more large-scale case-control studies are needed in order to establish the relationship between the discovered variants and phenotype of the disease. Some of the latest studies about MD genetics are described in Table 4.

Author	Study	Candidate Regions/Genes
Koyama et al, 1993 ⁶⁷	Case-control	HLA-Cw04DRB1*1602
Melchiorri et al, 2002 ⁶⁸	Case-control	HLA-Cw07
Lopez-Escamez et al, 2002 ⁶⁹	Case-control	HLA
Mhatre et al, 2002 ⁷⁰	Case-control	AQP2
Lynch et al, 2002 ⁷¹	Case-control	ATQ1
Doi et al, 2005 ⁷²	Case-control	KCNE1-3
Klar et, al 2006 ⁶⁶	Familial	12p12PIK3C2G
Lopez-Escamez et al, 2007 ⁷³	Case-control	HLA-DRB1*1011
Teggi et al, 2008 ⁷⁴	Case-control	ADD1
Kawaguchi et al, 2008 ⁷⁵	Case-control	HSPA1A
Vrabec et al, 2008 ⁷⁶	Case-control	HCFC1
Lopez-Escamez et al, 2009 ⁷⁷	Case-control	PARP1
Candreia et al, 2010 ⁷⁸	Case-control	AQP3
Maekawa et al, 2010 ⁷⁹	Case-control	AQP2
Campbell et al, 2010 ⁸⁰	Case-control	KCNE1-3
Lopez-Escamez et al, 2010 ⁸¹	Case-control	PTPN22
Khorsandi et al, 2011 ⁸²	Case-control	HLA-Cw04
Hietikko et al, 2011 ⁸³	Familial	12p12.3
Furuta et al, 2011 ⁸⁴	Case-control	IL1A
Lopez-Escamez et al, 2011 ⁸⁵	Case-control	CD16A/CD32
Arweiler-Harbeck et al, 2011 ⁶⁵	Familial	Chromosome 5
Gazquez et al, 2011 ⁸⁶	Case-control	NOS1-NOS2A
Gazquez et al, 2012 ⁸⁷	Case-control	MICA-STRA.4
Hietikko et al, 2012 ⁸⁸	Case-control	KCNE1
Yazdani et al, 2013 ⁸⁹	Case-control	MIF-173
Gazquez et al, 2013 ⁹⁰	Case-control	MIF,INFG,TFNA
Requena et al, 2013 ⁹¹	Case-control	TLR10
Teranishi et al, 2013 ⁹²	Case-control	Cav1
Cabrera et al, 2014 ⁹³	Case-control	NFKB1
Requena et al, 2015 ⁶²	Familial	DTNA,FAM136A
Yazdani et al, 2015 ⁹⁴	Case-control	RANTES
Martin-Sierra et al, 2016 ⁶³	Familial	PRKCB
Martin-Sierra et al, 2017 ⁶⁴	Familial	DPT,SEMA3D

Table 4. List of candidate genes/regions related to different MD cases studied in the last years.

Hypothesis

Hypothesis

Meniere's disease (MD) has a hereditary component affecting to the onset of the disease in different individuals and their families. The compared study between familial cases and sporadic cases using known genes related to MD symptomatology could allow us to discern which variants are likely pathogenic in these individuals. This study can complement the genetic background previously known about MD, so different pathways and protein interactions involved in the develop of this pathology. Discoveries of new variants can be also used as target for diagnostic panels and prognosis of the disease.

Goals

Goals

- The identification of rare variants segregating MD phenotype in affected familial cases through WES technology and validation of candidate variants through Sanger sequencing.
- To determine candidate variants in sensorineural hearing loss-related genes and vestibular phenotype-related genes by target-enrichment sequencing in sporadic MD cases retrieved from several Spanish populations. This will improve our understanding about genetics of MD in Spanish population through comparison with allele frequencies in different populations.
- To understand how selected variants affect the protein interactions of candidate genes and pathways involved in sporadic and familial cases.

Methods

1. Familial MD analysis

1. Diagnosis of cases

Two multicase families with autosomal dominant familial MD were analyzed. Both families were originally from the southeast of Spain. Pedigrees are added in Results. Institutional Review Board for Clinical Research approved this study, and an informed consent for donor of biological samples was obtained from all the subjects in the family study.

Diagnosis was established according to the diagnostic criteria defined by the Bárány Society (Table 1). Neurotology assessment was carried out in all cases, and a brain magnetic resonance imaging was performed to exclude other symptoms related to neurological issues. Patients were followed with serial pure tone audiograms at each visit, monitoring hearing loss from the initial diagnosis to the last recorded.

Pedigrees for both families are included in Results (Figure 11). The first family (F1) consisted of five generations where the two first generations are mostly deceased. Cases with MD included III-3, III-4, III-5. All of them had a complete MD phenotype in the same generation. The second family (F2) had three women affected with MD in the same generation with autoimmune background.

2. Familial samples

DNA samples were extracted from peripheral blood of patients using the manufacturer's protocol for DNA Isolation Kit QIAamp DNA Blood Kit (250) (QIAGEN, #50951106). DNA concentration was measured using two methods: Qubit dsDNA BR Assay kit (ThermoFisher Scientific) and Nanodrop 2000C (ThermoFisher Scientific). This last one added DNA quality ratios 260/230 and 280/260, although Qubit Assay Kit allows a better concentration measure. All sample ratios had quality ratios < 1.8 and 2.0 , respectively.

3. Whole-exome sequencing (WES)

DNA samples from familial cases and familial controls selected for each family were sequenced using SOLiD 5500xl platform for sequencing technology (Life Technologies). The sequencing includes exons and flanking intronic regions of most genes of the genome. The readings were captured by Agilent All Exon 50 MB capture kit. After that, the sequences were analyzed with Lifescape software 2.5 version (ThermoFisher). The results were mapped with GRCh37/hg19 human assembly as it were mostly used in annotation. Files with the variant calls (VCF) containing SNV for each individual were generated for further analyses.

4. Pipeline testing for candidate SNV prioritization

For each family, heterozygous SNVs found in all the affected cases with complete phenotype of the family were selected. The 1000 genome project⁹⁵, ExAC database⁹⁶, and Exome Variant Server (EVS)⁹⁷ were used to annotate the MAF and function for each variant (Table S4). All SNVs were filtered by MAF. For MD and autosomal dominant hearing loss (AD-SNHL), variants with $MAF \geq 0.001$ were discarded, since MD has a prevalence of 10–225 cases/100,000 individuals and the low prevalence described for AD-SNHL. For centronuclear myopathy (CNM), variants with $MAF \geq 0.0001$ were also discarded, since CNM is considered as a rare disease with a very low prevalence (1/25,000 males).

The pipeline was designed using different strategies to filter and prioritize SNVs (Figure 6): (a) the calculation of a pathogenic variant (PAVAR) risk composite score; (b) Exomiser v2 software⁹⁸; (c) VAAST annotation tool⁹⁹; (d) a combination of VAAST and Phevor tools¹⁰⁰ and a (e) list with other composite algorithms/tools. However, Phevor returns the same results than VAAST, but ranked by phenotype. In addition, other composite algorithms were used as CADD¹⁰¹ and FATHMM¹⁰². So, the shared candidate variants were selected. All variants were considered as potentially pathogenic according to the ACMG Standards and Guidelines⁵⁵, and all digital resources used are listed in Table S5.

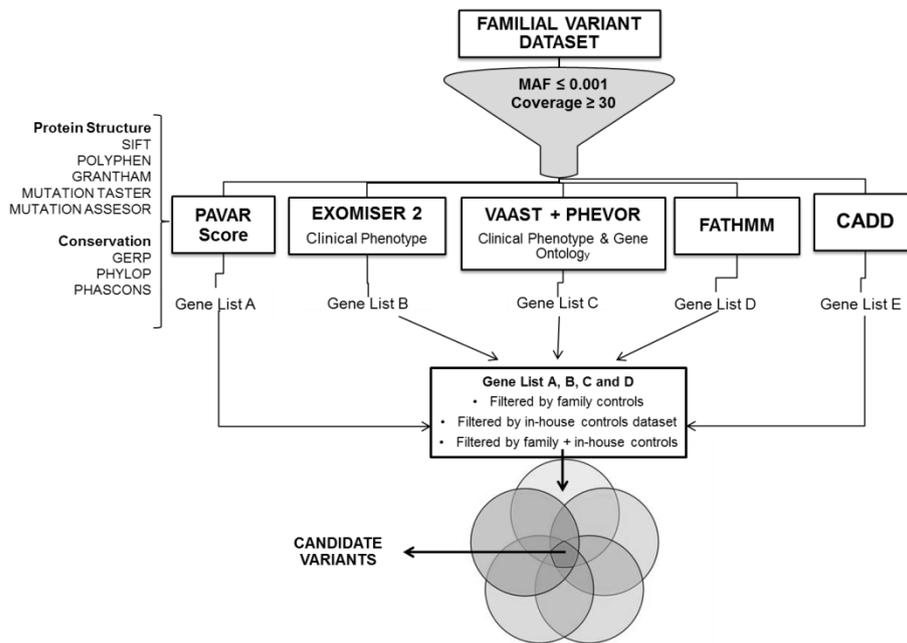


Figure 6. Workflow for candidate variants research in familial datasets.

In some AD diseases, incomplete penetrance was found; subsequently, familial controls could not be used to filter variants. Different control datasets collected for previous projects were used to evaluate the efficiency of our pipeline despite of the observed incomplete penetrance. F = family controls exome dataset, T-F = in-house control data exome dataset without familial control datasets, and T = in-house and family control datasets.

a) Pathogenic variant risk composite score (PAVAR score)

Functional annotation was used to prioritize SNVs, according to the effect on protein structure and phylogenetic conservation. Sequence conservation across species is a major criterium to assess how the variant, and the number of compared species varies according to the tool. To estimate the risk of a SNV to become a pathogenic variant, we used a seven-point scoring system based upon open-access prediction bioinformatics tools. ANNOVAR¹⁰³ and SeattleSeq Annotation¹⁰⁴ tools were used to achieve the score of SIFT (Sort Intolerant from Tolerant)¹⁰⁵, PolyPhen2 (Polymorphism Phenotyping v2)¹⁰⁶, Grantham's Matrix¹⁰⁷, GERP++ (Genomic Evolutionary Rate Profiling)¹⁰⁸, Mutation taster¹⁰⁹, PhastCons¹¹⁰, and PhyloP¹¹¹. The threshold to consider each variant as pathogenic is described in Table S6, according to the default settings suggested for each software developer. PAVAR score is calculated as the sum of the score obtained by seven systems. Each system adds one point if the variant is considered as potentially damaging and zero if it is benign. So, the higher the score is, the high the risk of

pathogenicity for a given variant. PAVAR score cannot be calculated for nonsense variants, since protein structure tools cannot assign any value. Since nonsense variants can modify dramatically the sequence of the protein, they were considered directly as the maximum PAVAR score = 7. All the variants with a score higher than or equal to 5 were not filtered, and they were considered as candidate variants.

b) Exomiser v2 software

Exomiser v2 prioritizes SNVs by comparing the phenotype across species, according to the inheritance pattern, using the mouse and fish as a model organism phenotype. Variant Call Format (VCF) files were analyzed with the following parameters: (a) HPO terms, Vertigo (HP:0002321), Tinnitus (HP:0000360), and Hearing Impairment (HP:0000365), were selected for Clinical Phenotype and (b) AD inheritance model. Since there are only three HPO terms associated with MD according to the public Human Phenotype Ontology database, but no gene is still included on it, the “ExomiserGene Combined Score” generated very low values. So, variants with a threshold $\geq 1.46 \times 10^{-5}$ were considered as candidate variants. Exomiser v2 allows the use of several HPO terms, but Phevor only allows five HPO terms. To compare both systems, only five HPO terms were selected for the benchmarking analyses. The five HPO terms most commonly associated with each disease were selected (Table S7 and S8).

c) VAAST annotation tool

The third approach was to annotate and filter SNVs, according to the dominant inheritance pattern by VAAST software. All case and control VCF files were processed according to the manual provided in the official website. Case files from the same pedigree were combined by the VAAST selection tool (VST) into a single condenser file; SNVs found in all the affected cases were selected. The quality of the resulting files was measured using the background provided: 1KGv3_CG_Div_NHLBI_dbSNP_RefSeq. cdr. A p value >0.05 indicates that there is no significant difference between the files (Table S9). The next step was to search for candidate genes and their potential disease-causing variants. Each family dataset was filtered with the following parameters: (a) dominant inheritance, (b) incomplete penetrance, (c) maximum combined population frequency for the disease-causing alleles $>0.0005^{112}$, and (d) 1×10^6 permutations per analysis to achieve a significant p value after Bonferroni correction. Variants with an alpha error less than or equal to 1 were considered as possibly pathogenic.

d) Phevor tool

In the fourth approach, the list of the resulting genes generated by VAAST tool was uploaded to the Phevor Webtool (phenotype driven variant ontological re-ranking tool) to prioritize candidate genes, according to phenotype and HPO terms¹¹³. To run the analyses for

MD, AD-SNHL, and CNM, the phenotypes were generated in Phevor using HPO terms described in Tables S7 and S8. Exomiser v2 only admits HPO term so to compare with Phevor, Disease Ontology Terms and Gene Ontology Terms were not used. No threshold value was applied in these analyses since the list of variants is generated from pre-filtered variants from VAAST.

e) Combined Annotation-Dependent Depletion (CADD)

CADD v1.3 is a pre-computed score database that is based on classifier algorithms. The major goal of CADD is to predict the deleterious, functionally significant and pathogenic variants from diversified classes of variants by integrative annotations. For each variant, CADD generates the combined annotation score (c-score) as an output and all scores were referenced against the pre-computed c-scores of 8.6 billion possible human SNPs. In CADD scoring criteria, functional variants should possess c-score greater than or equal to 10, whereas damaging variants show the c-score greater than or equal to 20 and the most lethal human variants show the c-score of greater than or equal to 30. To identify causal variants, a score greater than or equal to 15 was considered as potentially pathogenic.

f) Functional Analysis through Hidden Markov Models (FATHMM)

FATHMM predict the functional effects of protein missense mutations by combining sequence conservation within hidden Markov models (HMMs), representing the alignment of homologous sequences and conserved protein domains, with “pathogenicity weights”, representing the overall tolerance of the protein/domain to mutations. The prediction outputs are scored, and the majority of disease-associated AASs fell below -3 and -1.5 threshold. To identify potential causal variants, a score lesser than or equal to -1.5 was considered as potentially pathogenic.

5. Benchmarking procedures

The efficiency of the workflow was tested by benchmarking procedures in different synthetic family datasets with MD. In addition, a group of non-familial healthy controls was tested to identify any bias caused for MD that could influence in the analysis. Moreover, two AD disorders were selected: (a) autosomal dominant sensorineural hearing loss (AD-SNHL) and (b) Central nuclear myopathy (CNM). AD-SNHL has 33 genes associated to the disease, but the phenotype could overlap with MD. To avoid the bias of analyzing AD-SNHL and MD, we selected another disease (CNM) with no overlap in the phenotype with MD. CNM was selected because it has five different genes to perform the benchmarking analysis. The best characterized genes available for AD-SNHL included in the Hereditary Hearing Loss Homepage and CNM

genes described in Orphanet were selected (Table S5). For these genes, exome sequencing data of all exonic variants, in VCF format, were obtained from the public ESP database. Next, 200 variants for each disease were randomly selected to perform benchmarking analyses, but we also checked that at least part of them were described as pathogenic or associated with the disease in human mutation database (HGMD) (Table S1 and S2). To perform the analyses, the synthetic files were built inserting two random variants into real cases VCF files of each family. These synthetic family files for both diseases were analyzed with the six systems. The top 10, 20, and 50 ranked variants for AD- SNHL and CNM were analyzed by each separate system and by all combined strategies.

6. Statistical analysis

Logit regression model was built to assess the accuracy to predict correctly pathogenic variants associated with the phenotype. Firstly, variants selected for benchmarking analysis were classified as pathogenic or benign according to HGMD. The ranks conferred by each system were converted into ranks predictor-wise and normalized in $[0, 1]$, according to the top 10, 20, or 50. ROC curves were generated to determine the ability to predict real causal variants based on models consisting of the combination of the five systems (PAVAR, Exomiser v2, VAAST-Phevor, CADD, and FATHMM) and each individual system. In all the cases, the analyses were performed for the top 10, 20, and 50 ranked variants and using different control datasets to filter for private variants. AUCs were calculated for each ROC curves (Table S3). The statistical differences between AUCs were calculated by analysis of variance. The logit regression models obtained, according to the different combinations and ROC curves, were analyzed with R version 3.0.3 and RStudio version 0.98.1102.

7. Bioinformatics tools for rare SNV selection

VCF files contain raw variant calls and they could be highly heavy due to the large number of common variants in an exome. Therefore, filtering was performed following our previously tested pipeline to obtain significant rare and pathogenic SNVs on two different MD affected families.

8. Control datasets to filter by Spanish population variants

The first step required the filtering of shared non-pathogenic variants from controls. We used familial controls and an in-house control pool (35 control samples previously sequenced by the lab group) and the first stage 578 controls from the CIBERER Spanish Variant Server.

9. Minor allelic frequency filtering

Second filter to avoid common variants was to take variants with a minor allele frequency less than 0.0001. This frequency is defined for AD hearing loss genes.

10. Prioritization

The prioritization of candidate variants was addressed by scoring them according to their effects in protein structure, phylogenetic conservation and other parameters measured by different bioinformatic databases. We created a composite score using an additive method according to the assessment obtained from each tool. So, each time a variant did not pass the pathogenic threshold in each tool it sums a point in a multi-database score we called Pathogenic Variant or PAVAR score. Higher PAVAR scored variants were considered more pathogenic by most different databases. This previous score counts with only seven different databases covering different aspects of variant pathogenicity. Only variants with a total score > 5 were considered as likely pathogenic.

Accuracy of variant prioritization was contrasted by using two different tools for automatic prioritizing: Exomiser v.2 and Variant Annotation Analysis and Search Tool (VAAST) and Phevor plugin. Both tools were tested in conjunction as good predictors by previous testing. Exomiser and VAAST ranked all the variants according to different prediction values, prioritizing most likely pathogenic variants in the list.

11. Linkage analysis

Linkage information derived from WES-common SNVs in the pedigree was calculated in order to reduce candidate variants according to the method described by Gazal, et al. 2016¹¹⁴.

12. Validation by Sanger Sequencing

Candidate variants were validated using Sanger Sequencing. The Sanger sequencer used was a 3130 Genetic Analyzer (Applied Biosystems). Sequencing data was visualized in

Sequence Scanner Software v.1.0 (Applied Biosystems). Sequences were compared with the human consensus reference sequence to confirm variant existence in the cases and absence in controls. Primer pairs flanking SNV were designed using Primer3 software and checked with Primer-BLAST to confirm the exclusive amplification of the segment surrounding candidate variants avoiding any secondary product. Table 5 contains the sequences and characteristics of the chosen primer pairs.

NAME	SEQUENCE	PRODUCT SIZE (pb)
SEMA3D_Fw	GAGAGCTAGACGCCAAGATGTAA	249
SEMA3D_Rv	ATTCAATTAGGCACGTAGACAGG	249
DPT_Fw	AGCGATTCTTCCTGCCATGT	277
DPT_Rv	CAGGAAGTTGGCATTGCAGTTAC	277
SEMA3D.ex_Fw	TCATCTCAAGAAGGCAGTACCTC	213
SEMA3D.ex_Rv	TCTTTCATCTCTTGTGGGGAGTA	213
DPT.ex_Fw	CTGGTGGGAGGAGATCAACAG	250
DPT.ex_Rv	GGTTGTTGCTCCTCGGATATAGT	250

Table 5. Primer pairs designed for candidate variants validation through Sanger sequencing and expression analysis.

13. RNA extraction from cochlea and semicircular canals

Cochlear and semicircular canal human tissues were collected from shwannoma surgery patients (N=2) and were preserved immediately after extraction in RNAlater (Ambion). The tissue was homogenized mechanically by Qiazol (QIAGEN) protocol, lysing the tissue in TyssueLyser LT (QIAGEN) for 5 min at 50Hz.

RNA was isolated following QIAamp RNA miRNeasy Mini Kit (QIAGEN) protocol. RNA yield was measured in Nanodrop (Thermofisher Scientific) spectrophotometer. Quality and integrity of RNA were determined by Agilent RNA 6000 Nano Bioanalyzer chip. QuantiTect Reverse Transcription Kit (QIAGEN) was used for cDNA obtention following manufacturer protocol.

14. Expression analysis in tissue

To measure expression of selected genes in tissue, quantitative real-time PCR was performed in 7900HT PCR system using SYBR green RT-PCR techniques (Life Technologies). Data was analyzed with ABI RQ Manager Software (Applied Biosystems). As a housekeeping gene, HPRT1 was used to determine relative expression levels. Each sample was run in triplicate to calculate DeltaCT values for each sample along with their standard deviation (SD). Statistical analysis was performed with IBM SPSS software (SPSS Inc.). Primers for expression validation are listed in Table 5.

15. Protein 3D modelling

Models for selected proteins were designed through reference proteins in RCSB protein data bank. Reference proteins were used as similarity templates to construct the predicted protein and the altered protein. Templates were managed in SwissModel Deep View Software. This software was used to generate .pdb files too. Prediction software for protein modelling calculates P-value for the global quality of the model assembly, GDT/uGDT (global distance test and unnormalized global distance test), and RMSD for the absolute local quality of each residue in the model.

Visualization and comparison between models were addressed with PyMOL software. PyMOL software was used for calculation of RMS deviation between native model and altered model by superimposition of both predicted models along with THESEUS software. Possible interactions between aminoacid residues in the position involved and other residues were also determined by PyMOL. Quality of the model was measured checking complexity with Ramachandran plots using RAMPAGE webtool software. This plot details residue-by-residue geometry and overall structure geometry, noting when a residue is out of the predicted zone by using its own stereochemical properties.

16. Variant submission

The accession numbers for each tested candidate variant were submitted to ClinVar database (<http://www.ncbi.nlm.nih.gov/clinvar/>). Accession numbers for the variants in SEMA3D and DPT are SCV000266468 and SCV000266469 respectively. GRHL2 variant was also submitted under SCV000266470.

2. Sporadic MD analysis

1. Sporadic samples

Blood samples from 890 Spanish and Portuguese hospitals were obtained from patients with MD. Of them, 830 were sporadic MD cases while 50 were familial MD cases. Ten individuals also had otosclerosis, but shared hypoacusia symptoms. All the samples were collected by experts in Otoneurology following diagnostic criteria for Meniere Disease from 1995 (Table 1). Details for the selected cases are described in Table 6.

N (ind)	930
N (pools)	93
% Women	60
% Men	40
Groups	<i>N (individuals)</i>
South	290
North	480
Mid	90
Familial South	20
Familial North	30
Portugal	20
Diagnosed	<i>N (Individuals)</i>
SMD	830
FMD	50
Otosclerosis	10
Control	40

Table 6. Number of individuals and features of the selected cases and controls for targeted-gene sequencing.

2. DNA extraction

DNA samples from sporadic cases were extracted following the same protocol detailed in DNA extraction for familial MD cases. DNA concentration were also measured using two methods: Qubit dsDNA BR Assay kit (ThermoFisher Scientific) and Nanodrop 2000C (ThermoFisher Scientific). All sample ratios had quality ratios ranging 1.8 and 2.0 in 280/260 and 1.6 to 2.0 in 260/230.

3. Selection of target genes

Targeted genes were selected from a literature search attending to phenotype (hearing profile, comorbid vestibular symptoms) and pathogenicity observations in mouse models. Most of them were selected from HearingLoss.org website gene list for monogenic sensorineural hearing loss. Additional genes were selected if they have been previously found in familial MD, or allelic variations associated with hearing outcome in MD had been described, such as NFKB1 or TLR10 genes. Bibliography facts for those genes appears on Table S13.

Other genes are related to different vestibular disorders. Mitochondrial genes were added due to suspicion of possible mitochondrial involvement. Other relevant bibliography genes related to hearing loss were selected (Table S13).

The custom panel (Panel ID: 39351-1430751809) were designed by the Suredesign webtool (Agilent) to cover the exons and 50 bp in the flanking regions (5' and 3' UTR). This allowed the sequencing around 533.380kb with more than 98.46% coverage. Relevant information about location, size and other characteristics about each gene is added to results Table 9.

4. Preparation of pools

Haloplex Enrichment technology allows the selective amplification of candidate genes and sample pooling, reducing the costs of reagents and increasing sample size. We decided to pool patients according to their geographical origin. Each pool consists in 10 individuals of the same hospital. Only 4 pools were made with 40 healthy controls and 6 with 60 familial cases of different places as internal controls.

DNA concentration and quality were measured on each sample using two methods: Qubit dsDNA BR Assay kit (ThermoFisher Scientific) and Nanodrop 2000C (ThermoFisher Scientific). All samples had quality ratios ranging 1.8 and 2.0 in 280/260 and 1.6 to 2.0 in 260/230.

5. Haloplex protocol (capture, enrichment, barcoding)

The kit allows a total of 96 reactions simultaneously, including control reactions. Premade manufacturer protocol was followed for 12 samples (pools) together, and it was repeated until we obtained all 93 pools. The preparation of libraries including Indexing, capturing, PCR amplification and purification was made following Agilent protocol. Targeted-sequencing was performed in an Illumina NextSeq500 platform. Sequencing adapters were trimmed following manufacturer indications. Requested depth of coverage for the sequencing panel was 250X.

Validation of the protocol and library performing was analyzed with a 2100 Bioanalyzer High Sensitivity DNA Assay kit. Expected concentrations were between 1 and 10 ng/ul. Higher concentrations than 10 ng/ul were diluted 1:10 in 10mM TRIS, 1mM EDTA.

6. Data generation pipelines

Raw data downloaded and sequencing adapters were trimmed following manufacturer indications. Requested depth of coverage for the sequencing panel was 250X. The minimum coverage considered was 30X mean depth for nuclear genes, however mitochondrial sequences reached higher coverages with the enrichment technology. Bioinformatic analyses were performed according to the Good Practices recommended by Genome Analysis ToolKit (<https://software.broadinstitute.org/gatk/>). Mitochondrial genes were analyzed using the same pipeline.

Two methods were used to find differences in how UnifiedGenotyper and HaplotypeCaller (the old and the most recent tools for variant calling in GATK suite) address sequenced pools. Both custom pipelines use BWA-mem aligner and GATK suite tools following the GATK protocol for Variant Calling against GRCh37/hg19 human reference genome. Left normalization for multi-allelic variants were addressed by separated. Calling was made in the first pipeline with UnifiedGenotyper modifying number of chromosomes per sample (per pool, there are 20 chromosomes). The second pipeline used HaplotypeCaller, which cannot allow the same approach, but can automatically address high number of calls with a different approach. Variants with read depth (RD) <10 and genotype quality (GQ)<20 were excluded in all the calling pipelines following recommended hard filtering steps by GATK suite.

A third caller tool, VarScan, was used to filter and annotate quality strand data per variant to compare its output with GATK-based callers. VarScan allows the variant filtering using the information obtained according to each strand polarity. The method retrieves those variants that were only called in one strand, but not in the reverse strand, leading to false positive calls. This step was used as internal quality control to avoid strand bias usually generated in Haloplex data, as it has been reported in other studies

115

7. Positive control SNV validation

Positive control testing was addressed using samples from patients with familial MD with known variants on certain genes. These individuals come from previous familial studies with independently validated variants by Sanger sequencing. SEMA3D and DPT variants from the previous familial study were also included. Known variants were also sequenced and validated by Sanger. Coverage and mapping quality after each pipeline were annotated and measured. Chromatographs from SNV tested and validated are detailed in Figure S4.

8. Selection and prioritization of pathogenic SNV

In order to obtain more information of each SNV, we annotated the merged files using the ANNOVAR tool. Minor allele frequencies (MAF) were obtained for each candidate variant from gnomAD database and ExAC database (total individuals and non-finnish European individuals). Since the estimated prevalence of sporadic MD is 0.75/1000 individuals³⁴, we selected variants with MAF <0.001 for single rare variant analysis and prioritized them according to Combined Annotation Dependent Depletion (CADD) phred score. For burden analysis of common and rare variants, we chose a higher MAF value <0.1. The CVCS database for Spanish MAFs was also used for annotation of exonic rare variants.

KGGseq suite (grass.cgs.hku.hk/limx/kggseq) was used for the selection of rare variants to prioritize the most pathogenic variants according to the integrated model trained algorithm with known pathogenic variants and neutral control variants.

Enrichment analysis for each gene was made with all the exonic variants found with a MAF <0.1. This analysis required to divide total amount of variants in three groups: those present in total ExAC population, those present in NFE population, and finally those present in CVCS Spanish population. These three datasets were used for enrichment analysis comparison.

9. Validation of candidate pathogenic SNV

Candidate SNV were checked and validated in the different pools where they were called using Sanger sequencing. Primer pairs were designed with Primer3 tool and contrasted using Primer-BLAST tool in order to obtain selective and unique primer pairs that can amplify our region of interest. PCR was made following different cycles/temperatures attending to the different regions to amplify. Different amplification cycles were done for each primer pair.

10. Population statistics

Statistical analysis was performed with IBM SPSS v.20 program, Microsoft Excel suite tools, and diverse python and java encoded scripts. Due to overrepresentation of Spanish population in our dataset, most of the selected variants were filtered through exome sequencing data from Spanish controls of CSVS database. The MAF was calculated for each variant in our dataset and rare and previously unreported variants on MD patients were identified in our gene panel. Odd ratios with 95% confidence interval were calculated for each variant using MAF obtained from Spanish population (N=1579), ExAC (N=60706), and ExAC NFE (N=33370) populations as controls.

Gene burden analysis was addressed using 2x2 contingency tables counting total exonic alternate allele counts per gene in our cases against ExAC total controls, ExAC NFE controls and CSVS controls. Odds ratios per gene were calculated along with their 95% confidence intervals using Fisher's exact test and obtaining one-sided p-values. P-values were also corrected for the total amount of variants found per gene comparison following Bonferroni approach.

11. Position of variants in significant enriched genes

Several models were generated for rare variant-enriched domains in significant enriched genes by using the INSIDER modelling tool ¹¹⁶. The selected variants per gene are detailed in results. Prediction values were annotated with their calculated p values.

Results

Results

1. Prioritizing variants in exome datasets

Six prioritizing systems were selected and combined in the pipeline to filter and rank rare variants in exome sequencing data. Two of them were based upon protein structure and sequence conservation across species: (a) an in-house Pathogenic Variant (PAVAR) score and (b) the Variant Annotation Analysis and Search Tool (VAAST), and the other two prioritize according to the Phenotype Ontology information: (c) Exomiser v2 and (d) VAAST-Phevor. And finally, two integrated tools were compared and added to the system CADD and FATHMM.

2. Comparison of prioritizing strategies with FMD exome datasets

Table 1 shows the number of variants obtained for each FMD dataset with the six systems after filtering by several control datasets. We included the number of ranked variants with enough score to be prioritized, according to each of the six systems (thresholds are described in the “Material and methods” section). Mean values obtained for each family dataset were highly variable for each system, and they were dependent on the number of cases and controls available for each family (Table 7).

