UNIVERSIDAD DE GRANADA

PROGRAMA DE DOCTORADO EN BIOMEDICINA (B11.56.1)

Contribution of rare variants in coding regions to severe tinnitus

Contribución de variantes raras en regiones codificantes al acúfeno severo

International PhD Thesis

Tesis Doctoral con Mención Internacional

Alba Escalera Balsera

Directed by

José Antonio López Escámez

Granada, 2023

Editor: Universidad de Granada. Tesis Doctorales
Autor: Alba Escalera Balsera
ISBN: 978-84-1195-218-7
URI: https://hdl.handle.net/10481/89883

AGRADECIMIENTOS

Esta tesis, además de mi nombre, debe llevar el de todas estas personas que habéis hecho posible que llegue este momento. El camino que he recorrido estos años ha sido mucho más fácil gracias a vuestra ayuda y apoyo.

En primer lugar, quiero agradecer a Antonio por confiar en mí y haberme dado la oportunidad de hacer este doctorado. Eres el ejemplo de lo que es ilusionarse por la ciencia y no hay dudas del empeño que le pones a la investigación que llevas a cabo. Gracias por guiarme estos años y, sobre todo, gracias por haberme permitido formar parte de este increíble grupo.

Al jefe del *dry lab*, aunque siempre me digas NO o *nani*, gracias Álvaro por todos tus consejos y por todo lo que me has ayudado. Y a la jefa del *wet lab*, gracias Lidia por tu ayuda y por lo mucho que te preocupas por nosotros. Ojalá todos los limones fueran como tú.

Las dos cosas más bonitas que me ha ofrecido esta tesis han sido poder trabajar en lo que me gusta y encontrarme unos magníficos compañeros, compañeros que se han convertido en amigos. Marisa, siempre te digo que somos la dupla perfecta, entendiste mi código y me recordaste cómo se cogía una pipeta. Cuánto hemos pasado… pero qué bien nos lo hemos pasado. Obrigada menina. Paula, mi *compi* de tesis, juntas empezamos y juntas acabamos. Gracias por alegrarme los días, por contagiarme ese acento y por siempre estar cuando te he necesitado. Entre tres los problemas se resuelven rápido, las noches de ponernos al día se quedan cortas y todo lloro acaba en risas. Alberto Bernal, gracias por escucharme, por los viajes, por los *jajás* y por hacer que toda conversación mejore, aunque nos cierren las puertas del cielo. Alberto Parra, mi mano izquierda y el que más me aguanta, gracias por los cotilleos, la paciencia compartida y los consejitos *bioinfo.* Pablo el sevillano, por los exomas, las cocacolas y por acogerme desde el primer día. Pablo el gallego, porque marchaste, pero nos dejaste el marcho. A Jesús, por valorar mi vena artística, eres pura luz. Paqui, aunque con gustos opuestos, gracias por todo lo que me has ayudado. Mari por tus historias y tu “locura”. Elisheba, por tus ideales y porque es un placer tenerte en el *team* laboratorio. Por último, Olga, Estrella, Sana y Andreina, gracias por todos los momentos que compartimos.

Esta tesis también es gracias a mis queridos *bioinfos*, me acogisteis, me abristeis las puertas de la investigación, me ayudasteis y me seguís ayudando con mil dudas y, lo mejor, os convertisteis en mis amigos. A la cabeza que no para de Dani, al humor de Jordi, al corazón de Raúl y a la enciclopedia de Juan Antonio. A vosotros y a los que vinieron detrás que, sin duda, han mejorado el equipo. Iván, Marina y Samu, gracias.

A las personas que Genyo me ha dado la oportunidad de conocer, gracias por rescatarme cuando estaba perdida en cultivos, por la ayuda cuando hay que resolver problemas y por los consejos sobre qué haremos cuando acabemos la tesis. Pero lo que más, gracias por los viernes de tapas, sin duda es lo que más voy a echar de menos. Abel, Gonzalo, Iris, Jorge, Silvia, Araceli, Joselu, Joan, Agustín, Alberto, Juansan, Juanro, Félix, Marisol, Amador, Rita, Inés, Alba, Josemas y esas +40 personas que siempre se juntan.

A todo el personal de Genyo y de la Facultad de Medicina, por hacernos las cosas más fáciles.

A los clínicos que formáis parte del grupo en el Hospital Clínico San Cecilio, en el Hospital Universitario Virgen de las Nieves de Granada y en el resto de hospitales. En especial a Patricia, Marta, Juanma, Mari Carmen y Juan. A todos los pacientes que habéis hecho posible este estudio. Gracias por confiar en nosotros y por motivarme para avanzar tanto en el estudio de los acúfenos como de la Enfermedad de Ménière.

Thank you to the members of UNITI for allowing me to be part of this team, for teaching me so much about tinnitus and for the work you do. It has been a pleasure working with you.

To Nicky for welcoming me into your group so I could make my stay. And to her entire team, thanks to Alex, Maria, Elston, Hyung Chul and Nechama for teaching me about the non-coding regions and, also, the best pubs in Oxford.

A Granada por acogerme y a quienes me habéis acompañado en estos años. Natalia, Celia, Sergio, Inés, Yera, Paula; gracias. A los que me habéis enseñado que subiendo los pesos bajan los dramas, gracias.

Al primer regalo que me dio la biotecnología, esas amigas y amigos con los que parece que el tiempo no pase. No concibo una mejor forma de recargar las pilas que pasando unos días con vosotros. Estoy súper orgullosa de en quienes os estáis convirtiendo y más aún agradecida porque siempre sois un apoyo fundamental.

A Paula, por ser mi cable a tierra y por estar a mi lado desde que tengo uso de razón. A Rocío, por siempre inspirarme con tu visión del mundo. A vosotras, por confiar en mí como lo hacéis y, porque, aunque me vaya lejos, siempre os tengo cerca. Gracias infinitas a todas mis amigas, por ser casa y por siempre estar ahí.

Adrián, ¿con quién iba a picar código en dos teclados al mismo tiempo si no era contigo? Gracias por cuidarme, por apoyarme y por darme jamón cuando más lo necesito. Gracias por desestabilizar mi mundo cuadriculado y por compartir conmigo tantas y tantas risas. Nuestras tesis han ido de la mano y no se me ocurre mejor forma de continuar este camino.

A mi familia, por lo afortunada que soy de haber crecido con vosotros. Gracias por vuestro cariño y vuestra confianza en mí. Gracias Santi y gracias Florentina.

A mi madre y a mi padre, por acompañarme en todos los pasos de mi vida, por entenderme mejor que nadie y por quererme como lo hacéis. Vosotros me habéis enseñado las dos cosas más importantes: a cuestionarme el porqué de todo lo que me rodea, eso es la base de la ciencia; y a disfrutar y ser feliz, la base de la vida. Esta tesis también es vuestra.

Por último, gracias a mis compañeras y compañeros que luchan por la investigación y que, entre todos, hacen que vivamos con más conocimiento y rodeados de quienes queremos.

Gracias | Gràcies | Thank you

INDEX

TABLE INDEX

FIGURE INDEX

SUPPLEMENTARY TABLES INDEX

SUPPLEMENTARY FIGURE INDEX

ABSTRACT

Tinnitus is the subjective perception of tonal broad-band noise without an external acoustic source. The prevalence in the world population is between 10% and 30%; nevertheless, tinnitus is considered a disorder when it is associated with emotional distress, cognitive dysfunction and autonomic arousal. Meniere Disease (MD) is an inner ear disorder characterised by episodic vertigo associated with sensorineural hearing loss, tinnitus and aural fullness. Although vertigo attacks are considered the main symptom in the first years of the disease, persistent tinnitus is described as the most troublesome symptom by many MD patients. This thesis hypothesis was that tinnitus severity is related to an overload of rare variants in the coding regions of the individuals; therefore, the aim was to identify the central genes and biological processes associated with severe tinnitus.

Exome sequencing was performed for the 310 MD patients included in this study. According to the Tinnitus Handicap Inventory (THI) questionnaire, they were clustered into four subgroups. Individuals with a THI above the third quartile composed the severe tinnitus subgroup (I4) and those below the first quartile were the subgroup without disturbance caused by tinnitus (I1). Gene Burden Analyses (GBA) were carried out to target genes enriched in variants with high confidence of being loss-of-function (LoF) or variants with low confidence of being LoF and with a moderate impact in the protein, being missense most of them. Moreover, copy number variants (CNV) and structural variants (SV) were studied.

The results show: (1) Eleven genes have been associated with familial MD through a systematic review. (2) Twenty-eight genes were identified as candidates for severe tinnitus, clustering in four groups according to their expression and involved axogenesis, auditory brainstem response, synaptic transmission and hair cell stereocilia organisation as main biological functions. (3) Sixteen genes were identified as candidates for individuals without tinnitus, clustering also in four subgroups by the expression and being part of the protection of cochlear hair cells against damage and the structure of the tectorial membrane. (4) Tinnitus and hyperacusis were strongly associated in MD patients, which shared genetic causes. These results support that MD patients who experience bothersome tinnitus can be partially explained by differences in the rare genetic variants that they carry.

RESUMEN

El acúfeno es la percepción subjetiva de un ruido tonal en ausencia de una fuente acústica externa. La prevalencia en la población mundial oscila entre el 10% y el 30%; no obstante, el acúfeno se considera un trastorno cuando se asocia con angustia emocional, disfunción cognitiva y/o activación autonómica. La Enfermedad de Meniere (EM) es un trastorno del oído interno caracterizado por vértigo episódico, hipoacusia neurosensorial, acúfeno y/o sensación de plenitud ótica. Aunque los ataques de vértigo son el síntoma principal en los primeros años de la enfermedad, muchos pacientes con EM describen el acúfeno como el síntoma más molesto. La hipótesis de esta tesis fue que la gravedad del acúfeno está relacionada con una carga de variantes raras en las regiones codificantes de los individuos; por lo tanto, el objetivo fue identificar los principales genes y procesos biológicos asociados con el acúfeno severo.

Se realizó la secuenciación del exoma de los 310 pacientes con EM. En base al cuestionario *Tinnitus Handicap Inventory* (THI), se crearon cuatro subgrupos. Los individuos con un THI por encima del tercer cuartil conformaron el subgrupo de acúfeno severo (I4) y por debajo del primer cuartil el subgrupo sin molestias causadas por el acúfeno (I1). Se llevaron a cabo análisis de carga génica (*Gene Burden Analyses,* GBA) para identificar genes enriquecidos en variantes con alta probabilidad de ser de pérdida de función (*loss-of-function*, LoF) o variantes con baja probabilidad de ser LoF y con un impacto moderado en la proteína, siendo la mayoría de ellas de cambio de sentido (*missense*). Además, se estudiaron las variantes en el número de copias (*copy number variants,* CNVs) y las variantes estructurales (*structural variants,* SVs).

Los resultados muestran que: (1) Once genes fueron asociados con la EM familiar a través de una revisión sistemática. (2) Veintiocho genes se identificaron como candidatos para el acúfeno severo, participando en la axogénesis, la respuesta auditiva del tronco cerebral, la transmisión sináptica y la organización de los estereocilios de las células ciliadas. (3) Dieciséis genes se identificaron como candidatos para individuos sin acúfeno, formando parte de la protección contra el daño de las células ciliadas cocleares y de la estructura de la membrana tectorial. (4) El acúfeno y la hiperacusia estaban fuertemente asociados en pacientes con la EM, compartiendo causas genéticas. Estos resultados respaldan que la molestia causada por los acúfenos que algunos pacientes con EM experimentan puede explicarse parcialmente por diferencias en las variantes genéticas raras que presentan.

ACRONYMS AND ABBREVIATIONS

AAF	Anterior auditory fields
AAO-HNS	American Academy of Otolaryngology–Head and Neck Surgery
AB	Allele balance
AC	Allele count
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
AF	Allele frequency
AI	Primary area
AMP	Association for Molecular Pathology
AN	Allele number
AR	Autosomal recessive
AVCN	Anteroventral cochlear nucleus
BAM	Binary Alignment Map
bp	Base pair
BP	Biological Process
Ca^{2+}	Calcium ion
CADD	Combined Annotation Dependent Depletion
CC	Cellular Component
CI	Confidence interval
CNS	Segmented log2 ratios format
CNV	Copy number variant
CSVS	Collaborative Spanish Variant Server
DCN	Dorsal cochlear nucleus
DNA	Deoxyribonucleic acid
DNLL	Dorsal nuclei of the lateral lemniscus
DP	Depth

EF	Etiological fraction
ExAC	Exome Aggregation Consortium
FDR	False discovery rate
FLAGS	FrequentLy mutAted GeneS
FMD	Familial Meniere Disease
GATK	Genome Analysis Toolkit
GBA	Gene burden analysis
gDNA	Genomic DNA
gEAR	gene Expression Analysis Resource
GQ	Genotype quality
gnomAD	Genome Aggregation Database
GO	Gene Ontology
GTEx	Genotype-Tissue Expression
GÜF	Hypersensitivity to sound
GWAS	Genome-wide association study
HADS	Hospital Anxiety and Depression Scale
HPO	Human Phenotype Ontology
Hz	Hertz
IC	Inferior colliculus
IGV	Integrative Genomics Viewer
Indels	Insertions and deletions
K^+	Potassium ion
KEGG	Kyoto Encyclopedia of Genes and Genomes
LOEUF	Loss-of-function observed/expected upper bound fraction
LoF	Loss-of-function
LOFTEE	Loss-Of-Function Transcript Effect Estimator
LSO	Lateral superior olive
MD	Meniere Disease

MF	Molecular Function
MGB	Medial geniculate body
MGI	Mouse Genome Informatics
MNTB	Medial nucleus of the trapezoid body
MSO	Medial superior olivary nuclei
MD	Meniere Disease
Na^{2+}	Sodium ion
NFE	Non-Finnish European
OMIM	Online Mendelian Inheritance in Man
OR	Odds ratio
o/e	Observed/expected fraction
P0	Postnatal day 0
PAF	Posterior auditory fields
PHQ-9	Patient Health Questionnaire depression scale
pLI	Probability of being loss-of-function intolerant
PRISMA	Preferred Reporting Items for Systematic Reviews
PTA	Pure-tone audiometry
PVCN	Posteroventral cochlear nucleus
RNA	Ribonucleic acid
RNA-Seq	RNA-sequencing
SE	Systematic error
SEM	Standard error of the mean
SGN	Spiral ganglion neurons
SHIELD	Shared Harvard Inner-Ear Laboratory Database
SMD	Sporadic Meniere Disease
SNHL	Sensorineural hearing loss
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant

SPACNACS	Spanish Copy Number Alteration Collaborative Server
SV	Structural variant
RCT	Randomised Clinical Trial
RPKM	Reads per kilobase of transcript
rs	Reference single nucleotide polymorphism
THI	Tinnitus Handicap Inventory
TPM	Transcripts per million
T2T	Telomere-to-Telomere
UCSC	University of California, Santa Cruz
UNITI	Unification of Treatments and Interventions for Tinnitus Patients
VAS	Visual analogue scale
VCF	Variant Call Format
VEP	Variant Effect Predictor
VNLL	Ventral nuclei of the lateral lemniscus
VQSLOD	Variant quality score log-odds
VQSR	Variant Quality Score Recalibration
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WT	Wild type

1 INTRODUCTION

1.1 INNER EAR

The ear is responsible for maintaining the average balance and the sense of hearing. The ear is a complex series of interlinked structures divided into three different parts: the outer ear, the middle ear and the inner ear (Figure 1)[1].

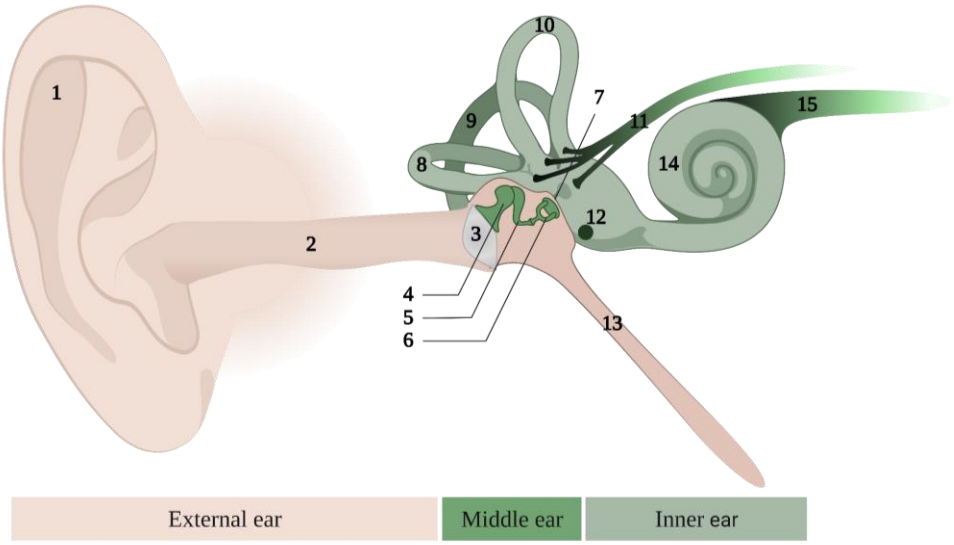

Figure 1 - Anatomy of the external, middle and inner ear.

(1) Auricle; (2) External auditory meatus; (3) Tympanic membrane; Ear ossicles: (4) Malleus, (5) Incus, (6) Stapes; (7) Oval window; Semicircular canals: (8) Horizontal, (9) Posterior, (10) Anterior; (11) Vestibular nerve; (12) Round window; (13) Eustachian tube; (14) Cochlea; (15) Cochlear nerve. Created with BioRender.com

The inner ear is a complex structure formed by an intricately shaped membranous tube suspended within a bony tube, the labyrinth. The inner ear comprises the cochlea and the vestibular system, the sensory organs responsible for hearing and sense of equilibrium, respectively[1,2]. The membranous labyrinth is filled by the endolymph and is surrounded by the bony labyrinth. The space between them contains the perilymph. Both fluids are essential for the correct excitation of hair cells responsible for sound and vestibular transmission[3,4]. In the inner ear there is a non-sensory structure named the endolymphatic sac, which is connected to the luminal space of the vestibular organ via the endolymphatic duct. The endolymphatic sac is involved in the immune reaction, regulating the inner ear homeostasis and the fluid volume[5].

1.1.1 Hair cells

Sensory hair cells are perhaps the most sensitive biological mechanosensors known. Within them, between 50 and 100 mechanically gated ion channel complexes are in charge of detecting hair bundle movements[6]. Although, the most relevant auditory and vestibular system cells are hair cells, they are meaningfully different in morphology, location, physiology and innervation[2].

1.1.2 Vestibular system

The vestibular system detects the angular and linear accelerations of the head and the body's gravitational forces; maintaining the balance and position[1,2]. It comprises three semicircular canals, utricle and saccule (Figure 2)[7].

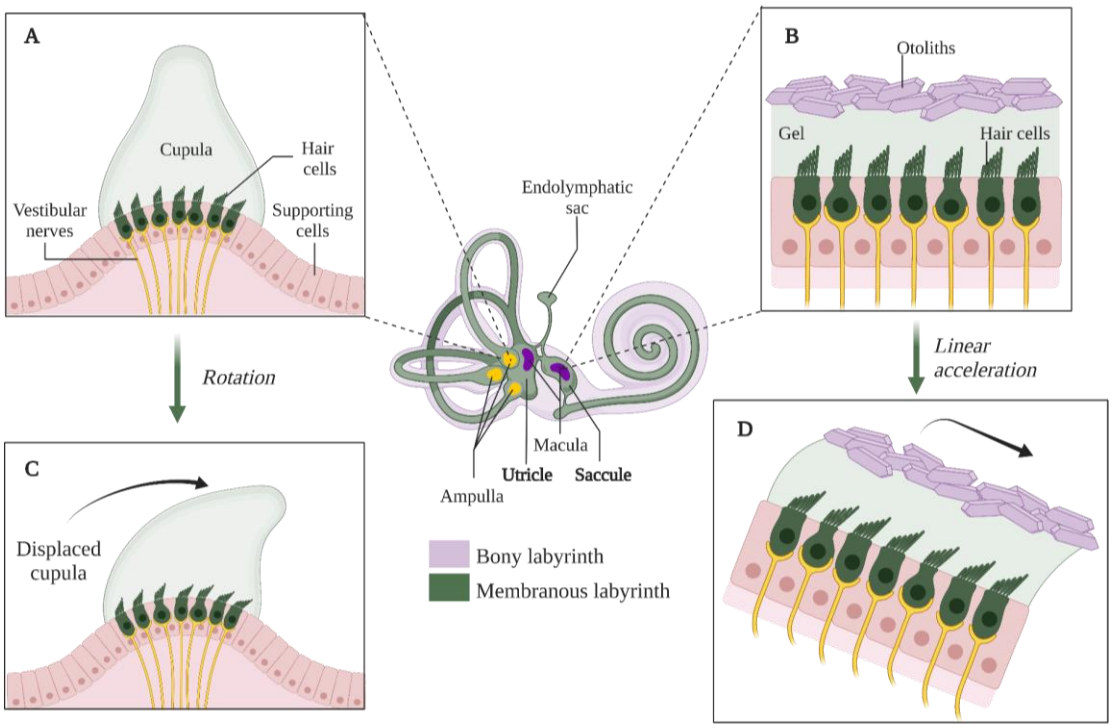

Figure 2 - Vestibular system.

Depolarisation (A, B) and hyperpolarisation (C, D) of the hair cells inside of the ampullas (A, C) and macula and saccule (B, D). Created with BioRender.com

The semicircular canals (anterior, posterior and horizontal), including their ampullas, detect angular accelerations. Each semicircular canal is connected to the utricle via an ampulla. Inside the ampulla is the cristae, which contains projections of hair cells surrounded by a gelatinous cupula. The head's rotation leads to the endolymph movement through the semicircular canals,

displacing the cupula and exciting the hair cells. The direction of the endolymph regulates the hair cells depolarisation or hyperpolarisation[3].

The utricle and saccule detect the linear acceleration; particularly, the macula - with sensory epithelia - that each of them contains is fundamental in this process. The utricle is involved in longitudinal acceleration and the saccule in vertical acceleration. Within the macula there are hair cells and supporting cells, besides above them the otolithic membrane with an extracellular protein matrix, termed gelatinous layer; on top of it there are found the otoconia, heavy calcium carbonate crystals. In the macula, the displacement of the hair cells leads to depolarisation or hyperpolarisation, depending on the direction of the movement, of the nerve fibres[3].

1.1.3 Cochlea

In the cochlea, the conversion from the vibrations generated by sound pressure into neurochemical impulses that travel to the brain along the cochlear-vestibular nerve[1,2].

The cochlea is a spiral-shaped fluid-filled organ divided into three segments: scala vestibuli, scala media and scala tympani (Figure 3B). In the centre of the cochlea is the scala media, separated from the scala vestibuli by the Reissner's membrane and from the scala tympani by the basilar membrane. The scala media contains endolymph, which is rich in potassium (K^+) and low in sodium (Na^{2+}) and calcium (Ca^{2+}), whereas the scala vestibuli and tympani are filled with perilymph, rich in Na^{2+} and low in K^+ and Ca^{2+}. This difference in concentration is indispensable for the cochlea's function because a positive endocochlear potential exists. Furthermore, the transport of K^+ allows the proper transduction caused by hair cells (Figure 3C)[1,3,7].

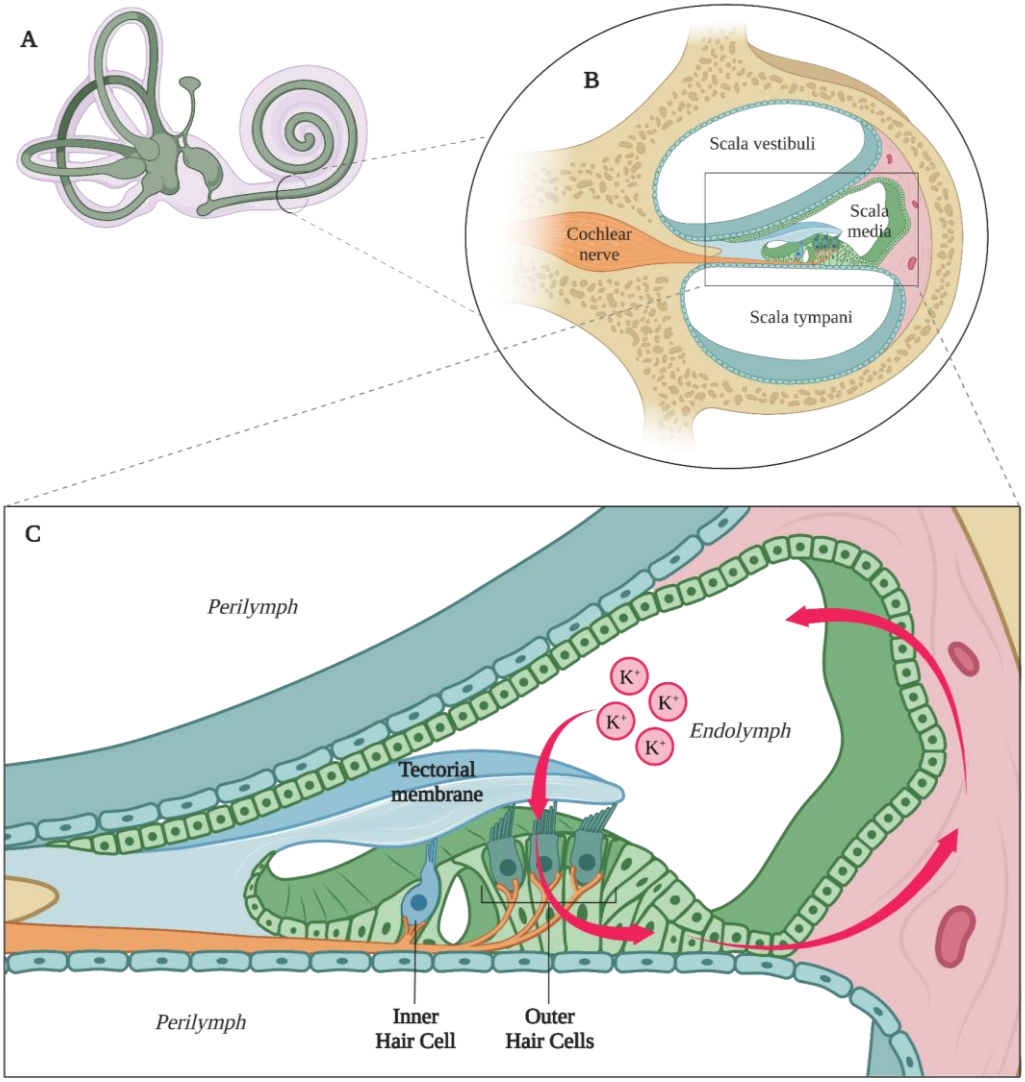

Figure 3 - Inner ear (A), a cross-section of the cochlea (B) and organ of Corti (C).

In the organ of Corti, the potassium ions (K^+) from the endolymph are transported inside the hair cells, which depolarises them. Then the K^+ are transported through the supporting cells to the stria vascularis, which canals return it to the endolymph. Created with BioRender.com.

1.1.3.1 Organ of Corti

The organ of Corti is located upon the basilar membrane, where the mechanical forces are converted into electrical impulses. It contains approximately 15,000 hair cells disposed into one row of inner hair cells and three rows of outer hair cells. The inner hair cells are the primary sensory cells of the auditory system, the most involved in the cochlear afferent innervation. In contrast, the outer hair cells are related to signal amplification. The hair cells are in contact with the tectorial membrane with projections named stereocilia and kinocilia (Figure 3C)[3,7].

Throughout the cochlea, the hair cells are distributed tonotopically, which allows them to discriminate between the different frequencies of the sounds. The tonotopy takes place because of the different physical properties of the basilar membrane: from base to apex, gradually, the stiffness decreases and the width increases. Therefore, high-frequency sounds entail maximal displacement of the basilar membrane at the base, while low-frequency sounds produce it at the apex[3,8].

1.1.3.2 Tectorial membrane

The tectorial membrane is a ribbon-like strip of extracellular matrix that spirals along the length of the organ of Corti. Its lower lateral surface is attached to the tips of the cochlea's hair bundles of mechanosensory hair cells. Due to the position above the bundles, it is suggested that the tectorial membrane is strongly related to the hair cells stimulation (Figure 3C). The tectorial membrane is mainly composed by water in a 97%; glycosaminoglycans; the II, V, IX and XI collagens; and the α-tectorin (Tecta), β-tectorin (Tectb), CEACAM16, otogelin and otogelin-like proteins[9–11].

1.2 AUDITORY SYSTEM

The auditory system is responsible for the perception and processing of understanding sounds. It comprises peripheral structures such as outer, middle and inner ear; and brain regions as the cochlear nuclei or auditory cortex. According to this, the auditory system is divided into peripheral and central auditory systems[12].

1.2.1 Peripheral auditory system

The peripheral auditory system is responsible for sound reaching the brain[12]. The sound waves are collected in the outer ear by the ear canal to the tympanic membrane, where air pressure oscillations produce vibrations. The vibrations are transmitted to the middle ear's ossicles: the malleus, incus and stapes. Via the oval window, the amplified energy of the vibrations arrives to the cochlea, into the inner ear (Figure 1)[7].

Within the cochlea, the basilar membrane vibration stimulates the hair cells causing a displacement of stereocilia and leading to the displacement of the adjacent kinocilia due to the tension created by tip links. This stiffness triggers the opening of the MET complex, located on the stereocilia surface, allowing the entrance of K^+, which depolarises the hair cell. Afterward, the entrance of Ca^{2+} leads to the release of neurotransmitters in the hair cell basolateral side to the synaptic space, which excites the afferent endings (Figure 4B). As a result, the mechanical energy is converted into electrical energy and the action potential generated travels along the cochlear nerve[3,7,12].

The auditory nerve is formed of peripheral axons from neurons in the spiral ganglion. Most (90%-95%) synapse at the base of the organ of Corti's hair cells. The outer hair cells synapse with the rest of the spiral ganglion neurons (SGN), which can contract the length of their cell body, altering the basilar membrane. In the presence of loud environments, the cortex produces changes to protect the health of the hair cells[12].

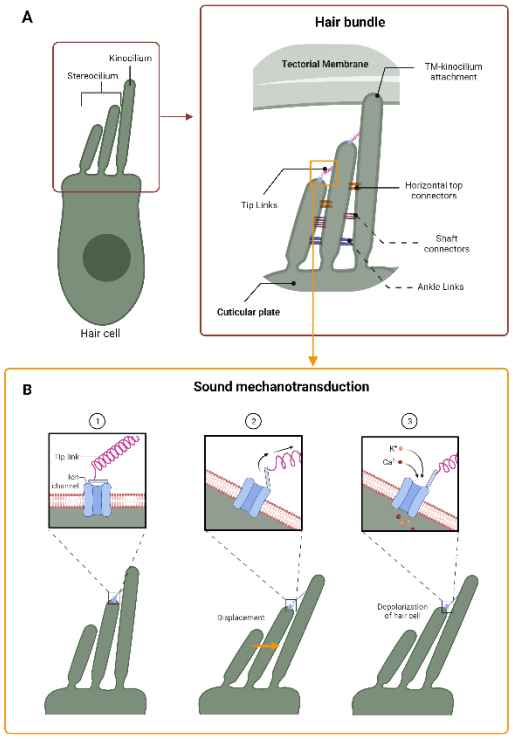

Figure 4 - Hair cell bundle to the tectorial membrane (A) and sound mechanotransduction due to the hair cell depolarisation (B).

1.2.2 Central auditory system

The signal pathways along the central auditory system diverge and converge, being parallel or overlapping. At every level of the auditory system there are crossing fibres; then, during the signal transport, the information can decussate or cross to the contralateral side of the brainstem. The ipsilateral and contralateral information assist in the localisation and interpretation of the sound quality[12,13].

The signals from the peripheral auditory system via the cochlear nerve (also named auditory nerve or VIII cranial nerve) arrive at the central auditory nuclei. The auditory information is transported via the auditory nerve across different nuclei, in the following order: cochlear nucleus, superior olivary nucleus, lateral lemniscus, inferior colliculus (IC) and medial geniculate body (MGB). In the cochlear nerve, the signal is divided into two pathways, when the auditory nerve reaches the brainstem each fibre branches forming the anterior and posterior branches of the nerve. The anteroventral cochlear nucleus (AVCN) is innervated by the anterior branch leading to the auditory stream of the brainstem, which localises the sound. Both dorsal cochlear nucleus (DCN) and posteroventral cochlear nucleus (PVCN) are innervated by the

posterior branch, bringing to the dorsal auditory stream of the brainstem, which is associated with the complex stimulus analysis[12–15].

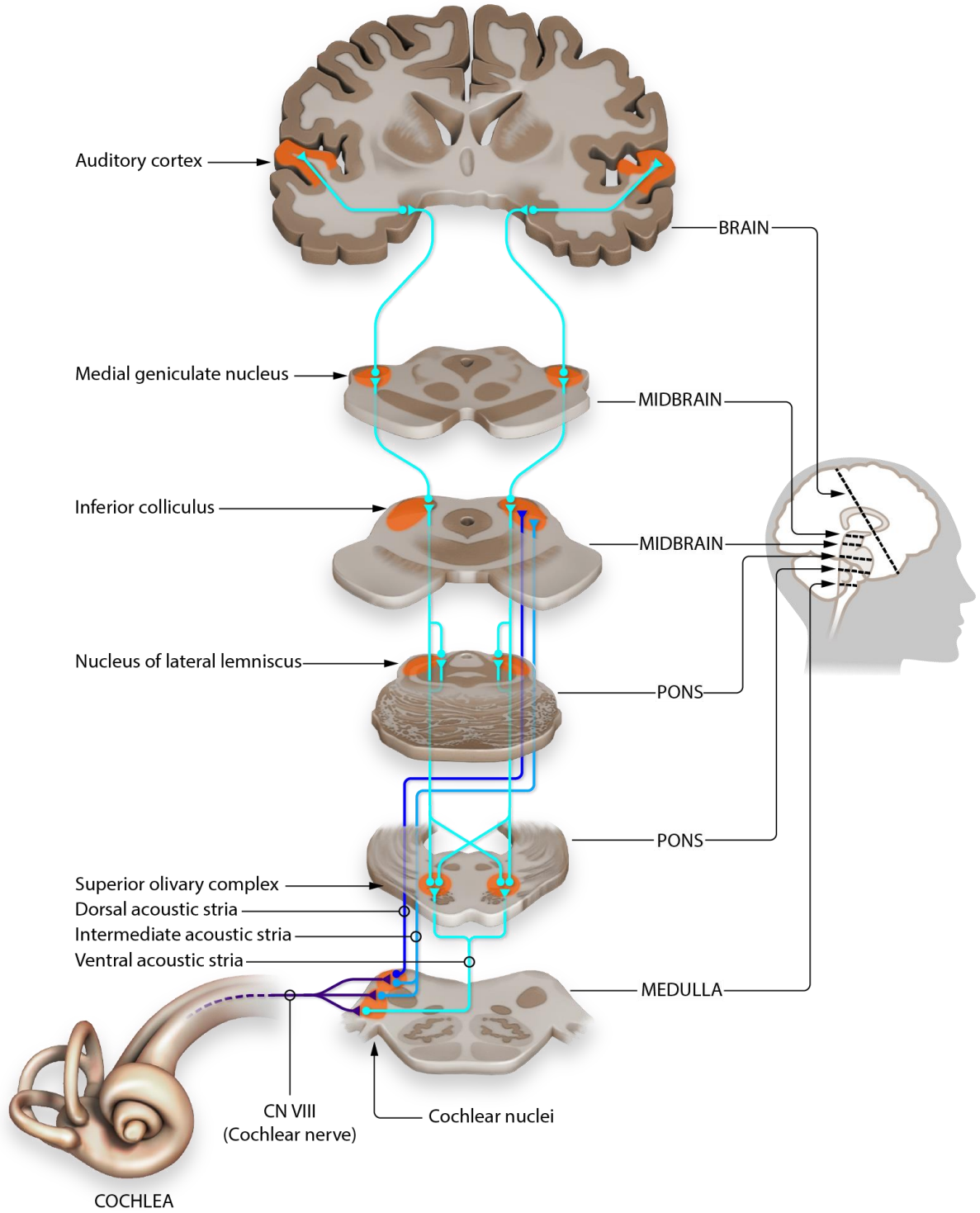

Figure 5 - Schematic graphic of the central auditory system.

Original author: Jonathan E. Peelle, License: CC BY 4.0.

On the one hand, the ventral auditory stream of the brainstem, which is mainly focused on the localisation of the sound by comparing the stimuli in both ears, starts in the AVCN. The spherical bushy cells of the AVCN project to the two medial superior olivary nuclei (MSO), which support the localisation of the sound by the comparison of the stimuli times between both ears; and to the lateral superior olive (LSO), in which are compared the intensities of the stimuli of both ears. The globular bushy cells of the AVCN project mainly to the contralateral medial nucleus of the trapezoid body (MNTB), being a step prior to the contralateral LSO. The MSO cells preferably react to sounds on the opposite side of the head, whereas the cells in the LSO to sound on the same side. The major output of the MSO goes to the ipsilateral IC, while the principal output of the LSO is crossed and projects to the contralateral IC. In the IC, the effects of the timing-based localisation signals and the intensity-based localisation signals - from the MSO and LSO, respectively - are summed. The lateral lemniscus is the track through which the fibres reach the IC, it includes the dorsal nuclei of the lateral lemniscus (DNLL) and the ventral nuclei of the lateral lemniscus (VNLL). DNLL receives signals from the contralateral cochlear nucleus, the ipsilateral MSO and both sites of the LSOs. The DNLL project crossed inhibitory inputs to both ICs and to the contralateral DNLL, increasing the accuracy, contrast and dynamic range of the sound source localisation[14–16].

On the other hand, the dorsal auditory stream of the brainstem begins in the DCN and PVCN. In contrast to the ventral stream, which is more direct, the dorsal stream performs more diverse functions, leading to a complex stimulus analysis. The DCN plays a very important role, not only in the excitation, but also in inhibitory responses. Moreover, the DCN is related to the start of the non-lemniscal auditory system division, which is a functional division involved in the complex multimodal integration and reflexes. The PVCN participates in timing and spectral analysis. The octopus' cells of the PVCN, specialised in extracting temporal fluctuations from the stimuli, arrive at the contralateral lateral lemniscus. The stellate cells of the PVCN, which could be excitatory and inhibitory, project to the contralateral VNLL and IC. The VNLL mainly extracts the temporal patterns from complex sound and projects them to the ipsilateral IC[14–16].

The first location in which the ventral and dorsal stream, therefore the sound localisation and sound pattern identification, converge is in the IC where the complexity is higher. The central nucleus of the IC, named central IC, is tonotopically organised due to a laminar anatomic structure. The central IC receives afferents from the LSO and MSO, because of that most of its cells have spatial directional selectivity for signals arriving from both ears, the pattern of

excitation and inhibition of these neurons is modified constantly due to the adaptation of the arrival of stimuli at different times. The external nucleus of the IC receives signals from the contralateral DCN and from the central IC, which excite its neurons and somatosensory information that inhibits its neurons. Whereas the dorsal cortex of the IC obtains stimuli from the contralateral IC and descending signals from the auditory cortex. Neurons in the dorsal cortex start at high excitation levels at the onset of the stimuli, but show a great adaptation to changes, which is critical in detecting new sounds[14–16].

The MGB receives inputs from the IC and projects to the auditory cortex, which is closely connected. The ventral division of the MGB is part of the lemniscal auditory pathway, while the medial and dorsal are part of the non-lemniscal auditory pathway. The ventral division mainly obtains stimuli from the ipsilateral central IC and projects to the ipsilateral auditory cortex areas that are tonotopically organised. The medial division receives afferents from all the IC, superior colliculus, vestibular system and spinal cord; and sends signals to diverse auditory cortex and the amygdala. Therefore, the medial division has a multisensory task, involving emotional conditioning too. The dorsal division has signals from the dorsal cortex of the IC and from the somatosensory system; it projects to the non-tonotopic cortical areas[14–16].

1.2.3 Auditory cortex

Through the auditory pathway, the signals reach the primary auditory cortex, where the perception occurs. The auditory cortex is a complex structure, composed of grey matter, that receives afferents from the MGB via the thalamocortical pathway. It is located bilaterally, deep in the superior portion of the temporal lobes within the Heschl gyrus. The auditory cortex is divided in different areas, the core areas are the primary area (AI) and the anterior and posterior auditory fields (AAF and PAF). The core region of the AI - defined by histologic criteria - receives inputs from the specific auditory division of the MGB (the ventral division) and is tonotopically arranged. The belt areas of the AI surround the core areas, to them arrive signals from the dorsal and media divisions of the MGB and from specific core areas[13,15,16].

In the auditory cortex a more complex integration and interpretation of the sound stimulus is carried out. The outputs of the AI cortex travel to different locations in the same cerebral hemisphere and between cerebral hemispheres. Depending on the final location of the signal produced by the sound stimulus, a diverse cognitive function will occur[13,15,17].

1.3 MENIERE DISEASE

Meniere Disease (MD, MIM 156000) is a chronic inner ear disease, characterised by spontaneous episodes of vertigo usually associated with sensorineural hearing loss (SNHL), tinnitus and/or aural fullness[18]. MD was first described in 1861 by Prosper Ménière, when he documented a series of patients with episodic vertigo and hearing loss. Besides, he reported the relation between vertigo and the inner ear[19].

Vertigo attacks are considered as the main symptom in the first years of the disease; however, persistent tinnitus is described as the most troublesome symptom by many MD patients[20,21]. MD is associated with other comorbidities such as migraine, benign paroxysmal positional vertigo and systemic autoimmune disease[18,22]. The disease's onset and the progression are provoked by the combination of genetic, epigenetics and environmental factors, leading to a multifactorial disorder[23].

MD is a rare disorder in the general population, nevertheless the occurrence in the different studies depends on the population and the diagnostic criteria chosen. The incidence of MD globally varies from 8.2 to 157 per 100,000 individuals per year; the prevalence of MD ranges from 3.5 per 100,000 in the Japanese population to 513 per 100,000 in the Finnish population. Despite the differences, it has been demonstrated a higher prevalence in Caucasians compared to other populations[22].

Nowadays, there is no consensus MD treatment, which varies from lifestyle and dietary modifications to surgery. The most useful dietary recommendation is the sodium restriction and high water ingestion. Besides, pharmacologic therapies include betahistine, diuretics, steroids or intratympanic gentamicin. Surgical treatments are carried out in less than 5% of patients, entailing endolymphatic sac surgery, semicircular canal occlusion, vestibular neurectomy, labyrinthectomy or cochlear implantation[22].

1.3.1 Diagnosis of Meniere Disease

The diagnostic criteria for MD were reformulated in 2015 by the Classification Committee of the Barany Society, the Korean Balance Society, the Japan Society for Equilibrium Research, the European Academy of Otology and Neurotology and the Equilibrium Committee of the American Academy of Otolaryngology-Head and Neck Surgery (AAO-HNS). The diagnostic criteria reformulation aims to develop an international consensus, to improve the homogeneity

of the clinical data across studies. The MD diagnostic criteria was based on clinical symptoms and was defined in two categories: definite MD and probable MD (Table 1)[18].

Table 1 - Criteria for diagnosis of Meniere Disease (MD). *Described by the International Classification Committee for Vestibular Disorders of the Barany Society in 2015.*

Definite MD
A. Two or more spontaneous episodes of vertigo, each lasting 20 minutes to 12 hours.
B. Audiometrically documented low- to medium frequency sensorineural hearing loss in one ear, defining the affected ear on at least one occasion before, during or after one of the episodes of vertigo.
C. Fluctuating aural symptoms (hearing, tinnitus or fullness) in the affected ear.
D. Not better accounted for by another vestibular diagnosis.
Probable MD
A. Two or more episodes of vertigo or dizziness, each lasting 20 minutes to 24 hours.
B. Fluctuating aural symptoms (hearing, tinnitus or fullness) in the affected ear.
C. Not better accounted for by another vestibular diagnosis.

1.3.2 Subgroups of Meniere Disease

MD is a heterogeneous condition based on the patient's clinical data, because of that, MD should be considered as a clinical syndrome with different aetiologies, instead of a single disease. This heterogeneity, added to the different comorbidities suffered by MD patients, considered in the therapeutic management[23].

MD individuals have been classified according to different criteria. Belinchon et al.[24] divided them in unilateral MD if they showed symptoms in only one ear; and bilateral MD if the symptoms evolved in both ears, presenting a more severe hearing loss. Furthermore, MD cases were subgrouped according to their cytokines levels, which could be classified into two groups by the IL-1β level[25]. In addition, Bächinger et al.[26] classified MD patients into degenerated and hypoplastic measuring the position of the vestibular aqueduct and the size of the endolymphatic sac.

Frejo et al.[27,28] presented 10 clinical subgroups of MD patients according to clinical data, divided in five groups with unilateral MD individuals and five with bilateral MD individuals. Both unilateral and bilateral clusters identified familial MD (FMD) cases, MD patients presenting migraine and MD patients with a comorbid autoimmune disease (Table 2). To be

considered as a FMD case at least one other relative (first or second degree) should fulfil all the criteria of definite or probable MD[18], the rest of them could be considered as sporadic MD (SMD).

Table 2 - Subgroups of Meniere Disease (MD) individuals according to clinical data.

Unilateral MD
Type 1 (53%): SMD, classical phenotype; the most common type.
Type 2 (8%): SMD, hearing loss precedes vertigo attacks; a rare condition.
Type 3 (13%): FMD.
Type 4 (15%): SMD with migraine.
Type 5 (11%): SMD with an autoimmune disease.
Bilateral MD
Type 1 (47%): SMD, unilateral hearing loss becomes bilateral.
Type 2 (17%): SMD, simultaneous hearing loss.
Type 3 (13%): FMD.
Type 4 (12%): SMD with migraine.
Type 5 (11%): SMD with an autoimmune disease.

SMD: Sporadic Meniere Disease; FMD: Familial Meniere Disease.

1.4 TINNITUS

Tinnitus is the conscious awareness of a tonal or composite noise for which there is no identifiable corresponding external acoustic source, which becomes Tinnitus Disorder when associated with emotional distress, cognitive dysfunction, and/or autonomic arousal, leading to behavioural changes and functional disability. Due to the lack of agreement on the tinnitus description, this definition was proposed by the Tinnitus Research Initiative (TRI)[29], a non-profit organisation composed by an international multidisciplinary group of tinnitus experts focused in scientific and clinical research. Commonly, tinnitus is explained as "ringing in the ears". The sounds experienced by individuals can be varied including ringing, buzzing, whistling, whooshing, or clicking sounds. Besides, the individual could feel the sound in one or both laterals, with different variability and pitch[29,30].

Owing to the absence of common criteria for the diagnosis of tinnitus, together with the fact that in many studies tinnitus is self-reported, it is difficult to calculate the prevalence of tinnitus in the population. Henry et al[31] described that tinnitus affects between 10% and 15% of adults worldwide and McCormack et al[32] showed that the prevalence ranges from 11.9% to 30.3%; although their results differed, it is confirmed the high prevalence of the symptom in the global population. Whereas only the 1% is severely distressed[29].

The clinical treatments of tinnitus depend on the clinician and mostly are focused on diminishing the awareness of tinnitus and its associated distress. In patients with hearing loss, the treatment of the hearing loss could be helpful with the tinnitus, it could be cochlear implantation or by drugs, or antidepressants if they also suffer from depression. Moreover, there are drugs in preclinical and clinical development that have as targets the NMDAR (NMDA receptor), Kv3.1 potassium channels, KCNQ2/3 channels or GABAergic transmission[33].

1.4.1 Tinnitus Disorder

The TRI recommends differentiating between "Tinnitus" and "Tinnitus Disorder", describing the first one as an auditory or sensory component that can be a secondary symptom caused by a trauma or disease. The second one is associated with suffering and should be considered as primary disorder[29].

In individuals with tinnitus disorder, tinnitus can affect the daily activities, prevent intellectual work, leading to an impairment in the quality of life. In some instances, tinnitus can motivate suicide. Furthermore, sudden severe tinnitus is related with hyperacusis and affective disorders such as phonophobia, anxiety or depression[29,34].

1.4.2 Tinnitus mechanism and auditory system

Despite its high prevalence, the underlying mechanisms of tinnitus remain poorly understood. For many years, it was assumed that the unique anatomical location of the pathology that caused tinnitus was the ear. Nevertheless, it was proposed that the abnormal function of neural circuits in the central nervous system was the cause of some forms of tinnitus[34–36].

Tinnitus is a phantom sensation of different types of sound; by definition, phantom sensations do not refer to the structure where the abnormal neural activity that causes the sensation is created. Moreover, in tinnitus, a network of several brain structures, neural transmitters and receptor types are involved. In most of the nervous system's disorders, one structure is pathologic and the rest receive abnormal signals, this cascade of structures complicates the study of this kind of disorders[34,37].

One of the possible causes of tinnitus is that it could originate at the synapses between hair cells and the auditory nerve, or within the auditory nerve itself. It has been studied that a slight injury to the auditory nerve has a variety of effects on the neurons in the cochlear nucleus, one of them is the rising in the excitation, which could be related with tinnitus[34,37].

Furthermore, the nervous system maintains a balance between excitement and inhibition by the sound. If this is altered and the inhibition is reduced, an "amplification" is produced in the neural networks entailing self-oscillations, which could lead to some forms of tinnitus[34].

The most common cause of tinnitus is the activation of neural plasticity[34,37]. Neural plasticity is the capacity of the nervous system to modify its function and structure in response to experience and injury[38]. The activation of neural plasticity can open (unmask) dormant synapses or close (mask) synapses that are conducting normally, in both cases it leads to abnormal changes in connectivity, producing the pathology. An example of change in connectivity, which often is related with tinnitus, is the activation of non-classical sensory pathways through re-routing the information, which entails a cross-modal interaction.

Moreover, the abnormalities in non-classical pathways could be related with symptoms that usually co-occur with tinnitus, such as hyperacusis, mood disorders, phantom sensations, improved perceptual capabilities, or atypical sensory experiences. The plasticity disorders – some forms of tinnitus, phantom sensations in other sensory systems, central neuropathic pain or spasm in motor systems – are developed by neural plasticity activation[34].

The lesions in the dorsal nucleus were studied in cats, it was reported that the cats with these lesions could not respond to sound elevations with reflexive responses, however they could learn to make responses through training. It was suspected that the dorsal nucleus is the beginning of certain auditory reflexive pathways due to the failure in reflexive responses but not the learned. Following these results, lesions of the dorsal nucleus could be one of the causes of the generation of tinnitus[15].

The final objective of understanding the tinnitus mechanisms is to find and develop a treatment. This aim is very challenging because of the complexity and heterogeneity of the disorder, each form of tinnitus could be caused by a different pathophysiology[37,39]. Besides, it is necessary to perform controlled studies to select well-defined endophenotypes of tinnitus patients.

1.4.3 UNITI project

The European Union funded a multidisciplinary project entitled "Unification of Treatments and Interventions for Tinnitus Patients" (UNITI, grant no.848261) under its Horizon 2020 framework. The ambition of the UNITI project was to identify predictive factors for the tinnitus treatment with the main objective of being able to provide a personalised treatment for tinnitus. To achieve this goal, data from existing biobanks, databases and collected in e/m-health applications were used to determine correlation between clinical variables and treatments[39,40].

To achieve this goal, the UNITI project has the following research purposes[39]:

1. Analysis of the existing clinical data to describe subgroups of tinnitus individuals.
2. Identification of genetic and blood biomarkers, performing Whole Exome Sequencing (WES) in well-characterised tinnitus cases; and Proximity Extension Assays.
3. Conduction of a Randomised Clinical Trial (RCT) with 500 patients from 5 European Centres.
4. Design a Decision Support System, a computational model trained with the generated data.

5. Analysis of the financial estimation via calculating the economic effects resulting from the interventions based on quality-adjusted life years.

To evaluate the best treatment or treatments for each type of affected individual, a multicentre Randomised Clinical Trial (RCT) was carried out in five clinical sites in different countries in the European Union[40]:

1. University of Regensburg, Regensburg, Germany (RCT coordinator).
2. Charité - Universitaetsmedizin Berlin, Berlin, Germany.
3. Ethniko Kai Kapodistriako Panepistimo Athinon, Athens, Greece.
4. Hospital Universitario Virgen de las Nieves/Hospital Clínico Universitario San Cecilio, Granada, Spain.
5. Katholieke Universiteit Leuven, Leuven, Belgium.

The cohort of participants was composed of 500 patients with mild to severe tinnitus distress, the study was designed to evaluate the effect of four treatments tested single or in combination during 12 weeks. The tinnitus therapy approaches were[40]:

1. Sound therapy: It is self-administered via a mobile application which can reproduce 64 different sound stimuli.
2. Hearing aids: It is applied based on the hearing function of the patient.
3. Structured counselling: It provides education, counselling and tips for a life with tinnitus on a daily basis.
4. Cognitive behavioural therapy for tinnitus: It is performed in weekly face-to-face group meetings and guided by psychologists or psychotherapists.

Finally, the Decision Support System was developed to reach the main objective of the UNITI project, which was to determine the optimal treatment or combination of treatments for each tinnitus patient based on selected variables. The scientific results are implemented in a user-friendly platform to support the daily practice of tinnitus experts[39,40].

1.4.4 Meniere Disease and tinnitus

MD is one of the many causes of tinnitus, which is often associated with the first episode of vertigo. Mostly, tinnitus in MD corresponds to the low-frequency SNHL. In the first stages of MD, the tinnitus complaint is secondary and usually intermittent. Most patients report a loud tinnitus before or during vertigo attacks; however, it could improve or disappear later on. As

the disease progresses, tinnitus can become persistent once the hearing loss becomes permanent and it is described as the most disturbing complaint by many MD patients[18,20,21,34].

Stephens et al[41] confirmed that tinnitus related to anxiety, sleep and depression was the main impact when the other MD's symptoms decline, along with complicating some listening situations and interactions with others. Individuals suffering from tinnitus presented a worst quality of life, mainly when studying the mood. Herraiz et al[20], identified a significant relation between the intensity of the tinnitus and the worse hearing loss or hyperacusis. Moreover, they described that MD cases showed more severe tinnitus than patients with other inner ear aetiologies. Perez-Carpena et al[42] found that in MD patients, the discomfort produced by tinnitus in those with tonal tinnitus was higher than in those with noise-like tinnitus; in addition it was permanent in 77.8% of unilateral and 80% of bilateral MD cases.

1.4.5. Hyperacusis and tinnitus

Hyperacusis is a rare hearing disorder characterised by the loudness perception of the sound, individuals with that phenotype perceive the sound that commonly are considered innocuous as intolerable. The patients report the sensation as painful or excessively loud, which in some cases leads to sound avoidant behaviour. The hyperacusis in some cases co-exists with tinnitus, which increases the impairment in their daily life. The prevalence in children and adolescents was measured between 3.2% and 17.1%, whereas in adults between 8% and 15.2%[43]. Furthermore, it was observed that about 40% of patients with moderately severe tinnitus suffer also from hyperacusis and it reaches 80% in the individuals with severe tinnitus[44].

1.5 GENETICS

Genetics is the study of heredity and the variation of inherited characteristics, how certain traits are passed from parents to offspring as a result of the sequence of their DNA[45]. The genome is the complete DNA sequence of an individual, whereas the exome consists of the coding regions (exons) that represent less than 2% of the genome[46].

1.5.1 Genetic variants

A reference human genome sequence is an established, high-quality and well-accepted human genome sequence. A variant is a permanent alteration in the DNA sequence that differs from what is found in the reference genome. The modification of a single base pair is called single nucleotide variant (SNV), the most common type of variant (Figure 6A). In addition, there are insertions or deletions (indels) of a certain number of nucleotides in the sequence (Figure 6B). If the indel is composed of a multiple of three nucleotides, it does not alter the reading frame (each codon is made up of three nucleotides) and they are called in frame insertion or deletion; however, if it alters the frame the indel is named frameshift. A missense variant is a SNV encodes a different amino acid at a particular position in the resulting protein. A loss-of-function (LoF) variant is predicted to disrupt protein-coding genes severely; it could be a SNV or an indel. There are different kinds of LoF depending on the consequence they produce in the protein, such as stop gain, stop loss, start loss or frameshift variants. Moreover, the genomic variation could affect larger chromosomal regions, which are called structural variants (SVs) and they involve at least 50 nucleotides and they could affect many thousands of nucleotides, as examples inversions and translocations. If the SV implies a change in the final number of nucleotides, it is called a copy number variant (CNV), such as insertions, deletions, duplications or tandem repeats (Figure 6C). In addition to the variants that fall in protein coding regions, there are variants in non-coding regions, such as promoters, splice sites and other regulatory elements[47–51].

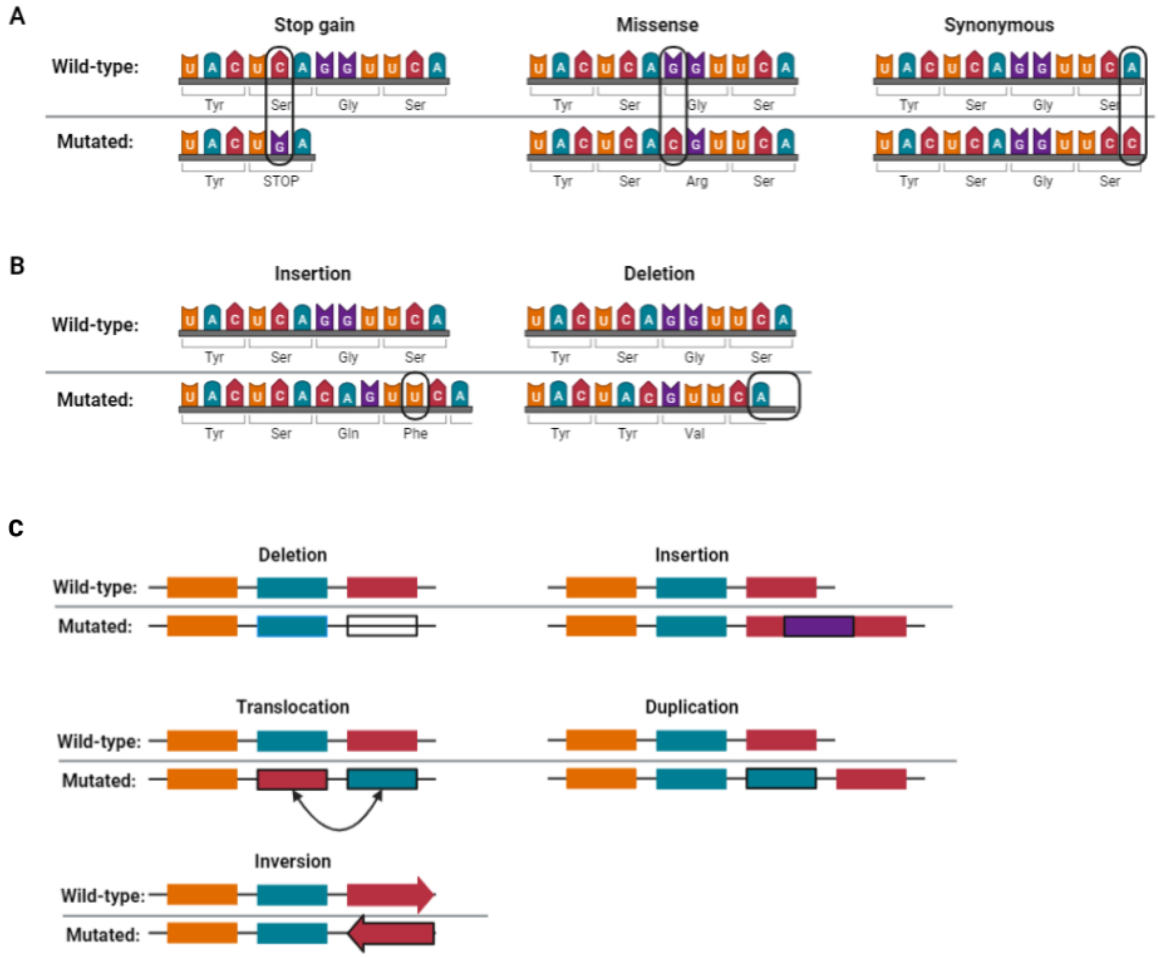

Figure 6 - Type of genetic single nucleotide variants (A), short insertions and deletions (B) and structural variants (C).

The inheritance pattern is autosomal dominant (AD) if the affected individual has a single mutant copy of the disease-associated gene and a normal copy; a variant is described as dominant if the phenotype is expressed in heterozygotes. Moreover, the penetrance could be considered incomplete when the heterozygous individual shows a phenotype less severe than that expressed in homozygotes. The inheritance pattern is autosomal recessive (AR) if the affected individual has the two copies of the disease-associated gene; a variant is considered recessive if the phenotype is observed only in homozygotes[51].

When the variant is compared with a reference population, it is possible to obtain its allele frequency (AF). Variants with an AF less than 0.01 are defined as rare variants and they are called single nucleotide polymorphism (SNP). Variants with an AF between 0.01 and 0.05 are considered less common and variants with an AF greater than 0.05 are common. Besides, variants not defined in the population are considered novel[52].

1.5.2 Genomics revolution

Genomics is the study of the genome, it encompasses a wide range of scientific disciplines, including genetics, computational biology or molecular biology. It has revolutionised the knowledge and the research in all biological areas, especially since the development of high-throughput DNA sequencing methods, one of the major technological advances of the past century. DNA sequencing is the process by which the exact sequence of bases in a DNA sample is determined[53].

The discovery of DNA as material genetic in the mid-20th century marked the beginning of a new era in genetics and the birth of molecular genetics. Since the development of the Sanger method - the first DNA sequencing technology - in 1977[54], next-generation sequencing tools have greatly improved to provide higher speed and accuracy and to reduce the cost hugely. These advances motivated the creation of the Human Genome Project, with the main objective of finding out the human reference genome for the first time[55], taking over a decade and costing $2.7 billion[56]. After 30 years of studies and several reference sequences described, the complete sequence of the genome could be fully uncovered last year with the presentation of the T2T-CHM13 sequence by the Telomere-to-Telomere (T2T) Consortium, which is formed by 3.055 billion base pair[57]. Despite the breakthrough of this publication, it does not represent the broad genomic diversity in the human population, to tackle this bias, the Human Pangenome Reference Consortium is working[58].

The rapid development of DNA sequencing methodology has transformed the genomics field, which plays an important role in biomedicine. The two methodologies more used are WES and whole genome sequencing (WGS). One of the main contributions is the analysis of genetic variants. The results are essential for understanding the genetic basis of groups of individuals and their differences, such as susceptibility to disease, leading to precision medicine[59].

1.5.3 Public genomics repositories

Due to the large amount of omics data generated by new technologies, creating several different omics databases, among them the genomics databases, was necessary. These repositories centralise and allow public access to genomics data and play a crucial role in data sharing, collaboration and reproducibility in genomics research.

Genomics repositories store DNA sequences, transcriptomes, epigenomes and metagenomes, associated with metadata. There is a large number of databases, focusing on different types of data or characteristics of their own. The National Center for Biotechnology Information (NCBI) contains different databases, such as GenBank with DNA sequences or Gene Expression Omnibus (GEO) that hosts gene expression data[60]. Similar databases are hosted in the European Bioinformatics Institute (EBI), as the European Nucleotide Archive (ENA) containing DNA and RNA sequences[61] or ArrayExpress with gene expression data[62].

Moreover, databases of genetic variants have been developed, of these, the most important in humans is the Genome Aggregation Database (gnomAD). The international coalition of investigators that develop gnomAD have the main objective of aggregating and harmonising both exome and genome sequencing data, which has been obtained from a wide variety of large-scale sequencing projects. One of the strengths of gnomAD is the high-quality data obtained by well-curated projects and strict quality control standards. The gnomAD database provides detailed information for each variant, such as the frequency in the global and in different populations[63]. In Spain, the Collaborative Spanish Variant Server (CSVS) has been created to provide information about the genomic variability of the Spanish population. CSVS contains SNVs and insertion/deletion variants data from a crowdsourcing initiative[64]. Besides, recently the Spanish Copy Number Alteration Collaborative Server (SPACNACS) has been published offering the Spanish population's CNV data[65]. In the United Kingdom, the UK Biobank is a large-scale database containing biomedical data and is a research resource. It includes genetic data and clinical information from a half a million participants[66].

Furthermore, some public genomics repositories also provide tools and resources for genomics research. For instance, BLAST - hosted by the NCBI - a tool that allows to look for similar sequences by similarly[60]; or the UCSC Genome Browser - developed by the University of California, Santa Cruz (UCSC) - a web tool for visualisation and analysis of genomics data[67].

1.5.4 Variants annotation assessment

It is essential to distinguish between benign and pathogenic variants in genetics studies to prioritise them. Henceforth, tools to predict the pathogenicity have been developed thanks to existing data and technologies. The Combined Annotation Dependent Depletion (CADD) is a tool for scoring the deleteriousness of SNV and insertion and deletion variants in the human genome[68]. As different tools were developed, to make the results more consistent, the

guidelines for the interpretation of sequence variants by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) were defined[69,70].

Furthermore, the variants can be assessed based on the confidence that variants classified as LoF indeed affect the protein with the Loss-Of-Function Transcript Effect Estimator (LOFTEE)[63].

The genomic constraint is the selective pressure that impedes that some genetic variations have a higher frequency in the population. It measures how a specific region or gene in the genome is evolutionarily conserved, which means that it has been stable over the time because of the important function that it has, usually they are related with essential biological processes. The regions with a high constraint are less tolerant to genetic variation, it is due to that variants in it would lead to important and pathogenic changes for the individual. Therefore, the regions with a low constraint are more tolerant to variations, being less crucial to selective pressure[63,71,72].

1.5.5 Variants analysis

The single-variant analyses (SVA) are less powerful for rare variants than for common variants with identical effect sizes; moreover, sample sizes should be large enough to be useful for rare variants. Gene burden analysis (GBA) is a type of aggregation test that collapses the frequencies for multiple genetic variants into a single score, then tests for association between the score and a trait. This research approach is powerful to study phenotypes caused by a large proportion of variants and all the effects are in a similar direction[73]. Finally, a genome-wide association study (GWAS) is used to identify candidate genomic variants statistically related with a risk for a particular trait. GWAS are useful with large cohorts and to study common variants[74].

1.6 GENETICS OF TINNITUS

The research community working on tinnitus has as one of the main goals to understand the condition's origin, a scientific challenge due to the heterogeneity of the symptom and the associated comorbidities. It is known that tinnitus has a multifactorial origin, as environmental factors, being the most studied noise exposure[75]. Analysing clinical data from adoptees demonstrated that the association between the risk of suffering from tinnitus for the adoptees and their biological parents was significant, whereas it was not significant between the adoptees and their adoptive parents. These results suggested that the genetic factors are involved in developing tinnitus[76]. In families, it was observed a significant effect of tinnitus of 15%[77]. Besides, the recurrence risk for tinnitus in siblings was greater in women than men, being significant in both cases[78]. Studies carried out in twins revealed that the heritability of bilateral tinnitus cases was greater than in unilateral tinnitus, 56% against 27%. Moreover, it was higher in men than women, 68% versus 41%[79]. Another study performed with male twin pairs, established that the relative proportion of additive genetic factors was approximately 40%[80].

Different approaches have been used in the past years to analyse the inheritance of tinnitus. One of the most common genomics analyses in the study of the diseases was also used in tinnitus. Four different GWAS revealed various results: (a) a SNV in *GPM6A* and 19 independent loci[81], (b) a SNV in an intergenic region and another in the intron of *TNFRSF1A*[82], (c) three variants close to *RCOR1*[83], (d) 17 suggestive SNVs spanning in 13 genes and a missense in *WDPCP*[84]; the last two using cohorts from UK Biobank. By genotyping, SNVs in the *GRM7, 5-HTTLPR, ADD1, BCR* and *KCNQ1* genes have correlated tinnitus severity[85–89]. Moreover, a mitochondrial variant was related to tinnitus[90]. An epigenetic study revealed that one CpG site for *BDNF* and three for *GDNF* were differentially methylated in chronic tinnitus individuals[91]. Using GBA to pinpoint candidate genes for tinnitus the *ANK2, TSC2* and *AKAP9* genes were enriched in missense variants[92]; by GBA and analysing also the SVs the candidate genes were *CACNA1E, NAV2* and *TMEM132D*[93].

Most of these studies have not been replicated, showing the limitations of GWAS based on Biobank data with poorly-defined phenotypes for tinnitus, hearing loss and other common associated comorbidities[94]. As an example of the complexity in identifying variants associated with tinnitus through GWAS, Trpchevska et al.[95] did not identify any variants associated with tinnitus in the largest cohort in which a GWAS has been conducted for this symptom. However, they did identify 48 risk variants related to hearing loss.

2 HYPOTHESIS

Tinnitus emerges as the most distressing symptom in the experience of numerous individuals with MD, furthermore the perceived severity of tinnitus is different in each patient. Recent studies have demonstrated that there exist variants in different genes related to severe tinnitus.

The working hypothesis is that in MD patients, the tinnitus severity is caused by multiple factors, including genetic and epigenetic variants for each individual. While some of these variants will have a pleiotropic effect on tinnitus and other associated comorbidities, such as hearing loss, other rare variants would have a tinnitus specific large effect size that will better explain severe or mild tinnitus. Therefore, those with a greater discomfort due to tinnitus and, in the opposite, those without annoyance caused by tinnitus, will have a burden of rare variants in candidate genes related to a risk, or protection, to develop tinnitus.

HIPÓTESIS

El acúfeno se describe como el síntoma más molesto para numerosos individuos con EM. Además, la severidad percibida del acúfeno varía en cada paciente. Estudios recientes han demostrado que existen variantes en diferentes genes relacionados con el acúfeno severo.

La hipótesis de trabajo es que, en los pacientes con EM, la severidad del acúfeno es causada por múltiples factores, incluyendo variantes genéticas y epigenéticas para cada individuo. Mientras que algunas de estas variantes tendrán un efecto pleiotrópico en el acúfeno y otras comorbilidades asociadas, como la pérdida de audición, otras variantes raras tendrán un gran efecto específico para el acúfeno que explicará mejor el acúfeno severo o leve. Por lo tanto, aquellos con una mayor incomodidad debido al acúfeno y, por otro lado, aquellos sin molestias causadas por el acúfeno, tendrán una carga de variantes raras en genes candidatos relacionados con un riesgo o protección para desarrollar acúfeno.

3 OBJECTIVES

The main objective in this project is to identify the main genes and biological processes associated with severe tinnitus.

Specific objectives:

1. To generate a curated dataset containing Exome Sequencing data from 429 MD patients with full clinical characterization, particularly for hearing loss and severe tinnitus.
2. To demonstrate the aggregated effect of rare variants in hearing loss and tinnitus in MD patients by GBA.
3. To identify novel mutations in the coding regions (exome) of MD patients with permanent severe tinnitus.
4. To determine SVs and CNVs in the coding regions of MD patients with permanent, severe tinnitus compared with non-permanent tinnitus.
5. To analyse the implications of the genetic biomarkers discovered via functional analyses using public databases.

OBJETIVOS

El objetivo principal de este proyecto es identificar los genes y procesos biológicos principales asociados con el acúfeno severo.

Objetivos específicos:

1. Generar un conjunto de información curado que contenga datos de secuenciación de exoma de 429 pacientes con EM con una caracterización clínica completa, particularmente para la hipoacusia y el acúfeno severo.
2. Demostrar el efecto agregado de las variantes raras en la hipoacusia y el acúfeno en pacientes con EM mediante el análisis de la carga génica.
3. Identificar nuevas mutaciones en las regiones codificantes (exoma) de pacientes con EM que presentan acúfeno severo permanente.
4. Determinar variantes estructurales (*structural variants*, SVs) y variantes en el número de copias (*copy number variants*, CNVs*)* en las regiones codificantes de pacientes con EM que tienen acúfeno severo permanente en comparación con aquellos que tienen acúfeno no permanente.
5. Analizar las implicaciones de los biomarcadores genéticos descubiertos a través de análisis funcionales utilizando bases de datos públicas.

4 MATERIALS & METHODS

4.1 SYSTEMATIC REVIEW OF THE GENETIC ARCHITECTURE OF FAMILIAL MENIERE DISEASE

4.1.1 Study design

The aim of this study was to perform a systematic review of sequencing studies conducted in individuals with FMD. This work was based on the Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines[96].

4.1.2 Research question and selection criteria

To depict the genes involved in the genetic architecture of FMD, the subsequent question was raised: *which genes have been found to be associated with FMD?* To address this query and, according to the systematic reviews methodology, the PICOS strategy was followed:

- Population: Patients diagnosed with FMD.
- Intervention: Genes related to FMD found by sequencing studies, mainly focused on rare variants.
- Comparison: The AF of the variants in the candidate genes were compared with the frequencies in reference population databases.
- Outcome: The candidate genes and the pathogenicity prediction of their rare variants were reported.
- Study design: Familial segregation studies were analysed.

4.1.3 Search strategies

The bibliographic search was conducted in the PubMed database, the query was: (familial [Title/Abstract] OR family [Title/Abstract] OR gene [Title/Abstract] OR genes [Title/Abstract] OR inheritance [Title/Abstract] OR variation [Title/Abstract] OR mutation [Title/Abstract]) AND (Meniere Disease [Title/Abstract] OR Meniere's Disease [Title/Abstract]). It was performed the 11th of August 2020; it was restricted to the last 20 years at the moment - from 2000 to 2020 - and written in English.

4.1.4 Exclusion criteria

From the retrieved papers, those meeting the following criteria were discarded from the study:

- Articles not published in English.

- Studies carried out in animals.

4.1.5 Quality assessment of selected studies

The titles and abstracts of the obtained articles were scrutinised to eliminate reviews, meta-analysis or those not relevant for this study, such as non-genetic studies, pharmacogenomics or clinical studies. Moreover, it was defined the three following criteria to remove those articles not optimal for the proposed research:

- Is the study performed with two or more members of a family diagnosed with MD or with patients from different families but all of them diagnosed with FMD?
- Has the study reported a gene or a position in the genome statistically significant when compared to genome reference databases?
- Has the study used an accurate methodology and is it described with enough details to validate its findings?

The articles that conducted a study with which it was possible to answer *yes* to all the previous questions were ultimately chosen for the review.

4.1.6 Data extraction and synthesis

The following data was obtained from each study selected to summarise the information: first author's last name, publication year, study design, population of the individuals, number of patients in the study, sex of the cases, criteria to diagnose MD, sequencing method used, candidate gene or genes and variant or variants reported. In addition, the genomic position, the variant type and the reference single nucleotide polymorphism (rs) was obtained for each variant. The AF of the variants were obtained from the original article and checked in the reference population databases. The AF from the global population from gnomAD was retrieved in all the cases, moreover it was used the AF from Exome Aggregation Consortium (ExAC) or country specific reference databases as CSVS and SweGen - from Spain and Sweden, respectively[63,64,97,98]. Besides, the pathogenicity of each variant was calculated according to the ACMG criteria[69,70] and predicted with the CADD score[99]. The inheritance pattern of the variants was defined based on the information described in each article. As a result, in the review, a list of candidate genes, their variants and their inheritance pattern of each of them was obtained as principal conclusion.

Therefore, those genes were considered from further analysis regarding tinnitus, because the patients used in this project were diagnosed with MD.

4.2 STUDY COHORT

4.2.1 Human subjects

In this work, a total of 429 individuals were used to create a database with exome sequencing data. The individuals were recruited from Spanish referral centres (N = 373) and Zurich University (N = 56); and a full clinical characterization, particularly for hearing loss and tinnitus were performed. Among all the individuals, the majority (N = 407) were MD patients, diagnosed as definite MD according to the diagnostic criteria described by the International Classification Committee for Vestibular Disorders of the Barany Society[18]. The rest of them (N = 22) were healthy controls who were relatives of the MD patients.

The Institutional Review Board approved the experimental protocols carried out during this study in all participating hospitals and every patient signed written informed consent. The study was carried out according to the principles of the Declaration of Helsinki revised in 2013 for investigation with humans.

4.2.2 Inclusion criteria

The inclusion criteria of individuals in this study, out of the total number of those in the dataset, were to be diagnosed with definite MD and the availability of the result for the Tinnitus Handicap Inventory (THI) questionnaire. Hence, a total of 310 MD patients were included in this cohort.

4.2.3 Clinical data

Among other clinical features, the main clinical data collected for the cohort was:

- Age.
- Sex.
- Sporadic of familiar MD.
- Tinnitus laterality.
- Tinnitus onset.
- Visual analogue scale (VAS) of annoyance for tinnitus[20].
- THI score.
- Hypersensitivity to sound (GÜF test) score.
- Patient Health Questionnaire depression scale (PHQ-9) score.
- Hospital Anxiety and Depression Scale (HADS)-A score.
- HADS-D score.

- Audiograms in both ears for air and bone conduction from 125 to 8,000 Hz. The 500 Hz, 1,000 Hz and 2,000 Hz frequencies were retrieved to calculate the pure-tone average (PTA). Moreover, the 4,000 Hz and 8,000 Hz frequencies were used as predictors of noise-induced hearing loss and age-related hearing loss, respectively.

4.2.3.1 Tinnitus Handicap Inventory (THI) questionnaire

All the individuals answered the Spanish validation of the THI. It evaluates the annoyance related to tinnitus and establishes the subscales functional, emotional and catastrophizing tinnitus. The questionnaire is composed of 25 items (Table S1, Table S2), which could be answered as: "yes" (4 points), "sometimes" (2 points), or "no" (0 points). The total score ranges from 0 to 100 and classifies the individuals in five levels of severity, explained in Table 3[20,100].

Table 3 - Classification according to Tinnitus Handicap Inventory (THI) score.

THI score	*Tinnitus handicap*	*Grade*
0-16	Slight or no handicap	I
18-36	Mild handicap	II
38-56	Moderate handicap	III
58-76	Severe handicap	IV
78-100	Catastrophic handicap	V

4.2.3.2 Visual analogue scale (VAS) of annoyance for tinnitus

The VAS is a rating scale in which the subject ranks its annoyance for tinnitus. It ranges from 0 to 10 and it has places inside the scale for each unity determining the intervals corresponding to the differences in preference as perceived by the subject[101].

4.2.3.3 Questionnaire of hypersensitivity to sound (GÜF test)

The Spanish version of the *Geräuschüberempfindlichkeit* (GÜF) test was asked to the individuals to assess the hyperacusis. It consists of 15 questions (Table S3, Table S4) with the options: "never" (0 points), "sometimes" (1 point), "frequently" (2 points) and "always" (3 points). Hence, the total score ranges between 0 and 45 and the individuals could be organised in the four grades of handicap, described in Table 4[102,103].

Table 4 - Classification according to questionnaire of hypersensitivity to sound (GÜF test) score.

GÜF score	*Hyperacusis handicap*	*Grade*
0-10	Slight handicap	I
11-17	Moderate handicap	II
18-25	Severe handicap	III
26-45	Very severe handicap	IV

4.2.3.4 Patient Health Questionnaire depression scale (PHQ-9)

To evaluate depressive symptoms, the Spanish version of the PHQ-9 questionnaire was used. It has 9 questions (Table S5, Table S6) with the next possible answers: "not at all" (0 points), "several days" (1 point), "more than half of the days" (2 points) and "nearly every day" (3 points). The total score ranges from 0 to 27, resulting in the extent of depressive symptoms, categorised in four groups, defined in Table 5[104,105].

Table 5 - Classification according to Patient Health Questionnaire depression scale (PHQ-9) score.

PHQ-9 score	*Level of depression severity*
0-4	Minimal symptoms
5-9	Mild symptoms
10-14	Moderate symptoms
15-19	Moderate severe symptoms
20-27	Severe symptoms

4.2.3.5 The Hospital Anxiety and Depression Scale (HADS)

The subjects answered the Spanish version of the HADS questionnaire, which detects the states of anxiety (HADS-A) and depression (HADS-D). It is formed by seven questions for anxiety and seven for depression (Table S7, Table S8), with different responses that are punctuated from 0 to 3. Hence, the total scores for each state ranges from 0 to 21, being the classification explained in Table 6[106,107].

Table 6 - Classification according to Hospital Anxiety and Depression Scale (HADS) A and D score.

HADS-A or HADS-D score	*State of anxiety or depression*
0-7	Normal

8-10	Borderline abnormal (borderline case)
11-21	Abnormal (case)

4.2.4 Subgroups of individuals

Due to the heterogeneity of MD, it is convenient to create subgroups of the sample, to make them as homogeneous as possible. In this study, the most relevant condition to investigate is tinnitus; however, hyperacusis is also of interest.

4.2.4.1 Subgroups according to Tinnitus Handicap Inventory (THI) distribution

To select subgroups of patients with a homogeneous annoyance related to tinnitus, the distribution of the THI score in the sample was used. Utilising the *quantile* function in R the first quartile, second quartile (median) and third quartile were calculated. As the THI must be an even number, in case a quartile was an odd number, one point was added to make a correct division.

Then, the individuals of the sample were separated in four subgroups according to the intervals, called THI-subgroups:

- Interval 1 (I1): The individuals between the minimum value and the quartile 1.
- Interval 2 (I2): The individuals between the quartile 1 and the quartile 2.
- Interval 3 (I3): The individuals between the quartile 2 and the quartile 3.
- Interval 4 (I4): The individuals between the quartile 3 and the maximum value.

4.2.4.2 Subgroups according to hypersensitivity to sound (GÜF) distribution

The same procedure was performed with the GÜF test scores, to classify the individuals according with the hyperacusis. In this case, odd numbers were allowed in the quartiles. The individuals also were classified (GÜF-subgroups) by intervals, called: I1-GÜF, I2-GÜF, I3-GÜF and I4-GÜF.

4.2.4.3 Subgroups according to the other variables

Moreover, the VAS score, PHQ-9, HADS-A and HADS-S questionnaires were used to subgroup the individuals, following the quartiles division.

4.2.5 Comparisons between variables

To study the relation between the different questionnaires used in this study (THI, GÜF, PHQ-9, HADS-A and HADS-D) and the VAS test, the Spearman's rank correlation coefficient (r) was calculated. It was calculated for all individuals as a whole and for each THI-subgroup separately. Moreover, the density of the THI-subgroups was represented for each test and the distribution between each pair of tests was studied using a scatter plot, indicating the THI-subgroups.

To identifying those individuals with a worse phenotype, not only in tinnitus, but also in hyperacusis, depression and anxiety, an alluvium plot separated by interval was designed.

4.2.6 Statistical analysis

A demographic descriptive analysis was conducted with the *base* R package[108] for the clinical data. Qualitative variables were compared using Pearson's Chi-squared Test for Count Data and presented as percentages. Quantitative variables were compared with the Wilcoxon Rank Sum test and presented as the mean ± standard error of the mean (SEM). The significance level considered was p-value < 0.05 and the confidence interval (CI) was 0.95.

4.3 SINGLE NUCLEOTIDE VARIANTS AND SHORT INSERTIONS AND DELETIONS

4.3.1 Dataset generation

4.3.1.1 DNA extraction

Blood or saliva samples were obtained from each individual. DNA was extracted from whole blood utilising the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany), following the manufacturer protocol. DNA concentration and quality parameters were checked by Nanodrop (Thermo Fisher, Waltham, MA, USA) and Qubit (Invitrogen, Waltham, MA, USA) as previously described[109]. Moreover, DNA integrity was verified by electrophoresis in a 2% agarose gel. For WES the minimum parameters considered were a concentration superior to 20 ng/µL, a 260/280 ratio superior to 1.8 and no observable smearing/DNA degradation by electrophoresis.

4.3.1.2 Whole exome sequencing library preparation

The standard exome capture libraries were created with the SureSelectXT Human All Exon V6 (Agilent Technologies, Santa Clara, CA, USA), utilising 1 µg of input genomic DNA (gDNA) from each sample. The gDNA were diluted with EB Buffer and sheared to a target peak size of 150–200 bp with the Covaris LE220 focused-ultrasonicator (Covaris, Woburn, MA), following the recommendations of the manufacturer. Then, the end-repair and the addition of an "A" tail were performed. They were followed by the ligation of the Agilent adapters to the fragments. When the efficiency of the ligations was assessed, the adapter ligated product was amplified by PCR. After purifying the final product, it was quantified by TapeStation DNA screentape D1000 (Agilent). Next, 250 ng of DNA library was mixed with hybridization buffers, blocking mixes, RNase block and 5 µl of SureSelect all exon capture library, according to the standard Agilent SureSelect Target Enrichment protocol. The captured DNA was washed and amplified. At the end, final purified product was quantified by qPCR following to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and quantified using the TapeStation DNA screentape D1000 (Agilent Technologies, Santa Clara, CA, USA).

4.3.1.3 Whole exome sequencing data analysis

The binary base calling files generated by the Illumina platform, through the integrated primary analysis called RTA (real time analysis), were converted into FASTQ files format with the Illumina package bcl2fastq v2.20.0. The demultiplexing option (--barcode-mismatches) was

set to as value: 0. Macrogen, an external company, performed the WES, as it was established in the UNITI procurement.

Resulting paired-end sequences had a length of 150 bp and a coverage of 100X. They were mapped to the human reference genome GRCh38/hg38 using the Nextflow Sarek v2.7.1 workflow[110], included in nf-core[111]. The alignment was done with BWA-MEM (Burrows-Wheeler Aligner - MEM)[112] and the preprocessing following GATK (Genome Analysis Toolkit)[113] best practices, utilising the GATK functionalities: *MarkDuplicates, BaseRecalibrator* and *ApplyBQSR*. Finally, a quality control was carried out with *FastQC, samtools stats, Qualimap bamqc* and summarised with *MultiQC*. The mapping resulted in a BAM (Binary Alignment Map) file format generated for each individual.

4.3.1.4 Variant calling, filtering and annotation

Using each BAM file previously generated as input, a variant genotyping for each sample was performed with *HaplotypeCaller* of GATK. SNVs and short indels candidates are detected at nucleotide resolution in this stage. The results were saved in a VCF (Variant Call Format) file format with the variants of each subject. Each VCF was normalised using *norm* function from bcftools[114] to split multiallelic variants into different variants and to align the indels to the left.

Afterwards, two different paths were followed:

A. To continue with the files obtained in the normalisation step.
B. With the aim of reproduce gnomAD genotype filtering, each VCF file was filtered according with the criteria followed to generate the gnomAD database using bcftools: Allele balance (AB) ≥ 0.2 and AB ≤ 0.8 (for heterozygous genotypes only), genotype quality (GQ) $\geq$ 20 and depth (DP) $\geq$ 10 (5 for haploid genotypes on sex chromosomes)[115]. Analyses carried out because of this approach will be called "AB-GQ-DP".

In both cases, using the *merge* function of bcftools, a big VCF file containing the variants from all the individuals of the study, was created.

The next step was the variant filtering. It was performed using Variant Quality Score Recalibration (VQSR) method from GATK suite. This method adds a variant quality score log-odds (VQSLOD) per call in the VCF file. This tool is a sophisticated filtering technique applied

on the variant call set that uses a machine learning approach to model the technical profile of variants in a training set and test the model to filter out probable artefacts. The variants were filtered using a threshold VQSLOD < 90.

Variants included in this file were annotated using the command line tool of Ensembl Variant Effect Predictor (VEP)[116], with the gene and transcript-related fields:

- Gene ID.
- Gene symbol.
- Transcript ID.
- RefSeq ID.
- CCDS ID.
- Biotype.
- cDNA coordinates.
- CDS coordinates.
- Distance.
- Consequence type (Figure 7)[117].
- Impact, determined by each consequence (Table S9)[117].
- Exon.
- Intron.

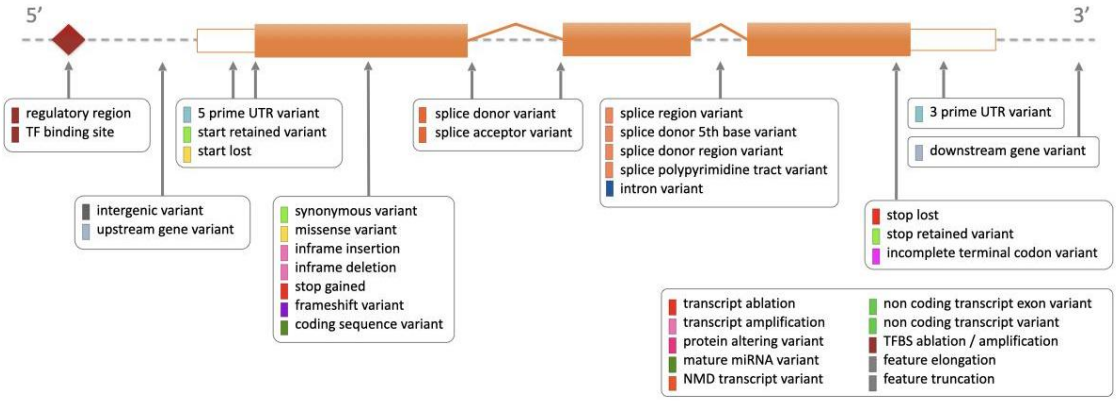

Figure 7 - Consequences terms annotated by VEP.

Diagram showing the location of each consequence term relative to the transcript structure. Image obtained from www.ensembl.org[117].

Additionally, variants were also annotated with databases of interest:

- AF, allele count (AC) and allele number (AN) of the non-Finnish European (NFE) population from gnomAD v3.0[118].
- AF, allele count (AC) and allele number (AN) of the global population from gnomAD v3.0[118].
- AF, AC and AN of the Spanish population from CSVS v3.0.1[64].
- CADD PHRED and CADD RAW scores[68].
- LOFTEE outputs: confidence prediction of being LoF (HC: high confidence, LC: low confidence), filters, flags and information[63].

The AC is the count of alternate alleles in the cohort, whereas the AN is the total number of alleles in the cohort. Therefore, the AF is the ratio between AC and AN.

To annotate with the Spanish reference population, it was necessary to ask for CSVS data, the v3.0.1 was obtained the 9th of November 2021. The study cohort has patients with MD, then the used data from CSVS was: "All except group VIII: Diseases of the ear and mastoid process". This data was in the GRCh19/hg19 reference genome, because of that the first step was to perform a liftover to GRCh38/hg38 reference genome, using the liftOver tool from UCSC[119] and the *hg19ToHg38.over.chain* file.

4.3.2 Gene burden analysis

To point out candidate genes and to demonstrate the aggregated effect of rare variants, GBA methodology was used. The GBAs were carried out with samples in the subgroups I4 and I1 concurrently from the subgroups obtained according to the THI score during the clinical characterisation, the selection of these is called the extreme phenotype approach[120]. In addition, it was done with the total of MD samples included in the whole cohort and with the I4-GÜF and I1-GÜF subgroups generated based on the GÜF score.

The first step in the GBA was to filter the variants by the effect they produce in the proteins. Seven different GBA were performed following different variant filtering criteria, then they were compared to select the best filter in order to target the candidate genes. The analyses were:

- HIGH: Variants with a high impact in the protein, according to the VEP annotation.
- HIGH HC (also abbreviated as HC): Variants with a high impact in the protein, according to the VEP annotation; and high confidence of being LoF, according to LOFTEE.

- HIGH LC + MODERATE CADD ≥ 20 (also abbreviated as LC+MOD): Variants with a high impact in the protein, according to the VEP annotation; and low confidence of being LoF, according to LOFTEE. Besides, variants with a moderate impact in the protein, according to the VEP annotation; and CADD PHRED score greater than 20, this cutoff retained those variants that are in the top 1% of the predicted deleteriousness throughout the human genome[99].
- MODERATE: Variants with a moderate impact in the protein, according to the VEP annotation.
- MODERATE CADD < 20: Variants with a moderate impact in the protein, according to the VEP annotation; and CADD PHRED score less than 20.
- LOW: Variants with a low impact in the protein according to the VEP annotation.

The second filter was by AF in the three reference populations: NFE population from gnomAD, global population in gnomAD and Spanish population from CSVS. As previously, the results were compared to choose the optimal filter, the following were studied:

- AF < 0.05: Frequency to filter out common variants[52].
- AF < 0.1: Minor prevalence of individuals suffering from tinnitus in the population[29].

After these filters, the formulas of the GBA will be applied for each reference population separately. For every variant, the AC and the AN of the studied subgroup and the reference population were used. The AC is the number of alternative alleles in the cohort and the AN is the total number of alleles in the cohort. The wild type (WT), which is the number of reference alleles in the cohort, was calculated as the difference between AN and AC.

The objective of the GBA is to study all the variants of each gene, for this purpose, the AC and the WT for each gene were added up. The WT, odds ratio (OR), the systematic error (SE) and the Z score (Z) were calculated with the following formulas[121]:

$$WT = AN - AC \quad (1)$$

$$OR = \frac{AC\ cases\ x\ WT\ control}{WT\ cases\ x\ AC\ control} \quad (2)$$

$$SE = \sqrt{\frac{1}{AC\ cases} + \frac{1}{AC\ controls} + \frac{1}{WT\ cases} + \frac{1}{WT\ controls}} \quad (3)$$

$$Z = \frac{ln(OR)}{SE} \quad (4)$$

The two-tailed *p*-value (*p*) was calculated using the complementary cumulative distribution function (CCDF), with the formula:

$$p = CCDF(|Z|)\ x\ 2 \quad (5)$$

The *p*-value was corrected by false discovery rate (FDR), using the total number of genes, with the next formula:

$$FDR = p\ x\ nGenes \quad (6)$$

The CI were calculated with:

$$CI = e^{ln(OR) \pm 1.96\ x\ SE} \quad (7)$$

Finally, the etiological fraction (EF) was obtained with the formula:

$$EF = \frac{OR - 1}{OR} \quad (8)$$

Variants without information in the reference dataset were considered as novel variants in that population. To be used in the GBA the AC in the control population was imputed as 0 and the AN was estimated for each novel variant as the mean of all the AN in the gene (without any filter) in the reference database.

The results of the GBA were filtered to select and prioritise candidate genes.

4.3.2.1 Genes prioritization

Genes significantly enriched in variants against the three reference populations were retained for further analysis. Genes without data in any variant in CSVS were also retained, due to

smaller number of controls in CSVS than in gnomAD there were many variants without information, especially in the short indels. The level of significance considered was *p*-value corrected by FDR < 0.05 and not including 1 in the CI.

The list of genes (N = 100) that most frequently present rare non-synonymous/splice-site variants in the general population, named FrequentLy mutAted GeneS (FLAGS)[122,123], were filtered in the results of the GBA. Moreover, olfactory receptor genes were filtered because they were relatively unconstrained[63].

Moreover, the number of individuals with variants in each gene was used to filter them. The threshold was established as the percentage of the total number of individuals, based on the AF filter used in the GBA. This means, 5% of the total in the GBA with the AF < 0.05 and 10% in the GBA with AF < 0.1.

To select the specific genes of each subgroup, the genes enriched in the same step of the workflow in both subgroups (I1 and I4) were filtered.

Finally, the genes were ranked by the number of individuals with variant.

4.3.2.2 Genes with novel variants

The genes with all the variants annotated as novel in the three populations were selected and studied deeper. Moreover, novel variants were filtered by being carried for two or more individuals. This was done for the variants retained after the filter by effect in the protein and AF.

4.4 COPY NUMBER VARIANTS AND STRUCTURAL VARIANTS

4.4.1 Dataset generation: Variant calling, filtering and annotation

To carry out the calling of CNVs and SVs, the BAM files generated previously for each individual were utilised.

The CNVs were inferred for each sample with CNVkit Python library and command-line software toolkit, v0.9.8[124]. The results were saved in a segmented log2 ratios (CNS) format files, exported as VCF considering the sex of the individuals. The Manta tool[125] was used to call SVs and indels. The pipeline was run using the Sarek v2.7.1 workflow[110], included in nf-core v1.14[111]. Moreover, TIDDIT was used to identify chromosomal rearrangements and report them as a SV, v2.3.1[126]. The variants obtained by the three tools were saved in VCF files, for each sample and for each tool, with the following data of interest: chromosome, position, type, length and genotype. The types of CNV are: deletions and duplications; the types of SV reported by Manta: deletions, duplications and insertions; and by TIDDIT: deletions, duplications, inversions and tandem-duplications. From CNVkit the call CNS file converted to VCF file was used, from Manta the *diploidSV* file and from TIDDIT the VCF file.

From this point, all the analyses were applied to the results of each tool (CNVkit, Manta and TIDDIT) independently, following three different paths. After the variant calling, variants were filtered by the length with bcftools, keeping those with less than 100,000 bp.

Filtered variants were annotated using AnnotSV v3.3.1[127], a command-line integrated tool for Structural Variations annotation and ranking. The VCFs were annotated with the *split* option, for each variant one annotation per gene. Among others, the following fields were annotated:

- Cytoband.
- Gene symbol.
- Transcript ID.
- Overlapped transcript length.
- Overlapped CDS length.
- Location in the gene.
- ACMG annotation (benign, likely benign, uncertain significance, likely pathogenic, pathogenic).

4.4.2 Variant prioritization

The identified variants for each individual were filtered and analysed to prioritise those related to the studied phenotype, therefore point candidate genes.

Variants were filtered by the ACMG annotation to select those pathogenic, likely pathogenic or with uncertain significance regarding the pathogenicity. Then, variants without an important effect in the protein (benign and likely benign) were discarded for further analysis. The positions of those variants were saved to filter the original VCF by them.

Until this step, the VCFs were studied separately for each sample, the next step was to merge all the files in one, with the aim of analysing all individuals of the subgroup together. In contrast to SNVs and indels, the CNVs and SVs could not be merged by a single position as they covered larger regions. Because of that, regions with an overlap of at least 60% in two or more samples of the same subgroup - I1 and I4, separately - were identified. The overlapping was analysed using SVDB v2.8.1[128], a toolkit for merging SV VCF files from multiple callers or individuals. As a result, the merge VCF - for both subgroups - contained the regions with overlapping variants and the samples with variants in each region. The VCFs were annotated using AnnotSV (as previously).

Finally, regions covering genes in the FLAGS list and the olfactory receptor genes were filtered. Moreover, only regions with variants in at least two individuals were maintained for further analysis. Besides, genes with variants in both subgroups (I1 and I4) were filtered at this step. As in the SNVs and short indels analysis, common genes were discarded to select specific genes for each subgroup.

4.5 VARIANTS IN CANDIDATE GENES

After the different pipelines, the candidate genes for both I4 and I1 subgroups were selected. The selected genes were retrieved for the following analyses:

- Genes enriched and prioritised in the GBA with HIGH HC variants.
- Genes enriched and prioritised in the GBA with HIGH LC + MODERATE CADD > 20 variants.
- Genes sharing SNV and SV in the same THI subgroup.

The variants in the candidate genes were validated using their BAM files by the Integrative Genomics Viewer (IGV), a high-performance interactive tool for the visual exploration of genomic data[129]. Only those in which it was possible to check the variant in the alignment were used for further analyses. Therefore, only the genes with validated variants were selected as a candidate for the phenotype.

To study the pathogenicity of some of the candidate variants, in addition to the annotated CADD score, their pathogenicity was predicted following the ACMG criteria[69,70].

The constraint was evaluated for the candidate genes to measure the tolerance to variation; it was obtained for each gene from the gnomAD database v.2.1.1[63]. Different scores were used for each type of variant[130].

For the LoF variants were used the following scores:

- Loss-of-function observed/expected upper bound fraction (LOEUF): For each gene, it is the upper bound of the 90% CI for the ratio between observed LoF variants and expected LoF variants predicted by the mutational model from gnomAD. Lower values indicate high constraint; for Mendelian disorders, to consider a gene constraint gnomAD suggests: LOEUF < 0.35. It is preferred over pLI because LOEUF considers the sample size.
- Probability of being loss-of-function intolerant (pLI): Greater values represent more constraint; to assume that a gene is constrained in Mendelian cases, gnomAD recommends: pLI > 0.9.

For the missense variants:

- Missense observed/expected fraction (missense o/e): It is the ratio between the observed and the expected missense variants in the gene, the expected is calculated based on the gnomAD mutational model. Then, the closer values are to zero, the more conserved the gene will be.
- Z score: Measures the deviation of observed counts from the expected number. Positive scores indicate increased constraint, the higher value, the less tolerance to variations.

Described variants in the reference databases overlapping the whole or a part of the CNV and SV from the candidate genes were searched. For the Spanish population SPACNACS[65] was used and for the global population gnomAD SVs v2.1[63]. As the human reference genome was GRCh19/hg19 in both databases, it was necessary to perform a liftover using liftOver tool from UCSC[119] in the variants. From the found variants overlapping the positions of the candidate regions - in addition to the genomic position - the length, type, frequency, number and population of the carrier individuals were retrieved.

4.6 EXPRESSION OF CANDIDATE GENES IN PUBLIC DATASETS

To analyse the implication of the candidate genes in the tissues of interest for the tinnitus phenotype, the expression of them from the following expression databases was retrieved:

- RNA-Seq in human brain tissues was retrieved: Amygdala, Anterior cingulate cortex (BA24), Caudate (basal ganglia), Cerebellar Hemisphere, Cerebellum, Cortex, Frontal Cortex (BA9), Hippocampus, Hypothalamus, Nucleus accumbens (basal ganglia), Putamen (basal ganglia), Spiral cord (cervical c-1) and Substantia nigra. The gene Transcripts per million (TPMs) were obtained from the Genotype-Tissue Expression (GTEx) project V8[131].
- RNA-Seq in postnatal day 0 (P0) mouse hair cells and non-hair cells from the cochlea[132], from gene Expression Analysis Resource (gEAR) portal (https://umgear.org). The Reads per kilobase of transcript (RPKMs) from the orthologous genes were extracted.
- RNA expression by microarray in P0 mouse SGN[133], from Shared Harvard Inner-Ear Laboratory Database (SHIELD, https://shield.hms.harvard.edu). The expression was obtained from the orthologous of the candidate genes.

The data of each dataset was normalised between 0 and 100, being 100 the maximum of each database. The data was also normalised for the I1 candidate genes between 0 and 100, but the 100 was the maximum of the I4 candidate genes for each database, to compare them easily. Then, the three databases were merged by gene name. The data was represented in a heatmap, where the values were scaled using the base-10 logarithm. The genes - the rows in the heatmap - were clustered based on the Euclidean distance.

4.7 FUNCTIONAL ENRICHMENT ANALYSIS OF CANDIDATE GENES USING PUBLIC DATABASES

Functional enrichment analyses were carried out to discover terms, phenotypes and pathways related to the candidate genes. The enrichment analyses were performed via an hypergeometric test utilising GeneCodis4 tool[134]. The genes universe was customised using all the genes with variants in the I4 or I1 individuals - depending on the input list of candidate genes - that pass the filters of the GBAs used:

- High impact in the protein and HC of being LoF, high impact in the protein and LC of being LoF and moderate impact in the protein and CADD score equal or greater than 20.
- AF less than 0.05 in the three reference population databases.

Moreover, the following databases were used:

- Biological Process (BP) from Gene Ontology (GO)[135,136].
- Human Phenotype Ontology (HPO)[137].
- Mouse Genome Informatics (MGI)[138].

From every term retrieved as a result of the different analysis the following information was obtained:

- Name.
- *p*-value.
- Number of genes.
- Relative enrichment: It measures the proportion between the number of genes found of an annotation in the input divided by the size of the input and how many genes are under that annotation divided by all the genes in the database.

4.8 VISUALISATIONS

The visualisations were conducted using R v4.2.1[108] and following R packages were used for the visualisations of the study cohort description: ggplot2 v3.4.1[139], GGally v2.1.2[140], ggalluvial v0.12.4[141], ComplexHeatmap v2.12.1[142], ComplexUpset v1.3.3[143], circlize v0.4.15[144], dplyr v1.0.9[145], tidyverse v1.3.2[146], stringr v1.4.0[147] and forcats v0.5.1[148].

5 RESULTS

5.1 SYSTEMATIC REVIEW OF THE GENETIC ARCHITECTURE OF FAMILIAL MENIERE DISEASE

The main goal in this thesis was to identify the principal genes and biological processes associated with severe tinnitus. As all the individuals in the cohort used in this work were diagnosed with MD, it was necessary to define the genetic background of MD. For this, a systematic review to identifying the rare variants or genes related to FMD was performed. Familial cases were studied, instead of sporadic, to enhance the robustness of the study of the genetic architecture of MD.

5.1.1 Selection and characteristics of familial Meniere Disease studies

A total of 191 articles were identified from PubMed, they were screened to exclude reviews, meta-analysis, non-genetic and animal studies. Then, the eligibility criteria of the 64 remaining articles were evaluated. Finally, eight of them were used in the qualitative synthesis (Figure 8).

One of the eligibility criteria was that all the reported cases in the families were diagnosed as FMD. Although it was desirable they were diagnosed following the diagnostic criteria described by the International Classification Committee for Vestibular Disorders of the Barany Society in 2015 (Table 1)[18], only three of the studies applied these criteria. Another three works followed the criteria defined earlier proposed by the Committee on Hearing and Equilibrium of the AAO-HNS published in 1995 (Table S10)[149]. Besides, two articles did not clarify the diagnostic criteria.

Moreover, to select the article, it was mandatory they identify at least one SNV related to individuals with FMD. Of the eight articles analysed, five reported candidate variants in one gene each of them - which were *PRKCB, COCH, STRC, OTOG* and *LSAMP* - in cases from Spain, South Korea, Sweden and Iran. Three studies confirmed candidate variants in two different genes: in the same Spanish family was found a variant in *FAM136A* and another in *DTNA*, another article described two variants in *DPT* and *SEMA3D* in two unrelated families, and variants in *HMX2* and *TMEM55B* were reported in the same Finnish family.

Regarding the eligibility criteria, in all the articles, variants were found by WES and/or Sanger sequencing. Besides, they reported information about the cases' population ancestry, sex and number of affected individuals (Table 7).

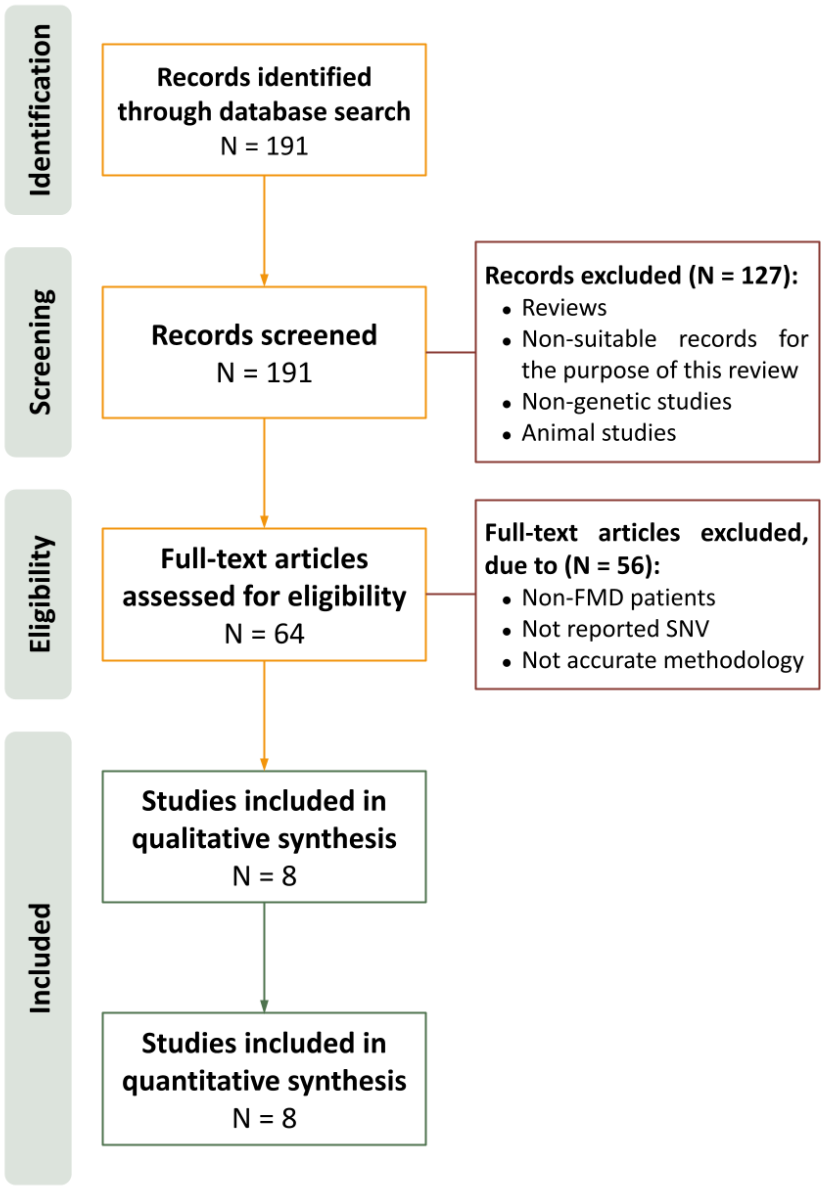

Figure 8 - Flowchart to select familial Meniere Disease (FMD) studies following the guidelines for systematic reviews.

SNV: Single nucleotide variants.

Table 7 - Summary of studies describing single nucleotide variants (SNVs) selected for quantitative synthesis.

Reference	Population	Patients	Sex	Diagnosis	Genetic findings	
					Gene	Variant
150	Spanish	3	F	AAO-HNS	*FAM136A*	chr2:70527974G>A
					DTNA	chr18:32462094G>T
151	Spanish	2	M	AAO-HNS	*PRKCB*	chr16:23999898G>T
152	Korean	3	F, M	Barany Society	*COCH*	chr14:31349796G>A
153	Spanish	3	F	Barany Society	*DPT*	chr1:168665849G>A
		3	F, M		*SEMA3D*	chr7:84642128G>A
154	Swedish-Norwegian	3	M	SNHL and episodic vertigo	*STRC*	chr15:43896948G>A
155	Finnish	2	M	AAO-HNS	*HMX2*	chr10:124909634T>A
					TMEM55B	chr14:20927370G>A
156	Spanish	73	F, M	Barany Society	*OTOG*	chr11:17574758G>A
						chr11:17578774G>A
						chr11:17594747C>A
						chr11:17621218C>T
						chr11:17627548G>A
						chr11:17631453C>T
						chr11:17632921C>T
						chr11:17656672G>A
						chr11:17663747G>A
						chr11:17667139G>C
157	Iranian	2	F, M	Definite MD	*LSAMP*	chr3:115561402T>C

F: Female; M: Male; AAO-HNS: American Academy of Otolaryngology–Head and Neck Surgery; SNHL: Sensorineural hearing loss; MD: Meniere Disease.

5.1.2 Inheritance of single nucleotide variant associated with familial Meniere Disease

Two heterozygous rare variants were identified in *FAM136A* and *DTNA* genes. Both variants were pathogenic, the stop gained variant chr2:70527974G>A in *FAM136A* was novel and the

splice gained variant chr18:32462094G>T in *DTNA* was ultrarare. They were found in three women in consecutive generations of the same family with definite MD, which proposed an AD inheritance pattern (Table 8).

In the *PRKCB* gene was reported the missense variant chr16:23999898G>T, which was novel and likely pathogenic. It was present in three relatives - two diagnosed with complete MD and one with SNHL, incomplete phenotype - therefore, it suggested an AD pattern of inheritance with incomplete penetrance.

The chr14:31349796G>A variant was found in the *COCH* gene, it was classified as likely pathogenic and described as novel because it had not been previously depicted in gnomAD nor ClinVar. It was identified in a South Korean family with DFNA9 phenotype, where the mother and two siblings exhibited symptoms consistent with definite MD and another two siblings showed an incomplete phenotype. The family met the criteria to be considered as FMD and it was proposed an AD inheritance pattern with incomplete penetrance.

In the *SEMA3D* gene was discovered the chr7:84642128G>A variant, it was missense and pathogenic. It segregates in three MD cases in the same generation of the family; besides they had another three relatives with incomplete phenotype in different generations; then, an AD inheritance with incomplete penetrance was suggested. The variant chr1:168665849G>A in the *DPT* gene was described in the same study. This missense and likely pathogenic variant was found in a family with three sisters diagnosed with definite MD and another seven relatives with incomplete phenotype (SNHL or episodic vertigo). The variant segregated in the three individuals with MD and two with SNHL, proposing an AD inheritance pattern with incomplete penetrance.

The *STRC* gene contained two variants: the stop gained chr15:43896948G>A variant and a deletion of approximately 97 kb. They were found in a Swedish-Norwegian family with two brothers and a first cousin with moderate, non-progressive SNHL and episodic vertigo. Both brothers presented the nonsense variant in homozygosis, whereas the cousin had the same variants in heterozygosis inherited from the mother and the deletion from the father. None of the parents showed any symptoms concordant with MD, proposing an AR inheritance pattern.

The missense variants chr10:124909634T>A and chr14:20927370G>A were described in the *HMX2* and *TMEM55B* genes, respectively; being classified as likely pathogenic and uncertain significance. Both were identified in an individual and his grandfather in heterozygosis. It was impossible to obtain a DNA sample from the father of the proband, which would be an obligated carrier of the variants, however he did not report any symptoms of MD. Since these two variants alone did not lead to MD with full penetrance, to confirm the AR inheritance pattern in this family, additional heterozygous variants should be necessary.

In the same article, ten missense variants were reported in the *OTOG* gene in various Spanish families with FMD. The rare variants chr11:17574758G>A (pathogenic) and chr11:17663747G>A were identified - both in heterozygosis - in two cases from two unrelated families and one of them also had the chr11:17627548G>A variant. Furthermore, the rare variants chr11:17578774G>A and chr11:17632921C>T were found - also in heterozygosis - in another four cases from another four unrelated families. Based on these data, a heterozygous compound recessive inheritance pattern existed in both sets of families. The remaining variants were identified in one, two or three unrelated patients with FMD.

Finally, in the *LSAMP* gene the novel likely pathogenic variant chr3:115561402T>C was found. It was reported in homozygosis in two affected siblings with definite MD; in addition, they had poor senses of smell, which suggested a MD-like phenotype. The parents were consanguineous and the other two siblings were unaffected, which described an AR inheritance pattern.

Table 8 - Genetic findings for each single nucleotide variant (SNV), pathogenicity and inheritance pattern.

Gene symbol	ID	Position	Protein	Variant effect	AF[1]		ACMG classification	CADD	Inheritance pattern
					gnomAD	Other			
FAM136A	rs690016537	chr2:70527974C>T	Q76*	Nonsense	Novel		Pathogenic (PS3, PS4, PM2, PM4, PP3)	41.00	AD
DTNA	rs533568822	chr18:32462094G>T	V715F	Missense	8.79E-06	NF (CSVS)	Pathogenic (PS3, PS4, BP1)	24.90	AD
PRKCB	rs1131692056	chr16:23999898G>T	G92V	Missense	Novel		Likely Pathogenic (PS4, PM2, PP3, PP5)	28.20	AD[2]
COCH	-	chr14:31349796	-	-	Novel		Likely pathogenic (PS4, PM2, PP2, PP3, PP5)	28.10	AD[2]
DPT	rs748718975	chr1:168665849C>T	R182C	Missense	1.72E-05	NF (CSVS)	Likely Pathogenic (PS4, PM1, PP3, PP5, BP1)	32.00	AD[2]
SEMA3D	rs1057519374	chr7:84642128C>T	P580S	Missense	Novel		Pathogenic (PS4, PM1, PM2, PP3, PP5)	25.00	AD[2]
STRC	rs144948296	chr15:43896948C>T	Q1343*	Nonsense	3.62E-04	0.001 (SweGen)	Pathogenic (PSV1, PS4, PM2, PP3, PP5)	40.00	AR
HMX2	rs1274867386	chr10:124909634T>A	Y273N	Missense	Novel		Likely Pathogenic (PS4, PM2, PP3)	31.00	AR[3]
TMEM55B	rs201529818	chr14:20927370C>T	L229F	Missense	9.56E-04	8.20E-05 (ExAC)	Uncertain Significance (PS4, PP3, BS1)	25.80	AR[3]
OTOG	rs552304627	chr11:17574758G>A	V141M	Missense	1.29E-03	4.00E-03 (CSVS)	Pathogenic (PVS1, PS4, PM2, PP3, BP1)	33.00	AR[3]
	rs61978648	chr11:17578774G>A	V269I	Missense	4.44E-03	1.40E-02 (CSVS)	Likely Benign (PS4, BP1, BP4, BP6)	19.12	AR[3]

Table 8 - Continuation.

Gene symbol	*ID*	*Position*	*Protein*	*Variant effect*	*AF[1]*		*ACMG classification*	*CADD*	*Inheritance pattern*
					gnomAD	*Other*			
OTOG	-	chr11:17594747	P747T	Missense	Novel		Uncertain Significance (PS4, PM2, BP1, BP4)	21.90	-
	rs117005078	chr11:17621218C>T	P1240L	Missense	5.74E-03	4.00E-03 (CSVS)	Likely Pathogenic (PS4, PM2, PP3, BP1)	33.00	-
	rs145689709	chr11:17627548G>A	R1353Q	Missense	4.04E-03	6.00E-03 (CSVS)	Uncertain Significance (PS4, PM2, BP1, BP4, BP6)	22.00	AR[3]
	rs117380920	chr11:17631453C>T	L1548F	Missense	1.24E-02	1.30E-02 (CSVS)	Benign (PS4, BS1, BS2, BP1, BP4, BP6)	12.42	-
	rs61736002	chr11:17632921C>T	A2037V	Missense	1.21E-03	4.00E-03 (CSVS)	Uncertain Significance (PS4, PM2, BP1, BP4)	7.61	AR[3]
	rs76461792	chr11:17656672G>A	R2556Q	Missense	4.67E-03	4.00E-03 (CSVS)	Benign (PS4, BS1, BS2, BP1, BP4, BP6)	23.50	-
	rs117315845	chr11:17663747G>A	R2802H	Missense	2.73E-03	6.00E-03 (CSVS)	Uncertain Significance (PS4, PM2, BP1, BP4, BP6)	16.79	AR[3]
	rs61997203	chr11:17667139G>C	L2842N	Missense	2.34E-02	1.90E-02 (CSVS)	Benign (PS4, BS1. BS2, BP1, BP6)	24.20	-
LSAMP	-	chr3:115561402	-	-	Novel		Likely Pathogenic (PS4, PM2)	25.90	AR

*ID: Reference single nucleotide polymorphism identifier; *: Stop codon; NF: Not found; [1]: Allelic frequencies reported in the original reports have been updated according to the available information in the last version of the reference database; gnomAD: Genome Aggregation Database; CSVS: Collaborative Spanish Variant Server; ExAC: Exome Aggregation Consortium; ACMG: American College of Medical Genetics and Genomics; CADD: Combined Annotation Dependent Depletion; AD: Autosomal dominant inheritance pattern; AR: Autosomal recessive inheritance pattern; [2]: Incomplete penetrance; [3]: Multiple inheritance.*

5.2 STUDY COHORT

5.2.1 Demographics

The MD individuals were clustered into four subgroups according to their THI score, to create groups of patients with a homogeneous annoyance related to tinnitus. The distribution of the THI score and the four subgroups is shown in Figure 9. The largest subgroup is the I1 (THI score ≤ quartile 1) with 88 patients, having 23 participants a THI score equal to 0, which means that those MD do not suffer from tinnitus. The most interesting subgroup in this study is the I4, which includes patients with severe or catastrophic tinnitus (THI score > quartile 3), with 75 individuals.

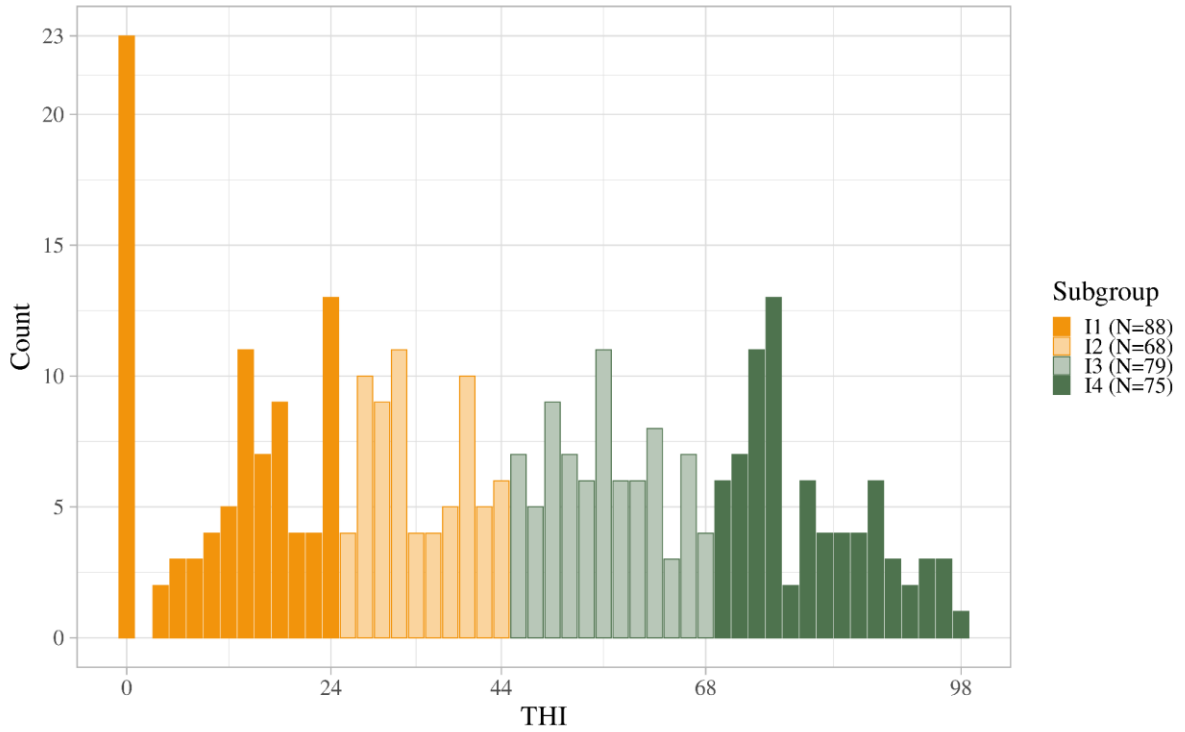

Figure 9 - Distribution of the Tinnitus Handicap Inventory (THI) in the study cohort.

The minimum value, quantiles and maximum value are defined in the x-axis; they allow to subgroup the sample in four intervals: I1, I2, I3 and I4.

A general overview of the demographics and clinical history of the I1 and I4 subgroups classified by THI is presented in Table 9. Significant differences between subgroups were found in the sex, where the percentage of females in I4 was bigger; GÜF, PHQ-9, HADS-A, HADS-D, VAS and PTA, all with higher scores in the I4 subgroup. As expected, the difference in the THI was highly significant; however, age was not different between patients in I1 and I4.

Table 9 - Clinical and demographic variables assessed in I1 and I4 subgroups, classified according to the Tinnitus Handicap Inventory (THI).

	THI		
	I1 (N = 88)	**I4 (N = 75)**	***p*-value**
Sex (% female)	64.29	52.94	0.44
Age (mean±SEM)	63.21±0.88	60.94±0.64	0.40
FMD (% yes)	21.43	29.41	0.59
Tinnitus laterality (% bilateral)	22.86	33.33	0.51
Tinnitus onset (mean±SEM)	51.38±0.79	45.77±0.94	0.08
THI (mean±SEM)	28.9±1.44	60.76±1.5	**1.61E-07**
GÜF (mean±SEM)	6.64±0.21	29.12±0.31	**7.99E-14**
PHQ-9 (mean±SEM)	5.32±0.37	13.32±0.48	**1.29E-06**
HADS-A (mean±SEM)	5.19±0.23	10.59±0.3	**2.21E-06**
HADS-D (mean±SEM)	5.47±0.25	10.03±0.28	**5.15E-05**
VAS (mean±SEM)	5±0.16	7.29±0.13	**1.71E-04**
PTA (mean±SEM)	38.45±0.9	54.97±1.24	**7.60E-05**
4 kHz (mean±SEM)	51.25±1.41	62.58±1.31	**1.60E-02**
8 kHz (mean±SEM)	60.65±1.53	83.41±6	0.07

Significant differences for p-value < 0.05 are highlighted in bold. MD: Meniere Disease; GÜF: Questionnaire of hypersensitivity to sound; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale; VAS: Visual analogue scale, PTA: Pure-tone audiometry; Hz: Hertz.

Furthermore, individuals were classified by the GÜF test, to determine those with a homogenous hyperacusis. In Figure 10 is presented the GÜF score distribution and the subgroups obtained. In this case, the I1 (N = 42) is also the bigger one; and the I4 (N = 34) is composed by those patients with a severe or very severe handicap due to the hyperacusis.

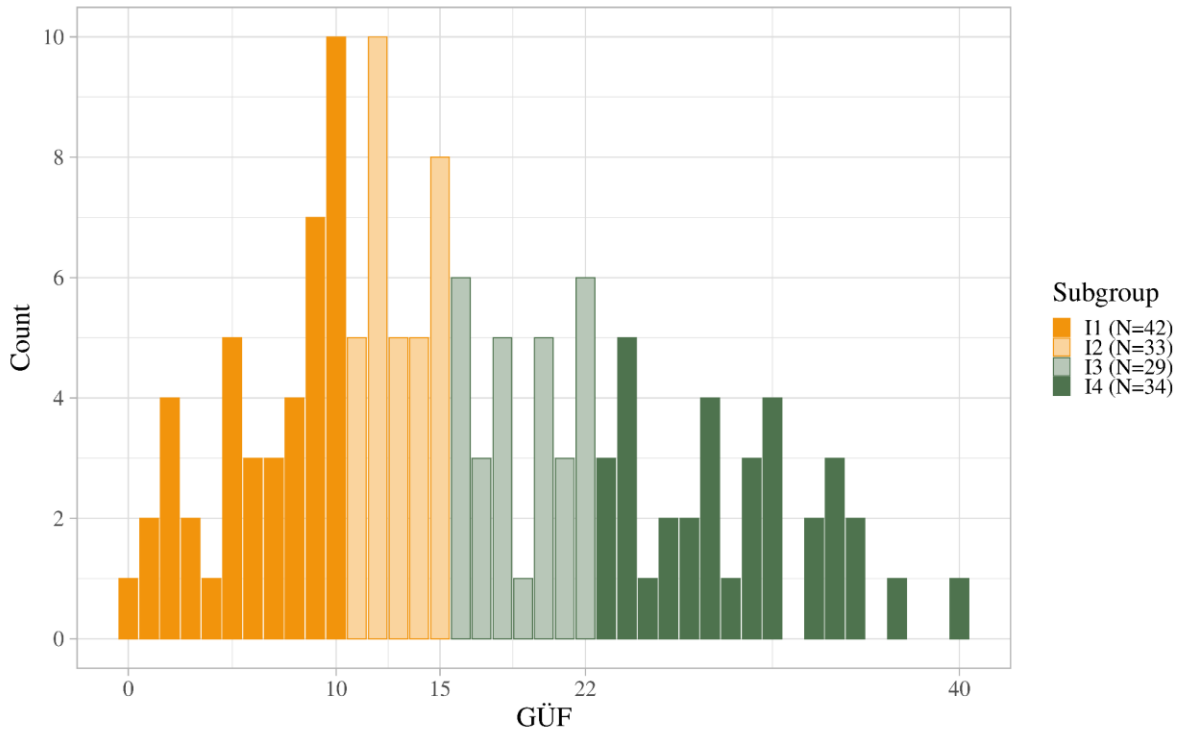

Figure 10 - Distribution of the questionnaire of hypersensitivity to sound (GÜF test) score in the study cohort.

The minimum value, quantiles and maximum value are defined in the x-axis; they allow to subgroup the sample in for intervals: I1, I2, I3 and I4.

A general overview of the demographics and clinical history of the I1 and I4 subgroups according to GÜF score is presented in Table 10. Significant differences between subgroups were found in the THI, PHQ-9, HADS-A, HADS-D, VAS, PTA and 4 kHz, all of them with higher scores in the I4 subgroup. As expected, the difference in the GÜF was highly significant.

Table 10 - Clinical and demographic variables assessed in I1 and I4 subgroups, classified according to the hypersensitivity to sound (GÜF test).

	GÜF		
	I1 (N = 88)	**I4 (N = 75)**	***p*-value**
Sex (% female)	64.29	52.94	0.44
Age (mean±SEM)	63.21±0.88	60.94±0.64	0.40
FMD (% yes)	21.43	29.41	0.59
Tinnitus laterality (% bilateral)	22.86	33.33	0.51
Tinnitus onset (mean±SEM)	51.38±0.79	45.77±0.94	0.08
THI (mean±SEM)	28.9±1.44	60.76±1.5	**1.61E-07**
GÜF (mean±SEM)	6.64±0.21	29.12±0.31	**7.99E-14**
PHQ-9 (mean±SEM)	5.32±0.37	13.32±0.48	**1.29E-06**
HADS-A (mean±SEM)	5.19±0.23	10.59±0.3	**2.21E-06**
HADS-D (mean±SEM)	5.47±0.25	10.03±0.28	**5.15E-05**
VAS (mean±SEM)	5±0.16	7.29±0.13	**1.71E-04**
PTA (mean±SEM)	38.45±0.9	54.97±1.24	**7.60E-05**
4 kHz (mean±SEM)	51.25±1.41	62.58±1.31	**1.60E-02**
8 kHz (mean±SEM)	60.65±1.53	83.41±6	0.07

Significant differences for p-value < 0.05 are highlighted in bold. MD: Meniere Disease; THI: Tinnitus Handicap Inventory; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale; VAS: Visual analogue scale; PTA: Pure-tone audiometry; Hz: Hertz.

Likewise, the VAS of annoyance for tinnitus; and the PHQ-9, HADS-A and HADS-S questionnaires, were analysed. In all cases, the subgroup composed by the large number of individuals is the I1, presented in Figure 11.

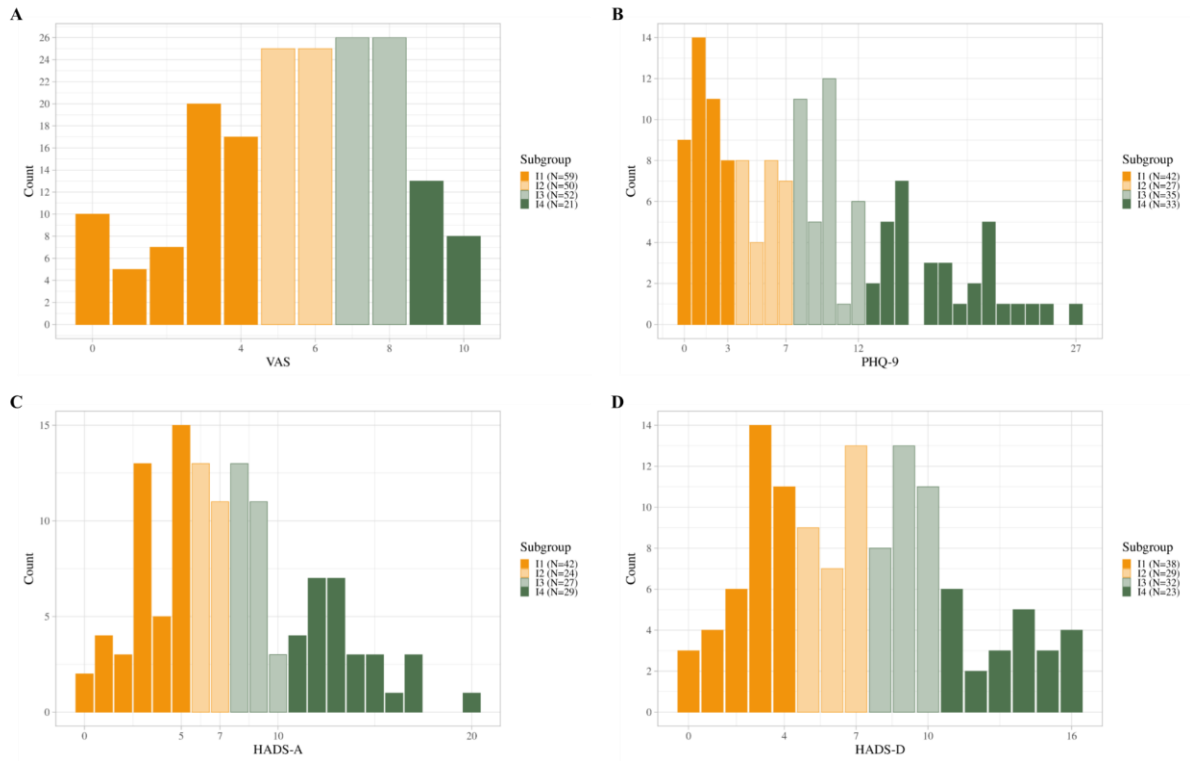

Figure 11 - Distribution of clinical variables in the study cohort.

A: the visual analogue scale (VAS) of annoyance for tinnitus; B: Patient Health Questionnaire depression scale (PHQ-9) score; C: Hospital Anxiety and Depression Scale (HADS)-A score; D: HADS-D score. The minimum value, quantiles and maximum value are defined in the x-axis; they allow to subgroup the sample in four intervals: I1, I2, I3 and I4.

5.2.2 Comparisons between variables

The correlation, density and frequency distribution between all the questionnaire scores (THI, GÜF, PHQ-9, HADS-A and HADS-D) and the VAS test were carried out (Figure 12).

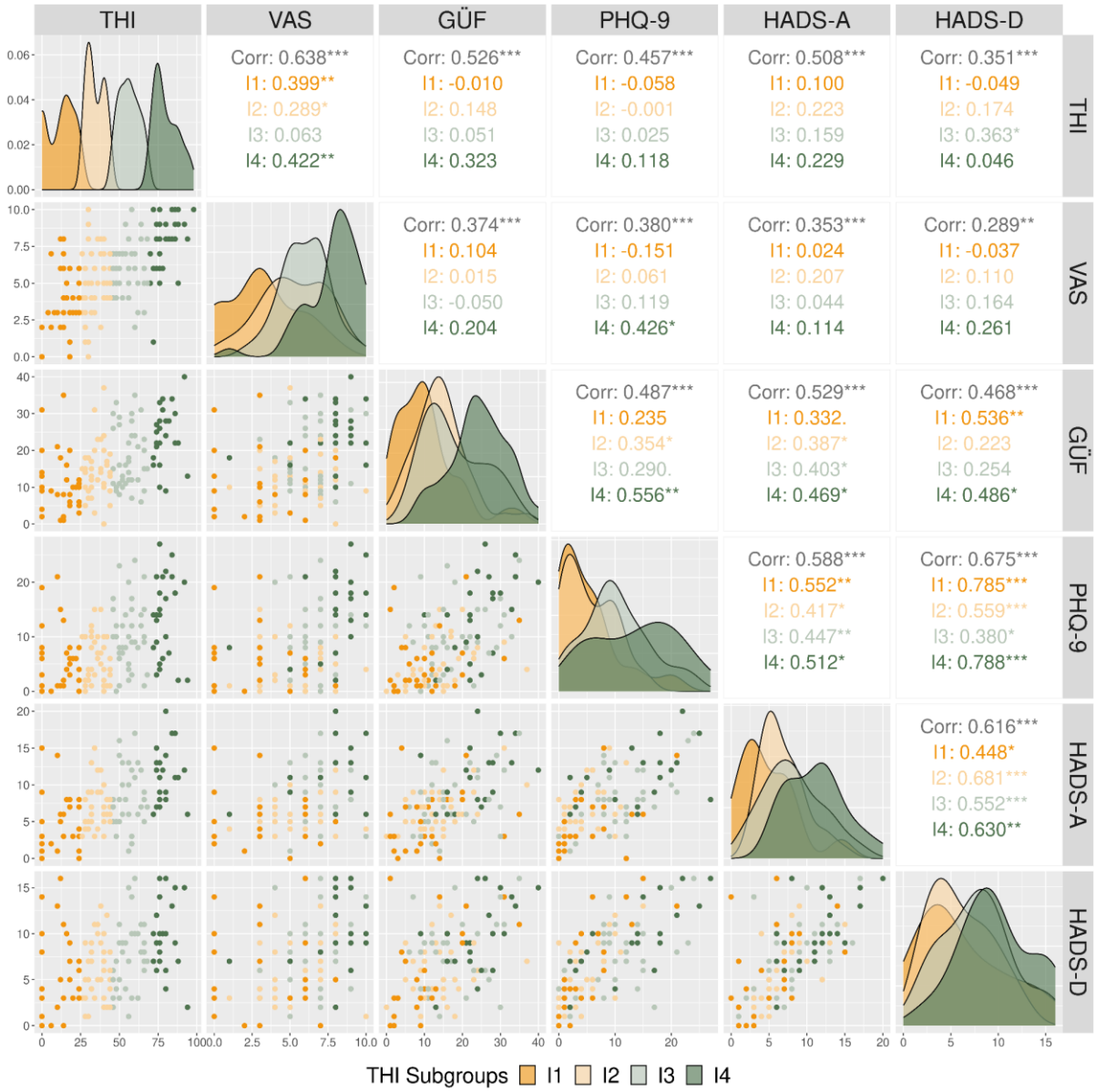

Figure 12 - Correlation, density and frequency distribution of clinical variables in the study cohort.

*Densities of each questionnaire (diagonal), correlations (upper part) and distributions (lower part) between the whole group of individuals and each subgroup defined by THI separately: I1, I2, I3, I4. THI: Tinnitus Handicap Inventory; VAS: visual analogue scale of annoyance for tinnitus; GÜF: hypersensitivity to sound; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale. The statistical significance of p-values is presented as "***": p-value < 0.001; "**": p-value < 0.01; "*": p-value < 0.05; ".": p-value < 0.10; "": otherwise.*

As Figure 12 presents, all the correlations between all the variables were statistically significant ($p < 0.01$) when comparing all the individuals in the sample as a whole; however, this does not happen when the correlations are performed by THI-subgroups separately. Focusing on THI, the highest correlation is with VAS (r = 0.638, $p < 0.001$), followed by GÜF (r = 0.526, $p < 0.001$). Moreover, as expected, the anxiety and depression questionnaires (PHQ-9, HADS-A and HADS-D) were strongly correlated, also when comparing the subgroups. The THI questionnaire and the VAS test measure the annoyance for tinnitus, its correlation was predicted. Nevertheless, THI and GÜF evaluate different symptoms - tinnitus and hyperacusis, respectively. This correlation confirmed the relation between these two symptoms in the studied cohort. In the density and scatter plots (Figure 12), it is observable that individuals in the I4 classified by THI, are more homogeneous regarding the GÜF score than the rest of them.

From the total of 310 MD patients studied in this cohort, it was only possible to obtain the result of all the questionnaires in 121 cases. Among them, seven individuals were classified in the I1 in the five tests, whereas five of them in the I4. Furthermore, regarding the previous correlation between THI and GÜF, it was interesting to study those individuals classified in the I1 and I4 subgroups in both questionnaires. A total of 21 MD cases were classified in the I1 and 15 in the I4, from this point the subgroups were termed I1-THI+GÜF and I4-THI+GÜF, respectively (Figure 13).

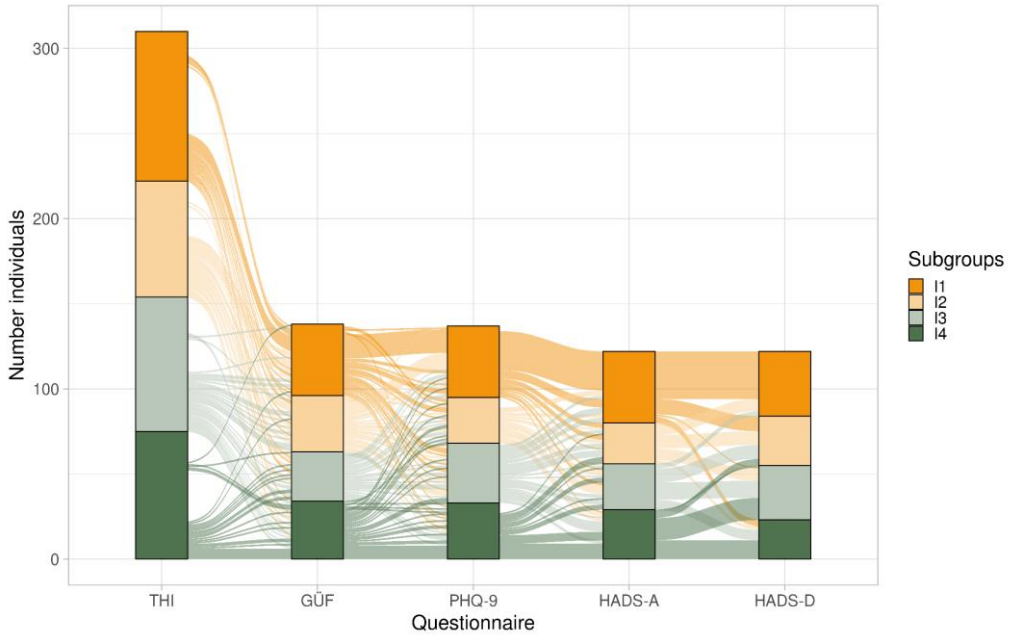

Figure 13 - Alluvium plot of the questionnaires from the study cohort.

THI: Tinnitus Handicap Inventory; GÜF: hypersensitivity to sound; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale.

5.3 SINGLE NUCLEOTIDE VARIANTS AND SHORT INSERTIONS AND DELETIONS ANALYSIS

5.3.1 Exome sequencing dataset

The first goal in this thesis was to generate a database containing Exome Sequencing data from 429 MD patients. The workflow summarising the main steps is presented in Figure 14.

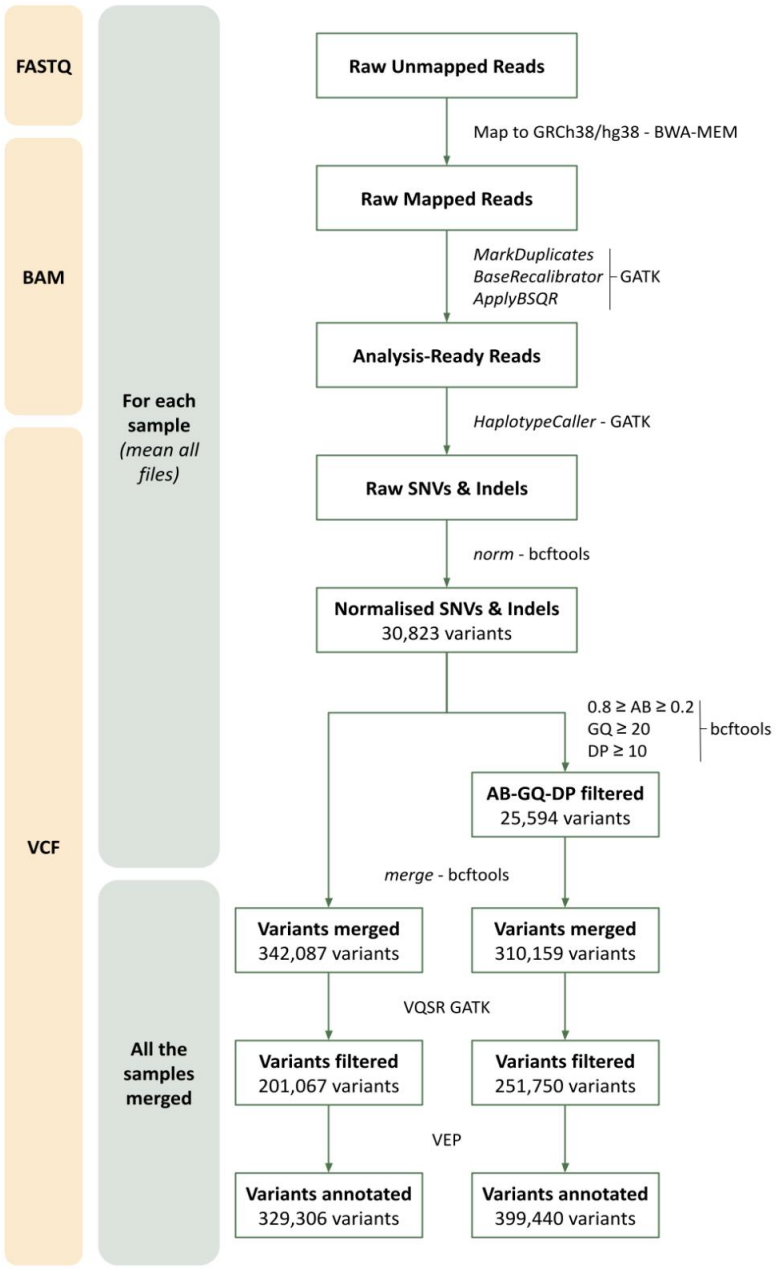

Figure 14 - Workflow to generate the exome sequencing dataset containing single nucleotide variants (SNVs) and insertions and deletions (Indels).

The diagram shows each step followed to generate the allelic variant dataset. From normalisation two different paths were followed: filtering only by VQSR (Variant Quality Score Recalibration) and filtering by AB-GQ-DP and VQSR. BAM: Binary Alignment Map; VCF: Variant Call Format; BWA-MEM: Burrows-Wheeler Aligner - MEM; GATK: Genome Analysis Toolkit; AB: Allele balance; GQ: Genotype quality; DP: Depth; VEP: Variant Effect Predictor.

After the alignment to the human reference genome GRCh38/hg38, variant calling and normalisation for each sample, a mean of 30,823 variants - SNVs and short indels - were obtained in each VCF file. From this step, two different workflows were followed to filter the variants by quality. In the AB-GQ-DP path, each VCF file was strictly filtered by AB, GQ and DP, obtaining a mean of 25,594 variants per sample. After merging all samples together with the AB-GQ-DP filters, the total number of variants was 310,159, compared to the 342,087 variants retrieved without any previous filter. Nevertheless, when both merge files were filtered by quality, using VQSR, the number of variants in the AB-GQ-DP pipeline was higher (N = 251,750) than in the other path (N = 201,067). In some cases, variants were annotated with more than one annotation per variant. As a result, 399,440 annotations were collected for further analysis in the AB-GQ-DP path; and 329,306 annotations in the path without previous filters. Considering these results, the following analyses were performed with the variants retrieved from the AB-GQ-DP path.