Family dataset	FMD Exomes (N)	Controls dataset (N)	PAVAR (score \geq 5, N)	Exomiser (Score \geq 1.46×10^{-5} (N))	VAAST (p-value \leq 1)	VAAST-Phevor (p-value \leq 1)	CADD score \geq 15(N)	FATHMM score \leq 1.5(N)
1	3	F (1)	17 (134)	308 (1437)	40	39	15 (38)	7 (35)
		T-F (29)	15 (106)	78 (296)	48	44	18 (36)	7 (34)
		T (30)	10 (68)	42 (175)	27	27	12 (25)	5 (23)
2	2	F (3)	4 (58)	60 (270)	53	22	9 (18)	1 (14)
		T-F (27)	9 (73)	89 (369)	146	135	12 (28)	1 (25)
		T (30)	2 (34)	9 (39)	19	16	5 (13)	0 (11)
3	3	F (2)	9 (68)	151 (862)	23	23	9 (20)	1 (14)
		T-F (28)	13 (92)	67 (309)	38	38	17 (25)	5 (20)
		T (30)	6 (32)	24 (104)	16	16	7 (10)	1 (7)
4	3	F (0)	31 (283)	394 (2198)	54	46	34 (90)	4 (86)
		T (30)	4 (34)	20 (72)	19	17	5 (14)	1 (14)
5	3	F (3)	16 (83)	93 (391)	68	22	7 (20)	1 (15)

		T-F (27)	14 (113)	89 (430)	52	45	14 (35)	7 (28)
		T (30)	5 (36)	18 (67)	11	9	4 (9)	1 (6)
Mean (1-5)	21	F	15.4 ± 10.21 (125)	251.5 ± 143.83 (1032)	47 ± 16.95	30.4 ± 11.33	14.8 9.96	2.8 2.4
		T-F	12.75 ± 2.63 (96)	85 ± 28.66 (351)	71 ± 50.35	65.5 ± 46.44	13.5 4.38	5.0 2.44
		T	5.2 ± 2.97 (51)	31 ± 13.94 (155)	28.2 ± 5.81	5.81 ± 6.44	6.60 2.87	1.6 1.74

Table 7. Number of remaining variants per family dataset according to the filtering strategy.

We selected the top 10, 20, and 50 ranked variants from each prioritizing system and filtered them using the different control datasets (F, T-F, and T) to analyse the concordance between methods. Figure 7 shows the concordance between all systems. Although PAVAR score and VAAST use a different methodology, both systems show the highest concordance rate to filter and prioritize the candidate variants. Between 20 and 55% of ranked variants were matched in top 10, top 20, and top 50. However, the observed variability in the ranked variants between the different systems is caused by the control datasets (F, T-F, or T) used to filter the variants. In contrast, Exomiser v2 and VAAST-Phevor prioritized according to the Phenotype Ontology information (HPO term)¹¹³, but the maximum correlation between systems was 28% when the largest control dataset (T) was used to filter. Therefore, only the variants located in genes previously associated with the phenotype were matched by different systems. Consequently, the combinations of PAVAR, VAAST-Phevor, and Exomiser v2 only matched in few variants (2–26%), which were top ranked and highly related with MD HPO terms. A similar concordance was obtained between the combination of those three and other combined systems as CADD or FATHMM.

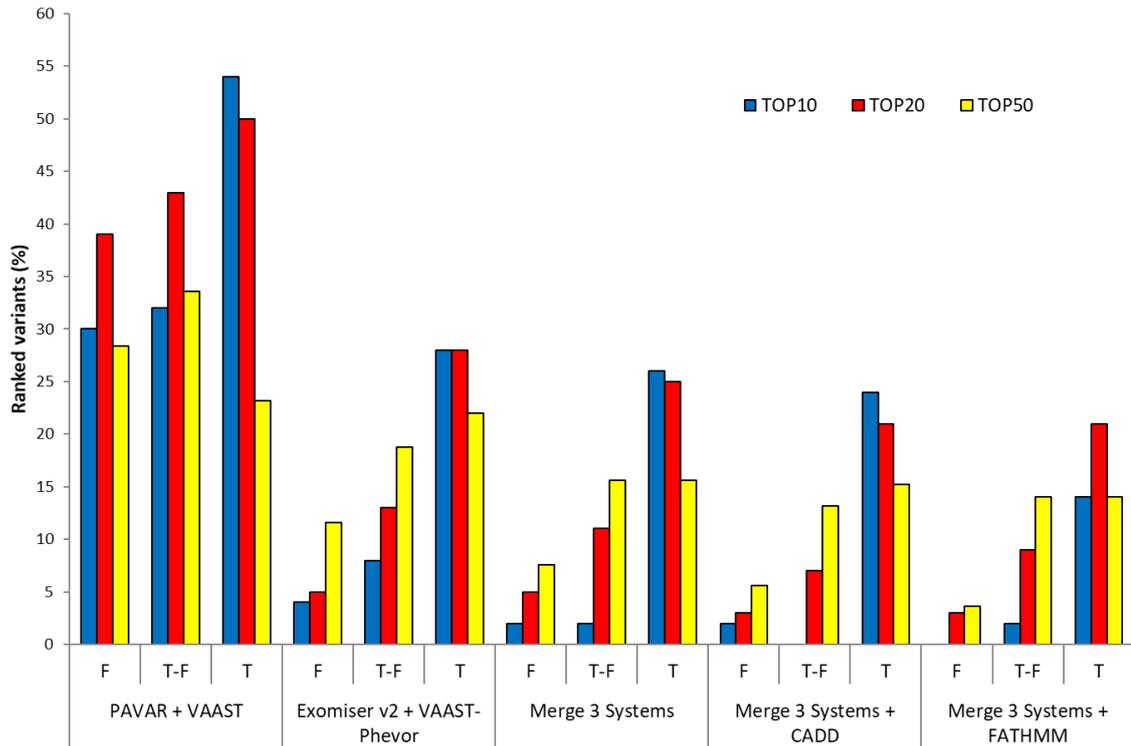


Figure 7. Prioritized variants in FMD datasets. Percentage of variants ranked and shared in top 10 (blue), 20 (red) and 50 (yellow) ranked variants by PAVAR+VAAST; Exomiser+Phevor; combination of three; combination of three+CADD; and combination of three+FATHMM. F = Familial datasets; T-F = in house controls without familial datasets; T = in house controls and familial datasets.

The maximum correlation between CADD and the merge of three systems was 24% in the top 10, whereas for FATHMM was 21% in the top 20. In both cases, this correlation was obtained after using the largest controls' dataset (T) to filter the variants.

3. Benchmark in exome datasets containing variants described in AD-SNHL and CNM genes

We compared the ability of these variant prioritizing tools to identify AD variants in small familial exome data files by a benchmarking procedure. Since the structure of the families as well as the number of cases and controls available for each pedigree could generate a bias in the benchmarking analyses, multiple families were tested.

Figure 3 shows the percentage of ranked variants in the top 10, 20, and 50 by the six systems for both, hearing loss variants (Figure 8a) and CNM variants (Figure 8b). In

the top 10 and 20, the observed percentages were highly variable between each system, particularly depending on the control dataset used.

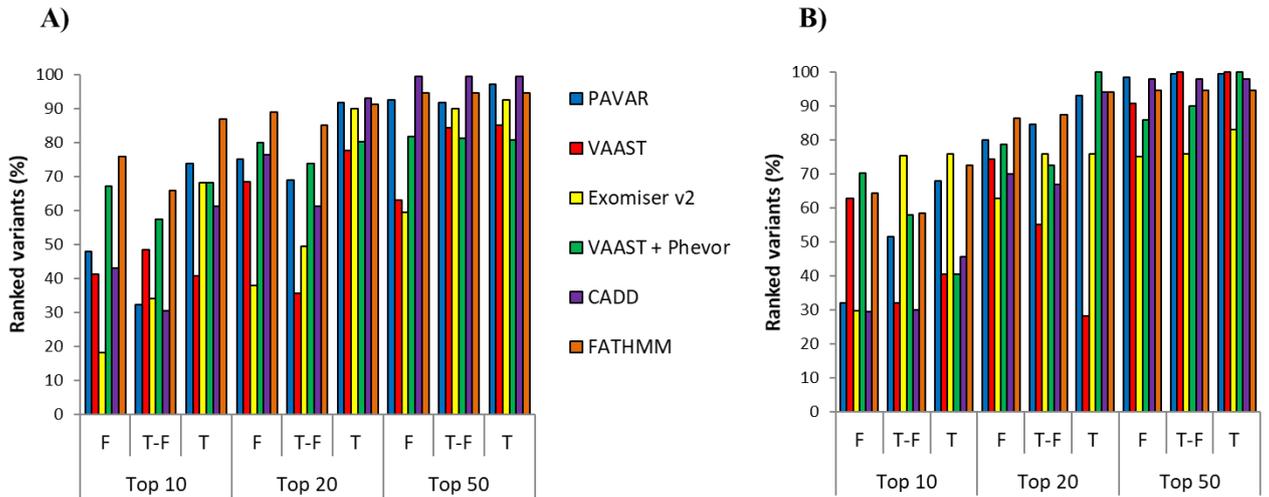


Figure 8. Benchmarking analyses for PAVAR (blue), VAAST (red), Exomiser (yellow), VAAST-Phevor (green), CADD (purple) and FATHMM (orange). Bar charts show the percentage of hearing loss (A) and CNM (B) variants ranked per strategy after filtering by each control dataset. F = Familial controls dataset; T-F = in house controls dataset without familial dataset; T = in house and family control datasets.

Next, we selected the top 10, 20, and 50 ranked variants from each prioritizing system and filtered them for the different datasets (F, T-F, and T) to analyse the concordance between the different methods. Figure 4 illustrates a progressive increase of concordance between systems in the top 10, 20, and 50 ranked variants for both disorders. Exomiser v2 and VAAST-Phevor yielded higher correlations in the top 10 and 20, highlighting that both tools identify similar genes associated with the HPO term for a given phenotype. This pattern was more prominent in the top 10 ranked variants for AD-SNHL datasets in the benchmarking, reaching a 50% of concordance (Figure 9a), whereas in CNM datasets, only 34% of concordance was found (Figure 9b). In contrast, low correlations were obtained between PAVAR score and VAAST (9–33%), mainly in the top 10 ranked, meaning that few variants are considered as candidates by both systems as real pathogenic variants. As a result, potentially pathogenic variants located in genes with HPO terms associated with the disease were shared by PAVAR, Exomiser v2, and VAAST-Phevor and tending to be ranked in the top 10.

A similar percentage was obtained when we add CADD to the combined system. However, the combination of multiple systems with CADD did not reduce the list of candidate variants in the top 10 ranking.

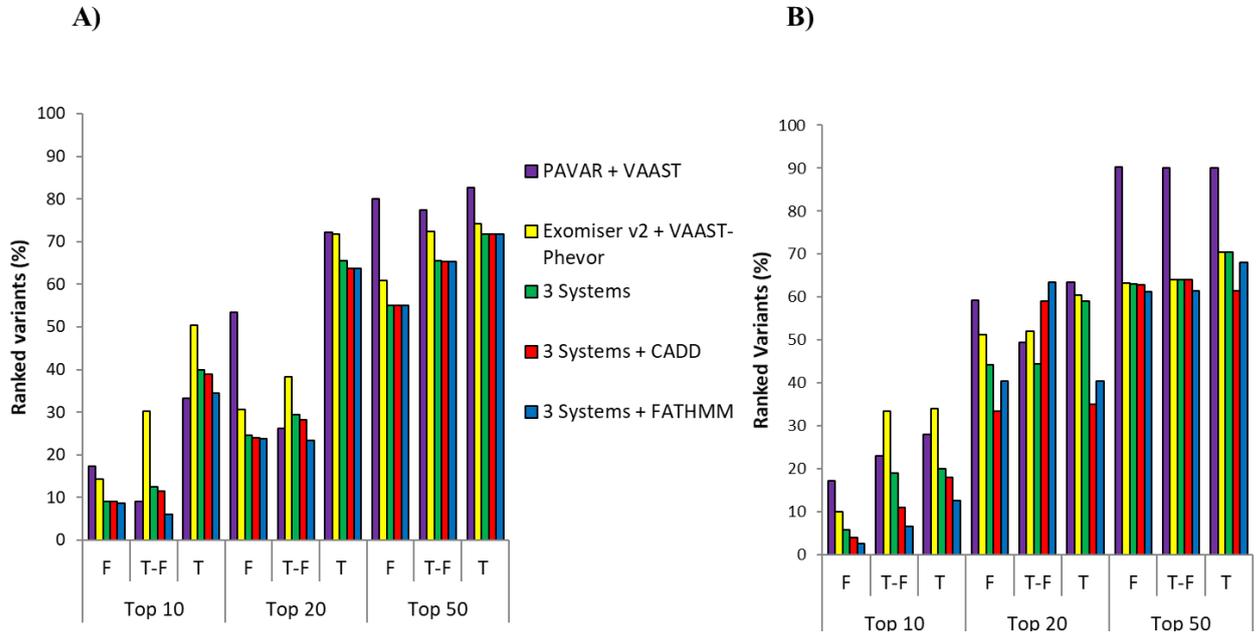


Figure 9. Benchmarking analyses combining prioritizing strategies. Bar charts show the percentage of shared variants for hearing loss (A) and CNM (B) ranked by PAVAR +VAAST (purple), Exomiser v2 + (VAAST-Phevor) (yellow), the three systems (green) the three systems and CADD (red), and the three systems and FATHMM (blue) among the top 10, 20 and 50 after filtering by different control datasets. F = Familial controls dataset; T-F = in house controls dataset without familial dataset; T = in house and family control datasets.

Next, 200 variants were randomly selected for each disease to build synthetic datasets. So, 42% for AD-SNHL and 25.5% CNM were previously described in HGDB as pathogenic (Table S1 and S2). So, multiple logit regression models were performed to assess the accuracy to predict correctly candidate variants associated with each phenotype. The area under the curve (AUC) for each system was calculated to assess the precision and accuracy to identify candidate variants for both diseases in several families (Table S3). On average, the combination of PAVAR, Exomiser v2, VAAST-Phevor, CADD, and FATHMM predicts potentially pathogenic variants associated with the phenotype between 68 and 71% of times in top 10, for both diseases (Figure 10a, b). These results were statistically significantly better than any single method (p values shown in Table S3).

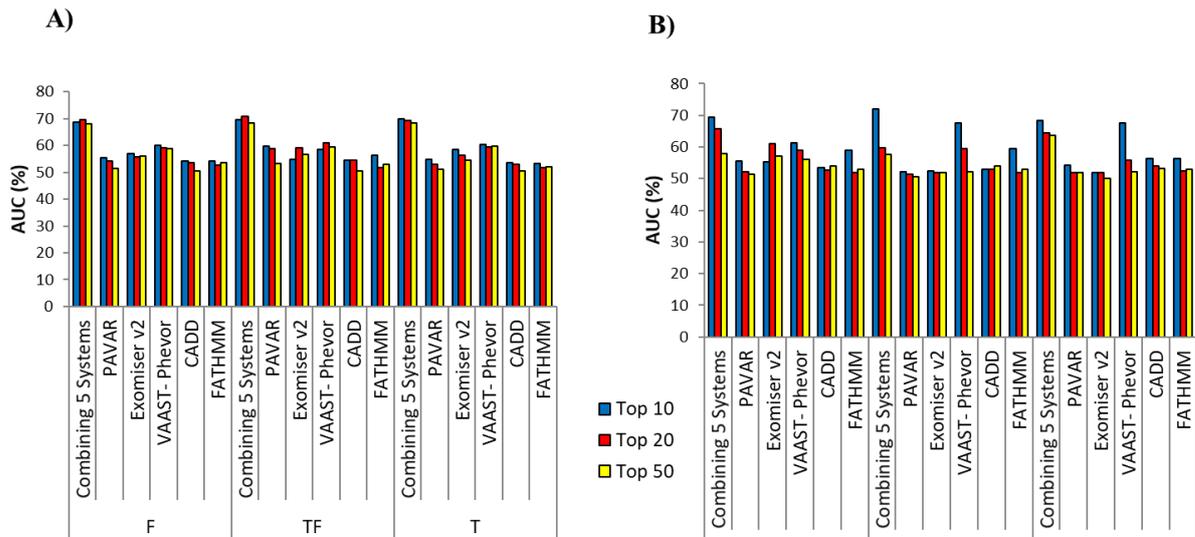


Figure 10. Precision and accuracy of the different systems by calculating AUC. Top 10 (blue), top 20 (red), top 50 (yellow). Bar charts show AUC percentages to identify real pathogenic variants for hearing loss (A) and CNM (B) after control dataset filtering.

4. Families study

A family consisting of 3 siblings (two men and one woman) in the same generation with the complete phenotype was selected (Figure 11a). None of the patients in this family have a history of migraine. The initial case (III: 3) was a 68 years old woman with bilateral MD with an onset at 50 years of age. She has an asymmetrical bilateral pantonal SNHL since the onset of the symptoms, with a mild SNHL at low frequencies and moderate hearing loss from 2000 to 8000 Hz in the left ear. In the right ear, her hearing loss was severe with a pure tone average (PTA) of 58-70 dB. During the first four years of the disease, the patient showed a hearing loss fluctuation in both ears at all frequencies (Figure 12a). She developed a progressive bilateral vestibular hypofunction and, after 15 years, she had reached a caloric areflexia in the right ear and severe hyporeflexia in the left ear. Her follow-up was over 19 years and she was controlled by low sodium diet, high water intake and betahistine. She also experienced occasional episodes of sudden falls that were considered Tumarkin otolithic crises. Although her father did not have a confirmed history of vertigo or early onset hearing loss, she had two maternal uncles with a history of SNHL (II: 1, II: 2). Her older brother (III: 4) developed a right ear MD at 63 years old and he has been followed for 3 years. This patient has a history of high blood pressure and severe obesity (Body Mass Index,

BMI=40). He initially presented a unilateral SNHL in the right ear and, a few months later, he started with episodes of vertigo and ipsilateral fluctuating auditory symptoms. Magnetic resonance imaging ruled out a brainstem or labyrinthine infarction, and he was diagnosed with delayed MD in the right ear, a clinical variant of the MD phenotype. Cervical vestibular evoked myogenic potentials (cVEMPs) showed an absence of

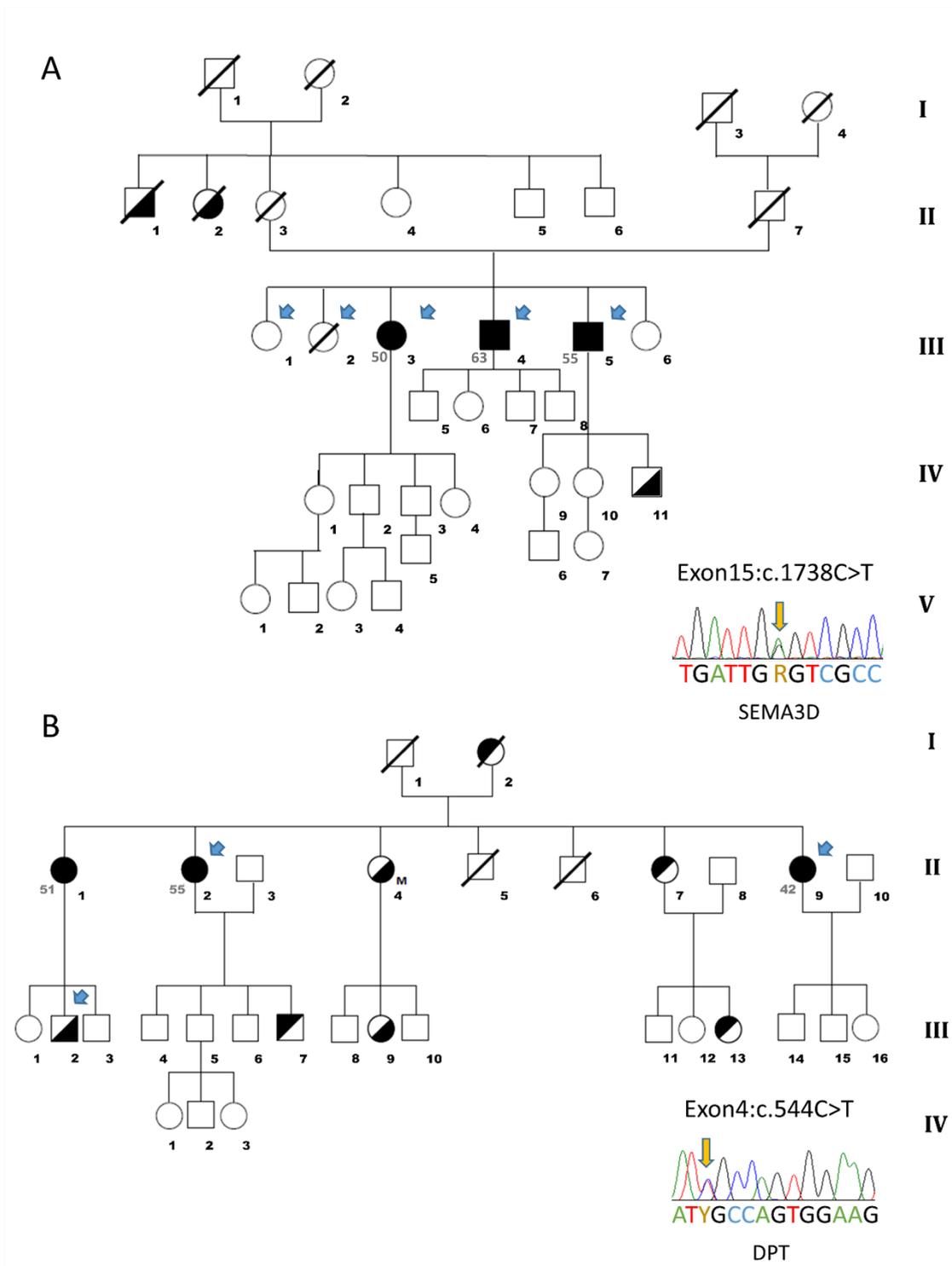


Figure 11. Pedigree for each selected family and their associated and validated variant.

vestibulo-collic reflexes in both ears, confirming the impairment of saccular function, and caloric testing confirmed a bilateral vestibular hypofunction. The younger brother (III: 5) presented a left ear MD that started at 55 years old and he has a follow-up of 9

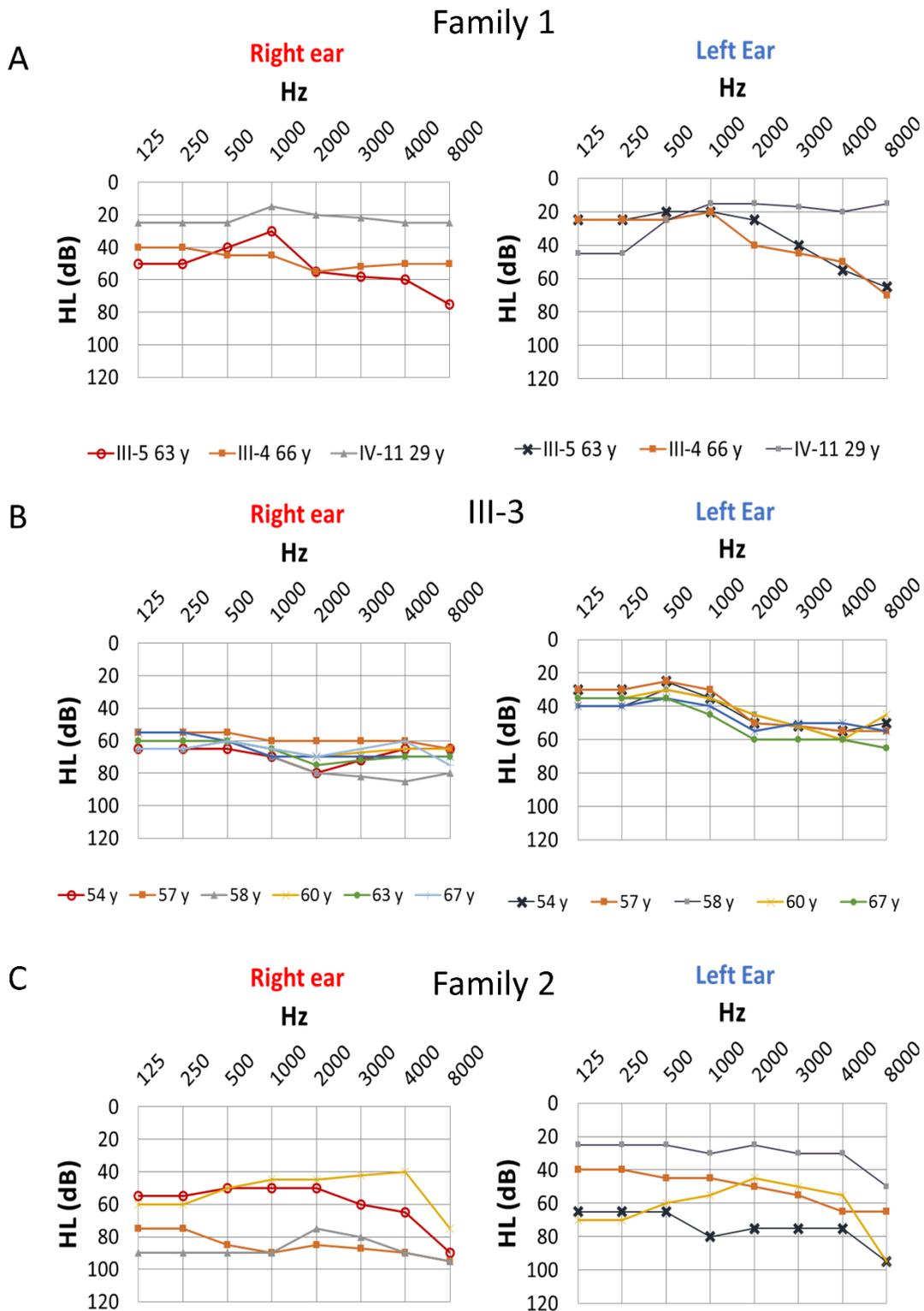


Figure 12. Audiograms and follow up for each case in each studied family. Audiograms were made for each ear.

years. His hearing profile was also a pantonal SNHL in the right ear and a mid-to high frequency SNHL in the left ear. cVEMPs confirmed a bilateral loss of saccular function and caloric tests demonstrated a right vestibular hypofunction in the horizontal canal. He also has high blood pressure and is severely overweight (BMI=35) and was treated with betahistine. This patient experienced a sudden drop of hearing that was treated with oral steroids, showing a sustained recovery of SNHL over several years. During the last two years, he has presented several episodes of sudden falls, highly suggestive of Tumarkin crisis. His 29 years-old son has a low-tone SNHL without vestibular symptoms.

The second family consisted of 3 women with MD in the same generation with an autoimmune background (Figure 11b). None of the patients in this family has a history of migraine. The initial proband was the younger sister (II: 9), who started with a right ear MD at 42 years old, with a low-frequency fluctuating SNHL and tinnitus. Her follow up was 19 years. She experienced a large number of attacks with no response to betahistine. Eight years after onset, she was treated with intratympanic gentamicin in the right ear with an immediate relief of vertigo attacks. cVEMPs and bithermal caloric testing confirmed an ipsilateral vestibular areflexia with normal response in the left ear. Hearing levels were maintained in the right ear with a PTA of 25 dB. Autoimmune screening found anti-ribonucleoprotein antibodies, suggesting a concomitant immune disorder without criteria for systemic lupus erythematosus. The elder sister (II: 1) developed a bilateral MD at 51 years old. She had a history of atrial fibrillation, high blood pressure and type 2 diabetes. She described episodes of vertigo, tinnitus and fluctuating hearing loss. Her bilateral SNHL showed a rapid progression in the left ear and later in the right ear involving all frequencies reaching a PTA of 70dB and 55 dB, respectively, after 5 years of follow up. She has been maintained on a low sodium diet and high water intake for the last two years without new episodes of vertigo. The third sibling (II: 2) was a woman that started with MD at 55 years old. She also had Sjögren syndrome, high blood pressure, type 2 diabetes and ischemic cardiac syndrome. She presented initially with a bilateral diachronic SNHL affecting all frequencies, which started in the right ear and advanced a few years later to include the left ear. She was treated with oral steroids producing a partial recovery of 25dB HL in both ears. After 19 years of follow up, hearing loss was permanent with a PTA of 85dB and 48dB in the

right and left ear, respectively. Two additional members of this pedigree presented SNHL without any vestibular symptoms. The first was a sister of the three affected patients (II: 4) who developed a progressive bilateral, synchronic, low-to-middle frequencies SNHL in her late fifties. She also had high blood pressure, type 2 diabetes, and cardiac failure secondary to an aortic valve double lesion. After 11 years of follow up, the hearing loss was permanent with a PTA of 45dB and 60dB in the right and left ears, respectively. The last patient was the second son of the elderly patient with MD (III: 2), who presented with sudden tinnitus with SNHL in the right ear and a PTA of 80dB at 51 years old. He was treated with oral steroids with no recovery. Currently, he has a permanent hearing loss, persistent tinnitus in the right ear and hyperacusis (Figure 12b).

Candidate variants. A list of candidate SNVs for each family is shown in Table 8. After the filtering and prioritizing process, the three individuals with MD in family 1 segregated one novel missense variant in the exon 15 of *SEMA3D* gene [NM_152754.2]. In addition, missense variants in the *GRHL2* [NM_024915.3] and *TRAK1* [NM_001042646.2] genes were identified in the three cases with definite MD and in one familial control (III: 2). However, linkage analysis using WES data on this pedigree excluded the variants in the *GRHL2* and *TRAK1* genes (Table 8). In the second family, the three sisters with MD shared a rare missense variant in exon 4 of the *DPT* gene [NM_001937.4]. This variant (chr1:g.168665849 G>A, c.544C>T) segregates the hearing loss phenotype, since it was also found in two individuals with incomplete phenotype (II: 4 and III: 2) presenting progressive bilateral SNHL and sudden SNHL, respectively (Table S11 and S12).

FAMILY	CHR	POS	GENE	PAVAR score	VAAST	PHEVOR	EXOMISER	MAF
1	7	84642128	SEMA3D	6	10	5	0.62	NR
	8	102555482	GRHL2	5	26	1	1	NR
	3	52455673	PHF7	5	22	11	0.84	2.48E-05
	4	170634382	CLCN3	5	38	23	0.86	8.24E-05
	3	42264873	TRAK1	6	6	5	0.58	1.63E-05
	6	89974214	GABRR2	6	42	26	0.86	8.24E-06
2	1	168665849	DPT	6	12	2	0.86	2.50E-05

	2	109086855	GCC2	6	96	NS	0	NR
	X	49034780	PRICKLE3	5	33	11	1	NR

Table 8. Candidate SNVs obtained with the different bioinformatics tools after filtering and prioritization process. Positios are relato to the GRCh37 reference.

Protein modelling. The protein sequences of Semaphorin-3D and Dermatopontin were aligned along with their ortholog sequences from different species to compare the degree of conservation of the involved domains for both proteins Figure 13. We observed that our amino-acid substitutions occur on highly conserved regions for each reference alignment. Three-dimensional protein models were constructed following previously crystallized related proteins from human in Protein Data Bank. In the case of Dermatopontin, the closest model described including our involved region was *Im8uA* (Bovine Gamma E model), which shares a 31% sequence identity with our query sequence, including a partial high confidence identity with our region of interest. To model Semaphorin-3D we used a crystallized model for Semaphorin-3A, *Iq47*, with a sequence identity of 58% that includes our amino-acid substitution. A Ramachandran plot for Dermatopontin model shows 11% of the residues (9 from the total query) found in a region disallowed for being considered correct by stereochemical quality, but all of them were found far from our region of interest (Figure S2). Similar considerations were made on a Semaphorin-3D protein model after finding 1.5% residues (8 from the total query) in the disallowed region (Figure S3). Our protein models show how variants can affect tertiary structure of proteins by changing a single amino-acid residue. In the Semaphorin-3D model, the substitution c.1738C>T (p.(Pro580Ser)) does not induce a major structural change due to the similar biochemical properties of both amino-acid residues (Figure 13a). However, the substitution c. 544C>T (p.(Arg182Cys)) in the Dermatopontin predicted model represents a relevant structural change in the properties of the protein (Figure 13b). Superposition for both, Dermatopontin and Semaphorin-3D models shows structural differences between the mutant and wild type proteins. These tools show a higher variability in Dermatopontin superposition models than in SEMA3D models. So, the distribution of the residues close to the variant (p.(Arg182Cys)) is highly affected by the change in the Dermatopontin model (positions 180 to 186 are mostly affected, especially p.Gln183 with a variance of 0.1180). In contrast, (p.(Pro580Ser)) seems to mildly affect the structural variance between the

Semaphorin-3D models (higher variance values reach 0.0001 on p.(Asp579)). Although the change in the sequence does not have any phosphate-binding sites close to the involved zone, the (p.(Arg182Cys)) change may induce the disappearance of some steric bonds, according to the Ramachandran Plot Explorer. However, arginine maintains more steric bonds with its proxy environment than cysteine in our Dermatopontin protein model.

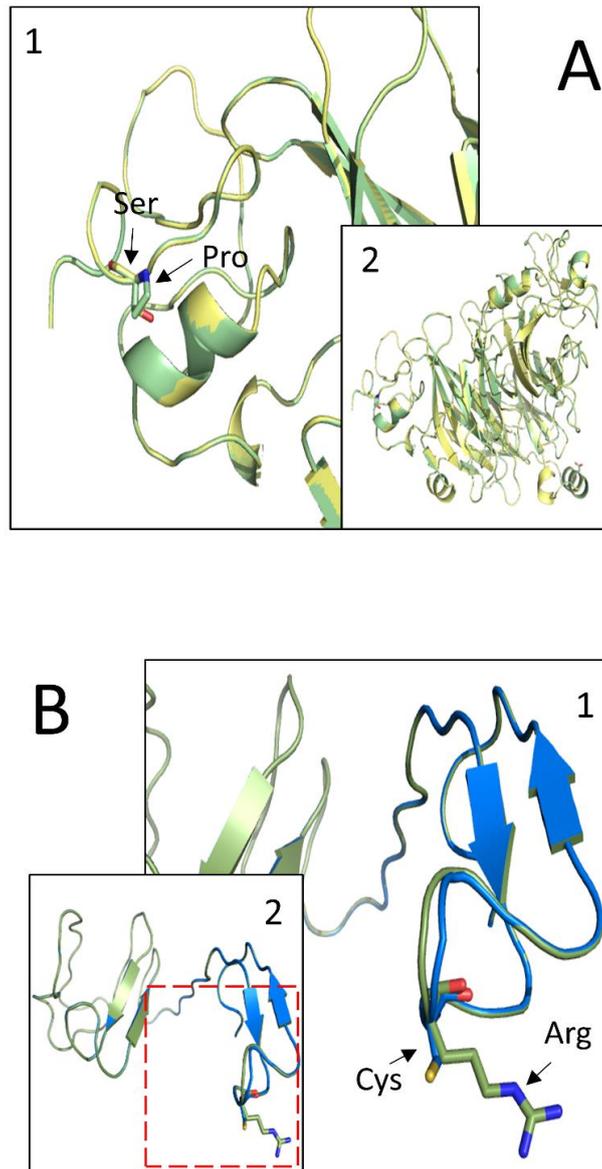


Figure 13. 3D modelling for SEMA3D and DPT variants. **A)** Structural differences between predicted mutated SEMA3D protein model. Superimposition of normal (green) and mutant (yellow). **B)** Structural differences between predicted mutant Dermatopontin protein models. Superimposition of normal Dermatopontin protein model (Green) and mutant Dermatopontin protein model (Blue). Amino-acid changes are marked with arrows.

Validation of expression levels of these genes in the cochlea

We observed significant expression levels of *SEMA3D* and *DPT* genes in human semicircular canals and cochlear tissues (Figure 14).

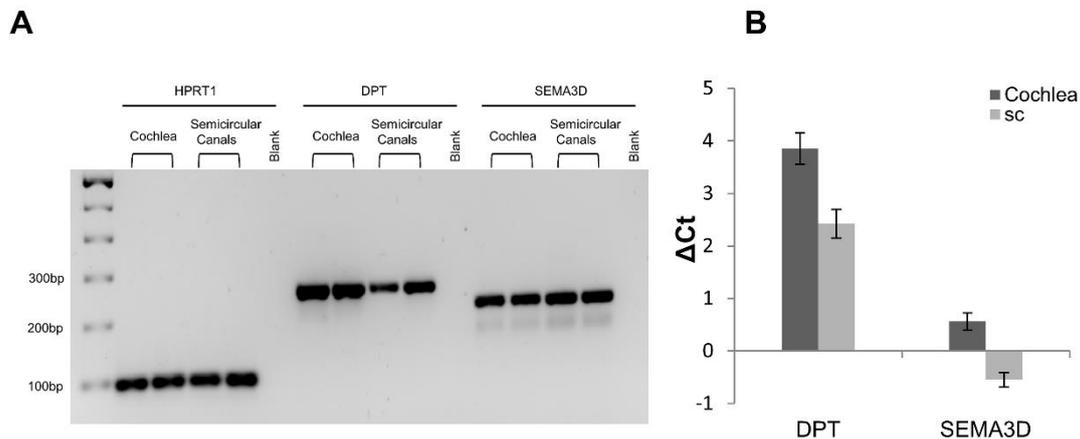


Figure 14. Expression of *DPT* and *SEMA3D* genes in the human cochlea and semicircular canals (sc) by qPCR. **A)** Validation of the expression using primers for *DPT* (250 bp), *SEMA3D* (213 bp) and a housekeeping gene *HPRT1* (92 bp) in agarose gel. Each sample has a technical replication. **B)** ΔCt values of *DPT* (ΔCt cochlea=3.85 \pm 0.3; ΔCt sc=2.42 \pm 0.27) and ΔCt values of *SEMA3D* (ΔCt cochlea=0.56 \pm 0.16; ΔCt sc=-0.55 \pm 0.13).

5. Sporadic cases study

Case and control selection

We performed targeted gene enrichment sequencing on 930 individuals. Sequencing was made in pools with 10 individuals each one for a total of 93 pools. Four pools of samples were prepared with 40 healthy controls, and 5 pools were completed with familial cases to a total of 50 familial cases. The rest of the selected individuals (740) were diagnosed as sporadic MD cases.

Design of potential MD-related genes panel

Haloplex target enrichment technology allows to analyze exonic and intronic regions, as much as some upstream and downstream gene regions that could be relevant

in gene regulation. The design of the custom panel includes a selected list of genes related with SNHL and other symptomatology related to MD (Table 9).

Gene	Genome location (hg19)	Exon N°.	Length (kb)	Gene	Genome location (hg19)	Exon N°.	Length (kb)
ACTG1	chr17:79476947-79479942	3	2666	MT-ATP8	chrM:8315-8621	1	307
ADD1	chr4:2845534-2931853	17	13391	ESPN	chr1:6484798-6521480	14	5685
ARNT2	chr15:80696642-80890328	21	8996	EYA4	chr6:133561686-133853308	23	11034
CCDC50	chr3:191046816-191116509	11	10398	FAM107B	chr10:14560506-14816946	21	9203
CEACAM1 6	chr19:45202371-45214036	7	2583	FAM136A	chr2:70523057-70529272	2	3286
CLDN14	chr21:37832869-37948917	8	3653	GJB2	chr13:20761554-20767164	2	2709
COCH	chr14:31343691-31364321	11	4442	GRHL2	chr8:102504610-102682004	17	7780
DPT	chr1:168664645-168698552	4	2188	KCNE1	chr21:35818936-35884623	7	6481
DTNA	chr18:32073204-32471858	30	13821	KCNE3	chr11:74165836-74178723	3	3707
POU4F3	chr5:145718537-145720133	2	1382	KCNJ10	chr1:160007207-160040101	2	5506
WHRN	chr9:117164310-117267780	14	6942	KCNQ1	chr11:2466171-2870390	19	5756
NR3B2	chr14:76776907-76968228	15	5584	KCNQ4	chr1:41249634-41306174	16	5783
MT-CO2	chrM:7535-8318	1	784	MARVELD2	chr5:68710889-68740207	8	5871
MT-TC	chrM:10008-10453	1	446	MICA	chr6:31367511-31384066	6	4618
MT-ND1	chrM:3256-4311	1	1056	MIF	chr22:24236141-24237464	2	1309
MT-ATP6	chrM:8476-9256	1	781	MSRB3	chr12:65672373-65860737	10	6387
MT-TL1	chrM:3179-3353	1	175	MYH14	chr19:50706835-50813852	43	11331
MT-TV	chrM:1550-1718	1	169	MYO7A	chr11:76839260-76926336	51	15511
MT-TI	chrM:4212-4380	1	169	NFKB1	chr4:103422436-103538509	29	8102
MT-TQ	chrM:4278-4449	1	172	P2RX2	chr12:133195316-133199022	7	2939
MT-TM	chrM:4351-4518	1	168	PNPT1	chr2:55861148-55921095	24	7916
MT-TW	chrM:5461-5628	1	168	PRKCB	chr16:23847250-24231982	20	12059
MT-TA	chrM:5536-5704	1	169	RDX	chr11:110045555-110167497	19	7564
MT-TN	chrM:5606-5778	1	173	SEMA3D	chr7:84624819-84816221	20	10062
MT-ND3	chrM:5710-5875	1	166	SLC12A2	chr5:127419408-127525430	27	13463
MT-TY	chrM:5775-5940	1	166	SLC26A4	chr7:107301030-107358304	24	8300
MT-TS1.	chrM:7395-7563	1	169	THAP1	chr8:42691767-42698524	4	2868
MT-TD	chrM:7467-7634	1	168	TJP2	chr9:71736130-71870174	25	9812
MT-TK	chrM:8244-8413	1	170	TLR10	chr4:38773810-38784661	4	4617

MT-TG	chrM:9940-10107	1	168	TPRN	chr9:140086019-140098695	3	3277
MT-TR	chrM:10354-10518	1	165	TRIOBP	chr22:38092945-38172613	26	15061
MT-TH	chrM:12087-12255	1	169	USH1C	chr11:17515392-17566013	29	6576
MT-TS2	chrM:12156-12314	1	159	USH1G	chr17:72912126-72919408	3	3868
MT-TL2	chrM:12215-12385	1	171	WFS1	chr4:6271526-6305042	8	5313
MT-TT	chrM:15837-16002	1	166				

Table 9. Gene panel designed for Meniere disease using Haloplex targeted-gene enrichment technology. Sizes are detailed in kb per regions.

Expected reads were near 140 million, however only 90 million were obtained per library (64% efficiency, with a mean coverage of 20X). Raw read depth per sample were approximately 390,000. There was a high number of reads of non-relevant regions of the genome.

The enrichment was confirmed with Bioanalyzer High Sensitivity DNA Assay kit and the 2100 Bioanalyzer with 2100 Expert Software. Libraries were validated following the manufacturer recommendations. Peak fragment size between 225 and 525 bp, concentration of the library DNA was calculated between the peaks at 175 and 625 bp. Some pools had to be diluted 1:10 in 10mM Tris-1mM EDTA when their concentration was higher than 10 ng/ul. Validated samples according to the Bioanalyzer profiles were selected and stored for sequencing.

Design validation

For the design validation, we used a control pool with known individuals with six previously validated SNVs by SANGER sequencing. After cutting the Illumina adapter attached to each read (AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTGGCTTCAATCTCGTA) we obtained 4538606 reads (raw). We proceed to use Cutadapt to cut the adapters and make two different trimming approaches according to the manufacturer recommendations. In the first trimming, we cut the adapter and reads shorter than 31 bp. The second trimming was harder and cut the adapter and reads shorter than 150 bp.

Alignment and calling

We compared the parallel trimming approaches following two different alignment tools, BWA mem and Bowtie. Pair-wise coverage was higher than expected due to mitochondrial genes included in the design. However, read mean coverage was around 562 and 833 reads for each pool, quite higher than expected.

Calling was performed for both trimming approaches and both aligners using SAMTOOLS calling and GATK UnifiedGenotyper caller. Total number of calls was annotated. GATK recommended workflow was vastly better in the quality filtering step than SAMTOOLS pipeline, allowing us to avoid a large number of false positives during calling. However, GATK filtering step filters ARNT2 SNV in the not-trimmed pool and in the hard-trimmed pool, although it was detected in the normal-trimmed pool. Finally, we decided to continue the pipeline including BWA mem as aligner and GATK suite as tools for calling and quality filtering. Hard-trimming was avoided and we decided to be conservative using reads larger than 31 bp.

As a last step, we followed the recommendations observed in other works¹¹⁵, making some considerations and corrections to our design. Enzyme cutting seems to cause some false positive calls in some genes. So, to avoid this situation during calling, we trimmed 5 nucleotides on the 3' and 5' extremes of the selected reads (corresponding to the zones recognized by enzymes) (Table 10).

	HALO36	HALO36_TRI	HALO36_TRI2
BWA+GATK	2138	2133	4371
BWA+SAMTOOLS	250340	251800	225096
Bowtie+SAMTOOLS	266010	271041	243024
Bowtie+GATK	4485	2031	2011
BWA+VARSCAN2	3122	2770	2770
Bowtie+VARSCAN2	5336	4258	2897

Table 10. Number of calls in the control pool with different combinations of aligners and callers.

After removing the trimmed adaptors, we tested again the known variants in our control pool. We tested known false positive calls in a pool chosen at random. Not a single previously seen false positive call was called again.

VarScan2 calling was performed following the trimming pipeline in order to obtain a measure of the strandedness of each call. A high number of false positives were still called after GATK UnifiedGenotyper due to the conception of the pooling. VarScan2 allowed us to add a strandedness value to each call, in order to filter those calls that were not found in both strands after alignment. VarScan2 uses a hard filtering for the entire pool of the call.

Agilent published after this approach his own property software to analyze Haloplex data called SureCall 2.1.1.13 software. According to the manufacturer, this software catches this entire false positive call problem. However, the generated output was not a vcf format file to annotate after all. So, we only took the number of final calls as a good marker of how much calls are true positive calls after the previous calling pipelines. The output concordance between Agilent's Surecall Software and VarScan2 calling was 72% lower than the concordance found between GATK and VarScan2 (89%). However, all the validated variants in the control pool were found in SureCall output after default parameters.

Finally, we merge the calling data obtained after GATK calling pipeline and the filtered merge obtained after VarScan2 strandedness step and it was used as a final merge of the SNV to be annotated and analyzed.

6. Rare variants analysis

Single rare variant filtering and analysis

We achieved an average capture efficiency rate (percentage of total on-target reads in total sequenced reads) of 69.01% on the target regions above 30X (minimum depth considered for quality filtering). Mean coverage percentage can be found in Table S15. A total of 2770 SNV in nuclear genes were selected from the raw merged dataset (18961 SNVs) after filtering by quality controls. The analysis workflow is summarized in Figure #. For rare variants analysis, SNV that were found in more than one pool were selected, remaining 1239 variants. After that, we filtered by variants observed in the control pools, leaving only 392 exonic SNV in cases (278 missense, 111 synonymous, 2 stopgain and 1 stoploss).

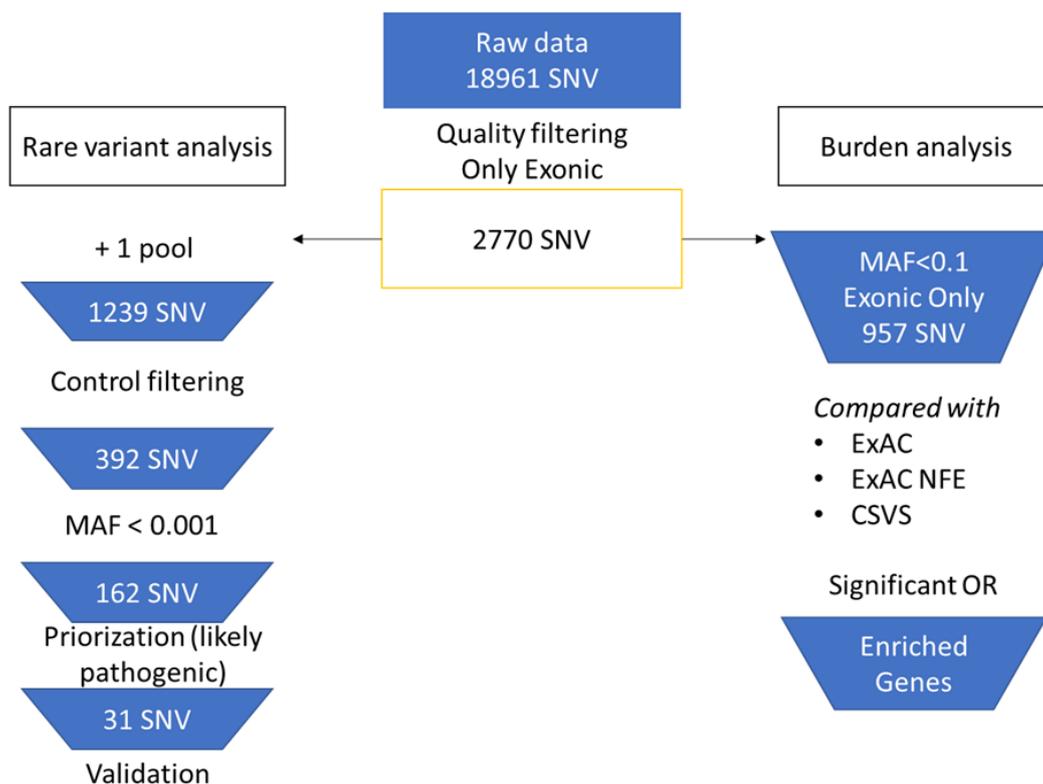


Figure 15. Flowchart of the bioinformatic analysis. On the left branch, rare variant analysis and prioritization workflow. On the right branch, burden analysis of selected variants.

A final set of 162 SNV with a MAF <0.001 were retrieved (143 missense, 18 synonymous, 1 stoploss, 1 stopgain). All the exonic variants were annotated and scored using different prioritization tools. Of them, 136 SNVs were not previously described in any population database and we considered them as potential novel variants.

After prioritizing the exonic variants by CADD phred, 31 rare variants remained (Table 11). Six of them were validated by Sanger sequencing in more than 2 individuals in the following genes: *GJB2*, *ESRRB*, *USH1G*, *SLC26A4* (Table S14). The rest of the variants were considered benign or likely benign since they did not reach the pathogenicity threshold predicted for KGGSeq. However, a novel synonymous variant in the *MARVELD2* gene was found and validated in 3 unrelated individuals.

Chr.	Position	Ref	Alt	Gene	Exonic Function	dbSNP	ExAC MAF	gnomAD MAF	Cohort MAF	CADD phred score
chr1	6488328	C	T	<i>ESPN</i>	Non synonymous	rs143577178	0.0005	0.0005	0.004819	35
chr4	6303574	G	T	<i>WFS1</i>	Synonymous	-	-	-	0.016867	-
chr4	6303946	C	A	<i>WFS1</i>	Non synonymous	-	-	0.004819	22.1	-
chr4	6304133	G	C	<i>WFS1</i>	Non synonymous	-	4.08E-06	0.010843	21	-
chr4	38775706	G	T	<i>TLR10</i>	Non synonymous	-	-	0.003614	20.7	-
chr4	38776070	C	A	<i>TLR10</i>	Non synonymous	-	-	0.024096	14.27	-
chr5	68715821	G	T	<i>MARVELD2</i>	Non synonymous	-	-	0.003614	23.3	-
chr7	107336408	A	C	<i>SLC26A4</i>	Non synonymous	rs200511789	0.0004	0.0004	0.003614	24.3
chr10	14563299	G	T	<i>FAM107B</i>	Non synonymous	-	-	0.003614	24.9	-
chr11	76885923	G	A	<i>MYO7A</i>	Non synonymous	rs781991817	0.0002	0.0003	0.004819	34
chr11	76892617	G	C	<i>MYO7A</i>	Non synonymous	rs200641606	0.0007	0.0007	0.003614	25.5
chr12	65672602	C	T	<i>MSRB3</i>	Synonymous	rs149757878	0.0002	0.0005	0.004819	-
chr13	20763612	C	A	<i>GJB2</i>	Non synonymous	rs72474224	-	0.007229	23.8	-
chr14	76957891	G	A	<i>ESRRB</i>	Non synonymous	rs201344770	0.0003	0.0002	0.008434	23.8
chr14	76966336	G	A	<i>ESRRB</i>	Non synonymous	rs200237229	0.0007	0.0005	0.003614	1.198
chr14	76966347	C	T	<i>ESRRB</i>	Non synonymous	rs201448899	0.001	0.0007	0.008434	15.41
chr16	24046852	C	T	<i>PRKCB</i>	Synonymous	rs115645964	0.0003	0.0003	0.003614	-
chr17	72915919	C	T	<i>USH1G</i>	Non synonymous	rs151242039	0.0006	0.0008	0.003614	8.91
chr17	72916543	T	G	<i>USH1G</i>	Non synonymous	-	-	0.004819	14.31	-
chr17	79478028	G	A	<i>ACTG1</i>	Synonymous	rs187127467	0.0002	0.0003	0.008434	-
chr18	32462094	G	T	<i>DTNA</i>	Non synonymous	rs533568822	2.47E-05	3.66E-05	0.003614	25
chr19	50784974	A	C	<i>MYH14</i>	Non synonymous	-	-	0.003614	20.3	-
chr22	38119197	G	T	<i>TRIOBP</i>	Non synonymous	-	-	0.003614	17.9	-
chr22	38119969	C	A	<i>TRIOBP</i>	Non synonymous	-	-	0.010843	15	-
chr22	38119977	G	T	<i>TRIOBP</i>	Non synonymous	-	-	0.003614	0.022	-
chr22	38120106	A	C	<i>TRIOBP</i>	Non synonymous	-	-	0.003614	2.968	-
chr22	38120116	C	A	<i>TRIOBP</i>	Non synonymous	-	-	0.024096	16.55	-
chr22	38120263	C	A	<i>TRIOBP</i>	Non synonymous	-	-	0.020482	22.3	-
chr22	38120302	C	A	<i>TRIOBP</i>	Non synonymous	-	-	0.003614	23.8	-
chr22	38168691	G	C	<i>TRIOBP</i>	Non synonymous	rs373236761	0.0001	0.0002	0.003614	26.6

Table 11. Prioritized rare SNVs found in the single rare variant analysis for sporadic MD cases. Minor allele frequency for each SNV is detailed as annotated by ExAC and gnomAD (exomes). Pathogenicity prediction is detailed according to CADD phred score.

One of the detected variants corresponded to a known previously documented familial variant in DTNA gene. This variant was validated in one of the two pools where it appeared, confirming that DTNA familial variant could appear in sporadic cases.

Validation of MD candidate rare variants

MD candidate rare variants were validated in their pools using Sanger sequencing method. For each variant, a pair of primer was tested by previous routine PCR amplification. All the primers were made specific for each region to test, avoiding secondary amplification of subproducts using Primer-BLAST tool.

7. Mitochondrial rare variants

Analysis of mitochondrial genes needed a different approach. Coverage of mitochondrial genes were higher than autosomal genes (200X). From the 69 genes added to the panel, 24 genes were mitochondrial genes. After the calling, 3886 variants were found (almost all positions recorded alternative alleles).

A filtering pipeline was made specific to manage mitochondrial data. Control filtering through our control pools left 2178 variants. Using HAPLOGREP2 tool, we could annotate a Soares pathogenicity score to each variant selected after control filtering. Considering a Soares score value >2 as likely pathogenic, we obtained 69 variants likely pathogenic. A conservation score using MITOMAP was used to obtain only variants with a high conservation value, obtaining 27 highly conserved and likely pathogenic variants. A quality filter taking care of strandedness of each variant left 22 variants. In this last selection we separated synonymous variants from non-synonymous obtaining 9 likely pathogenic non-synonymous variants in the genes ND1, ATP8, ATP6 and ND3. Two variants in tRNAs genes for Serine and Alanine were found too, but it's difficulty for validation left both of them for a later study (Table 12).

A population filter was address taking in consideration the population frequencies saved in GenBank (out of about 30k full length mitochondrial sequences) annotated with MITOMASTER tool. Only one of the variants was represented in 0.28% of the total GenBank mitochondrial sequences, so we filtered out. After that, we chose those present in more than 2 pools, obtaining two variants in ND1 and ATP6. A variant

in ND3 was also tested. Although it was only present in one pool, its possible pathogenicity made it relevant to validate (Stop gain codon variant).

Position	Pools	Ref	Alt	Gene	Effect	Frecuency Gen Bank
4011	5	C	A	MT-ND1	Non-syn: N → K	0 (0.00%)
8531	1	A	G	MT-ATP8/6	Non-syn: T → A	19 (0.06%)
8818	5	C	G	MT-ATP6	Non-syn: L → V	0 (0.00%)
8836	1	A	G	MT-ATP6	Non-syn: M → V	90 (0.28%)
8836	1	A	T	MT-ATP6	Non-syn: M → L	0 (0.00%)
10217	1	A	T	MT-ND3	Non-syn: M → I	0 (0.00%)
10262	2	A	T	MT-ND3	Non-syn: E → D	0 (0.00%)
10262	2	A	C	MT-ND3	Non-syn: E → D	0 (0.00%)
10370	1	T	A	MT-ND3	Non-syn: Y → term	0 (0.00%)

Table 12. Candidate variants in mitochondrial DNA.

Posición variante	Número de “pools”	Nt Ref.	Nt Alt.	Locus	Efecto mutagénico	Frecuencia Gen Bank
4011	5	C	A	MT-ND1	Non-syn: N → K	0 (0.00%)
8818	5	C	G	MT-ATP6	Non-syn: L → V	0 (0.00%)
10217	1	A	T	MT-ND3	Non-syn: M → I	0 (0.00%)
10370	1	T	A	MT-ND3	Non-syn: Y → term	0 (0.00%)

Table 13. Filtered candidate mitochondrial variants tested by Sanger sequencing.

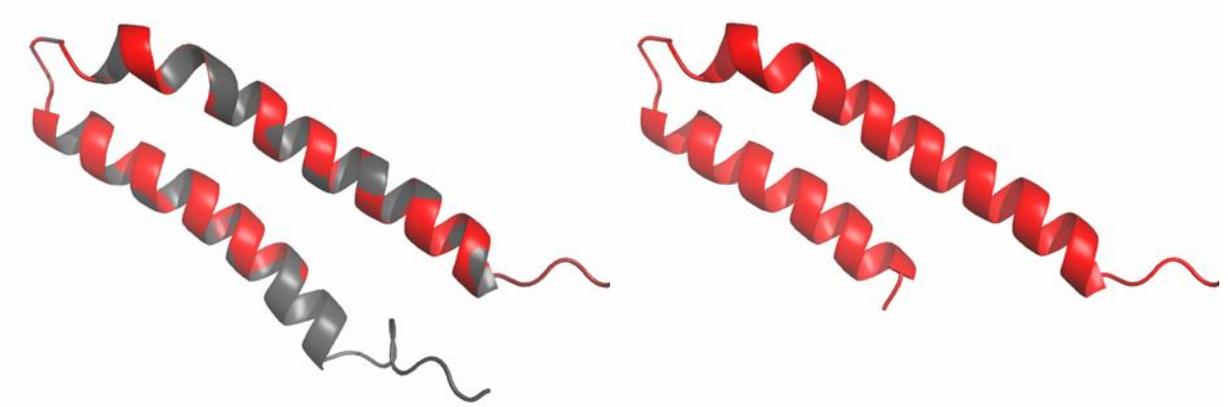


Figure 17. ND3 superposition protein model. *c. 10370T>A (p.(Tyr104TERM))*

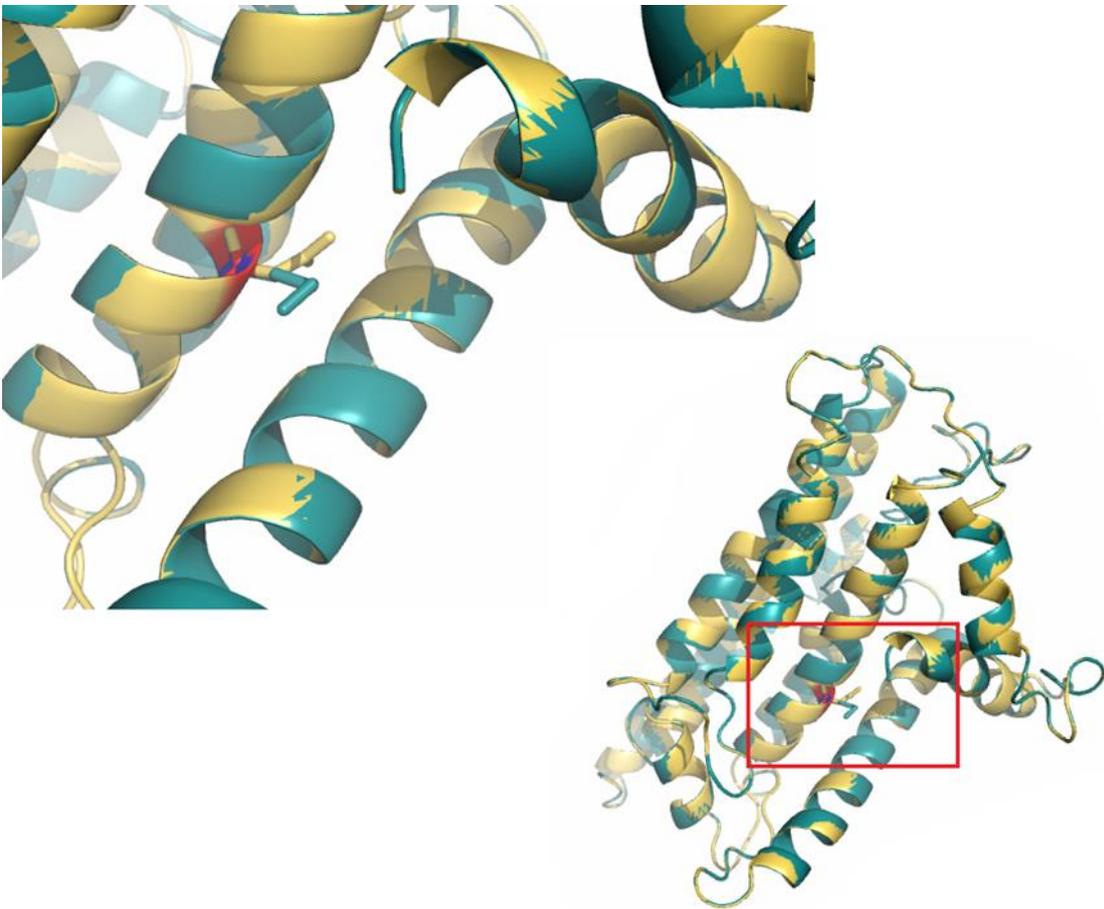


Figure 16. ND1 superposition protein model. *c. 4011C>A (p.(Asn235Lys))*

A 3D simulation of the effect in the protein of these variants was carried out using PyMol software and reference protein models from PDB database. ND1 p.(Asn235Lys) changes a negative charged polar acid for a positive charged polar acid, what supposes possible changes in tertiary bonds along the structure. ND3 p.(Tyr104TERM) provokes a deletion at the end coil of ND3 of 25 aminoacids. ATP6

p.(Leu98Val) is a minor change where an aliphatic amino acid changes for another one, making the structure practically equal.