5.3.2 Selection of filters for gene burden analysis

Before performing the GBA for each subgroup, defining the criteria to filter the analysed variants was necessary. To study homogenous groups, variants were classified according to the impact that they have in the protein, the confidence of being LoF by LOFTEE (for the HIGH impact variants) and the predicted deleteriousness by CADD PHRED score (for the MODERATE impact variants). The impact was chosen, instead of the consequence, because the impact encompasses different consequences with a similar effect in the protein, moreover the impact includes both SNVs and indels.

Taking these proposals into account, seven different GBAs were considered. The results of the GBAs were compared to select the best GBA to target candidate genes for tinnitus; for this, the subgroups I4 and I1 obtained according to the THI score were used (Figure 15).

Based on these results, filtering the variants with a HIGH impact in the protein by a high confidence of being LoF by LOFTEE allowed working with a more homogenous group of variants. However, filtering by a low confidence of being LoF generated a group with a small number of variants. Besides, comparing the results of the variants with a MODERATE impact in the protein with those filtered by CADD score ≥ 20, the second one was a more homogeneous group. Joining the variants with a HIGH impact in the protein with a low confidence of being LoF and the variants with a MODERATE impact in the protein with a CADD score ≥ 20, it

was possible to obtain a group of variants with a similar effect in the protein and with better results than studying each subgroup separately. Therefore, the elected filters for further analyses were:

- HIGH HC: Variants with a predicted high confidence of being LoF.
- HIGH LC + MODERATE CADD ≥ 20: Variants with a predicted low confidence of being LoF, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious.

Nevertheless, the rest of the results (HIGH, MODERATE and LOW) were considered in specific instances.

Moreover, the AF in the reference populations to filter the variants were discussed. The number of genes and variants resulting from each GBA, with the I4 and I1 subgroups, using the AF of 0.1 (Figure 16) were compared with the previous analysed with an AF lower than 0.05 (Figure 15).

Considering these findings, the cutoff of AF was established in 0.05 to obtain a reliable number of genes potentially candidate for tinnitus.

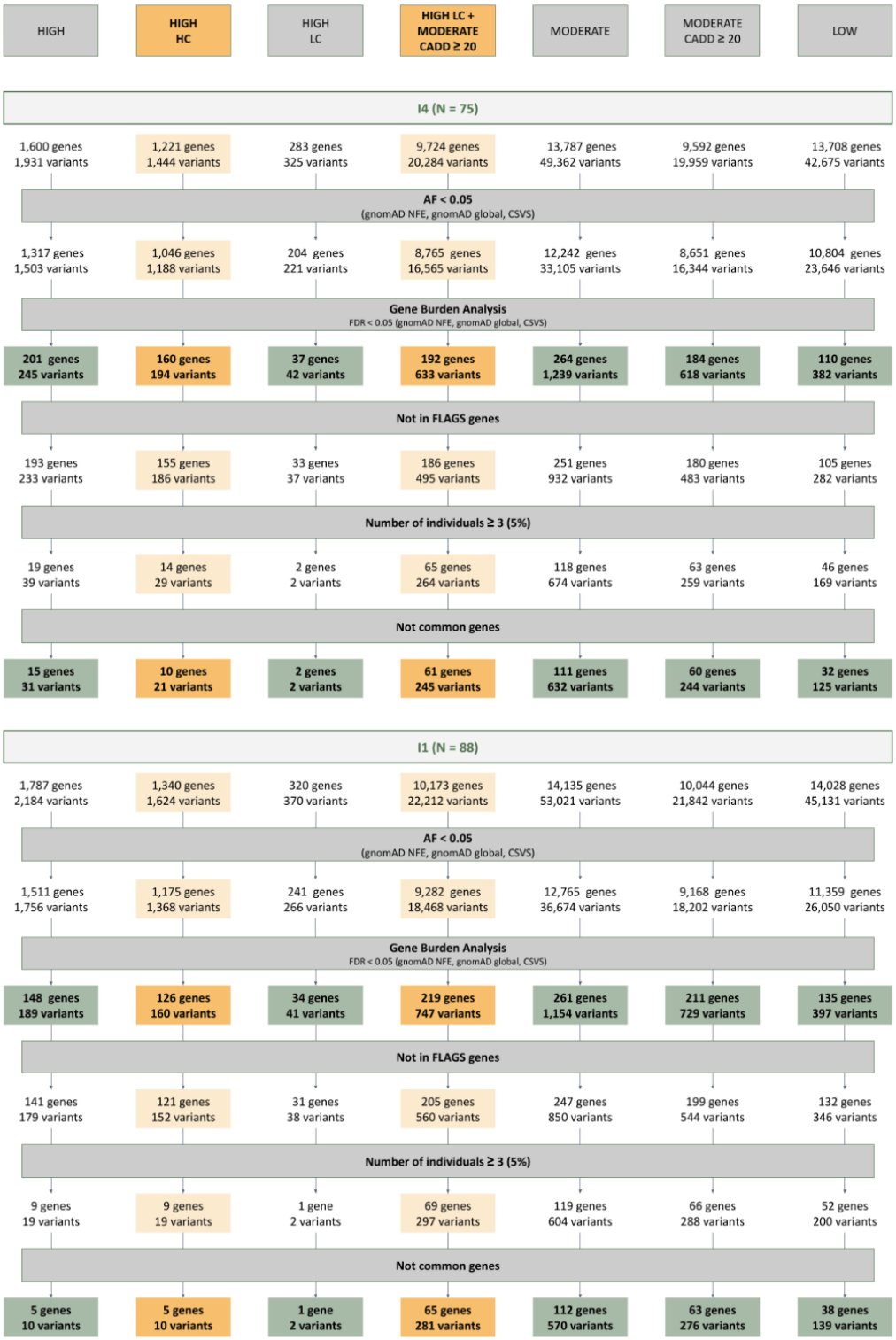

Figure 15 - Flow chart summarising the prioritisation strategy and the result of the gene burden analysis for I4 and I1 individuals, variants filtered by allele frequency (AF) < 0.05.

HC: High confidence; LC: Low confidence; CADD: Combined Annotation Dependent Depletion; FDR: p-value corrected by false discovery rate; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; FLAGS: FrequentLy mutAted GeneS.

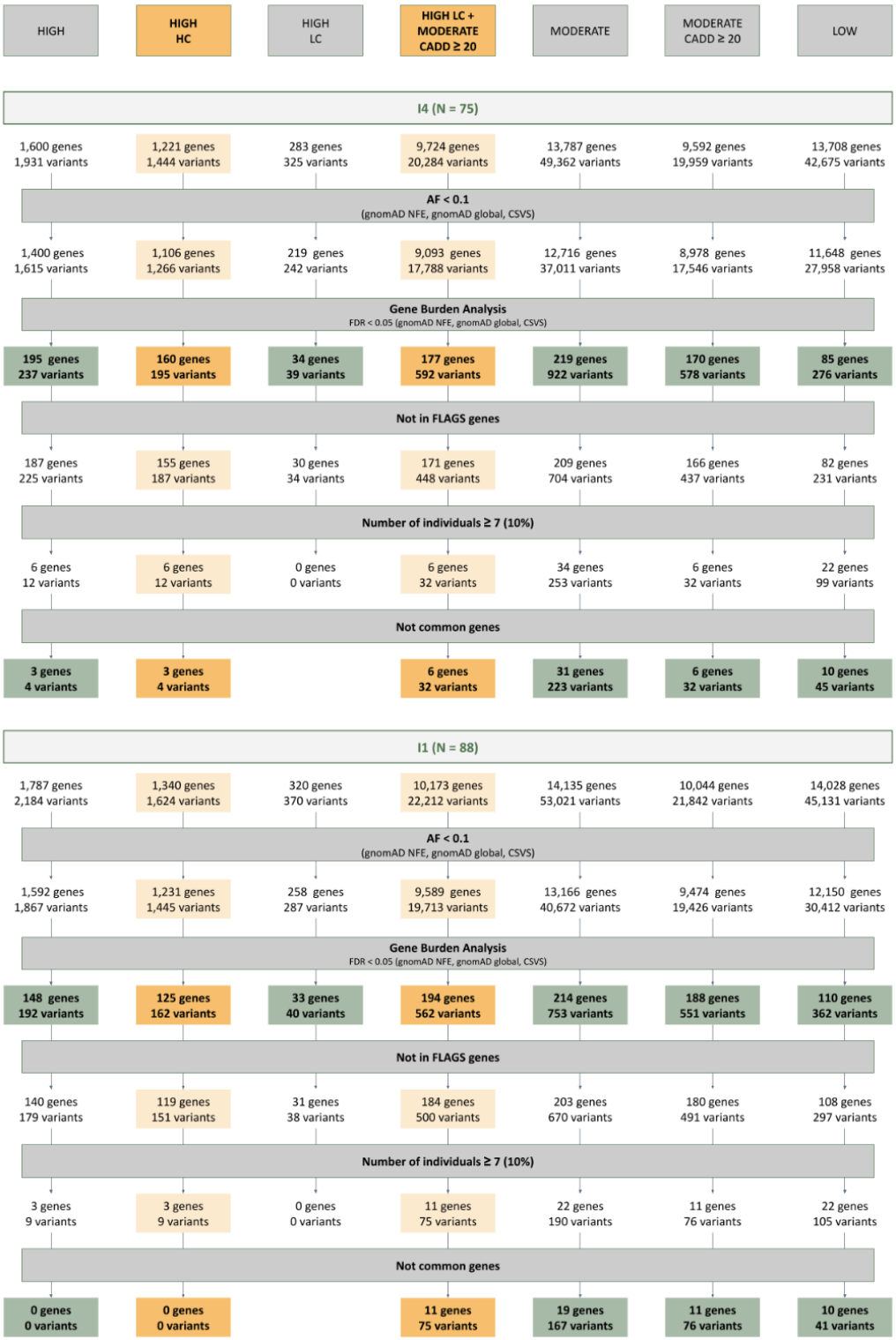

Figure 16 - Flow chart summarising the prioritisation strategy and the result of the gene burden analysis for I4 and I1 individuals, variants filtered by allele frequency (AF) < 0.1.

HC: High confidence; LC: Low confidence; CADD: Combined Annotation Dependent Depletion; FDR: p-value corrected by false discovery rate; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; FLAGS: FrequentLy mutAted GeneS.

5.3.3 Whole Meniere Disease cohort

The individuals in the study cohort suffer for MD, to define the genetic background associated with this disease two different GBA were performed with 310 patients (Figure 17). Consequently, these MD genes were studied regardless of the tinnitus phenotype.

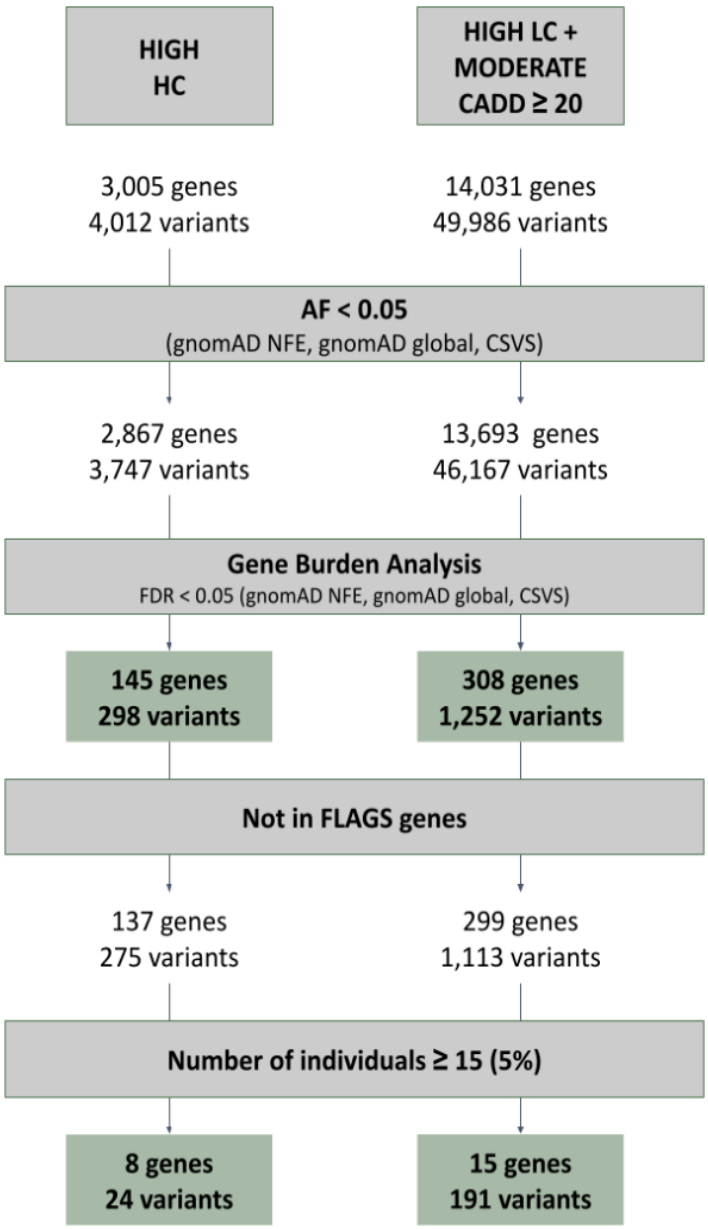

Figure 17 - Flow chart summarising the prioritisation strategy and the result of the gene burden analysis for all the MD individuals in the entire cohort, variants filtered by allelic frequency (AF) < 0.05.

HC: High confidence; LC: Low confidence; CADD: Combined Annotation Dependent Depletion; FDR: p-value corrected by false discovery rate; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; FLAGS: FrequentLy mutAted GeneS.

For the entire cohort, after the prioritisation, eight genes were enriched for variants with high confidence of being LoF, and 15 genes for variants with a low confidence of being LoF and variants with a moderate impact in the protein and predicted to be deleterious (Tables S11-S12).

5.3.4 Tinnitus Handicap Inventory (THI) subgroups

The main objective of this thesis was to identify the candidate genes and biological processes associated with severe tinnitus. GBAs were performed with the variants found in the individuals from the subgroups obtained according to their THI score (Figure 9) to identify candidate genes for tinnitus. The I4 subgroup (N = 75) was used to conduct a GBA to target genes with an overload of rare variants related to a worse progression of tinnitus. Moreover, the GBA was also carried out with the I1 (N = 88) to determine the genes associated with a burden of variants showing a protective effect to develop severe tinnitus.

5.3.4.1 Gene burden analysis: I4

Two GBAs were performed with variants with an AF less than 0.05 from the individuals in the subgroup I4 (Figure 18). Filtering by 0.05 allowed us to discard common variants in the population (if they are in more than 5% of the individuals in the reference populations). The AF of the variants were compared to the three reference populations: NFE from gnomAD, global from gnomAD and Spanish from CSVS; to which the individuals of the cohort belong according to their ancestry.

A total of 1,444 variants in 1,221 genes met the filter criteria of the HIGH HC group and AF less than 0.05. As a result of the GBA, 160 genes with 194 variants were found as enriched. After the defined filters (not FLAGS genes, number of individuals ≥ 3 and not being common genes), 10 genes with 21 variants were defined as candidate genes regarding LoF variants in the I4 subgroup (Figure 19, Table S13). The enriched, shortlisted genes, ranked by number of individuals with variants were: *PTTG2, FMN2, ENDOG, EFCAB5, LPIN3, MSH6, RECQL5, FBXO27, IQCN* and *RAD50.*

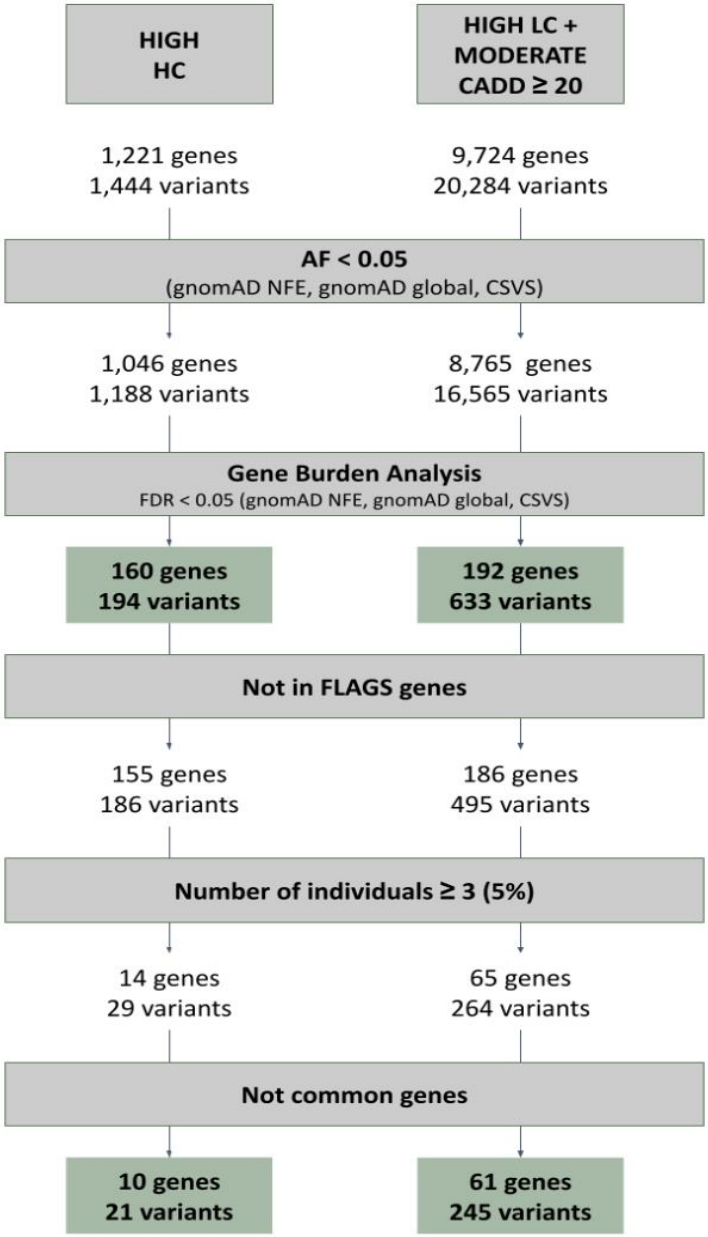

Figure 18 - Flow chart summarising the prioritisation strategy and the result of the gene burden analysis for I4 individuals, variants filtered by allelic frequency (AF) < 0.05.

HC: High confidence; LC: Low confidence; CADD: Combined Annotation Dependent Depletion; FDR: p-value corrected by false discovery rate; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; FLAGS: FrequentLy mutAted GeneS.

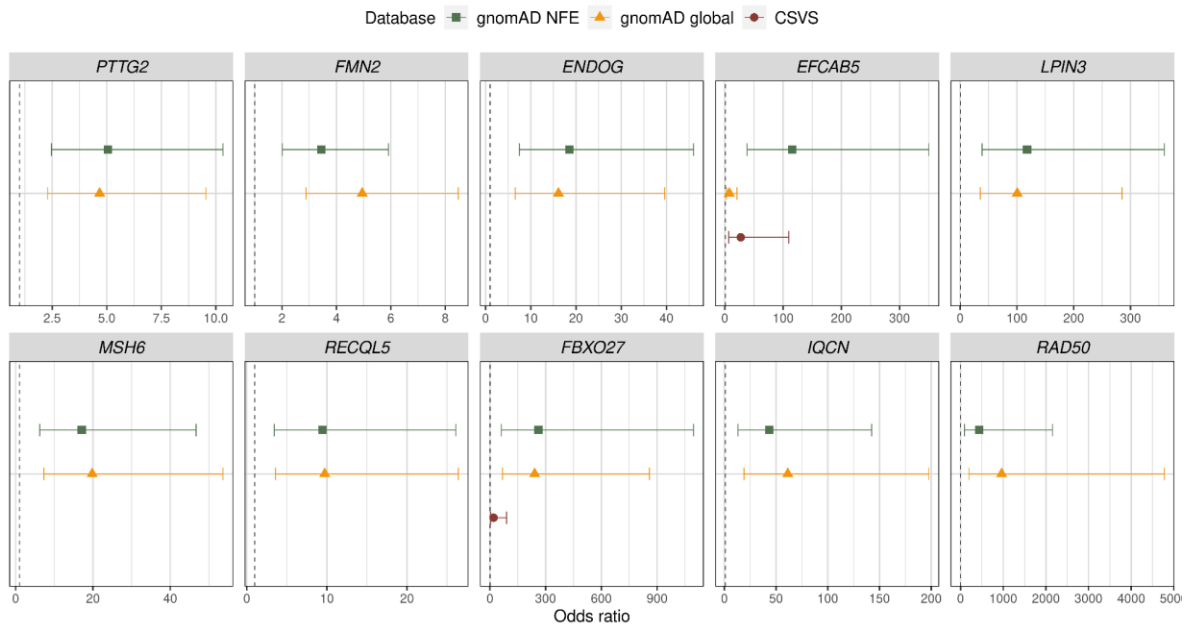

Figure 19 - Odds ratio (OR) of the genes enriched in variants with a predicted high confidence of being loss-of-function (HIGH HC), filtered by allelic frequency (AF) < 0.05, for the I4.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants.

The first step to study the variants of these genes was to validate them by IGV. The two variants in *FMN2* were not observed in any of the seven individuals reported by the calling. Regarding the *RECQL5* gene, two of the candidate variants were an insertion in two and four nucleotides in the same position; when the alignments were analysed it was observed an insertion - of two, four or six nucleotides - in most of the samples, also in I1 individuals and in gnomAD individuals, therefore this gene did not explain the tinnitus extreme phenotype. The variants of the remaining genes (N = 8) were validated by IGV and studied in more detail (Table 11).

Table 11 - Summary of variants found in the genes resulting from the gene burden analysis of variants with a high confidence of being loss-of-function for the I4 subgroup.

Gene symbol	Variant	Amino acid change	Consequence	AF				Individuals
				gnomAD NFE	gnomAD	CSVS	I4	
PTTG2	4:37960987 CAT>C	H185X	Frameshift	1.10E-02	1.19E-02	0	5.33E-02	I4-15, I4-32, I4-49, I4-57, I4-61, I4-62, I4-67, I4-73
ENDOG	9:128822601 CAGTA>C	SK296-297X	Frameshift	1.86E-03	2.14E-03	0	3.33E-02	I4-7, I4-15, I4-48, I4-63, I4-72
EFCAB5	17:30053751 CAG>C	R600X	Frameshift	0	0	0	1.33E-02	I4-32, I4-67
EFCAB5	17:30053806 C>T	R618*	Stop gained	1.08E-04	1.26E-04	0	6.67E-03	I4-49
EFCAB5	17:30078331 C>T	Q952*	Stop gained	1.24E-04	3.33E-03	1.00E-03	6.67E-03	I4-69
MSH6	2:47806652 G>GTAAC		Splice donor	2.64E-04	4.13E-04	0	1.33E-02	I4-52, I4-61
MSH6	2:47806652 GTAAC>G		Splice donor & Splice donor 5th base & Intron variant	1.32E-03	9.51E-04	0	1.33E-02	I4-3, I4-23
LPIN3	20:41349158 GC>G	P209X	Frameshift	2.32E-04	2.72E-04	0	2.67E-02	I4-8, I4-13, I4-40, I4-55
FBXO27	19:39026869 C>T		Splice donor	6.20E-05	7.00E-05	1.00E-03	1.33E-02	I4-10, I4-43
FBXO27	19:39031028 CCAGT>C	DW190-191X	Frameshift & Splice region	1.50E-05	1.40E-05	0	6.67E-03	I4-35
IQCN	19:18265085 CCT>C	Q818X	Frameshift	4.49E-04	3.14E-04	0	6.67E-03	I4-56
IQCN	19:18266016 T>TA	L508FX	Frameshift	0	0	0	6.67E-03	I4-42
IQCN	19:18266567 C>CT	-324-325X	Frameshift	1.60E-05	1.40E-05	0	6.67E-03	I4-53
RAD50	5:132595719 C>T	R706*	Stop gained	4.60E-05	2.10E-05	0	6.67E-03	I4-51
RAD50	5:132604038 T>TA	L839LX	Frameshift	0	0	0	6.67E-03	I4-74
RAD50	5:132604824 T>A	L848*	Stop gained	0	0	0	6.67E-03	I4-3

AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population.

Most of the variants found in the top genes were short indels, leading to a frameshift, splice acceptor or splice donor variants. As the summary table shows, the AF of almost all the variants in the Spanish population from CSVS was zero because in this database only SNVs were saved. Interestingly, the gene with variants in more individuals was *PTTG2*, where the same variant was identified in eight individuals.

Based on the LOEUF score, the *MSH6, EFCAB5, LPIN3, RAD50* and *IQCN* genes were slightly constrained. Any of the genes was constrained regarding the pLI score. In addition, no constraint data for the *PTTG2* gene were available in gnomAD (Table 12).

Table 12 - Constraint of the genes resulting from the gene burden analysis of variants with a high confidence of being loss-of-function for the I4 subgroup.

Gene symbol	LOEUF	pLI
PTTG2	-	-
ENDOG	1.638	0
EFCAB5	0.842	0
LPIN3	0.869	0
MSH6	0.498	0
FBXO27	1.760	0
IQCN	0.893	0
RAD50	0.873	0

pLI: Probability of being loss-of-function intolerant; LOEUF: Loss-of-function observed/expected upper bound fraction.

In the HIGH LC + MODERATE CADD ≥ 20 GBA were used 20,284 variants in 9,724 genes. The GBA identified a total of 192 genes with 633 variants. After the proposed filters, 61 genes with 245 variants were obtained as enriched for mainly missense variants in the I4 subgroup (Figure 20, Table S14). The enriched genes, ranked by number of individuals with variants, were: *THADA, UNC13C, DSCAML1, ITGB5, ANKRD44, WDFY4, ARVCF, CNTNAP2, EML6, ERCC6, GRIK4, MROH1, PDZD7, SPTBN4, CDH3, FURIN, KAT6A, MYBPC3, PPL, SASH1, SETD1A, STIM1, TAP1, TDRD6, VWA5B1, ZNF106, ARID5A, BIRC6, BRINP2, EPHA3, FKBP15, PDZRN3, PPFIBP1, RYR2, SALL3, SPTBN1, STAT6, STT3B, TRIB1, TRIM54, TTC22, ADAMTSL2, ADCY5, C12orf56, CASP14, CCDC14, CHD1, EIF4G2, FAM107A, GATA6, HDAC4, HOMER2, ILDR2, KCNH8, PKN3, RNF175, SLC9A3, SOWAHB, SRMS, TMPRSS13* and *ZNF554*.

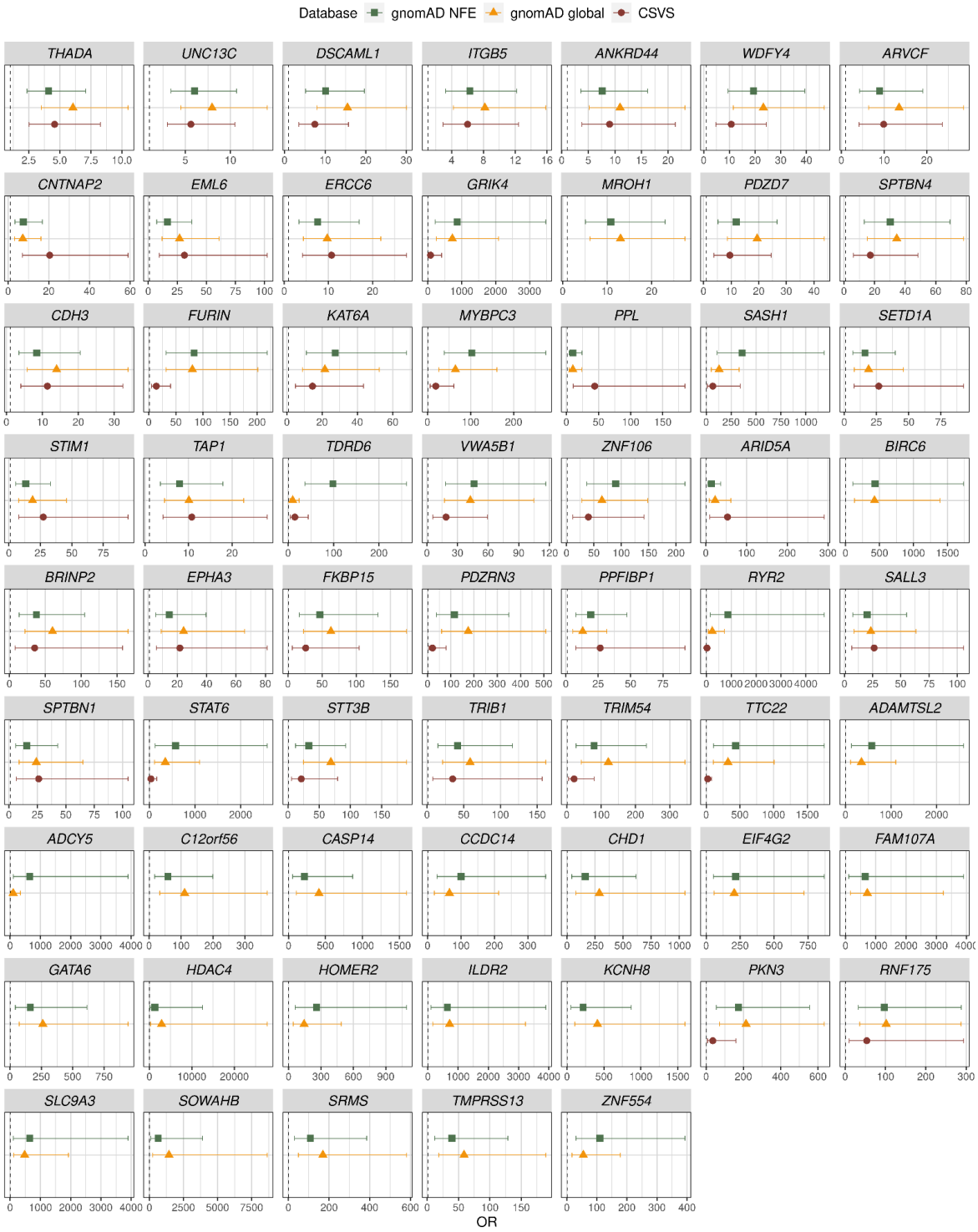

Figure 20 - Odds ratio (OR) of the genes enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20); filtered by allelic frequency (AF) < 0.05, for the I4.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants.

The 61 genes were used for further analysis; nevertheless, to study the more significant genes resulting from the GBA, an in-depth study of those with variants in more individuals was undertaken. The top 10 genes were selected, but as the last selected gene had variants in six individuals, all the genes with variants in at least six individuals - the top 14 - were studied (Table 13). All the variants in those genes were validated by IGV.

Table 13 - Summary of the variants found in the top 14 genes resulting from the gene burden analysis of variants with a low confidence of being loss-of-function and variants with a moderate impact in the protein for the I4 subgroup.

Gene symbol	Variant	Amino acid change	Consequence	CADD	AF				Individuals
					gnomAD NFE	gnomAD	CSVS	I4	
THADA	2:43279840C>T	E1741K	Missense	25.5	3.10E-03	2.22E-03	7.00E-03	1.33E-02	I4-10, I4-33
THADA	2:43291697T>G	E1670A	Missense & Splice region	24.1	0	0	0	6.67E-03	I4-31
THADA	2:43320457T>A	D1476V	Missense	21.4	6.20E-05	1.19E-04	1.00E-03	6.67E-03	I4-43
THADA	2:43320533G>A	P1451S	Missense	23.8	1.14E-02	7.08E-03	5.00E-03	2.00E-02	I4-19, I4-27, I4-30
THADA	2:43398139C>A	R1353S	Missense & Splice region	21.5	5.95E-03	4.41E-03	3.00E-03	2.00E-02	I4-3, I4-18, I4-40
THADA	2:43575013G>A	T351M	Missense	25.2	8.52E-04	6.42E-04	3.00E-03	2.00E-02	I4-34, I4-36, I4-54
UNC13C	15:54012994A>G	K31E	Missense	22.7	5.23E-03	3.39E-03	7.00E-03	1.33E-02	I4-18, I4-63
UNC13C	15:54013815C>G	D304E	Missense	22	5.69E-03	4.48E-03	6.00E-03	2.67E-02	I4-41, I4-68, I4-71, I4-75
UNC13C	15:54237666G>C	R1068S	Missense	23.5	3.10E-05	8.40E-05	0	6.67E-03	I4-5
UNC13C	15:54250290C>G	H1098Q	Missense	24.8	0	0	0	6.67E-03	I4-9
UNC13C	15:54333791G>A	E1507K	Missense	32	0	0	0	6.67E-03	I4-73
UNC13C	15:54626857C>A	A2130D	Missense	24	0	7.00E-06	0	6.67E-03	I4-1
UNC13C	15:54627055T>C	V2196A	Missense	26.8	2.46E-03	2.18E-03	2.00E-03	1.33E-02	I4-36, I4-57

Table 13 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I4*	
DSCAML1	*11:117437296C>G*	*E1516Q*	*Missense*	*24.1*	*0*	*0*	*0*	*6.67E-03*	*I4-69*
DSCAML1	11:117437352C>T	R1497Q	Missense	23.8	0	2.10E-05	0	6.67E-03	I4-32
DSCAML1	11:117439391C>T	R1340Q	Missense	25.3	5.38E-03	3.31E-03	7.00E-03	1.33E-02	I4-30, I4-52
DSCAML1	11:117458879G>A	T1148M	Missense	22.7	1.50E-05	8.40E-05	0	6.67E-03	I4-72
DSCAML1	11:117465081C>G	M1042I	Missense	23.3	0	0	0	6.67E-03	I4-8
DSCAML1	11:117505496C>T	A674T	Missense	20.4	4.65E-04	3.14E-04	1.00E-03	1.33E-02	I4-22, I4-73
DSCAML1	11:117518699G>A	S426L	Missense	23.3	1.24E-04	1.74E-04	0	6.67E-03	I4-4
ITGB5	3:124796649C>T	G478R	Missense	20.5	2.91E-03	2.68E-03	5.00E-03	2.00E-02	I4-39, I4-41, I4-49
ITGB5	3:124796747C>T	R445Q	Missense	25.8	1.55E-04	9.10E-05	1.00E-03	6.67E-03	I4-22
ITGB5	3:124859287C>A	V106F	Missense	23.9	0	0	0	6.67E-03	I4-32
ITGB5	3:124859365C>G	E80Q	Missense	24.5	6.60E-03	4.66E-03	4.00E-03	2.67E-02	I4-9, I4-28, I4-29, I4-51
WDFY4	10:48709925C>T	R65C	Missense	22.2	1.08E-04	7.00E-05	0	6.67E-03	I4-2
WDFY4	10:48720123C>A	A116E	Missense & Splice region	25.2	1.22E-03	1.24E-03	4.00E-03	6.67E-03	I4-61
WDFY4	10:48721347G>A	V146M	Missense	22.7	7.12E-04	3.77E-04	0	6.67E-03	I4-69
WDFY4	10:48774609A>T	H902L	Missense	25.3	3.10E-05	1.40E-05	0	6.67E-03	I4-8
WDFY4	10:48779970C>G	P1143A	Missense	23.3	2.79E-04	1.47E-04	0	6.67E-03	I4-34
WDFY4	10:48897507A>C	H2457P	Missense	20.6	0	2.10E-05	0	6.67E-03	I4-7
WDFY4	10:48957144G>A	V2665M	Missense	26	4.03E-04	4.33E-04	0	6.67E-03	I4-15
WDFY4	10:48981427C>T	T3146I	Missense	23.5	1.50E-05	7.00E-06	0	6.67E-03	I4-69

Table 13 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I4*	
ANKRD44	2:196989644T>C	R977G	Missense	23.2	1.86E-04	2.65E-04	0	6.67E-03	I4-21
ANKRD44	2:197005796C>T	A749T	Missense	22	6.20E-05	4.90E-05	0	6.67E-03	I4-52
ANKRD44	2:197078765T>G	K530Q	Missense	26.2	5.39E-03	3.67E-03	5.00E-03	2.00E-02	I4-6, I4-28, I4-54
ANKRD44	2:197083381T>C	D482G	Missense	30	5.42E-04	3.07E-04	0	6.67E-03	I4-23
ANKRD44	2:197125890C>T	D137N	Missense	26.9	0	0	0	6.67E-03	I4-13
MROH1	8:144239732C>T	P584L	Missense	22.4	3.10E-05	1.54E-04	0	6.67E-03	I4-17
MROH1	8:144248929T>C	M1058T	Missense	23.2	2.01E-04	1.68E-04	0	1.33E-02	I4-35, I4-52
MROH1	8:144254922C>T	R1180W	Missense	24.9	3.80E-03	2.87E-03	0	6.67E-03	I4-54
MROH1	8:144255693T>C	L1260P	Missense	23.9	3.10E-04	3.84E-04	0	6.67E-03	I4-57
MROH1	8:144260028G>A	A1388T	Missense	23.7	0	0	0	6.67E-03	I4-54
MROH1	8:144260223T>C	I1410T	Missense	21.9	0	4.20E-05	0	6.67E-03	I4-24
ARVCF	22:19971257T>C	K954E	Missense	24.6	0	7.00E-06	0	6.67E-03	I4-31
ARVCF	22:19973010G>A	A822V	Missense	24.8	2.01E-04	2.23E-04	0	6.67E-03	I4-62
ARVCF	22:19977463G>A	R608C	Missense	31	6.81E-04	5.51E-04	1.00E-03	1.33E-02	I4-55, I4-57
ARVCF	22:19977978G>A	R560W	Missense	24.6	6.04E-04	3.84E-04	2.00E-03	6.67E-03	I4-52
ARVCF	22:19981613C>T	R165Q	Missense	23.6	1.08E-04	2.65E-04	0	6.67E-03	I4-71
ARVCF	22:19981659C>T	D150N	Missense	27.5	3.64E-03	2.05E-03	1.00E-03	6.67E-03	I4-71

Table 13 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I4*	
SPTBN4	19:40497499G>A	V227M	Missense	25.6	3.10E-05	2.10E-05	0	6.67E-03	I4-18
SPTBN4	19:40502161A>T	I311L	Missense	22.9	4.65E-04	3.35E-04	2.00E-03	6.67E-03	I4-72
SPTBN4	19:40532673C>T	R1333W	Missense	23	3.10E-05	3.50E-05	0	6.67E-03	I4-58
SPTBN4	19:40532770G>A	R1365Q	Missense & Splice region	24.4	1.50E-05	1.40E-05	0	6.67E-03	I4-29
SPTBN4	19:40550266C>T	A1538V	Missense	25.8	6.66E-04	4.05E-04	0	6.67E-03	I4-69
SPTBN4	19:40557357G>A	R1875Q	Missense	29.9	1.24E-04	3.56E-04	0	6.67E-03	I4-68
EML6	2:54829435A>G	I269V	Missense	23.1	2.94E-04	2.09E-04	0	6.67E-03	I4-29
EML6	2:54866839G>A	R669Q	Missense	26.4	1.39E-04	6.30E-05	1.00E-03	6.67E-03	I4-69
EML6	2:54892634T>A	F907Y	Missense	26.7	1.89E-03	1.15E-03	1.00E-03	6.67E-03	I4-52
EML6	2:54894945G>A	V925M	Missense	26.5	0	0	0	6.67E-03	I4-56
EML6	2:54911019C>T	R1159W	Missense	26.2	1.08E-04	7.00E-05	0	6.67E-03	I4-43
EML6	2:54916835A>G	D1192G	Missense	25	0	0	0	6.67E-03	I4-26
CNTNAP2	7:147044039T>G	Y179D	Missense	25.4	0	0	0	6.67E-03	I4-5
CNTNAP2	7:147121078G>C	G285A	Missense	23.8	5.30E-03	3.97E-03	2.00E-03	6.67E-03	I4-17
CNTNAP2	7:147121123G>A	R300H	Missense	24.4	0	4.90E-05	0	6.67E-03	I4-64
CNTNAP2	7:147639255G>A	E683K	Missense	25.3	0	2.10E-05	0	6.67E-03	I4-26
CNTNAP2	7:147903589T>C	V708A	Missense	24.3	7.70E-05	1.47E-03	0	6.67E-03	I4-20
CNTNAP2	7:147903652C>T	A729V	Missense	23.3	0	4.90E-05	0	6.67E-03	I4-2

Table 13 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I4*	
GRIK4	11:120660376T>G	C20G	Missense	22	0	0	0	1.33E-02	I4-23, I4-71
GRIK4	11:120815389C>T	L87F	Missense	25.3	1.50E-05	7.00E-06	0	6.67E-03	I4-58
GRIK4	11:120952906T>A	W548R	Missense	32	0	0	0	6.67E-03	I4-57
GRIK4	11:120956827G>A	R583Q	Missense	25.1	0	7.00E-06	0	6.67E-03	I4-21
GRIK4	11:120982169T>C	I820T	Missense	24.1	3.10E-05	4.20E-05	0	6.67E-03	I4-34
ERCC6	10:49459074T>G	E1408A	Missense	22.3	1.87E-03	1.51E-03	0	6.67E-03	I4-49
ERCC6	10:49482775G>A	P694L	Missense	27.5	2.02E-04	1.19E-04	0	6.67E-03	I4-60
ERCC6	10:49524156T>G	D425A	Missense	22.8	2.94E-03	1.85E-03	3.00E-03	1.33E-02	I4-1, I4-22
ERCC6	10:49530733T>C	Q177R	Missense	26.9	1.24E-04	6.30E-05	0	6.67E-03	I4-38
ERCC6	10:49532565G>A	R134W	Missense	22.6	1.86E-04	5.93E-04	0	6.67E-03	I4-25
PDZD7	10:101009273C>T	A899T	Missense	25.6	1.50E-05	7.00E-06	0	6.67E-03	I4-34
PDZD7	10:101015713G>A	R558W	Missense	23.3	1.55E-04	1.05E-04	0	1.33E-02	I4-20, I4-52
PDZD7	10:101019175C>T	S324N	Missense	25.6	3.00E-03	1.78E-03	2.00E-03	1.33E-02	I4-39, I4-69
PDZD7	10:101024017G>A	P93L	Missense	23.4	2.48E-04	1.88E-04	2.00E-03	6.67E-03	I4-23

CADD: Combined Annotation Dependent Depletion; AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population.

As expected, all the variants in the top genes resulting from this analysis were missense, besides four of them also affected the splice site. A total of 22 variants were not found in the NFE reference population from gnomAD, 13 in the whole reference population from gnomAD and 40 in the Spanish reference population from CSVS. The variants with the highest number of individuals carrying them were chr15:54013815C>G in the *UNC13C* gene and chr3:124859365C>G in the *ITGB5* gene, both of them with four different individuals.

The most constrained genes were *DSCAML1, WDFY4* and *SPTBN4* according to the lower ratio between observed missense variants in the gene and the expected based on the gnomAD mutational model and the higher value of the Z score (Table 14).

Table 14 - Constraint of the top 14 genes resulting from the gene burden analysis of variants with a low confidence of being loss-of-function and variants with a moderate impact in the protein for the I4 subgroup.

Gene symbol	o/e	Z score
THADA	1.268 (1.209-1.330)	-2.964
UNC13C	1.012 (0.962-1.064)	-0.143
DSCAML1	0.759 (0.721-0.799)	3.15
ITGB5	0.862 (0.794-0.936)	1.064
WDFY4	0.785 (0.750-0.822)	3.124
ANKRD44	0.802 (0.707-0.910)	1.029
GRIK4	0.760 (0.702-0.823)	2.041
SPTBN4	0.701 (0.663-0.739)	3.934
EML6	0.903 (0.855-0.953)	1.116
PDZD7	1.058 (0.968-1.156)	-0.372
MROH1	0.899 (0.811-0.998)	0.598
ARVCF	1.030 (0.966-1.097)	-0.27
ERCC6	0.990 (0.932-1.051)	0.098
CNTNAP2	1.030 (0.970-1.093)	-0.291

o/e: Observed/expected ratio.

5.3.4.2 Gene burden analysis: I1

We also conducted two GBAs in the subgroup I1, following the same workflow previously explained for the subgroup I4 (Figure 21). The used variants were filtered by an AF less than 0.05.

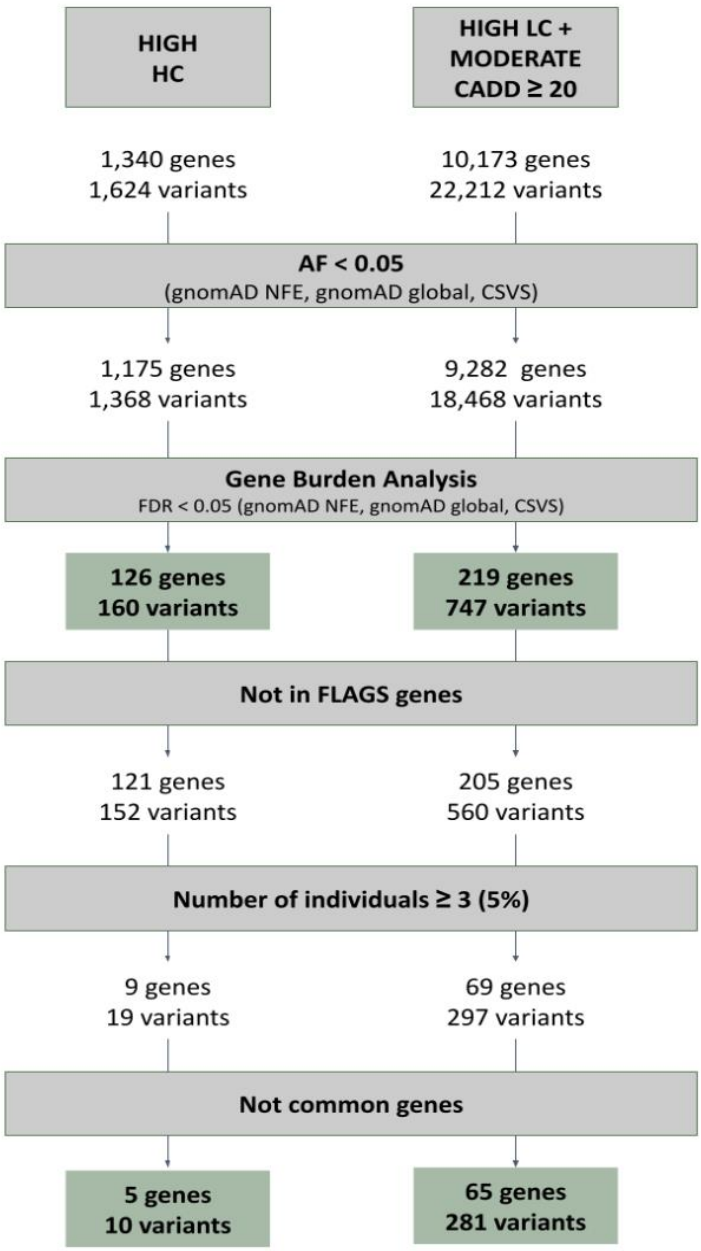

Figure 21 - Flow chart summarising the prioritisation strategy and the result of the gene burden analysis for I1 individuals, variants filtered by allelic frequency (AF) < 0.05.

HC: High confidence; LC: Low confidence; CADD: Combined Annotation Dependent Depletion; FDR: p-value corrected by false discovery rate; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; FLAGS: FrequentLy mutAted GeneS.

For the HIGH HC group, 1,624 variants in 1,340 genes pass the filter criteria of impact and AF. After the GBA, 126 genes with 160 variants were enriched. Finally, with the defined filters 5 genes with 10 variants were pointed as candidates for LoF variants in the I1 subgroup (Figure 22, Table S15): *ADAM2*, *TOMM20L*, *ENTPD8*, *PRKAGC* and *RBM5*.

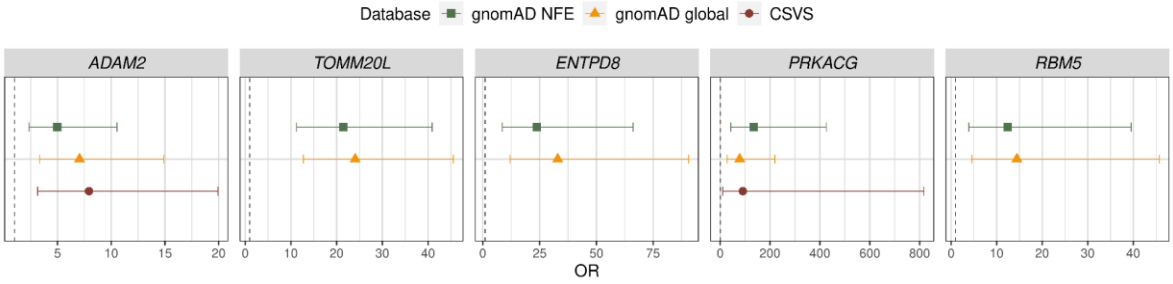

Figure 22 - Odds ratio (OR) of the genes enriched in variants with a predicted high confidence of being loss-of-function (HIGH HC), filtered by allelic frequency (AF) < 0.05, for the I1.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants.

Through the IGV validation, the variant in the *RBM5* gene was discarded because it appeared in most of the samples, including the individuals from I4, which did not explain the phenotype. Moreover, it was observed that in the *TOMM20L* gene two different frameshift variants in consecutive positions were reported by the calling, but when the alignment was analysed an in-frame deletion in the same position followed by a missense variant. As any of the two variants had a high impact in the protein, they were eliminated; and as the CADD scores were below 20, they were not included in the results of the following GBA. The variants in the remaining genes (N = 3) were confirmed by IGV and analysed in detail (Table 15).

Three of the seven variants found in the top genes were frameshifts caused by a short indel, moreover two stop gained and two splice acceptor variants. The AF of the short indels in the Spanish reference population from CSVS was zero because this database only collected SNVs. Remarkably, the frameshift variant in *ADAM2* was found in heterozygosis for one individual and in homozygosis for the other.

According to the described criteria, *ADAM* and *ENTPD8* genes were not constrained. Moreover, no constraint data for the *PRKACG* gene were available in gnomAD (Table 16).

Table 15 - Summary of variants found in the genes resulting from the gene burden analysis of variants with a high confidence of being loss-of-function for the I1 subgroup.

Gene symbol	Variant	Amino acid change	Consequence	AF				Individuals
				gnomAD NFE	gnomAD	CSVS	I1	
ADAM2	8:39755912 C>T		Splice acceptor	0	7.00E-06	1.00E-03	5.68E-03	I1-17
ADAM2	8:39769449 T>TAGTG	A385ATX	Frameshift	2.79E-04	4.33E-04	0	1.70E-02	I1-36, I1-54 (hom)
ADAM2	8:39788742 T>A		Splice acceptor	7.79E-03	5.25E-03	6.00E-03	1.14E-02	I1-27, I1-78
ADAM2	8:39821106 C>A	E137*	Stop gained	0	0	0	5.68E-03	I1-48
ENTPD8	9:137436928 GC>G	G165X	Frameshift	9.75E-04	7.04E-04	0	2.27E-02	I1-28, I1-50, I1-57, I1-66
PRKACG	9:69013330 T>TC	-254-255X	Frameshift	1.70E-04	2.93E-04	0	1.14E-02	I1-25, I1-42
PRKACG	9:69013503 C>T	W197*	Stop gained	0	0	0	1.14E-02	I1-3, I1-21

AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; hom: homozygous.

Table 16 - Constraint of the genes resulting from the gene burden analysis of variants with a high confidence of being loss-of-function for the I1 subgroup.

Gene symbol	LOEUF	pLI
ADAM2	1.176	0
ENTPD8	1.232	0
PRKACG	-	-

pLI: Probability of being loss-of-function intolerant; LOEUF: Loss-of-function observed/expected upper bound fraction.

A sum of 22,212 variants in 10,173 genes met the filter criteria of the HIGH LC + MODERATE CADD ≥ 20 GBA and AF less than 0.05. As a result of the GBA, 219 genes were found as enriched with 747 variants. After the filters, 65 genes with 281 variants were selected for the I1 subgroup (Figure 23, Table S16).

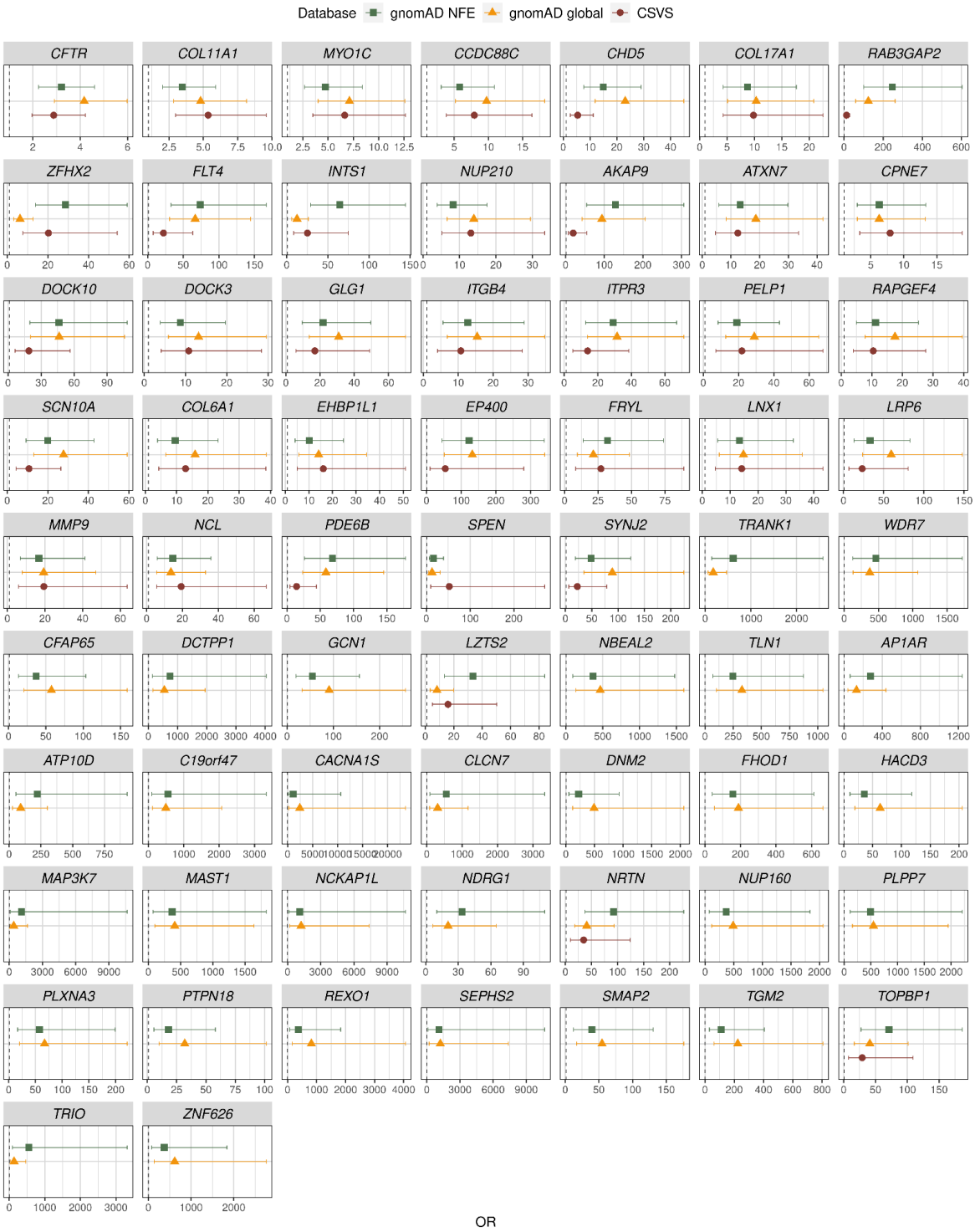

Figure 23 - Odds ratio (OR) of the genes enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20); filtered by allelic frequency (AF) < 0.05, for the I1.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants.

As in the analysis of the I4 subgroup, in addition to obtaining all the genes from the GBA for subsequent analyses, the genes with variants in a greater number of individuals were studied deeper. In this subgroup, the top 10 genes were selected, as the last selected gene had variants in seven individuals, all the genes with variants in seven or more individuals - the top 11 - were studied (Table 17).

Table 17 - Summary of variants found in the top 11 genes resulting from the gene burden analysis of variants with a low confidence of being loss-of-function and variants with a moderate impact in the protein for the I1 subgroup.

Gene symbol	Variant	Amino acid change	Consequence	CADD	AF				Individuals
					gnomAD NFE	gnomAD	CSVS	I1	
CFTR	7:117509089C>T	R74W	Missense	24.60	1.86E-04	4.34E-03	1.00E-03	1.70E-02	I1-15, I1-66 (hom)
CFTR	7:117509093G>A	R75Q	Missense	29.80	3.40E-02	1.89E-02	8.00E-03	1.70E-02	I1-22, I1-35, I1-58
CFTR	7:117531068T>C	I148T	Missense	22.60	1.55E-03	9.28E-04	3.00E-03	5.68E-03	I1-9
CFTR	7:117536576A>G	R258G	Missense	25.50	1.39E-04	1.12E-04	0	5.68E-03	I1-85
CFTR	7:117540282C>G	T351S	Missense	24.60	1.70E-04	1.75E-04	2.00E-03	5.68E-03	I1-83
CFTR	7:117559521C>T	H484Y	Missense	25.20	0	2.80E-05	0	5.68E-03	I1-59
CFTR	7:117559587A>G	I506V	Missense	23.80	3.72E-04	3.35E-04	1.00E-03	5.68E-03	I1-31
CFTR	7:117590400G>C	G576A	Missense	25.80	7.32E-03	5.08E-03	1.80E-02	3.41E-02	I1-13, I1-19, I1-28, I1-44, I1-59, I1-79
CFTR	7:117592169C>T	R668C	Missense	25.50	8.86E-03	6.12E-03	1.80E-02	3.98E-02	I1-13, I1-19, I1-28, I1-29, I1-44, I1-59, I1-79
CFTR	7:117592588A>G	I807M	Missense	24.00	6.35E-04	4.39E-04	1.00E-03	5.68E-03	I1-8
CFTR	7:117594945G>T	D836Y	Missense	30.00	3.10E-04	3.91E-04	5.00E-03	5.68E-03	I1-69
CFTR	7:117642528G>A	D1270N	Missense	28.90	6.20E-05	4.39E-03	1.00E-03	1.70E-02	I1-15, I1-66 (hom)
CFTR	7:117667023G>A	R1453Q	Missense	23.20	1.50E-05	1.40E-05	0	5.68E-03	I1-59

Table 17 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I1*	
COL11A1	1:102886930T>C	I1579V	Missense	23.00	0	0	0	5.68E-03	I1-56
COL11A1	1:102888579G>C	P1536A	Missense & Splice region	29.60	1.36E-02	8.82E-03	8.00E-03	4.55E-02	I1-2, I1-3, I1-17, I1-20, I1-24, I1-31 (hom), I1-42
COL11A1	1:102962756G>T	P974Q	Missense	27.60	3.52E-03	2.05E-03	3.00E-03	5.68E-03	I1-50
COL11A1	1:102979414A>T	F860I	Missense	24.60	4.21E-03	4.89E-03	3.00E-03	5.68E-03	I1-32
COL11A1	1:103014591G>A	R498C	Missense	29.70	0	7.00E-06	0	5.68E-03	I1-30
COL11A1	1:103074709G>T	T187K	Missense	28.10	3.10E-05	3.50E-05	0	5.68E-03	I1-16
COL11A1	1:103078818C>G	G110R	Missense	26.30	1.66E-03	8.60E-04	1.00E-03	5.68E-03	I1-70
MYO1C	17:1468481C>T	V876M	Missense	25.30	1.50E-05	1.40E-05	0	5.68E-03	I1-41
MYO1C	17:1469545C>T	E866K	Missense	23.30	1.08E-02	7.00E-03	4.00E-03	2.27E-02	I1-13, I1-25, I1-52, I1-67
MYO1C	17:1470651G>A	R751W	Missense	32.00	1.91E-03	1.51E-03	2.00E-03	5.68E-03	I1-7
MYO1C	17:1480564G>C	A290G	Missense	25.10	1.86E-04	1.12E-04	0	5.68E-03	I1-21
MYO1C	17:1480595C>T	D280N	Missense	25.90	1.55E-04	1.12E-04	0	5.68E-03	I1-85
MYO1C	17:1482889C>T	R173Q	Missense	22.80	1.43E-03	9.18E-04	3.00E-03	1.70E-02	I1-39, I1-68, I1-69
MYO1C	17:1484201C>T	E60K	Missense	25.90	1.50E-05	7.00E-06	0	5.68E-03	I1-16

Table 17 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I1*	
CCDC88C	14:91272686G>A	P2009L	Missense	25.00	1.22E-03	7.60E-04	1.00E-03	5.68E-03	I1-81
CCDC88C	14:91289303T>C	I1415V	Missense	25.00	0	0	0	5.68E-03	I1-75
CCDC88C	14:91294308C>T	R1326H	Missense	23.70	1.50E-05	7.00E-06	0	5.68E-03	I1-75
CCDC88C	14:91297376G>A	R1299C	Missense	29.80	7.29E-03	4.28E-03	5.00E-03	2.27E-02	I1-4, I1-17, I1-60, I1-86
CCDC88C	14:91313931G>A	R629W	Missense	24.20	1.09E-04	1.19E-04	1.00E-03	5.68E-03	I1-70
CCDC88C	14:91314101G>A	S572L	Missense	23.50	6.97E-04	4.12E-04	0	5.68E-03	I1-8
CCDC88C	14:91359660C>T	G108S	Missense	27.30	5.11E-04	2.86E-04	0	5.68E-03	I1-52
ZFHX2	14:23522249G>A	R2478W	Missense	25.00	2.17E-04	1.74E-04	2.00E-03	1.14E-02	I1-17, I1-43
ZFHX2	14:23523439C>T	R2168H	Missense	26.10	3.10E-05	6.27E-03	0	5.68E-03	I1-24
ZFHX2	14:23524433C>T	G1837S	Missense	23.60	1.24E-04	1.26E-04	0	5.68E-03	I1-3
ZFHX2	14:23525156G>A	R1596W	Missense	25.60	6.20E-05	3.50E-05	0	5.68E-03	I1-71
ZFHX2	14:23526113G>A	R1277W	Missense	23.50	0	1.40E-05	0	5.68E-03	I1-87
ZFHX2	14:23532748C>T	R793Q	Missense	24.00	1.50E-05	2.10E-05	0	5.68E-03	I1-30
ZFHX2	14:23533485C>T	G614E	Missense	22.80	1.15E-03	6.00E-04	1.00E-03	5.68E-03	I1-66
RAB3GAP2	1:220162252C>T	M1057I	Missense	23.90	9.30E-05	9.80E-05	1.00E-03	5.68E-03	I1-65
RAB3GAP2	1:220171992C>T	R825H	Missense	29.20	0	0	0	5.68E-03	I1-57
RAB3GAP2	1:220189752T>A	E577V	Missense	31.00	0	7.00E-06	0	5.68E-03	I1-11
RAB3GAP2	1:220190428G>A	P527L	Missense	27.40	4.60E-05	1.88E-04	2.00E-03	1.70E-02	I1-10, I1-63, I1-64
RAB3GAP2	1:220196260T>C	Y317C	Missense	24.70	1.50E-05	2.80E-05	0	5.68E-03	I1-75
RAB3GAP2	1:220232860T>C	K40R	Missense	22.90	3.10E-05	4.90E-05	0	5.68E-03	I1-6

Table 17 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I1*	
CHD5	1:6111775G>A	T1750M	Missense & Splice region	27.30	8.98E-04	6.28E-04	4.00E-03	1.14E-02	I1-45, I1-77
CHD5	1:6111799T>G	Y1742S	Missense	30.00	0	4.90E-05	0	5.68E-03	I1-17
CHD5	1:6128047A>G	V1301A	Missense & Splice region	22.70	1.60E-05	7.00E-06	0	5.68E-03	I1-83
CHD5	1:6142157G>A	R803W	Missense	32.00	0	7.00E-06	0	5.68E-03	I1-56
CHD5	1:6155653G>A	S151L	Missense	25.00	2.56E-03	1.54E-03	5.00E-03	2.27E-02	I1-38, I1-54, I1-80, I1-83
COL17A1	10:104032959G>A	A1435V	Missense	24.90	3.01E-03	2.69E-03	3.00E-03	2.84E-02	I1-9, I1-15, I1-26, I1-75, I1-78
COL17A1	10:104034137C>T	G1322S	Missense	22.80	4.80E-04	4.05E-04	1.00E-03	5.68E-03	I1-24
COL17A1	10:104034657C>T	D1244N	Missense	26.40	8.52E-04	6.49E-04	0	5.68E-03	I1-40
COL17A1	10:104057101C>A	G447C	Missense	21.80	9.29E-04	7.13E-04	1.00E-03	5.68E-03	I1-30
INTS1	7:1471578A>C	F2083C	Missense	22.60	5.11E-04	1.40E-03	1.00E-03	5.68E-03	I1-33
INTS1	7:1473168T>C	N1992D	Missense	20.20	0	0	0	5.68E-03	I1-41
INTS1	7:1474801G>A	A1847V	Missense	26.60	4.60E-05	1.64E-03	0	5.68E-03	I1-61
INTS1	7:1479449C>T	R1437H	Missense	28.20	0	7.00E-06	0	5.68E-03	I1-65
INTS1	7:1498539G>A	A433V	Missense	22.50	3.10E-05	9.80E-05	0	5.68E-03	I1-11
INTS1	7:1500295C>T	V141M	Missense	24.20	1.50E-05	2.80E-05	0	5.68E-03	I1-64
INTS1	7:1503909G>A	P18S	Missense	23.00	1.50E-05	4.20E-05	0	5.68E-03	I1-23

Table 17 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD*	*CSVS*	*I1*	
NUP210	3:13321694G>A	P1686L	Missense	28.70	0	4.90E-05	0	5.68E-03	I1-85
NUP210	3:13339966G>A	A1120V	Missense	25.70	4.60E-05	5.60E-05	0	5.68E-03	I1-47
NUP210	3:13341822C>T	G1052S	Missense	24.50	4.75E-03	2.73E-03	2.00E-03	5.68E-03	I1-44
NUP210	3:13341834C>T	G1048S	Missense	24.40	1.50E-05	7.00E-06	1.00E-03	5.68E-03	I1-53
NUP210	3:13371899C>T	S574N	Missense	23.30	1.50E-05	2.10E-05	1.00E-03	5.68E-03	I1-4
NUP210	3:13391293T>G	T151P	Missense	27.20	0	0	0	5.68E-03	I1-66
NUP210	3:13397363A>C	S144A	Missense	22.80	0	0	0	5.68E-03	I1-84
FLT4	5:180609010A>G	F1284S	Missense	26.50	1.50E-05	7.00E-06	0	5.68E-03	I1-40
FLT4	5:180616393C>G	D1065H	Missense	28.90	1.50E-05	7.00E-06	0	5.68E-03	I1-17
FLT4	5:180625965G>A	A442V	Missense	23.00	1.39E-04	2.16E-04	1.00E-03	1.14E-02	I1-31, I1-50
FLT4	5:180626236C>T	R378H	Missense	24.60	2.48E-04	2.44E-04	0	5.68E-03	I1-14
FLT4	5:180630035G>A	T195M	Missense	22.90	1.50E-05	1.40E-05	0	5.68E-03	I1-46
FLT4	5:180630056C>T	R188Q	Missense	21.60	1.08E-04	1.12E-04	0	5.68E-03	I1-78

CADD: Combined Annotation Dependent Depletion; AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; hom: homozygous.

All the variants in the top genes were missense and three of them also affected the splice site. Fourteen variants were not found in the NFE reference population from gnomAD, six in the whole reference population from gnomAD and 42 in the Spanish reference population from CSVS. The variant carried by a greater number of individuals, seven, was chr7:117592169C>T in the *CFTR* gene.

The *CHD5*, *ZFHX2* and *FLT4* genes, specially the first one, were those more constraint because of the lower ratio between observed and expected missense variants in the gene based on the mutational model of gnomAD and the higher value of the Z score (Table 18).

Table 18 - Constraint of the top 11 genes resulting from the gene burden analysis of variants with a low confidence of being loss-of-function and variants with a moderate impact in the protein for the I1 subgroup.

Gene symbol	o/e	Z score
CFTR	1.322 (1.254-1.393)	-3.14
COL11A1	0.909 (0.859-0.960)	1.024
MYO1C	1.070 (1.005-1.139)	-0.637
CCDC88C	0.942 (0.896-0.990)	0.708
ZFHX2	0.768 (0.728-0.809)	2.951
RAB3GAP2	0.908 (0.851-0.969)	0.868
CHD5	0.565 (0.529-0.602)	5.317
COL17A1	1.072 (1.014-1.132)	-0.748
INTS1	0.892 (0.850-0.936)	1.409
NUP210	0.935 (0.888-0.983)	0.782
FLT4	0.728 (0.681-0.778)	2.834

o/e: Observed/expected ratio.

5.3.5 Questionnaire of hypersensitivity to sound (GÜF) subgroups

Following the previously described workflow for the THI subgroups, a GBA in the individuals from the clusters obtained according to the GÜF score (Figure 10) was performed. The main cohort was I4-GÜF (N = 34), which was used to identify genes associated with severe hyperacusis. The I1-GÜF subgroup (N = 42) was studied to select genes with variants that would decrease the susceptibility to develop hyperacusis.

5.3.5.1 Gene burden analysis: I4-GÜF

Therefore, two GBAs were caried out with variants from the I4-GÜF individuals (Figure 24).

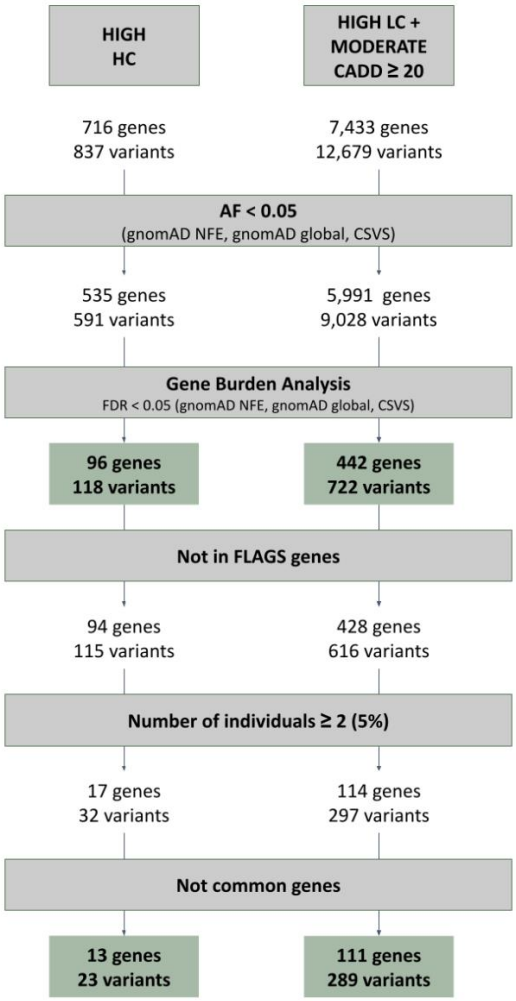

Figure 24 - Flow chart summarising the prioritisation strategy and the result of the gene burden analysis for I4-GÜF individuals, variants filtered by AF < 0.05.

HC: High confidence; LC: Low confidence; CADD: Combined Annotation Dependent Depletion; FDR: p-value corrected by false discovery rate; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; FLAGS: FrequentLy mutAted GeneS.

The candidate genes for the HIGH HC analysis (Figure 25) and for the HIGH LC + MODERATE CADD ≥ 20 (Figure 26) of the I4-GÜF cluster, were studied to compare them with those obtained in the I4 group at a later stage.