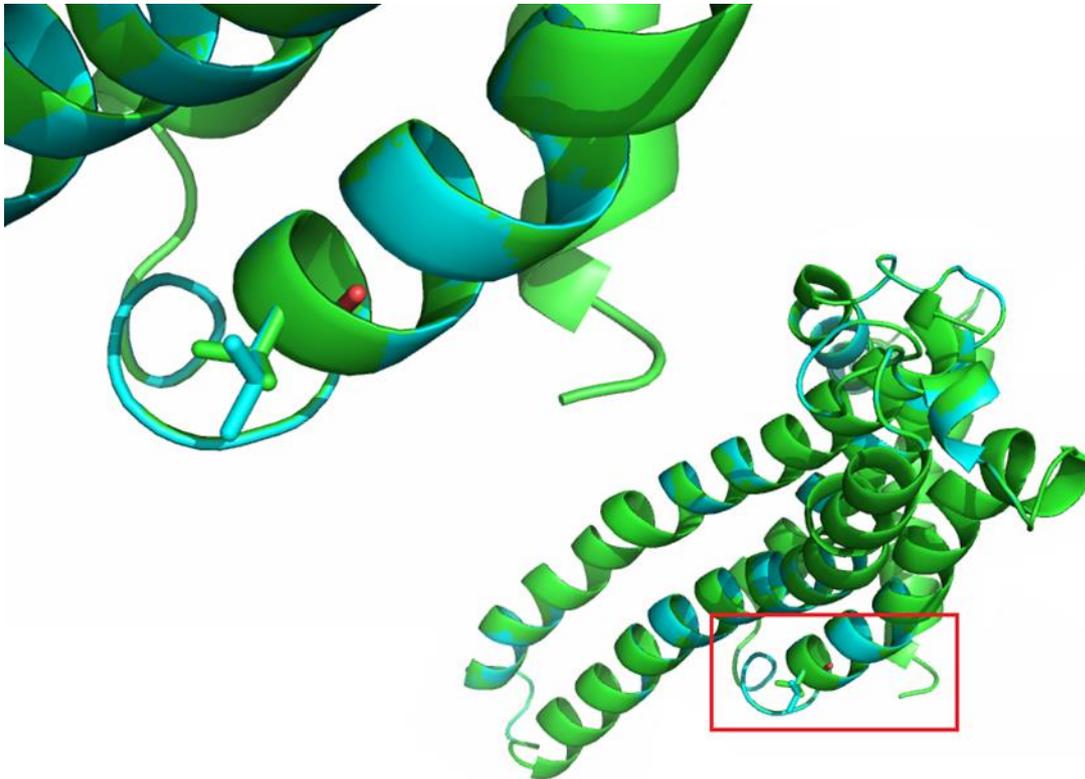


Figure 18. ATP6 superposition protein model. c. 8818C>G (p.(Leu98Val))

A primer pair for each one was made and both were tested in their respective pools using SANGER sequencing (Table 14). Nested PCR was performed to obtain the region of interest in ND1 due to difficulties for amplification of specific ND1 region using entire genomic DNA. ATP6 and ND3 were tested with a simple primer pair. However, no one of them were found in their respective pools, and were considered false positive variants.

GEN	Forward	Reverse	(bp)
MT-ND1 PCR 1	GCCGTTTACTCAATCCTCTGA	TTGTAATGGGTATGGAGACATATCA	1587
MT-ND1	GCCGTTTACTCAATCCTCTGA	TCATATTATGGCCAAGGGTCA	598

PCR 2			
MT-ND3	GATCTAGAAATTGCCCTCCTTTTAC	TAGTAGGGAGGATATGAGGTGTGAG	779
MT-ATP6	CTAGTATCCTTAATCATTTTTATTGCCACA	TAGATGGAGACATACAGAAATAGTCAAACC	1501

Table 14. Primers pair for sequencing candidate variants in mitochondrial genome. ND1 nested PCR needed two primer pairs. Secondary primer pair was used to amplify previous amplified region during primary PCR.

8. Gene burden analysis

The hypothesis is that an enrichment of common and rare variants in certain hearing loss genes may increase the risk to develop MD. To understand the implication of population frequencies in our pool of cases, we managed an association analysis between variants observed in MD cases against their respective frequencies on a healthy population per each gene of the panel. From the total amount of variants, we selected exonic variants for all the targeted genes. We filtered all the variants by a MAF of 0.1 threshold that allowed us to maintain a good emplacement of our population data (common variants including rare variants). A total of 957 exonic variants were evaluated against ExAC total population frequencies, ExAC non-finnish European frequencies, and CSVS Spanish population frequencies. We used ExAC cohorts as control in order to increase the statistical power of the analysis, due to the large cohort included in this database (N(global)=60,706, N(NFE)=33.370).

A gene burden analysis using our selected gene set was performed using these three comparisons. After Bonferroni correction, some genes showed a significant odds ratio in the three comparisons, making them desirable targets on the MD panel (Table 15).

Gene	# variants	Odds ratio ExAC ^a	P value	P corrected ^b	Odds ratio ExAC NFE ^a	P value	P corrected ^b	Odds ratio CSVS ^a	P value	P corrected ^b
MYH14	50	13.54 (5.85-31.37)	1.20E-09	6.01E-08	56.94 (25.13-128.98)	<1.00E-15	<1.00E-15	38.52 (16.95-87.56)	<1.00E-15	<1.00E-15
MYO7A	43	3.11 (1.73-5.57)	1.41E-04	6.08E-03	4.64 (2.65-8.13)	7.72E-08	3.32E-06	3.85 (2.18-6.81)	3.60E-06	1.55E-04
WFS1	36	1.24 (0.6-2.57)	5.67E-01	1	1.84 (0.94-3.62)	7.70E-02	1	3.28 (1.76-6.1)	1.83E-04	6.57E-03
ADD1	29	249.89 (96.17-649.29)	<1.00E-15	<1.00E-15	318.03 (122.45-826)	<1.00E-15	<1.00E-15	747.23 (287.95-1939.08)	<1.00E-15	<1.00E-15
WHRN	28	24.63 (7.96-76.17)	2.68E-08	7.51E-07	39.97 (13.03-122.58)	1.12E-10	3.13E-09	62.91 (20.62-191.98)	3.44E-13	9.64E-12
TPRN	26	36.63 (10.61-126.49)	1.24E-08	3.22E-07	46.93 (13.64-161.49)	1.03E-09	2.68E-08	81.9 (23.93-280.26)	2.24E-12	5.81E-11
USH1G	24	4.04 (1.91-8.51)	2.46E-04	5.90E-03	18.38 (9.26-36.48)	<1.00E-15	<1.00E-15	8.12 (4-16.49)	6.57E-09	1.58E-07
USH1C	22	5.12 (2.58-10.17)	3.13E-06	6.88E-05	10.68 (5.54-20.59)	1.54E-12	3.38E-11	5.12 (2.58-10.18)	3.06E-06	6.73E-05
P2RX2	22	20.34 (9.75-42.44)	<1.00E-15	<1.00E-15	23.74 (11.41-49.41)	<1.00E-15	<1.00E-15	23.51 (11.29-48.94)	<1.00E-15	<1.00E-15
ESPN	19	27.88 (17.06-45.55)	<1.00E-15	<1.00E-15	24.35 (14.88-39.84)	<1.00E-15	<1.00E-15	17.39 (10.59-28.56)	<1.00E-15	<1.00E-15
RDX	17	94.37 (36.06-247)	<1.00E-15	<1.00E-15	57.38 (21.85-150.66)	<1.00E-15	<1.00E-15	75.26 (28.72-197.23)	<1.00E-15	<1.00E-15
TJP2	17	2.78 (1.59-4.88)	3.58E-04	6.09E-03	7.99 (4.79-13.32)	<1.00E-15	<1.00E-15	2.92 (1.67-5.11)	1.67E-04	2.85E-03
SLC26A4	15	1.79 (1.02-3.15)	4.31E-02	6.46E-01	4.14 (2.5-6.86)	3.37E-08	5.06E-07	3.65 (2.19-6.09)	6.62E-07	9.93E-06
ESRRB	14	12.54 (7.26-21.63)	<1.00E-15	<1.00E-15	10.67 (6.16-18.48)	<1.00E-15	<1.00E-15	7.54 (4.31-13.18)	1.41E-12	1.98E-11
DTNA	13	6.07 (4.68-7.87)	<1.00E-15	<1.00E-15	4.87 (3.74-6.35)	<1.00E-15	<1.00E-15	4.78 (3.67-6.23)	<1.00E-15	<1.00E-15
PRKCB	13	27.62 (18.65-40.91)	<1.00E-15	<1.00E-15	17.6 (11.84-26.16)	<1.00E-15	<1.00E-15	40.79 (27.61-60.28)	<1.00E-15	<1.00E-15
SLC12A2	12	21.91 (15.95-30.1)	<1.00E-15	<1.00E-15	55.14 (40.31-75.43)	<1.00E-15	<1.00E-15	79.32 (58.04-108.41)	<1.00E-15	<1.00E-15
KCNQ1	11	7.93 (5.92-10.63)	<1.00E-15	<1.00E-15	13.75 (10.33-18.29)	<1.00E-15	<1.00E-15	6.96 (5.18-9.34)	<1.00E-15	<1.00E-15
NFKB1	10	1.87 (1.3-2.71)	7.81E-04	7.81E-03	1.92 (1.33-2.77)	4.69E-04	4.69E-03	3.79 (2.72-5.29)	4.44E-15	4.44E-14
DPT	10	1.88 (1.2-2.92)	5.36E-03	5.36E-02	8.03 (5.5-11.74)	<1.00E-15	<1.00E-15	4.38 (2.94-6.51)	3.01E-13	3.01E-12
GJB2	9	0.59 (0.31-1.12)	1.06E-01	9.51E-01	2.81 (1.77-4.44)	1.04E-05	9.32E-05	2.5 (1.57-3.98)	1.21E-04	1.09E-03
CEACAM16	9	1.25 (1.04-1.49)	1.56E-02	1.40E-01	1.33 (1.12-1.59)	1.48E-03	1.33E-02	2.45 (2.09-2.87)	<1.00E-15	<1.00E-15
FAM136A	8	31.18 (27.22-35.71)	<1.00E-15	<1.00E-15	38.36 (33.5-43.92)	<1.00E-15	<1.00E-15	14.71 (12.81-16.89)	<1.00E-15	<1.00E-15
GRHL2	8	11.72 (9.35-14.7)	<1.00E-15	<1.00E-15	9.28 (7.39-11.66)	<1.00E-15	<1.00E-15	6.45 (5.11-8.15)	<1.00E-15	<1.00E-15
EYA4	7	36.65 (28.83-46.59)	<1.00E-15	<1.00E-15	25.31 (19.88-32.22)	<1.00E-15	<1.00E-15	7.68 (5.97-9.88)	<1.00E-15	<1.00E-15
COCH	7	3.41 (2.54-4.58)	<1.00E-15	<1.00E-15	3.39 (2.52-4.56)	<1.00E-15	<1.00E-15	3.95 (2.95-5.28)	<1.00E-15	<1.00E-15
CCDC50	5	1.25 (0.99-1.59)	6.22E-02	3.11E-01	1.62 (1.3-2.03)	2.51E-05	1.25E-04	3.22 (2.63-3.94)	<1.00E-15	<1.00E-15
KCNJ10	5	4.38 (3.9-4.93)	<1.00E-15	<1.00E-15	3.14 (2.78-3.55)	<1.00E-15	<1.00E-15	2.3 (2.02-2.61)	<1.00E-15	<1.00E-15
SEMA3D	4	1.56 (1.03-2.38)	3.60E-02	1.44E-01	1.13 (0.72-1.77)	5.84E-01	1	3.79 (2.63-5.47)	1.09E-12	4.35E-12
CLDN14	4	4.51 (2.56-7.94)	1.74E-07	6.97E-07	22.86 (13.56-38.55)	<1.00E-15	<1.00E-15	4.72 (2.69-8.29)	6.61E-08	2.65E-07
POU4F3	4	7.31 (6.53-8.18)	<1.00E-15	<1.00E-15	4.92 (4.38-5.52)	<1.00E-15	<1.00E-15	9.2 (8.23-10.28)	<1.00E-15	<1.00E-15
MSRB3	4	13.35 (11.34-15.71)	<1.00E-15	<1.00E-15	10.85 (9.21-12.79)	<1.00E-15	<1.00E-15	4.06 (3.4-4.83)	<1.00E-15	<1.00E-15

Table 15. Gene burden analysis I. List of 29 genes showing a significant excess of missense exonic variants in patients with sporadic MD, according to the MAF observed in global ExAC population (N=60706), non-Finnish European ExAC population (NFE) (N=33370) and Spanish population from CSVS (N=1579). ^aOdds ratios were calculated in the 95% confidence interval. ^bP values were corrected with Bonferroni method.

We conducted a stronger filtering by the selection missense variants. Selection finishes in a better observation of known variants per gene significantly represented in MD cases. The total number of remaining variants for this analysis described in ExAC were 957 variants.

After Bonferroni correction, some genes showed a significant enrichment of rare variants in the three comparisons, making them candidate genes to be selected for a diagnosis panel for MD (Table 15). Moreover, 6 genes (FAM136A, ADD1, SLC12A2, POU4F3, RDX and PRKCB) presented some novel variants that were validated by Sanger, but they have not been described in global ExAC or CSVS datasets. Although these previously unreported variants could not be sequenced in all the parents of these patients, we considered them as potential de novo variants.

A second variant analysis using the missense variants described in CSVS Spanish population database was made (Table 16). Eighteen genes showed an excess of missense variants (a total of 46 variants, detailed in Table S16). Of note, five genes causing autosomal recessive SNHL showed the highest accumulation of missense variants when they were compared with NFE and Spanish population datasets: SLC26A4, GJB2, CLDN14, ESRRB and USH1G. The variants in these five genes were validated through Sanger sequencing and considered Spanish population-specific variants.

Gene	Number variants	% variants retained	Odds ratio ExAC ^a	P value	P corrected ^b	Odds ratio ExAC NFE ^a	P value	P corrected ^b	Odds ratio CSVS ^a	P value	P corrected ^b
GJB2	6	80	0.5 (0.27-0.93)	2.91E-02	1.75E-01	3.2 (2.12-4.83)	2.75E-08	1.65E-07	2.06 (1.33-3.19)	1.14E-03	6.85E-03
SEMA3D	2	50	1.1 (0.76-1.61)	6.13E-01	1	0.8 (0.53-1.21)	2.85E-01	5.70E-01	2.67 (1.94-3.68)	2.03E-09	4.06E-09
CLDN14	2	50	4.47 (2.55-7.83)	1.67E-07	3.35E-07	23.18 (13.81-38.9)	<1.00E-15	<1.00E-15	4.64 (2.65-8.11)	7.47E-08	1.49E-07
SLC26A4	6	40	1.18 (0.72-1.93)	5.04E-01	1	2.88 (1.89-4.38)	8.13E-07	4.88E-06	2.33 (1.51-3.59)	1.23E-04	7.37E-04
NFKB1	3	30	1.37 (0.99-1.91)	5.92E-02	1.78E-01	1.43 (1.03-1.98)	3.37E-02	1.01E-01	2.73 (2.03-3.66)	2.21E-11	6.62E-11
POU4F3	1	25	1.56 (1.47-1.66)	<1.00E-15	<1.00E-15	1.06 (1-1.14)	6.16E-02	6.16E-02	1.84 (1.73-1.95)	<1.00E-15	<1.00E-15
ESRRB	3	21	4.41 (3.31-5.89)	<1.00E-15	<1.00E-15	3.39 (2.52-4.55)	<1.00E-15	<1.00E-15	1.84 (1.33-2.54)	2.04E-04	6.12E-04
USH1G	5	21	2.51 (1.38-4.56)	2.66E-03	1.33E-02	20.27 (12.06-34.06)	<1.00E-15	<1.00E-15	4.67 (2.68-8.17)	6.11E-08	3.05E-07
CCDC50	1	20	0.54 (0.44-0.66)	7.23E-10	7.23E-10	0.7 (0.59-0.84)	1.41E-04	1.41E-04	1.38 (1.19-1.61)	3.38E-05	3.38E-05
P2RX2	3	14	2.57 (1.93-3.42)	8.71E-11	2.61E-10	3.14 (2.38-4.14)	<1.00E-15	<1.00E-15	2.67 (2.01-3.55)	1.12E-11	3.36E-11
FAM136A	1	13	140.36 (131.25-150.11)	<1.00E-15	<1.00E-15	76.75 (71.75-82.09)	<1.00E-15	<1.00E-15	3.68 (3.41-3.97)	<1.00E-15	<1.00E-15
RDX	2	12	4.49 (3.59-5.61)	<1.00E-15	<1.00E-15	2.69 (2.12-3.4)	<1.00E-15	<1.00E-15	3.33 (2.65-4.2)	<1.00E-15	<1.00E-15
TPRN	3	12	2.24 (1.7-2.96)	1.46E-08	4.37E-08	6.55 (5.1-8.4)	<1.00E-15	<1.00E-15	2.95 (2.26-3.86)	<1.00E-15	<1.00E-15
ESPN	2	11	10.69 (9-12.69)	0.00E+00	<1.00E-15	10.26 (8.64-12.19)	<1.00E-15	<1.00E-15	2.01 (1.64-2.46)	9.24E-12	1.85E-11
SLC12A2	1	8	281.71 (246.32-322.17)	<1.00E-15	<1.00E-15	308.08 (269.39-352.33)	<1.00E-15	<1.00E-15	14.77 (12.86-16.96)	<1.00E-15	<1.00E-15
PRKCB	1	8	15.1 (13.03-17.48)	<1.00E-15	<1.00E-15	10.5 (9.05-12.19)	<1.00E-15	<1.00E-15	5.54 (4.75-6.46)	<1.00E-15	<1.00E-15
MYH14	3	6	14.44 (10.42-20.03)	<1.00E-15	<1.00E-15	28.12 (20.38-38.79)	<1.00E-15	<1.00E-15	5.86 (4.16-8.25)	<1.00E-15	<1.00E-15
ADD1	1	3	563.42 (492.71-644.27)	<1.00E-15	<1.00E-15	308.08 (269.39-352.33)	<1.00E-15	<1.00E-15	14.77 (12.86-16.96)	<1.00E-15	<1.00E-15

Table 16. Gene burden analysis 2. List of 18 genes showing a significant excess of previously reported missense exonic variants in patients with sporadic MD, according to the MAF observed in CSVS Spanish database (N=1579), compared with global ExAC population (N=60706) and non-Finnish European ExAC population (N=33370). In bold, selected genes with higher percentage of variants retained (>20%) and significant OR on Spanish and NFE populations. ^aOdds ratios were calculated in the 95% confidence interval. ^bP values were corrected with Bonferroni method.

Excess of rare variants in hearing loss genes in familial cases

We used exome sequencing datasets from familial MD cases previously reported to search for rare variants identified in our panel in the sporadic cases. Although no single missense variant was found segregated in all the cases in the same family, we found several rare missense variants in at least one case per family in genes such as *GJB2*, *GRHL2*, *TRIOBP*, *RDX*, *KCNQ4*, *WFS1* and *ADD1*. These MD families show phenotypic differences in terms of age of onset, hearing profile and disease progression and the presence of rare variants can be addressed as potential modulators of the phenotype in each familial case (Table 17).

Cases	Family	SNV	Gene	RefSeq	Described in	Pathogenicity	MAF (gnomAD)	MAF (ExAC)
All	1	chr2:70527974 G>A	FAM1 36A	NM_03282 2.2	Requena et al, 2015	Pathogenic		
1	1	chr4:6303197 G>A	WFS1	NM_00600 5.3	No	Conflicting interpretations of pathogenicity	0.004 1	0.004 5
2	1	chr13:2076326 4 C>T	GJB2	NM_00400 4.5	No	Benign/Likely benign	0.009 4	0.010 6
All	1	chr18:3246209 4 G>T	DTNA	NM_00119 8938.1	Requena et al, 2015	Likely pathogenic	3.66E -05	2.47E -05
2	2	chr1:41296865 G>A	KCNQ 4	NM_00470 0.3	No	Unknown significance	2.13E -05	3.49E -05
All	2	chr1:16866584 9 G>A	DPT	NM_00193 7.4	Martin-Sierra et al, 2017	Likely pathogenic	2.03E -05	2.5E- 05
2	2	chr4:2900221 A>G	ADD1	NM_01418 9.3	No	Unknown significance	8.12E -06	8.24E -06
All	3	chr7:84642128 C>T	SEMA 3D	NM_15275 4.2	Martin-Sierra et al, 2017	Pathogenic		
All	3	chr8:10255548 2 G>T	GRHL 2	NM_02491 5.3	No	Unknown significance		
All	4	chr16:2399989 8 G>T	PRKC B	NM_00273 8.6	Martin-Sierra et al, 2016	Pathogenic		
1	4	chr22:3811940 5 C>T	TRIO BP	NM_00103 9141.2	No	Likely benign	4.06E -05	2.49E -05
1	4	chr3:19109866 0 A>G	CCDC 50	NM_17833 5.2	No	Benign	0.006 1	0.006 5

1	4	chr11:1101348 33 T>C	RDX	NM_00290 6.3	No	Unknown significance
---	---	-------------------------	-----	-----------------	----	----------------------

Table 17. Missense variants found in familial MD cases. Variants were retrieved from familial cases segregating a partial phenotype in different families.

Network-interactome analysis

We selected exonic variants from the gene burden analysis to analyze their potential additive effect at the protein-protein interaction interfaces by the tool INSIDER for our selected five genes. However, protein interfaces for ESRRB, CLDN14 and SLC26A4 genes could not be loaded and processed on the database (lacking predicted interfaces on ÉCLAIR database or crystalized protein structures on Protein Data Bank (PDB) database). Of note, most relevant affected interaction is observed in the self-interaction GJB2-GJB2 by the known variants observed in the burden analysis (significant spatial clustering with 4 SNV, $p=0.0009$) rs111033218:G>C (p.Phe83Leu), rs80338945:A>G(p.Leu90Pro), rs374625633:T>C(p.Ile30Val) and rs2274084:C>T(p.Val27Ile) (Figure 19).

Other interactions of interest were founded between the USH1G – USH1C genes, but the involved variants were not located in the known interaction surface of USH1G (Table 18).

To assess if the SNHL genes showing enrichment of missense variants were located in genomic regions with a higher recombination rates, we retrieved recombination rates from deCODE genetics maps for the ESRRB, GJB2, USH1G, CLDN14 and SLC26A4 genes and calculated linkage disequilibrium correlations for candidate missense variants in these five genes. USH1G and ESRRB genes have the highest recombination rates and they seem to be in genomic regions considered as hotspots (Table S17). However, most of the rare missense variants found were not clustered and showed a scattered distribution along the different exons with a low recombination rate (Table S18).

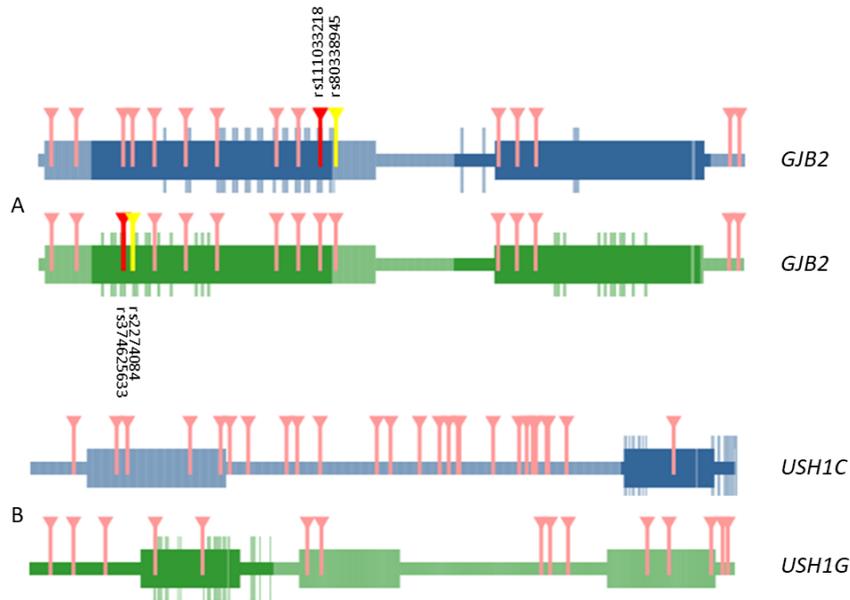


Figure 19. Representation of domains and interactive interface in GJB2 – GJB2 (A) and USH1G – USH1C (B) interaction. Marked in darker color boxes, interactive surfaces of the protein. Aminoacids in the interactive surface of the protein are highlighted in the same color. Variants that affect the interaction regions between both proteins are marked in red and yellow. Only in GJB2 – GJB2 self-interaction missense mutations are relevant in the interaction (dbSNP ids detailed in black). The rest of the variants affecting aminoacids tested in both interactions that are out in the interactive surface region are marked in pink.

	Domain	p-value	Residue	p-value	IC
Q495M9	-1.3	0.1999	N/A	N/A	No significant IC
Q9Y6N9	-0.1	0.9149	0.1	0.8872	No significant IC
Interaction	-0.7	0.2483	-0.8	0.4269	

Table 18. Interaction between both models of USH1G and USH1C. No significant values on domains or residues are found.

Discussion

Discussion

1. Candidate variant selection in singletons and small families

The use of bioinformatics tools to predict the pathogenicity of rare variants in the human genome is a well-established approach in mendelian disorders research. However, most of the family studies lack enough clinical information to segregate the candidate variants or enough number of cases inside an affected family to perform linkage analyses (small pedigrees or difficulty to access all the members of the family). In the case of MD, overlapping phenotypes and incomplete penetrance increase the difficulty of searching and filtering candidate variants⁶⁴. Finally, after a list of genes is obtained, the functional validation of candidate variants is still needed. So, the first goal is to reduce the list of candidate variants to be validated. Therefore, the ranking and prioritization of variants is an issue that need to be solved when prediction output is larger than expected.

Variant calling seems to be the first issue to resolve in variant analysis. A lot of different pipelines have been developed to cover different aspects derived from exome sequencing technology, including recommended workflows as GATK suggests, largely extended along different bioinformatics protocols^{117,118}. However, it is in annotation where we can start looking for biological approach to the variant filtering and selection. Our results show that the combination of multiple bioinformatics tools can be uses to improve the obtention of candidate variants and reduce the list of potential pathogenic variants to the most relevant for a given disease phenotype. The combination of multiple bioinformatics tools is a popular strategy to prioritize rare variants (REF), and our results are consistent with other studies for other prioritizing tools^{119,120}.

The list of candidate variants generated used to be too large for experimental validation. So, a two-steps filtering pipeline using population control cohorts and familial controls is needed to rule out population-specific variants and private familial variants of our list of candidate variants list. Our results show that the combination of

five tools (PAVAR, Exomiser v2, VAAST-Phevor, CADD and FATHMM) reduce the list of variants vastly and facilitates the identification of disease-causing variants.

Discrepancies between all the prioritization systems evaluated were found in the ranked results for all the diseases tested. Consequently, systems based on the same criteria, protein structure and sequence conservation, or phenotype ontology information, were clustered to analyze the concordance between them in the top 10, 20 and 50 ranked variants. Although PAVAR and VAAST used different methodologies, both prioritized variants according to the intrinsic effect on the protein of the variants. Of note, MD, AD-SNHL and CNM showed similar correlation scores between PAVAR and VAAST for the top 10 and 20 ranked variants. Both systems were more concordant when in-house control datasets or the merge of in-house and family control datasets were used to filter. Although familial controls are important to filter private variants, a large control dataset of the same population is more effective to reduce the list of candidate variants list.

In contrast, the concordance between VAAST-Phevor and Exomiser v2 varies depending on the disease studied. Although both systems are based on phenotype, VAAST-Phevor has a balanced score between potential pathogenicity and the association with the phenotype whereas Exomiser v2 assigns more weight to the phenotype than the potential pathogenicity. Diseases with a well-characterized phenotype by several HPO terms or diseases with known involved genes show a high correlation between VAAST-Phevor and Exomiser v2, as our results confirm for AD-SNHL and CNM. However, since MD only has few HPO terms and no gene associated in public databases, our data show a reduced concordance. In particular, our results show that the correlation between both systems in well-known diseases with many HPO terms is double the ones in disorders with limited phenotypic information such as diseases of the ear for all top 10, 20, and 50 ranked variants. Nevertheless, a high concordance between both systems does not indicate that those variants selected are disease-causing variants. The degree of concordance between both systems only demonstrates that the candidate genes are associated with the phenotype, but not necessarily its pathogenicity^{98,99}.

Initially, our pipeline joins both approaches by the identification of variants ranked as potentially pathogenic by the PAVAR score and associated them with the phenotype by Exomiser v2 and VAAST-Phevor. The combination of the three strategies gives few

variants ranked in the top 10 or 20, producing a short list that can be easily validated experimentally, as the ones validated in our lab for different families⁶²⁻⁶⁴. In addition, other combined systems were added to the list to improve the pipeline and to reduce more the number of candidate variants. Logit regression models and benchmarking analyses show that this combination not only reduce the list of candidate variants to be validated; this combined approach is more efficient to predict potential diseases-causing variants than each system separately. This enhanced efficiency is observed independently of the type of control dataset used. Our results confirm previous studies showing that prioritizing tools have less ability to rank variants in disorders with no previously known candidate gene¹²¹. Moreover, we demonstrate that the addition of more HPO terms improves the ranking of candidate genes. So, our pipeline allows to obtain a reduced list of variants when incomplete penetrance is found, and familial control datasets cannot be used.