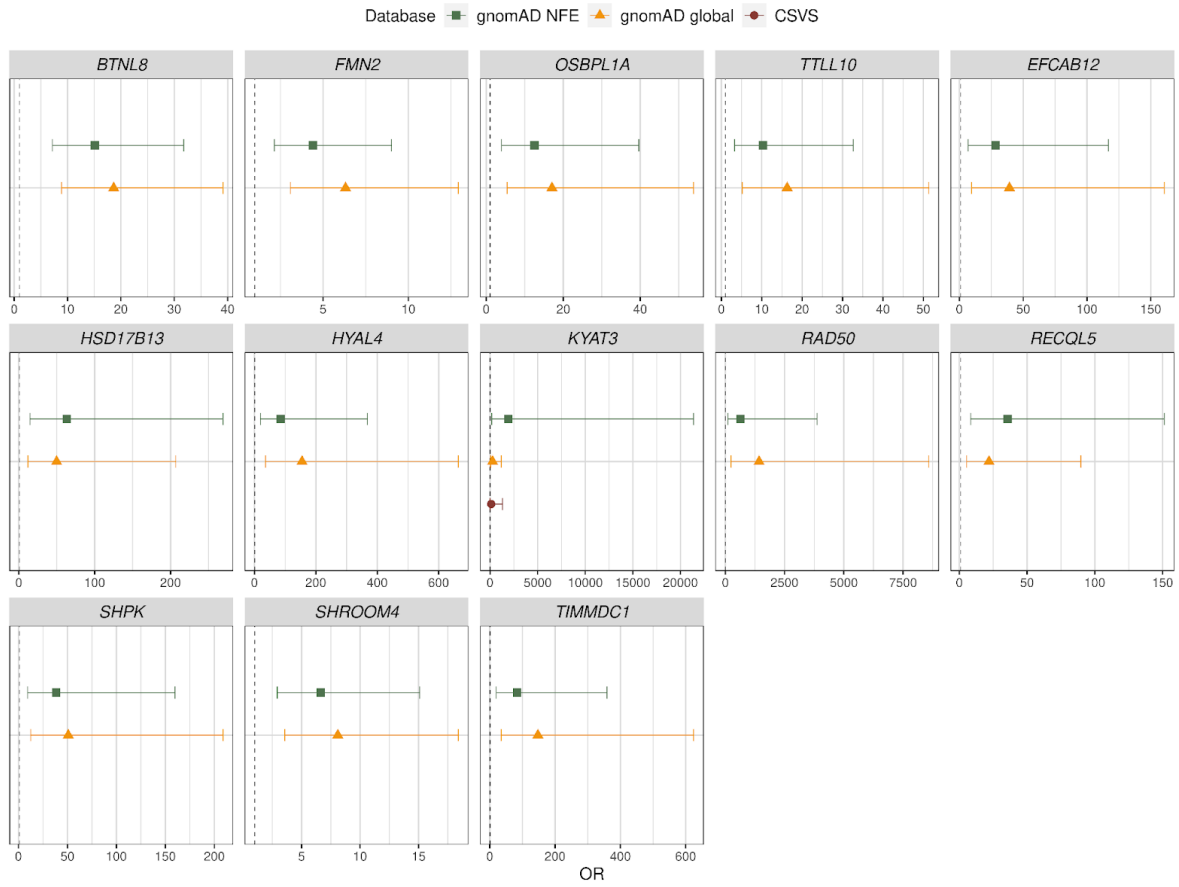

Figure 25 - Odds ratio (OR) of the genes enriched in variants with a predicted high confidence of being loss-of-function (HIGH HC), filtered by AF < 0.05, for the I4-GÜF.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants in them.

Figure 26 - Odds ratio (OR) of the genes enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20); filtered by AF < 0.05, for the I4-GÜF.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants in them.

5.3.5.2 Gene burden analysis: I1-GÜF

The I1-GÜF cohort was also studied, two different GBAs were performed (Figure 27).

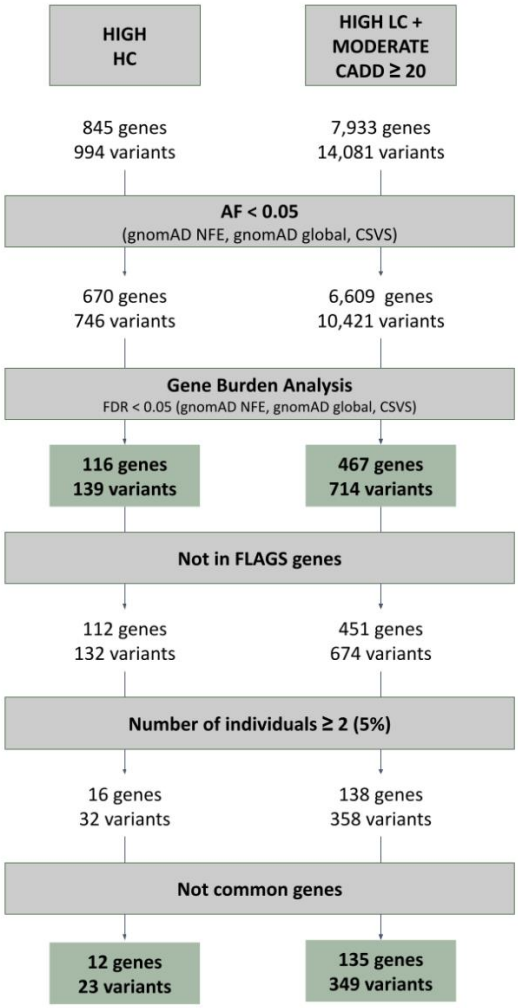

Figure 27 - Flow chart summarising the prioritisation strategy and the result of the gene burden analysis for I1-GÜF individuals, variants filtered by AF < 0.05.

HC: High confidence; LC: Low confidence; CADD: Combined Annotation Dependent Depletion; FDR: p-value corrected by false discovery rate; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; FLAGS: FrequentLy mutAted GeneS.

As mentioned above, the pointed genes for the HIGH HC analysis (Figure 28) and for the HIGH LC + MODERATE CADD ≥ 20 (Figure 29) of the I1-GÜF cluster, were identified to compare them with those obtained in the I1 group subsequently.

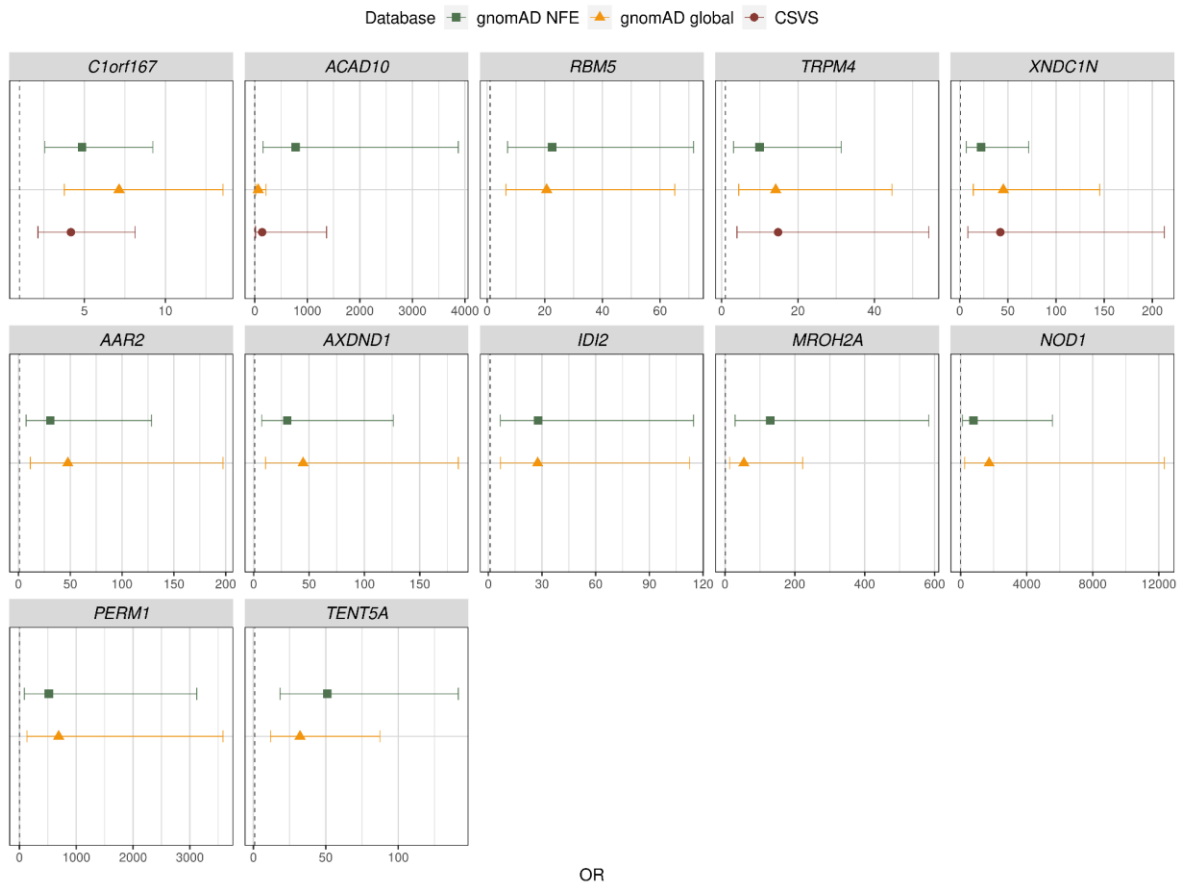

Figure 28 - Odds ratio (OR) of the genes enriched in variants with a predicted high confidence of being loss-of-function (HIGH HC), filtered by AF < 0.05, for the I1-GÜF.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants in them.

Figure 29 - Odds ratio (OR) of the genes enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20); filtered by AF < 0.05, for the I1-GÜF.

The OR were calculated against each reference population (gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population). Genes were ranked according to the number of individuals with variants in them.

5.3.6 Genes with novel variants

The first step in the GBA was to filter by AF less than 0.05 in the three reference populations (gnomAD NFE, gnomAD global and CSVS). In some genes, all the retained variants were novel for the three populations, which means that in any of them it was found an individual with the variant. Those genes could not be used for the GBA, as all their values would have been infinite.

5.3.6.1 Genes with novel variants in I4

In the analysis of variants with a high confidence of being LoF, 269 genes had all the variants annotated as novel and 258 of these genes only had one variant. In the analysis of variants with a low confidence of being LoF and variants with a moderate impact in the protein, 700 genes had all the variants annotated as novel and 645 of them with only one variant.

Furthermore, in these genes, novel variants present in two or more individuals in the I4 subgroup were further analysed. Four different variants were identified following these criteria: novel variants in the three reference populations and being in two or more individuals. The variant chr19:49365508G>C in the *DKKL1* gene, was the only one with a high impact in the protein and with a high confidence of being LoF, concretely it was a splice acceptor variant. Besides, it was predicted to be deleterious regarding the elevated CADD score. Three missense variants were also found in the *SETDB1*, *OR13D1* and *CNGA2* genes (Table 19).

Table 19 - Single nucleotide variants (SNVs) in genes with all novel variants.

Gene symbol	Variant	Amino acid change	Impact	Consequence	CADD	AF I4	Individuals
DKKL1	chr19:49365508G>C	-	High	Splice acceptor	32	0.013	I4-63, I4-12
SETDB1	chr1:150959220G>T	M792I	Moderate	Missense	26.5	0.013	I4-69, I4-68
OR13D1	chr9:104695248A>T	H244L	Moderate	Missense	22.8	0.013	I4-47, I4-72
CNGA2	chrX:151744098T>C	M532T	Moderate	Missense	22.4	0.013	I4-71, I4-32

CADD: Combined Annotation Dependent Depletion; ACMG: American College of Medical Genetics and Genomics; LP: Likely pathogenic; US: Uncertain significance; LB: Likely benign; AF: Allele frequency.

The *DKKL1* gene was not constrained based on the high LOEUF score and the pLI equal to zero (Table 20).

Table 20 - Constraint of loss-of-function (LoF) variants in the DKKL1 gene.

Gene symbol	LOEUF	pLI
DKKL1	1.428	0

pLI: Probability of being loss-of-function intolerant; LOEUF: Loss-of-function observed/expected upper bound fraction.

For the missense novel variants, the *SETDB1* gene was the only constraint because of the lower ratio between observed missense variants in the gene and the expected based on the gnomAD mutational model and the high value of the Z score (Table 21).

Table 21 - Constraint of missense variants in the SETDB1, OR13D1 and CNGA2 genes.

Gene symbol	o/e	Z score
SETDB1	0.586 (0.540-0.634)	4.008
OR13D1	1.102 (0.979-1.242)	-0.479
CNGA2	1.005 (0.909-1.111)	-0.028

o/e: Observed/expected ratio.

In addition, in the individuals with novel variants elevated THI values were observed, the ages of tinnitus onset were diverse and the hearing thresholds too. Both individuals with the splice acceptor variant in *DKKL1* had hyperacusis, anxiety (specially I4-63) and depression (more I4-12). The sole patient with available data for these questionnaires among those with novel missense variants had lower levels of hyperacusis, anxiety and depression (Table 22).

Table 22 - Summary of the studied questionnaire scores of the patients with the candidate variants.

Individual	Sex	Age	Tinnitus onset	PTA	4 kHz	8 kHz	THI	GÜF	VAS	PHQ-9	HADS-A	HADS-D
I4-63	F	76	53	40.00	50.00	60.00	88	22 (I3)	5 (I2)	2 (I1)	11 (I4)	7 (I2)
I4-12	F	50	32	39.17	35.00	35.00	72	16 (I3)	10 (I4)	14 (I4)	6 (I2)	10 (I3)
I4-69	M	80	-	-	-	-	94	-	-	-	-	-
I4-68	F	49	37	43.33	55.00	50.00	92	-	-	-	-	-
I4-47	F	56	-	-	-	-	82	-	-	-	-	-
I4-72	F	53	-	-	-	-	96	-	-	-	-	-
I4-71	F	66	60	39.17	60.00	82.50	94	13 (I2)	8 (I3)	2 (I1)	6 (I2)	2 (I1)
I4-32	F	47	38	42.50	42.50	55.00	76	-	-	-	-	-

M: Male; F: Female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory; VAS: Visual analogue scale of annoyance for tinnitus; GÜF: Hypersensitivity to sound; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale.

5.3.6.2 Genes with novel variants in I1

After filtering by variants with a high confidence of being LoF, 301 genes had all the variants novel and 288 of them only have one variant per gene. By low confidence of being LoF and with a moderate impact in the protein, 704 had all the variants novel and 673 of these genes had one variant. The novel variants found in two or more individuals in the I1 subgroup were studied more exhaustively. No variant with a high confidence of being LoF was detected in two or more individuals, whereas only one novel variant with the low confidence of being LoF and moderate impact in the protein filter was found. The chr19:55486810G>A missense variant in the *NAT14* gene identified in two individuals had a high CADD (Table 23).

Table 23 - Single nucleotide variants (SNVs) in genes with all novel variants.

Gene symbol	Variant	Amino acid change	Impact	Consequence	CADD	AF I1	Individuals
NAT14	chr19:55486810G>A	G159R	Moderate	Missense	27.4	0.011	I1-44, I1-67

CADD: Combined Annotation Dependent Depletion; ACMG: American College of Medical Genetics and Genomics; US: Uncertain significance; AF: Allele frequency.

According to the gnomAD mutational model, fewer missense variants than expected were observed in *NAT14*, which indicated a constraint of the gene (Table 24).

Table 24 - Constraint of missense variants in the SETDB1, OR13D1 and CNGA2 genes.

Gene symbol	o/e	Z score
NAT14	0.812 (0.677 - 0.890)	0.657

o/e: Observed/expected ratio.

Due to the THI value, individuals with the variant, although falling below the first quartile, still experience discomfort due to tinnitus. Moreover, I1-67 showed a bit of annoyance due to the hyperacusis and few symptoms of anxiety and depression (Table 25).

Table 25 - Summary of the studied questionnaire scores of the patients with the candidate variants.

Individual	Sex	Age	Tinnitus onset	PTA	4 kHz	8 kHz	THI	GUF	VAS	PHQ-9	HADS-A	HADS-D
I1-44	M	37	-	-	-	-	14	-	-	-	-	-
I1-67	F	53	45	40.83	55.00	60.00	18	8 (I1)	6 (I2)	6 (I2)	7 (I2)	5 (I2)

M: Male; F: Female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory; VAS: Visual analogue scale of annoyance for tinnitus; GÜF: Hypersensitivity to sound; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale.

5.4 COPY NUMBER VARIANTS AND STRUCTURAL VARIANTS ANALYSIS

5.4.1 Exome sequencing dataset

With the objective of including large structural variants in the database containing Exome Sequencing data, the CNV and SV were retrieved, filtered and added to the dataset. Figure 30 summarises the workflow taken to generate CNV and SV files.

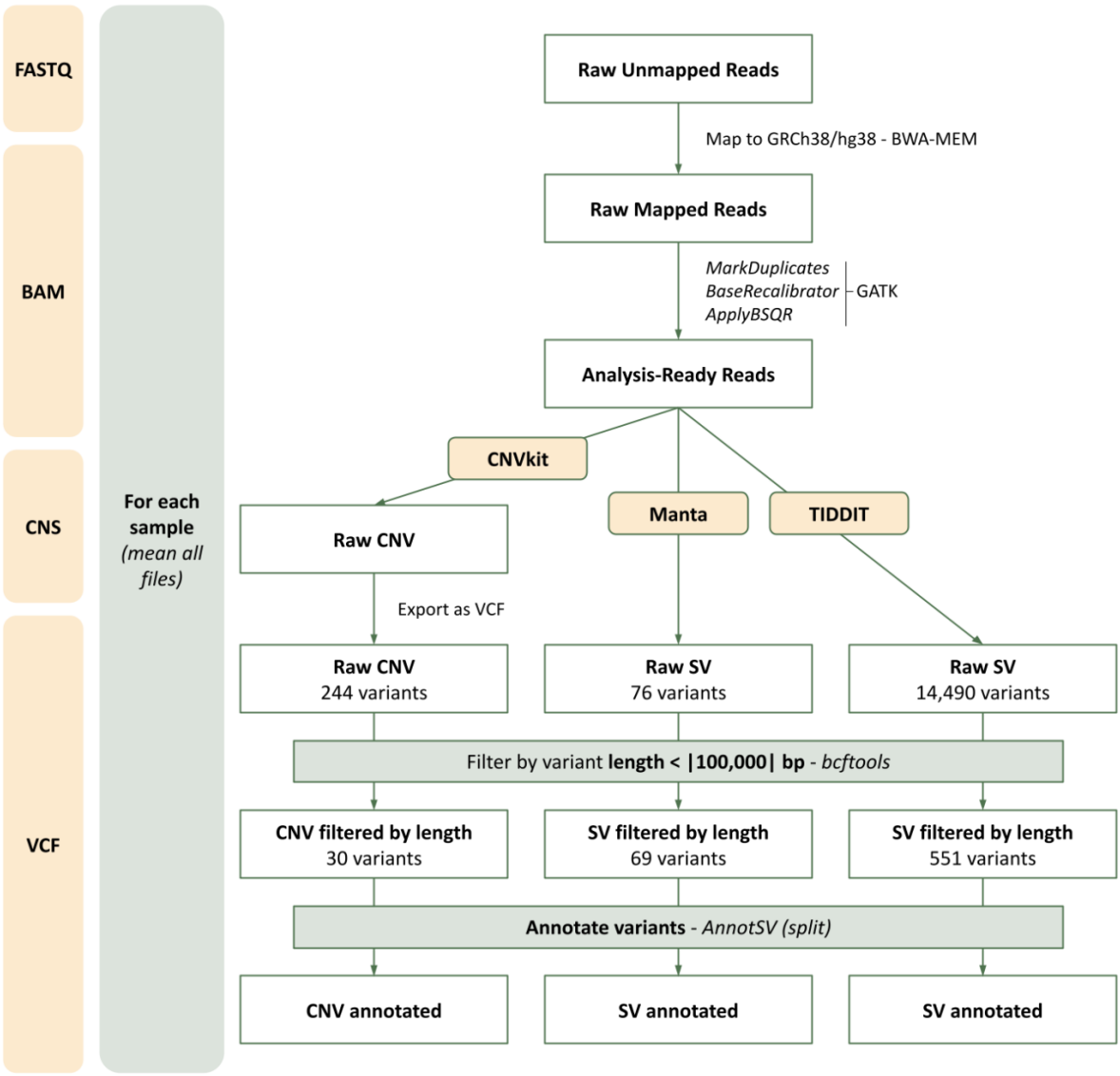

Figure 30 - Workflow to create the exome sequencing dataset containing copy number variants (CNVs) and structural variants (SVs).

The diagram shows each step followed to create the genetic dataset. BAM: Binary Alignment Map; CNS: Segmented log2 ratios; VCF: Variant Call Format; BWA-MEM: Burrows-Wheeler Aligner - MEM; GATK: Genome Analysis Toolkit; bp: base pair.

The BAM files obtained in the previous analysis were used to call the CNV and SV. A mean of 244 CNV were obtained in each VCF with CNVkit, 76 SV with Manta and 14,490 SV with TIDDIT. All the variants larger than 100,000 bp were filtered to avoid false positives. After this step, 30 CNV from CNVkit, 69 SV from Manta and 551 SV from TIDDIT were annotated.

5.4.2 Tinnitus Handicap Inventory (THI) subgroups

To determine genes with CNV or SV related to tinnitus, the variants of the individuals in the subgroups obtained by ranking their THI score (Figure 9) were studied. As in the SNV and short indels analysis, the principal cohort composed by I4 individuals with severe tinnitus (THI score > 68, N = 75) and the I1 subgroup (THI score ≤ 24, N = 88) were studied.

5.4.2.1 Copy number variants and structural variants: I4

The annotated variants were prioritised to find those regions and the genes covered by them, related to the individuals from the I4 subgroup, therefore with a worse progression of the tinnitus (Figure 31).

After the filter by pathogenicity, following the ACMG criteria, the number of candidate variants decreased: 26 from CNVkit and 48 from Manta; especially with the TIDDIT results, with 346 variants. To select regions for I4 individuals, the first step was determining CNV and SV with overlapping positions in more than 60% of each region. From this point, all the VCF from the I4 subgroup were merged into one. Then, 462 variants were identified from CNVkit results, 459 from Manta and 3,888 from TIDDIT. After filter variants covering low constrained genes that usually accumulate variants and singleton variants (only found in one individual); the number was reduced to 203, 193 and 1,386 variants - for CNVkit, Manta and TIDDIT, respectively. Finally, as the main objective in this work is to point candidate genes for each subgroup, those genes covered by variants in both subgroups (I1 and I4) were discarded.

Hence, at the end of this analysis with the I4 individuals, 55 variants in 49 samples covering 128 genes were determined in CNVkit, 19 variants in 33 samples involving 32 genes were identified in Manta, and 181 variants in 75 samples affecting 183 genes were defined in TIDDIT.

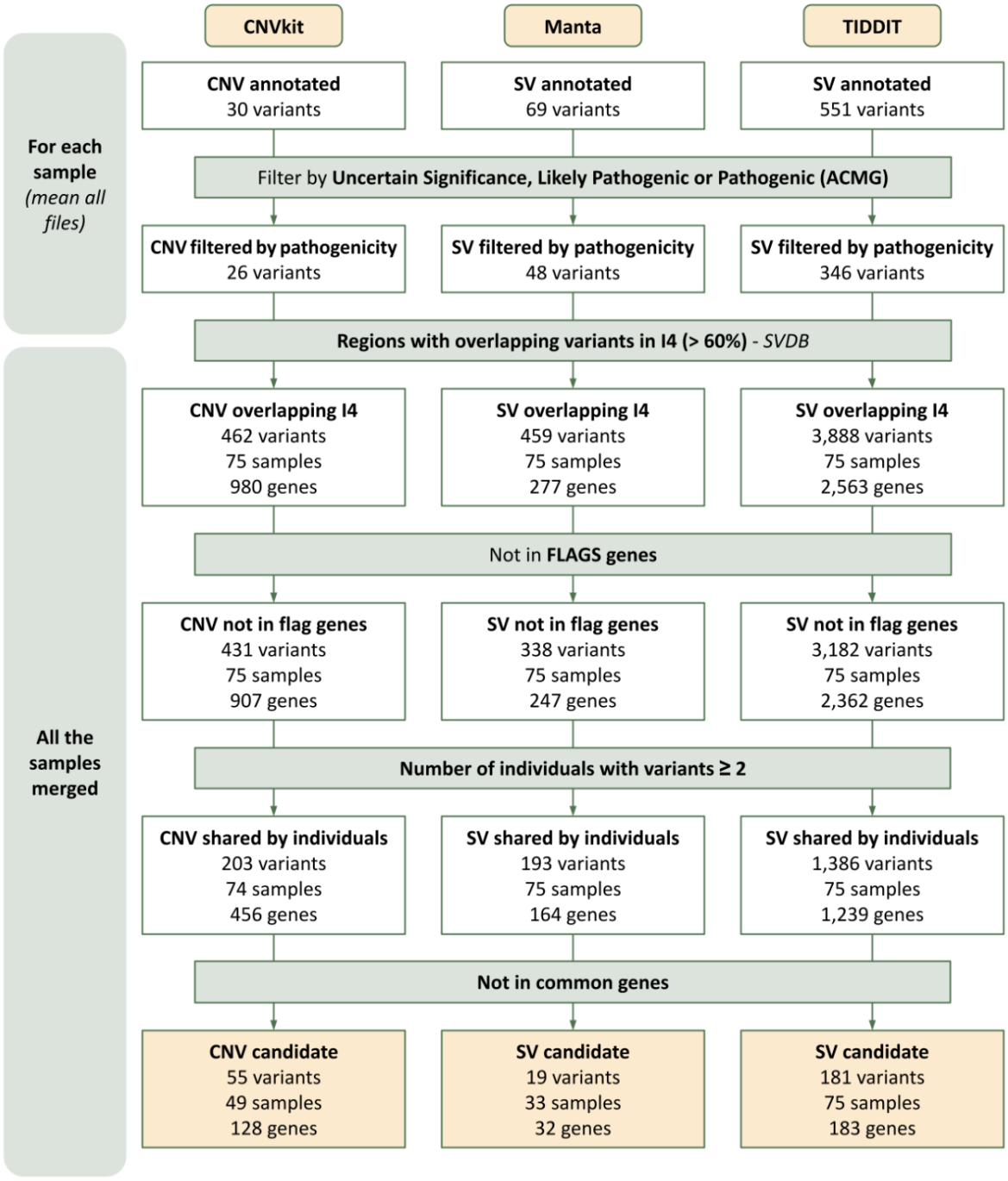

Figure 31 - Flow chart summarising the prioritisation strategy of the copy number variant (CNV) and structural variant (SV) analysis for the I4.

Variants were obtained with the CNVkit, Manta and TIDDIT tools. ACMG: American College of Medical Genetics and Genomics; FLAGS: FrequentLy mutAted GeneS.

5.4.2.2 Copy number variants and structural variants: I1

Similarly, annotated variants of the I1 individuals (low THI score) were prioritised to identify those genes related with a protection against tinnitus (Figure 32).

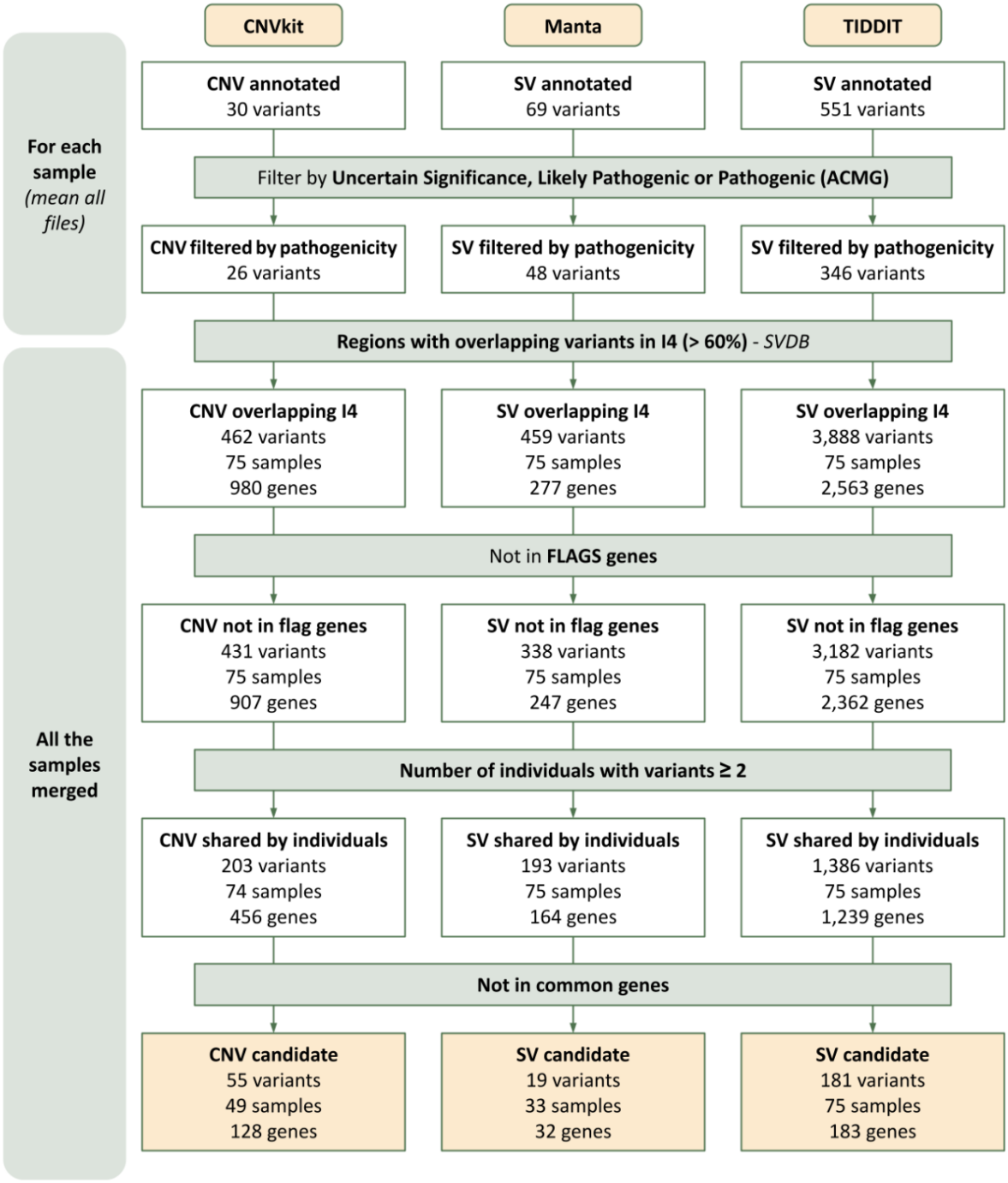

Figure 32 - Flow chart summarising the prioritisation strategy of the copy number variant (CNV) and structural variant (SV) analysis for the I1.

Variants were obtained with the CNVkit, Manta and TIDDIT tools. ACMG: American College of Medical Genetics and Genomics; FLAGS: FrequentLy mutAted GeneS.

Overlapping variants in the I1 individuals were retrieved and filtered by the ACMG criteria, as in the I4 analysis. A total of 511 regions from CNVkit, 457 from Manta and 4,553 from TIDDIT were selected. After filtering by FLAGS genes and several individuals with variants, 194 variants in the CNVkit path, 200 in the Manta and 1,387 in the TIDDIT were kept.

At the end of this analysis, 48 variants in 54 samples involving 107 genes were found with the CNVkit results, 13 variants in 31 samples covering 13 genes were identified in the Manta results, and 194 variants in 88 samples affecting 195 genes were selected from the TIDDIT results.

5.5 HEARING LOSS AND TINNITUS

The second specific objective of this thesis is to demonstrate the pleiotropic effect of rare variants in hearing loss and tinnitus phenotypes in MD by GBA.

To pinpoint genes showing an overload on rare variants associated with hearing loss and tinnitus, the results of the two GBAs performed in the whole MD cohort (Figure 17), in the I4 subgroup (Figure 18) and the I1 subgroup (Figure 21), were compared. The steps followed to prioritise genes after the GBA were the same for the three cohorts except for the last step, which was the filter to remove genes shared by I4 and I1 - those genes enriched in both cohorts -, to keep specific genes for each cohort. Then, the genes to perform the current analysis (enrichment of variants for hearing loss and tinnitus) were selected, those from the same step were chosen, which was the last step for the whole cohort and the genes before filtering by common genes in the subgroups.

First, genes enriched in variants with a high impact in the protein and high confidence of being LoF in the three cohorts were compared (Figure 33).

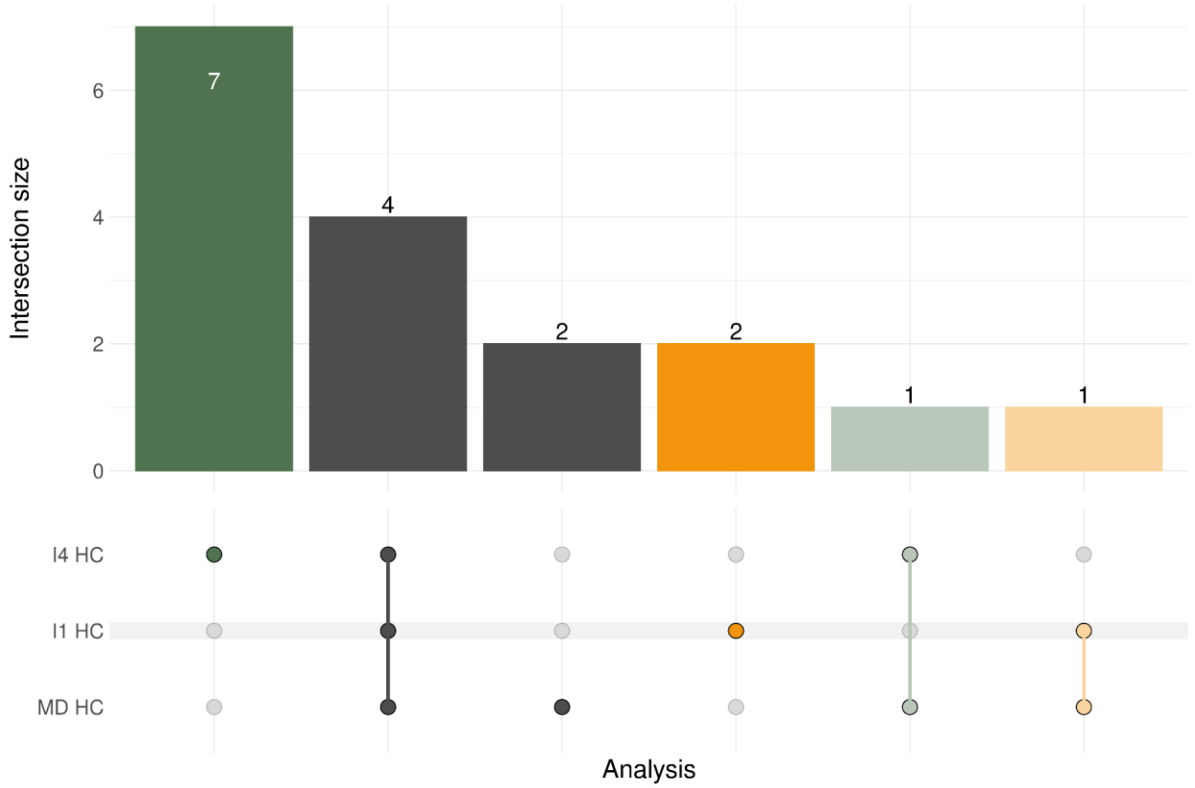

Figure 33 - Intersection of enriched genes in variants with a predicted high confidence of being loss-of-function (HC) for the I4, I1 and MD (Meniere Disease) individuals.

The *PTTG2* gene was shared between the whole MD cohort and I4 subgroup, which implied that this gene could be associated with both hearing loss and severe tinnitus. Moreover, the *ADAM2* gene was enriched in both MD and I1, which was expected to be involved in hearing loss and in a protection against severe tinnitus. In addition, by this intersection, seven genes were exclusive for the I4 subgroup and two for the I1 subgroup.

Regarding the genes enriched in variants with a low confidence of being LoF and with a moderate impact in the protein and predicted to be deleterious, the same comparisons were done with the three cohorts (Figure 34).

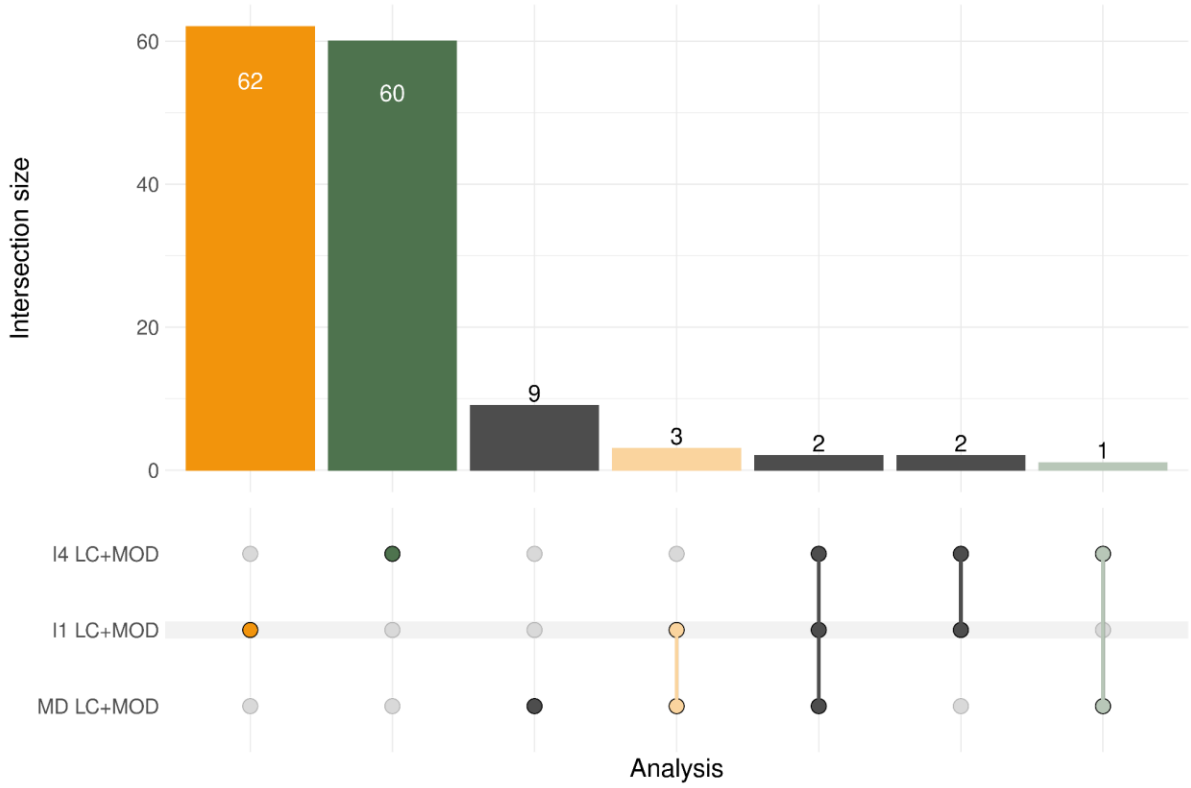

Figure 34 - Intersection of enriched genes in variants with a low confidence of being loss-of-function and moderate impact in the protein (mostly missense variants) and predicted to be deleterious (LC + MOD) for the I4, I1 and MD (Meniere Disease) individuals.

In this case, only one gene was identified as associated with hearing loss and worsen tinnitus: *CNTNAP2*. Besides, three genes were relevant for the hearing loss and any or less disturbance caused by tinnitus: *FRYL*, *FLT4* and *SCN10A*. For this type of variant, 60 genes were restricted to I4 and 62 to I1 subgroups.

Interestingly, none of the genes enriched in LoF variants or variants with moderate impact in the protein in the various subgroups were shared (Figure 35).

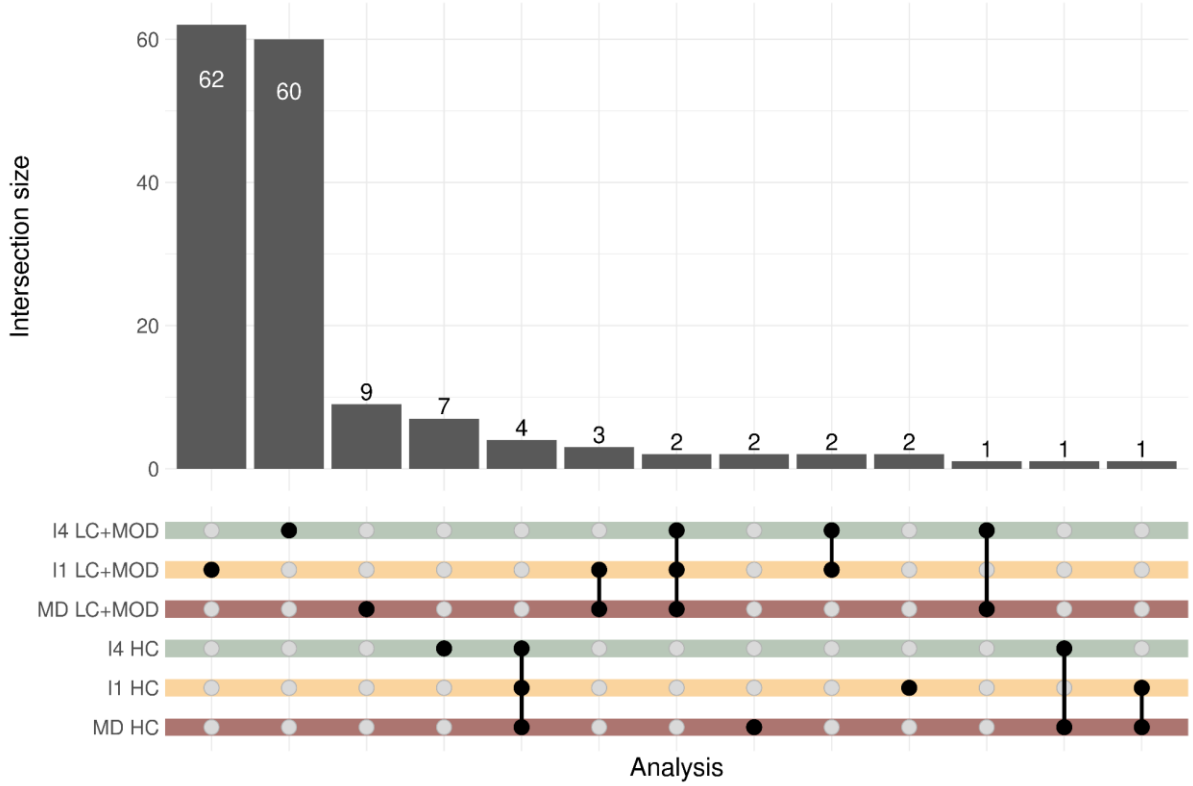

Figure 35 - Intersection of enriched genes in variants with a predicted high confidence of being loss-of-function (HC), and in variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (LC + MOD); for the I4, I1 and MD (Meniere Disease) individuals.

5.5.1 *PTTG2* and *CNTNAP2* genes

In the *PTTG2* gene, the same frameshift previously described in eight individuals from the I4 subgroup was found in another 12 individuals from the rest of the whole cohort: five for the I1 subgroup, five from the I2 and two from the I3 (Table S17).

Thirteen missense variants were identified in the *CNTNAP2* gene in the whole MD cohort, six of them were previously described for the I4 subgroup. The variants were carried by a total of 17 individuals: four for the I1 subgroup, one for the I2, five for the I3 and seven for the I4 (Table S18).

5.5.2 *ADAM2, FRYL, FLT4* and *SCN10A* genes

The same four LoF variants previously described in the I1 subgroup were identified in the whole MD cohort: two splice acceptor, one frameshift and one stop gain. In addition to the six individuals from the I1 subgroup, another one from the I2 subgroup, three from the I3 and five from the I4 (Table S19).

In the *FRYL* gene, 21 missense variants were found, six of them were previously described. They were carried by eight individuals from the I1 subgroup, eight from the I2, nine for the I3 and seven from the I4. Six of the 14 variants found in the whole MD cohort were previously described for the I1 subgroup. They were seven individuals from I1, one from I2, six from I3 and five from I4. Finally, 16 missense variants were described in the *SCN10A* gene, five of them were described in the I1 analysis. They were present in 13 individuals from the I1 subgroup, seven from the I2, eight for the I3 and six from the I4 (Table S20).

5.6 TINNITUS AND HYPERACUSIS

Due to the strong association between tinnitus and hyperacusis, the impact of rare variants in specific genes in both phenotypes by GBA was studied. To determine genes involved in both traits, the results of the two GBAs carried out with both subgroups obtained by THI (I4 and I1, Figures 18 and 21) and with both subgroups obtained by GÜF (I4-GÜF and I1-GÜF, Figures 24 and 27) were contrasted.

The GBAs's results were compared between the individuals from the fourth interval by THI against the individuals separated by GÜF (Figure 36). The *RAD50* gene was shared for the GBA of high confidence LoF variants. Moreover, eleven genes for the low confidence LoF variants and variants with moderate impact in the protein and predicted to be deleterious were shared between both subgroups: *KAT6A, SETD1A, PPL, FURIN, BIRC6, PKN3, HOMER2, FAM107A, HDAC4, RNF175* and *C12orf56.*

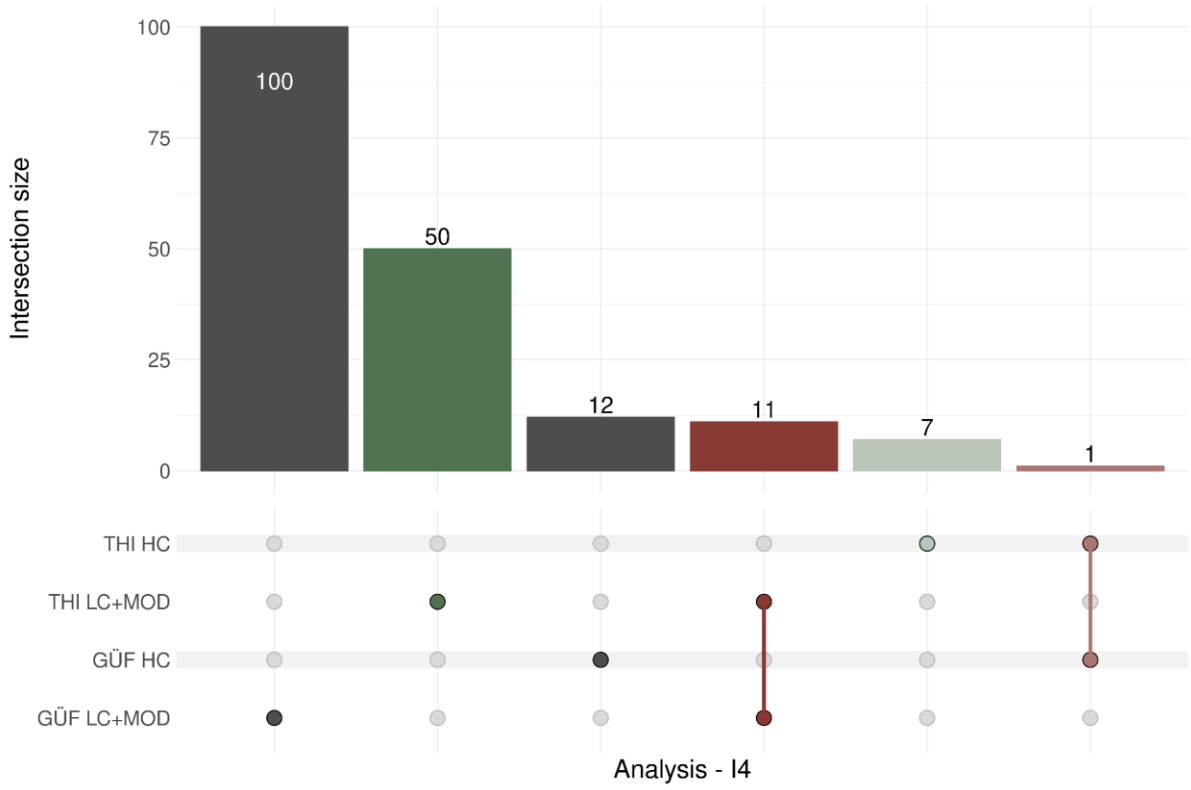

Figure 36 - Intersection of enriched genes in variants from individuals of the I4 grouped by the Tinnitus Handicap Inventory (THI) score, compared with individuals in the I4 grouped by the questionnaire of hypersensitivity to sound (GÜF) score.

HC: High confidence of being loss-of-function; LC+MOD: Low confidence of being loss-of-function and moderate impact in the protein (mostly missense variants) and predicted to be deleterious.

The same procedure was followed with the subgroups of individuals in the first interval, comparing the results of THI versus GÜF (Figure 37). In this case, no common genes were found between both subgroups for the GBA performed with high confidence LoF variants; and eleven genes for the analysis with low confidence LoF and deleterious moderate variants: *CFTR, INTS1, TRANK1, TLN1, DCTPP1, ATP10D, AP1AR, PLXNA3, PTPN18, ZNF626* and *HACD3.*

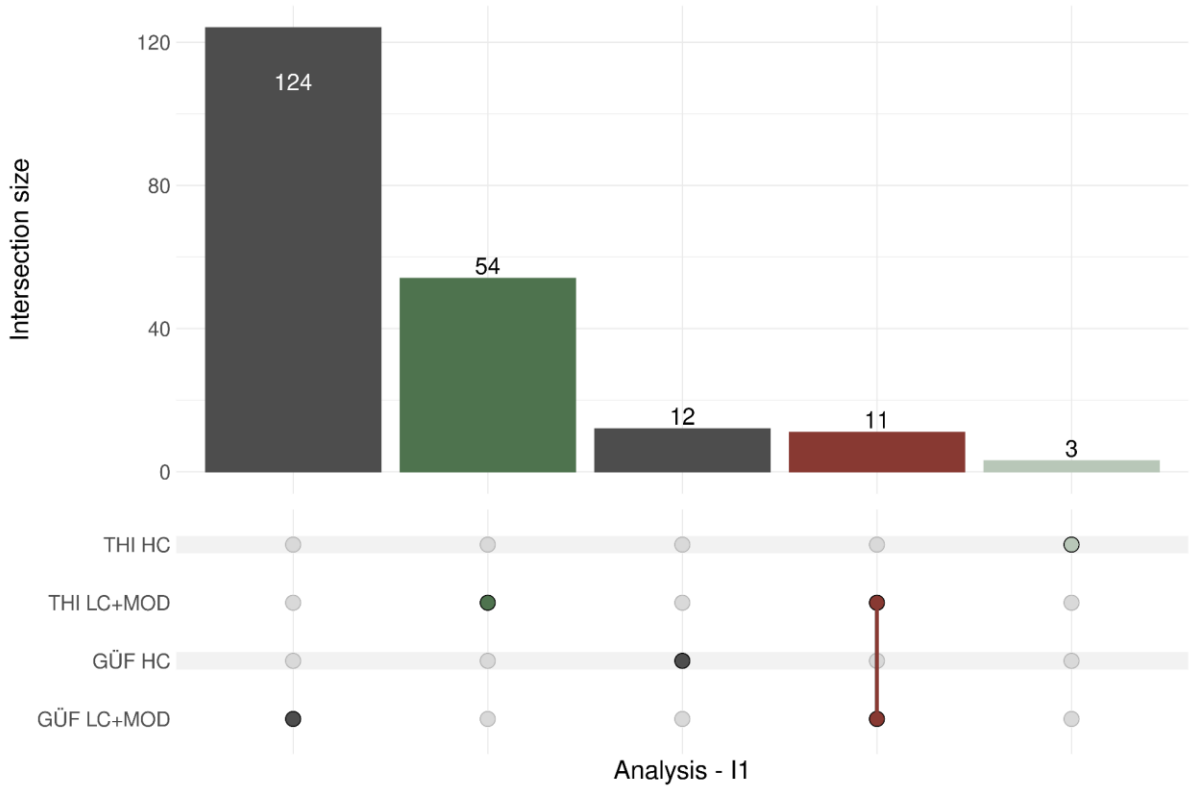

Figure 37 - Intersection of enriched genes in variants from individuals of the I1 grouped by the Tinnitus Handicap Inventory (THI) score, compared with individuals in the I1 grouped by the questionnaire of hypersensitivity to sound (GÜF) score.

HC: High confidence of being loss-of-function; LC+MOD: Low confidence of being loss-of-function and moderate impact in the protein (mostly missense variants) and predicted to be deleterious.

In addition, for the analysis with low confidence and moderate variants, between the I4 of THI and the I1-GÜF two genes were shared: *PPFIBP1* and *RYR2;* and three genes between the I1 of THI and the I4-GÜF: *ITPR3, MMP9* and *FRYL*.

5.7 IDENTIFICATION OF GENES SHARING SINGLE NUCLEOTIDE VARIANTS AND STRUCTURAL VARIANTS IN THE THI SUBGROUPS

The enriched genes obtained in the HIGH HC and HIGH LC + MODERATE CADD ≥ 20 by GBA; and the targeted genes with CNVs and SVs from CNVkit, Manta and TIDDIT were studied in the groups I4 and I1 separately (Figure 38). Shared genes in I4 and I1 individuals were discarded for further analyses.

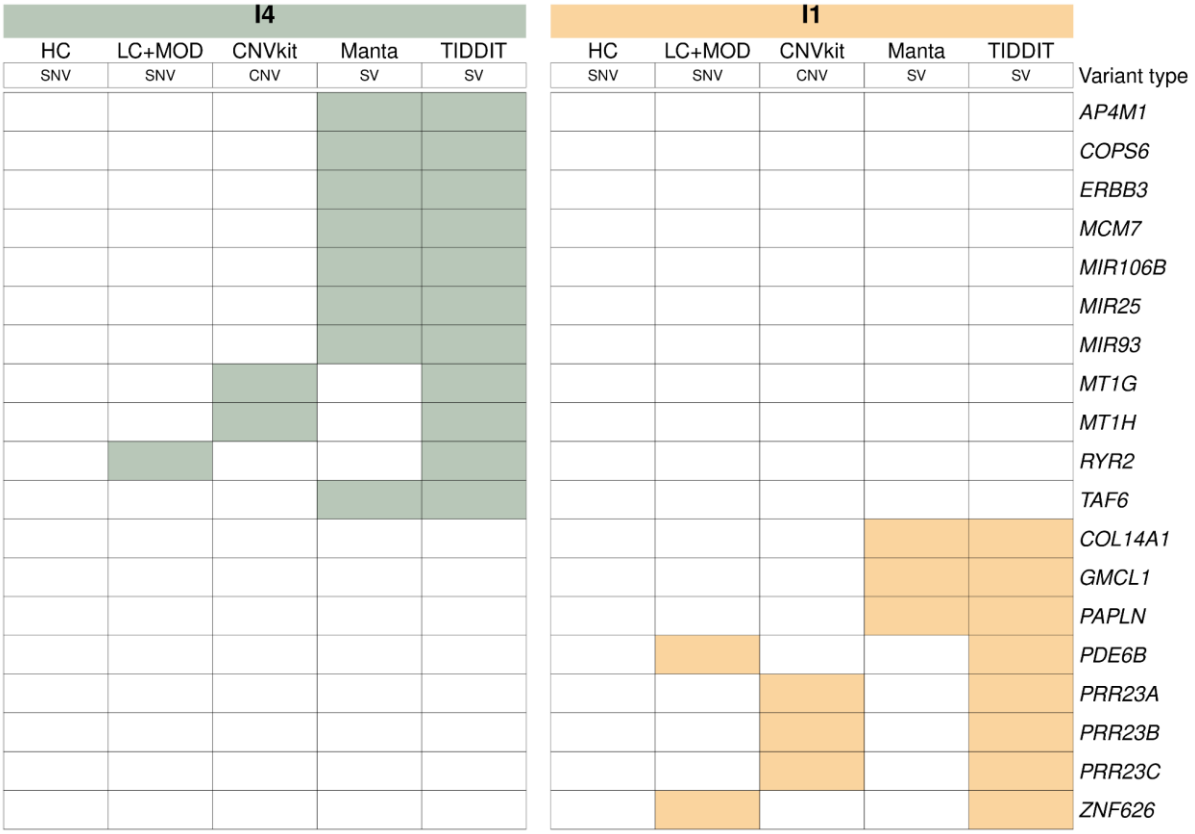

Figure 38 - Common genes for I4 and I1 subgroups.

Resulting from the gene burden analysis with variants high confidence of being loss-of-function (HC), GBA with variants with a low confidence of being loss-of-function and with a moderate impact in the protein and predicted to be deleterious (LC+MOD), copy number variants obtained with CNVkit, structural variants obtained with Manta and structural variants obtained with TIDDIT. SNV: single nucleotide variant; CNV: copy number variant; SV: structural variant.

The replication of genes - by following different workflows or using different tools - allowed for the shortlist of a few candidate genes.

As Figure 38 shown, regarding the I4 analyses, the higher number of replicated genes was found between Manta and TIDDIT. This was because both tools identified a duplication shared by two individuals covering seven genes. The first step was to validate the variants by IGV.

The *MT1G* and *MT1H* genes were discarded because it was not possible to validate the duplication found by CNVkit due to the big size (81 kb). Despite the four SNV found in *RYR2* were validated, the duplication reported by TIDDIT was in a tandem repeat region with low coverage, which also aligned in the chromosomes 4 and 6, suggesting that it might not be a duplication but rather an alignment error.

In reference to the I1 analyses, it was not achievable the validation of the two different duplications found in the *PRR23A*, *PRR23B* and *PRR23C* genes by CNVkit (98 kb) and TIDDIT (38 kb) due to the substantial size. In addition, there was not enough information in IGV to validate the duplication reported by TIDDIT in *ZNF626*.

The validated variants were analysed in detail.

5.7.1 *AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93* and *TAF6* genes

Almost the same duplication was found in two unrelated individuals (I4-28, I4-37) using Manta and TIDDIT tools, covering the *AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93* and *TAF6* genes. Comparing the results obtained by both tools (Table 26), they varied on the start position (only one base of difference) and in the end (70 bases of difference). Despite the discrepancy between the tools, which could be caused by the distinct criteria used to define a SV, the main of the SV covered the same region of the genome. Moreover, both tools found it in the same individuals and classified it as a duplication. The sequence was validated in IGV for both samples (Figure S1).

Table 26 - Summary of structural variants (SV) found in AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93 and TAF6 genes.

Tool	Chr	Start	End	Length	SV type	Individuals	Genes symbol	ACMG
Manta	7	100089053	100112257	23204	Duplication	I4-28, I4-37	*AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93, TAF6*	US
TIDDIT	7	100089054	100112327	23273	Duplication	I4-28, I4-37	*AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93, TAF6*	US

Chr: Chromosome; ACMG: American College of Medical Genetics and Genomics; US: Uncertain significance.

The pathogenicity of this duplication was uncertain based on the ACMG criteria (Table S21). The presence of this variant was checked in the reference populations. In the Spanish population from SPACNACS it was not found any SV covering the region; neither in the

gnomAD dataset, where only a shorter (280 bp) insertion present in an African control was encountered.

From the seven genes found in which the CNV was found, only *AP4M1, COPS6, MCM7* and *TAF6* encoded for a protein. Regarding the constraint of those genes (Table 27), the gene with the greatest constraint - the most intolerant to variants in it - was *COPS6,* because the LOEUF score was lower than 0.35 and the pLI score was very close to 1.

Table 27 - Constraint of AP4M1, COPS6, MCM7 and TAF6 genes.

Gene symbol	LOEUF	pLI
AP4M1	1.313	0
COPS6	0.272	0.987
MCM7	1.279	0
TAF6	0.784	0

pLI: Probability of being loss-of-function intolerant; LOEUF: Loss-of-function observed/expected upper bound fraction.

Moreover, the I4-28 and I4-37 individuals were women and the THI score of both was 76. It was observed that the hearing in individual I4-37 was more impaired and the rest of the questionnaire's data was not available for them (Table 28).

Table 28 - Summary of the clinical data of the patients with the candidate variants.

Individual	Sex	Age	PTA	4 kHz	8 kHz	THI
I4-28	F	72	51.67	67.50	65.00	76
I4-37	F	76	76.67	80.00	92.50	76

F: Female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory.

5.7.2 *ERBB3* gene

Three deletions and three duplications were found in the same region of the *ERBB3* gene in two individuals (I4-40 and I4-41) by Manta and TIDDIT, respectively (Table 29). Examining the findings obtained by both tools, they found the three SVs in almost the same region, but there was a discrepancy regarding the type of variant, because Manta identified the SVs as deletions and TIDDIT as duplications. Regarding this inconsistency, it was necessary to check

the two samples in IGV (Figure S2) and three deletions were noted covering the reported position by both tools.

Table 29 - Summary of structural variants (SV) found in ERBB3 gene.

Tool	Chr	Start	End	Length	SV type	Individuals	Gene symbol	ACMG
Manta	12	56100028	56100172	144	Deletion	I4-40, I4-41	*ERBB3*	US
Manta	12	56100243	56101058	815	Deletion	I4-40, I4-41	*ERBB3*	US
Manta	12	56101359	56101526	167	Deletion	I4-40, I4-41	*ERBB3*	US
TIDDIT	12	56100031	56100173	142	Duplication	I4-40, I4-41	*ERBB3*	US
TIDDIT	12	56100248	56101058	810	Duplication	I4-40, I4-41	*ERBB3*	US
TIDDIT	12	56101365	56101527	162	Duplication	I4-40, I4-41	*ERBB3*	US

Chr: Chromosome; ACMG: American College of Medical Genetics and Genomics; US: Uncertain significance.

The pathogenicity of all the regions was uncertain following the ACMG criteria (Table S22). Besides, these variants were checked in the reference populations. No SV covering the region was found in the Spanish population from SPACNACS. However, in the gnomAD dataset two different deletions of 812 bases and 163 bases - overlapping almost the same regions than two of the reported in the studied cohort - were found in only one European individual (Table 30).

Table 30 - Summary of structural variants (SV) found in gnomAD dataset in overlapping positions of the candidate regions.

Reference dataset	Chr	Start	End	Length	SV type	Frequency	Number individuals	Population
gnomAD	12	56100246 (hg38), 56494030 (hg19)	56101058 (hg38), 56494842 (hg19)	812	Deletion	4.61E-05	1	European
gnomAD	12	56101363 (hg38), 56495147 (hg19)	56101526 (hg38), 56495310 (hg19)	163	Deletion	4.61E-05	1	European

Chr: Chromosome.

The constraint of the *ERBB3* gene is low because the LOEUF is 0.685 and the pLI 0.

The THI score of the I4-40 and I4-41 individuals was 80 in both cases. The rest of the clinical data was only available for I4-41, exhibiting significant hyperacusis, anxiety and depression (Table 31).

Table 31 - Summary of the clinical data of the patients with the candidate variants.

Individual	Sex	Age	Tinnitus onset	PTA	4 kHz	8 kHz	THI	GÜF	VAS	PHQ-9	HADS-A	HADS-D
I4-40	M	63	-	-	-	-	80	-	-	-	-	-
I4-41	F	57	52	44.17	50.00	55.00	80	24 (I4)	8 (I3)	22 (I4)	20 (I4)	16 (I4)

M: Male; F: Female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory; GÜF: Hypersensitivity to sound; VAS: Visual analogue scale; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale.

5.7.3 *COL14A1* gene

A duplication covering almost the same region of the *COL14A1* gene was reported by Manta and TIDDIT (Table 32). It was found in two individuals by Manta and in the same two and another four by TIDDIT. Based on the ACMG criteria, the pathogenicity of the variant was uncertain (Table S23). It was validated by IGV in all the samples (Figure S3).

Table 32 - Summary of structural variants (SV) found in COL14A1 gene.

Tool	Chr	Start	End	Length	SV type	Individuals	Gene symbol	ACMG
Manta	8	120280928	120281118	190	Duplication	I1-18, I1-53	COL14A1	US
TIDDIT	8	120280930	120281123	193	Duplication	I1-18, I1-19, I1-53, I1-58, I1-66, I1-70	COL14A1	US

Chr: Chromosome; ACMG: American College of Medical Genetics and Genomics; US: Uncertain significance.

Any variant covering this region was found in the Spanish reference population from SPACNACS. In the gnomAD database, a duplication overlapping practically the same region was identified in 186 individuals from different populations (Table 33).

Table 33 - Summary of structural variants (SV) found in gnomAD dataset in overlapping positions of the candidate regions.

Reference dataset	Chr	Start	End	Length	SV type	Frequency	Number individuals	Population
gnomAD	8	120280928 (hg38), 121293167 (hg19)	120281124 (hg38), 121293363 (hg19)	196	Duplication	8.67E-03	186	Latino (9), European (151), African (26)

Chr: Chromosome.

The *COL14A1* gene did not exhibit significant constraint derived from the LOEUF score (LOEUF = 0.535) and any constraint from the pLI equal to 0.

Three females and three males presented this duplication, with a range of ages for the report of the tinnitus onset. Even though all individuals had a THI score lower than the first quartile, it was interesting the different scores, from 0 to 20, which means a considerable annoyance caused by the tinnitus. Moreover, there were differences in the hearing thresholds and in the rest of the questionnaires (Table 34).

Table 34 - Summary of the clinical data of the patients with the candidate variants.

Individual	Sex	Age	Tinnitus onset	PTA	4 kHz	8 kHz	THI	GÜF	VAS	PHQ-9	HADS-A	HADS-D
I1-18	F	94	76	62.50	90.00	112.50	0	14 (I2)	-	12 (I3)	0 (I1)	14 (I4)
I1-19	M	72	-	52.50	70.00	82.50	0	9 (I1)	-	1 (I1)	5 (I1)	3 (I1)
I1-53	M	60	54	26.67	50.00	50.00	16	16 (I3)	-	4 (I2)	8 (I3)	10 (I3)
I1-58	M	67	52	52.50	77.50	75.00	16	-	-	-	-	-
I1-66	F	79	60	60.83	67.50	72.50	18	20 (I3)	6 (I2)	3 (I1)	8 (I3)	9 (I3)
I1-70	F	54	-	-	-	-	20	-	-	-	-	-

M: male; F: female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory; GÜF: Hypersensitivity to sound; VAS: Visual analogue scale; PHQ-9: Patient Health Questionnaire depression scale, HADS: Hospital Anxiety and Depression Scale.

5.7.4 *GMCL1* gene

In a nearly identical region of the *GMCL1* gene a deletion and a duplication were found by Manta and TIDDIT, respectively, in the same two individuals (Table 35). As previously, the variant was checked in IGV in both individuals and it was observed that it was a deletion (Figure S4). According to the ACMG criteria, the pathogenicity was uncertain (Table S24).

Table 35 - Summary of structural variants (SV) found in GMCL1 gene.

Tool	Chr	Start	End	Length	SV type	Individuals	Gene symbol	ACMG
Manta	2	69847687	69847797	110	Deletion	I1-19, I1-46	GMCL1	US
TIDDIT	2	69847693	69847798	105	Duplication	I1-19, I1-46	GMCL1	US

Chr: Chromosome; ACMG: American College of Medical Genetics and Genomics; US: Uncertain significance.

Interestingly, almost the same deletion was found in both reference databases, in one individual in SPACNACS and in 73 in gnomAD (Table 36).

Table 36 - Summary of structural variants (SV) found in the gnomAD dataset in overlapping positions of the candidate regions.

Reference dataset	Chr	Start	End	Length	SV type	Frequency	Number individuals	Population
SPACNACS	2	69847686 (hg38), 70074818 (hg19)	69847802 (hg38), 70074934 (hg19)	116	Deletion	9.71E-03	1	Spanish
gnomAD	2	69847691 (hg38), 70074823 (hg19)	69847797 (hg38), 70074929 (hg19)	106	Deletion	3.46E-03	73 (2 hom)	Latino (3), European (23), African (9)

Chr: Chromosome.

In accordance with the constraint scores, the *GMCL1* gene was not constrained (LOEUF = 0.75 and pLI = 0).

The I1-19 patient did not report any disturbance due to the tinnitus, the levels of hyperacusis, anxiety and depression were low. The I1-46 individual had eight years of evolution for the tinnitus and her THI score was 14 (Table 37).

Table 37 - Summary of the clinical data of the patients with the candidate variants.

Individual	Sex	Age	Tinnitus onset	PTA	4 kHz	8 kHz	THI	GÜF	VAS	PHQ-9	HADS-A	HADS-D
I1-19	M	72	-	52.50	70.00	82.50	0	9 (I1)	-	1 (I1)	5 (I1)	3 (I1)
I1-46	F	59	51	38.33	45.00	55.00	14	-	6 (I2)	-	-	-

M: Male; F: Female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory; GÜF: Hypersensitivity to sound; VAS: Visual analogue scale; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale.

5.7.5 *PAPLN* gene

In the *PAPLN* gene, two different duplications were found in two different individuals each of them by Manta and TIDDIT (Table 38). The SV overlapped in 198 bp. In this case, following the ACMG criteria both SV were benign (Table S25). In addition, both were validated by IGV (Figure S5).

Table 38 - Summary of structural variants (SV) found in PAPLN gene.

Tool	Chr	Start	End	Length	SV type	Individuals	Gene symbol	ACMG
Manta	14	73263837	73264037	200	Duplication	I1-6, I1-48	*PAPLN*	B
TIDDIT	14	73263839	73264057	218	Duplication	I1-1, I1-88	*PAPLN*	B

Chr: Chromosome; ACMG: American College of Medical Genetics and Genomics; B: benign.

In SPACNACS database, any SV was identified covering a part of these regions. In gnomAD, two different duplications were detected overlapping with the candidate variants in a similar length, in 11 and one individual (Table 39).

Table 39 - Summary of structural variants (SV) found in gnomAD dataset in overlapping positions of the candidate regions.

Reference dataset	Chr	Start	End	Length	SV type	Frequency	Number individuals	Population
gnomAD	14	73263815 (hg38) - 73730523 (hg19)	73264058 (hg38) - 73730766 (hg19)	243	Duplication	5.11E-04	11	Latino (2), European (9)
gnomAD	14	73263838 (hg38) - 73730546 (hg19)	73264011 (hg38) - 73730719 (hg19)	173	Duplication	4.64E-05	1	East Asian

Chr: Chromosome.

The LOEUF score of the *PAPLN* gene was 1.141 and the pLI equal to zero, which denoted no constrain in the gene. The individuals I1-6 and I1-1 did not document any annoyance caused by tinnitus. The individual I1-48, with a THI score of 14, has been suffering from tinnitus for 30 years, moreover he presented hyperacusis, anxiety and depression. The individual I1-88, with seven years of tinnitus progression, had a THI score of 24, which was the cut-off between I1 and I2 (Table 40).

Table 40 - Summary of the clinical data of the patients with the candidate variants.

Individual	Sex	Age	Tinnitus onset	PTA	4 kHz	8 kHz	THI	GÜF	VAS	PHQ-9	HADS-A	HADS-D
I1-6	M	67	-	-	-	-	0	-	-	-	-	-
I1-48	M	63	36	82.5	95	97.5	14	35 (I4)	3 (I1)	13 (I4)	6 (I2)	11 (I4)
I1-1	F	20	-	-	-	-	0	-	-	-	-	-
I1-88	M	73	66	45	67.5	72.5	24	12 (I2)	6 (I2)	4 (I2)	8 (I3)	3 (I1)

M: Male; F: Female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory; GÜF: Hypersensitivity to sound; VAS: Visual analogue scale; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale.

5.7.6 *PDE6B* gene

The *PDE6B* gene was the only one with an enrichment in variants described by GBA and a validated SV. The burden was observed in variants with a low confidence of being LoF and with a moderate impact in the protein. Five missense variants in five different individuals were identified (Table 41). The chr4:662584G>A variant was pathogenic and it was not found in any reference population, the ACMG criteria for chr4:653900G>A described that its significance was uncertain with some pathogenic evidence, chr4:655980C>G and chr4:667929C>T were uncertain and chr4:654132G>A likely benign (Table S26). All the variants were validated by IGV.

Table 41 - Summary of single nucleotide variants (SNV) found in PDE6B gene.

Variant	Amino acid change	Consequence	CADD	ACMG	AF				Individuals
					gnomAD NFE	gnomAD	CSVS	I1	
chr4:653900G>A	E254K	Missense	31	US*	1.08E-04	6.30E-05	0	5.68E-03	I1-63
chr4:654132G>A	G302D	Missense	25.5	LB	2.63E-04	4.05E-04	1.00E-03	5.68E-03	I1-68
chr4:655980C>G	P345A	Missense	23.5	US	0	0	0	5.68E-03	I1-58
chr4:662584G>A	D600N	Missense	25.9	P	0	0	0	5.68E-03	I1-39
chr4:667929C>T	A809V	Missense	23	US	4.60E-05	2.10E-05	0	5.68E-03	I1-6

*CADD: Combined Annotation Dependent Depletion; ACMG: American College of Medical Genetics and Genomics; LP: Likely pathogenic; US: Uncertain significance; *: US with some pathogenic evidence; P: pathogenic; LB: likely benign; AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population.*

Besides, a deletion was found in the same gene in two individuals, one of them homozygous (Table 42). It was validated by IGV (Figure S6). The pathogenicity of this duplication was uncertain by the ACMG criteria (Table S27). It was not found in the reference datasets, but a bigger duplication overlapping in a region was detected in gnomAD in two African individuals.

Table 42 - Summary of structural variant (SV) found in PDE6B gene.

Tool	Chr	Start	End	Length	SV type	Individuals	Gene symbol	ACMG
TIDDIT	4	635416	635529	113	Deletion	I1-23 (het), I1-75 (hom)	*PDE6B*	US

Chr: Chromosome; ACMG: American College of Medical Genetics and Genomics; US: Uncertain significance; het: heterozygous; hom: homozygous.

The *PDE6B* gene was not constrained for missense due to the negative Z score and the ratio between observed and expected missense variants in the gene, being close to 1 (Z score = -0.798, o/e = 1.098 [1.025-1.177]). In the same way, it was not constrained for LoF variants (LOEUF = 1.187, pLI = 0).