However, this strategy has a major limitation: reduced phenotypic characterization of AD disorders will decrease the precision of the pipeline. Hence, deep phenotyping and updating the HPO terms in major databases will improve the yield of the system. Nowadays, besides the last update of HPO project, most of the ear diseases still have limited phenotype vocabulary and disease-phenotype annotations¹¹³. Further improvements to the pipeline is necessary to include other types of variants, such as structural variants, synonymous variants or copy number variants.

2. Familial MD

Familial MD is a rare condition and the estimated prevalence is 6-10% among sporadic cases^{60,122}. Some of the described families presented co-segregation with migraine and anticipation^{91,123,124}, but the two Spanish families that we report here do not show these features.

Our study shows that AD-FMD has an incomplete penetrance with variable expressivity and it confirms a clinical heterogeneity in FMD. First, the onset of MD in these families may vary from the most common syndrome (episodic vertigo, unilateral low-to-mid frequencies SNHL and tinnitus; case II: 9 in family 2) to bilateral diachronic

SNHL (the second ear develops the syndrome after several years, e.g., II: 1 and II: 2 in family 2), to bilateral pantonal asymmetrical SNHL (III: 3 and III: 5 in family 1) or delayed MD (hearing loss precedes the vestibular symptoms in months or years, e.g., III: 4 in family 1). Second, the progression of the hearing loss is also variable within each family, ranging from a rapid progression observed in case II: 1 to a slower one found in II: 9 in family 2, which also had an autoimmune background. These differences may rely on the cumulative effect of several regulatory variants in modifier genes. Accordingly, allelic variations in the *MICA*, *NFKB1* or *TLR10* genes have been reported to influence the hearing loss outcome in patients with sporadic MD^{87,91,93}. Third, several relatives in these families presented a partial syndrome with different types of SNHL, including sudden hearing loss (III: 2 in family 2) or pantonal SNHL (II: 2 in family 2) and no vestibular symptoms. These findings observed across different families with MD^{60,61,123} suggest that a) different genes can be involved in the development of the partial or complete phenotype, or b) the interaction of environmental or epigenetic factors can also determine the differences in expressivity within the phenotype.

Our findings have started to define two candidate genes associated with FMD and support the hypothesis of genetic heterogeneity in FMD. Requena *et al.* (2015)⁶² described novel variants in two other candidate genes for autosomal dominant FMD, *FAM136A* and *DTNA*, in a single family consisting of three women with a MD phenotype showing anticipation in consecutive generations.

In this study, we have identified two new candidate genes for FMD: *SEMA3D* and *DPT*, both being extracellular signals, which may be relevant to the formation or maintenance of inner ear structures^{125,126}.

SEMA3D gene encodes an axonal guiding protein, Semaphorin-3D. The protein is secreted and it inhibits the neural growth cones and the endothelial cell motility and migration, regulating cytoskeletal dynamics and cell adhesion¹²⁷. Semaphorins comprise seven different secreted proteins (designated by the letters A–G). These proteins contain a Sema domain involved in the formation of a complex with neuropilin (Nrp) and plexin transmembrane receptors. The Sema3–Nrp–plexin complexes activate the R-ras signaling pathway, and thus decrease the attachment of integrins to the extracellular matrix¹²⁸.

Family 1 presented a novel missense variant at chr7:g.84642128 G>A (c.1738C>T), which generates a (p.(Pro580Ser)) change in Semaphorin-3D. This variant segregated the complete MD phenotype in 3 patients within the same generation, suggesting an autosomal dominant inheritance. These patients were bilaterally affected and two of them responded to steroid therapy, but the functional effect of this novel variant is unknown. The amino-acid substitution occurs at the beginning of the $\alpha 5$ helix on the PSI domain. This PSI domain is an important repeated domain rich in disulphide motifs, found in different extracellular receptors like plexins, integrins and semaphorins¹²⁹. Our substitution seems to affect a non-highly conserved amino-acid on the most conserved part of the sequence (K-R-R-X-R-R-Q-D-V/I-R/K-X-G-D/N-P/A). Although potential interactions cannot be discarded, the sequence seems to be a zone of low probability for protein interactions. We also found in this family a rare missense variant in exon 2 of the *GRHL2* gene (chr8:g.102555482 G>T, c.34G>T), a gene associated with AD non-syndromic SNHL (DFNA28), causing low-frequency SNHL without vestibular symptoms¹³⁰. However, this variant was also found in an elderly relative without any clinical symptom (III: 2), and it had a lower prioritizing score, according to our pipeline.

The *DPT* gene, located at 1q12-q23, consists of 4 exons, with a coding region of 1786 bp. *DPT* encodes Dermatotin, an extracellular matrix protein of 201 AAs that interacts with integrins of the cell surface and proteoglycans such as dermatan-sulphate. Dermatotin is expressed in fibroblasts and it binds to TGF β regulating its activity by the formation of a complex with decorin, a leucine-rich proteoglycan, which also interacts with type I and II collagens¹²⁵. Dermatotin inhibits the formation of decorin-TGF β 1 complexes and probably also binds TGF β 1 on the surface of endothelial cells, thereby maintaining vascular homeostasis^{131,132}. We have identified a rare missense variant at chr1: g.168665849 G>A, c.544C>T (p.(Arg182Cys)), located at exon 4, in the 3 women with MD in the same generation, and in two individuals with SNHL in Family 2. This variant has a MAF = 2.5×10^{-5} in the Exome Aggregation Consortium (ExAC) and it probably produces a functional change in the protein sequence. It has been found in three individuals of Latino, South Asian and European (Non-Finnish) populations. We cannot determine if this rare variant changes the interaction with decorin or TGF β 1. However, it is found inside a highly conserved repeating sequence (D-R-E/Q-W-X-F/Y) of the known domain, specifically at the end of the fourth loop structure. Although no biological characteristics have been

demonstrated before¹³², this could represent important evidence of relevance to the maintenance of the protein functionality. Interestingly, both genes, *SEMA3D* and *DPT*, do not show a differential expression along the axis of the cochlea in the mouse transcriptome¹³³. The absence of a tonotopic gene expression gradient for these genes is consistent with the finding of a pantonal SNHL observed in most patients and it points to a mechanism involving the entire cochlea subsequent to the onset of the disease. However, this is not the case with all MD families. A recently reported family with low-frequency SNHL segregated a novel missense variant in the *PRKCB* gene, encoding PKCB II protein⁶³. Since PKCB II has an expression gradient in the tectal cells along the organ of Corti, that is highest at the apical turn of the cochlea, a loss of function in PKCB II may explain the onset of low-frequency SNHL. The observed phenotype in these families confirms a variable expressivity in FMD and the variants identified in the *SEMA3D* and *DPT* genes increase the number of candidate genes and suggest a genetic heterogeneity.

This Thesis also contribute to complete the list of genes associated with familial MD (Table 19).

Gene	Variant	gnomAD MAF	Validation studies	Reference
<i>COCH</i>	Chr14: 25 mutations	Low	Meniere- like HFHL	24
<i>DTNA</i>	chr18:32462094G>T	3.5x10 ⁻⁵	LCL, novel splice site, rat, drosophila	62
<i>FAM136A</i>	chr2:70527974C>T	Not found	LCL, gene expression	62
<i>PRKCB</i>	chr16: 23999898 G > T	Not found	Mouse, Rat, in silico	63
<i>DPT</i>	Chr7: 84642128 C>T	Not found	In silico	64

<i>SEMA3D</i>	Chr1: 168665849	2.4x10 ⁻⁵	In silico	64
	C>T			

Table 19. Updated familial variants related to FMD.

3. Sporadic MD

3.1. Multiallelic model for MD

The excess of missense variants in SNHL genes may point to core gene for hearing loss in MD. Our hypothesis is that common cis-regulatory variants and rare variants in one or more genes will contribute to the phenotype in MD. The model will need the additive effect of at least a common and a rare variant in the same gene in a given individual¹³⁴. In the simplest bi-allelic hypothesis, we will have:

$$\text{Ind 1} = \text{cv a} + \text{rv z (gene A)}$$

$$\text{Ind 2} = \text{cv b} + \text{rv y (gene B)}$$

$$\text{Ind 3} = \text{cv a} + \text{rv x (gene A)}$$

$$\text{Ind 4} = \text{cv b} + \text{rv w (gene B)}$$

Where *cv* is a common variant and *rv* represents a rare variant; however, this model could be more complex for a single gene:

$$\text{Ind 1} = \text{cv a} + \text{cv c} + \text{rv z (gene A)}$$

$$\text{Ind 2} = \text{cv b} + \text{cv d} + \text{rv z (gene B)}$$

So, several rare variants will be targeting the core genes (*rv z*, *rv x* for gene A; *rv y*, *rv w* for gene B) and common variants in the same genes will explain variable expressivity of the MD phenotype. Finally, in a more complex scenario, it could involve several genes (oligogenic multiallelic hypothesis):

$$\text{Ind } n = \text{cv } a + \text{rv } z (\text{gene A}) + \text{cv } b + \text{rv } y (\text{gene B}) + \dots + \text{cv } n + \text{rv } m (\text{gene N})$$

3.2. Panel design for familial MD

The Genomics England project (<https://www.genomicsengland.co.uk/>) has designed gene panels for the diagnosis of many genetic disorders including familial MD (<https://panelapp.genomicsengland.co.uk/panels/394/>). This panel is in an early stage of development.

For the design of our panel, we chose a total of 69 genes related to hearing loss. Most of genes were selected based on the hearing loss profile (low frequency or pantonal hearing loss). However, more than 90 genes have been described as related to hearing loss, making possible that more hearing loss genes could be involved in the phenotype (described in Hereditary Hearing Loss Homepage: <http://hereditaryhearingloss.org/>). Genetic evidence for hearing loss have been obtained from linkage analyses until the last decades, when classic sequencing techniques gave his way to the NGS techniques^{135,136}. Genetic diagnostic has improved notably in thanks to the developing of NGS technologies. Custom panels and microarrays have been the flag of a new and very powerful age of discovery of novel and rare variants for genetic diagnostic of hearing loss¹³⁷.

Our Thesis project presents an innovative approach. From the standard case-control WES o targeted sequencing studies performed in Mendelian disorders searching for rare variants in a single gene, we have studied the involvement and interaction of several rare variants in the same gene and the interaction with other genes.

One of the most known capture technique for hearing loss is the OtoSCOPE panel, where 66 genes causing hearing loss are used for diagnostics using SureSelect technology in Illumina platform¹³⁸. This panel, thought for diagnostics, recovered a diagnostic rate of 42%. However, data obtained from their study can be used as criteria for population genetics for their target genes. A new study¹³⁹ using the same platform OtoSCOPE analysed a lot more individuals form different planes (1119 patients), however only 141 showed hearing loss. Other known panel, as the one developed by Brownstein et al, 2011¹⁴⁰ included not only HL-related genes on human; they added a good number of HL- related genes in mice.

Our panel was designed considering hearing loss as main symptom shared for all the patients with MD, since the vestibular phenotype is more variable. Due to the clinical heterogeneity of the disease, most of the MD patients may present different ages of onset and other symptoms such as migraine or autoimmune disorders masking the disease phenotype. So, it will be recommendable to select sporadic patients with an early age of onset for future studies.

3.3. Rare variants in hearing loss genes in sporadic MD

It is well known that the frequency of hearing loss related genes depends on the study population. Herein, we present a study for MD patients in Spanish population covering different regions of the Iberian Peninsula including Portugal. As a part of the study, we consider a panel of genes related to hearing loss and other symptoms. Besides from the validated variants in singletons, only a few rare variants such as *ESRRB* rs201448899:C>T, *MARVELD2* rs369265136:G>A, *SLC26A4* rs200511789:A>C and *USH1G* rs151242039:C>T have been validated in more than one sporadic case in the entire cohort.

ESRRB encodes the estrogen-related receptor beta, also known as nuclear receptor subfamily 3, group B, member 2 or NR3B2. This gene encodes for a protein similar to the estrogen receptor but with a different and unknown role. Mutations in the mouse orthologue have been involved in the placental development and autosomal recessive SNHL^{141,142}.

MARVELD2 encodes a protein found in the tight junctions, between epithelial cells. The encoded protein seems to forge barriers between epithelial cells such the ones in the organ of Corti, where, in case of lack of these barriers, normal hearing is affected. Defects in this gene are a cause of deafness autosomal recessive type 49 (DFNB49)^{143,144}.

SLC26A4 gene encodes a protein known as pendrin. Pendrin is a protein very studied in hearing loss¹⁴⁵⁻¹⁴⁸. Its alteration is one of the most common causes of syndromic deafness, an autosomal recessive sensorineural hearing loss. It is highly associated with enlarged vestibular aqueduct syndrome (EVAS)¹⁴⁸⁻¹⁵⁰.

USHIG is a gene translating to a protein that contains three ankyrin domains, a class I PDZ-binding motif and a sterile alpha motif. This protein is well-known to interact with harmonin (*USH1C*), a protein associated with Usher syndrome type 1C¹⁵¹. This protein plays a role in the development and maintenance of the auditory and visual systems and functions in the cohesion of hair bundles formed by inner ear sensory cells^{152,153}. Alterations in the integrity of the protein seem to be cause of Usher syndrome type 1G^{152,154,155}.

However, ESRRB rs201448899:C>T has been observed in more Spanish controls than in global or NFE in ExAC. This increased frequency on the Iberian population when compared with other known largest frequencies as NFE, suggests that this is a population specific variant rather than a MD disease variant. Only the MARVELD2 rs369265136:G>A variant remains as a proper novel related to MD cases. However, the functional effect of a synonymous variant is unknown and functional studies will be required to decipher the relevancy of this variant in MD cases in the future.

3.4. Burden analysis of rare missense variants in sporadic MD

Our results demonstrate a burden of rare missense variants in few SNHL genes, including *GJB2*, *ESRRB*, *CLDN14*, *SLC26A* and *USHIG*. We speculate that the additive effect of several missense variants in the same gene could interact with the same or other genes at the protein level resulting in the hearing loss phenotype.

Population analysis was addressed in order to obtain a better image of our cohort. Besides of the bias that represents the own limited number of genes considered in our panel, we have found a significant increase in the accumulated frequencies for missense variants on several hearing loss genes in Table 16 are higher than in the Iberian population. These findings will demonstrate the involvement of multiple missense variants in the same gene and explain several clinical findings in MD. So, incomplete phenotype or even variable expressivity found in relatives of patients with familial MD could be explained on the differences found in multiple rare variants among individuals of the same family. In addition, some sporadic cases where a single low frequency variant with unknown significance can not explain the phenotype could be singletons individuals with several low frequency variants probably following a compound heterozygous recessive pattern of inheritance. Our results start to decipher the complex

interaction between rare and ultrarare variations (MAF < 0.0001) with common variants in the same or different genes in sporadic MD, adding more evidence to understand the genetic architecture of MD. However, one of the limitations of this study is the lack of availability of a replication cohort with different ethnicity in which to validate these findings.

Another limitation of our dataset is that the method used for resequencing mitochondrial genes may not be able to distinguish mitochondrial from nuclear sequences, as capture panels such as those based in the Haloplex technology may sequence all mitochondrial genome fragment replicas that are dispersed throughout the nuclear genome. Hence, variants observed may not belong to the genes targeted in the mitochondrial genome, but to their pseudogenes in the nuclear genome.

Several hypotheses could explain the excess of missense variants in SNHL genes in MD. First, the variable expressivity of SNHL in MD phenotype, could be the result of additive effect of low frequency or rare variants in the same gene. The combination of low frequency variants in the same gene can be a rare situation, as rare as the disease. As much changes are added to the protein, its integrity could be affected, showing a suboptimal functioning and finally, a loss of function. In our case, GJB2hexamer with a transmembrane channel function, has been determined as possible affected by these changes in their interactions. Previous studies have determined how certain changes in the monomer can affect to the develop of the hexamer hemichannel^{156,157} and its functionability¹⁵⁸. Here, bioinformatics models show how the interaction of low frequency variants found in MD patients can impact the interaction between two connexins monomers, but this effect could be amplified in a model including the 6 connexins that form the connexon. However, this hypothesis is difficult to reconcile with the fact that for some small genes such as GJB2, complex alleles with several point mutations are exceedingly rare.

A second hypothesis points interaction of common and rare variants in one or several genes in the disease phenotype, following its complex disease definition^{159,160}. So, the excess of rare variants will be targeting core genes for hearing loss in MD. In this case, high significant genes in our panel could be added to the pool of possible targets of the disease although not a single candidate variant could be enough to explain the disease. So, the interaction between cis-regulatory variants with rare variant in some of our candidate genes and other, a priori, not related SNHL genes could be relevant in

the expressivity. USH1G interacts with USH1C, known gene involved in Usher syndrome¹⁵¹. USH1G have been observed to have a minor role on Usher syndrome in Spanish population¹⁶¹, but not in MD, even though they share similar hearing loss profile. Although no one of the missense variants in USH1G were in an interaction domain, this could be a step considering interaction between different proteins as a main factor to develop a mild phenotype. This hypothesis was reinforced through the data found in familial cases. For instance, the variant rs748718975 in DPT gene was only associated with the SNHL phenotype in the family where it was described, but these cases showed different characteristics in the age of onset or hearing loss outcome. These differences between the cases can be explained with other variants found in KCNQ4 (rs574794136:G>A) and ADD1 (rs372777117:A>G) genes, although these variants were previously described as variants of unknown significance. So, the variant rs574794136:G>A was found in two sisters with MD, but not in the third one, that was carrier of rs372777117:A>G. This excess of rare variants in certain genes observed in familial cases could explain the differences in expressivity in a given family.

This panel was made as an early screening diagnostic panel. Here we can show how certain SNHL gene variants can be related to MD in the Iberian population and show that multiple rare variation in the same gene should be considered as likely pathogenic. Although there are large differences in the coverage for some genes between the MD panel and the WES databases, these are not the ones with excess of missense variants. Our results may contribute to define new criteria for the genetic diagnosis of MD.

A future new diagnostic panel for familial and sporadic MD will include some of the genes that we have investigated.

Conclusions

Conclusions

1. A pipeline combining multiple variant prioritization algorithms and tools is an excellent method to be used in small family-based studies. We have tested and confirmed that this workflow can reduce the number of variants in exome datasets with incomplete phenotypes without the use of familial controls to test private variants. This approach allows the study of small families or with incomplete data in the pedigree.
2. Familial MD shows genetic heterogeneity with incomplete penetrance and variable expressivity. We have characterized two new families with MD cases illustrating the clinical heterogeneity of the disease. The analysis determined two variants, one in *SEMA3D* in one family, and another one in *DPT* segregating the phenotype in the second.
3. The study of sporadic cases determined how some SNHL-related genes have an enrichment of missense variants in Spanish population. These genes include well known genes as *GJB2*, *ESRRB*, *CLDN14*, *USH1G* and *SLC26A4*. Some of the variants found in *GJB2* have relevancy in the hexamer formation, integrity of the protein and its function as transmembrane channel.

Bibliography

1. Marieb, E. N. & Hoehn, K. *Human anatomy & physiology*. (Pearson Education, 2009).
2. Hudspeth, A. J. Making an effort to listen: mechanical amplification in the ear. *Neuron* **59**, 530–545 (2008).
3. Purves, D. *et al.* Two Kinds of Hair Cells in the Cochlea. (2001).
4. Wan, G., Corfas, G. & Stone, J. S. Inner ear supporting cells: rethinking the silent majority. *Semin. Cell Dev. Biol.* **24**, 448–459 (2013).
5. Müller, U. Cadherins and Mechanotransduction by Hair Cells. *Curr. Opin. Cell Biol.* **20**, 557–566 (2008).
6. Chan, D. K. & Hudspeth, A. J. Ca²⁺ current-driven nonlinear amplification by the mammalian cochlea in vitro. *Nat. Neurosci.* **8**, 149–155 (2005).
7. Gulley, R. L. & Reese, T. S. Intercellular junctions in the reticular lamina of the organ of Corti. *J. Neurocytol.* **5**, 479–507 (1976).
8. Slepecky, N. B., Henderson, C. G. & Saha, S. Post-translational modifications of tubulin suggest that dynamic microtubules are present in sensory cells and stable microtubules are present in supporting cells of the mammalian cochlea. *Hear. Res.* **91**, 136–147 (1995).
9. Lopez-Escamez, J. A. *et al.* Diagnostic criteria for Menière’s disease. *J. Vestib. Res. Equilib. Orientat.* **25**, 1–7 (2015).
10. Clinical Anatomy and Physiology of the Vestibular System - Oxford Medicine. Available at: <http://oxfordmedicine.com/view/10.1093/med/9780199608997.001.0001/med-9780199608997-chapter-8>. (Accessed: 10th May 2018)
11. Sajjadi, H. & Paparella, M. M. Meniere’s disease. *Lancet Lond. Engl.* **372**, 406–414 (2008).
12. House, J. W., Doherty, J. K., Fisher, L. M., Derebery, M. J. & Berliner, K. I. Meniere’s disease: prevalence of contralateral ear involvement. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **27**, 355–361 (2006).
13. Paparella, M. M. & Griebie, M. S. Bilaterality of Meniere’s Disease. *Acta Otolaryngol. (Stockh.)* **97**, 233–237 (1984).

14. Kamei, T. Delayed endolymphatic hydrops as a clinical entity. *Int. Tinnitus J.* **10**, 137–143 (2004).
15. Müller, U. & Barr-Gillespie, P. G. New treatment options for hearing loss. *Nat. Rev. Drug Discov.* **14**, 346–365 (2015).
16. Baguley, D. M., Cope, T. E. & McFerran, D. J. Functional auditory disorders. *Handb. Clin. Neurol.* **139**, 367–378 (2016).
17. McCormack, A., Edmondson-Jones, M., Somerset, S. & Hall, D. A systematic review of the reporting of tinnitus prevalence and severity. *Hear. Res.* **337**, 70–79 (2016).
18. Lempert, T. *et al.* Vestibular migraine: diagnostic criteria. *J. Vestib. Res. Equilib. Orientat.* **22**, 167–172 (2012).
19. Requena, T., Espinosa-Sanchez, J. M. & Lopez-Escamez, J. A. Genetics of dizziness: cerebellar and vestibular disorders. *Curr. Opin. Neurol.* **27**, 98–104 (2014).
20. Mijovic, T., Zeitouni, A. & Colmegna, I. Autoimmune sensorineural hearing loss: the otology–rheumatology interface. *Rheumatology* **52**, 780–789 (2013).
21. George, D. L. & Pradhan, S. Idiopathic sensorineural hearing disorders in adults--a pragmatic approach. *Nat. Rev. Rheumatol.* **5**, 505–512 (2009).
22. Committee on Hearing and Equilibrium guidelines for the diagnosis and evaluation of therapy in Menière’s disease. American Academy of Otolaryngology-Head and Neck Foundation, Inc. *Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.* **113**, 181–185 (1995).
23. Robertson, N. G. *et al.* Mapping and characterization of a novel cochlear gene in human and in mouse: a positional candidate gene for a deafness disorder, DFNA9. *Genomics* **46**, 345–354 (1997).
24. Robertson, N. G. *et al.* Mutations in a novel cochlear gene cause DFNA9, a human nonsyndromic deafness with vestibular dysfunction. *Nat. Genet.* **20**, 299–303 (1998).
25. Bae, S.-H. *et al.* Identification of pathogenic mechanisms of COCH mutations, abolished cochlin secretion, and intracellular aggregate formation: genotype-phenotype correlations in DFNA9 deafness and vestibular disorder. *Hum. Mutat.* **35**, 1506–1513 (2014).

26. Foster, C. A. & Breeze, R. E. The Meniere attack: an ischemia/reperfusion disorder of inner ear sensory tissues. *Med. Hypotheses* **81**, 1108–1115 (2013).
27. Kamat, P. K., Kalani, A., Metreveli, N., Tyagi, S. C. & Tyagi, N. A possible molecular mechanism of hearing loss during cerebral ischemia in mice. *Can. J. Physiol. Pharmacol.* **93**, 505–516 (2015).
28. Friberg, U., Stahle, J. & Svedberg, A. The natural course of Meniere's disease. *Acta Oto-Laryngol. Suppl.* **406**, 72–77 (1984).
29. Alexander, T. H. & Harris, J. P. Current epidemiology of Meniere's syndrome. *Otolaryngol. Clin. North Am.* **43**, 965–970 (2010).
30. Tyrrell, J. S., Whinney, D. J. D., Ukoumunne, O. C., Fleming, L. E. & Osborne, N. J. Prevalence, associated factors, and comorbid conditions for Ménière's disease. *Ear Hear.* **35**, e162-169 (2014).
31. Morrison, A. W. & Johnson, K. J. Genetics (molecular biology) and Meniere's disease. *Otolaryngol. Clin. North Am.* **35**, 497–516 (2002).
32. Kotimäki, J., Sorri, M., Aantaa, E. & Nuutinen, J. Prevalence of Meniere disease in Finland. *The Laryngoscope* **109**, 748–753 (1999).
33. Shojaku, H. & Watanabe, Y. The prevalence of definite cases of Ménière's disease in the Hida and Nishikubiki districts of central Japan: a survey of relatively isolated areas of medical care. *Acta Oto-Laryngol. Suppl.* **528**, 94–96 (1997).
34. Morales Angulo, C., Gómez Castellanos, R., García Mantilla, J., Bezos Capelastegui, J. T. & Carrera, F. [Epidemiology of Ménière's disease in Cantabria]. *Acta Otorrinolaringol. Esp.* **54**, 601–605 (2003).
35. Hallpike, C. S. & Cairns, H. Observations on the Pathology of Ménière's Syndrome: (Section of Otology). *Proc. R. Soc. Med.* **31**, 1317–1336 (1938).
36. Paparella, M. M. Pathology of Meniere's disease. *Ann. Otol. Rhinol. Laryngol. Suppl.* **112**, 31–35 (1984).
37. Paparella, M. M. Pathogenesis of Meniere's disease and Meniere's syndrome. *Acta Oto-Laryngol. Suppl.* **406**, 10–25 (1984).

38. Nacional, S. E. de O. y de P. C.-F. (España) C. *Enfermedad de Ménière: desde las ciencias básicas hacia la medicina clínica : ponencia oficial del LX Congreso Nacional de la Sociedad Española de Otorrinolaringología y Patología Cérvico-Facial 2009*. (EUROMEDICE, 2009).
39. Yoda, S. *et al.* Round window membrane in Ménière's disease: a human temporal bone study. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **32**, 147–151 (2011).
40. Semaan, M. T., Alagramam, K. N. & Megerian, C. A. The basic science of Meniere's disease and endolymphatic hydrops. *Curr. Opin. Otolaryngol. Head Neck Surg.* **13**, 301–307 (2005).
41. Merchant, S. N., Adams, J. C. & Nadol, J. B. Pathophysiology of Meniere's syndrome: are symptoms caused by endolymphatic hydrops? *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **26**, 74–81 (2005).
42. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122**, 565–581 (2008).
43. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
44. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
45. Pei, B. *et al.* The GENCODE pseudogene resource. *Genome Biol.* **13**, R51 (2012).
46. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001).
47. Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C. & Gelbart, W. M. *Transcription: an overview of gene regulation in eukaryotes*. (2000).
48. Lagerkvist, U. 'Two out of three': an alternative method for codon reading. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 1759–1762 (1978).
49. Park, J.-H. *et al.* Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18026–18031 (2011).

50. Buske, O. J., Manickaraj, A., Mital, S., Ray, P. N. & Brudno, M. Identification of deleterious synonymous variants in human genomes. *Bioinforma. Oxf. Engl.* **29**, 1843–1850 (2013).
51. Kim, B.-Y., Park, J. H., Jo, H.-Y., Koo, S. K. & Park, M.-H. Optimized detection of insertions/deletions (INDELs) in whole-exome sequencing data. *PLoS One* **12**, e0182272 (2017).
52. Zhang, N., Huang, T. & Cai, Y.-D. Discriminating between deleterious and neutral non-frameshifting indels based on protein interaction networks and hybrid properties. *Mol. Genet. Genomics MGG* **290**, 343–352 (2015).
53. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
54. Puig, M., Casillas, S., Villatoro, S. & Cáceres, M. Human inversions and their functional consequences. *Brief. Funct. Genomics* **14**, 369–379 (2015).
55. Richards, S. *et al.* Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).
56. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).
57. Nakagawa, H. & Fujita, M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* **109**, 513–522 (2018).
58. Goh, G. & Choi, M. Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. *Genomics Inform.* **10**, 214–219 (2012).
59. Birgeron, L., Gustavson, K.-H. & Stahle, J. FAMILIAL MENIERE’S DISEASE: A GENETIC INVESTIGATION. *Otol. Neurotol.* **8**, 319 (1987).
60. Requena, T. *et al.* Familial clustering and genetic heterogeneity in Meniere’s disease. *Clin. Genet.* **85**, 245–252 (2014).

61. Hietikko, E., Kotimäki, J., Sorri, M. & Männikkö, M. High incidence of Meniere-like symptoms in relatives of Meniere patients in the areas of Oulu University Hospital and Kainuu Central Hospital in Finland. *Eur. J. Med. Genet.* **56**, 279–285 (2013).
62. Requena, T. *et al.* Identification of two novel mutations in FAM136A and DTNA genes in autosomal-dominant familial Meniere’s disease. *Hum. Mol. Genet.* **24**, 1119–1126 (2015).
63. Martín-Sierra, C. *et al.* A novel missense variant in PRKCB segregates low-frequency hearing loss in an autosomal dominant family with Meniere’s disease. *Hum. Mol. Genet.* **25**, 3407–3415 (2016).
64. Martín-Sierra, C. *et al.* Variable expressivity and genetic heterogeneity involving DPT and SEMA3D genes in autosomal dominant familial Meniere’s disease. *Eur. J. Hum. Genet. EJHG* **25**, 200–207 (2017).
65. Arweiler-Harbeck, D., Horsthemke, B., Jahnke, K. & Hennies, H. C. Genetic aspects of familial Ménière’s disease. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **32**, 695–700 (2011).
66. Klar, J., Frykholm, C., Friberg, U. & Dahl, N. A Meniere’s disease gene linked to chromosome 12p12.3. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* **141B**, 463–467 (2006).
67. Koyama, S., Mitsuishi, Y., Bibee, K., Watanabe, I. & Terasaki, P. I. HLA Associations with Meniere’s Disease. *Acta Otolaryngol. (Stockh.)* **113**, 575–578 (1993).
68. Melchiorri, L. *et al.* Human leukocyte antigen-A, -B, -C and -DR alleles and soluble human leukocyte antigen class I serum level in Ménière’s disease. *Acta Oto-Laryngol. Suppl.* 26–29 (2002).
69. López-Escámez, J. A., López-Nevot, A., Cortes, R., Ramal, L. & López-Nevot, M. A. Expression of A, B, C and DR antigens in definite Meniere’s disease in a Spanish population. *Eur. Arch. Otorhinolaryngol.* **259**, 347–350 (2002).
70. Mhatre, A. N. *et al.* Aquaporin-2 expression in the mammalian cochlea and investigation of its role in Meniere’s disease. *Hear. Res.* **170**, 59–69 (2002).

71. Lynch, M. *et al.* Structural and mutational analysis of antiqutin as a candidate gene for Ménière disease. *Am. J. Med. Genet.* **110**, 397–399 (2002).
72. Doi, K. *et al.* Ménière's disease is associated with single nucleotide polymorphisms in the human potassium channel genes, KCNE1 and KCNE3. *ORL J. Oto-Rhino-Laryngol. Its Relat. Spec.* **67**, 289–293 (2005).
73. Lopez-Escamez, J. *et al.* HLA-DRB1*1101 Allele May Be Associated With Bilateral Ménière's Disease in Southern European Population. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **28**, 891–5 (2007).
74. Teggi, R. *et al.* Gly460Trp alpha-adducin mutation as a possible mechanism leading to endolymphatic hydrops in Ménière's syndrome. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **29**, 824–828 (2008).
75. Kawaguchi, S., Hagiwara, A. & Suzuki, M. Polymorphic analysis of the heat-shock protein 70 gene (HSPA1A) in Ménière's disease. *Acta Otolaryngol. (Stockh.)* **128**, 1173–1177 (2008).
76. Vrabec, J. T., Liu, L., Li, B. & Leal, S. M. Sequence variants in host cell factor C1 are associated with Ménière's disease. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **29**, 561–566 (2008).
77. Lopez-Escamez, J. A. *et al.* Poly(ADP-ribose) polymerase-1 (PARP-1) longer alleles spanning the promoter region may confer protection to bilateral Meniere's disease. *Acta Otolaryngol. (Stockh.)* **129**, 1222–1225 (2009).
78. Candreia, C., Schmuziger, N. & Gürtler, N. Molecular analysis of aquaporin genes 1 to 4 in patients with Ménière's disease. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **26**, 787–792 (2010).
79. Maekawa, C. *et al.* Expression and translocation of aquaporin-2 in the endolymphatic sac in patients with Meniere's disease. *J. Neuroendocrinol.* **22**, 1157–1164 (2010).
80. Campbell, C. A. *et al.* Polymorphisms in KCNE1 or KCNE3 are not associated with Ménière disease in the Caucasian population. *Am. J. Med. Genet. A.* **152A**, 67–74 (2010).

81. Lopez-Escamez, J. A. *et al.* Association of a functional polymorphism of PTPN22 encoding a lymphoid protein phosphatase in bilateral Meniere's disease. *The Laryngoscope* **120**, 103–107 (2010).
82. Khorsandi, M.-T. *et al.* Associations between HLA-C Alleles and Definite Meniere's Disease. *Iran. J. Allergy Asthma Immunol.* **10**, 119–122 (2011).
83. Hietikko, E. *et al.* Finnish familial Meniere disease is not linked to chromosome 12p12.3, and anticipation and cosegregation with migraine are not common findings. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **13**, 415–420 (2011).
84. Furuta, T. *et al.* Association of interleukin-1 gene polymorphisms with sudden sensorineural hearing loss and Ménière's disease. *Int. J. Immunogenet.* **38**, 249–254 (2011).
85. Lopez-Escamez, J. A. *et al.* Polymorphisms of CD16A and CD32 Fc γ receptors and circulating immune complexes in Ménière's disease: a case-control study. *BMC Med. Genet.* **12**, 2 (2011).
86. Gazquez, I. *et al.* Functional variants in NOS1 and NOS2A are not associated with progressive hearing loss in Ménière's disease in a European Caucasian population. *DNA Cell Biol.* **30**, 699–708 (2011).
87. Gazquez, I. *et al.* MICA-STR A.4 is associated with slower hearing loss progression in patients with Ménière's disease. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **33**, 223–229 (2012).
88. Hietikko, E., Kotimäki, J., Okuloff, A., Sorri, M. & Männikkö, M. A replication study on proposed candidate genes in Ménière's disease, and a review of the current status of genetic studies. *Int. J. Audiol.* **51**, 841–845 (2012).
89. Yazdani, N. *et al.* Association between MIF gene variation and Meniere's disease. *Int. J. Immunogenet.* **40**, 488–491 (2013).
90. Gázquez, I. *et al.* Functional variants of MIF, INFG and TFNA genes are not associated with disease susceptibility or hearing loss progression in patients with Ménière's disease.

- Eur. Arch. Oto-Rhino-Laryngol. Off. J. Eur. Fed. Oto-Rhino-Laryngol. Soc. EUFOS Affil. Ger. Soc. Oto-Rhino-Laryngol. - Head Neck Surg.* **270**, 1521–1529 (2013).
91. Requena, T. *et al.* Allelic variants in TLR10 gene may influence bilateral affectation and clinical course of Meniere's disease. *Immunogenetics* **65**, 345–355 (2013).
 92. Teranishi, M. *et al.* Polymorphisms in genes involved in the free-radical process in patients with sudden sensorineural hearing loss and Ménière's disease. *Free Radic. Res.* **47**, 498–506 (2013).
 93. Cabrera, S. *et al.* Intronic variants in the NFKB1 gene may influence hearing forecast in patients with unilateral sensorineural hearing loss in Meniere's disease. *PloS One* **9**, e112171 (2014).
 94. Yazdani, N. *et al.* Sex-specific association of RANTES gene -403 variant in Meniere's disease. *Eur. Arch. Oto-Rhino-Laryngol. Off. J. Eur. Fed. Oto-Rhino-Laryngol. Soc. EUFOS Affil. Ger. Soc. Oto-Rhino-Laryngol. - Head Neck Surg.* **272**, 2221–2225 (2015).
 95. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
 96. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
 97. Fu, W. *et al.* Analysis of 6,515 exomes reveals a recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
 98. Smedley, D. *et al.* Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat. Protoc.* **10**, 2004–2015 (2015).
 99. Kennedy, B. *et al.* Using VAAST to Identify Disease-Associated Variants in Next-Generation Sequencing Data. *Curr. Protoc. Hum. Genet. Editor. Board Jonathan Haines AI* **81**, 6.14.1-6.14.25 (2014).
 100. Singleton, M. V. *et al.* Phevor Combines Multiple Biomedical Ontologies for Accurate Identification of Disease-Causing Alleles in Single Individuals and Small Nuclear Families. *Am. J. Hum. Genet.* **94**, 599–610 (2014).

101. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
102. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).
103. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
104. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
105. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
106. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
107. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
108. Davydov, E. V. *et al.* Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLOS Comput. Biol.* **6**, e1001025 (2010).
109. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
110. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
111. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
112. Shearer, A. E. *et al.* Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. *Am. J. Hum. Genet.* **95**, 445–453 (2014).
113. Köhler, S. *et al.* The Human Phenotype Ontology in 2017. *Nucleic Acids Res.* **45**, D865–D876 (2017).

114. Gazal, S. *et al.* Can whole-exome sequencing data be used for linkage analysis? *Eur. J. Hum. Genet.* **24**, 581–586 (2016).
115. Collet, A. *et al.* Pros and cons of HaloPlex enrichment in cancer predisposition genetic diagnosis. *Genet. 2015 Vol 2 Pages 263-280* (2015). doi:10.3934/genet.2015.4.263
116. Meyer, M. J. *et al.* Interactome INSIDER: a structural interactome browser for genomic studies. *Nat. Methods* **15**, 107–114 (2018).
117. Bao, R. *et al.* Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform.* **13**, 67–82 (2014).
118. Precone, V. *et al.* Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives. *BioMed Research International* (2015). doi:10.1155/2015/161648
119. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
120. Smedley, D. *et al.* Walking the interactome for candidate prioritization in exome sequencing studies of Mendelian diseases. *Bioinforma. Oxf. Engl.* **30**, 3215–3222 (2014).
121. Javed, A., Agrawal, S. & Ng, P. C. Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nat. Methods* **11**, 935–937 (2014).
122. Lee, J. M. *et al.* Genetic aspects and clinical characteristics of familial Meniere’s disease in a South Korean population. *The Laryngoscope* **125**, 2175–2180 (2015).
123. Morrison, A. W., Bailey, M. E. S. & Morrison, G. a. J. Familial Ménière’s disease: clinical and genetic aspects. *J. Laryngol. Otol.* **123**, 29–37 (2009).
124. Frykholm, C. *et al.* Familial Ménière’s disease in five generations. *Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol.* **27**, 681–686 (2006).
125. Kato, A. *et al.* Dermatotopontin interacts with fibronectin, promotes fibronectin fibril formation, and enhances cell adhesion. *J. Biol. Chem.* **286**, 14861–14869 (2011).
126. Kuhn, T. B. *et al.* Regulating actin dynamics in neuronal growth cones by ADF/cofilin and rho family GTPases. *J. Neurobiol.* **44**, 126–144 (2000).

127. Aghajanian, H. *et al.* Semaphorin 3d and semaphorin 3e direct endothelial motility through distinct molecular signaling pathways. *J. Biol. Chem.* **289**, 17971–17979 (2014).
128. Webber, A. & Raz, Y. Axon guidance cues in auditory development. *Anat. Rec. A. Discov. Mol. Cell. Evol. Biol.* **288**, 390–396 (2006).
129. Love, C. A. *et al.* The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D. *Nat. Struct. Mol. Biol.* **10**, 843–848 (2003).
130. Vona, B., Nanda, I., Neuner, C., Müller, T. & Haaf, T. Confirmation of GRHL2 as the gene for the DFNA28 locus. *Am. J. Med. Genet. A.* **161A**, 2060–2065 (2013).
131. Okamoto, O. & Fujiwara, S. Dermatotin, a novel player in the biology of the extracellular matrix. *Connect. Tissue Res.* **47**, 177–189 (2006).
132. Okamoto, O., Fujiwara, S., Abe, M. & Sato, Y. Dermatotin interacts with transforming growth factor beta and enhances its biological activity. *Biochem. J.* **337 (Pt 3)**, 537–541 (1999).
133. Yoshimura, H. *et al.* Deafness Gene Expression Patterns in the Mouse Cochlea Found by Microarray Analysis. *PLOS ONE* **9**, e92547 (2014).
134. Castel, S. E. *et al.* Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet.* **50**, 1327–1334 (2018).
135. Marazita, M. L. *et al.* Genetic epidemiological studies of early-onset deafness in the U.S. school-age population. *Am. J. Med. Genet.* **46**, 486–491 (1993).
136. Shearer, A. E., Hildebrand, M. S. & Smith, R. J. Hereditary Hearing Loss and Deafness Overview. in *GeneReviews®* (eds. Adam, M. P. *et al.*) (University of Washington, Seattle, 1993).
137. Shearer, A. E. & Smith, R. J. H. Massively Parallel Sequencing for Genetic Diagnosis of Hearing Loss: The New Standard of Care. *Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.* **153**, 175–182 (2015).
138. Shearer, A. E. *et al.* Advancing Genetic Testing for Deafness with Genomic Technology. *J. Med. Genet.* **50**, (2013).

139. Sloan-Heggen, C. M. *et al.* Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum. Genet.* **135**, 441–450 (2016).
140. Brownstein, Z. *et al.* Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in Middle Eastern families. *Genome Biol.* **12**, R89 (2011).
141. Weber, M. L. *et al.* Role of estrogen related receptor beta (ESRRB) in DFN35B hearing impairment and dental decay. *BMC Med. Genet.* **15**, 81 (2014).
142. Collin, R. W. J. *et al.* Mutations of ESRRB Encoding Estrogen-Related Receptor Beta Cause Autosomal-Recessive Nonsyndromic Hearing Impairment DFNB35. *Am. J. Hum. Genet.* **82**, 125–138 (2008).
143. Mašindová, I. *et al.* MARVELD2 (DFNB49) Mutations in the Hearing Impaired Central European Roma Population - Prevalence, Clinical Impact and the Common Origin. *PLoS ONE* **10**, (2015).
144. Nayak, G. *et al.* Molecular genetics of MARVELD2 and clinical phenotype in Pakistani and Slovak families segregating DFNB49 hearing loss. *Hum. Genet.* **134**, 423–437 (2015).
145. Tekin, M. *et al.* Screening the SLC26A4 gene in probands with deafness and goiter (Pendred syndrome) ascertained from a large group of students of the schools for the deaf in Turkey. *Clin. Genet.* **64**, 371–374 (2003).
146. Tsukamoto, K. *et al.* Distribution and frequencies of PDS (SLC26A4) mutations in Pendred syndrome and nonsyndromic hearing loss associated with enlarged vestibular aqueduct: a unique spectrum of mutations in Japanese. *Eur. J. Hum. Genet. EJHG* **11**, 916–922 (2003).
147. Li, X. C. *et al.* A mutation in PDS causes non-syndromic recessive deafness. *Nat. Genet.* **18**, 215–217 (1998).
148. Usami, S. *et al.* Non-syndromic hearing loss associated with enlarged vestibular aqueduct is caused by PDS mutations. *Hum. Genet.* **104**, 188–192 (1999).

149. Yang, T. *et al.* Transcriptional control of SLC26A4 is involved in Pendred syndrome and nonsyndromic enlargement of vestibular aqueduct (DFNB4). *Am. J. Hum. Genet.* **80**, 1055–1063 (2007).
150. Yang, T. *et al.* 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 (2009).
151. Millán, J. M. *et al.* An Update on the Genetics of Usher Syndrome. *J. Ophthalmol.* **2011**, (2011).
152. Weil, D. *et al.* Usher syndrome type I G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with theUSH1C protein, harmonin. *Hum. Mol. Genet.* **12**, 463–471 (2003).
153. Miyasaka, Y. *et al.* Heterozygous mutation of Ush1g/Sans in mice causes early-onset progressive hearing loss, which is recovered by reconstituting the strain-specific mutation in Cdh23. *Hum. Mol. Genet.* **25**, 2045–2059 (2016).
154. Zheng, Q. Y. *et al.* Digenic inheritance of deafness caused by 8J allele of myosin-VIIA and mutations in other Usher I genes. *Hum. Mol. Genet.* **21**, 2588–2598 (2012).
155. Aparisi, M. J. *et al.* Targeted next generation sequencing for molecular diagnosis of Usher syndrome. *Orphanet J. Rare Dis.* **9**, 168 (2014).
156. Bennett, B. C. *et al.* An electrostatic mechanism for Ca²⁺-mediated regulation of gap junction channels. *Nat. Commun.* **7**, 8770 (2016).
157. Bicego, M. *et al.* Pathogenetic role of the deafness-related M34T mutation of Cx26. *Hum. Mol. Genet.* **15**, 2569–2587 (2006).
158. Jara, O. *et al.* Critical role of the first transmembrane domain of Cx26 in regulating oligomerization and function. *Mol. Biol. Cell* **23**, 3299–3311 (2012).
159. Becker, K. G. The common variants/multiple disease hypothesis of common complex genetic disorders. *Med. Hypotheses* **62**, 309–317 (2004).
160. Mitchell, K. J. What is complex about complex disorders? *Genome Biol.* **13**, 237 (2012).

161. Aller, E. *et al.* Screening of the USH1G gene among Spanish patients with Usher syndrome. Lack of mutations and evidence of a minor role in the pathogenesis of the syndrome. *Ophthalmic Genet.* **28**, 151–155 (2007).

Supplementary data

Table S1 Two hundred randomly selected SNV located in genes causing autosomal dominant sensorineural hearing loss.

CHR.	POSITION	Rs	REF	ALT	FUNTION	GEN	MAF	HMDB	PHENOTYPE	MIM NUMBER	DISEASE
1	35250451	rs373725070	G	A	missense	GJB3	8,245E-06	yes			
1	35250457	rs1805063	C	T	missense	GJB3	0,0242	yes			
1	35250652	rs140829910	T	C	missense	GJB3	8,242E-06				
1	35250673	rs375681439	C	T	missense	GJB3	0,00004946			612644	DFNA2B
1	35250862	rs376748531	G	A	missense	GJB3	0,0003				
1	35250892	rs80297119	T	G	missense	GJB3	0,0016	yes			
1	35250910	rs74315318	G	A	missense	GJB3	0,0005	yes			
1	35251030	rs373815705	C	A	missense	GJB3	0,0004	yes			
1	41296788	rs142453905	T	C	missense	KCNQ4	0,0005				
1	41296828	rs34287852	T	G	missense	KCNQ4	0,2028	yes		600101	DFNA2A
1	41300706	rs370248473	G	A	missense	KCNQ4	0				
3	191098660	rs114502673	A	G	missense	CCDC50	0,0065			607453	DFNA44
4	6279336	rs111773340	C	A	missense	WFS1	0,0002				
4	6290774	rs145639028	G	A	missense	WFS1	0,00004427	yes			
4	6290847	rs113651985	C	T	missense	WFS1	0,0005				
4	6292945	rs115346085	G	A	missense	WFS1	1	yes			
4	6293040	rs41264699	A	C	missense	WFS1	0,0041	yes			
4	6293659	rs141233896	C	G	missense	WFS1	0,0002				
4	6296872	rs142428158	G	A	missense	WFS1	0,0002	yes			
4	6302499	rs369795224	C	T	missense	WFS1	0,00002471	yes			
4	6302816	rs35031397	C	G	missense	WFS1	0,0036	yes			
4	6302843	rs150894674	G	A	missense	WFS1	0,0003	yes		600965	DFNA6/14/38
4	6302889	rs1801208	G	A	missense	WFS1	0,057	yes			
4	6302955	rs377726402	G	A	Nonsense	WFS1	0,00000825	yes			
4	6303011	rs141254874	G	C	missense	WFS1	8,249E-06				
4	6303033	rs28937892	C	T	missense	WFS1	0,0000495	yes			
4	6303080	rs377544135	C	G	missense	WFS1	0,00004951	yes			

4	6303179	rs150840308	G	A	missense	WFS1	0,00009071		
4	6303194	rs199946797	C	T	missense	WFS1	0,0005	yes	
4	6303248	rs1805069	G	A	missense	WFS1	0,0089	yes	
4	6303278	rs143084511	G	T	missense	WFS1	0	yes	
4	6303281	rs138968466	C	T	missense	WFS1	0,00007419		
4	6303306	rs148544389	C	T	missense	WFS1	0		
4	6303361	rs143064649	G	A	Nonsense	WFS1	0,00000824	yes	
4	6303422	rs140213376	A	C	missense	WFS1	8,241E-06		
4	6303516	rs138258392	C	T	missense	WFS1	0		
4	6303534	rs71530907	C	T	missense	WFS1	0,007	yes	
4	6303641	rs71524377	G	A	missense	WFS1	0,00004207	yes	
4	6303644	rs200099217	C	T	missense	WFS1	0,0001	yes	
4	6303660	rs143280847	A	G	missense	WFS1	0,00004235		
4	6303680	rs1805070	A	G	missense	WFS1	0,0074	yes	
4	6303776	rs201239579	G	T	Nonsense	WFS1	8,691E-06	yes	
4	6303869	rs71526461	T	C	missense	WFS1	0,0001		
4	6303887	rs376974936	G	A	missense	WFS1	0,00005197		
4	6303891	rs369107336	C	G	missense	WFS1	0,0000347		
4	6304188	rs147934586	C	T	missense	WFS1	0,00001705		
<hr/>									
4	88533540	rs36094464	A	T	missense	DSPP	0,0905	yes	
4	88533843	rs368559431	G	A	missense	DSPP	0,00001658		
4	88534138	rs200819405	C	A	missense	DSPP	0,00003313		
4	88534326	rs201942511	G	A	missense	DSPP	0,0004		
4	88535112	rs368812371	G	T	missense	DSPP	0,00002485		
4	88536188	rs201148490	A	G	missense	DSPP	0,0014		605594 DFNA39
4	88536269	rs371825362	G	A	missense	DSPP	0,0004		
4	88536362	rs111205174	G	A	missense	DSPP	0,505		
4	88536650	rs370270012	G	T	missense	DSPP	0,0001		
4	88537715	rs148827799	G	T	missense	DSPP	0,0003		

5	140908057	rs376328260	A	C	missense	DIAPH1	0,00001656		124900	DFNA1
5	145719411	rs139312280	C	A	missense	POU4F3	0,00005821			
5	145719480	rs372436251	C	T	missense	POU4F3	0,00004173		602459	DFNA15
5	145719481	rs367737951	C	T	missense	POU4F3	0,00001669			
5	145719516	rs368239745	T	G	missense	POU4F3	0,00002509			
6	33133557	rs377656039	G	C	missense	COL11A2	0			
6	33137619	rs142500487	G	A	missense	COL11A2	0,00001648			
6	33141825	rs121912949	G	A	missense	COL11A2	0,0001	yes		
6	33142318	rs376355040	G	A	missense	COL11A2	8,946E-06			
6	33144056	rs141023125	T	G	missense	COL11A2	0,0000113		601868	DFNA13
6	33146726	rs149697159	G	C	missense	COL11A2	0,0000934	yes		
6	33146747	rs145499142	G	A	missense	COL11A2	0,0011	yes		
6	33147579	rs144862714	G	A	missense	COL11A2	0,0001	yes		
6	33156764	rs138305560	G	C	missense	COL11A2	0,00004969			
6	76527343	rs371575926	G	A	missense	MYO6	0,00003295			
6	76550343	rs150820400	C	T	missense	MYO6	0,00004149			
6	76572432	rs369889326	C	T	Nonsense	MYO6	0,00000824		606346	DFNA22
6	76599811	rs370750657	A	G	missense	MYO6	0,00001666			
6	76621394	rs141925339	G	A	missense	MYO6	0,00001661			
6	76624662	rs367978681	A	G	missense	MYO6	0,00005818			
8	102643928	rs200016612	G	C	missense	GRHL2	0,0001		608641	DFNA28
9	71845108	rs142684074	C	T	missense	TJP2	0,00002471			
9	71855051	rs143346845	C	T	missense	TJP2	0,00001785		613558	DFNA51
9	71863070	rs28556975	T	C	missense	TJP2	0,0029			
9	75303654	rs140437301	G	A	missense	TMC1	0,0002			
9	75366775	rs199560971	G	A	missense	TMC1	0,00001648			
9	75387348	rs375919123	T	C	missense	TMC1	8,237E-06			
9	75403306	rs148443938	T	G	missense	TMC1	0,00001648		606705	DFNA36
9	75404123	rs367924428	G	A	missense	TMC1	0,00007425	yes		

9	75404174	rs151001642	C	T	Nonsense	TMC1	0,00001648	yes		
9	75406910	rs372710475	C	T	missense	TMC1	0,00008237	yes		
9	75435758	rs368084452	G	A	missense	TMC1	8,247E-06	yes		
9	117803271	rs2274750	C	T	missense	TNC	0,0516	yes		
9	117808785	rs2104772	T	A	missense	TNC	0,4385	yes		
9	117819465	rs200005353	G	A	missense	TNC	0,00006629			
9	117822050	rs148749117	C	T	missense	TNC	0,00006629			
9	117825276	rs373148389	G	T	missense	TNC	8,288E-06			
9	117827085	rs141417605	C	T	missense	TNC	8,237E-06		615629	DFNA56
9	117827169	rs369874534	C	T	missense	TNC	0,0000825			
9	117840353	rs142334930	G	A	missense	TNC	0,00008237			
9	117848284	rs371055558	C	T	missense	TNC	0,00002537			
9	117848760	rs141624690	C	T	missense	TNC	0,00002471			
9	117849138	rs141281085	C	T	missense	TNC	8,238E-06			
9	117849280	rs150493993	C	T	missense	TNC	0,00009069			
9	117853183	rs143586851	C	T	missense	TNC	0,0002			
11	76853783	rs1052030	T	C	missense	MYO7A	0,4348	yes		
11	76853790	rs371849195	G	C	missense	MYO7A	0,00005013			
11	76867062	rs370395532	C	G	missense	MYO7A	0,00001716	yes		
11	76867967	rs201539845	G	A	missense	MYO7A	0,00003318	yes		
11	76868016	rs370897466	A	C	missense	MYO7A	8,325E-06			
11	76868392	rs184866544	A	G	missense	MYO7A	0,0011	yes		
11	76870496	rs45629132	G	A	missense	MYO7A	0,0011	yes		
11	76871254	rs368716988	A	G	missense	MYO7A	0,0003			
11	76873225	rs200304238	A	G	missense	MYO7A	0,0002			
11	76873944	rs375350389	C	G	missense	MYO7A	8,283E-06			
11	76885871	rs111033201	C	T	Nonsense	MYO7A	0,00003682	yes		
11	76885947	rs200057810	C	T	missense	MYO7A	0,00005899		601317	DFNA11
11	76890889	rs368341987	G	A	missense	MYO7A	0,0039	yes		
11	76892489	rs373089701	C	T	missense	MYO7A	0,00003807			

11	76892561	rs375668125	G	A	missense	MYO7A	8,569E-06		
11	76892613	rs199575418	G	A	missense	MYO7A	0,0002		
11	76893620	rs375050157	T	A	missense	MYO7A	0,00002515	yes	
11	76901153	rs111033178	G	A	missense	MYO7A	0,0004	yes	
11	76903189	rs376291076	G	A	missense	MYO7A	0,0001	yes	
11	76910708	rs41298747	C	T	missense	MYO7A	0,005	yes	
11	76912636	rs2276288	A	T	missense	MYO7A	0,544	yes	
11	76914163	rs111033287	C	T	missense	MYO7A	0,0021	yes	
11	76915143	rs201008835	C	A	missense	MYO7A	0,0002		
11	76915183	rs376674270	G	A	missense	MYO7A	0,00007701		
11	76916599	rs368657015	T	C	missense	MYO7A	0,00003079	yes	
11	76924054	rs367647666	G	A	missense	MYO7A	0,00007811		
11	76925708	rs200359303	G	A	missense	MYO7A	0,0003		
<hr/>									
11	120976653	rs376541939	G	A	missense	TECTA	8,236E-06		
11	120979969	rs145898158	C	T	missense	TECTA	0,00003301	yes	
11	120998925	rs371892292	C	T	missense	TECTA	0,00001658		
11	120999013	rs374863954	A	T	missense	TECTA	0,00003398		
11	121000423	rs111759871	C	T	missense	TECTA	0,0002	yes	
11	121000636	rs146175803	A	G	missense	TECTA	0,0004	yes	
11	121000716	rs143998942	G	A	missense	TECTA	0,00007547		
11	121000878	rs374229006	G	C	missense	TECTA	0,00002517		601543 DFNA8/12
11	121008565	rs369690173	T	G	missense	TECTA	8,324E-06		
11	121008594	rs147890616	G	C	missense	TECTA	0,0001	yes	
11	121008681	rs373132598	G	A	missense	TECTA	0,00003336		
11	121016729	rs373655409	G	A	missense	TECTA	0,00004191		
11	121023709	rs375984509	G	C	missense	TECTA	0		
11	121028738	rs374996667	C	A	missense	TECTA	0		
11	121038716	rs367589125	T	A	missense	TECTA	0,00002473		
11	121038773	rs140236996	C	T	missense	TECTA	8,237E-06	yes	

12	57422595	rs370014993	T	C	missense	MYO1A	0,00005766		
12	57424918	rs113470661	G	A	missense	MYO1A	0,0047	yes	
12	57430769	rs138855953	C	T	missense	MYO1A	0,00002471		
12	57430791	rs373952237	G	A	missense	MYO1A	0,00002471		
12	57431366	rs148808080	C	T	missense	MYO1A	0,0005	yes	
12	57431698	rs144320005	C	T	Nonsense	MYO1A	0,00006627		
12	57431785	rs368223948	C	T	missense	MYO1A	0		608652 DFNA48
12	57432715	rs367561406	T	C	missense	MYO1A	0,00002475		
12	57435225	rs61753849	C	A	missense	MYO1A	0,00009884	yes	
12	57437119	rs55679042	C	T	missense	MYO1A	0,0055	yes	
12	57437952	rs137975387	G	C	missense	MYO1A	0,0004		
12	57440417	rs146269737	G	A	Nonsense	MYO1A	0,00001648		
12	57441459	rs121909305	G	A	Nonsense	MYO1A	0,0032	yes	
12	100806604	rs373873276	G	A	missense	SLC17A8	0,00001647		605583 DFNA25
12	100813782	rs372802080	T	G	missense	SLC17A8	0		
12	133197122	rs149982621	T	C	missense	P2RX2	0,00006615		
12	133197906	rs147592928	G	A	missense	P2RX2	0,00002474		608224 DFNA41
12	133198367	rs140087499	C	T	missense	P2RX2	0,00009463		
13	20763045	rs370868313	C	T	missense	GJB2	0		
13	20763051	rs111033194	T	G	missense	GJB2	0,00009471	yes	
13	20763104	rs111033294	T	C	missense	GJB2	0,00008391	yes	
13	20763222	rs111033360	C	T	missense	GJB2	0,00003304	yes	
13	20763246	rs373684994	C	T	missense	GJB2	0,0001	yes	
13	20763269	rs370044106	A	G	missense	GJB2	0,00000826	yes	
13	20763294	rs80338948	G	A	missense	GJB2	0,0002	yes	
13	20763341	rs111033196	C	T	missense	GJB2	0,0154	yes	601544 DFNA3A
13	20763366	rs150529554	C	T	missense	GJB2	0,00009913	yes	
13	20763395	rs374572413	C	T	missense	GJB2	0,00002477	yes	
13	20763452	rs80338945	A	G	missense	GJB2	0,0009	yes	

13	20763472	rs111033218	G	C	missense	GJB2	0,0018	yes		
13	20763483	rs199883710	G	A	Nonsense	GJB2	8,239E-06	yes		
13	20763534	rs370696868	C	T	missense	GJB2	0,00001648	yes		
13	20763602	rs111033296	G	T	missense	GJB2	8,244E-06	yes		
13	20763627	rs371024165	G	A	missense	GJB2	0,0000412	yes		
13	20763642	rs2274084	C	T	missense	GJB2	0,0454	yes		
13	20797001	rs146231737	C	T	missense	GJB6	0,00005767		612643	DFNA3B
14	31344166	rs200935305	G	A	missense	COCH	0,0002		601369	DFNA9
14	31355287	rs367884240	C	T	missense	COCH	0,0000412			
14	61113177	rs144481204	C	A	missense	SIX1	0,00006589	yes	605192	DFNA23
16	2546346	rs371245371	C	T	missense	TBC1D24	0,00002508			
16	2547101	rs370233833	G	A	missense	TBC1D24	8,516E-06		613577	DFNA65
16	2550904	rs372995761	A	G	missense	TBC1D24	8,397E-06			
19	50720992	rs138001307	G	A	missense	MYH14	0,0002			
19	50728854	rs371766484	C	T	missense	MYH14	0,00001999			
19	50747534	rs119103280	G	T	missense	MYH14	0,0029	yes	600652	DFNA4
19	50771512	rs113993956	G	A	missense	MYH14	0,0004	yes		
19	50792886	rs368076336	G	A	missense	MYH14	0			
19	50794165	rs375795690	C	T	missense	MYH14	0,00008387			
22	36678809	rs142565774	C	T	missense	MYH9	0,00007722			
22	36682852	rs375515914	C	T	missense	MYH9	0,00002473			
22	36682873	rs142094977	A	G	missense	MYH9	0,0013	yes		
22	36684873	rs373393111	C	T	missense	MYH9	0,00001648	yes	160775	DFNA17
22	36688178	rs76368635	G	A	missense	MYH9	0,001	yes		
22	36691696	rs200901330	A	G	missense	MYH9	0,0003	yes		
22	36692971	rs147911658	T	A	missense	MYH9	8,291E-06			
22	36710207	rs375899392	C	T	missense	MYH9	0,00001648			

Table S2 Two hundred randomly selected SNV located in genes causing Centro Nuclear Myopathy.