The individuals with available data showed an age of onset between 52 and 61 years old, the range of THI was from zero to 22 and the individuals with questionnaire data reported low levels of hyperacusis, anxiety and depression (Table 43).

Table 43 - Summary of the clinical data of the patients with the candidate variants.

Individual	Sex	Age	Tinnitus onset	PTA	4 kHz	8 kHz	THI	GÜF	VAS	PHQ-9	HADS-A	HADS-D
I1-63	F	62	52	30.83	45.00	52.50	18	6 (I1)	3 (I1)	5 (I2)	-	-
I1-68	M	60	61	38.33	57.50	62.50	20	10 (I1)	3 (I1)	6 (I2)	3 (I1)	3 (I1)
I1-58	M	67	52	52.50	77.50	75.00	16	-	-	-	-	-
I1-39	F	72	-	-	-	-	12	-	-	-	-	-
I1-6	M	67	-	-	-	-	0	-	-	-	-	-
I1-23	M	57	-	15.83	32.50	25.00	0	-	-	-	-	-
I1-75	F	63	57	23.33	12.50	27.50	22	-	-	-	-	-

M: Male; F: Female; PTA: Pure-tone audiometry; Hz: Hertz; THI: Tinnitus Handicap Inventory; GÜF: Hypersensitivity to sound; VAS: Visual analogue scale; PHQ-9: Patient Health Questionnaire depression scale; HADS: Hospital Anxiety and Depression Scale.

5.8 EXPRESSION OF CANDIDATE GENES IN BRAIN, COCHLEA AND SPIRAL GANGLION NEURON

The expression of the candidate genes obtained in the previous analyses was studied using three different databases. To study the expression in brain tissues GTEx database with data from humans was utilised. The RNA-seq data from P0 mice from gEAR database were obtained to study the expression in hair cells and non-hair cells from the cochlea. The expression of the SGN from P0 mice was extracted from the auditory and vestibular ganglion neurons from the SHIELD database. The candidate genes were: the genes enriched in the HIGH HC analysis, the top genes resulting from the LC + MOD analysis and the genes sharing SNV and SV. The genes shared between the GBA with the whole MD cohort and the GBA with the I4 and I1 subgroups were ruled out as candidate genes. For the mouse databases, the orthologous of each gene was used.

5.8.1 Expression of I4 candidate genes

The heatmap (Figure 39) shows the expression of the 28 candidate genes for the subgroup I4 according to the THI. The genes were clustered based on their expression; it was postulated that genes highly expressed in some tissues could fulfil an important role in that tissue. Four clusters were calculated containing different numbers of genes. Cluster 1 included seven genes, which exhibited almost no or negligible expression in the brain, cochlea or SGN; except for the *LPIN3* gene, which was slightly expressed in the brain and the *WDFY4* gene in the cochlea and SGN. Cluster 2 comprised nine genes with low expression in the brain and medium expression in the cochlea and SGN. It was interesting the high expression of the *ERBB3* gene in the SGN and the strong expression of the *ERCC6* gene in the hair cells. Cluster 3 was formed by eight genes more expressed in the brain but less in the cochlea and SGN, the *SPTBN4* gene presented a strong expression in cerebellar hemisphere and cerebellum. Notably, the *ENDOG* and *ITGB5* genes had a high expression in the SGN; and the *UNC13C* and *PDZD7* genes were expressed in cochlea. Finally, cluster 4 consisted of four genes, three of them sharing SVs between Manta and TIDDIT, with the greatest expression in all tissues. Especially the *COP6* gene in the brain and SGN and *ARVCF* in the cerebellar hemisphere and cerebellum. It was remarkable that the expression of the cerebellar hemisphere and the cerebellum varies with respect to the rest of the tissues in most genes, often being greater. In addition, the cortex and frontal cortex expression were quite different, also in the spinal cord.

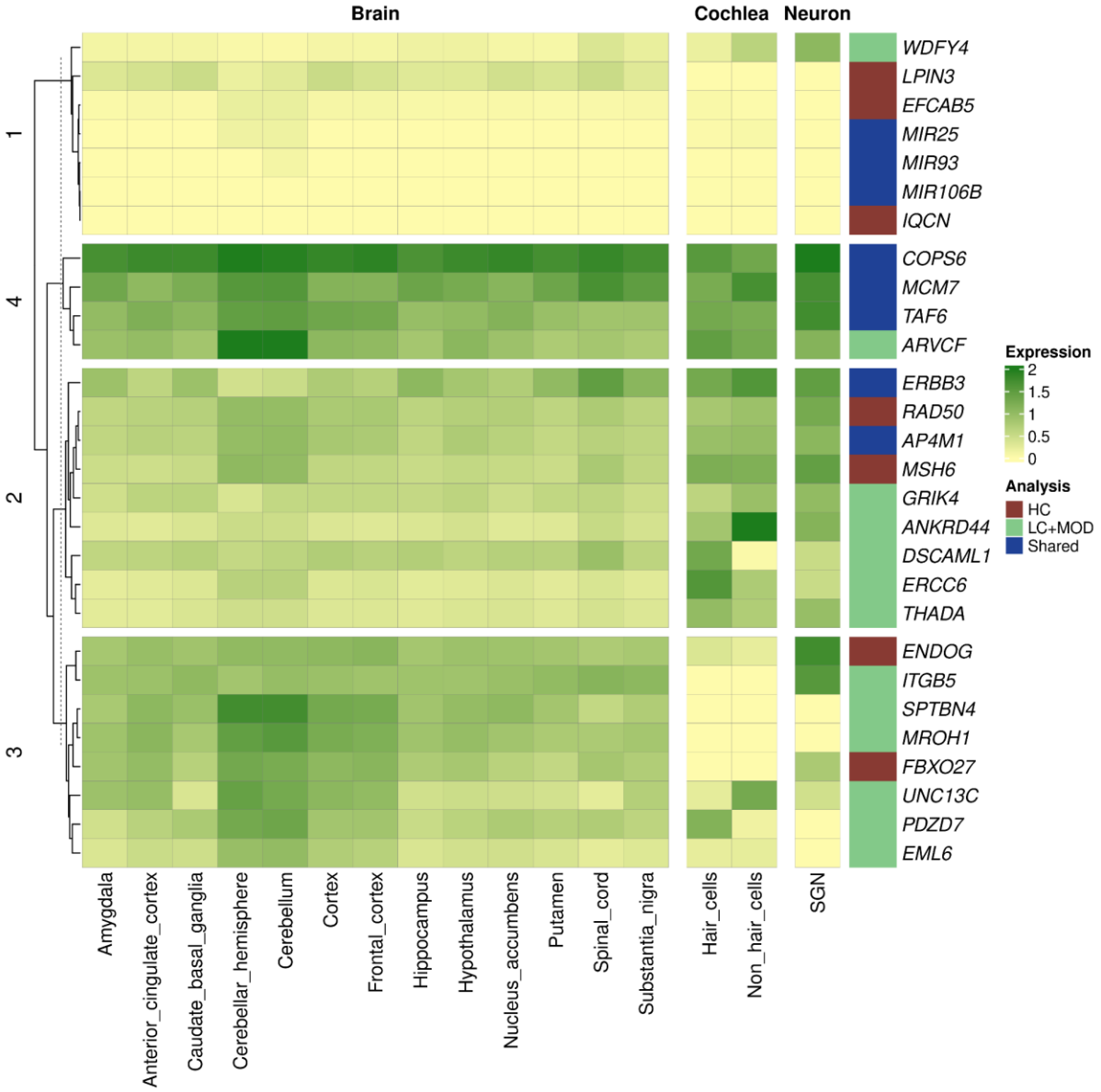

Figure 39 - Expression of the 28 candidate genes for the I4.

The heatmap shows the expression levels of candidate genes in 13 brain tissues obtained from GTEx in humans, the expression of the orthologous genes in the cochlear hair cells and non-hair cells from gEAR in postnatal day 0 (P0) mice and in the spiral ganglion neuron (SGN) from SHIELD in P0 mice. The expression for each dataset was normalised from 0 to 100, it was represented as the base-10 logarithm. Each gene was labelled by the analysis by which it was found: enrichment of variants with a high confidence of being loss-of-function (HC), enrichment of variants with a low confidence of being loss-of-function and with moderate impact in the protein predicted to be deleterious (LC+MOD), and shared between different analysis.

5.8.2 Expression of I1 candidate genes

Following the same methodology as with the I4 subgroup, the heatmap (Figure 40) depicts the expression of the 16 candidate genes for the subgroup I1, according to the THI. To compare these results against the expression of the candidate genes for I4, each dataset with the I1 genes

was normalised between 0 and 100, taking the value 100 as the maximum value for the same dataset with the I4 genes. When comparing the expression of the candidate genes for I4, it was observed that the I1 genes had lower expression in the brain tissues. Regarding the four clusters defined, cluster 1 was formed by four genes, containing the two obtained in the analysis of variants with a high confidence of being LoF, none of them were expressed in the tissues studied. Cluster 2 had two genes, without expression and low expression in the brain but highly expressed in SGN and cochlea, especially in non-hair cells. Cluster 3, composed of five genes, showed expression in the brain but there were differences in the expression in the cochlea and SGN. Cluster 4 was constituted by five genes, which were expressed in the brain, cochlea and SGN, aside from *CHD5* without expression in the cochlea.

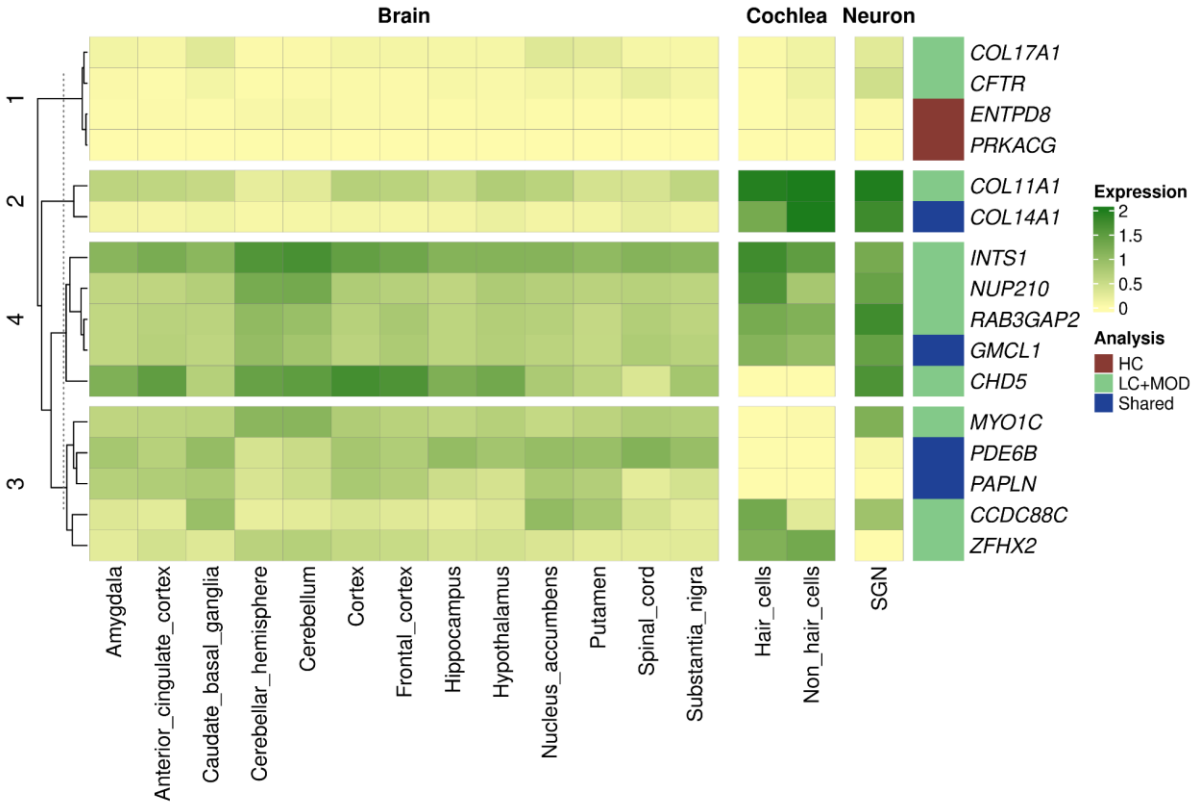

Figure 40 - Expression of the 16 candidate genes for the I1.

The heatmap shows the expression levels of candidate genes in 13 brain tissues obtained from GTEx in humans, the expression of the orthologous genes in the cochlear hair cells and non-hair cells from gEAR in postnatal day 0 (P0) mice and in the spiral ganglion neuron (SGN) from SHIELD in P0 mice. The expression for each dataset was normalised from 0 to 100 - being 100 the maximum value in the I4 subgroup) -, it was represented as the base-10 logarithm. Each gene was labelled by the analysis by which it was found: enrichment of variants with a high confidence of being loss-of-function (HC), enrichment of variants with a low confidence of being loss-of-function and with moderate impact in the protein predicted to be deleterious (LC+MOD), and shared between different analysis.

5.9 FUNCTIONAL ENRICHMENT ANALYSIS OF CANDIDATE GENES

To analyse the implications of the 28 candidate genes for I4 and the 16 for I1 discovered, functional analyses using public databases were performed. A total of three enrichment analyses were carried out with the candidate genes for the I4 and I1 subgroup using the following databases: BP from GO, HPO and MGI.

5.9.1 Enrichment analysis of I4 candidate genes

Using the BP ontology from the GO database, 17 terms were obtained as significant ($p < 0.05$) with at least two candidate genes in the term (Figure 41, Table S28). From the top 15, the term with a lower *p*-value was *pyrimidine dimer repair* and the term with a higher number of genes *cellular response to DNA damage stimulus.* Interestingly for the severe tinnitus phenotype, the terms *synaptic transmission, glutamatergic, axonogenesis, nervous system development* and *chemical synaptic transmission* were enriched. Moreover, *sensory perception of sound, detection of mechanical stimulus involved in sensory perception of sound, auditory receptor cell development* and *auditory receptor cell stereocilium organisation* were also enriched.

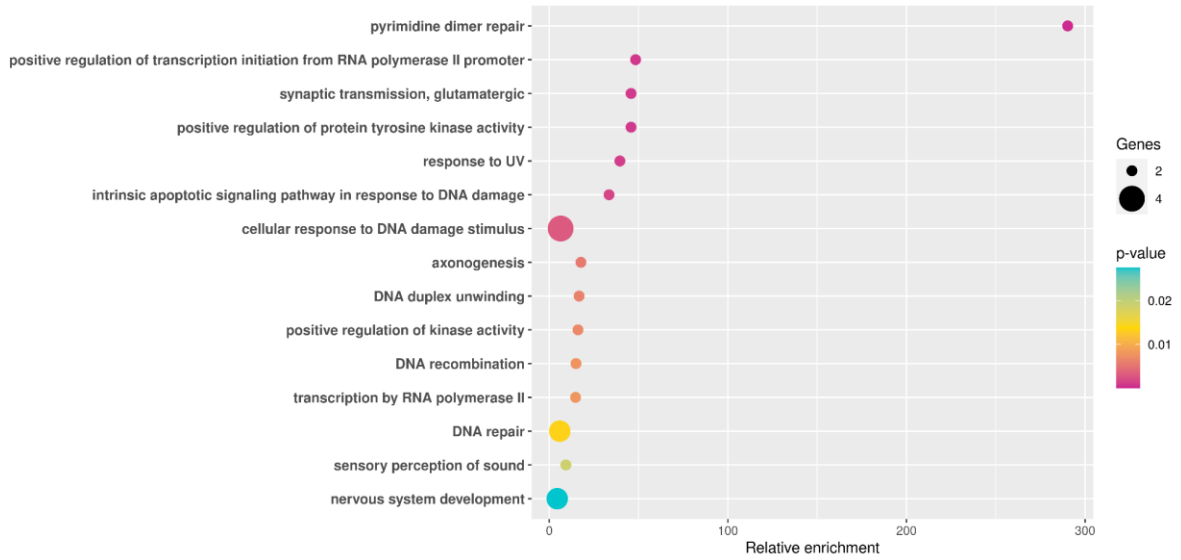

Figure 41 - Biological processes from Gene Ontology database associated with the 28 candidate genes for the I4 subgroup.

Dot-plot representing the relative enrichment, p-value and number of genes associated with each term. The top 15 terms with at least two genes.

With the HPO database, 36 phenotypes with two or more candidate genes were statistically related ($p < 0.05$) to the genes from the I4 subgroup (Figure 42, Table S29). The most significant phenotype (lower *p*-value and higher relative enrichment) was *absent brainstem auditory responses*, strongly related to the tinnitus severe phenotype. Besides, *anxiety*, *abnormal auditory evoked potentials*, *sensorineural hearing impairment* and *adult onset sensorineural hearing impairment* were enriched.

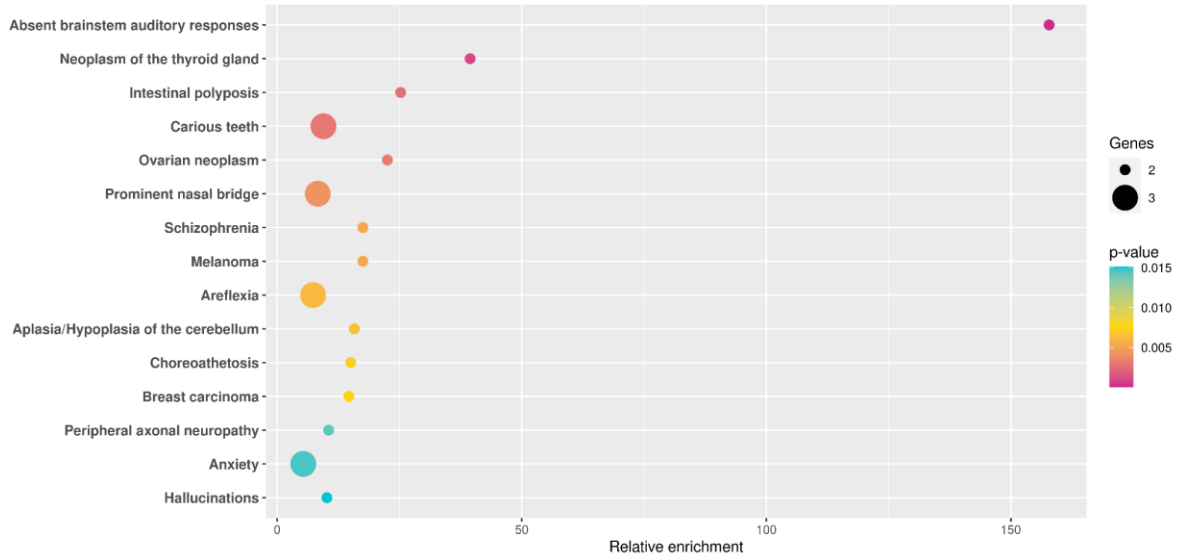

Figure 42 - Phenotypes from Human Phenotype Ontology database associated with the 28 candidate genes for the I4 subgroup.

Dot-plot representing the relative enrichment, p-value and number of genes associated with each term. The top 15 terms with at least two genes.

Finally, when the MGI database was used, 20 mouse phenotypes with at least two candidate genes were significant ($p < 0.05$) for the I4 candidate genes (Figure 43, Table S30). In this case, the most significant term was *increased cellular sensitivity to gamma-irradiation.* In addition, the mouse phenotypes *increased or absent threshold for auditory brainstem response*, *neuron degeneration*, *abnormal excitatory postsynaptic currents* and *abnormal auditory cortex morphology*, among others, were obtained as enriched. Besides, *abnormal hair cell mechanoelectric transduction*, *abnormal outer hair cell kinocilium location or orientation* and *decreased cochlear microphonics* and *deafness* were also found as enriched.

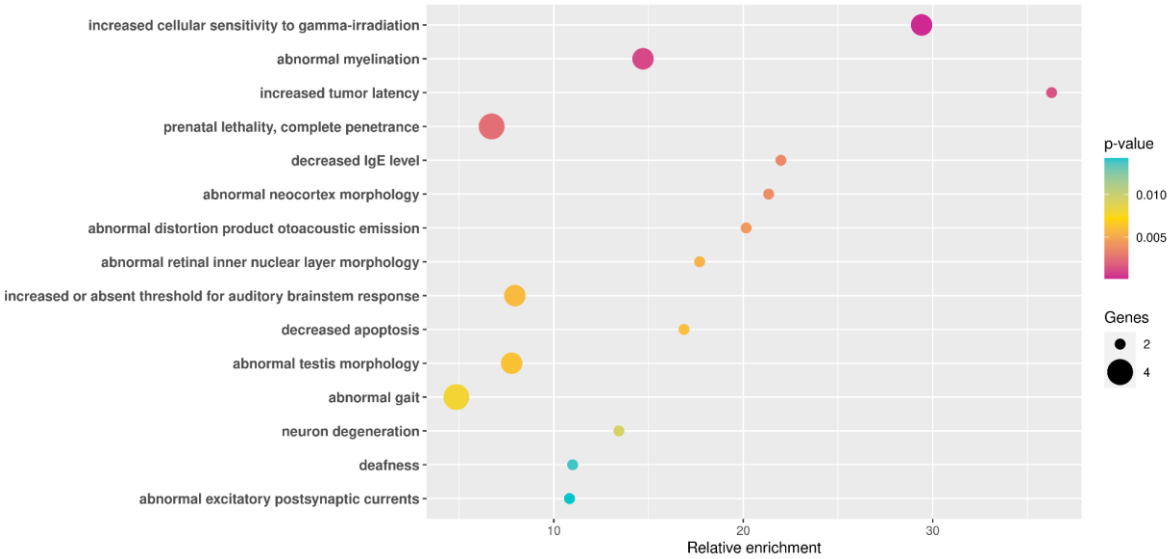

Figure 43 - Phenotypes from Mouse Genome Informatics database associated with the 28 candidate genes for the I4 subgroup.

Dot-plot representing the relative enrichment, p-value and number of genes associated with each term. The top 15 terms with at least two genes.

5.9.2 Enrichment analysis of I1 candidate genes

The enrichment analysis of the candidate genes for the I1 subgroup using the BP ontology from the GO database reported four significant terms ($p < 0.05$) with at least two genes (Figure 44, Table S31). The most significant term, with four genes, was *extracellular matrix organization*, the term *cerebral cortex neuron differentiation* with one gene was also interesting.

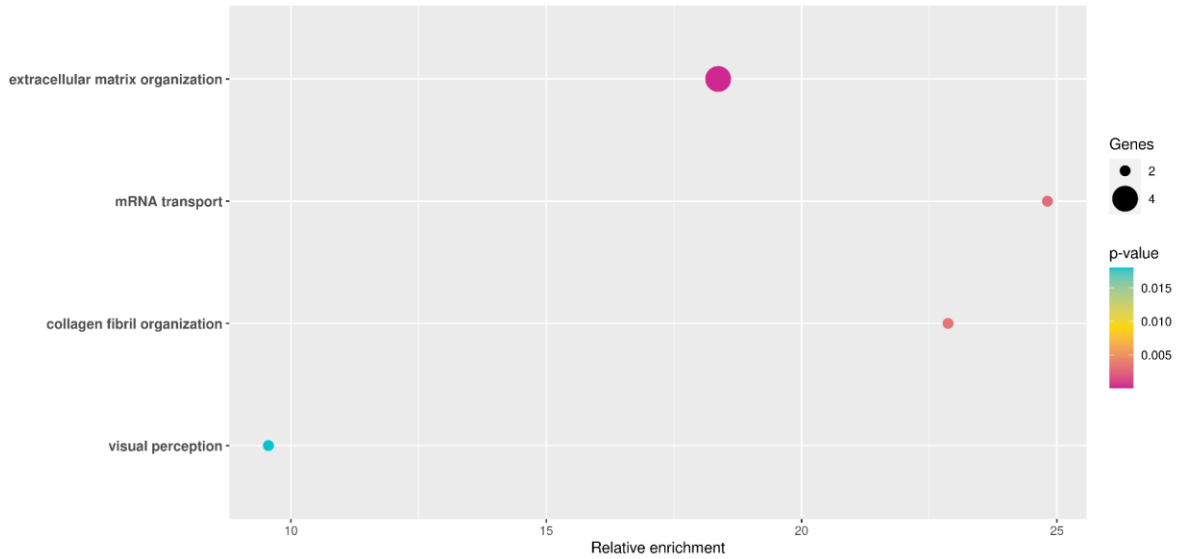

Figure 44 - Biological processes from Gene Ontology database associated with the 16 candidate genes for the I1 subgroup.

Dot-plot representing the relative enrichment, p-value and number of genes associated with each term. Terms with at least two genes.

A total of 12 phenotypes with two or more candidate genes were statistically related ($p < 0.05$) to the genes from the I1 subgroup when using the HPO database (Figure 45, Table S32). In this case, the most interesting phenotypes were *pain* and *pain insensitivity* - with one gene.

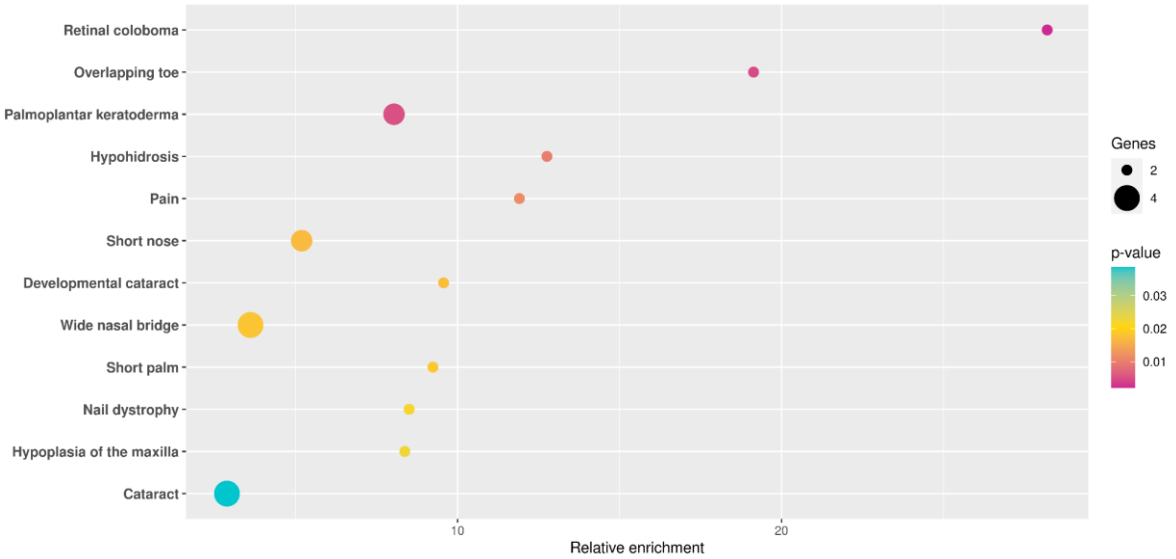

Figure 45 - Phenotypes from Human Phenotype Ontology database associated with the 16 candidate genes for the I1 subgroup.

Dot-plot representing the relative enrichment, p-value and number of genes associated with each term. Terms with at least two genes.

With the MGI database, four mouse phenotypes with at least two candidate genes were significant ($p < 0.05$) for the I4 candidate genes (Figure 46, Table S33). Similarly, to the I4 analysis, the term *absent cochlear inner hair cells* and *absent cochlear outer hair cells* - with one gene each of them - was found as enriched.

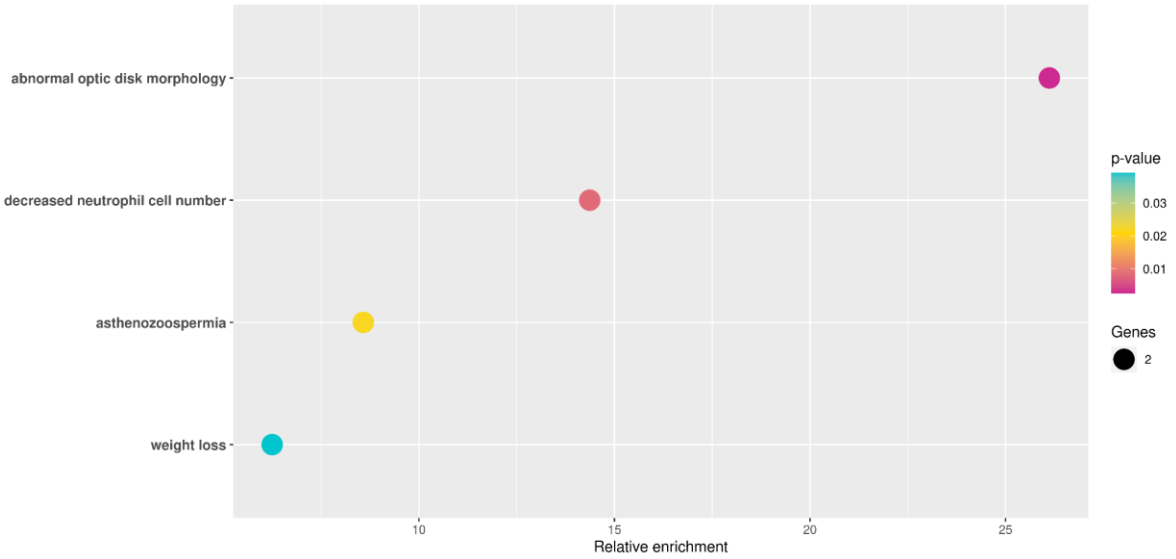

Figure 46 - Phenotypes from Mouse Genome Informatics database associated with the 16 candidate genes for the I1 subgroup.

Dot-plot representing the relative enrichment, p-value and number of genes associated with each term. Terms with at least two genes.

6 DISCUSSION

This thesis aims to study the contribution of rare variants in the human genome to tinnitus by selecting patients with MD and genes with a burden of rare variants in both extremes of the phenotypic spectrum.

This research was based on the hypothesis that MD individuals with a greater tinnitus discomfort and, on the opposite, those without tinnitus or with little annoyance caused by their tinnitus, have an enrichment of rare genetic variants in certain genes that modify the susceptibility to develop tinnitus. For this, the exome of 310 MD individuals was sequenced and analysed. The patients with and without severe tinnitus were clustered according to the THI score (subsets I4 and I1). To investigate the aggregated effect of rare variants, various GBAs were conducted. In addition, the novel variants, the CNV and the SV were analysed to improve the knowledge between these variants and their relationship with tinnitus. Finally, the expression and the biological functions of the candidate genes were studied through public databases.

6.1 MAIN FINDINGS IN THE CANDIDATE GENES OF FAMILIAL MENIERE DISEASE

The systematic review of the FMD revealed 11 candidate genes associated with the disease; it demonstrated that it is not a monogenic disorder due to genetic heterogeneity. Interestingly, some SNVs reported in the *OTOG* gene were found in various families with unrelated MD individuals[156]. In contrast, the SNVs identified in the *FAM136A, DTNA, PRKCB, COCH, DPT, SEMA3D, STRC, HMX2, TMEM55B* and *LSAMP* were found in one family each[150–155,157]. Those results revealed that it is necessary to carry out further analysis in cellular and animal models to improve the evidence of the association between *OTOG* and FMD and that finding more families carrying the proposed variants for the rest of the genes should be essential to confirm the role of the genes in the disorder. The genes associated with FMD have functions in different organs and structures. Due to the functional differences and the genetic heterogeneity reported, it was concluded that diverse molecular mechanisms should lead to the same syndrome.

Moreover, the inheritance patterns of each family were studied in the systematic review. It was demonstrated that in FMD, the inheritance could be AD or AR. The inheritance pattern of the described SNVs in *FAM136A* and *DTNA* was AD. The variants in *PRKCB, COCH, DPT* and *SEMA3D* genes were segregated with an AD pattern and incomplete penetrance since some the family relatives showed an MD incomplete phenotype, only SNHL or vertigo. In addition, the SNVs in *STRC* and *LSAMP* genes followed an AR inheritance pattern. However, the variants in *HMX, TMEM55B* and five from *OTOG* had an AR multiple inheritance.

6.1.1 Genes sharing enrichment of variants in the whole cohort of Meniere Disease and in the I4 subgroup

Furthermore, a GBA using the whole MD cohort studied in this work revealed eight genes enriched in variants with a high confidence of being LoF and 15 genes enriched in mostly missense variants, some splice acceptor and donor and variants with a high impact in the protein but with a low confidence of being LoF. None of the 11 genes associated with FMD were found enriched in the GBA, which could be because only 61 of the 310 patients were diagnosed as familial cases. However, the main interest in performing the GBA with the MD cohort was to compare these results against the GBA from the I4 and I1 subgroups. Two genes were shared between the whole cohort and I4: *PTTG2* (pituitary tumour-transforming 2) gene is part of the *cell cycle* pathway (hsa04110)[158] and it has been involved with different cancer types[159]; while *CNTNAP2* (contactin associated protein 2) is essential to neurocognitive development, being associated with the normal development and with complex neurological disorders[160]. Four genes shared between the whole cohort and I1: *ADAM2* (ADAM metallopeptidase domain 2) plays an important role in mammalian fertilisation[161]; *FRYL* (FRY-like transcription coactivator) is a gene that acts as a transcriptional activator in different cancers[162]; *FLT4* (fms related receptor tyrosine kinase 4) encoding the VEGFR3 (vascular endothelial growth factor receptor 3) is related to the formation of the circulatory vessels[163]; and the *SCN10A* (sodium voltage-gated channel alpha subunit 10) gene, encoding the sodium channel isoform Nav1.8, is associated to nociceptive and cardiac neurons[164]. These results suggest that the burden of variants in the whole MD cohort and in one of the subgroups may be because the genes are not specific only for tinnitus. In addition, the multilevel origin of tinnitus could involve different brain regions, but also the auditory organ and cochlear nucleus. Regarding the expression in human brain tissues, mouse cochlea and mouse SGN; *CNTNAP2* was expressed in the brain, highly expressed in SGN and did not show any expression in cochlea. *FRYL* was expressed in all the studied tissues, especially in cochlear hair cells. *FLT4* had low expression in the studied

tissues; while *PTTG2*, *ADAM2* and *SCN10A* were not expressed in any[131–133]. Further functional studies will be needed at the cellular level and with animal models to demonstrate the role of rare variants in the *CNTNAP2* and *FRYL* genes in non-severe tinnitus.

As the major aim of this thesis was to identify the genes contributing to severe tinnitus and all the individuals in the cohort suffer from MD, the 11 genes found as associated with FMD by the systematic review and the six genes enriched in variants in the whole MD cohort and the I4 or I1 subgroups were considered in the analysis for tinnitus.

6.2 COHORT DEMOGRAPHICS

In this work, it was observed that the percentage of females with a greater annoyance due to the tinnitus was higher than the percentage of females without tinnitus or with lower disturbance. Discrepancies were found in previous studies regarding the prevalence according to sex. For bothersome tinnitus - severe/moderate worry, annoyance or upset caused by the noises - the prevalence in females was significantly higher than in men[165].

The major correlation between the THI and the other questionnaires was with the GÜF questionnaire, which measures hypersensitivity to sound or hyperacusis. As previously reported, about 40% of individuals with moderately severe tinnitus suffer from hyperacusis and it reaches 80% of individuals when they present severe tinnitus[166,167]. In addition, the correlation between anxiety - using the HADS-A questionnaire - and depression - using both PHQ-9 and HADS-D questionnaires - was found significant. A cross-sectional analysis showed that 26.1% of individuals with tinnitus reported anxiety and 25.6% depression, being in both cases significantly greater than in the rest of the population[168]. Despite this, only 15 patients from the studied cohort were classified in the I4 subgroup for tinnitus and hyperacusis and 21 patients in the I1 subgroup.

The number of variants obtained after filtering by AB, GQ and DP - reproducing the gnomAD genotype filtering[115] - and by VQSR, was higher than filtering only by VQSR. This difference would be caused because around 5,000 variants in each sample were removed with the AB, GQ and DP filter, which could be false positives. The absence of a high number of false positives would improve the machine learning model of VQSR.

6.3 RARE VARIANTS SHOW AN AGGREGATE EFFECT IN TINNITUS

Regarding the criteria to filter the variants for the GBA, the selection of filtering by variants with a high impact in the protein and high confidence of being LoF, and variants with a high impact in the protein and low confidence of being LoF and variants with a moderate impact in the protein and predicted to be deleterious, allowed to obtain homogeneous groups improving the results of the analysis. Concerning the frequency in the reference databases, the AF to determine rare variants was initially set as 1%[52]. However, this AF (and the minors to it) was discarded because the lowest AF found in a variant in the study cohort was 0.006 (1/77*2, being 77 the number of individuals in the I4 subgroup), too close to 0.01. Filtering by a frequency less than 5% allows the discard of common variants[52] and filtering by a frequency less than 10% keeps variants with a similar to or lower frequency than the prevalence of tinnitus in the population[29]. The number of variants after filtering by 0.05 or 0.1 was almost the same, then the number of enriched genes was also very similar. Notwithstanding, a large difference was observed in the number of genes after filtering by the number of individuals (three for 0.05 against seven for 0.1), obtaining a small number of candidate genes with the filter of AF by 0.1.

The total number of samples in CSVS v3.0.1 was 2,048[64], whereas the number of genomes in gnomAD v3.1 was 76,156[118]. Moreover, the CSVS database only included SNVs. Then, due to the small sample size and the lack of short indels, it was decided not to filter out the genes with variants not included in CSVS as not significant after the GBA. However, the genes with a *p*-value adjusted by FDR greater than 0.05 for the CSVS database were discarded because, with the available data, it was possible to state that these variants were not associated with tinnitus, but rather they were characteristics of the Spanish population.

The gnomAD database does not provide clinical information about the individuals; due to the high prevalence of tinnitus in the global population, it was expected that a considerable number of individuals in the database presented this phenotype. Besides, variants that confer a greater risk of developing tinnitus in patients with MD were searched. In consequence, it was not expected to find very rare variants.

6.4 CANDIDATE GENES FOR TINNITUS EXTREME PHENOTYPE

To pinpoint candidate genes from the GBA, the genes with a significant burden of high confidence LoF variants and the top genes enriched in low confidence LoF variants and variants with moderate impact in the protein (mostly missense variants) and predicted to be deleterious, were selected. To create the SNV exome database, after the variant calling and normalisation, it was essential to filter the variants. Despite the previous filter to VQSR by AB, GQ and DP - following the filter to create the gnomAD database[115] - improved the ratio of true positives; few variants in the potential candidate genes were discarded because IGV did not confirm them. Particularly, in regions with low quality or repetition of the same nucleotides. However, most of them were confirmed.

Furthermore, genes sharing SNVs, CNVs and SVs from different analyses were selected as candidate genes. It was difficult to validate some of the CNV and SV. Although the tools were prepared to identify those variants in exome sequencing, most of them cover non-coding regions, which could lead to failure in the lectures or the impossibility of validating them with IGV. Interestingly, all the SV in the individuals from the I1 subgroup were found in the reference populations, which could be because they were not important to develop the tinnitus.

The novel variants analyses were carried out to investigate any association between genes with novel variants and the tinnitus severe phenotype. The most interesting result was that four novel variants for I4 and one for I1 were shared by two individuals. However, these results would not explain the phenotype, consequently they were not selected as candidate genes.

6.4.1 Severe tinnitus (I4)

The expression data of the I4 candidate genes in 13 brain tissues, mouse cochlea and mouse SGN allowed the clustering of the genes in four groups[131–133]. The genes from cluster 1 were either not expressed at all or expressed to a minimal extent in the studied tissues. For this reason, it was difficult to understand their involvement in the phenotype, at least with the available data. The *IQCN* (IQ motif containing N) gene was related to male infertility in humans and mice[169]. The *MIR25, MIR93* and *MIR106B* genes were non-coding. All three were transcribed to microRNAs, which had in common the biological process *gene silencing by miRNA* and the molecular function *mRNA binding involved in posttranscriptional gene silencing*, both described by GO database[135,136]. None of the microRNAs had one of the I4 candidate genes as target gene[170]. The *WDFY4* (WDFY family member 4) gene, poorly

expressed in cochlear non-hair cells and expressed in SGN, was related to the decreasing IL-1β and IL-6 and increasing the antinuclear antibody in systemic lupus erythematosus patients[171]. A duplication in this gene was found in neurodevelopmental diseases[172] and it was associated with clinically amyopathic dermatomyositis[173]. The *LPIN3* (lipin 3) gene was related to dementia[174]; however, few expression was observed in the brain tissues, it was found as differentially expressed in the cochlea compared against the vestibule of three human donors[175]. The *EFCAB5* (EF-hand calcium binding domain 5) gene was associated with the neuronal ageing rates in five human brain regions[176]. Remarkably, a burden of missense variants was obtained for the *EFCAB5* gene in 97 patients from Sweden with chronic or constant tinnitus[93].

Two of the three genes associated with *absent brainstem auditory responses* (HP:0004463), *ERCC6* (ERCC excision repair 6, chromatin remodeling factor) and *SPTBN4* (spectrin beta, non-erythrocytic 4), were candidate genes for the I4 subgroup according to the HPO database[137]. In the same way, based on the mammalian phenotype ontology from MGI database, both genes and *PDZD7* (PDZ domain containing 7) were involved in the *increased or absent threshold for auditory brainstem response* (MP:0011967) mouse phenotype[138]. Besides, *ERCC6* was related to *abnormal auditory evoked potentials* (HP:0006958)[137]. Although the absence of auditory brainstem response, also known as brainstem auditory evoked potentials, indicates that the individual suffers from hearing loss[17]; the enrichment of this term in I4 candidate genes might suggest that individuals carrying variants in both genes could have altered brainstem auditory responses. As it has been demonstrated previously, tinnitus patients showed a significant prolongation in the latencies of waves I, III and V of the auditory brainstem responses. The prolongation of each wave could be caused by an alteration of nerve fibres in the following regions of the central auditory pathway, respectively: distal part of the auditory nerve, cochlear nucleus and lateral lemniscus and inferior colliculus[177,178]. Both genes were also linked to the human phenotype *peripheral axonal neuropathy* (HP:0003477), the mouse phenotype *abnormal myelination* (MP:0000920) with *ERBB3* (erb-b2 receptor tyrosine kinase 3), and the mouse phenotype *abnormal gait* (MP:0001406) with *RAD50* (RAD50 double strand break repair protein) and *ARVCF* (ARVCF delta catenin family member). All of them confirm the previous relationship between those genes and abnormal function of the nerve fibres. Moreover, it was reported the association between *ERBB3*, *ERCC6*, *PDZD7* and *SPTBN4* and *sensorineural hearing impairment* (HP:0000407) in humans; and between *ERCC6* and *SPTBN4* in *impaired hearing* (MP:0006325) and *impaired balance* (MP:0001525)

in mice. Taking all these genes' information together, it was supposed that the phenotypes related to the inner ear could be caused by a malfunction of the nerve impulse transmission. Separately, in mice, *SPTBN4* was related to *abnormal auditory cortex morphology* (MP:0004631), *ERBB3* and *ERCC6* to *abnormal neocortex morphology* (MP:0008547) and to *neuron degeneration* (MP:0003224), and *ERCC6* and *PDZD7* to *abnormal distortion product otoacoustic emission* (MP:0004736). Interestingly, *ERCC6* was expressed in the brain, cochlea and SGN, especially in cochlear hair cells; whereas, *SPTBN4* only was expressed in the brain, particularly in the cerebellum. The mouse orthologue Sptbn4 encodes the beta-4 (β4)-spectrin. The absence of β4-spectrin disrupts the clustering of the voltage-gated sodium channels (GO:0045162) at the heminode along the nerve terminal and temporal failures in the presynaptic spikes, which results in a deficit of the central auditory processing[179].

Even though it seemed that *ERCC6* was related to tinnitus by alteration of nerve fibres in different regions of the central auditory pathway, it was remarkable that *ERCC6* and *MSH6* (mutS homolog 6) were two of the three genes involved in the *pyrimidine dimer repair* (GO:0006290) according to the Biological Processes ontology from GO[135,136]. In addition, both of them were also related with the *intrinsic apoptotic signaling pathway in response to DNA damage* (GO:0008630) and *response to UV* (GO:0009411). It has been demonstrated the importance of the dimer repairing after damage caused by UV, principally in the skin[180]. Due to the fact they repair the DNA[181], both genes were associated with multiple diseases as different cancers[182,183]. Because of that, it makes sense that both genes were expressed in all the studied tissues, being part of the second defined cluster.

ERCC6, COPS6 (COP9 signalosome subunit 6) and *RAD50* were involved in the *increased cellular sensitivity to gamma-irradiation* (MP:0002007) mouse phenotype from MGI[138]. This increase improved the safety and efficacy of the irradiation treatment against the vestibular schwannomas, which are benign tumours that originate from Schwann cells, which surround the vestibulocochlear nerve; these tumours, among other symptoms, produce tinnitus[184]. Of note, *RAD50* was also enriched in missense variants in the previously described cohort of 97 patients with chronic or constant tinnitus[93]. Notwithstanding the three genes were expressed in all the studied tissues, it was remarkable the high expression of the *COPS6* gene in all the tissues, being the most expressed gene in all the brain tissues and in SGN, which is consistent with its function.

The ErbB3 protein, encoded by the *ERBB3* gene, is located in the nucleus of Schwann cells[185], it plays an important role in the *Schwann cells differentiation* (GO:0014037) according to GO[135,136] and it is the only gene participating in the *absent Schwann cell precursors* based on MGI[138]. The Schwann cells are the myelinating cell of the peripheral nervous system, the myelination increases the conduction velocity along the axon[186]. Torii et al.[187] demonstrated that the ErbB3 mouse protein was essential to the migration and myelination of the Schwann cells. In addition, Schwann cells are necessary in the response to axon damage and axon regeneration in the peripheral nervous system[186]. This gene was expressed in all the studied tissues; however, it was interesting the differences in the expression pattern compared with the other genes in the same cluster.

The *PDZD7* was well characterised as a nonsyndromic hearing loss gene[188]. In the MGI database, it was associated to *decreased cochlear microphonics* (MP:0004414), *abnormal outer hair cell kinocilium location or orientation* (MP:0030961), *decreased outer hair cell stereocilia number* (MP:0004529), *abnormal cochlear outer hair cell physiology* (MP:0004434) and *abnormal hair cell mechanoelectric transduction* (MP:0004431). This was consistent with the finding that, among the tissues studied, the cochlear hair cells showed the highest expression. The *PDZD7* gene encodes for the PDZD7 protein, which is a scaffolding protein highly expressed in the stereocilia of the inner hair cells[189]. Despite the fact that PDZD7 function has not been linked to tinnitus; in rodents, it was demonstrated that the development of tinnitus was very related to synaptopathy and deafferentation at the inner hair cells[190,191]. In addition, the *PDZD7*, *MSH6* and *ARVCF* genes were involved in *anxiety* (HP:0000739) and *depression* (HP:0000716) human phenotypes, according to HPO[137].

GRIK4 and *UNC13C* were linked to the *synaptic transmission glutamatergic* (GO:0035249) and *chemical synaptic transmission* (GO:0007268) biological processes by GO. In addition, *GRIK4* was involved in *glutamate receptor signaling pathway* (GO:0007215) and *ionotropic glutamate receptor signaling pathway* (GO:0035235). *UNC13C* was part of *presynaptic dense core vesicle exocytosis* (GO:0099525), *negative regulation of synaptic plasticity* (GO:0031914) and different *synaptic vesicle* terms (GO:0016188, GO:0016081, GO:0016079 and GO:0016082)[135,136]. *GRIK4* (glutamate ionotropic receptor kainate type subunit 4) encodes the GluK4 protein, which modulates the memory, cognition and baseline affect and also regulates the excitotoxic cell death cascades[192]. Moreover, variants in this gene have been related to mental diseases, such as bipolar disorder or autism[193]. Then, *GRIK4* is involved in

neuropsychiatric and neurodegenerative disorders[192]. Shin et al.[194] exposed that, in mice, the mediodorsal thalamic neurons positive to GluK4 were excited by the auditory-induced arousal during the slow-wave sleep. *UNC13C* (unc-13 homolog C, Munc13-3) is part of the Munc13 mammalian protein family; it is expressed in the caudal areas of the brain and enriched in synapsis, especially in presynaptic terminals[195,196]. It has a neuroprotective function in the brain[197], besides potentially damaging variants in the gene were related to Alzheimer's disease and frontotemporal dementia[198]. Interestingly, *UNC13C* was also enriched in both LoF and missense variants, separately, in the depicted Swedish cohort suffering from chronic or constant tinnitus[93].

SPTBN4 and *DSCAML1* participate in *axonogenesis* (GO:0007409) and *DSCAML1* in *dendrite self-avoidance* (GO:0070593) according to GO[135,136]. *DSCAML1* (DS cell adhesion molecule like 1) encodes for the cell adhesion molecule DSCAML1 protein, an immunoglobulin-like transmembrane protein and a receptor. It was observed that in neurons with variants in *DSCAML1* the regulatory function in the number of synapsis was lost, which could lead to the pathophysiology of neurodevelopmental diseases[199]. In zebrafish, it was demonstrated that *dscaml1* was fundamental for stress axis development and its dysregulation could be related to neuropsychiatric disorders[200]. Furthermore, *DSCAML1* was enriched in missense variants in a tinnitus Swedish cohort with a THI score greater than 58[93].

The nuclear *ENDOG* (endonuclease G) gene encodes for the protein ENDOG, a mitochondrial nuclease. It plays a vital role in the apoptosis and regulates the autophagy[201]. In Guinea pigs it was found that after a noise exposure the endoG protein was translocated from the mitochondria to the nucleus in apoptotic outer hair cells[202]. The relation of the *ENDOG* apoptosis inducing factor with the noise-induced hair cell death pathways suggests that variants in this gene could be related to failure in the protection of the hair cells against noise trauma leading to tinnitus.

FBXO27, encoded by the *FBXO27* (F-box protein 27) gene, is a glycoprotein-specific F-box protein. It is part of the SCF ubiquitin ligase complex, in charge of signalisation to carry out the autophagy. Concretely, SCF^{FBXO27} ubiquitinates glycoproteins in damaged lysosomes to induce lysophagy. *FBXO27* is highly expressed in the nervous system, it has been demonstrated that the abnormal accumulation of proteins could lead to neurodegenerative disorders. Alterations in the FBXO27 protein may hinder proper lysophagy, leading to the accumulation of aberrant proteins in the tissues where it is expressed[203].

Biallelic LoF variants in the *AP4M1* (adaptor related protein complex 4 subunit mu 1) were described as caused of spastic paraplegia 50, which is a neurological disorder[204]. Besides, a missense homozygous variant in the same gene was associated to individuals spastic paraplegia, intellectual disability, hearing loss and microcephaly[205].

MROH1 (maestro heat like repeat family member 1) gene was related to progression-free survival in prostate cancer[206]. Variants in *EML6* (EMAP like 6) were associated with keratoconus in two families[207] and one variant to hypertension[208]. The *TAF6* (TATA-box binding protein associated factor 6) gene encodes for the protein TAF6, which is one of the principal components of the TATA-box-binding protein (TBP) essential to initiating gene transcription by RNA polymerase II[209]. *TAF6* was related to the Cornelia de Lange syndrome[210] and the Alazami-Yuan syndrome[211]. Despite the fact that those genes have described functions unrelated to tinnitus, both *EML6* and *MROH1* genes were also enriched in missense variants in the explained tinnitus cohort of 97 individuals, and *TAF6* was also enriched in missense variants but in a replication cohort composed of 147 patients with chronic and constant tinnitus[93].

The *THADA* (THADA armadillo repeat containing) gene was related to the function of pancreatic β-cells [212], therefore with type 2 diabetes[213]. Moreover, it was involved in the maturation and maintenance in tumour cells[214], and with polycystic ovary syndrome[215]. *ITGB5* (integrin subunit beta 5) is one of the eight integrins, taking part in the regulation of different cellular processes, being one of the most studied carcinogenesis[216]. It has been related to various cancers, such as pancreatic[217] or colorectal[218]. By transcriptome-wide association study, the *ANKRD44* (ankyrin repeat domain 44) gene was related to osteoarthritis[219] and it had the function of osteogenic differentiation[220]. In addition, variants in this gene were related to resistance to the Trastuzumab, a drug to treat breast cancer[221]. *MCM7* (minichromosome maintenance complex component 7) encodes for the protein MCM7, which is part of the hexameric complex composed of MCM proteins. These proteins are essential to the replication and cell cycle progression; due to this function, *MCM7* has been associated with different cancers[222,223]. To sum up, these genes are associated with multiple functions, thus necessitating further analysis to confirm whether they are false positives or genes related to tinnitus.

6.4.2 Few disturbances caused by tinnitus (I1)

The genetics of the individuals from the I1 subgroup were studied to discover genes that would protect against tinnitus. Two of the I1 candidate genes had previously been associated with the protection of hair cells from damage to the cochlea. In the *CFTR* (CF transmembrane conductance regulator) gene, 13 variants in 18 individuals from the I1 subgroup were identified. In mouse cells, the treatment with a potentiator of CFTR demonstrated a protection of cochlear hair cells against noise trauma[224]. Although it was not expressed in the studied tissues in this work, a prior analysis revealed the presence of mRNA and protein of CFTR in the outer hair cells[225]. Deeper experiments will be necessary to check the protection against tinnitus caused by variants in *CFTR*. Furthermore, in null mice for the ectonucleotidase CD39, encoded by Entpd1, compensation by the up-regulation of Entpd8 was observed[226]. Based on these results, the variants in the *ENTPD8* (ectonucleoside triphosphate diphosphohydrolase 8) gene could improve the compensation against cochlea damages that would provoke tinnitus.

In addition, two candidate genes are essential to the correct movement and arrangement of the stereocilia and, therefore, to the correct excitation of the hair cells. The function of the unconventional myosin-Ic protein, encoded by *MYO1C* (myosin IC), in the cochlea is the adaptation motor in the hair bundles between the stereocilia of the hair cells[227]. Variants in *MYO1C* were related to hearing loss[228]. Although no expression in the cochlea was observed in the studied datasets, in three human inner ear samples, it was found that the expression of *MYO1C* was significantly greater than in the vestibule[175]. *CCDC88C* (coiled-coil domain containing 88C) encodes for the Daple protein. Variants in the Ccdc88c mouse gene caused errors in the correct arrangement of hair bundles in the coc656635hlear hair cells due to a malfunction of the apical microtubule distribution[229,230].

The enrichment analysis of candidate genes for the I1 subgroup revealed that four of them were involved with the *extracellular matrix organization* (GO:0030198) biological process by GO: *COL11A1* (collagen type XI alpha 1 chain), *COL14A1* (collagen type XIV alpha 1 chain), *COL17A1* (collagen type XVII alpha 1 chain) and *PAPLN* (papilin, proteoglycan like sulfated glycoprotein)[135,136]. The tectorial membrane is an extracellular matrix in the cochlea[11], which has previously been linked to MD, as in the case of variants in the *TECTA* gene[231]. The collagens that have been described in the tectorial membrane, for the moment, are collagens II, IV, IX and XI[11]; however, it is possible that XIV and XVII collagens, and even papilin, take part in its correct structure. *COL11A1* and *PAPLN* were only expressed in the brain, *COL11A1*

and *COL14A1* were highly expressed in cochlea and SGN. Although changes in the proteins of the tectorial membrane suggest that they affect hearing[232], the results obtained in this work suggest they could protect against tinnitus. In addition, according to the MGI database, the *COL11A1* gene is related to *absent cochlear inner hair cells* (MP:0004397), *absent cochlear outer hair cells* (MP:0004403) and *abnormal auditory brainstem response waveform shape* (MP:0011966) phenotypes[138]. Variants in *COL11A1* were associated with hearing loss in Stickler syndrome[233], nonsyndromic hearing loss[234] and adult genetic SNHL[235], among others.

The rest of the I1 candidate genes were associated with various phenotypes and functions, which makes necessary deeper analyses to link them to the protection against tinnitus. *ZFHX2* (zinc finger homeobox 2) was related to human pain insensitivity disorder[236] and in mice it controls emotional aspects through the function of monoaminergic neurons[237]. *CHD5* (chromodomain helicase DNA binding protein 5), encoding for the nucleosome remodelling and deacetylation (NuRD) complex, was essential for neuronal development. It was related to a neurodevelopmental disorder and different cancers, such as neuroblastomas or hepatocellular carcinoma. Variants in *RAB3GAP2* (RAB3 GTPase activating non-catalytic protein subunit 2) caused Martsolf syndrome among other neurodevelopmental, neuromuscular and neurodegenerative disorders. Variants in *PDE6B* (phosphodiesterase 6B) were related to retinitis pigmentosa due to the disruption of the cGMP hydrolysis[238]. *PRKACG* (protein kinase cAMP-activated catalytic subunit gamma) was associated with macrothrombocytopenia[239]. *INTS1* (integrator complex subunit 1) was described for developmental delays[240]. The depletion of *NUP210* (nucleoporin 210) suppressed metastasis[241]. *GMCL1* (germ cell-less 1, spermatogenesis associated) was linked to carotid paragangliomas[242] and asthenozoospermia[243].

6.5 THE INTEGRATION WITH PREVIOUS STUDIES DEPICTS THE GENETIC ARCHITECTURE OF TINNITUS

To define not only the genes but also the biological functions and processes associated with tinnitus, the results of previous genetic studies on tinnitus were analysed and compared with those obtained in this project.

The results of this work, together with the antecedents, suggest that variants in genes involved in the organisation and activation of voltage-gated channels in the axonal projections would be one of the causes of tinnitus. Particularly, the β4-spectrin, encoded by the candidate gene found in the I4 subset *SPTBN4*, is necessary for the correct clustering and function of the voltage-gated sodium channels along the nerve terminal. In addition, the ankyrin-B protein, encoded by *ANK2* and highly expressed in neurons, presents the modulation activity of Nav1.6 (sodium channel protein type 8 subunit alpha) inactivation[244]. Whereas, the voltage-dependent R-type calcium channel subunit alpha-1E, encoded by *CACNA1E,* is part of the high-voltage calcium channel, which is in charge of the correct modulation of neurons to process the information[245]. The *ANK2* gene was enriched in missense variants in a Spanish MD cohort and in a replication Swedish cohort, both with severe tinnitus[92]. A burden of missense variants was reported in individuals with chronic or constant tinnitus in a Swedish cohort of 97 individuals and with a severe tinnitus subgroup[93].

Another possible cause is the incorrect development of the neurons, leading to tinnitus. The candidate genes for I4 subgroup *SPTBN4* and *DSCAML1* are closely related to the axogenesis. In the same way, the previously described *NAV2* gene is involved in the outgrowth of the neurite and the elongation of the axon[93]; and *ANK2* in the neural development, differentiation and migration[92]. *NAV2* was enriched in missense variants in two different tinnitus cohorts[93]. Interestingly, *DSCAML1* shows an overload of rare variants not only in the cohort of this thesis; moreover, it was found enriched in the Swedish cohort with severe tinnitus.

An interesting cause is the deafferentation at the hair cells caused by proteins necessary for the correct function of the hair cells[156,157]. In this project, the protein PDZD7 and ENDOG have been related to severe tinnitus. Those proteins are an essential scaffold in the stereocilia and a nuclease involved in the apoptosis of hair cells, respectively. Besides, the protein FBF1, encoded by the *FBF1* (Fas binding factor 1) gene, is essential to the formation, assembly and maintenance of the cilia[246,247]. And *LOXHD1* (lipoxygenase homology PLAT domains 1) is

expressed in the cochlear hair cells and LoF variants in it were related to auditory impairment[248]. A missense variant in *LOXHD1* and a tandem duplication in *FBF1* covering 4 exons were related to age-related hearing impairment, which shares genetic causes with tinnitus, according to the authors[249].

Tinnitus disorder has also been related to anxiety and depression[29]; interestingly the genes *PDZD7*, *MSH6* and *ARVCF* were associated with the anxiety and depression human phenotypes. Also, variants in *TMEM132D* (transmembrane protein 132D) were found to be related to anxiety phenotypes and major depressive disorder[250,251]. *TMEM132D* was also enriched in missense variants in the Swedish cohort with tinnitus and in the subset with severe tinnitus[93]. Further functional genomic studies will be needed to demonstrate that common and rare variants in these four genes, expressed in the brain, have pleiotropic effects on anxiety and tinnitus phenotypes. It is not possible to confirm if those variants are related to tinnitus or anxiety and depression.

Finally, the genes replicated between this thesis and the previous studies performed on tinnitus were summarised in Table 44.

Table 44 - Shared genes between candidate genes for I4 and I1 subgroups and previous results.

Article	Individuals	Type of variants	I4			I1	
			HC	LC+MOD	Shared	LC+MOD	Shared
Gallego-Martinez et al (2022)[93]	Tinnitus cohort with chronic or constant tinnitus (N = 97)	LoF		*UNC13C*			
		Missense	*EFCAB5, RAD50*	*EML6, MROH1, UNC13C*		*MYO1C*	
	Tinnitus cohort severe tinnitus (N = 34)	Missense		*DSCAML1*		*NUP210, RAB3GAP2*	*COL14A1*
	Replication tinnitus cohort with chronic or constant tinnitus (N = 147)	LoF				*CFTR*	
		Missense			*TAF6*		
Amanat et al (2021)[92]	Meniere Disease cohort with tinnitus almost extreme phenotype (N = 29)	Missense				*MYO1C*	

Resulting from the gene burden analysis with variants with high confidence of being loss-of-function (HC), GBA with variants with a low confidence of being loss-of-function and with a moderate impact in the protein and predicted to be deleterious (LC+MOD), and shared between different analysis. LoF: loss-of-function.

6.6 SIMILARITIES AND DIFFERENCES IN THE GENETICS OF TINNITUS AND HYPERACUSIS

Due to the critical relationship between tinnitus and hyperacusis, the enriched genes through the GBA in common between the I4 and I1 subgroups obtained by both questionnaires were analysed. Although the sample size of the GÜF subgroups was small to assure significant results, the results of the GBA were used to compare those obtained with the THI subgroups.

Firstly, the initial number of variants for the THI subgroups was greater than for the GÜF subgroups. However, the final number of candidate genes for the THI subgroups was lower than for the GÜF subgroups. The number of samples for the I4 and I1 subgroups for GÜF were 34 and 42, respectively, lower than those for the same subgroups but classified according to the THI, which were 75 and 88, respectively. The limited sample size could lead to false positives. In addition, the less stringent filter by the number of individuals, two for GÜF and three for THI, also contributed to the final number of candidate genes.

For severe tinnitus, 12 genes were enriched in LoF or missense variants in the I4 subgroups for both tinnitus and hyperacusis. Those genes would explain the genetics shared between tinnitus and hyperacusis phenotypes. Variants in *KAT6A* were associated with developmental delay and SNHL[252]. A truncating variant in *HOMER2*, which codes for a scaffolding protein of the stereocilia, was related to DFNA68 hearing loss[253,254]. However, the orthologue in mice Homer2 did not exhibit significant alteration after tinnitus induction in the auditory cortex[255]. Besides, the *HDAC4* gene is a histone deacetylase. It was demonstrated that the inhibition of HDAC4 protected the cochlear hair cell in mice suffering from hearing loss, due to maintenance of acetylation level[256–258].

Remarkably, two genes were enriched in missense variants in the subgroup with severe tinnitus and in the subgroup without hyperacusis. Deeper analyses will be necessary to explain its relation with the 20% of individuals with severe tinnitus and without hyperacusis[44]. One of them, the *RYR2* gene encodes for a ryanodine receptor, a calcium channel carrying out the regulation of the free calcium concentration in the cell. Those receptors have been found in the cochlear and vestibular hair cells and are involved in the cochlear maturity[259,260].

Eleven genes had a burden of LoF or missense variants in individuals without disturbance caused by tinnitus or hyperacusis. These genes are important candidate protectors against both

tinnitus and hyperacusis. It was shown that CFTR preserves cochlear hair cells when there is noise damage. Moreover, the *TLN1* (talin 1) gene promotes axonal regeneration after axon injury[261]. *PLXNA3* is expressed in the VIII cranial nerve neurons, where it plays an important role in the afferent projections of the neurons in the inner ear[262].

Furthermore, three genes had an overload of missense variants in the individuals without tinnitus but with severe hyperacusis. The *MMP9* gene is essential to developing sensory circuits[263]. Rats exposed to noise presented an increase in the *MMP9* expression in the primary auditory cortex[264]. Moreover, in mice with sensory hypersensitivity, the inhibition of *MMP9* leads to reduced auditory processing deficits[265].

6.7 FUTURE DIRECTIONS

In this study, the genetic data have been generated by performing exome sequencing, which covers only the coding regions. As a future approach, it would be ideal to carry out genome sequencing to study not only modifications in the proteins but also in non-coding regions, such as untranslated regions (UTRs) or promoters, which have an essential function in the expression. Besides, sequencing the whole genome will be essential to analyse CNVs and SVs, since it is challenging to obtain results in exomes. In addition, a replication cohort with proper clinical data would be interesting as a replication cohort, to select more confident candidate genes.

Regarding the candidate genes identified in this thesis, further analysis in cell or animal models would be necessary to confirm the role that they play in the tinnitus phenotype. Candidate genes for severe tinnitus and the pathways in which they are involved are potentially targets to a drug treatment. Besides, genes involved in the auditory brainstem and mid-latency responses have been described. To study these responses would select individuals to be treated in different levels of the auditory pathway. Moreover, some of the candidate genes for the phenotype with an absence of disturbance caused by tinnitus have been previously related with to protection of the hair cell against cochlear damage, so analysing the modifications in them could be interesting to future treatments.

7 CONCLUSIONS

1. MD and tinnitus have a complex genetic architecture. Through the systematic review, 11 genes have been identified in FMD. Moreover, six new genes show a burden of rare variants for MD and tinnitus phenotype.

2. Finding LoF or missense rare SNVs and SVs overlapping the same genes in both extremes of the tinnitus phenotypic spectrum supports the use of extreme phenotype strategies for pinpointing candidate genes.

3. The main severe tinnitus candidate genes sharing an overload of missense variants and SVs are *AP4M1, COPS6, ERBB3, MCM7, MIR106B, MIR25, MIR93* and *TAF6.*

4. The candidate genes identified in severe tinnitus patients are associated with different biological processed: *SPTBN4* and *DSCAML1* with axogenesis, *GRIK4* and *UNC13C* with synaptic transmission, and *PDZD7* and *ENDOG* with the depolarisation of the hair cells.

5. The candidate genes found in MD patients without tinnitus reveal variants in proteins that protect hair cells against damage, encoded by the *CFTR, ENTPD8, MYO1C* and *CCDC88C* genes, and in structural proteins of the tectorial membrane, as the collagen XI.

6. The clinical association between tinnitus and hyperacusis is partially based on the shared genetic structure observed in the MD cohort.

CONCLUSIONES

1. La EM y el acúfeno presentan una arquitectura genética compleja. A través de una revisión sistemática, se han identificado 11 genes en la EM familiar. Además, seis nuevos genes muestran una carga de variantes raras para la EM y el fenotipo del acúfeno.

2. El hallazgo de SNVs raras LoF o de cambio de sentido (*missense*) y SVs que se superponen en los mismos genes en ambos extremos del espectro fenotípico del acúfeno respalda el uso de estrategias de fenotipo extremo para señalar genes candidatos.

3. Los principales genes candidatos para el acúfeno severo que comparten una sobrecarga de variantes de cambio de sentido (*missense*) y SVs son *AP4M1*, *COPS6*, *ERBB3*, *MCM7*, *MIR106B*, *MIR25*, *MIR93* y *TAF6*.

4. Los genes candidatos identificados en pacientes con acúfeno severo están asociados con diferentes procesos biológicos: *SPTBN4* y *DSCAML1* con la axogénesis, *GRIK4* y *UNC13C* con la transmisión sináptica, y *PDZD7* y *ENDOG* con la despolarización de las células ciliadas.

5. Los genes candidatos encontrados en pacientes con EM sin acúfeno revelan variantes en proteínas que protegen las a células ciliadas contra el daño, codificadas por los genes *CFTR*, *ENTPD8*, *MYO1C* y *CCDC88C*, y en proteínas estructurales de la membrana tectoria, como el colágeno XI.

6. La asociación clínica entre el acúfeno y la hiperacusia se basa en parte en la estructura genética compartida observada en la cohorte con EM.

8 REFERENCES

1. Newton VE, Vallely PJ. *Infection and Hearing Impairment*. John Wiley & Sons; 2006.

2. Lim R, Brichta AM. Anatomical and physiological development of the human inner ear. *Hear Res*. 2016;338:9-21. doi:10.1016/j.heares.2016.02.004

3. Bruss DM, Shohet JA. Neuroanatomy, Ear. In: *StatPearls*. StatPearls Publishing; 2022. Accessed January 18, 2023. http://www.ncbi.nlm.nih.gov/books/NBK551658/

4. Ekdale EG. Form and function of the mammalian inner ear. *J Anat*. 2016;228(2):324-337. doi:10.1111/joa.12308

5. Kim SH, Nam GS, Choi JY. Pathophysiologic Findings in the Human Endolymphatic Sac in Endolymphatic Hydrops: Functional and Molecular Evidence. *Ann Otol Rhinol Laryngol*. 2019;128(6_suppl):76S-83S. doi:10.1177/0003489419837993

6. Holt JR, Tobin M, Elferich J, et al. Putting the Pieces Together: the Hair Cell Transduction Complex. *JARO J Assoc Res Otolaryngol*. 2021;22(6):601-608. doi:10.1007/s10162-021-00808-0

7. LeMasurier M, Gillespie PG. Hair-Cell Mechanotransduction and Cochlear Amplification. *Neuron*. 2005;48(3):403-415. doi:10.1016/j.neuron.2005.10.017

8. Celesia GG. *Disorders of Peripheral and Central Auditory Processing1: Disorders of Peripheral and Central Auditory Processing*. Elsevier Health Sciences; 2013.

9. Sellon JB, Ghaffari R, Freeman DM. The Tectorial Membrane: Mechanical Properties and Functions. *Cold Spring Harb Perspect Med*. 2019;9(10):a033514. doi:10.1101/cshperspect.a033514

10. Richardson G, Lukashkin A, Russell I. The tectorial membrane: One slice of a complex cochlear sandwich. *Curr Opin Otolaryngol Head Neck Surg*. 2008;16(5):458-464. doi:10.1097/MOO.0b013e32830e20c4

11. Goodyear RJ, Richardson GP. Structure, Function, and Development of the Tectorial Membrane: An Extracellular Matrix Essential for Hearing. *Curr Top Dev Biol*. 2018;130:217-244. doi:10.1016/bs.ctdb.2018.02.006

12. Peterson DC, Reddy V, Hamel RN. Neuroanatomy, Auditory Pathway. In: *StatPearls*. StatPearls Publishing; 2022. Accessed January 19, 2023. http://www.ncbi.nlm.nih.gov/books/NBK532311/

13. Mangold SA, M Das J. Neuroanatomy, Cortical Primary Auditory Area. In: *StatPearls*. StatPearls Publishing; 2022. Accessed January 24, 2023. http://www.ncbi.nlm.nih.gov/books/NBK554521/

14. Musiek FE, Baran JA. *The Auditory System: Anatomy, Physiology, and Clinical Correlates; Second Edition*. Plural Publishing; 2018.

15. Pickles JO. Auditory pathways: anatomy and physiology. *Handb Clin Neurol*. 2015;129:3-25. doi:10.1016/B978-0-444-62630-1.00001-9

16. Møller AR. *Hearing: Anatomy, Physiology, and Disorders of the Auditory System, Third Edition*. Plural Publishing; 2012.

17. Young A, Cornejo J, Spinner A. Auditory Brainstem Response. In: *StatPearls*. StatPearls Publishing; 2023. Accessed October 9, 2023. http://www.ncbi.nlm.nih.gov/books/NBK564321/

18. Lopez-Escamez JA, Carey J, Chung WH, et al. Diagnostic criteria for Menière's disease. *J Vestib Res Equilib Orientat*. 2015;25(1):1-7. doi:10.3233/VES-150549

19. Baloh RW. Prosper Ménière and his disease. *Arch Neurol*. 2001;58(7):1151-1156. doi:10.1001/archneur.58.7.1151

20. Herraiz C, Tapia MC, Plaza G. Tinnitus and Ménière's disease: characteristics and prognosis in a tinnitus clinic sample. *Eur Arch Oto-Rhino-Laryngol Off J Eur Fed Oto-Rhino-Laryngol Soc EUFOS Affil Ger Soc Oto-Rhino-Laryngol - Head Neck Surg*. 2006;263(6):504-509. doi:10.1007/s00405-006-0019-9

21. Ueberfuhr MA, Wiegrebe L, Krause E, Gürkov R, Drexl M. Tinnitus in Normal-Hearing Participants after Exposure to Intense Low-Frequency Sound and in Ménière's Disease Patients. *Front Neurol*. 2017;7:239. doi:10.3389/fneur.2016.00239

22. Espinosa-Sanchez JM, Lopez-Escamez JA. Chapter 19 - Menière's disease. In: Furman JM, Lempert T, eds. *Handbook of Clinical Neurology*. Vol 137. Neuro-Otology. Elsevier; 2016:257-277. doi:10.1016/B978-0-444-63437-5.00019-4

23. Frejo L, Lopez-Escamez JA. Cytokines and Inflammation in Meniere Disease. *Clin Exp Otorhinolaryngol*. 2022;15(1):49-59. doi:10.21053/ceo.2021.00920

24. Belinchon A, Perez-Garrigues H, Tenias JM, Lopez A. Hearing assessment in Menière's disease. *The Laryngoscope*. 2011;121(3):622-626. doi:10.1002/lary.21335

25. Frejo L, Gallego-Martinez A, Requena T, et al. Proinflammatory cytokines and response to molds in mononuclear cells of patients with Meniere disease. *Sci Rep*. 2018;8:5974. doi:10.1038/s41598-018-23911-4

26. Bächinger D, Brühlmann C, Honegger T, et al. Endotype-Phenotype Patterns in Meniere's Disease Based on Gadolinium-Enhanced MRI of the Vestibular Aqueduct. *Front Neurol*. 2019;10:303. doi:10.3389/fneur.2019.00303

27. Frejo L, Soto-Varela A, Santos-Perez S, et al. Clinical Subgroups in Bilateral Meniere Disease. *Front Neurol*. 2016;7:182. doi:10.3389/fneur.2016.00182

28. Frejo L, Martin-Sanz E, Teggi R, et al. Extended phenotype and clinical subgroups in unilateral Meniere disease: A cross-sectional study with cluster analysis. *Clin Otolaryngol Off J ENT-UK Off J Neth Soc Oto-Rhino-Laryngol Cervico-Facial Surg*. 2017;42(6):1172-1180. doi:10.1111/coa.12844

29. De Ridder D, Schlee W, Vanneste S, et al. Tinnitus and tinnitus disorder: Theoretical and operational definitions (an international multidisciplinary proposal). *Prog Brain Res*. 2021;260:1-25. doi:10.1016/bs.pbr.2020.12.002

30. Levine RA, Oron Y. Chapter 23 - Tinnitus. In: Aminoff MJ, Boller F, Swaab DF, eds. *Handbook of Clinical Neurology*. Vol 129. The Human Auditory System. Elsevier; 2015:409-431. doi:10.1016/B978-0-444-62630-1.00023-8

31. Henry JA, Reavis KM, Griest SE, et al. Tinnitus: An Epidemiologic Perspective. *Otolaryngol Clin North Am*. 2020;53(4):481-499. doi:10.1016/j.otc.2020.03.002

32. McCormack A, Edmondson-Jones M, Somerset S, Hall D. A systematic review of the reporting of tinnitus prevalence and severity. *Hear Res*. 2016;337:70-79.

33. Henton A, Tzounopoulos T. What's the buzz? The neuroscience and the treatment of tinnitus. *Physiol Rev*. 2021;101(4):1609-1632. doi:10.1152/physrev.00029.2020

34. Møller AR, Langguth B, DeRidder D, Kleinjung T. *Textbook of Tinnitus*. Springer Science & Business Media; 2010.

35. Tonndorf J. The analogy between tinnitus and pain: a suggestion for a physiological basis of chronic tinnitus. *Hear Res*. 1987;28(2-3):271-275. doi:10.1016/0378-5955(87)90054-2

36. Jastreboff PJ. Phantom auditory perception (tinnitus): mechanisms of generation and perception. *Neurosci Res*. 1990;8(4):221-254. doi:10.1016/0168-0102(90)90031-9

37. Henry JA, Roberts LE, Caspary DM, Theodoroff SM, Salvi RJ. Underlying Mechanisms of Tinnitus: Review and Clinical Implications. *J Am Acad Audiol*. 2014;25(1):5-126. doi:10.3766/jaaa.25.1.2

38. von Bernhardi R, Bernhardi LE von, Eugenín J. What Is Neural Plasticity? *Adv Exp Med Biol*. 2017;1015:1-15. doi:10.1007/978-3-319-62817-2_1

39. Schlee W, Schoisswohl S, Staudinger S, et al. Towards a unification of treatments and interventions for tinnitus patients: The EU research and innovation action UNITI. *Prog Brain Res*. 2021;260:441-451. doi:10.1016/bs.pbr.2020.12.005

40. Schoisswohl S, Langguth B, Schecklmann M, et al. Unification of Treatments and Interventions for Tinnitus Patients (UNITI): a study protocol for a multi-center randomized clinical trial. *Trials*. 2021;22:875. doi:10.1186/s13063-021-05835-z

41. Stephens D, Pyykkö I, Yoshida T, et al. The consequences of tinnitus in long-standing Ménière's disease. *Auris Nasus Larynx*. 2012;39(5):469-474. doi:10.1016/j.anl.2011.10.011

42. Perez-Carpena P, Martinez-Martinez M, Carranza RAM, Batuecas-Caletrio A, Lopez-Escamez JA. A tinnitus symphony in 100 patients with Meniere's disease. *Clin Otolaryngol*. 2019;44(6):1176-1180. doi:10.1111/coa.13438

43. Coey JG, De Jesus O. Hyperacusis. In: *StatPearls*. StatPearls Publishing; 2023. Accessed October 25, 2023. http://www.ncbi.nlm.nih.gov/books/NBK557713/

44. Aazh H, Taylor L, Danesh AA, Moore BCJ. The Effectiveness of Unguided Internet-Based Cognitive Behavioral Therapy for Tinnitus for Patients with Tinnitus Alone or Combined with Hyperacusis and/or Misophonia: A Preliminary Analysis. *J Am Acad Audiol*. Published online May 5, 2023. doi:10.1055/a-2087-0262

45. National Institute of General Medical Sciences. National Institute of General Medical Sciences (NIGMS). Accessed February 6, 2023. https://nigms.nih.gov/

46. Sun H, Shen XR, Fang ZB, et al. Next-Generation Sequencing Technologies and Neurogenetic Diseases. *Life*. 2021;11(4):361. doi:10.3390/life11040361

47. Cardoso JGR, Andersen MR, Herrgård MJ, Sonnenschein N. Analysis of Genetic Variation and Potential Applications in Genome-Scale Metabolic Modeling. *Front Bioeng Biotechnol*. 2015;3:13. doi:10.3389/fbioe.2015.00013

48. McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*. 2008;18(1):148-161. doi:10.1038/cr.2008.4

49. Human Genomic Variation. Genome.gov. Accessed October 19, 2023. https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genomic-variation

50. MacArthur DG, Balasubramanian S, Frankish A, et al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science*. 2012;335(6070):823-828.

51. National Academies of Sciences E, Division H and M, Services B on HC, Populations B on the H of S, Testing C on the EB for G. Understanding Genetic Variance and Phenotype Expression. In: *An Evidence Framework for Genetic Testing*. National Academies Press (US); 2017. Accessed October 19, 2023. https://www.ncbi.nlm.nih.gov/books/NBK425811/

52. Goswami C, Chattopadhyay A, Chuang EY. Rare variants: data types and analysis strategies. *Ann Transl Med*. 2021;9(12):961. doi:10.21037/atm-21-1635

53. A Brief Guide to Genomics. Genome.gov. Published September 14, 2022. Accessed February 6, 2023. https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics

54. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci*. 1977;74(12):5463-5467.

55. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921. doi:10.1038/35057062

56. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. *Hum Immunol*. 2021;82(11):801-811. doi:10.1016/j.humimm.2021.02.012

57. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science*. 2022;376(6588):44-53. doi:10.1126/science.abj6987

58. Miga KH, Wang T. The Need for a Human Pangenome Reference Sequence. *Annu Rev Genomics Hum Genet*. 2021;22:81-102. doi:10.1146/annurev-genom-120120-081921

59. Lightbody G, Haberland V, Browne F, et al. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Brief Bioinform*. 2019;20(5):1795-1811. doi:10.1093/bib/bby051

60. Sayers EW, Beck J, Bolton EE, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2021;49(D1):D10-D17. doi:10.1093/nar/gkaa892

61. Leinonen R, Akhtar R, Birney E, et al. The European Nucleotide Archive. *Nucleic Acids Res*. 2011;39(Database issue):D28-D31. doi:10.1093/nar/gkq967

62. Athar A, Füllgrabe A, George N, et al. ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res*. 2019;47(D1):D711-D715. doi:10.1093/nar/gky964

63. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

64. Peña-Chilet M, Roldán G, Perez-Florido J, et al. CSVS, a crowdsourcing database of the Spanish population genetic variability. *Nucleic Acids Res*. 2021;49(D1):D1130-D1137. doi:10.1093/nar/gkaa794

65. López-López D, Roldán G, Fernández-Rueda JL, et al. A crowdsourcing database for the copy-number variation of the Spanish population. *Hum Genomics*. 2023;17(1):20. doi:10.1186/s40246-023-00466-8

66. UK Biobank - UK Biobank. https://www.ukbiobank.ac.uk/

67. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006. doi:10.1101/gr.229102

68. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892

69. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15(7):565-574. doi:10.1038/gim.2013.73

70. Richards S, Aziz N, Bale 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*. 2015;17(5):405-423. doi:10.1038/gim.2015.30

71. Hodgkinson A, Casals F, Idaghdour Y, Grenier JC, Hernandez RD, Awadalla P. Selective constraint, background selection, and mutation accumulation variability within and between human populations. *BMC Genomics*. 2013;14:495. doi:10.1186/1471-2164-14-495

72. Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the human genome. *Nat Genet*. 2019;51(1):88-95. doi:10.1038/s41588-018-0294-6

73. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet*. 2014;95(1):5-23. doi:10.1016/j.ajhg.2014.06.009

74. Genome-Wide Association Studies (GWAS). https://www.genome.gov/genetics-glossary/Genome-Wide-Association-Studies

75. Masterson EA, Themann CL, Luckhaupt SE, Li J, Calvert GM. Hearing difficulty and tinnitus among U.S. workers and non-workers in 2007. *Am J Ind Med*. 2016;59(4):290-300. doi:10.1002/ajim.22565

76. Cederroth CR, PirouziFard M, Trpchevska N, et al. Association of Genetic vs Environmental Factors in Swedish Adoptees With Clinically Significant Tinnitus. *JAMA Otolaryngol-- Head Neck Surg*. 2019;145(3):222-229. doi:10.1001/jamaoto.2018.3852

77. Hendrickx JJ, Huyghe JR, Demeester K, et al. Familial aggregation of tinnitus: a European multicentre study. *B-ENT*. 2007;3 Suppl 7:51-60.

78. Trpchevska N, Bulla J, Prada Hellberg M, et al. Sex-Dependent Aggregation of Tinnitus in Swedish Families. *J Clin Med*. 2020;9(12):3812. doi:10.3390/jcm9123812

79. Maas IL, Brüggemann P, Requena T, et al. Genetic susceptibility to bilateral tinnitus in a Swedish twin cohort. *Genet Med Off J Am Coll Med Genet*. 2017;19(9):1007-1012. doi:10.1038/gim.2017.4

80. Bogo R, Farah A, Karlsson KK, Pedersen NL, Svartengren M, Skjönsberg Å. Prevalence, Incidence Proportion, and Heritability for Tinnitus: A Longitudinal Twin Study. *Ear Hear*. 2017;38(3):292-300. doi:10.1097/AUD.0000000000000397

81. Bhatt IS, Wilson N, Dias R, Torkamani A. A genome-wide association study of tinnitus reveals shared genetic links to neuropsychiatric disorders. *Sci Rep*. 2022;12:22511. doi:10.1038/s41598-022-26413-6

82. Xie C, Niu Y, Ping J, et al. Genome-wide association study identifies new loci associated with noise-induced tinnitus in Chinese populations. *BMC Genomic Data*. 2021;22(1):31. doi:10.1186/s12863-021-00987-y

83. Wells HRR, Abidin FNZ, Freidin MB, Williams FMK, Dawson SJ. Genome-wide association study suggests that variation at the RCOR1 locus is associated with tinnitus in UK Biobank. *Sci Rep*. 2021;11(1):6470. doi:10.1038/s41598-021-85871-6

84. Urbanek ME, Zuo J. Genetic predisposition to tinnitus in the UK Biobank population. *Sci Rep*. 2021;11(1):18150. doi:10.1038/s41598-021-97350-z

85. Haider HF, Flook M, Aparicio M, et al. Biomarkers of Presbycusis and Tinnitus in a Portuguese Older Population. *Front Aging Neurosci*. 2017;9:346. doi:10.3389/fnagi.2017.00346

86. Jeong JE, Jeon S, Han JS, et al. The Mediating Effect of Psychological Distress on the Association between BDNF, 5-HTTLPR, and Tinnitus Severity. *Psychiatry Investig*. 2021;18(3):187-195. doi:10.30773/pi.2020.0295

87. Yuce S, Sancakdar E, Bağcı G, et al. Angiotensin-Converting Enzyme (ACE) I/D and Alpha-Adducin (ADD1) G460W Gene Polymorphisms in Turkish Patients with Severe Chronic Tinnitus. *J Int Adv Otol*. 2016;12(1):77-81. doi:10.5152/iao.2016.1732

88. Watabe T, Kanzaki S, Sato N, Matsunaga T, Muramatsu M, Ogawa K. Single nucleotide polymorphisms in tinnitus patients exhibiting severe distress. *Sci Rep*. 2020;10(1):13023. doi:10.1038/s41598-020-69467-0

89. Bhatt IS, Dias R, Torkamani A. Association Analysis of Candidate Gene Polymorphisms and Tinnitus in Young Musicians. *Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol*. 2021;42(9):e1203-e1212. doi:10.1097/MAO.0000000000003279

90. Lechowicz U, Pollak A, Raj-Koziak D, et al. Tinnitus in patients with hearing loss due to mitochondrial DNA pathogenic variants. *Eur Arch Oto-Rhino-Laryngol Off J Eur Fed Oto-Rhino-Laryngol Soc EUFOS Affil Ger Soc Oto-Rhino-Laryngol - Head Neck Surg*. 2018;275(8):1979-1985. doi:10.1007/s00405-018-5028-y

91. Orenay-Boyacioglu S, Caliskan M, Boyacioglu O, Coskunoglu A, Bozkurt G, Cam FS. Chronic tinnitus and BDNF/GDNF CpG promoter methylations: a case-control study. *Mol Biol Rep*. 2019;46(4):3929-3936. doi:10.1007/s11033-019-04837-0

92. Amanat S, Gallego-Martinez A, Sollini J, et al. Burden of rare variants in synaptic genes in patients with severe tinnitus: An exome based extreme phenotype study. *EBioMedicine*. 2021;66:103309. doi:10.1016/j.ebiom.2021.103309

93. Gallego-Martinez A, Escalera-Balsera A, Trpchevska N, et al. Using coding and non-coding rare variants to target candidate genes in patients with severe tinnitus. *NPJ Genomic Med*. 2022;7:70. doi:10.1038/s41525-022-00341-w

94. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20(8):467-484. doi:10.1038/s41576-019-0127-1

95. Trpchevska N, Freidin MB, Broer L, et al. Genome-wide association meta-analysis identifies 48 risk variants and highlights the role of the stria vascularis in hearing loss. *Am J Hum Genet*. 2022;109(6):1077-1091. doi:10.1016/j.ajhg.2022.04.010

96. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *J Clin Epidemiol*. 2009;62(10):1006-1012. doi:10.1016/j.jclinepi.2009.06.005

97. Karczewski KJ, Weisburd B, Thomas B, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45(Database issue):D840-D845. doi:10.1093/nar/gkw971

98. Ameur A, Dahlberg J, Olason P, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur J Hum Genet*. 2017;25(11):1253-1260. doi:10.1038/ejhg.2017.130

99. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(Database issue):D886-D894. doi:10.1093/nar/gky1016

100. Newman CW, Jacobson GP, Spitzer JB. Development of the Tinnitus Handicap Inventory. *Arch Otolaryngol Neck Surg*. 1996;122(2):143-148. doi:10.1001/archotol.1996.01890140029007

101. Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? *Med Decis Mak Int J Soc Med Decis Mak*. 2001;21(4):329-334. doi:10.1177/0272989X0102100408

102. Herráiz C, de los Santos G, Diges I, Díez R, Aparicio JM, Herráiz C. Evaluación de la hiperacusia: test de hipersensibilidad al sonido. *Acta Otorrinolaringológica Esp*. 2006;57(7):303-306. doi:10.1016/S0001-6519(06)78716-7

103. Nelting M, Rienhoff NK, Hesse G, Lamparter U. [The assessment of subjective distress related to hyperacusis with a self-rating questionnaire on hypersensitivity to sound]. *Laryngorhinootologie*. 2002;81(5):327-334. doi:10.1055/s-2002-28342

104. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. *J Gen Intern Med*. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x

105. Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and Utility of the Patient Health Questionnaire in Diagnosing Mental Disorders in 1003 General Hospital Spanish Inpatients. *Psychosom Med*. 2001;63(4):679.

106. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 1983;67(6):361-370. doi:10.1111/j.1600-0447.1983.tb09716.x

107. Cabrera V, Martín-Aragón M, Terol M del C, Núñez R, Pastor M de los Á. La Escala de Ansiedad y Depresión Hospitalaria (HAD) en fibromialgia: Análisis de sensibilidad y especificidad. *Ter Psicológica*. 2015;33(3):181-193. doi:10.4067/S0718-48082015000300003

108. R Core Team (2022). R: A language and environment for statistical computing. Published online 2022. https://www.R-project.org/

109. Szczepek AJ, Frejo L, Vona B, et al. Recommendations on Collecting and Storing Samples for Genetic Studies in Hearing and Tinnitus Research. *Ear Hear*. 2019;40(2):219-226. doi:10.1097/AUD.0000000000000614

110. Garcia M, Juhos S, Larsson M, et al. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Research*. 2020;9:63. doi:10.12688/f1000research.16665.2

111. Ewels PA, Peltzer A, Fillinger S, et al. nf-core: Community curated bioinformatics pipelines. Published online April 16, 2019:610741. doi:10.1101/610741

112. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324

113. Van der Auwera GA, O'Connor BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st Edition. O'Reilly Media; 2020.

114. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993. doi:10.1093/bioinformatics/btr509

115. Tiao G, Goodrich J. gnomAD v3.1 New Content, Methods, Annotations, and Data Availability. gnomAD v3.1 New Content, Methods, Annotations, and Data Availability. Published 2020. Accessed March 16, 2023. gnomad.broadinstitute.org/news/2020-10-gnomad-v3-1-new-content-methods-annotations-and-data-availability/

116. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17:122. doi:10.1186/s13059-016-0974-4

117. Ensembl Variation - Calculated variant consequences. Ensembl Variation - Calculated variant consequences. Accessed March 20, 2023. www.ensembl.org/info/genome/variation/prediction/predicted_data.html

118. Chen S, Francioli LC, Goodrich JK, et al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. Published online October 10, 2022:2022.03.20.485034. doi:10.1101/2022.03.20.485034

119. Lift Genome Annotations - UCSC. Lift Genome Annotations. Accessed November 9, 2021. genome.ucsc.edu/cgi-bin/hgLiftOver

120. Selection of extreme phenotypes: the role of clinical observation in translational research - PMC. Accessed October 25, 2023. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2997959/

121. Walsh R, Mazzarotto F, Whiffin N, et al. Quantitative approaches to variant classification increase the yield and precision of genetic testing in Mendelian diseases: the case of hypertrophic cardiomyopathy. *Genome Med*. 2019;11:5. doi:10.1186/s13073-019-0616-z

122. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJ, van Karnebeek C, Wasserman WW. FLAGS, frequently mutated genes in public exomes. *BMC Med Genomics*. 2014;7:64. doi:10.1186/s12920-014-0064-y

123. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJ, van Karnebeek C, Wasserman WW. Correction to: FLAGS, frequently mutated genes in public exomes. *BMC Med Genomics*. 2017;10:69. doi:10.1186/s12920-017-0309-7

124. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol*. 2016;12(4):e1004873. doi:10.1371/journal.pcbi.1004873

125. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinforma Oxf Engl*. 2016;32(8):1220-1222. doi:10.1093/bioinformatics/btv710

126. Eisfeldt J, Vezzi F, Olason P, Nilsson D, Lindstrand A. TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Research*. 2017;6:664. doi:10.12688/f1000research.11168.2

127. Geoffroy V, Lamouche JB, Guignard T, et al. The AnnotSV webserver in 2023: updated visualization and ranking. *Nucleic Acids Res*. 2023;51(W1):W39-W45. doi:10.1093/nar/gkad426

128. Eisfeldt J. SVDB. Published online July 15, 2023. Accessed July 17, 2023. https://github.com/J35P312/SVDB

129. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178-192. doi:10.1093/bib/bbs017

130. Gene constraint | gnomAD. Accessed September 29, 2023. https://gnomad.broadinstitute.org/help/constraint

131. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

132. Cai T, Jen HI, Kang H, Klisch TJ, Zoghbi HY, Groves AK. Characterization of the transcriptome of nascent hair cells and identification of direct targets of the Atoh1 transcription factor. *J Neurosci Off J Soc Neurosci*. 2015;35(14):5870-5883. doi:10.1523/JNEUROSCI.5083-14.2015

133. Lu CC, Appler JM, Houseman EA, Goodrich LV. Developmental Profiling of Spiral Ganglion Neurons Reveals Insights into Auditory Circuit Assembly. *J Neurosci*. 2011;31(30):10903-10918. doi:10.1523/JNEUROSCI.2358-11.2011

134. Garcia-Moreno A, López-Domínguez R, Villatoro-García JA, et al. Functional Enrichment Analysis of Regulatory Elements. *Biomedicines*. 2022;10(3):590. doi:10.3390/biomedicines10030590

135. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556

136. Gene Ontology Consortium, Aleksander SA, Balhoff J, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1):iyad031. doi:10.1093/genetics/iyad031

137. Köhler S, Gargano M, Matentzoglu N, et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res*. 2021;49(D1):D1207-D1217. doi:10.1093/nar/gkaa1043

138. Smith CL, Eppig JT. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med*. 2009;1(3):390-399. doi:10.1002/wsbm.44

139. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Published online 2016. https://ggplot2.tidyverse.org

140. Schloerke B, Cook D, Larmarange J, et al. GGally: Extension to “ggplot2.” Published online 2021. https://CRAN.R-project.org/package=GGally

141. Brunson JC, Read QD. ggalluvial: Alluvial Plots in “ggplot2.” Published online 2023. http://corybrunson.github.io/ggalluvial/

142. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinforma Oxf Engl*. 2016;32(18):2847-2849. doi:10.1093/bioinformatics/btw313

143. Krassowski M, Arts M, Lagger C, Max. krassowski/complex-upset: v1.3.5. Published online November 11, 2022. doi:10.5281/zenodo.7314197

144. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinforma Oxf Engl*. 2014;30(19):2811-2812. doi:10.1093/bioinformatics/btu393

145. Wickham H, Francois R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. Published online 2022. https://CRAN.R-project.org/package=dplyr

146. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4(43):1686. doi:10.21105/joss.01686

147. Wickham H. stringr: Simple, Consistent Wrappers for Common String Operations. Published online 2019. https://CRAN.R-project.org/package=stringr

148. Wickham H. forcats: Tools for Working with Categorical Variables (Factors). Published online 2021. https://CRAN.R-project.org/package=forcats

149. 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*. 1995;113(3):181-185. doi:10.1016/S0194-5998(95)70102-8

150. Requena T, Cabrera S, Martín-Sierra C, Price SD, Lysakowski A, Lopez-Escamez JA. Identification of two novel mutations in FAM136A and DTNA genes in autosomal-dominant familial Meniere’s disease. *Hum Mol Genet*. 2015;24(4):1119-1126. doi:10.1093/hmg/ddu524

151. Martín-Sierra C, Requena T, Frejo L, 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*. 2016;25(16):3407-3415. doi:10.1093/hmg/ddw183

152. Kim BJ, Kim AR, Han KH, et al. Distinct vestibular phenotypes in DFNA9 families with COCH variants. *Eur Arch Oto-Rhino-Laryngol Off J Eur Fed Oto-Rhino-Laryngol Soc EUFOS Affil Ger Soc Oto-Rhino-Laryngol - Head Neck Surg*. 2016;273(10):2993-3002. doi:10.1007/s00405-015-3885-1

153. Martín-Sierra C, Gallego-Martinez A, Requena T, Frejo L, Batuecas-Caletrío A, Lopez-Escamez JA. Variable expressivity and genetic heterogeneity involving DPT and SEMA3D genes in autosomal dominant familial Meniere’s disease. *Eur J Hum Genet EJHG*. 2017;25(2):200-207. doi:10.1038/ejhg.2016.154

154. Frykholm C, Klar J, Tomanovic T, Ameur A, Dahl N. Stereocilin gene variants associated with episodic vertigo: expansion of the DFNB16 phenotype. *Eur J Hum Genet EJHG*. 2018;26(12):1871-1874. doi:10.1038/s41431-018-0256-6

155. Skarp S, Kanervo L, Kotimäki J, Sorri M, Männikkö M, Hietikko E. Whole-exome sequencing suggests multiallelic inheritance for childhood-onset Ménière's disease. *Ann Hum Genet*. 2019;83(6):389-396. doi:10.1111/ahg.12327

156. Roman-Naranjo P, Gallego-Martinez A, Soto-Varela A, et al. Burden of Rare Variants in the OTOG Gene in Familial Meniere's Disease. *Ear Hear*. 2020;41(6):1598-1605. doi:10.1097/AUD.0000000000000878

157. Mehrjoo Z, Kahrizi K, Mohseni M, et al. Limbic System Associated Membrane Protein Mutation in an Iranian Family Diagnosed with Ménière's Disease. *Arch Iran Med*. 2020;23(5):319-325. doi:10.34172/aim.2020.21

158. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30.

159. Gui Y, Liu X, Wang C, Yang P. Overexpressing PTTG family genes predict poor prognosis in kidney renal clear cell carcinoma. *World J Surg Oncol*. 2021;19:111. doi:10.1186/s12957-021-02225-2

160. Rodenas-Cuadrado P, Ho J, Vernes SC. Shining a light on CNTNAP2: complex functions to complex disorders. *Eur J Hum Genet*. 2014;22(2):171-178. doi:10.1038/ejhg.2013.100

161. Cho C. Testicular and epididymal ADAMs: expression and function during fertilization. *Nat Rev Urol*. 2012;9(10):550-560. doi:10.1038/nrurol.2012.167

162. Naidoo M, Levine F, Gillot T, et al. MicroRNA-1205 Regulation of FRYL in Prostate Cancer. *Front Cell Dev Biol*. 2021;9:647485. doi:10.3389/fcell.2021.647485

163. Althali NJ, Hentges KE. Genetic insights into non-syndromic Tetralogy of Fallot. *Front Physiol*. 2022;13:1012665. doi:10.3389/fphys.2022.1012665

164. Bezzina CR, Barc J, Mizusawa Y, et al. Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat Genet*. 2013;45(9):1044-1049. doi:10.1038/ng.2712

165. Biswas R, Lugo A, Akeroyd MA, Schlee W, Gallus S, Hall DA. Tinnitus prevalence in Europe: a multi-country cross-sectional population study. *Lancet Reg Health - Eur*. 2021;12:100250. doi:10.1016/j.lanepe.2021.100250

166. Cederroth CR, Lugo A, Edvall NK, et al. Association between Hyperacusis and Tinnitus. *J Clin Med*. 2020;9(8):2412. doi:10.3390/jcm9082412

167. Aazh H, Moore BCJ. Factors related to uncomfortable loudness levels for patients seen in a tinnitus and hyperacusis clinic. *Int J Audiol*. 2017;56(10):793-800. doi:10.1080/14992027.2017.1335888

168. Bhatt JM, Bhattacharyya N, Lin HW. Relationships Between Tinnitus And The Prevalence Of Anxiety And Depression. *The Laryngoscope*. 2017;127(2):466-469. doi:10.1002/lary.26107

169. Wang Y, Chen G, Tang Z, et al. Loss-of-function mutations in IQCN cause male infertility in humans and mice owing to total fertilization failure. *Mol Hum Reprod*. 2023;29(7):gaad018. doi:10.1093/molehr/gaad018

170. Huang HY, Lin YCD, Cui S, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res*. 2022;50(D1):D222-D230. doi:10.1093/nar/gkab1079

171. Zhai J, Zhang P, Zhang N, Luo Y, Wu Y. Analysis of WDFY4 rs7097397 and PHLDB1 rs7389 polymorphisms in Chinese patients with systemic lupus erythematosus. *Clin Rheumatol*. 2022;41(7):2035-2042. doi:10.1007/s10067-022-06103-4

172. Coe BP, Stessman HAF, Sulovari A, et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat Genet*. 2019;51(1):106-116. doi:10.1038/s41588-018-0288-4

173. Kochi Y, Kamatani Y, Kondo Y, et al. Splicing variant of WDFY4 augments MDA5 signalling and the risk of clinically amyopathic dermatomyositis. *Ann Rheum Dis*. 2018;77(4):602-611. doi:10.1136/annrheumdis-2017-212149

174. Potashkin JA, Bottero V, Santiago JA, Quinn JP. Bioinformatic Analysis Reveals Phosphodiesterase 4D-Interacting Protein as a Key Frontal Cortex Dementia Switch Gene. *Int J Mol Sci*. 2020;21(11):3787. doi:10.3390/ijms21113787

175. Schrauwen I, Hasin-Brumshtein Y, Corneveaux JJ, et al. A comprehensive catalogue of the coding and non-coding transcripts of the human inner ear. *Hear Res*. 2016;333:266-274. doi:10.1016/j.heares.2015.08.013

176. Lu AT, Hannon E, Levine ME, et al. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat Commun*. 2017;8:15353. doi:10.1038/ncomms15353

177. Ravikumar G, Ashok Murthy V. A Study of Brainstem Auditory Evoked Responses in Normal Hearing Patients with Tinnitus. *Indian J Otolaryngol Head Neck Surg*. 2016;68(4):429-433. doi:10.1007/s12070-015-0917-5

178. Jacxsens L, De Pauw J, Cardon E, et al. Brainstem evoked auditory potentials in tinnitus: A best-evidence synthesis and meta-analysis. *Front Neurol*. 2022;13:941876. doi:10.3389/fneur.2022.941876

179. Nip K, Kashiwagura S, Kim JH. Loss of β4-spectrin impairs Nav channel clustering at the heminode and temporal fidelity of presynaptic spikes in developing auditory brain. *Sci Rep*. 2022;12(1):5854. doi:10.1038/s41598-022-09856-9

180. Douki T, von Koschembahr A, Cadet J. Insight in DNA Repair of UV-induced Pyrimidine Dimers by Chromatographic Methods. *Photochem Photobiol*. 2017;93(1):207-215. doi:10.1111/php.12685

181. Salem ME, Bodor JN, Puccini A, et al. Relationship between MLH1, PMS2, MSH2 and MSH6 gene-specific alterations and tumor mutational burden in 1057 microsatellite instability-high solid tumors. *Int J Cancer*. 2020;147(10):2948-2956. doi:10.1002/ijc.33115

182. Liu YL, Cadoo KA, Maio A, et al. Early age of onset and broad cancer spectrum persist in MSH6- and PMS2-associated Lynch syndrome. *Genet Med Off J Am Coll Med Genet*. 2022;24(6):1187-1195. doi:10.1016/j.gim.2022.02.016

183. Lin X, Wu Y, Li Q, et al. Genetic Association of ERCC6 rs2228526 Polymorphism with the Risk of Cancer: Evidence from a Meta-Analysis. *BioMed Res Int*. 2022;2022:2662666. doi:10.1155/2022/2662666

184. Yue WY, Clark JJ, Telisak M, Hansen MR. Inhibition of c-Jun N-terminal kinase activity enhances vestibular schwannoma cell sensitivity to γ-irradiation. *Neurosurgery*. 2013;73(3):506-516. doi:10.1227/01.neu.0000431483.10031.89

185. Raabe TD, Deadwyler G, Varga JW, Devries GH. Localization of neuregulin isoforms and erbB receptors in myelinating glial cells. *Glia*. 2004;45(2):197-207. doi:10.1002/glia.10311

186. Fallon M, Tadi P. Histology, Schwann Cells. In: *StatPearls*. StatPearls Publishing; 2023. Accessed October 10, 2023. http://www.ncbi.nlm.nih.gov/books/NBK544316/

187. Torii T, Miyamoto Y, Takada S, et al. In vivo knockdown of ErbB3 in mice inhibits Schwann cell precursor migration. *Biochem Biophys Res Commun*. 2014;452(3):782-788. doi:10.1016/j.bbrc.2014.08.156

188. Vona B, Lechno S, Hofrichter MAH, et al. Confirmation of PDZD7 as a Nonsyndromic Hearing Loss Gene. *Ear Hear*. 2016;37(4):e238-246. doi:10.1097/AUD.0000000000000278

189. Liu H, Pecka JL, Zhang Q, Soukup GA, Beisel KW, He DZZ. Characterization of transcriptomes of cochlear inner and outer hair cells. *J Neurosci Off J Soc Neurosci*. 2014;34(33):11085-11095. doi:10.1523/JNEUROSCI.1690-14.2014

190. Tziridis K, Forster J, Buchheidt-Dörfler I, et al. Tinnitus development is associated with synaptopathy of inner hair cells in Mongolian gerbils. *Eur J Neurosci*. 2021;54(3):4768-4780. doi:10.1111/ejn.15334

191. Singer W, Zuccotti A, Jaumann M, et al. Noise-induced inner hair cell ribbon loss disturbs central arc mobilization: a novel molecular paradigm for understanding tinnitus. *Mol Neurobiol*. 2013;47(1):261-279. doi:10.1007/s12035-012-8372-8

192. Lowry ER, Kruyer A, Norris EH, Cederroth CR, Strickland S. The GluK4 Kainate Receptor Subunit Regulates Memory, Mood, and Excitotoxic Neurodegeneration. *Neuroscience*. 2013;235:215-225. doi:10.1016/j.neuroscience.2013.01.029

193. Aller MI, Pecoraro V, Paternain AV, Canals S, Lerma J. Increased Dosage of High-Affinity Kainate Receptor Gene grik4 Alters Synaptic Transmission and Reproduces Autism Spectrum Disorders Features. *J Neurosci*. 2015;35(40):13619-13628. doi:10.1523/JNEUROSCI.2217-15.2015

194. Shin A, Park S, Shin W, et al. A brainstem-to-mediodorsal thalamic pathway mediates sound-induced arousal from slow-wave sleep. *Curr Biol CB*. 2023;33(5):875-885.e5. doi:10.1016/j.cub.2023.01.033

195. Augustin I, Betz A, Herrmann C, Jo T, Brose N. Differential expression of two novel Munc13 proteins in rat brain. *Biochem J*. 1999;337 (Pt 3)(Pt 3):363-371.

196. Dudenhöffer-Pfeifer M, Schirra C, Pattu V, et al. Different Munc13 isoforms function as priming factors in lytic granule release from murine cytotoxic T lymphocytes. *Traffic Cph Den*. 2013;14(7):798-809. doi:10.1111/tra.12074

197. Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med*. 2013;5(5):48. doi:10.1186/gm452

198. Pasanen P, Myllykangas L, Pöyhönen M, et al. Genetics of dementia in a Finnish cohort. *Eur J Hum Genet*. 2018;26(6):827-837. doi:10.1038/s41431-018-0117-3

199. Ogata S, Hashizume K, Hayase Y, et al. Potential involvement of DSCAML1 mutations in neurodevelopmental disorders. *Genes Cells Devoted Mol Cell Mech*. 2021;26(3):136-151. doi:10.1111/gtc.12831

200. Ma M, Brunal AA, Clark KC, et al. Deficiency in the cell-adhesion molecule dscaml1 impairs hypothalamic CRH neuron development and perturbs normal neuroendocrine stress axis function. *Front Cell Dev Biol*. 2023;11:1113675. doi:10.3389/fcell.2023.1113675

201. Wang W, Li J, Tan J, et al. Endonuclease G promotes autophagy by suppressing mTOR signaling and activating the DNA damage response. *Nat Commun*. 2021;12:476. doi:10.1038/s41467-020-20780-2

202. Han W, Shi X, Nuttall AL. AIF and endoG translocation in noise exposure induced hair cell death. *Hear Res*. 2006;211(1-2):85-95. doi:10.1016/j.heares.2005.10.004

203. Yoshida Y, Yasuda S, Fujita T, et al. Ubiquitination of exposed glycoproteins by SCFFBXO27 directs damaged lysosomes for autophagy. *Proc Natl Acad Sci U S A*. 2017;114(32):8574-8579. doi:10.1073/pnas.1702615114

204. Verkerk AJMH, Schot R, Dumee B, et al. Mutation in the AP4M1 Gene Provides a Model for Neuroaxonal Injury in Cerebral Palsy. *Am J Hum Genet*. 2009;85(1):40-52. doi:10.1016/j.ajhg.2009.06.004

205. Akbar W, Ullah A, Haider N, et al. Identification of novel homozygous variants in FOXE3 and AP4M1 underlying congenital syndromic anophthalmia and microphthalmia. *J Gene Med*. Published online September 27, 2023:e3601. doi:10.1002/jgm.3601

206. Alfahed A, Ebili HO, Almoammar NE, et al. Prognostic Values of Gene Copy Number Alterations in Prostate Cancer. *Genes*. 2023;14(5):956. doi:10.3390/genes14050956

207. Shinde V, Sobreira N, Wohler ES, et al. Pathogenic alleles in microtubule, secretory granule and extracellular matrix-related genes in familial keratoconus. *Hum Mol Genet*. 2021;30(8):658-671. doi:10.1093/hmg/ddab075

208. Kwon YJ, Kim JO, Park JM, et al. Identification of Genetic Factors Underlying the Association between Sodium Intake Habits and Hypertension Risk. *Nutrients*. 2020;12(9):2580. doi:10.3390/nu12092580

209. Bieniossek C, Papai G, Schaffitzel C, et al. The architecture of human general transcription factor TFIID core complex. *Nature*. 2013;493(7434):699-702. doi:10.1038/nature11791

210. Yuan B, Pehlivan D, Karaca E, et al. Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. *J Clin Invest*. 2015;125(2):636-651. doi:10.1172/JCI77435

211. Lin SZ, Feng JH, Sun LP, Ma HW, Wang WQ, Li JY. Novel compound heterozygous variants in the TAF6 gene in a patient with Alazami-Yuan syndrome: A case report. *World J Clin Cases*. 2022;10(6):1889-1895. doi:10.12998/wjcc.v10.i6.1889

212. Simonis-Bik AM, Nijpels G, van Haeften TW, et al. Gene Variants in the Novel Type 2 Diabetes Loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B Affect Different Aspects of Pancreatic β-Cell Function. *Diabetes*. 2010;59(1):293-301. doi:10.2337/db09-1048

213. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44(9):981-990. doi:10.1038/ng.2383

214. Li C, Chi H, Deng S, et al. THADA drives Golgi residency and upregulation of PD-L1 in cancer cells and provides promising target for immunotherapy. *J Immunother Cancer*. 2021;9(8):e002443. doi:10.1136/jitc-2021-002443

215. Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*. 2011;43(1):55-59. doi:10.1038/ng.732

216. Bianconi D, Unseld M, Prager GW. Integrins in the Spotlight of Cancer. *Int J Mol Sci*. 2016;17(12):2037. doi:10.3390/ijms17122037

217. Zhuang H, Zhou Z, Ma Z, et al. Characterization of the prognostic and oncologic values of ITGB superfamily members in pancreatic cancer. *J Cell Mol Med*. 2020;24(22):13481-13493. doi:10.1111/jcmm.15990

218. Shi W, He J, Huang Y, et al. Integrin β5 enhances the malignancy of human colorectal cancer by increasing the TGF-β signaling. *Anticancer Drugs*. 2021;32(7):717-726. doi:10.1097/CAD.0000000000001050

219. Xu J, Zeng Y, Si H, et al. Integrating transcriptome-wide association study and mRNA expression profile identified candidate genes related to hand osteoarthritis. *Arthritis Res Ther*. 2021;23(1):81. doi:10.1186/s13075-021-02458-2

220. Li M, Shen YJ, Chai S, Bai YL, Li ZH. miR-133a-3p inhibits the osteogenic differentiation of bone marrow mesenchymal stem cells by regulating ankyrin repeat domain 44. *Gen Physiol Biophys*. 2021;40(4):329-339. doi:10.4149/gpb_2020038

221. La Ferla M, Lessi F, Aretini P, et al. ANKRD44 Gene Silencing: A Putative Role in Trastuzumab Resistance in Her2-Like Breast Cancer. *Front Oncol*. 2019;9:547. doi:10.3389/fonc.2019.00547

222. Gou K, Liu J, Feng X, Li H, Yuan Y, Xing C. Expression of Minichromosome Maintenance Proteins (MCM) and Cancer Prognosis: A meta-analysis. *J Cancer*. 2018;9(8):1518-1526. doi:10.7150/jca.22691

223. Snyder M, Huang XY, Zhang JJ. The Minichromosome Maintenance Proteins 2-7 (MCM2-7) Are Necessary for RNA Polymerase II (Pol II)-mediated Transcription. *J Biol Chem*. 2009;284(20):13466-13472. doi:10.1074/jbc.M809471200

224. F W, R H, X H, et al. CFTR potentiator ivacaftor protects against noise-induced hair cell loss by increasing Nrf2 and reducing oxidative stress. *Biomed Pharmacother Biomedecine Pharmacother*. 2023;166. doi:10.1016/j.biopha.2023.115399

225. K H, Kk M, Ct A, et al. Interaction between CFTR and prestin (SLC26A5). *Biochim Biophys Acta*. 2010;1798(6). doi:10.1016/j.bbamem.2010.02.001

226. Vlajkovic SM, Housley GD, Thorne PR, et al. Preservation of cochlear function in Cd39 deficient mice. *Hear Res*. 2009;253(1-2):77-82. doi:10.1016/j.heares.2009.03.009

227. Batters C, Arthur CP, Lin A, et al. Myo1c is designed for the adaptation response in the inner ear. *EMBO J*. 2004;23(7):1433-1440. doi:10.1038/sj.emboj.7600169

228. Lin T, Greenberg MJ, Moore JR, Ostap EM. A Hearing-Loss Associated Myo1c Mutation (R156W) Decreases the Myosin Duty Ratio and Force Sensitivity. *Biochemistry*. 2011;50(11):1831-1838. doi:10.1021/bi1016777

229. Siletti K, Tarchini B, Hudspeth AJ. Daple coordinates organ-wide and cell-intrinsic polarity to pattern inner-ear hair bundles. *Proc Natl Acad Sci U S A*. 2017;114(52):E11170-E11179. doi:10.1073/pnas.1716522115

230. Ozono Y, Tamura A, Nakayama S, et al. Daple deficiency causes hearing loss in adult mice by inducing defects in cochlear stereocilia and apical microtubules. *Sci Rep*. 2021;11:20224. doi:10.1038/s41598-021-96232-8

231. Roman-Naranjo P, Parra-Perez AM, Escalera-Balsera A, et al. Defective α-tectorin may involve tectorial membrane in familial Meniere disease. *Clin Transl Med*. 2022;12(6):e829. doi:10.1002/ctm2.829

232. Kammerer R, Rüttiger L, Riesenberg R, et al. Loss of mammal-specific tectorial membrane component carcinoembryonic antigen cell adhesion molecule 16 (CEACAM16) leads to hearing impairment at low and high frequencies. *J Biol Chem*. 2012;287(26):21584-21598. doi:10.1074/jbc.M111.320481

233. Acke FRE, De Leenheer EMR. Hearing Loss in Stickler Syndrome: An Update. *Genes*. 2022;13(9):1571. doi:10.3390/genes13091571

234. Rad A, Schade-Mann T, Gamerdinger P, et al. Aberrant COL11A1 splicing causes prelingual autosomal dominant nonsyndromic hearing loss in the DFNA37 locus. *Hum Mutat*. 2021;42(1):25-30. doi:10.1002/humu.24136

235. Corriols-Noval P, López Simón EC, Cadiñanos J, et al. Clinical Impact of Genetic Diagnosis of Sensorineural Hearing Loss in Adults. *Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol*. 2022;43(10):1125-1136. doi:10.1097/MAO.0000000000003706

236. Habib AM, Matsuyama A, Okorokov AL, et al. A novel human pain insensitivity disorder caused by a point mutation in ZFHX2. *Brain J Neurol*. 2018;141(2):365-376. doi:10.1093/brain/awx326

237. Komine Y, Takao K, Miyakawa T, Yamamori T. Behavioral abnormalities observed in Zfhx2-deficient mice. *PloS One*. 2012;7(12):e53114. doi:10.1371/journal.pone.0053114

238. Dong Y, Yan J, Xu W, Paquet-Durand F, Hu Z, Jiao K. HDAC inhibition delays photoreceptor loss in Pde6b mutant mice of retinitis pigmentosa: insights from scRNA-seq and CUT&Tag. *PeerJ*. 2023;11:e15659. doi:10.7717/peerj.15659

239. Manchev VT, Hilpert M, Berrou E, et al. A new form of macrothrombocytopenia induced by a germ-line mutation in the PRKACG gene. *Blood*. 2014;124(16):2554-2563. doi:10.1182/blood-2014-01-551820

240. Krall M, Htun S, Schnur RE, et al. Biallelic sequence variants in INTS1 in patients with developmental delays, cataracts, and craniofacial anomalies. *Eur J Hum Genet EJHG*. 2019;27(4):582-593. doi:10.1038/s41431-018-0298-9

241. Nuclear pore protein NUP210 depletion suppresses metastasis through heterochromatin-mediated disruption of tumor cell mechanical response - PubMed. Accessed October 16, 2023. https://pubmed.ncbi.nlm.nih.gov/34903738/

242. Snezhkina AV, Lukyanova EN, Zaretsky AR, et al. Novel potential causative genes in carotid paragangliomas. *BMC Med Genet*. 2019;20(Suppl 1):48. doi:10.1186/s12881-019-0770-6

243. Liu XX, Cai L, Liu FJ. An in silico analysis of human sperm genes associated with asthenozoospermia and its implication in male infertility. *Medicine (Baltimore)*. 2018;97(49):e13338. doi:10.1097/MD.0000000000013338

244. Shirahata E, Iwasaki H, Takagi M, et al. Ankyrin-G regulates inactivation gating of the neuronal sodium channel, Nav1.6. *J Neurophysiol*. 2006;96(3):1347-1357. doi:10.1152/jn.01264.2005

245. Schneider T, Neumaier F, Hescheler J, Alpdogan S. Cav2.3 R-type calcium channels: from its discovery to pathogenic de novo CACNA1E variants: a historical perspective. *Pflugers Arch*. 2020;472(7):811-816. doi:10.1007/s00424-020-02395-0

246. Tanos BE, Yang HJ, Soni R, et al. Centriole distal appendages promote membrane docking, leading to cilia initiation. *Genes Dev*. 2013;27(2):163-168. doi:10.1101/gad.207043.112

247. Wei Q, Xu Q, Zhang Y, et al. Transition fibre protein FBF1 is required for the ciliary entry of assembled intraflagellar transport complexes. *Nat Commun*. 2013;4:2750. doi:10.1038/ncomms3750

248. Grillet N, Schwander M, Hildebrand MS, et al. Mutations in LOXHD1, an evolutionarily conserved stereociliary protein, disrupt hair cell function in mice and cause progressive hearing loss in humans. *Am J Hum Genet*. 2009;85(3):328-337. doi:10.1016/j.ajhg.2009.07.017

249. Ivarsdottir EV, Holm H, Benonisdottir S, et al. The genetic architecture of age-related hearing impairment revealed by genome-wide association analysis. *Commun Biol*. 2021;4:706. doi:10.1038/s42003-021-02224-9

250. Inoue A, Akiyoshi J, Muronaga M, et al. Association of TMEM132D, COMT, and GABRA6 genotypes with cingulate, frontal cortex and hippocampal emotional processing in panic and major depressive disorder. *Int J Psychiatry Clin Pract*. 2015;19(3):192-200. doi:10.3109/13651501.2015.1043133

251. Erhardt A, Czibere L, Roeske D, et al. TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. *Mol Psychiatry*. 2011;16(6):647-663. doi:10.1038/mp.2010.41

252. Murray CR, Abel SN, McClure MB, et al. Novel Causative Variants in DYRK1A, KARS, and KAT6A Associated with Intellectual Disability and Additional Phenotypic Features. *J Pediatr Genet*. 2017;6(2):77-83. doi:10.1055/s-0037-1598639

253. Lachgar M, Morín M, Villamar M, Del Castillo I, Moreno-Pelayo MÁ. A Novel Truncating Mutation in HOMER2 Causes Nonsyndromic Progressive DFNA68 Hearing Loss in a Spanish Family. *Genes*. 2021;12(3):411. doi:10.3390/genes12030411

254. Azaiez H, Decker AR, Booth KT, et al. HOMER2, a stereociliary scaffolding protein, is essential for normal hearing in humans and mice. *PLoS Genet*. 2015;11(3):e1005137. doi:10.1371/journal.pgen.1005137

255. Yan W, Zhu H, Yu B, et al. Effects of two inhibitors of metabolic glutamate receptor 5 on expression of endogenous homer scaffold protein 1 in the auditory cortex of mice with tinnitus. *Bioengineered*. 2021;12(1):7156-7164. doi:10.1080/21655979.2021.1979354

256. Nam YS, Choi YM, Lee S, Cho HH. Valproic Acid Inhibits Progressive Hereditary Hearing Loss in a KCNQ4 Variant Model through HDAC1 Suppression. *Int J Mol Sci*. 2023;24(6):5695. doi:10.3390/ijms24065695

257. Chen FQ, Schacht J, Sha SH. Aminoglycoside-induced histone deacetylation and hair cell death in the mouse cochlea. *J Neurochem*. 2009;108(5):1226-1236. doi:10.1111/j.1471-4159.2009.05871.x

258. Wen LT, Wang J, Wang Y, Chen FQ. Association between histone deacetylases and the loss of cochlear hair cells: Role of the former in noise-induced hearing loss. *Int J Mol Med*. 2015;36(2):534-540. doi:10.3892/ijmm.2015.2236

259. Liang Y, Huang L, Yang J. Differential expression of ryanodine receptor in the developing rat cochlea. *Eur J Histochem EJH*. 2009;53(4):e30. doi:10.4081/ejh.2009.e30

260. Perin P, Botta L, Tritto S, Laforenza U. Expression and localization of ryanodine receptors in the frog semicircular canal. *J Biomed Biotechnol*. 2012;2012:398398. doi:10.1155/2012/398398

261. Sakai Y, Tsunekawa M, Ohta K, et al. The Integrin Signaling Network Promotes Axon Regeneration via the Src-Ephexin-RhoA GTPase Signaling Axis. *J Neurosci Off J Soc Neurosci*. 2021;41(22):4754-4767. doi:10.1523/JNEUROSCI.2456-20.2021

262. Katayama K ichi, Imai F, Suto F, Yoshida Y. Deletion of Sema3a or plexinA1/plexinA3 causes defects in sensory afferent projections of statoacoustic ganglion neurons. *PloS One*. 2013;8(8):e72512. doi:10.1371/journal.pone.0072512

263. Reinhard SM, Razak K, Ethell IM. A delicate balance: role of MMP-9 in brain development and pathophysiology of neurodevelopmental disorders. *Front Cell Neurosci*. 2015;9:280. doi:10.3389/fncel.2015.00280

264. Park SS, Lee DH, Lee SM, Lee CH, Kim SY. Noise exposure alters MMP9 and brevican expression in the rat primary auditory cortex. *BMC Neurosci*. 2020;21(1):16. doi:10.1186/s12868-020-00567-3

265. Pirbhoy PS, Rais M, Lovelace JW, et al. Acute pharmacological inhibition of matrix metalloproteinase-9 activity during development restores perineuronal net formation and normalizes auditory processing in Fmr1 KO mice. *J Neurochem*. 2020;155(5):538-558. doi:10.1111/jnc.15037

9 SUPPLEMENTARY MATERIAL

9.1 SUPPLEMENTARY TABLES

Table S1 - Tinnitus Handicap Inventory (THI) questionnaire.

TINNITUS HANDICAP INVENTORY				
The purpose of the scale is to identify the problems your tinnitus may be causing you.				
1F	Because of your tinnitus is it difficult for you to concentrate?	YES	SOMETIMES	NO
2F	Does the loudness of your tinnitus make it difficult for you to hear people?	YES	SOMETIMES	NO
3F	Does your tinnitus make you angry?	YES	SOMETIMES	NO
4F	Does your tinnitus make you feel confuse?	YES	SOMETIMES	NO
5C	Because of your tinnitus do you feel desperate?	YES	SOMETIMES	NO
6E	Do you complain a great deal about your tinnitus?	YES	SOMETIMES	NO
7F	Because of your tinnitus do you have trouble falling to sleep at night?	YES	SOMETIMES	NO
8C	Do you feel as though you cannot escape your tinnitus?	YES	SOMETIMES	NO
9F	Does your tinnitus interfere with your ability to enjoy social activities (such as going out to dinner, to the movies)?	YES	SOMETIMES	NO
10E	Because of your tinnitus do you feel frustrated?	YES	SOMETIMES	NO
11C	Because of your tinnitus do you feel that you have a terrible disease?	YES	SOMETIMES	NO
12F	Does your tinnitus make it difficult for you to enjoy life?	YES	SOMETIMES	NO
13F	Does your tinnitus interfere with your job or household responsibilities?	YES	SOMETIMES	NO
14F	Because of your tinnitus do you find that you are often irritable?	YES	SOMETIMES	NO
15F	Because of your tinnitus is it difficult for you to read?	YES	SOMETIMES	NO
16E	Does your tinnitus make you upset?	YES	SOMETIMES	NO
17E	Do you feel that your tinnitus problem has placed stress on your relationship with members of your family and friends?	YES	SOMETIMES	NO
18F	Do you find it difficult to focus your attention away from your tinnitus and on other things?	YES	SOMETIMES	NO
19C	Do you feel that you have no control over your tinnitus?	YES	SOMETIMES	NO
20F	Because of your tinnitus do you often feel tired?	YES	SOMETIMES	NO
21E	Because of your tinnitus do you feel depressed?	YES	SOMETIMES	NO
22E	Does your tinnitus make you feel anxious?	YES	SOMETIMES	NO
23C	Do you feel that you can no longer cope with your tinnitus?	YES	SOMETIMES	NO
24F	Does your tinnitus get worse when you are under stress?	YES	SOMETIMES	NO
25E	Does your tinnitus make you feel insecure?	YES	SOMETIMES	NO

Table S2 - Spanish version of the Tinnitus Handicap Inventory (THI) questionnaire.

	CUESTIONARIO ACÚFENOS THI ADAPTADO Contesta a las preguntas en función de su propia valoración.	
1F	¿Le resulta difícil concentrarse por culpa de su acúfeno?	SÍ A VECES NO
2F	Debido a la intensidad del acúfeno ¿Le cuesta oír a los demás?	SÍ A VECES NO
3F	¿Se enoja a causa del acúfeno?	SÍ A VECES NO
4F	¿Le produce confusión su acúfeno?	SÍ A VECES NO
5C	¿Se encuentra desesperado por tener el acufeno?	SÍ A VECES NO
6E	¿Se queja mucho por tener su acúfeno?	SÍ A VECES NO
7F	¿Tiene problemas para conciliar el sueño por su acúfeno?	SÍ A VECES NO
8C	¿Cree que su problema de acúfeno es insolucionable?	SÍ A VECES NO
9F	¿Interfiere su acúfeno en su vida social (salir a cenar, al cine)?	SÍ A VECES NO
10E	¿Se siente frustrado por su acúfeno?	SÍ A VECES NO
11C	¿Cree que tiene una enfermedad incurable?	SÍ A VECES NO
12F	¿Su acúfeno le impide disfrutar de la vida?	SÍ A VECES NO
13F	¿Infiere su acúfeno en su trabajo o tareas del hogar?	SÍ A VECES NO
14F	¿Se siente a menudo irritable por culpa de su acúfeno?	SÍ A VECES NO
15F	¿Tiene dificultades para leer por culpa de su acúfeno?	SÍ A VECES NO
16E	¿Se encuentra usted triste debido a su acúfeno?	SÍ A VECES NO
17E	¿Cree que su acúfeno le crea tensiones o interfiere en su relación con la familia o amigos?	SÍ A VECES NO
18F	¿Es difícil, para usted, fijar su atención en cosas distintas a su acúfeno?	SÍ A VECES NO
19C	¿Cree que su acúfeno es incurable?	SÍ A VECES NO
20F	¿Se siente a menudo cansado por culpa de su acúfeno?	SÍ A VECES NO
21E	¿Se siente deprimido por culpa de su acúfeno?	SÍ A VECES NO
22E	¿Se siente ansioso por culpa de su acúfeno?	SÍ A VECES NO
23C	¿Cree que su problema de acúfenos le desborda?	SÍ A VECES NO
24F	¿Empeora su acúfeno cuando tiene estrés?	SÍ A VECES NO
25E	¿Se siente usted inseguro por culpa de su acúfeno?	SÍ A VECES NO

Table S3 - English version of the questionnaire of hypersensitivity to sound (GÜF test).

	Mark the corresponding box with an X	**Never**	**Sometimes**	**Often**	**Always**
1	Sounds that didn't disturb me earlier frighten me now.				
2	I worry that I will never succeed in getting used to loud/uncomfortable sounds.				
3	I cannot listen for a long time when I am surrounded by loud/uncomfortable sounds.				
4	Because of my hypersensitivity to sound, there is tension between my partner and/or my family and myself.				
5	I have to avoid certain sounds.				
6	I am very scared of noise.				
7	I think the hypersensitivity to sound has ruined my life.				
8	When surrounded by a lot of sounds, I don't understand anything.				
9	Other people distance themselves from me because I can't stand loud/uncomfortable sounds.				
10	I am annoyed by sound that are too loud/uncomfortable for me.				
11	Loud/uncomfortable sounds cause physical pain in my ears.				
12	I believe I won't be able to cope in everyday life if hypersensitivity to sound continues to be this bad.				
13	I immediately withdraw when there are loud/uncomfortable sounds.				
14	I am afraid that loud/uncomfortable sounds damage my hearing.				
15	Since becoming hypersensitivity to sound, I no longer enjoy music.				

Table S4 - Spanish version of the questionnaire of hypersensitivity to sound (GÜF test).

	Marque con una X la casilla correspondiente	**Nunca o no es cierto**	**En ocasiones**	**Frecuen-temente**	**Siempre**
1	Ciertos ruidos que antes no me molestaban ahora me provocan miedo.				
2	Me preocupa la idea de que nunca voy a ser capaz de acostumbrarme a estos sonidos fuertes o desagradables.				
3	Cuanto tengo alrededor ruidos fuertes o desagradables no puedo escuchar o prestar atención.				
4	Tengo problemas con mi pareja o mi familia por mi mayor sensibilidad a los sonidos.				
5	Ante la presencia de ciertos sonidos, tengo la necesidad de manifestarlo o decírselo a los demás.				
6	Tengo mucho miedo a los ruidos intensos.				
7	Pienso que la hipersensibilidad a los sonidos que tengo me ha arruinado mi vida.				
8	Cuando tengo muchos ruidos alrededor, no oigo ni entiendo nada.				
9	Algunas personas me evitan porque no soporto los ruidos fuertes o desagradables.				
10	Los sonidos fuertes o desagradables me provocan enfado.				
11	Tengo dolor de oídos cuando hay ruidos intensos o desagradables.				
12	Pienso que voy a ser incapaz de enfrentarme a la vida diaria si persiste mi hipersensibilidad a los ruidos.				
13	Cuando hay ruidos intensos o desagradables, me retiro o me retraigo inmediatamente.				
14	Tengo miedo porque los ruidos fuertes o desagradables deterioren mi audición.				
15	Desde que tengo esta hipersensibilidad a los sonidos ya no disfruto de la música.				

Table S5 - Patient Health Questionnaire depression scale (PHQ-9).

	Over the last 2 weeks, how often have you been bothered by any of the following problems?	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1	Little interest or pleasure in doing things.				
2	Feeling down, depressed, or hopeless.				
3	Trouble falling or staying asleep, or sleeping too much.				
4	Feeling tired or having little energy.				
5	Poor appetite or overeating.				
6	Feeling bad about yourself - or that you are a failure or have let yourself or your family down.				
7	Trouble concentrating on things, such as reading the newspaper or watching television.				
8	Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual.				
9	Thoughts that you would be better off dead or of hurting yourself in some way.				

Table S6 - Spanish version of the Patient Health Questionnaire depression scale (PHQ-9).

	Durante las *últimas 2 semanas*, ¿qué tan seguido ha tenido molestias debido a los siguientes problemas?	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1	Poco interés o placer en hacer cosas.				
2	Se ha sentido decaído(a), deprimido(a) o sin esperanzas.				
3	Ha tenido dificultad para quedarse o permanecer dormido(a), o ha dormido demasiado.				
4	Se ha sentido cansado(a) o con poca energía.				
5	Sin apetito o ha comido en exceso.				
6	Se ha sentido mal con usted mismo(a) – o que es un fracaso o que ha quedado mal con usted mismo(a) o con su familia.				
7	Ha tenido dificultad para concentrarse en ciertas actividades, tales como leer el periódico o ver la televisión.				
8	¿Se ha movido o hablado tan lento que otras personas podrían haberlo notado? o lo contrario – muy inquieto(a) o agitado(a) que ha estado moviéndose mucho más de lo normal.				
9	Pensamientos de que estaría mejor muerto(a) o de lastimarse de alguna manera.				

Table S7 - The Hospital Anxiety and Depression Scale (HADS) questionnaire.

Tick the box beside the reply that is closest to how you have been feeling in the past week. Don't take too long over you replies: your immediate is best.			
A.1. I feel tenso or 'wound up':			
3) Most of the time ☐	2) A lot of the time ☐	1) From time to time, occasionally ☐	0) Not at all ☐
D.1. I still enjoy the things I used to enjoy:			
0) Definitely as much ☐	1) No quite so much ☐	2) Only a little ☐	3) Hardly at all ☐
A.2. I get a sort of frightened feeling as if something awful is about to happen:			
3) Very definitely and quite badly ☐	2) Yes, but not too badly ☐	1) A little, but it doesn't worry me ☐	0) Not at all ☐
D.2. I can laugh and see the funny side of things:			
0) As much as I always could ☐	1) Not quite so much now ☐	2) Definitely not so much now ☐	3) Not at all ☐
A.3. Worrying thoughts go through my mind:			
3) A great deal of the time ☐	2) A lot of the time ☐	1) From time to time, but not too often ☐	0) Only occasionally ☐
D.3. I feel cheerful:			
3) Not at all ☐	2) Not often ☐	1) Sometimes ☐	0) Most of the time ☐
A.4. I can sit at ease and feel relaxed:			
0) Definitely ☐	1) Usually ☐	2) Not often ☐	3) Not at all ☐
D.4. I feel as if I am slowed down:			
3) Nearly all the time ☐	2) Very often ☐	1) Sometimes ☐	0) Not at all ☐
A.5. I get a sort of frightened feeling like 'butterflies' in the stomach:			
0) Not at all ☐	1) Occasionally ☐	2) Quite often ☐	3) Very often ☐
D.5. I have lost interest in my appearance:			
3) Definitely ☐	2) I don't take as much care as I should ☐	1) I may not take quite as much care ☐	0) I take just as much care as ever☐
A.6. I feel restless as I have to be on the move:			
3) Very much indeed ☐	2) Quite a lot ☐	1) Not very much ☐	0) Not at all ☐
D.6. I look forward with enjoyment to things:			
0) As much as I ever did ☐	1) Rather less than I used to ☐	2) Definitely less than I used to☐	3) Hardly at all ☐
A.7. I get sudden feelings of panic:			
3) Very often indeed ☐	2) Quite often ☐	1) Not very often ☐	0) Not at all ☐
D.7. I can enjoy a good book or radio or TV program:			
0) Often ☐	1) Sometimes ☐	2) Not often ☐	3) Very seldom ☐

Table S8 - Spanish version of the The Hospital Anxiety and Depression Scale (HADS) questionnaire.

Lea cada pregunta y marque la que usted considere que coincide con su propio estado emocional en la última semana. **No es necesario que piense mucho tiempo cada respuesta, en este cuestionario las respuestas espontáneas tienen más valor que las que se piensan mucho.**			
A.1. Me siento tenso/a o nervioso/a:			
3) Casi todo el día ☐	2) Gran parte del día ☐	1) De vez en cuando ☐	0) Nunca ☐
D.1. Sigo disfrutando de las cosas como siempre:			
0) Ciertamente igual que antes ☐	1) No tanto como antes ☐	2) Solamente un poco ☐	3) Ya no disfruto con nada ☐
A.2. Siento una especie de temor como si algo malo fuera a suceder:			
3) Sí, y muy intenso ☐	2) Sí, pero no muy intenso ☐	1) Sí, pero no me preocupa ☐	0) No siento nada de eso ☐
D.2. Soy capaz de reírme y ver el lado gracioso de las cosas:			
0) Igual que siempre ☐	1) Actualmente algo menos ☐	2) Actualmente mucho menos ☐	3) Actualmente en absoluto ☐
A.3. Tengo la cabeza llena de preocupaciones:			
3) Casi todo el día ☐	2) Gran parte del día ☐	1) De vez en cuando ☐	0) Nunca ☐
D.3. Me siento alegre:			
3) Nunca ☐	2) Muy pocas veces ☐	1) En algunas ocasiones ☐	0) Gran parte del día ☐
A.4. Soy capaz de permanecer sentado/a, tranquilo/a y relajado/a:			
0) Siempre ☐	1) A menudo ☐	2) A veces ☐	3) Nunca ☐
D.4. Me siento lento/a y torpe:			
3) Gran parte del día ☐	2) A menudo ☐	1) A veces ☐	0) Nunca ☐
A.5. Experimento una desagradable sensación de "nervios y hormigueos" en el estómago:			
0) Nunca ☐	1) Sólo en algunas ocasiones ☐	2) A menudo ☐	3) Muy a menudo ☐
D.5. He perdido el interés por mi aspecto personal:			
3) Completamente ☐	2) No me cuido como debería hacerlo ☐	1) Es posible que no me cuide como debiera ☐	0) Me cuido como siempre lo he hecho ☐
A.6. Me siento inquieto/a como si no pudiera parar de moverme:			
3) Realmente mucho ☐	2) Bastante ☐	1) No mucho ☐	0) En absoluto ☐
D.6. Espero las cosas con ilusión:			
0) Como siempre ☐	1) Algo menos que antes ☐	2) Mucho menos que antes ☐	3) En absoluto ☐
A.7. Experimento de repente sensaciones de gran angustia o temor:			
3) Muy a menudo ☐	2) Con cierta frecuencia ☐	1) Raramente ☐	0) Nunca ☐
D.7. Soy capaz de disfrutar con un buen libro o con un buen programa de televisión:			
0) A menudo ☐	1) Algunas veces ☐	2) Pocas veces ☐	3) Casi nunca ☐

Table S9 - Consequences terms and their corresponding impacts annotated by VEP.

Consequence term	Impact
Transcript ablation	HIGH
Splice acceptor variant	HIGH
Splice donor variant	HIGH
Stop gained	HIGH
Frameshift variant	HIGH
Stop lost	HIGH
Start lost	HIGH
Transcript amplification	HIGH
Inframe insertion	MODERATE
Inframe deletion	MODERATE
Missense variant	MODERATE
Protein altering variant	MODERATE
Splice region variant	LOW
Splice donor 5th base variant	LOW
Splice donor region variant	LOW
Splice polypyrimidine tract variant	LOW
Incomplete terminal codon variant	LOW
Start retained variant	LOW
Stop retained variant	LOW
Synonymous variant	LOW
Coding sequence variant	MODIFIER
Mature miRNA variant	MODIFIER
5 prime UTR variant	MODIFIER
3 prime UTR variant	MODIFIER
Non coding transcript exon variant	MODIFIER
Intron variant	MODIFIER
NMD transcript variant	MODIFIER
Non coding transcript variant	MODIFIER
Upstream gene variant	MODIFIER
Downstream gene variant	MODIFIER
TFBS ablation	MODIFIER
TFBS amplification	MODIFIER
TF binding site variant	MODIFIER

Table S9 - Continuation.

Consequence term	Impact
Regulatory region ablation	MODIFIER
Regulatory region amplification	MODIFIER
Feature elongation	MODIFIER
Regulatory region variant	MODIFIER
Feature truncation	MODIFIER
Intergenic variant	MODIFIER

Consequence terms relative to the transcript structure and their corresponding impacts. Table modified obtained from www.ensembl.org [117].

Table S10 - Criteria for diagnosis of Meniere Disease (MD). Described by the Committee on Hearing and Equilibrium of the American Academy of Otolaryngology–Head and Neck Surgery (AAO-HNS) published in 1995.

Certain MD
A. Definite MD, plus histopathologic confirmation.
Definite MD
A. Two or more definitive spontaneous episodes of vertigo 20 minutes or longer.
B. Audiometrically documented hearing loss on at least one occasion.
C. Tinnitus or aural fullness in the treated ear.
D. Other causes excluded.
Probable MD
A. One definitive episode of vertigo.
B. Audiometrically documented hearing loss on at least one occasion.
C. Tinnitus or aural fullness in the treated ear.
D. Other causes excluded.
Possible MD
A. Episodic vertigo of the MD type without documented hearing loss, or
B. SNHL, fluctuating or fixed, with disequilibrium but without definitive episodes.
C. Other causes excluded.

Table S11 - Summary of the genes enriched in variants with a predicted high confidence of being loss-of-function (HIGH HC), filtered by allelic frequency < 0.05, for the whole Meniere Disease cohort.

Gene symbol	Number variants	Number individuals	gnomAD NFE					gnomAD global					CSVS				
			OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF
SYNGAP1	1	61	12.07	9.15	15.92	0	0.92	12.09	9.23	15.84	0	0.92	-	-	-	-	-
KIR3DL1	2	54	5.96	4.61	7.71	0	0.83	8.56	6.64	11.04	0	0.88	-	-	-	-	-
SKA3	9	49	4.63	3.63	5.90	0	0.78	4.96	3.90	6.31	0	0.80	-	-	-	-	-
BTNL8	1	28	11.24	8.40	15.04	0	0.91	13.88	10.46	18.43	0	0.93	-	-	-	-	-
STK33	2	27	2.35	1.78	3.09	2.30E-06	0.57	3.18	2.42	4.18	2.81E-13	0.69	-	-	-	-	-
FMN2	4	22	2.12	1.57	2.85	1.25E-03	0.53	2.99	2.23	4.02	6.49E-10	0.67	-	-	-	-	-
PTTG2	1	20	2.99	1.90	4.69	3.53E-03	0.67	2.77	1.77	4.33	1.67E-02	0.64	-	-	-	-	-
ADAM2	4	15	3.22	1.96	5.31	7.48E-03	0.69	4.56	2.78	7.49	4.03E-06	0.78	5.26	2.53	10.96	6.38E-03	0.81

gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; OR: Odds ratio; CI: Confidence interval; FDR: p-value corrected by false discovery rate; EF: Etiological fraction.

Table S12 - Summary of the genes enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20), filtered by allelic frequency < 0.05, for the whole Meniere Disease cohort.

Gene symbol	Number variants	Number individuals	gnomAD NFE					gnomAD global					CSVS				
			OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF
ADGRL2	15	67	1.74	1.38	2.19	3.38E-02	0.42	2.24	1.78	2.82	8.10E-08	0.55	1.85	1.42	2.39	3.90E-02	0.46
PRKRA	2	54	23.43	17.33	31.67	0	0.96	23.09	17.35	30.73	0	0.96	3.06	2.20	4.24	2.57E-07	0.67
CACNA1H	23	51	4.14	3.15	5.43	0	0.76	3.79	2.90	4.96	0	0.74	2.49	1.82	3.40	1.37E-04	0.60
PIEZO1	23	44	2.73	2.05	3.64	8.58E-08	0.63	3.94	2.96	5.25	1.00E-16	0.75	2.20	1.58	3.06	3.03E-02	0.55
DNHD1	23	33	3.06	2.17	4.30	1.70E-06	0.67	2.61	1.86	3.66	3.65E-04	0.62	2.66	1.78	3.97	1.98E-02	0.62
C15orf39	11	31	4.92	3.43	7.06	6.01E-14	0.80	5.36	3.75	7.66	4.00E-16	0.81	2.83	1.86	4.33	1.55E-02	0.65
PCNX1	14	24	2.77	1.85	4.16	1.03E-02	0.64	4.44	2.96	6.65	6.51E-09	0.77	3.47	2.11	5.71	1.05E-02	0.71
GXYLT1	2	19	7.88	4.92	12.64	1.19E-13	0.87	6.73	4.25	10.67	6.04E-12	0.85	###	10.31	190.62	3.76E-03	0.98
TNK2	9	18	4.98	3.10	8.00	3.67E-07	0.80	5.87	3.67	9.38	1.67E-09	0.83	4.13	2.28	7.51	3.46E-02	0.76
FRYL	17	17	20.85	12.74	34.13	0	0.95	10.19	6.43	16.13	0	0.90	8.66	4.34	17.27	9.84E-06	0.88
FLT4	12	17	3.93	2.42	6.38	3.83E-04	0.75	6.45	3.98	10.44	4.73E-10	0.84	5.57	2.90	10.73	2.93E-03	0.82
AMZ1	7	16	6.63	4.06	10.83	5.34E-10	0.85	9.41	5.79	15.30	1.90E-15	0.89	5.35	2.80	10.23	4.09E-03	0.81
SCN10A	12	15	3.65	2.22	6.01	4.39E-03	0.73	5.65	3.44	9.29	1.00E-07	0.82	5.23	2.71	10.09	9.05E-03	0.81
CNTNAP2	11	15	3.78	2.26	6.33	5.30E-03	0.74	3.85	2.31	6.41	2.98E-03	0.74	6.46	3.15	13.21	3.61E-03	0.85
STOX2	10	15	4.73	2.86	7.81	1.53E-05	0.79	5.59	3.40	9.18	1.39E-07	0.82	4.80	2.50	9.20	2.52E-02	0.79

gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; OR: Odds ratio; CI: Confidence interval; FDR: p-value corrected by false discovery rate; EF: Etiological fraction.