CHR.	POSITION	Rs	REF	ALT	FUNTION	GEN	MAF	HMDB	PHENOTYPE MIM NUMBER	DISEASE
2	127806106	rs375004668	G	A	missense	BIN1	0,0001			
2	127806143	rs147655157	G	A	missense	BIN1	0,0002			
2	127806161	rs121909275	T	A	nosense	BIN1	0,0001	yes		
2	127806176	rs368983991	C	T	missense	BIN1	0,0001			
2	127808046	rs138047593	T	C	missense	BIN1	0,003			
2	127808076	rs112318500	G	A	missense	BIN1	0,034			
2	127808410	rs148422103	G	A	missense	BIN1	0,0008			
2	127808434	rs371571307	C	T	missense	BIN1	0,0001			
2	127808458	rs368238742	A	G	missense	BIN1	0,0001			
2	127808470	rs144459969	C	T	missense	BIN1	0,0002			
2	127808746	rs372650268	G	C	missense	BIN1	0,0001		255200	CNM2
2	127808749	rs140410496	G	A	missense	BIN1	0,0001			
2	127809920	rs200124094	C	T	missense	BIN1	0,0002			
2	127811539	rs375322787	T	C	missense	BIN1	0,0001			
2	127811566	rs368616652	G	A	missense	BIN1	0,0001			
2	127811582	rs200887814	C	T	missense	BIN1	0,0004			
2	127815174	rs76037557	G	A	missense	BIN1	0,0002			
2	127816664	rs374565677	C	T	missense	BIN1	0,0003			
2	127818175	rs367585396	T	A	missense	BIN1	0,0001			
2	127818193	rs117721706	C	T	missense	BIN1	0,0046			
2	127818194	rs144391901	G	A	missense	BIN1	0,0002			
2	127818197	rs148473945	A	G	missense	BIN1	0,0005			
2	127819743	rs372072916	C	T	missense	BIN1	0,0001			
2	127821184	rs375697182	G	A	missense	BIN1	0,0001			
2	127821206	rs146573197	C	T	missense	BIN1	0,0004			

2	127821511	rs143820618	G	T	missense	BIN1	0,0004		
2	127826543	rs371755655	G	T	missense	BIN1	0,0001		
2	127826558	rs267606681	C	T	missense	BIN1	0,000001	yes	
2	127826568	rs121909274	C	T	missense	BIN1	0,000001	yes	
2	127834212	rs369549551	T	A	missense	BIN1	0,0001		
2	127834262	rs121909273	C	A	missense	BIN1	0,000001	yes	
2	127864463	rs142657993	C	G	missense	BIN1	0,0001		
<hr/>									
3	9695311	rs377332766	C	G	missense	MTMR14	0,0001		
3	9695332	rs375944156	G	T	nosense	MTMR14	0,0001		
3	9704013	rs368196455	G	A	missense	MTMR14	0,0001		
3	9704024	rs372268047	C	T	missense	MTMR14	0,0001		
3	9711119	rs373227805	G	C	missense	MTMR14	0,0001		
3	9711141	rs189614064	T	A	missense	MTMR14	0,001		
3	9712833	rs374591212	A	G	missense	MTMR14	0,0001		
3	9714412	rs200924533	G	A	missense	MTMR14	0,0001		
3	9714418	rs201904466	A	G	missense	MTMR14	0,0002		
3	9719029	rs142525507	T	A	missense	MTMR14	0,0008		
3	9719057	rs374251047	T	C	missense	MTMR14	0,0001		160150 CNM1
3	9719695	rs372538745	C	T	missense	MTMR14	0,0001		
3	9726277	rs375469777	G	A	missense	MTMR14	0,0001		
3	9726311	rs121434509	G	A	missense	MTMR14	0,000001	yes	
3	9726588	rs183134138	C	T	missense	MTMR14	0,0014		
3	9726918	rs115607360	G	A	missense	MTMR14	0,0006		
3	9729559	rs372498357	G	A	missense	MTMR14	0,0002		
3	9730400	rs369183361	G	A	missense	MTMR14	0,0001		
3	9730639	rs376068526	C	T	missense	MTMR14	0,0002		
3	9730643	rs377445755	G	A	missense	MTMR14	0,0001		
3	9730675	rs370380809	C	G	missense	MTMR14	0,0001		
3	9730678	rs374725262	G	A	missense	MTMR14	0,0001		
3	9730693	rs371569636	C	T	missense	MTMR14	0,0001		

3	9730709	rs375373181	G	A	missense	MTMR14	0,0001			
3	9730718	rs121434510	G	A	missense	MTMR14	0,000001	yes		
3	9730758	rs371363549	G	C	missense	MTMR14	0,0001			
3	9731692	rs201206576	G	A	missense	MTMR14	0,0002			
3	9739406	rs371144090	A	G	missense	MTMR14	0,0001			
3	9739439	rs201626220	A	G	missense	MTMR14	0,0006			
3	9739479	rs376164405	G	C	missense	MTMR14	0,0001			
3	9739498	rs370895091	A	G	missense	MTMR14	0,0001			
3	9739526	rs200360764	C	T	missense	MTMR14	0,0001			
3	9743502	rs368605936	C	T	missense	MTMR14	0,0001			
3	9743503	rs370811714	G	A	missense	MTMR14	0,0001			
3	9743528	rs374180282	C	G	missense	MTMR14	0,0001			
3	9743616	rs375966737	C	T	missense	MTMR14	0,0001			
3	9743632	rs202121982	G	A	missense	MTMR14	0,0005			
<hr/>										
12	81101577	rs147184101	G	A	missense	MYF6	0,0002			
12	81101682	rs190471225	G	A	missense	MYF6	0,0004			
12	81101767	rs138296448	C	A	missense	MYF6	0,0004	yes		
12	81101770	rs372392737	C	T	missense	MYF6	0,0001		614408	CNM3
12	81101786	rs377370090	A	T	missense	MYF6	0,0001			
12	81101832	rs28928909	G	T	missense	MYF6	0,0006	yes		
12	81101845	rs200372502	T	C	missense	MYF6	0,0001			
12	81101886	rs370270818	A	G	missense	MYF6	0,0002			
12	81101976	rs368477055	C	G	missense	MYF6	0,0001			
12	81102342	rs143677057	T	A	missense	MYF6	0,0004			
12	81102358	rs146824657	A	C	missense	MYF6	0,0002			
12	81102363	rs375170162	T	C	missense	MYF6	0,0001			
12	81102373	rs143786238	T	C	missense	MYF6	0,0001			
12	81102385	rs375228457	C	A	missense	MYF6	0,0001			
<hr/>										
19	10870442	rs144250390	G	A	missense	DNM2	0,0022			

19	10883157	rs148790687	C	T	missense	DNM2	0,0002			
19	10883206	rs369347296	A	G	missense	DNM2	0,0001			
19	10883235	rs375151459	G	A	missense	DNM2	0,0002			
19	10886432	rs370086632	G	A	missense	DNM2	0,0001			
19	10887847	rs145478270	G	A	missense	DNM2	0,0004			
19	10893725	rs145607989	C	G	missense	DNM2	0,0002			
19	10893758	rs138128705	C	T	missense	DNM2	0,0001			
19	10893786	rs202155679	C	T	missense	DNM2	0,0006			
19	10897265	rs140963588	C	T	missense	DNM2	0,0001			
19	10897334	rs148105340	A	G	missense	DNM2	0,0001			
19	10904505	rs121909092	G	A	missense	DNM2	0,0001	yes	160150	CNM1
19	10904508	rs121909090	C	T	missense	DNM2	0,01	yes		
19	10904509	rs121909089	G	A	missense	DNM2	0,03	yes		
19	10906825	rs140208362	G	A	missense	DNM2	0,0001			
19	10908100	rs199927590	A	G	missense	DNM2	0,0001			
19	10908108	rs148633841	A	G	missense	DNM2	0,0001			
19	10908190	rs371514802	A	G	missense	DNM2	0,0001			
19	10909177	rs372593558	C	T	missense	DNM2	0,0001			
19	10909184	rs140043676	G	A	missense	DNM2	0,0003			
19	10909199	rs375653221	G	A	missense	DNM2	0,0001			
19	10909204	rs143992936	G	A	missense	DNM2	0,0001			
19	10909219	rs121909091	C	T	missense	DNM2	0,000001	yes		
19	10916639	rs141132980	A	G	missense	DNM2	0,0001			
19	10922976	rs368752035	A	G	missense	DNM2	0,0001			
19	10922991	rs121909093	G	T	missense	DNM2	0,000001	yes		
19	10923027	rs144763522	T	A	missense	DNM2	0,0001			
19	10930668	rs121909088	A	G	missense	DNM2	0,11	yes		
19	10930693	rs121909094	T	A	missense	DNM2	0,14	yes		
19	10934538	rs121909095	C	T	missense	DNM2	0,02	yes		
19	10934538	rs121909096	C	G	missense	DNM2	0,000001	yes		

19	10939829	rs370459176	T	C	missense	DNM2	0,0002	
19	10939902	rs375350902	C	G	missense	DNM2	0,0001	
19	10939910	rs368325934	G	A	missense	DNM2	0,0001	
19	10939922	rs375820696	C	T	missense	DNM2	0,0002	
19	10940816	rs374864354	C	T	missense	DNM2	0,0002	
19	10940928	rs138527500	C	T	missense	DNM2	0,0001	
19	10941659	rs139930306	G	C	missense	DNM2	0,0001	
19	10941677	rs149825590	G	A	missense	DNM2	0,0003	
19	10941695	rs369312570	G	A	missense	DNM2	0,0001	
19	10943688	rs139213045	G	A	missense	DNM2	0,0005	
19	10943750	rs146430642	G	A	missense	DNM2	0,0004	
19	10943768	rs373157246	C	T	nosense	DNM2	0,0001	
19	10943807	rs376139740	C	T	missense	DNM2	0,0002	
19	10943828	rs370918190	C	T	missense	DNM2	0,0001	
19	10943855	rs151223408	C	T	missense	DNM2	0,0002	
19	10943856	rs373835440	G	A	missense	DNM2	0,0001	
19	10943882	rs372270914	A	G	missense	DNM2	0,0001	
19	38939430	rs140037232	C	T	missense	RYR1	0,0002	yes
19	38945887	rs147723844	A	G	missense	RYR1	0,001	yes
19	38946103	rs111888148	G	A	missense	RYR1	0,0002	yes
19	38946112	rs144336148	G	A	missense	RYR1	0,0005	yes
19	38948186	rs193922772	G	A	missense	RYR1	0,0001	yes
19	38948830	rs144845360	G	A	missense	RYR1	0,0002	
19	38951020	rs200069592	G	A	missense	RYR1	0,0002	
19	38951140	rs372652716	G	A	missense	RYR1	0,0001	
19	38951205	rs375669412	G	A	missense	RYR1	0,0006	
19	38954087	rs138020885	C	T	missense	RYR1	0,0002	
19	38954405	rs374492243	A	G	missense	RYR1	0,0001	
19	38954473	rs144935444	G	A	missense	RYR1	0,0001	
19	38955289	rs148623597	G	A	missense	RYR1	0,0012	yes

19	38955362	rs201827275	C	T	missense	RYR1	0,0001		
19	38956847	rs367860207	C	T	missense	RYR1	0,0001		
19	38956858	rs138209392	G	A	missense	RYR1	0,0001	180901	CNM
19	38956954	rs374477216	C	T	missense	RYR1	0,0001		
19	38956988	rs374776563	G	A	missense	RYR1	0,0002		
19	38958382	rs149096607	C	T	missense	RYR1	0,0002		
19	38959666	rs201174268	G	A	missense	RYR1	0,0002		
19	38964116	rs377185497	C	T	missense	RYR1	0,0001		
19	38964275	rs34694816	A	G	missense	RYR1	0,054		
19	38966001	rs141678782	C	G	missense	RYR1	0,0001		
19	38966014	rs187496208	C	T	missense	RYR1	0,0002		
19	38966056	rs150499158	G	A	missense	RYR1	0,0001		
19	38968395	rs370851779	G	A	missense	RYR1	0,0001		
19	38976235	rs368726019	T	C	missense	RYR1	0,0001		
19	38976331	rs146504767	G	A	missense	RYR1	0,0004		
19	38976478	rs193922781	C	T	missense	RYR1	0,0001	yes	
19	38976529	rs377476955	C	T	missense	RYR1	0,0001		
19	38976636	rs372958050	T	C	missense	RYR1	0,0001		
19	38976655	rs34934920	C	T	missense	RYR1	0,017		
19	38976783	rs147603571	G	A	missense	RYR1	0,0001		
19	38980791	rs145801146	C	T	missense	RYR1	0,0001		
19	38985195	rs143398211	G	A	missense	RYR1	0,0002	yes	
19	38986946	rs193922795	G	A	missense	RYR1	0,0001	yes	
19	38989817	rs34390345	A	G	missense	RYR1	0,0002	yes	
19	38989881	rs147213895	A	G	missense	RYR1	0,0008	yes	
19	38989882	rs202061237	C	T	missense	RYR1	0,0004	yes	
19	38990295	rs193922802	G	A	missense	RYR1	0,0001	yes	
19	38990311	rs144526634	G	A	missense	RYR1	0,0002	yes	
19	38990346	rs146306934	G	A	missense	RYR1	0,0002	yes	
19	38990594	rs193922808	G	T	missense	RYR1	0,0001	yes	

19	38991258	rs375148516	G	A	missense	RYR1	0,0001	
19	38993605	rs371447916	C	T	missense	RYR1	0,0001	
19	38995510	rs2915951	T	C	missense	RYR1	0,32	yes
19	38995998	rs35180584	C	G	missense	RYR1	0,009	yes
19	38996982	rs138647599	G	A	missense	RYR1	0,0001	
19	39002892	rs375292503	A	G	missense	RYR1	0,0001	
19	39003006	rs61739911	C	T	missense	RYR1	0,001	yes
19	39006812	rs377541724	A	G	missense	RYR1	0,0001	
19	39009877	rs118204421	C	T	missense	RYR1	0,0002	yes
19	39010075	rs148892609	C	T	missense	RYR1	0,0001	
19	39016132	rs143987857	G	A	missense	RYR1	0,0006	yes
19	39018342	rs148130880	G	A	missense	RYR1	0,0001	
19	39026638	rs140616359	G	A	missense	RYR1	0,0002	yes
19	39034191	rs147136339	A	G	missense	RYR1	0,002	yes
19	39038899	rs144685735	C	T	missense	RYR1	0,0002	yes
19	39055615	rs370527763	G	A	missense	RYR1	0,0001	
19	39057615	rs73933023	C	T	missense	RYR1	0,011	yes
19	39061260	rs118192130	G	A	missense	RYR1	0,0002	yes
19	39062672	rs143520367	C	T	missense	RYR1	0,0001	yes
19	39070706	rs143988412	A	G	missense	RYR1	0,0001	yes
19	39070708	rs200442804	C	T	missense	RYR1	0,0012	yes
19	39070725	rs148540135	C	T	missense	RYR1	0,0001	yes
19	39070731	rs193922875	G	A	missense	RYR1	0,0001	yes
19	39071022	rs193922879	G	A	missense	RYR1	0,0001	yes
19	39075653	rs118192153	C	T	missense	RYR1	0,0001	yes
19	39076780	rs146876145	C	T	missense	RYR1	0,0002	yes

Table S3 Logit regression model to predict pathogenic variants is based on models consisting of single or multiple prediction tools for the top 10, 20 and 50 ranked variants for each tool, respectively. ROC curves and areas under the curve (AUC) values were used to compare the accuracy of each strategy. All p-value were corrected by Bonferroni's method.

	Pedigree	Area under curve	Top 10 AUC (p-value)	Top 20 AUC (p-value)	Top 50 AUC (p-value)
AD-SNHL	1 - F	Combining 5 Systems	69%	71%	68%
		PAVAR	58% (6x10 ⁻⁰⁴)	59% (7x10 ⁻⁰⁵)	53% (4x10 ⁻⁰⁵)
		Exomiser v2	51% (5x10 ⁻⁰⁵)	58% (4x10 ⁻⁰⁵)	58% (2x10 ⁻⁰⁴)
		VAAST- Phevor	59% (6x10 ⁻⁰⁴)	59% (1x10 ⁻⁰³)	58% (5x10 ⁻⁰⁴)
		CADD	55% (8x10 ⁻⁰⁵)	57% (1x10 ⁻⁰⁵)	50% (1x10 ⁻⁰⁵)
		FATHMM	57% (2x10 ⁻⁰⁴)	51% (1x10 ⁻⁰⁶)	53% (9x10 ⁻⁰⁵)
	1 - TF	Combining 5 Systems	68%	69%	69%
		PAVAR	59% (2x10 ⁻⁰³)	60% (1x10 ⁻⁰³)	53% (2x10 ⁻⁰⁶)
		Exomiser v2	51% (1x10 ⁻⁰⁴)	58% (2x10 ⁻⁰⁴)	57% (5x10 ⁻⁰⁴)
		VAAST- Phevor	58% (1x10 ⁻⁰³)	59% (1x10 ⁻⁰³)	58% (6x10 ⁻⁰⁵)
		CADD	52% (1x10 ⁻⁰⁴)	53% (1x10 ⁻⁰⁵)	50% (1x10 ⁻⁰⁶)
		FATHMM	58% (1x10 ⁻⁰³)	51% (1x10 ⁻⁰⁵)	53% (6x10 ⁻⁰⁶)
	1 - T	Combining 5 Systems	66%	70%	65%
		PAVAR	55% (1x10 ⁻⁰³)	55% (1x10 ⁻⁰⁵)	53% (5x10 ⁻⁰⁴)
		Exomiser v2	57% (4x10 ⁻⁰³)	59% (3x10 ⁻⁰³)	54% (4x10 ⁻⁰⁴)
		VAAST- Phevor	58% (4x10 ⁻⁰³)	59% (1x10 ⁻⁰⁴)	59% (0.02)
		CADD	54% (7x10 ⁻⁰⁴)	54% (7x10 ⁻⁰⁶)	50% (2x10 ⁻⁰⁴)
		FATHMM	55% (1x10 ⁻⁰³)	51% (2x10 ⁻⁰⁶)	51% (1x10 ⁻⁰⁴)
3 - F	Combining 5 Systems	68%	70%	67%	
	PAVAR	54% (1x10 ⁻⁰³)	52% (1x10 ⁻⁰⁴)	50% (8x10 ⁻⁰⁶)	

	Exomiser v2	59% (1x10 ⁻⁰³)	54% (2x10 ⁻⁰⁴)	54% (2x10 ⁻⁰³)
	VAAST- Phevor	60% (1x10 ⁻⁰³)	59% (0.01)	59% (2x10 ⁻⁰⁴)
	CADD	52% (1x10 ⁻⁰⁴)	50% (7x10 ⁻⁰⁵)	50% (8x10 ⁻⁰⁶)
	FATHMM	52% (1x10 ⁻⁰⁴)	53% (3x10 ⁻⁰⁴)	53% (4x10 ⁻⁰⁵)
	Combining 5 Systems	67%	71%	66%
3 - TF	PAVAR	60% (0.02)	54% (4x10 ⁻⁰⁶)	52% (2x10 ⁻⁰⁴)
	Exomiser v2	54% (1x10 ⁻⁰³)	59% (2x10 ⁻⁰³)	54% (2x10 ⁻⁰³)
	VAAST- Phevor	58% (8x10 ⁻⁰³)	62% (5x10 ⁻⁰⁴)	60% (0.03)
	CADD	55% (3x10 ⁻⁰³)	53% (1x10 ⁻⁰⁶)	50% (1x10 ⁻⁰⁴)
	FATHMM	53% (1x10 ⁻⁰³)	51% (1x10 ⁻⁰⁶)	51% (1x10 ⁻⁰⁴)
	Combining 5 Systems	73%	66%	66%
3 - T	PAVAR	54% (7x10 ⁻⁰⁷)	52% (1x10 ⁻⁰⁴)	50% (1x10 ⁻⁰⁴)
	Exomiser v2	59% (4x10 ⁻⁰⁴)	54% (1x10 ⁻⁰³)	54% (1x10 ⁻⁰³)
	VAAST- Phevor	61% (2x10 ⁻⁰⁵)	60% (7x10 ⁻⁰³)	60% (7x10 ⁻⁰³)
	CADD	54% (4x10 ⁻⁰⁷)	50% (9x10 ⁻⁰⁵)	50% (9x10 ⁻⁰⁵)
	FATHMM	50% (2x10 ⁻⁰⁷)	50% (6x10 ⁻⁰⁵)	50% (6x10 ⁻⁰⁵)
	Combining 5 Systems	67%	66%	68%
5 - F	PAVAR	54% (5x10 ⁻⁰⁴)	53% (3x10 ⁻⁰³)	52% (8x10 ⁻⁰⁶)
	Exomiser v2	58% (25x10 ⁻⁰³)	54% (2x10 ⁻⁰³)	57% (1x10 ⁻⁰³)
	VAAST- Phevor	59% (3x10 ⁻⁰³)	58% (0.06)	58% (1x10 ⁻⁰⁴)
	CADD	54% (2x10 ⁻⁰⁴)	50% (1x10 ⁻⁰³)	50% (4x10 ⁻⁰⁶)
	FATHMM	52% (1x10 ⁻⁰⁴)	53% (7x10 ⁻⁰³)	53% (2x10 ⁻⁰⁵)
	Combining 5 Systems	72%	70%	69%
5 - TF	PAVAR	59% (1x10 ⁻⁰⁵)	60% (2x10 ⁻⁰⁴)	53% (1x10 ⁻⁰⁶)
	Exomiser v2	58% (5x10 ⁻⁰⁶)	58% (8x10 ⁻⁰⁵)	57% (2x10 ⁻⁰⁴)

	VAAST- Phevor	58% (5×10^{-06})	61% (9×10^{-04})	58% (8×10^{-05})
	CADD	54% (7×10^{-07})	57% (3×10^{-05})	50% (9×10^{-07})
	FATHMM	57% (2×10^{-06})	52% (4×10^{-06})	53% (4×10^{-06})
	Combining 5 Systems	71%	71%	71%
5 - T	PAVAR	53% (2×10^{-06})	51% (3×10^{-05})	50% (6×10^{-05})
	Exomiser v2	57% (1×10^{-04})	54% (6×10^{-04})	53% (1×10^{-01})
	VAAST- Phevor	61% (2×10^{-04})	59% (3×10^{-03})	59% (3×10^{-03})
	CADD	55% (2×10^{-06})	50% (3×10^{-05})	50% (3×10^{-05})
	FATHMM	53% (2×10^{-06})	53% (1×10^{-04})	53% (1×10^{-04})
		Combining 5 Systems	67%	68%
FCONTROL - F	PAVAR	57% (4×10^{-03})	58% (6×10^{-05})	53% (8×10^{-07})
	Exomiser v2	51% (3×10^{-04})	58% (1×10^{-03})	59% (4×10^{-04})
	VAAST- Phevor	59% (0.03)	59% (8×10^{-02})	59% (3×10^{-05})
	CADD	50% (4×10^{-04})	56% (2×10^{-04})	50% (3×10^{-07})
	FATHMM	53% (4×10^{-04})	50% (6×10^{-05})	52% (1×10^{-06})
		Combining 5 Systems	69%	65%
1 - F	PAVAR	57% (6×10^{-03})	51% (2×10^{-02})	50% (0.04)
	Exomiser v2	60% (0.02)	60% (0.32)	52% (0.01)
	VAAST- Phevor	62% (0.08)	57% (0.15)	52% (0.02)
	CADD	53% (1×10^{-03})	52% (0.02)	53% (0.89)
	FATHMM	58% (0.01)	51% (0.02)	52% (0.03)
		Combining 5 Systems	71%	59%
1 - TF	PAVAR	52% (1×10^{-04})	51% (0.08)	50% (0.50)
	Exomiser v2	52% (1×10^{-04})	51% (0.08)	51% (0.01)
	VAAST- Phevor	67% (0.39)	59% (0.80)	52% (0.02)
	CADD	53% (1×10^{-04})	52% (0.09)	53% (0.90)

	FATHMM	59% (1×10^{-03})	51% (0.08)	52% (0.04)
	Combining 5 Systems	73%	57%	57%
	PAVAR	54% (4×10^{-05})	51% (0.19)	51% (5×10^{-02})
1 - T	Exomiser v2	51% (9×10^{-06})	51% (0.20)	50% (1×10^{-02})
	VAAST- Phevor	67% (4×10^{-02})	55% (0.61)	52% (2×10^{-02})
	CADD	52% (8×10^{-05})	53% (0.60)	53% (0.91)
	FATHMM	60% (7×10^{-04})	52% (0.32)	52% (4×10^{-02})
	Combining 5 Systems	63%	71%	70%
	PAVAR	53% (3×10^{-03})	53% (5×10^{-05})	52% (5×10^{-05})
FCONTROL - T	Exomiser v2	50% (5×10^{-03})	61% (2×10^{-02})	61% (6×10^{-03})
	VAAST- Phevor	60% (0.45)	60% (1×10^{-02})	60% (4×10^{-03})
	CADD	50% (1×10^{-03})	54% (8×10^{-05})	52% (1×10^{-04})
	FATHMM	52% (2×10^{-03})	52% (6×10^{-05})	52% (4×10^{-05})

Table S4 Number of SNV obtained in 21 exome datasets according to its effect on protein sequence and position on the reference genome (GRCh37 hg19).

	MD		Controls	
	Mean ± SD	%	Mean ± SD	%
Total variants	44703 ± 7831	100	36098 ± 1 2263	100
Missense variants	8602 ± 1034	19.24	6726 ± 2131	18.64
Nonsense variants	116 ± 55	0.26	75 ± 15	0.21
Synonymous variants	9116 ± 917	20.39	7296 ± 2417	20.22
Intronic variants	24977 ± 5888	55.87	20470 ± 7413	56.72
UTR 3' or 5' variants	1890 ± 350	4.23	1559 ± 541	4.32
Novel variants	3319 ± 2302	7.42	2125 ± 1063	5.88

Table S5. Web Resources, the URLs for software presented are as follows:

Web Resources	URLs
ANNOVAR	http://www.openbioinformatics.org/annovar/
Mutation Taster	http://www.mutationtaster.org/
PhastCons	http://compgen.bscb.cornell.edu/phast/phastCons-HOWTO.html
PhyloP	http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt
POLY-PHEN2	http://genetics.bwh.harvard.edu/pph2/index.shtml
SIFT	http://provean.jcvi.org/genome_submit.php
GERP++	http://mendel.stanford.edu/SidowLab/downloads/gerp/
Exomiser v2	https://www.sanger.ac.uk/resources/databases/exomiser/
CADD	http://cadd.gs.washington.edu/
FATHMM	http://fathmm.biocompute.org.uk/
1000 Genomes	http://www.1000genomes.org/dbSNP
NHLBI Exome Sequencing Project Exome Variant Server	http://evs.gs.washington.edu/EVS/
SHIELD: Shared Harvard Inner-Ear Laboratory Database	https://shield.hms.harvard.edu/
EMBL-EBI	http://www.ebi.ac.uk/
Orphanet	http://www.orpha.net/consor4.01/www/cgi-bin/OC_Exp.php?lng=EN&Expert=169189
Whole exome sequencing protocols	http://www.genomics.agilent.com/en/home.jsp
FastQC software	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Hereditary Hearing Loss Homepage	http://hereditaryhearingloss.org/
Human Phenotype Ontology database	http://human-phenotype-ontology.github.io/

Table S6 Pathogenic variants scoring system (PAVAR). To calculate the score for each variant, one point was assigned for each tool which score exceed the predefined threshold.

	Score = 0	Score = 1
SIFT	> 0.05 : tolerated	< 0.05 : deleterious
Mutation Taster	Polymorphism (p_value)	Disease causing (p_value)
Grantham Matrix	(0-50) : conservative (51-100) : moderately conservative	(101-150) : moderately radical (≥151) : radical
PolyPhen-2	(0.956,0.453) : possibly damaging (0.453,0) : benign	(>0.957):probably damaging
PhyloP	< 0.700	> 0.700
PhastCons	< 0.700	> 0.700
GERP	< 3.0 little evolutionary conservation	> 3.0 strong evolutionary conservation

Table S7: HPO terms used to describe the AD-SNHLs.

HPO code	HPO terms	Number of AD-SNHLs including this HPO term
HP:0000407	Sensorineural hearing impairment	11
HP:0000365	Hearing impairment	5
HP:0000360	Tinnitus	5
HP:0003676	Progressive	4
HP:0005101	High-frequency hearing impairment	3
HP:0002321	Vertigo	2
HP:0000408	Progressive sensorineural hearing impairment	2
HP:0008619	Bilateral sensorineural hearing impairment	2
HP:0001730	Progressive hearing impairment	2
HP:0008573	Low-frequency sensorineural hearing impairment	1
HP:0000703	Dentinogenesis imperfecta	1
HP:0011463	Childhood onset	1
HP:0008542	Low-frequency hearing loss	1
HP:0008615	Adult onset sensorineural hearing impairment	1
HP:0009591	Abnormality of the vestibulocochlear nerve	1
HP:0005102	Cochlear degeneration	1
HP:0008596	Postlingual sensorineural hearing impairment	1
HP:0011462	Young adult onset	1
HP:0000405	Conductive hearing impairment	1
HP:0004467	Preauricular pit	1
HP:0003621	Juvenile onset	1

Table S8 HPO terms used to describe the CNMs.

HPO code	HPO terms	Number of CNMs including this HPO term
HP:0001371	Flexion contracture	3
HP:0001270	Motor delay	2
HP:0000007	Autosomal recessive inheritance	2
HP:0000508	Ptosis	2
HP:0000218	High palate	2
HP:0003687	Centrally nucleated skeletal muscle fibers	2
HP:0010628	Facial palsy	2
HP:0001284	Areflexia	2
HP:0003677	Slow progression	2
HP:0003236	Elevated serum creatine phosphokinase	2
HP:0003691	Scapular winging	1
HP:0003391	Gowers sign	1
HP:0000602	Ophthalmoplegia	1
HP:0002747	Respiratory insufficiency due to muscle weakness	1
HP:0001319	Neonatal hypotonia	1
HP:0003458	EMG: myopathic abnormalities	1
HP:0003674	Onset	1
HP:0001256	Intellectual disability, mild	1
HP:0002460	Distal muscle weakness	1
HP:0001260	Dysarthria	1
HP:0002808	Kyphosis	1
HP:0001618	Dysphonia	1
HP:0002515	Waddling gait	1
HP:0003307	Hyperlordosis	1
HP:0003700	Generalized amyotrophy	1
HP:0003327	Axial muscle weakness	1
HP:0000276	Long face	1
HP:0001761	Pes cavus	1
HP:0001762	Talipes equinovarus	1
HP:0002650	Scoliosis	1
HP:0008872	Feeding difficulties in infancy	1
HP:0000544	External ophthalmoplegia	1
HP:0003701	Proximal muscle weakness	1
HP:0003712	Skeletal muscle hypertrophy	1
HP:0003388	Easy fatigability	1
HP:0005335	Sleepy facial expression	1
HP:0003394	Muscle cramps	1
HP:0001324	Muscle weakness	1
HP:0100305	Ring fibers	1
HP:0002063	Rigidity	1
HP:0003557	Increased variability in muscle fiber diameter	1
HP:0003798	Nemaline bodies	1
HP:0001374	Congenital hip dislocation	1
HP:0003593	Infantile onset	1
HP:0001252	Muscular hypotonia	1
HP:0003198	Myopathy	1
HP:0003680	Nonprogressive	1
HP:0001380	Ligamentous laxity	1
HP:0002905	Hyperphosphatemia	1

HP:0001945	Fever	1
HP:0003803	Type 1 muscle fiber predominance	1
HP:0002913	Myoglobinuria	1
HP:0001789	Hydrops fetalis	1

Table S9 VAAST files. P-value of quality, no significant differences were found between WES data and the background.

Pedigree	Cases (n)	Controls (n)	p-value
1	3	1	0.872
2	2	2	0.409
3	3	1	1
4	3	0	0.560
5	3	3	1

Figure S2. Ramachandran plot for Dermatopontin protein model. Percentage of residues on the favored region is 84,1%; percentage of residues on the allowed region is 4,9%; percentage of residues on the disallowed region is 11,0%. Residue substitution p.Arg182Cys is on favored region.

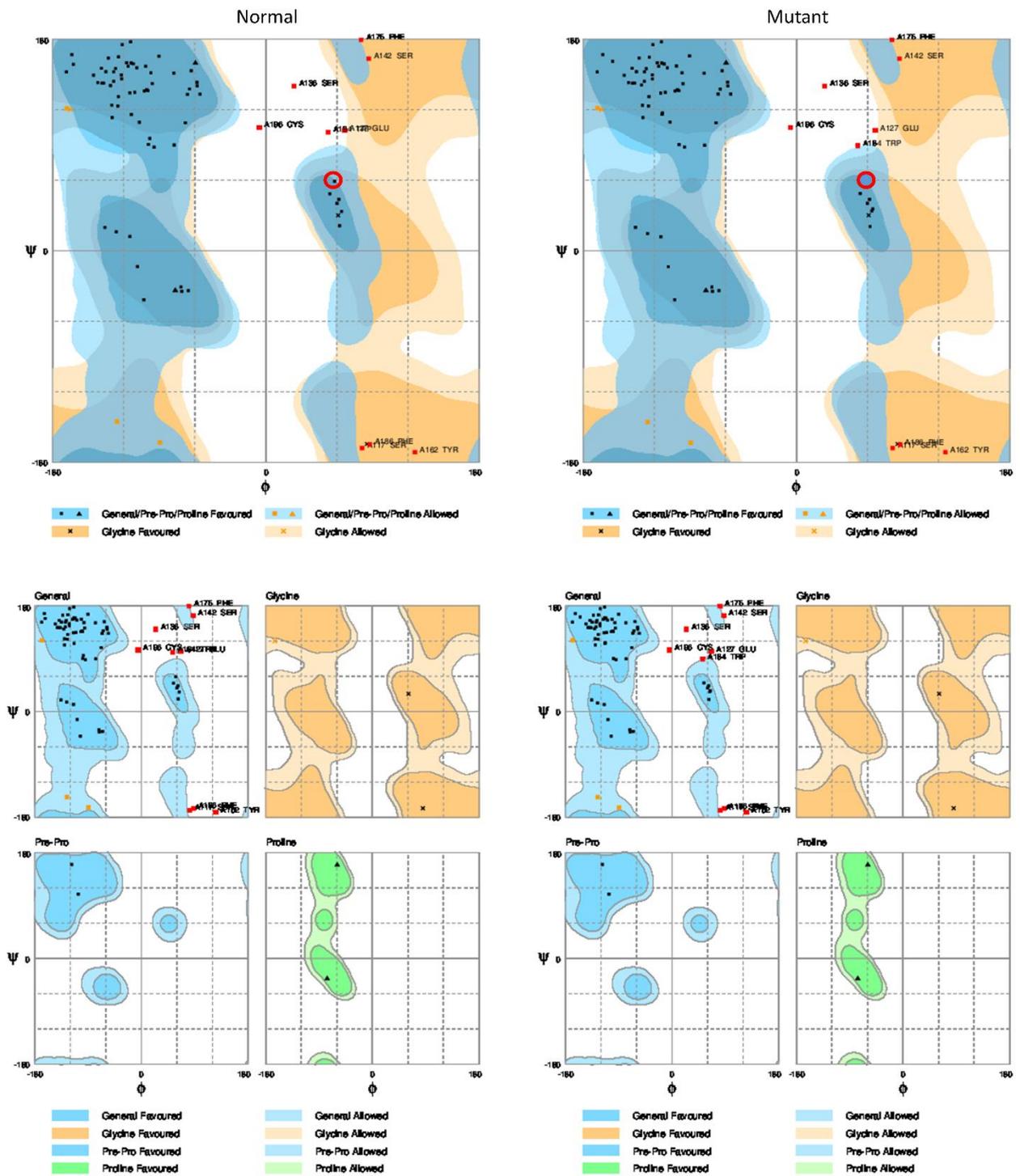


Figure S3. Ramachandran plot for SEMA3D protein model. Percentage of residues on the favored region 90.8 %; percentage of residues on the allowed region is 7.7%; percentage of residues on the disallowed region is 1,5%. Residue substitution p.Pro580Ser is on the favoured region.

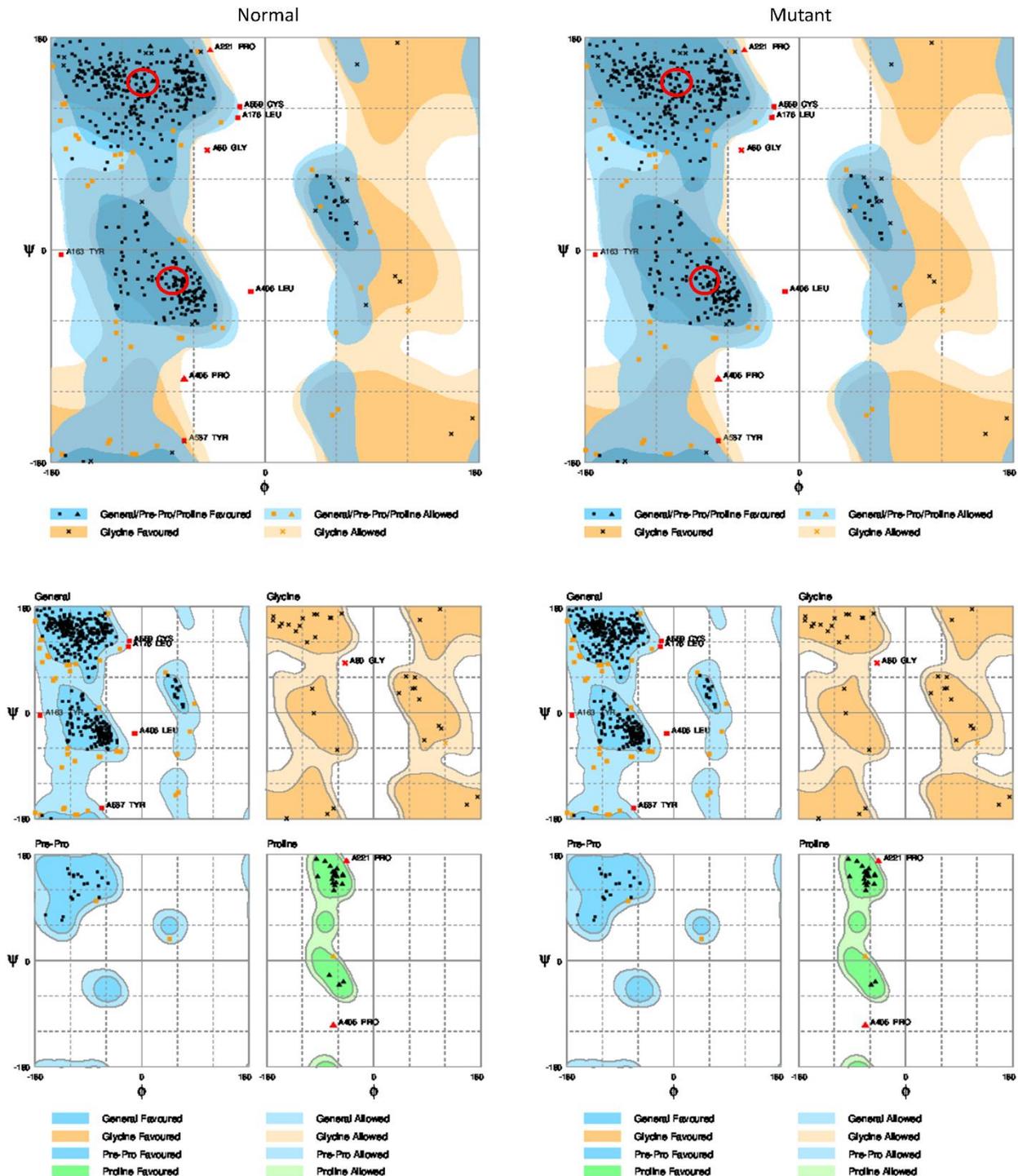


Table S10. Primers sequences used to validate the SNVs identified and the expression levels of SEMA3D and DPT genes.

NAME	SEQUENCE	PRODUCT SIZE (pb)
SEMA3D_Fw	GAGAGCTAGACGCCAAGATGTAA	249
SEMA3D_Rv	ATTCAATTAGGCACGTAGACAGG	249
DPT_Fw	AGCGATTCTTCCTGCCATGT	277
DPT_Rv	CAGGAAGTTGGCATTGCAGTTAC	277
SEMA3D.ex_Fw	TCATCTCAAGAAGGCAGTACCTC	213
SEMA3D.ex_Rv	TCTTTCATCTCTTGTTGGGGAGTA	213
DPT.ex_Fw	CTGGTGGGAGGAGATCAACAG	250
DPT.ex_Rv	GGTTGTTGCTCCTCGGATATAGT	250

Table S11. LOD scores obtained for the final list of candidate variants from WES genomic markers analysed in family 1.

F1	CHR	MARKER	POS	LOD
SEMA3D	7	M1	83764309	0.9601
	7	M2	84628989	0.9534
	7	M3	84642128	0.9447
	7	M4	84644346	0.9332
	7	M5	84644500	0.9195
TRAK1	3	M1	42251263	0.5453
	3	M2	42251329	0.6648
	3	M3	42264873	0.6577
	3	M4	42305131	0.6428
	3	M5	42560882	0.6281
PHF7	3	M1	52453893	0.1742
	3	M2	52454262	0.1794
	3	M3	52455673	0.1847
	3	M4	52456973	0.1812
	3	M5	52469941	0.1847
GABBR2	6	M1	89974066	0.1955
	6	M2	89974115	0.1895
	6	M3	89974214	0.1671
	6	M4	89975569	0.133

	6	M5	89981413	0.0904
CLCN	4	M1	170601449	-0.1258
	4	M2	170628474	-0.1289
	4	M3	170634382	-0.1321
	4	M4	170634390	-0.1288
	4	M5	170640993	-0.1251
GHRL2	8	M1	101937374	-0.1575
	8	M2	102504974	-0.1659
	8	M3	102555482	-0.1747
	8	M4	102678798	-0.1747
	8	M5	102678972	-0.1741

Table S12. LOD scores obtained for the final list of candidate variants from WES genomic markers analysed in family 2.

F2	CHR	MARKER	POS	LOD
DPT	1	M1	168549422	0.4573
	1	M2	168550535	0.4616
	1	M3	168665849	0.4663
	1	M4	168698173	0.4664
	1	M5	169094100	0.4619
GCC2	2	M1	108922036	0.2215
	2	M2	109069008	0.2236
	2	M3	109086855	0.2259
	2	M4	109087765	0.2238
	2	M5	109087885	0.2169
PRICKLE3	X	M1	49021537	0.0177
	X	M2	49032208	0.0184
	X	M3	49034780	0.0205
	X	M4	49069366	0.0251
	X	M5	49071964	0.0241

Table S13. Bibliography related to gene panel selection. DFNA, DFNB and others genes are separated in three tables.

Locus (OMIM)	Gene (OMIM)	Reference
DFNA2A	KCNQ4	Kubisch et al., 1999
DFNA3A	GJB2	Kelsell et al., 1997
DFNA4	MYH14	Donaudy et al., 2004
	CEACAM16	Zheng et al., 2011
DFNA6/14/38	WFS1	Bespalova et al., 2001 ; Young et al., 2001
DFNA9	COCH	Robertson et al., 1998
DFNA10	EYA4	Wayne et al., 2001
DFNA11	MYO7A	Liu et al., 1997
DFNA15	POU4F3	Vahava et al., 1998
DFNA20/26	ACTG1	Zhu et al., 2003 ; van Wijk et al., 2003
DFNA28	GRHL2	Peters et al., 2002
DFNA41	P2RX2	Yan et al., 2013
DFNA44	CCDC50	Modamio-Hoybjor et al., 2007
DFNA51	TJP2	Walsh et al., 2010

Locus (OMIM)	Gene (OMIM)	Reference (OMIM)
DFNB1A	GJB2	Kelsell et al., 1997
DFNB2	MYO7A	Liu et al., 1997 ; Weil et al., 1997
DFNB4	SLC26A4	Li et al., 1998
DFNB18	USH1C	Ouyang et al., 2002 ; Ahmed et al., 2002
DFNB24	RDX	Khan et al., 2007
DFNB28	TRIOBP	Shahin et al., 2006 ; Riazuddin et al., 2006
DFNB29	CLDN14	Wilcox et al., 2001
DFNB31	WHRN	Mburu et al., 2003
DFNB35	ESRRB	Collin et al., 2008
DFNB36	ESPN	Naz et al., 2004
DFNB49	MARVELD2	Riazuddin et al., 2006
DFNB70	PNPT1	von Ameln et al., 2012
DFNB74	MSRB3	Waryah et al., 2009 ; Ahmed et al., 2011
DFNB79	TPRN	Rehman et al., 2010 ; Li et al., 2010
	ESRP1	Rohacek et al., 2017

Gene (OMIM)	Reference (OMIM)
ADD1	Teggi et al., 2088
ARNT2	Own unpublished data
DPT	Martin-Sierra C et al., 2017
DTNA	Requena T et al., 2015
NR3B2	Chen J. et al., 2007
FAM107B	Scoles D et al., 2017
FAM136A	Requena T et al., 2015
KCNE1	Abbott GW 2016
KCNE3	Abbott GW 2016
KCNJ10	Smith RJH, 1998
KCNQ1	Splawski I et al., 1997

MICA	Gazquez I et al., 2012
MIF	Gazquez I et al., 2013
PRKCB	Martin-Sierra C et al., 2016
SEMA3D	Martín-Sierra et al, 2017
SLC12A2	Dixon MJ et al., 1999
THAP1	Own unpublished data
TLR10	Requena T et al., 2013
NFKB1	Cabrera S et al., 2014
USH1G	Miyasaka Y et al. 2016

Table S14. Novel and rare variants in patients with sporadic MD. List of rare variants validated through Sanger sequencing in MD patients. DbSNP accession numbers are detailed for each tested variant. Number of individuals with the variant in our MD cohort are described between parenthesis.

Gene	Location	dbSNP	MAF (gnomAD)	Case Freq (n)
<i>ESRRB</i>	14:76957891	rs201344770	0.0002096	0.0068181 (6)
	14:76966336	rs200237229	0.0004445	0.0034090 (3)
	14:76966347	rs201448899	0.0006266	0.0079545 (7)
<i>MARVELD2</i>	5:68715821	rs369265136	0.0000040	0.0034090 (3)
<i>SLC26A4</i>	7:107336408	rs200511789	0.0003572	0.0034090 (3)
<i>USH1G</i>	17:72915919	rs151242039	0.0006846	0.0034090 (3)
	17:72916543	rs111033465	0.0003479	0.0010752 (1)
<i>GJB2</i>	13:20763264	rs111033186	0.007274	0.0079545 (7)
	13:20763612	rs72474224	0.006587	0.0068181 (6)
	13:20763452	rs80338945	0.0008818	0.0034090 (3)
	13:20763642	rs2274084	0.04538	0.0118279 (11)
	13:20763633	rs374625633	0.0000057	0.0010752 (1)

Table S15. Mean coverage percentage per gene region in ExAC, gnomAD and our MD panel.
Only the 18 genes with significant excess of missense variants in the Spanish population from Table 4 are detailed.

GENE	MEAN COVERAGE PERCENTAGE		
	ExAC	gnomAD	MD panel
<i>GJB2</i>	70.49	82.3	79.56
<i>ESRRB</i>	49.55	72.47	69.38
<i>CLDN14</i>	52.71	74.71	76.95
<i>USH1G</i>	56.66	71.04	53.53
<i>SLC26A4</i>	64.16	76.55	71.41
<i>MYH14</i>	28.94	48.89	46.44
<i>SEMA3D</i>	60.53	56.44	44.48
<i>NFKB1</i>	61.76	57.32	80.97
<i>CCDC50</i>	54.61	68.64	61.9
<i>P2RX2</i>	52.84	57.25	74.36
<i>FAM136A</i>	60.35	56.13	82.4
<i>RDX</i>	55.01	64.22	84.33
<i>TPRN</i>	29.88	47.96	64.23
<i>ESPN</i>	27.79	48.8	65.01
<i>SLC12A2</i>	56.83	48.99	74.8
<i>PRKCB</i>	65.12	58.28	71.61
<i>ADD1</i>	69.31	65.04	71.74

Table S16. Selection of missense variants found in excess in the MD Spanish cohort.

Chr	Pos	R	Al	Func.refG	Gene.refG	ExonicFunc.ref	ExAC	ExAC NFE
		ef	t	ene	ene	Gene	MAF	MAF
chr 1	6488328	C	T	exonic	ESPN	nonsynonymous SNV	0.0005	0.0004
chr 1	6511753	C	T	exonic	ESPN	nonsynonymous SNV	0.0002	0.0003
chr 2	70524494	A	C	exonic	FAM136A	nonsynonymous SNV	8.24E-03	1.50E-02
chr 3	1.91E+08	A	G	exonic	CCDC50	nonsynonymous SNV	0.0065	0.0050
chr 4	2900221	A	G	exonic	ADD1	nonsynonymous SNV	8.24E-03	0
chr 4	1.04E+08	C	A	exonic	NFKB1	nonsynonymous SNV	9.99E-02	6.03E-02
chr 4	1.04E+08	A	G	exonic	NFKB1	nonsynonymous SNV	0.0097	0.0084
chr 4	1.04E+08	G	T	exonic	NFKB1	nonsynonymous SNV	0.0019	0.0028
chr 5	1.27E+08	A	T	exonic	SLC12A2	nonsynonymous SNV	1.66e-05	0

chr 5	1.46E+08	C	G	exonic	POU4F3	nonsynonymous SNV	0.0004	0.0005
chr 7	84636125	C	T	exonic	SEMA3D	nonsynonymous SNV	0.0002	0.0002
chr 7	84727240	A	G	exonic	SEMA3D	nonsynonymous SNV	0.0172	0.0235
chr 7	1.07E+08	A	G	exonic	SLC26A4	nonsynonymous SNV	4.13e-05	0
chr 7	1.07E+08	T	C	exonic	SLC26A4	nonsynonymous SNV	0.0005	0.0006
chr 7	1.07E+08	T	C	exonic	SLC26A4	nonsynonymous SNV	0.0083	0.0075
chr 7	1.07E+08	T	G	exonic	SLC26A4	nonsynonymous SNV	0.0132	0.0005
chr 7	1.07E+08	G	A	exonic	SLC26A4	nonsynonymous SNV	0.0044	0.0002
chr 7	1.07E+08	C	T	exonic	SLC26A4	nonsynonymous SNV	0.0020	0.0029
chr 9	1.4E+08	G	A	exonic	TPRN	nonsynonymous SNV	0.0022	0.0036
chr 9	1.4E+08	A	G	exonic	TPRN	nonsynonymous SNV	0.0070	0.0004
chr 9	1.4E+08	G	C	exonic	TPRN	nonsynonymous SNV	0.0037	0.0002
chr 11	1.1E+08	G	A	exonic	RDX	nonsynonymous SNV	0.0023	0.0038
chr 11	1.1E+08	G	C	exonic	RDX	nonsynonymous SNV	2.48E-02	4.52e-05
chr 12	1.33E+08	G	A	exonic	P2RX2	nonsynonymous SNV	0.0015	0.0024
chr 12	1.33E+08	A	C	exonic	P2RX2	nonsynonymous SNV	0.0028	0.0029
chr 12	1.33E+08	C	T	exonic	P2RX2	nonsynonymous SNV	0.0018	0.0003
chr 13	20763264	C	T	exonic	GJB2	nonsynonymous SNV	0.0106	0.0045
chr 13	20763452	A	G	exonic	GJB2	nonsynonymous SNV	0.0009	0.0015
chr 13	20763612	C	T	exonic	GJB2	nonsynonymous SNV	0.0066	0.0019
chr 13	20763642	C	T	exonic	GJB2	nonsynonymous SNV	0.0454	0.0023
chr 13	20763686	C	A	exonic	GJB2	nonsynonymous SNV	5.01E-02	0
chr 13	20763710	C	T	exonic	GJB2	nonsynonymous SNV	0.0004	0
chr 14	76957891	G	A	exonic	ESRRB	nonsynonymous SNV	0.0003	0.0001
chr 14	76966275	T	C	exonic	ESRRB	nonsynonymous SNV	0.0035	0.0050
chr 14	76966347	C	T	exonic	ESRRB	nonsynonymous SNV	0.0010	0.0014

chr 16	24046832	C	G	exonic	PRKCB	nonsynonymous SNV	0.0003	0.0005
chr 17	72915919	C	T	exonic	USH1G	nonsynonymous SNV	0.0006	0.0007
chr 17	72916365	C	T	exonic	USH1G	nonsynonymous SNV	0.0023	0.0018
chr 17	72916507	C	T	exonic	USH1G	nonsynonymous SNV	0.0118	0.0004
chr 17	72916543	T	C	exonic	USH1G	nonsynonymous SNV	0.0117	0.0003
chr 17	72916621	T	C	exonic	USH1G	nonsynonymous SNV	0.0001	0.0002
chr 19	50720992	G	A	exonic	MYH14	nonsynonymous SNV	0.0002	0.0003
chr 19	50766628	C	T	exonic	MYH14	nonsynonymous SNV	0.0005	0.0008
chr 19	50770231	G	A	exonic	MYH14	nonsynonymous SNV	0.0041	0.0022
chr 21	37833809	T	C	exonic	CLDN14	nonsynonymous SNV	0.0004	0.0007
chr 21	37833983	G	A	exonic	CLDN14	nonsynonymous SNV	0.0403	0.0256

Table S17. Average recombination rates (cM/Mb) in deCODE genetic maps for the chosen genes. Only *USH1G* genomic location recombination rate seems to be notably higher than the human average recombination rate (1.2).

RR (Avg cM/Mb)	
<i>ESRRB</i>	1.8
<i>GJB2</i>	0.6
<i>USH1G</i>	2.9
<i>CLDN14</i>	1.4
<i>SLC26A4</i>	1.5

Table S18. Linkage disequilibrium and pairwise correlations (R^2) between pairs of missense variants reported in 1KGenomes populations for the genes with higher excess of missense variants in the Spanish cohort against global 1KGenomes population.

ESRRB	rs61744548	rs201344770	rs553650212	rs188462546	rs201726554	rs201448899
rs61744548	1	0	0	0	0	0
rs201344770	0	1	0	0	0	0
rs553650212	0	0	1	0	0	0
rs188462546	0	0	0	1	0	0
rs201726554	0	0	0	0	1	0
rs201448899	0	0	0	0	0	1

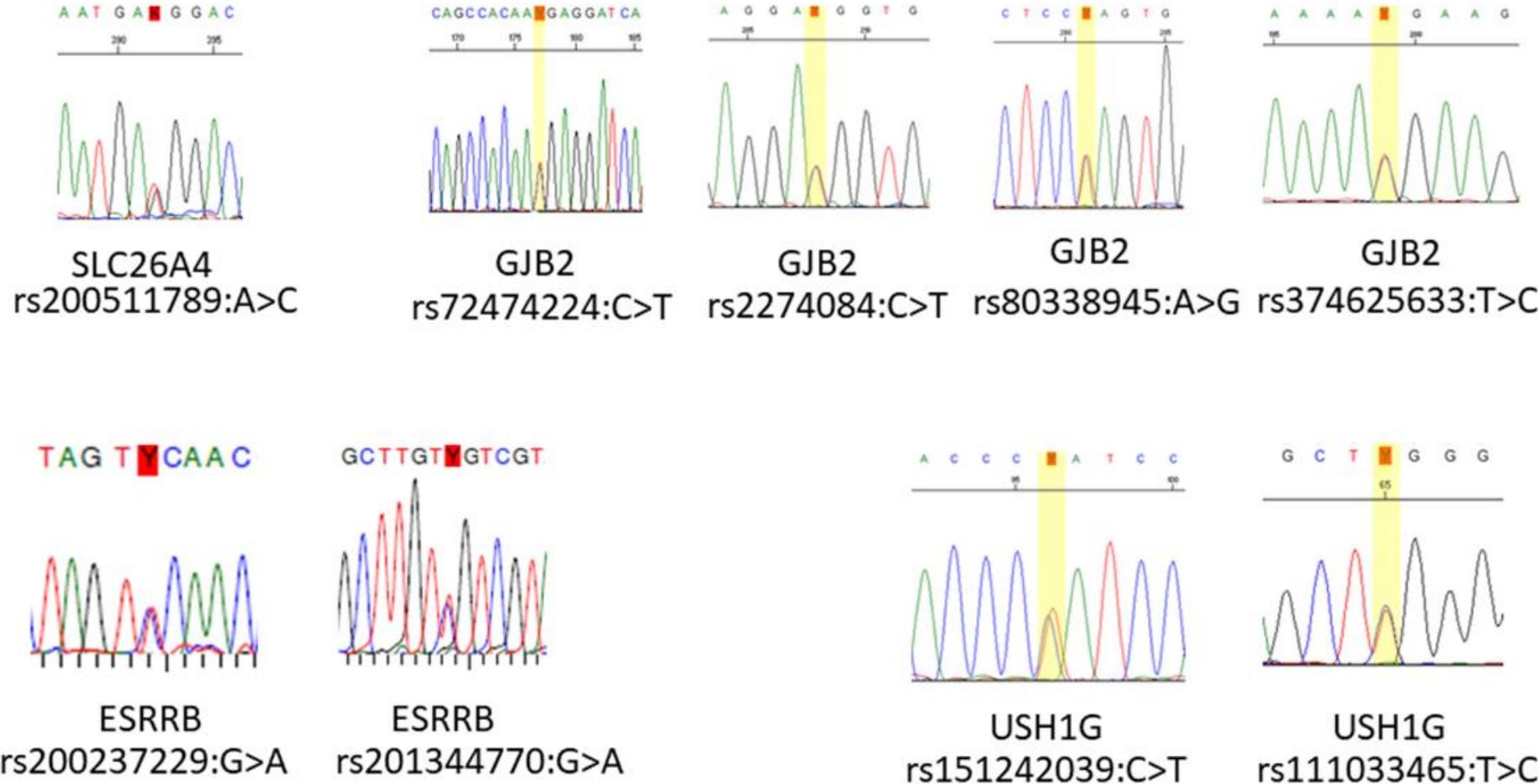
GJB2	rs111033186	rs111033218	rs72474224	rs2274084	rs111033222
rs111033186	1	0	0	0.001	0
rs111033218	0	1	0	0	0
rs72474224	0	0	1	0.001	0
rs2274084	0.001	0	0.001	1	0
rs111033222	0	0	0	0	1

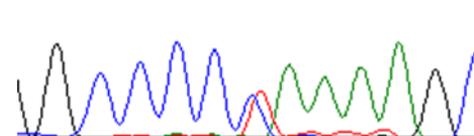
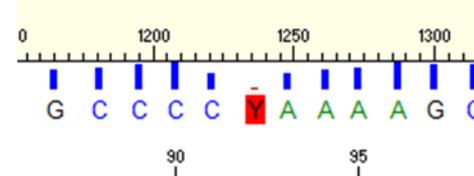
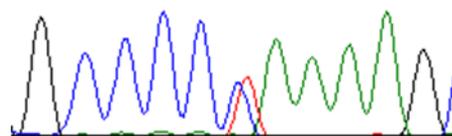
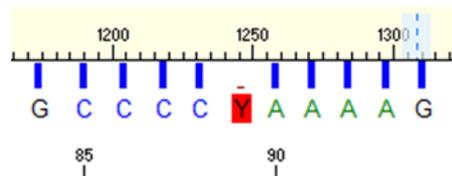
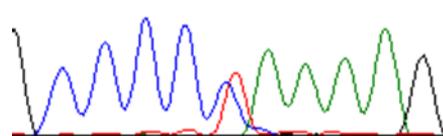
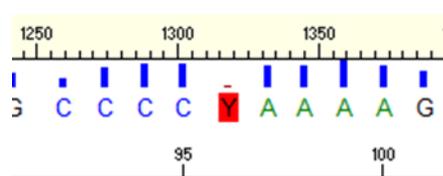
USH1G	rs151242039	rs149002004	rs201644674	rs141688757	rs111033466	rs111033465
rs151242039	1	0	0	0	0	0
rs149002004	0	1	0	0	0	0
rs201644674	0	0	1	0	0	0
rs141688757	0	0	0	1	0	0
rs111033466	0	0	0	0	1	1
rs111033465	0	0	0	0	1	1

CLDN14	rs61745291	rs139437157	rs113350364	rs148223897	rs146395322	rs113831133
rs61745291	1	0	0	0	0	0.001
rs139437157	0	1	0	0	0	0
rs113350364	0	0	1	0	0	0
rs148223897	0	0	0	1	0	0
rs146395322	0	0	0	0	1	0
rs113831133	0.001	0	0	0	0	1

SLC26A4	rs200431470	rs111033243	rs200511789	rs55638457	rs17154335	rs17154347	rs17154353	rs111033255
rs200431470	1	0	0	0.046	0	0	0	0
rs111033243	0	1	0	0	0	0	0	0
rs200511789	0	0	1	0	0	0	0	0
rs55638457	0.046	0	0	1	0	0	0	0
rs17154335	0	0	0	0	1	0.412	0.309	0
rs17154347	0	0	0	0	0.412	1	0.749	0
rs17154353	0	0	0	0	0.309	0.749	1	0
rs111033255	0	0	0	0	0	0	0	1

Figure S4. Chromatographs from validated candidate variants for SMD. qPCR curves for ESRRB variants in healthy controls are added.





MARVELD2	rs369265136	G/A
-----------------	-------------	-----