Table S13 - Summary of the genes enriched in variants with a predicted high confidence of being loss-of-function (HIGH HC), filtered by allelic frequency < 0.05, for the I4.

Gene symbol	Number variants	Number individuals	gnomAD NFE					gnomAD global					CSVS				
			OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF
PTTG2	1	8	5.05	2.47	10.33	6.45E-03	0.80	4.67	2.29	9.54	1.76E-02	0.79	-	-	-	-	-
FMN2	2	7	3.45	2.01	5.91	4.72E-03	0.71	4.95	2.89	8.48	4.50E-06	0.80	-	-	-	-	-
ENDOG	1	5	18.52	7.46	45.97	2.17E-07	0.95	16.11	6.56	39.56	1.03E-06	0.94	-	-	-	-	-
EFCAB5	3	4	115.77	38.27	350.18	2.73E-14	0.99	7.77	2.89	20.88	3.66E-02	0.87	27.37	6.82	109.80	8.33E-04	0.96
RECQL5	3	4	9.46	3.43	26.09	9.74E-03	0.89	9.73	3.59	26.40	6.08E-03	0.90	-	-	-	-	-
MSH6	2	4	17.08	6.25	46.68	2.18E-05	0.94	19.81	7.31	53.65	3.29E-06	0.95	-	-	-	-	-
LPIN3	1	4	117.86	38.65	359.37	3.48E-14	0.99	100.57	35.49	285.04	3.20E-15	0.99	-	-	-	-	-
RAD50	3	3	433.28	87.21	2152.56	7.85E-11	1.00	961.00	193.44	4774.26	3.50E-14	1.00	-	-	-	-	-
IQCN	3	3	43.30	13.17	142.41	3.78E-07	0.98	61.32	19.01	197.72	4.27E-09	0.98	-	-	-	-	-
FBXO27	2	3	260.89	62.07	1096.66	2.12E-11	1.00	241.10	67.69	858.77	1.99E-14	1.00	20.00	4.46	89.76	2.54E-02	0.95

gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; OR: Odds ratio; CI: Confidence interval; FDR: p-value corrected by false discovery rate; EF: Etiological fraction.

Table S14 - Summary of the genes enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20), filtered by allelic frequency (AF) < 0.05, for the I4.

Gene symbol	Number variants	Number individuals	gnomAD NFE					gnomAD global					CSVS				
			OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF
THADA	6	13	4.09	0.00	2.36	7.10E+00	0.76	6.06	0.00	3.50	1.05E+01	0.84	4.58	0.00	2.54	8.28E+00	0.78
UNC13C	7	12	6.02	0.00	3.39	1.07E+01	0.83	7.96	0.00	4.50	1.41E+01	0.87	5.61	0.00	3.00	1.05E+01	0.82
DSCAML1	7	9	10.11	0.00	5.21	1.96E+01	0.90	15.50	0.00	8.00	3.00E+01	0.94	7.47	0.00	3.55	1.57E+01	0.87
ITGB5	4	9	6.29	0.00	3.24	1.22E+01	0.84	8.19	0.00	4.23	1.59E+01	0.88	6.01	0.01	2.90	1.24E+01	0.83
WDFY4	8	7	19.36	0.00	9.51	3.94E+01	0.95	23.23	0.00	11.50	4.70E+01	0.96	10.73	0.00	4.72	2.44E+01	0.91
ANKRD44	5	7	7.61	0.00	3.59	1.61E+01	0.87	10.97	0.00	5.19	2.32E+01	0.91	9.00	0.00	3.79	2.13E+01	0.89
MROH1	6	6	10.84	0.00	5.10	2.30E+01	0.91	13.00	0.00	6.15	2.75E+01	0.92	-	-	-	-	-
ARVCF	6	6	8.97	0.00	4.23	1.90E+01	0.89	13.50	0.00	6.38	2.85E+01	0.93	9.89	0.00	4.15	2.36E+01	0.90
SPTBN4	6	6	30.22	0.00	13.18	6.93E+01	0.97	34.53	0.00	15.25	7.82E+01	0.97	17.18	0.00	6.10	4.84E+01	0.94
EML6	6	6	16.55	0.00	7.30	3.75E+01	0.94	26.94	0.00	11.94	6.08E+01	0.96	31.19	0.00	9.50	1.02E+02	0.97
CNTNAP2	6	6	7.48	0.01	3.33	1.68E+01	0.87	7.23	0.01	3.23	1.62E+01	0.86	20.42	0.00	7.07	5.90E+01	0.95
GRIK4	5	6	867.87	0.00	216.64	3.48E+03	1.00	722.16	0.00	249.96	2.09E+03	1.00	81.64	0.00	16.45	4.05E+02	0.99
ERCC6	5	6	7.56	0.01	3.36	1.70E+01	0.87	9.74	0.00	4.35	2.18E+01	0.90	10.72	0.01	4.15	2.77E+01	0.91
PDZD7	4	6	11.79	0.00	5.22	2.66E+01	0.92	19.42	0.00	8.62	4.38E+01	0.95	9.54	0.02	3.72	2.45E+01	0.90
ZNF106	6	5	89.64	0.00	37.13	2.16E+02	0.99	64.80	0.00	28.28	1.49E+02	0.98	39.86	0.00	11.23	1.42E+02	0.97
SASH1	5	5	361.03	0.00	109.94	1.19E+03	1.00	129.86	0.00	50.90	3.31E+02	0.99	66.17	0.00	12.82	3.42E+02	0.98
MYBPC3	5	5	103.16	0.00	38.80	2.74E+02	0.99	64.95	0.00	26.18	1.61E+02	0.98	19.40	0.00	6.14	6.13E+01	0.95
TDRD6	5	5	98.48	0.00	37.19	2.61E+02	0.99	9.99	0.00	4.13	2.42E+01	0.90	14.77	0.01	4.94	4.42E+01	0.93
KAT6A	5	5	27.42	0.00	11.07	6.79E+01	0.96	21.55	0.00	8.86	5.24E+01	0.95	14.54	0.01	4.86	4.35E+01	0.93
SETD1A	5	5	16.15	0.00	6.59	3.95E+01	0.94	18.91	0.00	7.78	4.59E+01	0.95	26.74	0.00	7.72	9.26E+01	0.96
PPL	5	5	10.02	0.00	4.12	2.44E+01	0.90	10.01	0.00	4.13	2.42E+01	0.90	44.06	0.00	10.51	1.85E+02	0.98
VWA5B1	4	5	46.14	0.00	18.29	1.16E+02	0.98	42.59	0.00	17.33	1.05E+02	0.98	18.80	0.00	5.95	5.94E+01	0.95

Table S14 – Continuation.

Gene symbol	*Number variants*	*Number individuals*	*gnomAD NFE*					*gnomAD global*					*CSVS*				
			OR	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*
STIM1	4	5	13.47	0.00	5.51	3.29E+01	0.93	18.81	0.00	7.73	4.57E+01	0.95	27.34	0.00	7.89	9.47E+01	0.96
CDH3	4	5	8.44	0.02	3.47	2.05E+01	0.88	13.95	0.00	5.75	3.39E+01	0.93	11.39	0.04	4.00	3.24E+01	0.91
TAP1	3	5	7.94	0.00	3.52	1.79E+01	0.87	10.09	0.00	4.49	2.27E+01	0.90	10.75	0.01	4.11	2.81E+01	0.91
FURIN	3	5	83.69	0.00	31.99	2.19E+02	0.99	80.49	0.00	32.17	2.01E+02	0.99	13.72	0.01	4.67	4.03E+01	0.93
PPFIBP1	4	4	19.20	0.00	7.81	4.72E+01	0.95	13.08	0.00	5.39	3.17E+01	0.92	26.58	0.00	7.67	9.21E+01	0.96
BIRC6	4	4	433.16	0.00	108.08	1.74E+03	1.00	426.89	0.00	131.10	1.39E+03	1.00	-	-	-	-	-
PDZRN3	4	4	115.55	0.00	38.24	3.49E+02	0.99	174.83	0.00	60.06	5.09E+02	0.99	21.13	0.04	5.66	7.89E+01	0.95
STAT6	4	4	577.62	0.00	129.00	2.59E+03	1.00	349.45	0.00	110.96	1.10E+03	1.00	36.07	0.02	8.05	1.62E+02	0.97
RYR2	4	4	866.60	0.00	158.42	4.74E+03	1.00	240.30	0.00	80.10	7.21E+02	1.00	26.84	0.02	6.70	1.08E+02	0.96
FKBP15	4	4	46.84	0.00	16.64	1.32E+02	0.98	63.02	0.00	22.84	1.74E+02	0.98	26.08	0.03	6.51	1.05E+02	0.96
BRINP2	4	4	37.68	0.00	13.52	1.05E+02	0.97	60.10	0.00	21.82	1.66E+02	0.98	35.23	0.02	7.87	1.58E+02	0.97
STT3B	4	4	33.31	0.00	12.01	9.24E+01	0.97	68.66	0.00	24.82	1.90E+02	0.99	21.27	0.04	5.70	7.94E+01	0.95
SPTBN1	4	4	15.75	0.00	5.79	4.29E+01	0.94	24.03	0.00	8.88	6.50E+01	0.96	26.13	0.03	6.52	1.05E+02	0.96
EPHA3	4	4	14.55	0.00	5.35	3.95E+01	0.93	24.29	0.00	8.97	6.57E+01	0.96	21.77	0.03	5.83	8.13E+01	0.95
ARID5A	4	4	13.43	0.00	4.95	3.64E+01	0.93	22.62	0.00	8.37	6.12E+01	0.96	53.17	0.03	9.72	2.91E+02	0.98
SALL3	3	4	20.15	0.00	7.36	5.52E+01	0.95	23.43	0.00	8.65	6.35E+01	0.96	26.33	0.03	6.56	1.06E+02	0.96
TTC22	2	4	436.25	0.00	108.59	1.75E+03	1.00	322.78	0.00	103.51	1.01E+03	1.00	21.37	0.04	5.71	8.00E+01	0.95
TRIM54	2	4	79.31	0.00	27.16	2.32E+02	0.99	121.01	0.00	42.53	3.44E+02	0.99	21.34	0.04	5.70	7.99E+01	0.95
TRIB1	2	4	41.52	0.00	14.79	1.17E+02	0.98	58.65	0.00	21.24	1.62E+02	0.98	34.96	0.02	7.79	1.57E+02	0.97
PKN3	4	3	173.32	0.00	54.21	5.54E+02	0.99	213.69	0.00	72.10	6.33E+02	1.00	35.52	0.02	7.93	1.59E+02	0.97
ADAMTSL2	4	3	577.77	0.00	129.03	2.59E+03	1.00	349.73	0.00	111.05	1.10E+03	1.00	-	-	-	-	-
HOMER2	3	3	259.93	0.00	61.93	1.09E+03	1.00	144.21	0.00	42.70	4.87E+02	0.99	-	-	-	-	-
CASP14	3	3	216.67	0.00	54.02	8.69E+02	1.00	411.94	0.00	106.18	1.60E+03	1.00	-	-	-	-	-
EIF4G2	3	3	216.66	0.00	54.02	8.69E+02	1.00	206.02	0.00	59.00	7.19E+02	1.00	-	-	-	-	-
KCNH8	3	3	216.66	0.00	54.02	8.69E+02	1.00	412.04	0.00	106.21	1.60E+03	1.00	-	-	-	-	-

Table S14 – Continuation.

Gene symbol	*Number variants*	*Number individuals*	*gnomAD NFE*					*gnomAD global*					*CSVS*				
			OR	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*
CHD1	3	3	162.44	0.00	42.95	6.14E+02	0.99	288.38	0.00	79.10	1.05E+03	1.00	-	-	-	-	-
GATA6	3	3	162.31	0.00	42.92	6.14E+02	0.99	261.89	0.00	72.82	9.42E+02	1.00	-	-	-	-	-
SRMS	3	3	108.31	0.00	30.46	3.85E+02	0.99	169.70	0.00	49.56	5.81E+02	0.99	-	-	-	-	-
ADCY5	3	3	650.00	0.00	108.35	3.90E+03	1.00	103.02	0.00	31.21	3.40E+02	0.99	-	-	-	-	-
FAM107A	3	3	649.90	0.00	108.33	3.90E+03	1.00	721.15	0.00	160.94	3.23E+03	1.00	-	-	-	-	-
ILDR2	3	3	649.79	0.00	108.31	3.90E+03	1.00	720.66	0.00	160.83	3.23E+03	1.00	-	-	-	-	-
SOWAHB	3	3	649.52	0.00	108.27	3.90E+03	1.00	1441.59	0.00	240.30	8.65E+03	1.00	-	-	-	-	-
SLC9A3	3	3	649.34	0.00	108.24	3.90E+03	1.00	480.02	0.00	119.68	1.93E+03	1.00	-	-	-	-	-
HDAC4	3	3	1299.75	0.00	134.94	1.25E+04	1.00	2883.16	0.00	299.32	2.78E+04	1.00	-	-	-	-	-
TMPRSS13	3	3	39.35	0.00	12.02	1.29E+02	0.97	58.76	0.00	18.25	1.89E+02	0.98	-	-	-	-	-
RNF175	2	3	96.92	0.00	32.61	2.88E+02	0.99	101.84	0.00	36.12	2.87E+02	0.99	53.53	0.03	9.77	2.93E+02	0.98
CCDC14	2	3	100.29	0.00	28.43	3.54E+02	0.99	65.75	0.00	20.30	2.13E+02	0.98	-	-	-	-	-
C12orf56	2	3	59.28	0.00	17.65	1.99E+02	0.98	111.32	0.00	33.51	3.70E+02	0.99	-	-	-	-	-
ZNF554	1	3	109.80	0.00	30.67	3.93E+02	0.99	55.13	0.00	17.04	1.78E+02	0.98	-	-	-	-	-

gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; OR: Odds ratio; CI: Confidence interval; FDR: p-value corrected by false discovery rate; EF: Etiological fraction.

Table S15 - Summary of the genes enriched in variants with a predicted high confidence of being loss-of-function (HIGH HC), filtered by allelic frequency < 0.05, for the I1.

Gene symbol	Number variants	Number individuals	gnomAD NFE					gnomAD global					CSVS				
			OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF
ADAM2	4	6	4.98	0.02	2.35	1.05E+01	0.80	7.05	0.00	3.34	1.49E+01	0.86	7.93	0.00	3.15	1.99E+01	0.87
TOMM20L	2	5	21.43	0.00	11.23	4.09E+01	0.95	24.06	0.00	12.72	4.55E+01	0.96	-	-	-	-	-
PRKACG	2	4	134.86	0.00	42.74	4.26E+02	0.99	78.37	0.00	27.95	2.20E+02	0.99	91.07	0.02	10.15	8.17E+02	0.99
ENTPD8	1	4	23.82	0.00	8.57	6.62E+01	0.96	32.99	0.00	12.01	9.06E+01	0.97	-	-	-	-	-
RBM5	1	3	12.38	0.02	3.88	3.95E+01	0.92	14.44	0.00	4.57	4.57E+01	0.93	-	-	-	-	-

gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; OR: Odds ratio; CI: Confidence interval; FDR: p-value corrected by false discovery rate; EF: Etiological fraction.

Table S16 - Summary of the genes enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20), filtered by allelic frequency < 0.05, for the I1.

Gene symbol	Number variants	Number individuals	gnomAD NFE					gnomAD global					CSVS				
			OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF	OR	low CI	high CI	FDR	EF
CFTR	13	18	3.21	0.00	2.23	4.60E+00	0.69	4.17	0.00	2.91	5.98E+00	0.76	2.88	0.00	1.96	4.21E+00	0.65
COL11A1	7	13	3.47	0.03	2.05	5.90E+00	0.71	4.80	0.00	2.83	8.14E+00	0.79	5.35	0.00	2.98	9.59E+00	0.81
MYO1C	7	12	4.73	0.00	2.67	8.38E+00	0.79	7.11	0.00	4.01	1.26E+01	0.86	6.64	0.00	3.49	1.26E+01	0.85
CCDC88C	7	9	5.81	0.00	3.10	1.09E+01	0.83	9.75	0.00	5.21	1.82E+01	0.90	7.95	0.00	3.87	1.64E+01	0.87
ZFHX2	7	8	28.67	0.00	13.94	5.90E+01	0.97	6.32	0.00	3.14	1.27E+01	0.84	20.29	0.00	7.60	5.41E+01	0.95
RAB3GAP2	6	8	246.40	0.00	100.5	6.04E+02	1.00	123.78	0.00	58.71	2.61E+02	0.99	12.41	0.00	5.25	2.93E+01	0.92
CHD5	5	8	14.88	0.00	7.61	2.91E+01	0.93	23.11	0.00	11.88	4.50E+01	0.96	5.39	0.05	2.59	1.12E+01	0.81
COL17A1	4	8	8.72	0.00	4.31	1.76E+01	0.89	10.31	0.00	5.11	2.08E+01	0.90	9.81	0.00	4.28	2.25E+01	0.90
INTS1	7	7	64.54	0.00	28.85	1.44E+02	0.98	12.42	0.00	5.88	2.63E+01	0.92	25.14	0.00	8.44	7.49E+01	0.96
NUP210	7	7	8.26	0.00	3.90	1.75E+01	0.88	13.93	0.00	6.59	2.95E+01	0.93	13.12	0.00	5.15	3.34E+01	0.92
FLT4	6	7	73.84	0.00	32.73	1.67E+02	0.99	66.70	0.00	30.80	1.44E+02	0.99	22.16	0.00	7.76	6.33E+01	0.95
AKAP9	7	6	129.09	0.00	54.49	3.06E+02	0.99	93.90	0.00	42.86	2.06E+02	0.99	19.91	0.00	7.21	5.50E+01	0.95
ITPR3	6	6	29.12	0.00	12.66	6.70E+01	0.97	31.48	0.00	13.90	7.13E+01	0.97	13.95	0.00	5.06	3.85E+01	0.93
SCN10A	5	6	19.90	0.00	9.28	4.27E+01	0.95	27.74	0.00	13.02	5.91E+01	0.96	10.73	0.00	4.37	2.64E+01	0.91
DOCK10	5	6	46.16	0.00	19.70	1.08E+02	0.98	46.38	0.00	20.32	1.06E+02	0.98	18.86	0.00	6.32	5.62E+01	0.95
GLG1	5	6	21.72	0.00	9.51	4.96E+01	0.95	30.88	0.00	13.63	7.00E+01	0.97	16.97	0.00	5.87	4.90E+01	0.94
ATXN7	5	6	13.18	0.00	5.82	2.98E+01	0.92	18.68	0.00	8.29	4.21E+01	0.95	12.39	0.01	4.57	3.36E+01	0.92
RAPGEF4	5	6	11.19	0.00	4.95	2.53E+01	0.91	17.56	0.00	7.80	3.95E+01	0.94	10.45	0.02	3.96	2.76E+01	0.90
CPNE7	4	6	6.29	0.01	2.97	1.33E+01	0.84	6.28	0.01	2.97	1.33E+01	0.84	7.94	0.02	3.35	1.88E+01	0.87
PELP1	4	6	18.95	0.00	8.32	4.32E+01	0.95	28.93	0.00	12.77	6.55E+01	0.97	21.86	0.00	7.03	6.79E+01	0.95
ITGB4	4	6	12.75	0.00	5.63	2.89E+01	0.92	15.44	0.00	6.86	3.47E+01	0.94	10.75	0.01	4.07	2.84E+01	0.91
DOCK3	4	6	8.73	0.00	3.87	1.97E+01	0.89	13.16	0.00	5.85	2.96E+01	0.92	10.77	0.01	4.08	2.84E+01	0.91
FRYL	6	5	32.02	0.00	13.87	7.39E+01	0.97	21.45	0.00	9.52	4.84E+01	0.95	27.13	0.00	8.27	8.90E+01	0.96

Table S16 – Continuation.

Gene symbol	*Number variants*	*Number individuals*	*gnomAD NFE*					*gnomAD global*					*CSVS*				
			OR	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*
EP400	5	5	122.93	0.00	44.58	3.39E+02	0.99	131.96	0.00	51.19	3.40E+02	0.99	54.27	0.01	10.5	2.80E+02	0.98
WDR7	5	5	461.07	0.00	123.6	1.72E+03	1.00	371.97	0.00	128.9	1.07E+03	1.00	-	-	-	-	-
TRANK1	5	5	614.95	0.00	146.7	2.58E+03	1.00	177.95	0.00	67.50	4.69E+02	0.99	-	-	-	-	-
PDE6B	5	5	68.30	0.00	26.24	1.78E+02	0.99	58.46	0.00	23.54	1.45E+02	0.98	14.44	0.02	4.71	4.42E+01	0.93
SPEN	5	5	16.03	0.00	6.53	3.93E+01	0.94	12.90	0.00	5.32	3.13E+01	0.92	52.38	0.02	10.1	2.70E+02	0.98
LNX1	5	5	13.35	0.00	5.46	3.27E+01	0.93	14.75	0.00	6.08	3.58E+01	0.93	14.14	0.03	4.62	4.33E+01	0.93
COL6A1	5	5	9.55	0.01	3.92	2.33E+01	0.90	15.92	0.00	6.55	3.87E+01	0.94	12.88	0.04	4.31	3.85E+01	0.92
SYNJ2	4	5	48.59	0.00	19.07	1.24E+02	0.98	89.04	0.00	35.28	2.25E+02	0.99	22.63	0.01	6.53	7.83E+01	0.96
LRP6	4	5	32.98	0.00	13.17	8.26E+01	0.97	59.40	0.00	23.89	1.48E+02	0.98	23.26	0.01	6.72	8.05E+01	0.96
MMP9	4	5	16.77	0.00	6.82	4.12E+01	0.94	19.30	0.00	7.93	4.70E+01	0.95	19.38	0.01	5.90	6.37E+01	0.95
NCL	4	5	14.64	0.00	5.97	3.59E+01	0.93	13.55	0.00	5.58	3.29E+01	0.93	19.30	0.02	5.57	6.68E+01	0.95
EHBP1L1	4	5	10.14	0.00	4.16	2.47E+01	0.90	14.16	0.00	5.83	3.44E+01	0.93	16.12	0.02	5.10	5.09E+01	0.94
LZTS2	5	4	33.53	0.00	13.39	8.40E+01	0.97	8.23	0.02	3.40	1.99E+01	0.88	15.89	0.02	5.03	5.02E+01	0.94
TLN1	4	4	245.94	0.00	69.25	8.73E+02	1.00	327.48	0.00	102.4	1.05E+03	1.00	-	-	-	-	-
NBEAL2	4	4	368.90	0.00	92.08	1.48E+03	1.00	467.84	0.00	136.6	1.60E+03	1.00	-	-	-	-	-
GCN1	4	4	54.65	0.00	19.07	1.57E+02	0.98	90.94	0.00	32.28	2.56E+02	0.99	-	-	-	-	-
CFAP65	4	4	36.89	0.00	13.16	1.03E+02	0.97	57.46	0.00	20.79	1.59E+02	0.98	-	-	-	-	-
DCTPP1	3	4	739.39	0.00	135.1	4.05E+03	1.00	546.96	0.00	153.9	1.94E+03	1.00	-	-	-	-	-
TOPBP1	3	3	71.21	0.00	27.24	1.86E+02	0.99	41.08	0.00	16.66	1.01E+02	0.98	29.14	0.00	7.80	1.09E+02	0.97
PLPP7	3	3	492.64	0.00	109.9	2.21E+03	1.00	546.61	0.00	153.8	1.94E+03	1.00	-	-	-	-	-
DNM2	3	3	221.32	0.00	52.76	9.29E+02	1.00	491.08	0.00	117.1	2.06E+03	1.00	-	-	-	-	-
ATP10D	3	3	221.31	0.00	52.75	9.28E+02	1.00	90.93	0.00	27.50	3.01E+02	0.99	-	-	-	-	-
AP1AR	3	3	276.60	0.00	61.75	1.24E+03	1.00	129.20	0.00	38.12	4.38E+02	0.99	-	-	-	-	-
FHOD1	3	3	158.12	0.00	40.78	6.13E+02	0.99	188.85	0.00	53.66	6.65E+02	0.99	-	-	-	-	-
NUP160	3	3	368.87	0.00	74.28	1.83E+03	1.00	490.94	0.00	117.0	2.06E+03	1.00	-	-	-	-	-

Table S16 - Continuation.

Gene symbol	*Number variants*	*Number individuals*	*gnomAD NFE*					*gnomAD global*					*CSVS*				
			OR	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*	*OR*	*low CI*	*high CI*	*FDR*	*EF*
REXO1	3	3	368.82	0.00	74.27	1.83E+03	1.00	818.51	0.00	164.8	4.06E+03	1.00	-	-	-	-	-
MAST1	3	3	368.80	0.00	74.26	1.83E+03	1.00	409.07	0.00	102.0	1.64E+03	1.00	-	-	-	-	-
TRIO	3	3	552.73	0.00	92.17	3.31E+03	1.00	136.24	0.00	40.01	4.64E+02	0.99	-	-	-	-	-
PLXNA3	3	3	57.38	0.00	16.56	1.99E+02	0.98	66.94	0.00	20.24	2.21E+02	0.99	-	-	-	-	-
SEPHS2	3	3	1107.13	0.00	114.9	1.07E+04	1.00	1228.69	0.00	204.9	7.37E+03	1.00	-	-	-	-	-
NCKAP1L	3	3	1106.67	0.00	114.9	1.07E+04	1.00	1227.38	0.00	204.7	7.36E+03	1.00	-	-	-	-	-
CACNA1S	3	3	1106.44	0.00	114.9	1.07E+04	1.00	2453.59	0.00	254.8	2.36E+04	1.00	-	-	-	-	-
MAP3K7	3	3	1105.81	0.00	114.8	1.06E+04	1.00	408.99	0.00	102.0	1.64E+03	1.00	-	-	-	-	-
SMAP2	3	3	39.52	0.00	11.98	1.30E+02	0.97	54.57	0.00	16.90	1.76E+02	0.98	-	-	-	-	-
NDRG1	3	3	33.50	0.00	10.24	1.10E+02	0.97	20.59	0.00	6.53	6.50E+01	0.95	-	-	-	-	-
PTPN18	3	3	18.14	0.01	5.67	5.80E+01	0.94	31.90	0.00	10.03	1.01E+02	0.97	-	-	-	-	-
NRTN	2	3	92.31	0.00	37.50	2.27E+02	0.99	40.83	0.00	17.87	9.33E+01	0.98	34.91	0.00	9.81	1.24E+02	0.97
ZNF626	2	3	370.05	0.00	74.43	1.84E+03	1.00	615.93	0.00	137.3	2.76E+03	1.00	-	-	-	-	-
TGM2	2	3	110.99	0.00	30.42	4.05E+02	0.99	223.88	0.00	62.19	8.06E+02	1.00	-	-	-	-	-
CLCN7	2	3	555.04	0.00	92.46	3.33E+03	1.00	308.00	0.00	81.37	1.17E+03	1.00	-	-	-	-	-
C19orf47	2	3	554.96	0.00	92.44	3.33E+03	1.00	492.51	0.00	117.2	2.07E+03	1.00	-	-	-	-	-
HACD3	2	3	35.80	0.00	10.89	1.18E+02	0.97	63.15	0.00	19.42	2.05E+02	0.98	-	-	-	-	-

gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; OR: Odds ratio; CI: Confidence interval; FDR: p-value corrected by false discovery rate; EF: Etiological fraction.

Table S17 - Summary of variants found in the PTTG2 gene, enriched in variants with a high confidence of being loss-of-function (HIGH HC), for the I4 subgroup and the whole Meniere Disease (MD) cohort.

Gene symbol	Variant	Amino acid change	Consequence	CADD	AF				Individuals
					gnomAD NFE	gnomAD global	CSVS	MD	
PTTG2	chr4:37960987CAT>C	H185X	Frameshift	19.08	1.10E-02	1.19E-02	-	3.23E-02	I1-19, I1-23, I1-28, I1-63, I1-84, I2-7, I2-12, I2-13, I2-20, I2-25, I3-15, I3-25, I4-15, I4-32, I4-49, I4-57, I4-61, I4-62, I4-67, I4-73

CADD: Combined Annotation Dependent Depletion; AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population.

Table S18 - Summary of variants found in the CNTNAP2 gene, enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20), for the I4 subgroup and the whole Meniere Disease cohort.

Gene symbol	Variant	Amino acid change	Consequence	CADD	AF				Individuals
					gnomAD NFE	gnomAD global	CSVS	MD	
CNTNAP2	chr7:146839902T>G	W134G	Missense & Splice region	26.3	3.41E-04	2.09E-04	1.00E-03	1.61E-03	I1-81
	chr7:147043983G>A	R160H	Missense	28.5	6.19E-04	4.68E-04	0	1.61E-03	I1-19
	chr7:147044015C>T	R171C	Missense	25.5	6.20E-05	3.50E-05	0	1.61E-03	I2-2
	chr7:147044039T>G	Y179D	Missense	25.4	-	-	-	1.61E-03	I4-5
	chr7:147121078G>C	G285A	Missense	23.8	5.30E-03	3.97E-03	2.00E-03	4.84E-03	I3-50, I3-75, I4-17
	chr7:147121086A>G	I288V	Missense	16.92	4.60E-05	2.10E-05	-	1.61E-03	I4-18
	chr7:147121123G>A	R300H	Missense	24.4	0	4.90E-05	-	1.61E-03	I4-64
	chr7:147395676G>C	Q522H	Missense	24.9	-	-	0	1.61E-03	I3-10
	chr7:147639255G>A	E683K	Missense	25.3	0	2.10E-05	-	1.61E-03	I4-26
	chr7:147903589T>C	V708A	Missense	24.3	7.70E-05	1.47E-03	-	3.23E-03	I1-4, I4-20
	chr7:147903652C>T	A729V	Missense	23.3	0	4.90E-05	0	1.61E-03	I4-2
	chr7:147977966G>C	G787A	Missense	25.8	1.50E-05	2.10E-05	-	3.23E-03	I1-66, I3-27
	chr7:148383834C>G	L1221V	Missense	15.53	3.10E-05	1.40E-05	-	1.61E-03	I3-56

CADD: Combined Annotation Dependent Depletion; AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population.

Table S19 - Summary of variants found in the ADAM2 gene, enriched in variants with a high confidence of being loss-of-function (HIGH HC), for the I1 subgroup and the whole Meniere Disease (MD) cohort.

Gene symbol	Variant	Amino acid change	Consequence	CADD	AF				Individuals
					gnomAD NFE	gnomAD global	CSVS	MD	
ADAM2	chr8:39755912C>T	-	Splice acceptor	23.2	0	7.00E-06	1.00E-03	1.61E-03	I1-17
	chr8:39769449T>TAGTG	A385ATX	Frameshift	22.8	2.79E-04	4.33E-04	-	1.45E-02	I1-36, I1-54(hom), I2-47, I3-21, I3-32, I4-8, I4-40, I4-72
	chr8:39788742T>A	-	Splice acceptor	27.1	7.79E-03	5.25E-03	6.00E-03	8.07E-03	I1-27, I1-78, I3-67, I4-33, I4-54
	chr8:39821106C>A	E137*	Stop gained	35	-	-	-	1.61E-03	I1-48

*CADD: Combined Annotation Dependent Depletion; AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; *: Stop codon; hom: homozygous.*

Table S20 - Summary of variants found in the FRYL, FLT4 and SCN10A genes, enriched in variants with a predicted low confidence of being loss-of-function, and variants with a moderate impact in the protein (mostly missense variants) and predicted to be deleterious (HIGH LC + MODERATE CADD ≥ 20), for the I1 subgroup and the whole Meniere Disease (MD) cohort.

Gene symbol	Variant	Amino acid change	Consequence	CADD	AF				Individuals
					gnomAD NFE	gnomAD global	CSVS	MD	
FRYL	chr4:48499487C>T	V2993M	Missense	23.7	1.50E-05	1.40E-05	0	1.61E-03	I4-35
	chr4:48500184T>C	S2877G	Missense	19.45	8.05E-04	4.74E-04	1.00E-03	1.61E-03	I3-17
	chr4:48505561T>G	N2817H	Missense	23.8	-	-	-	1.61E-03	I2-46
	chr4:48510858T>C	T2758A	Missense	23.9	3.10E-05	1.40E-05	0	1.61E-03	I3-47
	chr4:48510945C>T	V2729I	Missense	21.9	7.70E-05	4.90E-05	0	1.61E-03	I2-6
	chr4:48527999G>A	P2371L	Missense	31	-	-	-	1.61E-03	I4-5
	chr4:48540000A>T	C2122S	Missense	27.8	0	2.10E-05	-	1.61E-03	I1-77
	chr4:48540568T>C	N2027S	Missense	19.04	3.25E-04	1.95E-04	1.00E-03	3.23E-03	I1-67, I4-49
	chr4:48540784C>T	R1955Q	Missense	20.9	6.20E-05	1.05E-04	1.00E-03	1.61E-03	I3-30
	chr4:48547590A>G	F1690L	Missense	22.2	1.08E-04	1.15E-03	1.00E-03	1.61E-03	I3-73
	chr4:48553248T>C	T1468A	Missense	22.2	8.36E-04	1.46E-03	1.00E-03	3.23E-03	I1-48, I2-7
	chr4:48557067C>A	A1393S	Missense	22.4	2.03E-04	1.00E-04	0	1.61E-03	I1-40
	chr4:48562903T>C	M1228V	Missense	22.1	-	-	-	1.61E-03	I4-55
	chr4:48576086A>C	S889A	Missense	22.1	1.08E-04	8.40E-05	-	3.23E-03	I2-17, I4-21
	chr4:48579072T>C	K810R	Missense	18.47	6.20E-05	3.30E-03	2.00E-03	1.61E-03	I2-42
	chr4:48581473C>T	V707I	Missense	19.47	1.49E-02	9.86E-03	1.20E-02	1.45E-02	I1-13, I2-7, I2-20, I2-67, I3-45, I3-66, I3-77, I4-7, I4-29
	chr4:48589875T>C	M504V	Missense & Splice region	22.3	1.50E-05	7.00E-06	-	1.61E-03	I1-54
	chr4:48595990G>A	P349L	Missense	24.5	1.60E-05	7.00E-06	-	1.61E-03	I1-48

Table S20 - Continuation.

Gene symbol	Variant	Amino acid change	Consequence	CADD	AF				Individuals
					gnomAD NFE	gnomAD global	CSVS	MD	
FRYL	chr4:48609804G>T	P144H	Missense	22.1	-	-	-	1.61E-03	I1-65
	chr4:48620714T>C	D80G	Missense	24.8	-	-	-	1.61E-03	I3-47
	chr4:48623151C>A	R50M	Missense	25.5	-	-	-	1.61E-03	I3-63
FLT4	chr5:180603326G>A	R1320W	Missense	23.6	4.18E-04	2.58E-04	0	1.61E-03	I4-43
	chr5:180609010A>G	F1284S	Missense	26.5	1.50E-05	7.00E-06	-	1.61E-03	I1-40
	chr5:180609924G>A	T1263M	Missense	15.66	4.60E-05	4.90E-05	-	1.61E-03	I2-16
	chr5:180616393C>G	D1065H	Missense	28.9	1.50E-05	7.00E-06	-	1.61E-03	I1-17
	chr5:180616920C>T	E1026K	Missense	29.6	-	-	-	1.61E-03	I3-1
	chr5:180618911G>A	P954S	Missense	23.1	5.95E-03	3.36E-03	3.00E-03	6.45E-03	I3-6, I3-15, I3-38, I4-67
	chr5:180620608CCCT>C	R802-	In frame deletion & Splice region	25.6	1.50E-05	7.00E-06	-	3.23E-03	I3-63, I4-52
	chr5:180620888C>G	V763L	Missense	22.6	-	-	-	1.61E-03	I4-36
	chr5:180621158G>C	H705Q	Missense	19.97	-	-	-	1.61E-03	I3-2
	chr5:180625965G>A	A442V	Missense	23	1.39E-04	2.16E-04	1.00E-03	3.23E-03	I1-31, I1-50
	chr5:180626236C>T	R378H	Missense	24.6	2.48E-04	2.44E-04	-	1.61E-03	I1-14
	chr5:180629409C>T	V279M	Missense	25.9	-	-	-	1.61E-03	I4-50
	chr5:180630035G>A	T195M	Missense	22.9	1.50E-05	1.40E-05	0	1.61E-03	I1-46
	chr5:180630056C>T	R188Q	Missense	21.6	1.08E-04	1.12E-04	-	1.61E-03	I1-78

Table S20 - Continuation.

Gene symbol	*Variant*	*Amino acid change*	*Consequence*	*CADD*	*AF*				*Individuals*
					gnomAD NFE	*gnomAD global*	*CSVS*	*MD*	
SCN10A	chr3:38697520A>T	N1900K	Missense	15.4	0	1.40E-05	0	1.61E-03	I3-42
	chr3:38698236C>T	G1662S	Missense	25.1	8.52E-04	6.14E-04	1.00E-03	3.23E-03	I1-35, I3-66
	chr3:38698265A>T	L1652Q	Missense	28.4	-	-	-	1.61E-03	I2-61
	chr3:38698484C>T	R1579Q	Missense	26.3	3.10E-05	6.30E-05	-	1.61E-03	I1-28
	chr3:38701915C>T	M1527I	Missense	15.33	9.30E-05	1.33E-04	0	1.61E-03	I4-65
	chr3:38713959C>T	R1268Q	Missense & Splice region	25.5	4.20E-03	2.53E-03	-	1.61E-03	I2-14
	chr3:38713999G>C	L1255V	Missense	21.3	-	-	-	1.61E-03	I3-53
	chr3:38718660A>G	I1225T	Missense	27.6	6.04E-04	4.68E-04	2.00E-03	3.23E-03	I1-4, I1-56
	chr3:38722404G>A	R1121C	Missense	23.8	3.25E-04	2.30E-04	1.00E-03	1.61E-03	I4-52
	chr3:38728754C>A	G810W	Missense	24.8	5.27E-04	2.93E-04	1.00E-03	4.84E-03	I1-22, I1-64, I3-54
	chr3:38739519C>A	R759L	Missense	27.8	0	7.00E-06	-	1.61E-03	I1-22
	chr3:38752313A>G	L554P	Missense	16.92	1.39E-04	9.80E-05	1.00E-03	1.61E-03	I3-25
	chr3:38757154C>A	C319F	Missense	26.3	-	-	-	1.61E-03	I2-58
	chr3:38763578T>C	I206M	Missense	16.19	1.52E-02	1.40E-02	1.80E-02	2.74E-02	I1-8, I1-14, I1-26, I1-37, I1-53, I1-76, I2-38, I2-61, I3-12, I3-15(hom), I3-41, I4-24(hom), I4-28, I4-41, I4-54
	chr3:38771406A>C	Y158D	Missense & Splice region	24.2	5.11E-04	3.21E-04	0	1.61E-03	I2-39
	chr3:38789020T>C	I136V	Missense	23.2	3.10E-05	4.20E-05	-	1.61E-03	I2-26

CADD: Combined Annotation Dependent Depletion; AF: Allele frequency; gnomAD NFE: Non-Finnish European for gnomAD; gnomAD: Global population for gnomAD; CSVS: Collaborative Spanish Variant Server, Spanish population; hom: homozygous.

Table S21 - American College of Medical Genetics and Genomics (ACMG) criteria for the structural variants found in AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93 and TAF6 genes.

Variant		**chr7:100089053-100112261-DUP**
Gene(s)		*AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93, TAF6*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 11 domains in 4 proteins: 5 x MCM7, HUMAN, 3 x TAF6, HUMAN, 2 x CSN6, HUMAN and 1 x AP4M1, HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 4 coding genes: AP4M1, COPS6, MCM7 and TAF6.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Uncertain significance	Found 0 benign CNVs, 0 common variants by DGV and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	No condition is met.

Table S22 - American College of Medical Genetics and Genomics (ACMG) criteria for the structural variants found in ERBB3 gene.

Variant		**chr12:56100028-56100172-DEL**
Gene(s)		*ERBB3*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 1 domain in 1 protein: 1 x ERBB3_HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 1 coding gene: ERBB3.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Benign	Found 1 common variant by DGV, 0 benign CNVs and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	CNV has both break points in a loss-of-function causing gene, but it doesn't contain any known loss-of-function variant.

Variant		**chr12:56100243-56101058-DEL**
Gene(s)		*ERBB3*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 1 domain in 1 protein: 1 x ERBB3_HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 1 coding gene: ERBB3.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Benign	Found 1 common variant by DGV, 0 benign CNVs and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	CNV has both break points in a loss-of-function causing gene, but it doesn't contain any known loss-of-function variant.

Table S22 - Continuation.

Variant		**chr12:56101359-56101526-DEL**
Gene(s)		*ERBB3*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant doesn't affect any known domain, but affects 2 coding-genes.
Gene	Uncertain significance	This structural variant affects 2 coding genes: ENSG00000257411 and ERBB3.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Benign	Found 1 common variant by DGV, 0 benign CNVs and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	CNV has both break points in a loss-of-function causing gene, but it doesn't contain any known loss-of-function variant.

Variant		**chr12:56100173-56101208-DEL**
Gene(s)		*ERBB3*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 1 domain in 1 protein: 1 x ERBB3_HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 1 coding gene: ERBB3.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Benign	Found 1 common variant by DGV, 0 benign CNVs and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	CNV has both breakpoints in the same loss-of-function causing gene, contains 2 loss-of-function causing variants and a coding region of a loss-of-function gene.

Table S23 - American College of Medical Genetics and Genomics (ACMG) criteria for the structural variants found in COL14A1 gene.

Variant		**chr8:120280928-120281118-DUP**
Gene(s)		*COL14A1*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 2 domains in 1 protein: 2 x COEA1_HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 1 coding gene: COL14A1.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Uncertain significance	Found 0 benign CNVs, 0 common variants by DGV and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	No condition is met.

Table S24 - American College of Medical Genetics and Genomics (ACMG) criteria for the structural variants found in GMCL1 gene.

Variant		**chr2:69847687-69847797-DEL**
Gene(s)		*GMCL1*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant doesn't affect any known domain, but affects 1 coding-gene.
Gene	Uncertain significance	This structural variant affects 1 coding gene: GMCL1.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Uncertain significance	Found 0 benign CNVs, 0 common variants by DGV and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	No condition is met.

Table S25 - American College of Medical Genetics and Genomics (ACMG) criteria for the structural variants found in PAPLN gene.

Variant		**chr14:73263837-73264037-DUP**
Gene(s)		*PAPLN*
Pathogenicity		Benign
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 1 domain in 1 protein: 1 x PPN_HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 1 coding gene: PAPLN.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Benign	Found 1 common variant by DGV, 0 benign CNVs and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	No condition is met.

Variant		**chr14:73263839-73264057-DUP**
Gene(s)		*PAPLN*
Pathogenicity		Benign
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 1 domain in 1 protein: 1 x PPN_HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 1 coding gene: PAPLN.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Benign	Found 1 common variant by DGV, 0 benign CNVs and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	No condition is met.

Table S26 - American College of Medical Genetics and Genomics (ACMG) criteria for the single nucleotide variants found in PDE6B gene.

Variant		**chr4:653900G>A**
Gene(s)		*PDE6B*
Pathogenicity		Uncertain significance with some pathogenic evidence
Criteria:		
PM1	Moderate	Hot-spot of length 17 amino-acids has 9 missense/in-frame variants (5 pathogenic variants, 4 uncertain variants and no benign), which qualifies as moderate pathogenic.
		UniProt protein PDE6B_HUMAN domain 'GAF 2' has 96 missense/in-frame variants (14 pathogenic variants, 80 uncertain variants and 2 benign variants), which qualifies as supporting pathogenic.
PP3	Moderate	MetaRNN = 0.893 is between 0.841 and 0.939 ⇒ moderate pathogenic.
PM2	Supporting	GnomAD genomes homozygous allele count = 1 is less than 2 for AD/AR gene PDE6B, good gnomAD genomes coverage = 32.7.
		GnomAD exomes homozygous allele count = 1 is less than 2 for AD/AR gene PDE6B, gnomAD exomes coverage is unavailable.

Variant		**chr4:654132G>A**
Gene(s)		*PDE6B*
Pathogenicity		Likely benign
Criteria:		
PM1	Supporting	UniProt protein PDE6B_HUMAN domain 'GAF 2' has 96 missense/in-frame variants (14 pathogenic variants, 80 uncertain variants and 2 benign variants), which qualifies as supporting pathogenic.
PM2	Supporting	GnomAD genomes homozygous allele count = 0 is less than 2 for AD/AR gene PDE6B, good gnomAD genomes coverage = 32.7.
		GnomAD exomes homozygous allele count = 0 is less than 2 for AD/AR gene PDE6B, gnomAD exomes coverage is unavailable.
BP4	Strong	MetaRNN = 0.0198 is between 0.00692 and 0.108 ⇒ strong benign.

Table S26 - Continuation.

Variant		**chr4:655980C>G**
Gene(s)		*PDE6B*
Pathogenicity		Uncertain significance
Criteria:		
PM1	Supporting	UniProt protein PDE6B_HUMAN domain 'GAF 2' has 96 missense/in-frame variants (14 pathogenic variants, 80 uncertain variants and 2 benign variants), which qualifies as supporting pathogenic.
PM2	Supporting	Variant not found in gnomAD genomes, good gnomAD genomes coverage = 33.0.
		Variant not found in gnomAD exomes, gnomAD exomes coverage is unavailable.
PP3	Supporting	MetaRNN = 0.837 is between 0.748 and 0.841 ⇒ supporting pathogenic.

Variant		**chr4:662584G>A**
Gene(s)		*PDE6B*
Pathogenicity		Pathogenic
Criteria:		
PP5	Very strong	ClinVar classifies this variant as Pathogenic, 2 stars (multiple consistent, reviewed Feb '23, 4 submissions), citing 3 articles (31877679, 26868535 and 18723146), associated with Congenital Stationary Night Blindness Autosomal Dominant 2 and Retinitis Pigmentosa 40.
PP3	Strong	MetaRNN = 0.956 is greater than 0.939 ⇒ strong pathogenic.
PM1	Supporting	UniProt protein PDE6B_HUMAN metal ion binding 'Divalent metal cation 1' has 102 missense/in-frame variants (23 pathogenic variants, 78 uncertain variants and 1 benign variant), which qualifies as supporting pathogenic.
		UniProt protein PDE6B_HUMAN domain 'PDEase' has 203 missense/in-frame variants (54 pathogenic variants, 147 uncertain variants and 2 benign variants), which qualifies as supporting pathogenic.
PM2	Supporting	Variant not found in gnomAD genomes, good gnomAD genomes coverage = 31.9.
		GnomAD exomes homozygous allele count = 0 is less than 2 for AD/AR gene PDE6B, gnomAD exomes coverage is unavailable.

Table S26 - Continuation.

Variant		**chr4:667929C>T**
Gene(s)		*PDE6B*
Pathogenicity		Uncertain significance
Criteria:		
PM1	Moderate	Hot-spot of length 17 amino-acids has 10 missense/in-frame variants (4 pathogenic variants, 6 uncertain variants and no benign), which qualifies as moderate pathogenic.
		UniProt protein PDE6B_HUMAN domain 'PDEase' has 203 missense/in-frame variants (54 pathogenic variants, 147 uncertain variants and 2 benign variants), which qualifies as supporting pathogenic.
PM2	Supporting	GnomAD genomes homozygous allele count = 0 is less than 2 for AD/AR gene PDE6B, good gnomAD genomes coverage = 32.2.
		GnomAD exomes homozygous allele count = 0 is less than 2 for AD/AR gene PDE6B, gnomAD exomes coverage is unavailable.
BP4	Supporting	MetaRNN = 0.41 is between 0.267 and 0.43 ⇒ supporting benign.

Table S27 - American College of Medical Genetics and Genomics (ACMG) criteria for the structural variants found in PDE6B gene.

Variant		**chr4:635416-635529-DEL**
Gene(s)		*PDE6B*
Pathogenicity		Uncertain significance
Criteria:		
Genomic Content	Uncertain significance	This structural variant affects 1 domain in 1 protein: 1 x PDE6B_HUMAN reported in UniProt Regions.
Gene	Uncertain significance	This structural variant affects 1 coding gene: PDE6B.
Inheritance	Uncertain significance	No phenotypes or diseases provided.
Literature	Benign	Found 1 benign CNV reported by 1 x dbVar, 3 common variants by DGV and 0 pathogenic CNVs reported.
Gene/Regions Overlap	Uncertain significance	CNV has both break points in a loss-of-function causing gene, but it doesn't contain any known loss-of-function variant.

Table S28 - Biological processes from Gene Ontology database associated with the 28 candidate genes for the I4 subgroup.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
pyrimidine dimer repair	GO:0006290	2	19	3	8272	1.50E-05	290.25	*ERCC6, MSH6*
positive regulation of transcription initiation from RNA polymerase II promoter	GO:0060261	2	19	18	8272	7.48E-04	48.37	*ERCC6, TAF6*
positive regulation of protein tyrosine kinase activity	GO:0061098	2	19	19	8272	8.35E-04	45.83	*ERBB3, ERCC6*
synaptic transmission, glutamatergic	GO:0035249	2	19	19	8272	8.35E-04	45.83	*GRIK4, UNC13C*
response to UV	GO:0009411	2	19	22	8272	1.12E-03	39.58	*ERCC6, MSH6*
intrinsic apoptotic signaling pathway in response to DNA damage	GO:0008630	2	19	26	8272	1.57E-03	33.49	*ERCC6, MSH6*
positive regulation of apoptotic DNA fragmentation	GO:1902512	1	19	1	8272	2.30E-03	435.37	*ENDOG*
mitochondrial DNA catabolic process	GO:0032043	1	19	1	8272	2.30E-03	435.37	*ENDOG*
positive regulation of cardiac muscle tissue development	GO:0055025	1	19	1	8272	2.30E-03	435.37	*ERBB3*
negative regulation of synaptic plasticity	GO:0031914	1	19	1	8272	2.30E-03	435.37	*UNC13C*
positive regulation of mitochondrial DNA replication	GO:0090297	1	19	1	8272	2.30E-03	435.37	*ENDOG*
cellular response to DNA damage stimulus	GO:0006974	4	19	272	8272	3.00E-03	6.40	*MCM7, ERCC6, RAD50, MSH6*
double-strand break repair via classical nonhomologous end joining	GO:0097680	1	19	2	8272	4.59E-03	217.68	*ERCC6*
regulation of mitotic recombination	GO:0000019	1	19	2	8272	4.59E-03	217.68	*RAD50*
negative regulation of ATPase-coupled calcium transmembrane transporter activity	GO:1901895	1	19	2	8272	4.59E-03	217.68	*THADA*
cranial nerve development	GO:0021545	1	19	2	8272	4.59E-03	217.68	*ERBB3*
CD8-positive, alpha-beta T cell activation	GO:0036037	1	19	2	8272	4.59E-03	217.68	*WDFY4*
negative regulation of secretion	GO:0051048	1	19	2	8272	4.59E-03	217.68	*ERBB3*
negative regulation of endoplasmic reticulum calcium ion concentration	GO:0032471	1	19	2	8272	4.59E-03	217.68	*THADA*
chromosome organization involved in meiotic cell cycle	GO:0070192	1	19	2	8272	4.59E-03	217.68	*RAD50*
axonogenesis	GO:0007409	2	19	49	8272	5.51E-03	17.77	*SPTBN4, DSCAML1*
DNA duplex unwinding	GO:0032508	2	19	52	8272	6.19E-03	16.74	*MCM7, RAD50*
positive regulation of kinase activity	GO:0033674	2	19	54	8272	6.66E-03	16.12	*ERBB3, RAD50*
reproductive process	GO:0022414	1	19	3	8272	6.88E-03	145.12	*SPTBN4*
response to superoxide	GO:0000303	1	19	3	8272	6.88E-03	145.12	*ERCC6*
negative regulation of heart rate	GO:0010459	1	19	3	8272	6.88E-03	145.12	*SPTBN4*
telomeric 3' overhang formation	GO:0031860	1	19	3	8272	6.88E-03	145.12	*RAD50*
meiotic mismatch repair	GO:0000710	1	19	3	8272	6.88E-03	145.12	*MSH6*

Table S28 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
single strand break repair	GO:0000012	1	19	3	8272	6.88E-03	145.12	*ERCC6*
positive regulation of hydrogen peroxide-mediated programmed cell death	GO:1901300	1	19	3	8272	6.88E-03	145.12	*ENDOG*
neuron death in response to oxidative stress	GO:0036475	1	19	3	8272	6.88E-03	145.12	*ENDOG*
dense core granule priming	GO:0061789	1	19	3	8272	6.88E-03	145.12	*UNC13C*
adaptive thermogenesis	GO:1990845	1	19	3	8272	6.88E-03	145.12	*THADA*
presynaptic dense core vesicle exocytosis	GO:0099525	1	19	3	8272	6.88E-03	145.12	*UNC13C*
negative regulation of telomere capping	GO:1904354	1	19	3	8272	6.88E-03	145.12	*RAD50*
negative regulation of double-strand break repair via nonhomologous end joining	GO:2001033	1	19	3	8272	6.88E-03	145.12	*ERCC6*
DNA recombination	GO:0006310	2	19	58	8272	7.65E-03	15.01	*RAD50, ENDOG*
transcription by RNA polymerase II	GO:0006366	2	19	59	8272	7.91E-03	14.76	*ERCC6, TAF6*
positive regulation of helicase activity	GO:0051096	1	19	4	8272	9.16E-03	108.84	*MSH6*
synaptic vesicle maturation	GO:0016188	1	19	4	8272	9.16E-03	108.84	*UNC13C*
auditory receptor cell development	GO:0060117	1	19	4	8272	9.16E-03	108.84	*PDZD7*
regulation of peptidyl-serine phosphorylation	GO:0033135	1	19	4	8272	9.16E-03	108.84	*SPTBN4*
regulation of transcription elongation from RNA polymerase II promoter	GO:0034243	1	19	4	8272	9.16E-03	108.84	*ERCC6*
endocardial cushion development	GO:0003197	1	19	4	8272	9.16E-03	108.84	*ERBB3*
synaptic vesicle docking	GO:0016081	1	19	4	8272	9.16E-03	108.84	*UNC13C*
transcription elongation from RNA polymerase I promoter	GO:0006362	1	19	4	8272	9.16E-03	108.84	*ERCC6*
protein deneddylation	GO:0000338	1	19	4	8272	9.16E-03	108.84	*COPS6*
DNA protection	GO:0042262	1	19	4	8272	9.16E-03	108.84	*ERCC6*
regulation of DNA-templated transcription, elongation	GO:0032784	1	19	5	8272	1.14E-02	87.07	*ERCC6*
Schwann cell differentiation	GO:0014037	1	19	5	8272	1.14E-02	87.07	*ERBB3*
maintenance of DNA repeat elements	GO:0043570	1	19	5	8272	1.14E-02	87.07	*MSH6*
apoptotic DNA fragmentation	GO:0006309	1	19	5	8272	1.14E-02	87.07	*ENDOG*
response to UV-B	GO:0010224	1	19	5	8272	1.14E-02	87.07	*ERCC6*
central nervous system projection neuron axonogenesis	GO:0021952	1	19	5	8272	1.14E-02	87.07	*SPTBN4*
somatic recombination of immunoglobulin gene segments	GO:0016447	1	19	5	8272	1.14E-02	87.07	*MSH6*
clustering of voltage-gated sodium channels	GO:0045162	1	19	5	8272	1.14E-02	87.07	*SPTBN4*
DNA repair	GO:0006281	3	19	221	8272	1.33E-02	5.91	*ERCC6, RAD50, MSH6*
DNA strand elongation involved in DNA replication	GO:0006271	1	19	6	8272	1.37E-02	72.56	*MCM7*

Table S28 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
telomere maintenance via recombination	GO:0000722	1	19	7	8272	1.60E-02	62.20	*RAD50*
positive regulation of peptidyl-serine phosphorylation of STAT protein	GO:0033141	1	19	7	8272	1.60E-02	62.20	*ERCC6*
positive regulation of DNA-templated transcription, elongation	GO:0032786	1	19	7	8272	1.60E-02	62.20	*ERCC6*
transcription-coupled nucleotide-excision repair	GO:0006283	1	19	7	8272	1.60E-02	62.20	*ERCC6*
fatty acid catabolic process	GO:0009062	1	19	7	8272	1.60E-02	62.20	*LPIN3*
positive regulation of calcineurin-NFAT signaling cascade	GO:0070886	1	19	7	8272	1.60E-02	62.20	*ERBB3*
stress fiber assembly	GO:0043149	1	19	7	8272	1.60E-02	62.20	*ITGB5*
positive regulation of telomere maintenance	GO:0032206	1	19	7	8272	1.60E-02	62.20	*RAD50*
regulation of phosphorylation	GO:0042325	1	19	8	8272	1.82E-02	54.42	*MCM7*
positive regulation of protein autophosphorylation	GO:0031954	1	19	8	8272	1.82E-02	54.42	*RAD50*
glycoprotein catabolic process	GO:0006516	1	19	8	8272	1.82E-02	54.42	*FBXO27*
synaptic vesicle exocytosis	GO:0016079	1	19	8	8272	1.82E-02	54.42	*UNC13C*
epithelial cell-cell adhesion	GO:0090136	1	19	8	8272	1.82E-02	54.42	*ITGB5*
cardiac conduction	GO:0061337	1	19	8	8272	1.82E-02	54.42	*SPTBN4*
synaptic vesicle priming	GO:0016082	1	19	8	8272	1.82E-02	54.42	*UNC13C*
sensory perception of sound	GO:0007605	2	19	93	8272	1.89E-02	9.36	*PDZD7, SPTBN4*
cell fate determination	GO:0001709	1	19	9	8272	2.05E-02	48.37	*DSCAML1*
telomere maintenance via telomerase	GO:0007004	1	19	9	8272	2.05E-02	48.37	*RAD50*
RNA polymerase II preinitiation complex assembly	GO:0051123	1	19	9	8272	2.05E-02	48.37	*TAF6*
isotype switching	GO:0045190	1	19	9	8272	2.05E-02	48.37	*MSH6*
response to X-ray	GO:0010165	1	19	10	8272	2.27E-02	43.54	*ERCC6*
somatic hypermutation of immunoglobulin genes	GO:0016446	1	19	10	8272	2.27E-02	43.54	*MSH6*
dendrite self-avoidance	GO:0070593	1	19	10	8272	2.27E-02	43.54	*DSCAML1*
positive regulation of response to DNA damage stimulus	GO:2001022	1	19	10	8272	2.27E-02	43.54	*ENDOG*
transmission of nerve impulse	GO:0019226	1	19	11	8272	2.50E-02	39.58	*SPTBN4*
negative regulation of DNA recombination	GO:0045910	1	19	11	8272	2.50E-02	39.58	*MSH6*
double-strand break repair via break-induced replication	GO:0000727	1	19	11	8272	2.50E-02	39.58	*MCM7*
extrinsic apoptotic signaling pathway in absence of ligand	GO:0097192	1	19	11	8272	2.50E-02	39.58	*ERBB3*
positive regulation of transcription by RNA polymerase III	GO:0045945	1	19	12	8272	2.72E-02	36.28	*ERCC6*
response to tumor necrosis factor	GO:0034612	1	19	12	8272	2.72E-02	36.28	*ENDOG*
triglyceride biosynthetic process	GO:0019432	1	19	12	8272	2.72E-02	36.28	*LPIN3*

Table S28 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
auditory receptor cell stereocilium organization	GO:0060088	1	19	12	8272	2.72E-02	36.28	*PDZD7*
regulation of sodium ion transport	GO:0002028	1	19	12	8272	2.72E-02	36.28	*SPTBN4*
intrinsic apoptotic signaling pathway	GO:0097193	1	19	12	8272	2.72E-02	36.28	*MSH6*
nervous system development	GO:0007399	3	19	291	8272	2.75E-02	4.49	*ERBB3, ERCC6, DSCAML1*
response to gamma radiation	GO:0010332	1	19	13	8272	2.95E-02	33.49	*ERCC6*
detection of mechanical stimulus involved in sensory perception of sound	GO:0050910	1	19	13	8272	2.95E-02	33.49	*PDZD7*
cellular response to organic substance	GO:0071310	1	19	13	8272	2.95E-02	33.49	*MCM7*
determination of adult lifespan	GO:0008340	1	19	13	8272	2.95E-02	33.49	*MSH6*
monoubiquitinated histone deubiquitination	GO:0035521	1	19	13	8272	2.95E-02	33.49	*TAF6*
DNA-templated transcription, initiation	GO:0006352	1	19	13	8272	2.95E-02	33.49	*TAF6*
positive regulation of double-strand break repair via homologous recombination	GO:1905168	1	19	14	8272	3.17E-02	31.10	*ERCC6*
glutamate receptor signaling pathway	GO:0007215	1	19	14	8272	3.17E-02	31.10	*GRIK4*
negative regulation of TOR signaling	GO:0032007	1	19	14	8272	3.17E-02	31.10	*ENDOG*
DNA unwinding involved in DNA replication	GO:0006268	1	19	14	8272	3.17E-02	31.10	*MCM7*
actin filament capping	GO:0051693	1	19	14	8272	3.17E-02	31.10	*SPTBN4*
tRNA methylation	GO:0030488	1	19	14	8272	3.17E-02	31.10	*THADA*
ionotropic glutamate receptor signaling pathway	GO:0035235	1	19	15	8272	3.39E-02	29.02	*GRIK4*
DNA replication initiation	GO:0006270	1	19	15	8272	3.39E-02	29.02	*MCM7*
monoubiquitinated histone H2A deubiquitination	GO:0035522	1	19	15	8272	3.39E-02	29.02	*TAF6*
chemical synaptic transmission	GO:0007268	2	19	129	8272	3.47E-02	6.75	*GRIK4, UNC13C*
positive regulation of defense response to virus by host	GO:0002230	1	19	16	8272	3.62E-02	27.21	*ERCC6*
transcription initiation from RNA polymerase II promoter	GO:0006367	1	19	16	8272	3.62E-02	27.21	*TAF6*
positive regulation of intrinsic apoptotic signaling pathway	GO:2001244	1	19	17	8272	3.84E-02	25.61	*TAF6*
positive regulation of multicellular organism growth	GO:0040018	1	19	17	8272	3.84E-02	25.61	*SPTBN4*
peripheral nervous system development	GO:0007422	1	19	17	8272	3.84E-02	25.61	*ERBB3*
response to antibiotic	GO:0046677	1	19	17	8272	3.84E-02	25.61	*ENDOG*
cellular response to virus	GO:0098586	1	19	18	8272	4.06E-02	24.19	*WDFY4*
mismatch repair	GO:0006298	1	19	18	8272	4.06E-02	24.19	*MSH6*
adult walking behavior	GO:0007628	1	19	18	8272	4.06E-02	24.19	*SPTBN4*
positive regulation of apoptotic process	GO:0043065	2	19	144	8272	4.24E-02	6.05	*ENDOG, TAF6*

Table S28 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
base-excision repair	GO:0006284	1	19	19	8272	4.28E-02	22.91	*ERCC6*
cellular lipid metabolic process	GO:0044255	1	19	19	8272	4.28E-02	22.91	*LPIN3*
DNA damage checkpoint signaling	GO:0000077	1	19	19	8272	4.28E-02	22.91	*ERCC6*
positive regulation of DNA repair	GO:0045739	1	19	19	8272	4.28E-02	22.91	*ERCC6*
adult behavior	GO:0030534	1	19	20	8272	4.50E-02	21.77	*SPTBN4*
antigen processing and presentation	GO:0019882	1	19	20	8272	4.50E-02	21.77	*WDFY4*
neuron apoptotic process	GO:0051402	1	19	21	8272	4.72E-02	20.73	*ERBB3*
response to mechanical stimulus	GO:0009612	1	19	21	8272	4.72E-02	20.73	*ENDOG*
dorsal/ventral pattern formation	GO:0009953	1	19	21	8272	4.72E-02	20.73	*DSCAML1*
interstrand cross-link repair	GO:0036297	1	19	22	8272	4.94E-02	19.79	*MSH6*
mRNA transcription by RNA polymerase II	GO:0042789	1	19	22	8272	4.94E-02	19.79	*TAF6*
endodermal cell differentiation	GO:0035987	1	19	22	8272	4.94E-02	19.79	*ITGB5*
reciprocal meiotic recombination	GO:0007131	1	19	22	8272	4.94E-02	19.79	*RAD50*

Table S29 - Phenotypes from Human Phenotype Ontology database associated with the 28 candidate genes for the I4 subgroup.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
Absent brainstem auditory responses	HP:0004463	2	8	4	2525	5.26E-05	157.81	*ERCC6, SPTBN4*
Neoplasm of the thyroid gland	HP:0100031	2	8	16	2525	1.03E-03	39.45	*ERBB3, MSH6*
Intestinal polyposis	HP:0200008	2	8	25	2525	2.54E-03	25.25	*ERBB3, MSH6*
Carious teeth	HP:0000670	3	8	100	2525	2.92E-03	9.47	*ERCC6, PDZD7, ARVCF*
Second metatarsal posteriorly placed	HP:0008125	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Vascular calcification	HP:0004934	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Lentiglobus	HP:0011527	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Subcortical white matter calcifications	HP:0007346	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Dense calcifications in the cerebellar dentate nucleus	HP:0002461	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Increased blood pressure	HP:0032263	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Patchy demyelination of subcortical white matter	HP:0002545	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Pigmentation anomalies of sun-exposed skin	HP:0007623	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Gonadal hypoplasia	HP:0008639	1	8	1	2525	3.17E-03	315.63	*ERCC6*
Ovarian neoplasm	HP:0100615	2	8	28	2525	3.19E-03	22.54	*RAD50, MSH6*
Prominent nasal bridge	HP:0000426	3	8	114	2525	4.25E-03	8.31	*ERCC6, ARVCF, TAF6*
Melanoma	HP:0002861	2	8	36	2525	5.24E-03	17.53	*ERCC6, RAD50*
Schizophrenia	HP:0100753	2	8	36	2525	5.24E-03	17.53	*PDZD7, ARVCF*
Areflexia	HP:0001284	3	8	129	2525	6.04E-03	7.34	*ERBB3, ERCC6, SPTBN4*
Abnormal peripheral myelination	HP:0003130	1	8	2	2525	6.33E-03	157.81	*ERCC6*
Degenerative vitreoretinopathy	HP:0007964	1	8	2	2525	6.33E-03	157.81	*ERBB3*
Refractory anemia with ringed sideroblasts	HP:0004828	1	8	2	2525	6.33E-03	157.81	*ERBB3*
Recurrent infections due to aspiration	HP:0004891	1	8	2	2525	6.33E-03	157.81	*SPTBN4*
Deep longitudinal plantar crease	HP:0004681	1	8	2	2525	6.33E-03	157.81	*ERCC6*
Widely spaced primary teeth	HP:0006313	1	8	2	2525	6.33E-03	157.81	*ERCC6*
Normal pressure hydrocephalus	HP:0002343	1	8	2	2525	6.33E-03	157.81	*ERCC6*
Peripheral dysmyelination	HP:0003469	1	8	2	2525	6.33E-03	157.81	*ERCC6*
Lisch nodules	HP:0009737	1	8	2	2525	6.33E-03	157.81	*MSH6*
Square pelvis bone	HP:0003278	1	8	2	2525	6.33E-03	157.81	*ERCC6*
Aplasia/Hypoplasia of the cerebellum	HP:0007360	2	8	40	2525	6.45E-03	15.78	*ERCC6, PDZD7*
Choreoathetosis	HP:0001266	2	8	42	2525	7.10E-03	15.03	*ERCC6, SPTBN4*
Breast carcinoma	HP:0003002	2	8	43	2525	7.43E-03	14.68	*RAD50, MSH6*
Laryngeal carcinoma	HP:0012118	1	8	3	2525	9.48E-03	105.21	*MSH6*

Table S29 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
Hypoplasia of the primary teeth	HP:0006334	1	8	3	2525	9.48E-03	105.21	*ERCC6*
Axillary freckling	HP:0000997	1	8	3	2525	9.48E-03	105.21	*MSH6*
Slender nose	HP:0000417	1	8	3	2525	9.48E-03	105.21	*ERCC6*
Adult onset sensorineural hearing impairment	HP:0008615	1	8	3	2525	9.48E-03	105.21	*ERCC6*
Subdural hemorrhage	HP:0100309	1	8	3	2525	9.48E-03	105.21	*ERCC6*
Increased cellular sensitivity to UV light	HP:0003224	1	8	3	2525	9.48E-03	105.21	*ERCC6*
Cerebellar calcifications	HP:0007352	1	8	3	2525	9.48E-03	105.21	*ERCC6*
Salivary gland neoplasm	HP:0100684	1	8	3	2525	9.48E-03	105.21	*MSH6*
Ivory epiphyses of the phalanges of the hand	HP:0010234	1	8	3	2525	9.48E-03	105.21	*ERCC6*
Functional intestinal obstruction	HP:0005249	1	8	4	2525	1.26E-02	78.91	*ERBB3*
Peripheral axonal neuropathy	HP:0003477	2	8	60	2525	1.42E-02	10.52	*ERCC6, SPTBN4*
Anxiety	HP:0000739	3	8	177	2525	1.46E-02	5.35	*MSH6, PDZD7, ARVCF*
Hallucinations	HP:0000738	2	8	62	2525	1.51E-02	10.18	*MSH6, PDZD7*
Constipation	HP:0002019	3	8	180	2525	1.53E-02	5.26	*ERBB3, MSH6, ARVCF*
Splenomegaly	HP:0001744	3	8	181	2525	1.55E-02	5.23	*ERBB3, ERCC6, ARVCF*
Occipital myelomeningocele	HP:0007271	1	8	5	2525	1.58E-02	63.13	*ARVCF*
Decreased lacrimation	HP:0000633	1	8	5	2525	1.58E-02	63.13	*ERCC6*
Chromosomal breakage induced by ionizing radiation	HP:0010997	1	8	5	2525	1.58E-02	63.13	*RAD50*
T-cell lymphoma	HP:0012190	1	8	5	2525	1.58E-02	63.13	*MSH6*
Olivopontocerebellar atrophy	HP:0002542	1	8	5	2525	1.58E-02	63.13	*ERCC6*
Loss of facial adipose tissue	HP:0000292	1	8	5	2525	1.58E-02	63.13	*ERCC6*
Abnormal dental enamel morphology	HP:0000682	2	8	64	2525	1.61E-02	9.86	*PDZD7, ARVCF*
Incomplete penetrance	HP:0003829	2	8	66	2525	1.70E-02	9.56	*ERBB3, MSH6*
Gastrointestinal hemorrhage	HP:0002239	2	8	68	2525	1.80E-02	9.28	*MSH6, ARVCF*
Depression	HP:0000716	3	8	193	2525	1.85E-02	4.91	*MSH6, PDZD7, ARVCF*
Abnormality of the tonsils	HP:0100765	1	8	6	2525	1.89E-02	52.60	*ARVCF*
Erythroid hyperplasia	HP:0012132	1	8	6	2525	1.89E-02	52.60	*ERBB3*
Malignant genitourinary tract tumor	HP:0006758	1	8	6	2525	1.89E-02	52.60	*MSH6*
Hematological neoplasm	HP:0004377	1	8	6	2525	1.89E-02	52.60	*MSH6*
Hirsutism	HP:0001007	2	8	70	2525	1.90E-02	9.02	*ERCC6, TAF6*
Aganglionic megacolon	HP:0002251	2	8	70	2525	1.90E-02	9.02	*ERBB3, ARVCF*
Arthrogryposis multiplex congenita	HP:0002804	2	8	72	2525	2.01E-02	8.77	*ERBB3, ERCC6*

Table S29 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
Renal hypoplasia	HP:0000089	2	8	72	2525	2.01E-02	8.77	*ERCC6, ARVCF*
Hypoplastic pelvis	HP:0008839	1	8	7	2525	2.20E-02	45.09	*ERCC6*
Tricuspid atresia	HP:0011662	1	8	7	2525	2.20E-02	45.09	*ARVCF*
Abnormality of the ear	HP:0000598	1	8	7	2525	2.20E-02	45.09	*ERCC6*
Dry hair	HP:0011359	1	8	7	2525	2.20E-02	45.09	*ERCC6*
Type 2 muscle fiber atrophy	HP:0003554	1	8	7	2525	2.20E-02	45.09	*SPTBN4*
Contractures involving the joints of the feet	HP:0008366	1	8	7	2525	2.20E-02	45.09	*ERCC6*
Defective DNA repair after ultraviolet radiation damage	HP:0003079	1	8	7	2525	2.20E-02	45.09	*ERCC6*
Abnormality of the inner ear	HP:0000359	1	8	7	2525	2.20E-02	45.09	*PDZD7*
Multiple suture craniosynostosis	HP:0011324	1	8	7	2525	2.20E-02	45.09	*ARVCF*
Long philtrum	HP:0000343	3	8	210	2525	2.32E-02	4.51	*ERCC6, ARVCF, TAF6*
Intestinal malrotation	HP:0002566	2	8	79	2525	2.39E-02	7.99	*ERBB3, ARVCF*
Sensorineural hearing impairment	HP:0000407	4	8	394	2525	2.43E-02	3.20	*ERBB3, ERCC6, PDZD7, SPTBN4*
Type 1 muscle fiber atrophy	HP:0011807	1	8	8	2525	2.51E-02	39.45	*SPTBN4*
Urinary tract neoplasm	HP:0010786	1	8	8	2525	2.51E-02	39.45	*MSH6*
Hypertensive crisis	HP:0100735	1	8	8	2525	2.51E-02	39.45	*ARVCF*
Bird-like facies	HP:0000320	1	8	8	2525	2.51E-02	39.45	*RAD50*
Cardiac diverticulum	HP:0100571	1	8	8	2525	2.51E-02	39.45	*MSH6*
Abnormal auditory evoked potentials	HP:0006958	1	8	8	2525	2.51E-02	39.45	*ERCC6*
Short stature	HP:0004322	5	8	625	2525	2.60E-02	2.53	*ERBB3, ERCC6, RAD50, ARVCF, TAF6*
Abnormal aortic arch morphology	HP:0012303	1	8	9	2525	2.82E-02	35.07	*ARVCF*
Renal neoplasm	HP:0009726	1	8	9	2525	2.82E-02	35.07	*MSH6*
Delayed eruption of primary teeth	HP:0000680	1	8	9	2525	2.82E-02	35.07	*ERCC6*
Cerebral white matter atrophy	HP:0012762	1	8	9	2525	2.82E-02	35.07	*ERCC6*
Neoplasm of the skeletal system	HP:0010622	1	8	9	2525	2.82E-02	35.07	*MSH6*
Abnormality of dental color	HP:0011073	1	8	9	2525	2.82E-02	35.07	*PDZD7*
Nausea and vomiting	HP:0002017	2	8	89	2525	3.00E-02	7.09	*ERBB3, MSH6*
Mild postnatal growth retardation	HP:0001530	1	8	10	2525	3.13E-02	31.56	*ERCC6*
Basal ganglia calcification	HP:0002135	1	8	10	2525	3.13E-02	31.56	*ERCC6*
Abnormality of peripheral nerve conduction	HP:0003134	1	8	10	2525	3.13E-02	31.56	*ERCC6*
Lung adenocarcinoma	HP:0030078	1	8	10	2525	3.13E-02	31.56	*ERCC6*

Table S29 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
Abnormality of temperature regulation	HP:0004370	1	8	10	2525	3.13E-02	31.56	*ERCC6*
Primary peritoneal carcinoma	HP:0030406	1	8	10	2525	3.13E-02	31.56	*RAD50*
Abnormality of the uterus	HP:0000130	1	8	10	2525	3.13E-02	31.56	*ARVCF*
Scarring	HP:0100699	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Retinal hemorrhage	HP:0000573	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Alveolar cell carcinoma	HP:0006519	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Hypoplasia of the iris	HP:0007676	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Progeroid facial appearance	HP:0005328	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Agnosia	HP:0010524	1	8	11	2525	3.44E-02	28.69	*MSH6*
Poikiloderma	HP:0001029	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Increased blood urea nitrogen	HP:0003138	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Hypoplasia of teeth	HP:0000685	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Severe failure to thrive	HP:0001525	1	8	11	2525	3.44E-02	28.69	*ERCC6*
Tetany	HP:0001281	1	8	11	2525	3.44E-02	28.69	*ARVCF*
Somatic mutation	HP:0001428	2	8	97	2525	3.52E-02	6.51	*ERCC6, MSH6*
Gastroesophageal reflux	HP:0002020	3	8	247	2525	3.57E-02	3.83	*ERCC6, ARVCF, SPTBN4*
Anodontia	HP:0000674	1	8	12	2525	3.74E-02	26.30	*ERCC6*
Platybasia	HP:0002691	1	8	12	2525	3.74E-02	26.30	*ARVCF*
Seborrheic dermatitis	HP:0001051	1	8	12	2525	3.74E-02	26.30	*ARVCF*
Reduced subcutaneous adipose tissue	HP:0003758	1	8	12	2525	3.74E-02	26.30	*ERCC6*
Small earlobe	HP:0000385	1	8	12	2525	3.74E-02	26.30	*ARVCF*
Impaired T cell function	HP:0005435	1	8	12	2525	3.74E-02	26.30	*ARVCF*
Benign neoplasm of the central nervous system	HP:0100835	1	8	12	2525	3.74E-02	26.30	*MSH6*
Premature coronary artery atherosclerosis	HP:0005181	1	8	12	2525	3.74E-02	26.30	*ERCC6*
Neoplasm of the rectum	HP:0100743	1	8	12	2525	3.74E-02	26.30	*MSH6*
Neonatal hypotonia	HP:0001319	2	8	103	2525	3.93E-02	6.13	*TAF6, SPTBN4*
Hypermetropia	HP:0000540	2	8	104	2525	4.00E-02	6.07	*ERCC6, RAD50*
Subcortical cerebral atrophy	HP:0012157	1	8	13	2525	4.05E-02	24.28	*PDZD7*
Hyperthyroidism	HP:0000836	1	8	13	2525	4.05E-02	24.28	*ARVCF*
Corneal ulceration	HP:0012804	1	8	13	2525	4.05E-02	24.28	*ERCC6*
Abnormality of the pharynx	HP:0000600	1	8	13	2525	4.05E-02	24.28	*ARVCF*
Adenoma sebaceum	HP:0009720	1	8	13	2525	4.05E-02	24.28	*MSH6*

Table S29 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p*-value**	**Relative enrichment**	**Genes**
Entropion	HP:0000621	1	8	13	2525	4.05E-02	24.28	*ERCC6*
Pancreatic adenocarcinoma	HP:0006725	1	8	13	2525	4.05E-02	24.28	*MSH6*
Turricephaly	HP:0000262	1	8	14	2525	4.36E-02	22.54	*ARVCF*
Elevated circulating follicle stimulating hormone level	HP:0008232	1	8	14	2525	4.36E-02	22.54	*ERCC6*
Astrocytosis	HP:0002446	1	8	14	2525	4.36E-02	22.54	*ERCC6*
Corneal neovascularization	HP:0011496	1	8	14	2525	4.36E-02	22.54	*ARVCF*
Hemianopia	HP:0012377	1	8	14	2525	4.36E-02	22.54	*PDZD7*
Pituitary adenoma	HP:0002893	1	8	14	2525	4.36E-02	22.54	*MSH6*
Astrocytoma	HP:0009592	1	8	14	2525	4.36E-02	22.54	*MSH6*
Narrow nose	HP:0000460	1	8	14	2525	4.36E-02	22.54	*ERCC6*
Glioblastoma multiforme	HP:0012174	1	8	14	2525	4.36E-02	22.54	*MSH6*
Death in infancy	HP:0001522	2	8	109	2525	4.36E-02	5.79	*ERCC6, MSH6*
Hypertonia	HP:0001276	2	8	111	2525	4.51E-02	5.69	*ERCC6, MSH6*
Abnormal fallopian tube morphology	HP:0011027	1	8	15	2525	4.66E-02	21.04	*RAD50*
Hypoparathyroidism	HP:0000829	1	8	15	2525	4.66E-02	21.04	*ARVCF*
Hypoplastic iliac wing	HP:0002866	1	8	15	2525	4.66E-02	21.04	*ERCC6*
Truncus arteriosus	HP:0001660	1	8	15	2525	4.66E-02	21.04	*ARVCF*
Thickened calvaria	HP:0002684	1	8	16	2525	4.97E-02	19.73	*ERCC6*
Urinary retention	HP:0000016	1	8	16	2525	4.97E-02	19.73	*ERCC6*
Dermal atrophy	HP:0004334	1	8	16	2525	4.97E-02	19.73	*ERCC6*
Uveitis	HP:0000554	1	8	16	2525	4.97E-02	19.73	*ERCC6*
Neoplasm of the liver	HP:0002896	1	8	16	2525	4.97E-02	19.73	*MSH6*

Table S30 - Phenotypes from Mouse Genome Informatics database associated with the 28 candidate genes for the I4 subgroup.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
increased cellular sensitivity to gamma-irradiation	MP:0002007	3	18	37	6530	1.29E-04	29.41	*ERCC6, COPS6, RAD50*
abnormal myelination	MP:0000920	3	18	74	6530	1.01E-03	14.71	*ERBB3, ERCC6, SPTBN4*
increased tumor latency	MP:0009828	2	18	20	6530	1.32E-03	36.28	*ERBB3, RAD50*
prenatal lethality, complete penetrance	MP:0011091	4	18	216	6530	2.47E-03	6.72	*MCM7, ERBB3, COPS6, RAD50*
abnormal pontine flexure morphology	MP:0000844	1	18	1	6530	2.76E-03	362.78	*ERBB3*
absent Schwann cell precursors	MP:0001109	1	18	1	6530	2.76E-03	362.78	*ERBB3*
thin atrioventricular cushion	MP:0000300	1	18	1	6530	2.76E-03	362.78	*ERBB3*
abnormal mandibular nerve morphology	MP:0009800	1	18	1	6530	2.76E-03	362.78	*ERBB3*
increased skin pigmentation	MP:0030926	1	18	1	6530	2.76E-03	362.78	*RAD50*
decreased IgE level	MP:0002492	2	18	33	6530	3.60E-03	21.99	*RAD50, MSH6*
abnormal neocortex morphology	MP:0008547	2	18	34	6530	3.82E-03	21.34	*ERBB3, ERCC6*
abnormal distortion product otoacoustic emission	MP:0004736	2	18	36	6530	4.28E-03	20.15	*ERCC6, PDZD7*
photophobia	MP:0013787	1	18	2	6530	5.51E-03	181.39	*ERCC6*
decreased cochlear microphonics	MP:0004414	1	18	2	6530	5.51E-03	181.39	*PDZD7*
dysmetria	MP:0003314	1	18	2	6530	5.51E-03	181.39	*SPTBN4*
absent adrenergic chromaffin cells	MP:0000645	1	18	2	6530	5.51E-03	181.39	*ERBB3*
absent cerebellum vermis	MP:0000865	1	18	2	6530	5.51E-03	181.39	*ERBB3*
abnormal prevertebral ganglion morphology	MP:0008316	1	18	2	6530	5.51E-03	181.39	*ERBB3*
long limbs	MP:0000548	1	18	2	6530	5.51E-03	181.39	*ERCC6*
abnormal testis size	MP:0004849	1	18	2	6530	5.51E-03	181.39	*RAD50*
skin photosensitivity	MP:0001202	1	18	2	6530	5.51E-03	181.39	*ERCC6*
absence of NMDA-mediated synaptic currents	MP:0001901	1	18	2	6530	5.51E-03	181.39	*GRIK4*
abnormal retinal inner nuclear layer morphology	MP:0003733	2	18	41	6530	5.52E-03	17.70	*ERCC6, DSCAML1*
increased or absent threshold for auditory brainstem response	MP:0011967	3	18	137	6530	5.85E-03	7.94	*ERCC6, PDZD7, SPTBN4*
decreased apoptosis	MP:0006043	2	18	43	6530	6.06E-03	16.87	*RAD50, ENDOG*
abnormal testis morphology	MP:0001146	3	18	140	6530	6.22E-03	7.77	*RAD50, ENDOG, EFCAB5*
abnormal gait	MP:0001406	4	18	299	6530	7.92E-03	4.85	*ERCC6, RAD50, ARVCF, SPTBN4*
decreased incidence of tumors by ionizing radiation induction	MP:0004503	1	18	3	6530	8.25E-03	120.93	*COPS6*
abnormal PNS glial cell morphology	MP:0001105	1	18	3	6530	8.25E-03	120.93	*ERBB3*
abnormal cerebellar plate morphology	MP:0000856	1	18	3	6530	8.25E-03	120.93	*ERBB3*

Table S30 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
abnormal sympathetic system morphology	MP:0001007	1	18	3	6530	8.25E-03	120.93	*ERBB3*
absent Schwann cells	MP:0001108	1	18	3	6530	8.25E-03	120.93	*ERBB3*
abnormal airway resistance	MP:0002328	1	18	3	6530	8.25E-03	120.93	*RAD50*
hippocampus atrophy	MP:0030003	1	18	3	6530	8.25E-03	120.93	*ERCC6*
absent bone marrow cell	MP:0000175	1	18	3	6530	8.25E-03	120.93	*RAD50*
increased hepatoblastoma incidence	MP:0010054	1	18	3	6530	8.25E-03	120.93	*RAD50*
neuron degeneration	MP:0003224	2	18	54	6530	9.44E-03	13.44	*ERBB3, ERCC6*
Deiters cell degeneration	MP:0004628	1	18	4	6530	1.10E-02	90.69	*ERCC6*
increased late pro-B cell number	MP:0003134	1	18	4	6530	1.10E-02	90.69	*RAD50*
increased basal cell carcinoma incidence	MP:0004208	1	18	4	6530	1.10E-02	90.69	*MSH6*
abnormal brain ventricular system morphology	MP:0002200	1	18	4	6530	1.10E-02	90.69	*ERBB3*
abnormal ophthalmic nerve morphology	MP:0009798	1	18	4	6530	1.10E-02	90.69	*ERBB3*
atrioventricular valve regurgitation	MP:0006046	1	18	4	6530	1.10E-02	90.69	*ERBB3*
abnormal paravertebral ganglion morphology	MP:0008317	1	18	4	6530	1.10E-02	90.69	*ERBB3*
abnormal auditory cortex morphology	MP:0004631	1	18	4	6530	1.10E-02	90.69	*SPTBN4*
increased hemolymphoid system tumor incidence	MP:0010296	1	18	5	6530	1.37E-02	72.56	*MSH6*
increased tail pigmentation	MP:0011276	1	18	5	6530	1.37E-02	72.56	*RAD50*
pillar cell degeneration	MP:0004586	1	18	5	6530	1.37E-02	72.56	*ERCC6*
abnormal base-excision repair	MP:0009796	1	18	5	6530	1.37E-02	72.56	*ERCC6*
increased uterus tumor incidence	MP:0009222	1	18	5	6530	1.37E-02	72.56	*MSH6*
deafness	MP:0001967	2	18	66	6530	1.39E-02	10.99	*PDZD7, SPTBN4*
abnormal excitatory postsynaptic currents	MP:0002910	2	18	67	6530	1.43E-02	10.83	*GRIK4, DSCAML1*
abnormal retina morphology	MP:0001325	3	18	199	6530	1.62E-02	5.47	*ERCC6, ITGB5, DSCAML1*
priapism	MP:0003415	1	18	6	6530	1.64E-02	60.46	*SPTBN4*
embryonic lethality before implantation	MP:0006204	1	18	6	6530	1.64E-02	60.46	*TAF6*
abnormal dorsal striatum morphology	MP:0004102	1	18	6	6530	1.64E-02	60.46	*ERCC6*
increased amacrine cell number	MP:0008105	1	18	6	6530	1.64E-02	60.46	*DSCAML1*
abnormal outer hair cell kinocilium location or orientation	MP:0030961	1	18	6	6530	1.64E-02	60.46	*PDZD7*
photosensitivity	MP:0001999	1	18	6	6530	1.64E-02	60.46	*ERCC6*
degeneration of organ of Corti supporting cells	MP:0004465	1	18	7	6530	1.91E-02	51.83	*ERCC6*
abnormal mating frequency	MP:0001377	1	18	7	6530	1.91E-02	51.83	*SPTBN4*
abnormal cranial flexure morphology	MP:0004203	1	18	7	6530	1.91E-02	51.83	*ERBB3*

Table S30 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
increased mammary gland apoptosis	MP:0014126	1	18	7	6530	1.91E-02	51.83	*ERBB3*
abnormal pro-B cell morphology	MP:0005432	1	18	7	6530	1.91E-02	51.83	*RAD50*
small nodose ganglion	MP:0001088	1	18	8	6530	2.19E-02	45.35	*ERBB3*
abnormal mismatch repair	MP:0009797	1	18	8	6530	2.19E-02	45.35	*MSH6*
abnormal glial cell morphology	MP:0003634	1	18	8	6530	2.19E-02	45.35	*ERBB3*
increased foot pad pigmentation	MP:0000575	1	18	8	6530	2.19E-02	45.35	*RAD50*
hypermyelination	MP:0010050	1	18	8	6530	2.19E-02	45.35	*ERBB3*
decreased noradrenaline level	MP:0012574	1	18	8	6530	2.19E-02	45.35	*ERBB3*
small petrosal ganglion	MP:0001085	1	18	8	6530	2.19E-02	45.35	*ERBB3*
impaired hearing	MP:0006325	2	18	85	6530	2.24E-02	8.54	*ERCC6, SPTBN4*
hypolactation	MP:0013716	1	18	9	6530	2.46E-02	40.31	*ERBB3*
decreased cerebellar granule cell precursor proliferation	MP:0013551	1	18	9	6530	2.46E-02	40.31	*ERCC6*
lipodystrophy	MP:0011174	1	18	9	6530	2.46E-02	40.31	*ERCC6*
reduced linear vestibular evoked potential	MP:0004814	1	18	9	6530	2.46E-02	40.31	*SPTBN4*
decreased Schwann cell number	MP:0001107	1	18	10	6530	2.72E-02	36.28	*ERBB3*
increased interleukin-5 secretion	MP:0008702	1	18	10	6530	2.72E-02	36.28	*RAD50*
abnormal mammary gland epithelium morphology	MP:0009504	1	18	10	6530	2.72E-02	36.28	*ERBB3*
abnormal enteric ganglia morphology	MP:0001045	1	18	10	6530	2.72E-02	36.28	*ERBB3*
abnormal sympathetic ganglion morphology	MP:0001008	1	18	10	6530	2.72E-02	36.28	*ERBB3*
impaired balance	MP:0001525	2	18	96	6530	2.81E-02	7.56	*ERCC6, SPTBN4*
decreased myelin sheath thickness	MP:0011731	1	18	11	6530	2.99E-02	32.98	*ERBB3*
poor circulation	MP:0001633	1	18	11	6530	2.99E-02	32.98	*ERBB3*
increased interleukin-13 secretion	MP:0008672	1	18	11	6530	2.99E-02	32.98	*RAD50*
abnormal postural reflex	MP:0002980	1	18	11	6530	2.99E-02	32.98	*GRIK4*
axonal dystrophy	MP:0003225	1	18	11	6530	2.99E-02	32.98	*SPTBN4*
abnormal fat cell differentiation	MP:0011168	1	18	11	6530	2.99E-02	32.98	*LPIN3*
ovary atrophy	MP:0004833	1	18	11	6530	2.99E-02	32.98	*RAD50*
abnormal superior cervical ganglion morphology	MP:0001011	1	18	12	6530	3.26E-02	30.23	*ERBB3*
increased incidence of tumors by ionizing radiation induction	MP:0004500	1	18	12	6530	3.26E-02	30.23	*COPS6*
abnormal stomach epithelium morphology	MP:0000471	1	18	12	6530	3.26E-02	30.23	*ERBB3*
abnormal oligodendrocyte physiology	MP:0008917	1	18	12	6530	3.26E-02	30.23	*ERCC6*

Table S30 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
absent linear vestibular evoked potential	MP:0004813	1	18	12	6530	3.26E-02	30.23	*SPTBN4*
abnormal ventral spinal root morphology	MP:0003993	1	18	13	6530	3.53E-02	27.91	*ERBB3*
thin cerebellar granule layer	MP:0006099	1	18	13	6530	3.53E-02	27.91	*ERCC6*
decreased outer hair cell stereocilia number	MP:0004529	1	18	13	6530	3.53E-02	27.91	*PDZD7*
decreased airway responsiveness	MP:0002335	1	18	13	6530	3.53E-02	27.91	*RAD50*
abnormal embryonic neuroepithelial layer differentiation	MP:0000786	1	18	13	6530	3.53E-02	27.91	*ERBB3*
abnormal adipose tissue development	MP:0011167	1	18	13	6530	3.53E-02	27.91	*LPIN3*
abnormal cranial nerve morphology	MP:0001056	1	18	13	6530	3.53E-02	27.91	*ERBB3*
decreased interleukin-5 secretion	MP:0008703	1	18	14	6530	3.79E-02	25.91	*RAD50*
pancytopenia	MP:0005152	1	18	14	6530	3.79E-02	25.91	*RAD50*
abnormal mammary gland duct morphology	MP:0009503	1	18	14	6530	3.79E-02	25.91	*ERBB3*
embryonic lethality between somite formation and embryo turning, incomplete penetrance	MP:0011107	1	18	14	6530	3.79E-02	25.91	*COPS6*
abnormal abdominal fat pad morphology	MP:0000010	1	18	14	6530	3.79E-02	25.91	*ERCC6*
abnormal sperm number	MP:0002673	1	18	14	6530	3.79E-02	25.91	*RAD50*
abnormal cochlear outer hair cell physiology	MP:0004434	1	18	14	6530	3.79E-02	25.91	*PDZD7*
increased lymphoma incidence	MP:0012431	2	18	114	6530	3.85E-02	6.36	*RAD50, MSH6*
abnormal mammary gland alveolus morphology	MP:0009506	1	18	15	6530	4.06E-02	24.19	*ERBB3*
abnormal cardiac epithelial to mesenchymal transition	MP:0008825	1	18	15	6530	4.06E-02	24.19	*ERBB3*
paresis	MP:0000754	1	18	15	6530	4.06E-02	24.19	*SPTBN4*
abnormal hair cell mechanoelectric transduction	MP:0004431	1	18	15	6530	4.06E-02	24.19	*PDZD7*
decreased body weight	MP:0001262	6	18	978	6530	4.14E-02	2.23	*LPIN3, MCM7, ERBB3, ERCC6, RAD50, SPTBN4*
hypopigmentation	MP:0005408	1	18	16	6530	4.33E-02	22.67	*ERBB3*
abnormal cerebral aqueduct morphology	MP:0005537	1	18	17	6530	4.59E-02	21.34	*ERBB3*
abnormal neural crest cell migration	MP:0002950	1	18	17	6530	4.59E-02	21.34	*ERBB3*
intestine polyps	MP:0008011	1	18	17	6530	4.59E-02	21.34	*ERBB3*
abnormal retinal ganglion layer morphology	MP:0005241	1	18	17	6530	4.59E-02	21.34	*DSCAML1*
abnormal limb posture	MP:0004263	1	18	18	6530	4.85E-02	20.15	*ERCC6*
decreased paired-pulse facilitation	MP:0002920	1	18	18	6530	4.85E-02	20.15	*GRIK4*
premature aging	MP:0003786	1	18	18	6530	4.85E-02	20.15	*ERCC6*

Table S30 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p*-value	Relative enrichment	Genes
increased liver glycogen level	MP:0010400	1	18	18	6530	4.85E-02	20.15	*ERCC6*
increased liver adenoma incidence	MP:0003324	1	18	18	6530	4.85E-02	20.15	*RAD50*
decreased interleukin-13 secretion	MP:0008673	1	18	18	6530	4.85E-02	20.15	*RAD50*
increased intestinal adenocarcinoma incidence	MP:0002957	1	18	18	6530	4.85E-02	20.15	*MSH6*
abnormal retinal inner plexiform layer morphology	MP:0003734	1	18	18	6530	4.85E-02	20.15	*DSCAML1*
abnormal retinal rod bipolar cell morphology	MP:0006074	1	18	18	6530	4.85E-02	20.15	*DSCAML1*
decreased hepatocyte proliferation	MP:0004001	1	18	18	6530	4.85E-02	20.15	*RAD50*
abnormal axon fasciculation	MP:0009450	1	18	18	6530	4.85E-02	20.15	*ERBB3*
abnormal atrioventricular valve morphology	MP:0002745	1	18	18	6530	4.85E-02	20.15	*ERBB3*
increased cellular sensitivity to ultraviolet irradiation	MP:0008410	1	18	18	6530	4.85E-02	20.15	*ERCC6*

Table S31 - Biological processes from Gene Ontology database associated with the 16 candidate genes for the I1 subgroup.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
extracellular matrix organization	GO:0030198	4	15	127	8747	5.11E-05	18.37	*COL17A1, COL11A1, COL14A1, PAPLN*
U2 snRNA 3'-end processing	GO:0034474	1	15	1	8747	1.71E-03	583.13	*INTS1*
nucleoside diphosphate biosynthetic process	GO:0009133	1	15	1	8747	1.71E-03	583.13	*ENTPD8*
nucleoside monophosphate biosynthetic process	GO:0009124	1	15	1	8747	1.71E-03	583.13	*ENTPD8*
positive regulation of cyclic nucleotide-gated ion channel activity	GO:1902161	1	15	1	8747	1.71E-03	583.13	*CFTR*
positive regulation of cellular response to insulin stimulus	GO:1900078	1	15	1	8747	1.71E-03	583.13	*MYO1C*
mRNA transport	GO:0051028	2	15	47	8747	2.84E-03	24.81	*NUP210, MYO1C*
collagen fibril organization	GO:0030199	2	15	51	8747	3.33E-03	22.87	*COL11A1, COL14A1*
positive regulation of protein lipidation	GO:1903061	1	15	2	8747	3.43E-03	291.57	*RAB3GAP2*
retinal cell apoptotic process	GO:1990009	1	15	2	8747	3.43E-03	291.57	*PDE6B*
establishment of protein localization to endoplasmic reticulum membrane	GO:0097051	1	15	2	8747	3.43E-03	291.57	*RAB3GAP2*
detection of light stimulus	GO:0009583	1	15	2	8747	3.43E-03	291.57	*PDE6B*
positive regulation of voltage-gated chloride channel activity	GO:1902943	1	15	2	8747	3.43E-03	291.57	*CFTR*
positive regulation of endoplasmic reticulum tubular network organization	GO:1903373	1	15	3	8747	5.14E-03	194.38	*RAB3GAP2*
positive regulation of signal transduction by p53 class mediator	GO:1901798	1	15	3	8747	5.14E-03	194.38	*CHD5*
positive regulation of cell migration by vascular endothelial growth factor signaling pathway	GO:0038089	1	15	3	8747	5.14E-03	194.38	*MYO1C*
transepithelial water transport	GO:0035377	1	15	3	8747	5.14E-03	194.38	*CFTR*
positive regulation of enamel mineralization	GO:0070175	1	15	3	8747	5.14E-03	194.38	*CFTR*
tendon development	GO:0035989	1	15	3	8747	5.14E-03	194.38	*COL11A1*
stress-activated protein kinase signaling cascade	GO:0031098	1	15	3	8747	5.14E-03	194.38	*CCDC88C*
membrane hyperpolarization	GO:0060081	1	15	4	8747	6.84E-03	145.78	*CFTR*
regulation of bicellular tight junction assembly	GO:2000810	1	15	4	8747	6.84E-03	145.78	*MYO1C*
positive regulation of vascular endothelial growth factor signaling pathway	GO:1900748	1	15	4	8747	6.84E-03	145.78	*MYO1C*
spermatogenesis, exchange of chromosomal proteins	GO:0035093	1	15	4	8747	6.84E-03	145.78	*CHD5*
apical constriction	GO:0003383	1	15	4	8747	6.84E-03	145.78	*CCDC88C*
histone H3-K27 trimethylation	GO:0098532	1	15	4	8747	6.84E-03	145.78	*CHD5*

Table S31 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p-value*	Relative enrichment	Genes
intracellular pH elevation	GO:0051454	1	15	4	8747	6.84E-03	145.78	*CFTR*
hemidesmosome assembly	GO:0031581	1	15	5	8747	8.55E-03	116.63	*COL17A1*
high-density lipoprotein particle assembly	GO:0034380	1	15	6	8747	1.02E-02	97.19	*PRKACG*
phototransduction, visible light	GO:0007603	1	15	6	8747	1.02E-02	97.19	*PDE6B*
nucleoside diphosphate catabolic process	GO:0009134	1	15	6	8747	1.02E-02	97.19	*ENTPD8*
proteoglycan metabolic process	GO:0006029	1	15	7	8747	1.19E-02	83.30	*COL11A1*
multicellular organismal water homeostasis	GO:0050891	1	15	7	8747	1.19E-02	83.30	*CFTR*
amelogenesis	GO:0097186	1	15	7	8747	1.19E-02	83.30	*CFTR*
cerebral cortex neuron differentiation	GO:0021895	1	15	7	8747	1.19E-02	83.30	*CHD5*
cellular response to forskolin	GO:1904322	1	15	8	8747	1.36E-02	72.89	*CFTR*
renal water homeostasis	GO:0003091	1	15	9	8747	1.53E-02	64.79	*PRKACG*
regulation of sensory perception of pain	GO:0051930	1	15	10	8747	1.70E-02	58.31	*ZFHX2*
entrainment of circadian clock by photoperiod	GO:0043153	1	15	10	8747	1.70E-02	58.31	*PDE6B*
protein kinase A signaling	GO:0010737	1	15	10	8747	1.70E-02	58.31	*PRKACG*
visual perception	GO:0007601	2	15	122	8747	1.80E-02	9.56	*COL11A1*, *PDE6B*
snRNA processing	GO:0016180	1	15	11	8747	1.87E-02	53.01	*INTS1*
cartilage condensation	GO:0001502	1	15	11	8747	1.87E-02	53.01	*COL11A1*
positive regulation of exocytosis	GO:0045921	1	15	12	8747	2.04E-02	48.59	*CFTR*
cytoskeleton-dependent intracellular transport	GO:0030705	1	15	12	8747	2.04E-02	48.59	*CCDC88C*
positive regulation of protein targeting to membrane	GO:0090314	1	15	13	8747	2.21E-02	44.86	*MYO1C*
sperm capacitation	GO:0048240	1	15	13	8747	2.21E-02	44.86	*CFTR*
ventricular cardiac muscle tissue morphogenesis	GO:0055010	1	15	13	8747	2.21E-02	44.86	*COL11A1*
positive regulation of transcription by RNA polymerase III	GO:0045945	1	15	14	8747	2.38E-02	41.65	*MYO1C*
positive regulation of autophagosome assembly	GO:2000786	1	15	14	8747	2.38E-02	41.65	*RAB3GAP2*
non-canonical Wnt signaling pathway	GO:0035567	1	15	14	8747	2.38E-02	41.65	*CCDC88C*
detection of mechanical stimulus involved in sensory perception of sound	GO:0050910	1	15	14	8747	2.38E-02	41.65	*COL11A1*
histone H4 acetylation	GO:0043967	1	15	16	8747	2.71E-02	36.45	*CHD5*
vesicle transport along actin filament	GO:0030050	1	15	16	8747	2.71E-02	36.45	*MYO1C*
chondrocyte development	GO:0002063	1	15	16	8747	2.71E-02	36.45	*COL11A1*
vesicle docking involved in exocytosis	GO:0006904	1	15	16	8747	2.71E-02	36.45	*CFTR*
cholesterol transport	GO:0030301	1	15	17	8747	2.88E-02	34.30	*CFTR*

Table S31 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
cholesterol biosynthetic process	GO:0006695	1	15	17	8747	2.88E-02	34.30	*CFTR*
positive regulation of insulin secretion involved in cellular response to glucose stimulus	GO:0035774	1	15	17	8747	2.88E-02	34.30	*CFTR*
regulation of neuron differentiation	GO:0045664	1	15	19	8747	3.21E-02	30.69	*ZFHX2*
adult behavior	GO:0030534	1	15	20	8747	3.38E-02	29.16	*ZFHX2*
endodermal cell differentiation	GO:0035987	1	15	22	8747	3.71E-02	26.51	*COL11A1*
protein targeting to membrane	GO:0006612	1	15	22	8747	3.71E-02	26.51	*MYO1C*
bicarbonate transport	GO:0015701	1	15	22	8747	3.71E-02	26.51	*CFTR*
protein destabilization	GO:0031648	1	15	27	8747	4.53E-02	21.60	*CCDC88C*
protein targeting	GO:0006605	1	15	27	8747	4.53E-02	21.60	*MYO1C*
regulation of protein phosphorylation	GO:0001932	1	15	28	8747	4.70E-02	20.83	*CCDC88C*
embryonic skeletal system morphogenesis	GO:0048704	1	15	28	8747	4.70E-02	20.83	*COL11A1*
skeletal system morphogenesis	GO:0048705	1	15	29	8747	4.86E-02	20.11	*COL11A1*

Table S32 - Phenotypes from Human Phenotype Ontology database associated with the 16 candidate genes for the I1 subgroup.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
Retinal coloboma	HP:0000480	2	10	19	2679	2.07E-03	28.20	*INTS1, RAB3GAP2*
Meningeal calcification	HP:0100250	1	10	1	2679	3.73E-03	267.90	*COL11A1*
White papule	HP:0031289	1	10	1	2679	3.73E-03	267.90	*CFTR*
Small distal femoral epiphysis	HP:0012283	1	10	1	2679	3.73E-03	267.90	*COL11A1*
Widely patent sagittal suture	HP:0005476	1	10	1	2679	3.73E-03	267.90	*COL11A1*
Abnormal phalangeal joint morphology of the hand	HP:0006261	1	10	1	2679	3.73E-03	267.90	*CFTR*
Posterior vertebral hypoplasia	HP:0008451	1	10	1	2679	3.73E-03	267.90	*COL11A1*
Irregular proximal tibial epiphyses	HP:0006456	1	10	1	2679	3.73E-03	267.90	*COL11A1*
Wide tufts of distal phalanges	HP:0006095	1	10	1	2679	3.73E-03	267.90	*COL11A1*
Widely patent coronal suture	HP:0005442	1	10	1	2679	3.73E-03	267.90	*COL11A1*
Slender ulna	HP:0003992	1	10	1	2679	3.73E-03	267.90	*RAB3GAP2*
Overlapping toe	HP:0001845	2	10	28	2679	4.50E-03	19.14	*INTS1, RAB3GAP2*
Palmoplantar keratoderma	HP:0000982	3	10	100	2679	5.00E-03	8.04	*COL17A1, COL14A1, CFTR*
Absent vas deferens	HP:0012873	1	10	2	2679	7.45E-03	133.95	*CFTR*
Thin clavicles	HP:0006645	1	10	2	2679	7.45E-03	133.95	*COL11A1*
Adermatoglyphia	HP:0007455	1	10	2	2679	7.45E-03	133.95	*COL17A1*
Undetectable visual evoked potentials	HP:0007965	1	10	2	2679	7.45E-03	133.95	*RAB3GAP2*
Excessive skin wrinkling on dorsum of hands and fingers	HP:0007407	1	10	2	2679	7.45E-03	133.95	*CFTR*
Broad fingertip	HP:0011300	1	10	2	2679	7.45E-03	133.95	*RAB3GAP2*
Irregular distal femoral epiphysis	HP:0006407	1	10	2	2679	7.45E-03	133.95	*COL11A1*
Prominent nipples	HP:0004405	1	10	2	2679	7.45E-03	133.95	*RAB3GAP2*
Skin detachment	HP:0032156	1	10	2	2679	7.45E-03	133.95	*COL17A1*
Decreased forced expiratory flow 25-75%	HP:0032359	1	10	2	2679	7.45E-03	133.95	*CFTR*
Abnormal vitreous humor morphology	HP:0004327	1	10	2	2679	7.45E-03	133.95	*COL11A1*
Atrophic, patchy alopecia	HP:0004529	1	10	2	2679	7.45E-03	133.95	*COL17A1*
Verrucous papule	HP:0012500	1	10	2	2679	7.45E-03	133.95	*COL14A1*
Subepithelial corneal opacities	HP:0008039	1	10	2	2679	7.45E-03	133.95	*COL17A1*
Irregular astigmatism	HP:0031792	1	10	2	2679	7.45E-03	133.95	*COL17A1*
Hypohidrosis	HP:0000966	2	10	42	2679	9.97E-03	12.76	*COL11A1, ZFHX2*
Pancreatic pseudocyst	HP:0005206	1	10	3	2679	1.12E-02	89.30	*CFTR*
Localized skin lesion	HP:0011355	1	10	3	2679	1.12E-02	89.30	*COL17A1*
Broad ischia	HP:0100865	1	10	3	2679	1.12E-02	89.30	*COL11A1*

Table S32 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
Asymmetry of the ears	HP:0010722	1	10	3	2679	1.12E-02	89.30	*RAB3GAP2*
Orthokeratotic hyperkeratosis	HP:0025080	1	10	3	2679	1.12E-02	89.30	*CFTR*
Dumbbell-shaped long bone	HP:0000947	1	10	3	2679	1.12E-02	89.30	*COL11A1*
Adenocarcinoma of the small intestine	HP:0040274	1	10	3	2679	1.12E-02	89.30	*COL14A1*
Meconium ileus	HP:0004401	1	10	3	2679	1.12E-02	89.30	*CFTR*
Prominent antitragus	HP:0008593	1	10	3	2679	1.12E-02	89.30	*RAB3GAP2*
Small proximal tibial epiphyses	HP:0012284	1	10	3	2679	1.12E-02	89.30	*COL11A1*
Pain	HP:0012531	2	10	45	2679	1.14E-02	11.91	*COL17A1, COL14A1*
Abnormal distal phalanx morphology of finger	HP:0009832	1	10	4	2679	1.49E-02	66.98	*RAB3GAP2*
Posterior rib cupping	HP:0000922	1	10	4	2679	1.49E-02	66.98	*COL11A1*
Short nose	HP:0003196	3	10	155	2679	1.69E-02	5.19	*INTS1, COL11A1, RAB3GAP2*
Developmental cataract	HP:0000519	2	10	56	2679	1.73E-02	9.57	*COL11A1, RAB3GAP2*
Wide nasal bridge	HP:0000431	4	10	297	2679	1.81E-02	3.61	*INTS1, COL11A1, RAB3GAP2, PDE6B*
Acute infectious pneumonia	HP:0011949	1	10	5	2679	1.85E-02	53.58	*CFTR*
Pontocerebellar atrophy	HP:0006879	1	10	5	2679	1.85E-02	53.58	*CCDC88C*
Recurrent Haemophilus influenzae infections	HP:0005376	1	10	5	2679	1.85E-02	53.58	*CFTR*
Palmar hyperhidrosis	HP:0006089	1	10	5	2679	1.85E-02	53.58	*COL17A1*
Palmar pruritus	HP:0031248	1	10	5	2679	1.85E-02	53.58	*CFTR*
Hyperkeratotic papule	HP:0045059	1	10	5	2679	1.85E-02	53.58	*COL14A1*
Elevated sweat chloride	HP:0012236	1	10	5	2679	1.85E-02	53.58	*CFTR*
Splanchnic vein thrombosis	HP:0030247	1	10	5	2679	1.85E-02	53.58	*CFTR*
Inferior vermis hypoplasia	HP:0007068	1	10	5	2679	1.85E-02	53.58	*INTS1*
Abnormal light- and dark-adapted electroretinogram	HP:0008323	1	10	5	2679	1.85E-02	53.58	*PDE6B*
Halitosis	HP:0100812	1	10	5	2679	1.85E-02	53.58	*CFTR*
Short palm	HP:0004279	2	10	58	2679	1.85E-02	9.24	*COL11A1, RAB3GAP2*
Nail dystrophy	HP:0008404	2	10	63	2679	2.17E-02	8.50	*COL17A1, COL14A1*
Calcification of falx cerebri	HP:0005462	1	10	6	2679	2.22E-02	44.65	*COL11A1*
Absent frontal sinuses	HP:0002688	1	10	6	2679	2.22E-02	44.65	*COL11A1*
Lens luxation	HP:0012019	1	10	6	2679	2.22E-02	44.65	*COL11A1*
Ileus	HP:0002595	1	10	6	2679	2.22E-02	44.65	*CFTR*
Painless fractures due to injury	HP:0002661	1	10	6	2679	2.22E-02	44.65	*ZFHX2*
Acral blistering	HP:0031045	1	10	6	2679	2.22E-02	44.65	*COL17A1*

Table S32 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
Recurrent pancreatitis	HP:0100027	1	10	6	2679	2.22E-02	44.65	*CFTR*
Broad long bones	HP:0005622	1	10	6	2679	2.22E-02	44.65	*COL11A1*
Hypoplastic frontal sinuses	HP:0002738	1	10	6	2679	2.22E-02	44.65	*COL11A1*
Macrothrombocytopenia	HP:0040185	1	10	6	2679	2.22E-02	44.65	*PRKACG*
Hypoplasia of the maxilla	HP:0000327	2	10	64	2679	2.23E-02	8.37	*COL11A1, RAB3GAP2*
Reduced forced expiratory volume in one second	HP:0032342	1	10	7	2679	2.59E-02	38.27	*CFTR*
Recurrent bronchopulmonary infections	HP:0006538	1	10	7	2679	2.59E-02	38.27	*CFTR*
Pancreatic calcification	HP:0005213	1	10	7	2679	2.59E-02	38.27	*CFTR*
Decreased pulmonary function	HP:0005952	1	10	7	2679	2.59E-02	38.27	*CFTR*
Hypoplastic dermoepidermal hemidesmosomes	HP:0020117	1	10	7	2679	2.59E-02	38.27	*COL17A1*
Brain neoplasm	HP:0030692	1	10	7	2679	2.59E-02	38.27	*COL14A1*
Adenocarcinoma of the colon	HP:0040276	1	10	7	2679	2.59E-02	38.27	*COL14A1*
Macrodontia of permanent maxillary central incisor	HP:0000675	1	10	7	2679	2.59E-02	38.27	*COL11A1*
Abnormal diaphysis morphology	HP:0000940	1	10	7	2679	2.59E-02	38.27	*COL11A1*
Reduced FEV1/FVC ratio	HP:0030877	1	10	7	2679	2.59E-02	38.27	*CFTR*
Dental enamel pits	HP:0009722	1	10	7	2679	2.59E-02	38.27	*COL17A1*
Scarring alopecia of scalp	HP:0004552	1	10	7	2679	2.59E-02	38.27	*COL17A1*
Cor pulmonale	HP:0001648	1	10	8	2679	2.95E-02	33.49	*CFTR*
Hearing abnormality	HP:0000364	1	10	8	2679	2.95E-02	33.49	*COL11A1*
Thick upper lip vermilion	HP:0000215	1	10	8	2679	2.95E-02	33.49	*COL11A1*
Mitten deformity	HP:0004057	1	10	8	2679	2.95E-02	33.49	*COL17A1*
Anterior rib cupping	HP:0000907	1	10	9	2679	3.31E-02	29.77	*COL11A1*
Esophageal neoplasm	HP:0100751	1	10	9	2679	3.31E-02	29.77	*COL14A1*
Scanning speech	HP:0002168	1	10	9	2679	3.31E-02	29.77	*CCDC88C*
Pierre-Robin sequence	HP:0000201	1	10	9	2679	3.31E-02	29.77	*COL11A1*
Hodgkin lymphoma	HP:0012189	1	10	9	2679	3.31E-02	29.77	*COL14A1*
Palmoplantar hyperhidrosis	HP:0007410	1	10	9	2679	3.31E-02	29.77	*CFTR*
Abnormal thrombosis	HP:0001977	1	10	9	2679	3.31E-02	29.77	*CFTR*
Decreased corneal reflex	HP:0008000	1	10	9	2679	3.31E-02	29.77	*ZFHX2*
Plantar hyperkeratosis	HP:0007556	1	10	9	2679	3.31E-02	29.77	*COL17A1*
Limb joint contracture	HP:0003121	1	10	9	2679	3.31E-02	29.77	*COL17A1*
Upper airway obstruction	HP:0002781	1	10	9	2679	3.31E-02	29.77	*COL11A1*

Table S32 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
Ocular pain	HP:0200026	1	10	9	2679	3.31E-02	29.77	*COL17A1*
Long clavicles	HP:0000890	1	10	10	2679	3.68E-02	26.79	*COL11A1*
Abnormality of dental color	HP:0011073	1	10	10	2679	3.68E-02	26.79	*COL17A1*
Orthokeratosis	HP:0040162	1	10	10	2679	3.68E-02	26.79	*COL14A1*
Obstructive azoospermia	HP:0011962	1	10	10	2679	3.68E-02	26.79	*CFTR*
Neoplasm of the skeletal system	HP:0010622	1	10	10	2679	3.68E-02	26.79	*COL14A1*
Spondyloepiphyseal dysplasia	HP:0002655	1	10	10	2679	3.68E-02	26.79	*COL11A1*
Abnormal renal morphology	HP:0012210	1	10	10	2679	3.68E-02	26.79	*CFTR*
Narrow greater sciatic notch	HP:0003375	1	10	10	2679	3.68E-02	26.79	*COL11A1*
Cataract	HP:0000518	4	10	372	2679	3.86E-02	2.88	*INTS1, COL11A1, RAB3GAP2, PDE6B*
Congenital stationary night blindness	HP:0007642	1	10	11	2679	4.04E-02	24.35	*PDE6B*
Hypoplastic ischia	HP:0003175	1	10	11	2679	4.04E-02	24.35	*COL11A1*
Vertical supranuclear gaze palsy	HP:0000511	1	10	11	2679	4.04E-02	24.35	*CCDC88C*
Reduced amplitude of dark-adapted bright flash electroretinogram a-wave	HP:0030483	1	10	11	2679	4.04E-02	24.35	*PDE6B*
Congenital stationary night blindness with abnormal fundus	HP:0030639	1	10	11	2679	4.04E-02	24.35	*PDE6B*
Congenital stationary night blindness with normal fundus	HP:0030638	1	10	11	2679	4.04E-02	24.35	*PDE6B*
Recurrent corneal erosions	HP:0000495	1	10	11	2679	4.04E-02	24.35	*COL17A1*
Radial bowing	HP:0002986	1	10	11	2679	4.04E-02	24.35	*COL11A1*
Arthropathy	HP:0003040	1	10	11	2679	4.04E-02	24.35	*COL11A1*
Compensatory head posture	HP:0031705	1	10	11	2679	4.04E-02	24.35	*PDE6B*
Oral mucosal blisters	HP:0200097	1	10	11	2679	4.04E-02	24.35	*COL17A1*
Ulnar bowing	HP:0003031	1	10	12	2679	4.40E-02	22.33	*COL11A1*
Corneal scarring	HP:0000559	1	10	12	2679	4.40E-02	22.33	*COL17A1*
Talipes valgus	HP:0004684	1	10	12	2679	4.40E-02	22.33	*RAB3GAP2*
Pain insensitivity	HP:0007021	1	10	12	2679	4.40E-02	22.33	*ZFHX2*
Glossoptosis	HP:0000162	1	10	12	2679	4.40E-02	22.33	*COL11A1*
Short femur	HP:0003097	1	10	12	2679	4.40E-02	22.33	*COL11A1*
Stomach cancer	HP:0012126	1	10	12	2679	4.40E-02	22.33	*COL14A1*
Electronegative electroretinogram	HP:0007984	1	10	13	2679	4.76E-02	20.61	*PDE6B*
Horizontal eyebrow	HP:0011228	1	10	13	2679	4.76E-02	20.61	*INTS1*

Table S32 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
Fibular hypoplasia	HP:0003038	1	10	13	2679	4.76E-02	20.61	*COL11A1*
Epiphora	HP:0009926	1	10	13	2679	4.76E-02	20.61	*COL17A1*
Hypergranulosis	HP:0025114	1	10	13	2679	4.76E-02	20.61	*COL14A1*

Table S33 - Phenotypes from Mouse Genome Informatics database associated with the 16 candidate genes for the I1 subgroup.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
intestinal mucus accumulation	MP:0014231	1	12	1	6896	1.74E-03	574.67	*CFTR*
increased respiratory epithelial sodium ion transmembrane transport	MP:0014183	1	12	1	6896	1.74E-03	574.67	*CFTR*
dilated crypts of Lieberkuhn	MP:0014230	1	12	1	6896	1.74E-03	574.67	*CFTR*
pancreatic acinus dilation	MP:0014232	1	12	1	6896	1.74E-03	574.67	*CFTR*
bile duct epithelium hyperplasia	MP:0014233	1	12	1	6896	1.74E-03	574.67	*CFTR*
abnormal submucosal gland morphology	MP:0014033	1	12	1	6896	1.74E-03	574.67	*CFTR*
lacrimal gland atrophy	MP:0013455	1	12	1	6896	1.74E-03	574.67	*CFTR*
meconium ileus	MP:0014036	1	12	1	6896	1.74E-03	574.67	*CFTR*
abnormal Brunner's gland morphology	MP:0012518	1	12	1	6896	1.74E-03	574.67	*CFTR*
dilated Brunner's glands	MP:0012519	1	12	1	6896	1.74E-03	574.67	*CFTR*
decreased respiratory epithelial chloride transmembrane transport	MP:0014028	1	12	1	6896	1.74E-03	574.67	*CFTR*
abnormal trachea gland morphology	MP:0013494	1	12	1	6896	1.74E-03	574.67	*CFTR*
abnormal vital capacity	MP:0002309	1	12	1	6896	1.74E-03	574.67	*CFTR*
thick retinal outer nuclear layer	MP:0008517	1	12	1	6896	1.74E-03	574.67	*PDE6B*
serous retinal detachment	MP:0020442	1	12	1	6896	1.74E-03	574.67	*PDE6B*
ileum hypertrophy	MP:0009484	1	12	1	6896	1.74E-03	574.67	*CFTR*
abnormal optic disk morphology	MP:0008259	2	12	44	6896	2.52E-03	26.12	*RAB3GAP2, PDE6B*
coiled cecum	MP:0009478	1	12	2	6896	3.48E-03	287.33	*CFTR*
abnormal ileal goblet cell morphology	MP:0013797	1	12	2	6896	3.48E-03	287.33	*CFTR*
decreased respiratory epithelial sodium ion transmembrane transport	MP:0014182	1	12	2	6896	3.48E-03	287.33	*CFTR*
gallbladder inflammation	MP:0003251	1	12	2	6896	3.48E-03	287.33	*CFTR*
abnormal uterine cervix morphology	MP:0001135	1	12	2	6896	3.48E-03	287.33	*CFTR*
abnormal small intestine goblet cell morphology	MP:0013792	1	12	2	6896	3.48E-03	287.33	*CFTR*
decreased intestinal epithelial sodium ion transmembrane transport	MP:0014206	1	12	3	6896	5.21E-03	191.56	*CFTR*
pulmonary epithelial necrosis	MP:0010858	1	12	3	6896	5.21E-03	191.56	*CFTR*
abnormal joint mobility	MP:0008069	1	12	3	6896	5.21E-03	191.56	*COL11A1*
decreased intestinal epithelial chloride transmembrane transport	MP:0014208	1	12	3	6896	5.21E-03	191.56	*CFTR*

Table S33 - Continuation.

Description	Annotation	Genes found	Input size	Term genes	Universe	*p-value*	Relative enrichment	Genes
decreased small intestinal villus number	MP:0014079	1	12	3	6896	5.21E-03	191.56	*CFTR*
pancreatic acinar cell atrophy	MP:0009150	1	12	3	6896	5.21E-03	191.56	*CFTR*
abnormal alveolar macrophage physiology	MP:0014226	1	12	3	6896	5.21E-03	191.56	*CFTR*
abnormal brain ependyma motile cilium location or orientation	MP:0030963	1	12	3	6896	5.21E-03	191.56	*CCDC88C*
acute pancreas inflammation	MP:0003340	1	12	3	6896	5.21E-03	191.56	*CFTR*
pulmonary interstitial fibrosis	MP:0003426	1	12	3	6896	5.21E-03	191.56	*CFTR*
abnormal vestibular hair cell physiology	MP:0004438	1	12	3	6896	5.21E-03	191.56	*MYO1C*
abnormal nasal mucosa morphology	MP:0002238	1	12	3	6896	5.21E-03	191.56	*CFTR*
abnormal paranasal sinus morphology	MP:0002240	1	12	3	6896	5.21E-03	191.56	*CFTR*
abnormal CD4-positive, alpha beta T cell number	MP:0008073	1	12	4	6896	6.94E-03	143.67	*NUP210*
absent retinal rod cells	MP:0008454	1	12	4	6896	6.94E-03	143.67	*PDE6B*
abnormal costal cartilage morphology	MP:0006432	1	12	4	6896	6.94E-03	143.67	*COL11A1*
increased respiratory mucosa goblet cell number	MP:0010861	1	12	4	6896	6.94E-03	143.67	*CFTR*
abnormal visual contrast sensitivity	MP:0011831	1	12	4	6896	6.94E-03	143.67	*PDE6B*
abnormal mucociliary clearance	MP:0001947	1	12	4	6896	6.94E-03	143.67	*CFTR*
dilated gallbladder	MP:0009343	1	12	4	6896	6.94E-03	143.67	*CFTR*
decreased neutrophil cell number	MP:0000222	2	12	80	6896	8.14E-03	14.37	*CFTR, PDE6B*
abnormal alpha-beta T cell morphology	MP:0012762	1	12	5	6896	8.67E-03	114.93	*NUP210*
podocyte hypertrophy	MP:0011871	1	12	5	6896	8.67E-03	114.93	*COL17A1*
abnormal small intestinal crypt cell physiology	MP:0010156	1	12	5	6896	8.67E-03	114.93	*CFTR*
absent cochlear inner hair cells	MP:0004397	1	12	5	6896	8.67E-03	114.93	*COL11A1*
abnormal epidermal-dermal junction morphology	MP:0011159	1	12	5	6896	8.67E-03	114.93	*COL17A1*
abnormal knee joint morphology	MP:0030837	1	12	5	6896	8.67E-03	114.93	*COL11A1*
micromelia	MP:0008736	1	12	5	6896	8.67E-03	114.93	*COL11A1*
abnormal gland morphology	MP:0002163	1	12	5	6896	8.67E-03	114.93	*CFTR*
small cecum	MP:0009477	1	12	5	6896	8.67E-03	114.93	*CFTR*
abnormal hyaline cartilage morphology	MP:0006429	1	12	5	6896	8.67E-03	114.93	*COL11A1*
disorganized retinal outer plexiform layer	MP:0008520	1	12	6	6896	1.04E-02	95.78	*PDE6B*
thin uterus	MP:0009081	1	12	6	6896	1.04E-02	95.78	*CFTR*
decreased tendon stiffness	MP:0003098	1	12	6	6896	1.04E-02	95.78	*COL14A1*
abnormal lacrimal gland morphology	MP:0001346	1	12	6	6896	1.04E-02	95.78	*CFTR*
small Meckel's cartilage	MP:0030026	1	12	6	6896	1.04E-02	95.78	*COL11A1*

Table S33 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
embryonic lethality before implantation	MP:0006204	1	12	6	6896	1.04E-02	95.78	*INTS1*
pancreatic acinar cell zymogen granule accumulation	MP:0009161	1	12	6	6896	1.04E-02	95.78	*CFTR*
abnormal gallbladder physiology	MP:0005085	1	12	7	6896	1.21E-02	82.10	*CFTR*
decreased autoantibody level	MP:0003726	1	12	7	6896	1.21E-02	82.10	*COL17A1*
increased mast cell degranulation	MP:0008764	1	12	8	6896	1.38E-02	71.83	*COL17A1*
abnormal alveolar macrophage morphology	MP:0008245	1	12	8	6896	1.38E-02	71.83	*CFTR*
abnormal bronchus morphology	MP:0002264	1	12	8	6896	1.38E-02	71.83	*CFTR*
decreased uterus weight	MP:0004905	1	12	9	6896	1.56E-02	63.85	*CFTR*
abnormal colon goblet cell morphology	MP:0013795	1	12	9	6896	1.56E-02	63.85	*CFTR*
abnormal brain ependyma morphology	MP:0010152	1	12	9	6896	1.56E-02	63.85	*CCDC88C*
abnormal extracellular matrix morphology	MP:0013258	1	12	9	6896	1.56E-02	63.85	*PDE6B*
crypts of Lieberkuhn abscesses	MP:0000491	1	12	9	6896	1.56E-02	63.85	*CFTR*
impaired sperm capacitation	MP:0003666	1	12	9	6896	1.56E-02	63.85	*CFTR*
buphthalmos	MP:0009274	1	12	9	6896	1.56E-02	63.85	*PDE6B*
abnormal circadian behavior phase	MP:0020473	1	12	9	6896	1.56E-02	63.85	*PDE6B*
abnormal ion homeostasis	MP:0001765	1	12	9	6896	1.56E-02	63.85	*CFTR*
decreased retinal rod cell number	MP:0008453	1	12	9	6896	1.56E-02	63.85	*PDE6B*
dermal-epidermal separation	MP:0011160	1	12	9	6896	1.56E-02	63.85	*COL17A1*
abnormal locomotor circadian rhythm	MP:0020477	1	12	10	6896	1.73E-02	57.47	*PDE6B*
peritoneal inflammation	MP:0003303	1	12	10	6896	1.73E-02	57.47	*CFTR*
dilated pancreatic duct	MP:0009144	1	12	10	6896	1.73E-02	57.47	*CFTR*
abnormal submandibular gland morphology	MP:0003793	1	12	10	6896	1.73E-02	57.47	*CFTR*
absent cochlear outer hair cells	MP:0004403	1	12	10	6896	1.73E-02	57.47	*COL11A1*
abnormal vas deferens morphology	MP:0002769	1	12	10	6896	1.73E-02	57.47	*CFTR*
increased CD4-positive, CD25-positive, alpha-beta regulatory T cell number	MP:0010168	1	12	10	6896	1.73E-02	57.47	*NUP210*
abnormal circardian behavior entrainment	MP:0020476	1	12	11	6896	1.90E-02	52.24	*PDE6B*
ocular hypertension	MP:0005258	1	12	11	6896	1.90E-02	52.24	*PDE6B*
abnormal fluid regulation	MP:0001784	1	12	11	6896	1.90E-02	52.24	*CFTR*
increased pruritus	MP:0010072	1	12	11	6896	1.90E-02	52.24	*COL17A1*
arrest of spermiogenesis	MP:0008279	1	12	11	6896	1.90E-02	52.24	*CHD5*
intestinal obstruction	MP:0003270	1	12	12	6896	2.07E-02	47.89	*CFTR*
retinal detachment	MP:0003099	1	12	12	6896	2.07E-02	47.89	*PDE6B*

Table S33 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
increased diameter of long bones	MP:0008151	1	12	12	6896	2.07E-02	47.89	*COL11A1*
decreased ear pigmentation	MP:0011279	1	12	12	6896	2.07E-02	47.89	*PDE6B*
increased T-helper 1 cell number	MP:0008086	1	12	12	6896	2.07E-02	47.89	*NUP210*
increased effector memory T-helper cell number	MP:0013772	1	12	12	6896	2.07E-02	47.89	*NUP210*
increased lung compliance	MP:0010895	1	12	12	6896	2.07E-02	47.89	*CFTR*
asthenozoospermia	MP:0002675	2	12	134	6896	2.18E-02	8.58	*CHD5, CFTR*
decreased visual acuity	MP:0006149	1	12	13	6896	2.24E-02	44.21	*PDE6B*
abnormal spermiation	MP:0004182	1	12	13	6896	2.24E-02	44.21	*CHD5*
delayed sexual maturation	MP:0001938	1	12	13	6896	2.24E-02	44.21	*CFTR*
respiratory system inflammation	MP:0002405	1	12	13	6896	2.24E-02	44.21	*CFTR*
abnormal jejunum morphology	MP:0004002	1	12	13	6896	2.24E-02	44.21	*CFTR*
absent photoreceptor outer segment	MP:0008585	1	12	14	6896	2.41E-02	41.05	*PDE6B*
enlarged gallbladder	MP:0009342	1	12	14	6896	2.41E-02	41.05	*CFTR*
small pancreas	MP:0004247	1	12	14	6896	2.41E-02	41.05	*CFTR*
excessive scratching	MP:0001412	1	12	15	6896	2.58E-02	38.31	*COL17A1*
abnormal thymus involution	MP:0001824	1	12	15	6896	2.58E-02	38.31	*CFTR*
hippocampal neuron degeneration	MP:0000811	1	12	15	6896	2.58E-02	38.31	*PDE6B*
abnormal salivary gland morphology	MP:0000613	1	12	16	6896	2.75E-02	35.92	*CFTR*
protruding tongue	MP:0009908	1	12	16	6896	2.75E-02	35.92	*COL11A1*
blindness	MP:0002001	1	12	16	6896	2.75E-02	35.92	*PDE6B*
premature hair loss	MP:0005114	1	12	16	6896	2.75E-02	35.92	*PDE6B*
abnormal bile duct morphology	MP:0002928	1	12	16	6896	2.75E-02	35.92	*CFTR*
abnormal tendon morphology	MP:0005503	1	12	16	6896	2.75E-02	35.92	*COL14A1*
decreased corpora lutea number	MP:0002680	1	12	16	6896	2.75E-02	35.92	*CFTR*
abnormal trachea morphology	MP:0002282	1	12	17	6896	2.92E-02	33.80	*CFTR*
abnormal retinal cone cell morphology	MP:0001006	1	12	17	6896	2.92E-02	33.80	*PDE6B*
decreased alpha-beta T cell number	MP:0012765	1	12	17	6896	2.92E-02	33.80	*NUP210*
prolonged estrous cycle	MP:0009006	1	12	17	6896	2.92E-02	33.80	*CFTR*
absent estrous cycle	MP:0009009	1	12	18	6896	3.09E-02	31.93	*CFTR*
macrocytosis	MP:0000248	1	12	18	6896	3.09E-02	31.93	*MYO1C*
thin retinal outer plexiform layer	MP:0008519	1	12	18	6896	3.09E-02	31.93	*PDE6B*
increased susceptibility to infection	MP:0002406	1	12	18	6896	3.09E-02	31.93	*CFTR*

Table S33 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
abnormal Muller cell morphology	MP:0005547	1	12	19	6896	3.26E-02	30.25	*PDE6B*
abnormal cecum morphology	MP:0000494	1	12	20	6896	3.43E-02	28.73	*CFTR*
retinal rod cell degeneration	MP:0008451	1	12	20	6896	3.43E-02	28.73	*PDE6B*
abnormal pancreas physiology	MP:0002693	1	12	20	6896	3.43E-02	28.73	*CFTR*
shortened head	MP:0000435	1	12	21	6896	3.60E-02	27.37	*COL11A1*
abnormal gallbladder morphology	MP:0005084	1	12	21	6896	3.60E-02	27.37	*CFTR*
intestinal ulcer	MP:0000512	1	12	21	6896	3.60E-02	27.37	*CFTR*
abnormal ocular fundus morphology	MP:0002864	1	12	21	6896	3.60E-02	27.37	*PDE6B*
abnormal retinal rod cell morphology	MP:0001005	1	12	21	6896	3.60E-02	27.37	*PDE6B*
decreased ovary weight	MP:0004856	1	12	21	6896	3.60E-02	27.37	*CFTR*
abnormal bone marrow cavity morphology	MP:0000065	1	12	22	6896	3.76E-02	26.12	*COL11A1*
increased central memory CD8 positive, alpha-beta T cell number	MP:0010847	1	12	22	6896	3.76E-02	26.12	*NUP210*
increased fibroblast proliferation	MP:0011703	1	12	22	6896	3.76E-02	26.12	*COL17A1*
decreased ovulation rate	MP:0003355	1	12	22	6896	3.76E-02	26.12	*CFTR*
disorganized photoreceptor outer segment	MP:0008586	1	12	22	6896	3.76E-02	26.12	*PDE6B*
intestinal inflammation	MP:0001858	1	12	22	6896	3.76E-02	26.12	*CFTR*
osteoarthritis	MP:0003560	1	12	22	6896	3.76E-02	26.12	*COL11A1*
abnormal platelet activation	MP:0006298	1	12	22	6896	3.76E-02	26.12	*PDE6B*
weight loss	MP:0001263	2	12	184	6896	3.92E-02	6.25	*CCDC88C, CFTR*
abnormal small intestine crypts of Lieberkuhn morphology	MP:0004841	1	12	23	6896	3.93E-02	24.99	*CFTR*
abnormal long bone diaphysis morphology	MP:0004214	1	12	23	6896	3.93E-02	24.99	*COL11A1*
abnormal tracheal cartilage morphology	MP:0003120	1	12	23	6896	3.93E-02	24.99	*COL11A1*
abnormal basement membrane morphology	MP:0004272	1	12	24	6896	4.10E-02	23.94	*COL17A1*
decreased skin tensile strength	MP:0003089	1	12	24	6896	4.10E-02	23.94	*COL14A1*
decreased retinal cone cell number	MP:0008446	1	12	24	6896	4.10E-02	23.94	*PDE6B*
focal hair loss	MP:0000418	1	12	24	6896	4.10E-02	23.94	*COL17A1*
abnormal intestinal goblet cell morphology	MP:0003449	1	12	24	6896	4.10E-02	23.94	*CFTR*
abnormal lymphocyte morphology	MP:0002619	1	12	25	6896	4.27E-02	22.99	*NUP210*
small uterus	MP:0002637	1	12	25	6896	4.27E-02	22.99	*CFTR*
aphagia	MP:0001438	1	12	25	6896	4.27E-02	22.99	*CFTR*
abnormal phospholipid level	MP:0004777	1	12	25	6896	4.27E-02	22.99	*CFTR*

Table S33 - Continuation.

Description	**Annotation**	**Genes found**	**Input size**	**Term genes**	**Universe**	***p-value***	**Relative enrichment**	**Genes**
photoreceptor outer segment degeneration	MP:0008584	1	12	26	6896	4.44E-02	22.10	*PDE6B*
abnormal renal glomerulus basement membrane morphology	MP:0011348	1	12	26	6896	4.44E-02	22.10	*COL17A1*
transmission ratio distortion	MP:0004179	1	12	26	6896	4.44E-02	22.10	*CFTR*
abnormal endochondral bone ossification	MP:0008272	1	12	26	6896	4.44E-02	22.10	*COL11A1*
abnormal crypts of Lieberkuhn morphology	MP:0000490	1	12	26	6896	4.44E-02	22.10	*CFTR*
reddish skin	MP:0001190	1	12	27	6896	4.60E-02	21.28	*COL17A1*
abnormal auditory brainstem response waveform shape	MP:0011966	1	12	27	6896	4.60E-02	21.28	*COL11A1*
abnormal retinal photoreceptor layer morphology	MP:0003728	1	12	27	6896	4.60E-02	21.28	*PDE6B*
blistering	MP:0001208	1	12	28	6896	4.77E-02	20.52	*COL17A1*
abnormal reproductive system physiology	MP:0001919	1	12	29	6896	4.94E-02	19.82	*COL17A1*
decreased retinal ganglion cell number	MP:0006309	1	12	29	6896	4.94E-02	19.82	*PDE6B*

9.2 SUPPLEMENTARY FIGURES

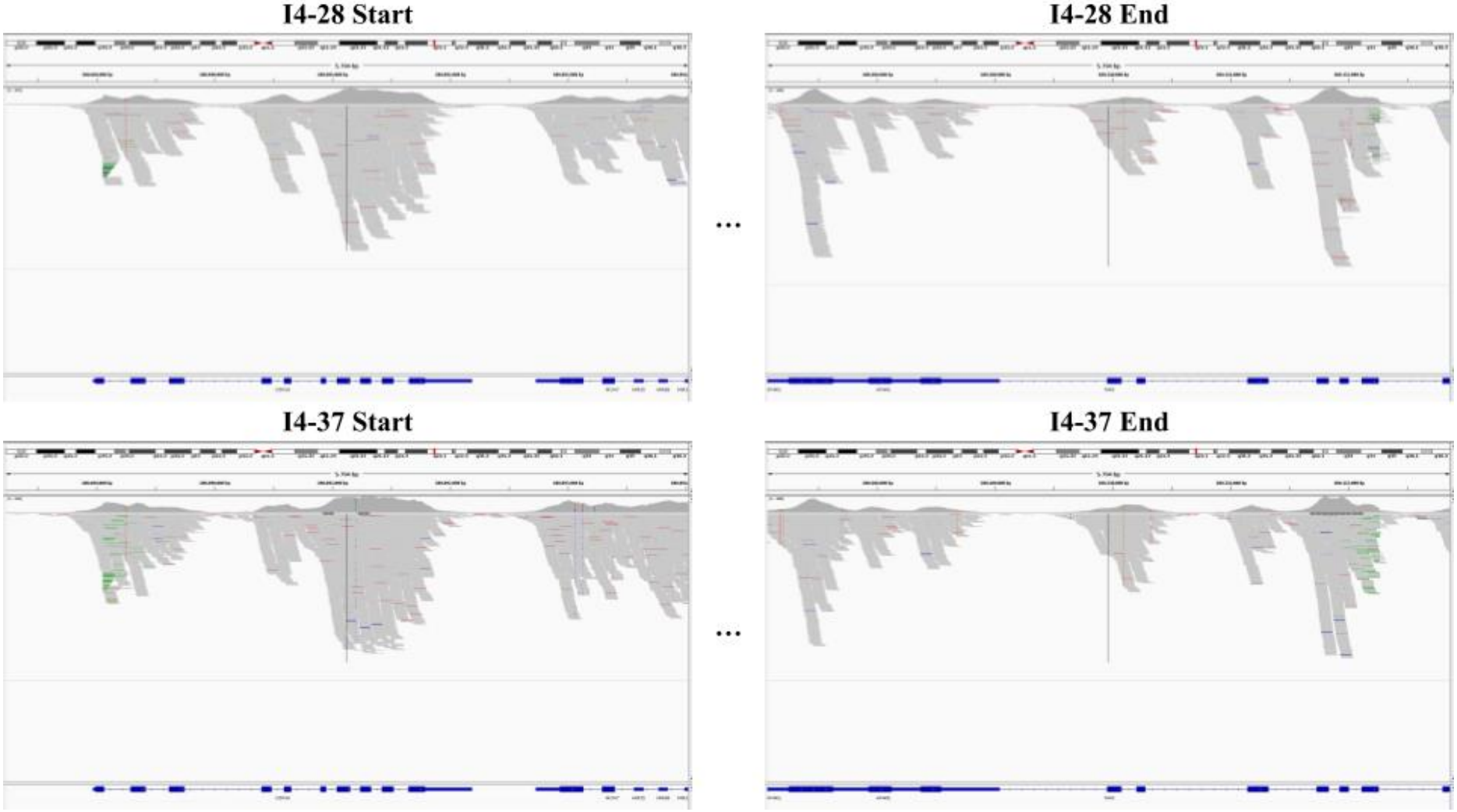

Figure S1 - Validation IGV AP4M1, COPS6, MCM7, MIR106B, MIR25, MIR93 and TAF6 genes.

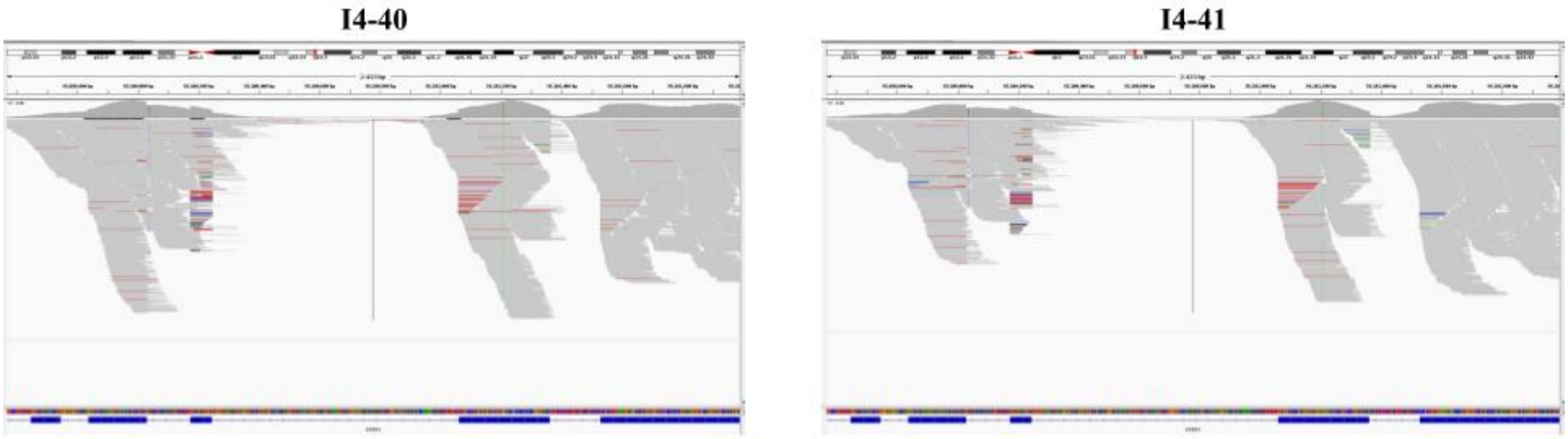

Figure S2 - Validation IGV ERBB3 gene.

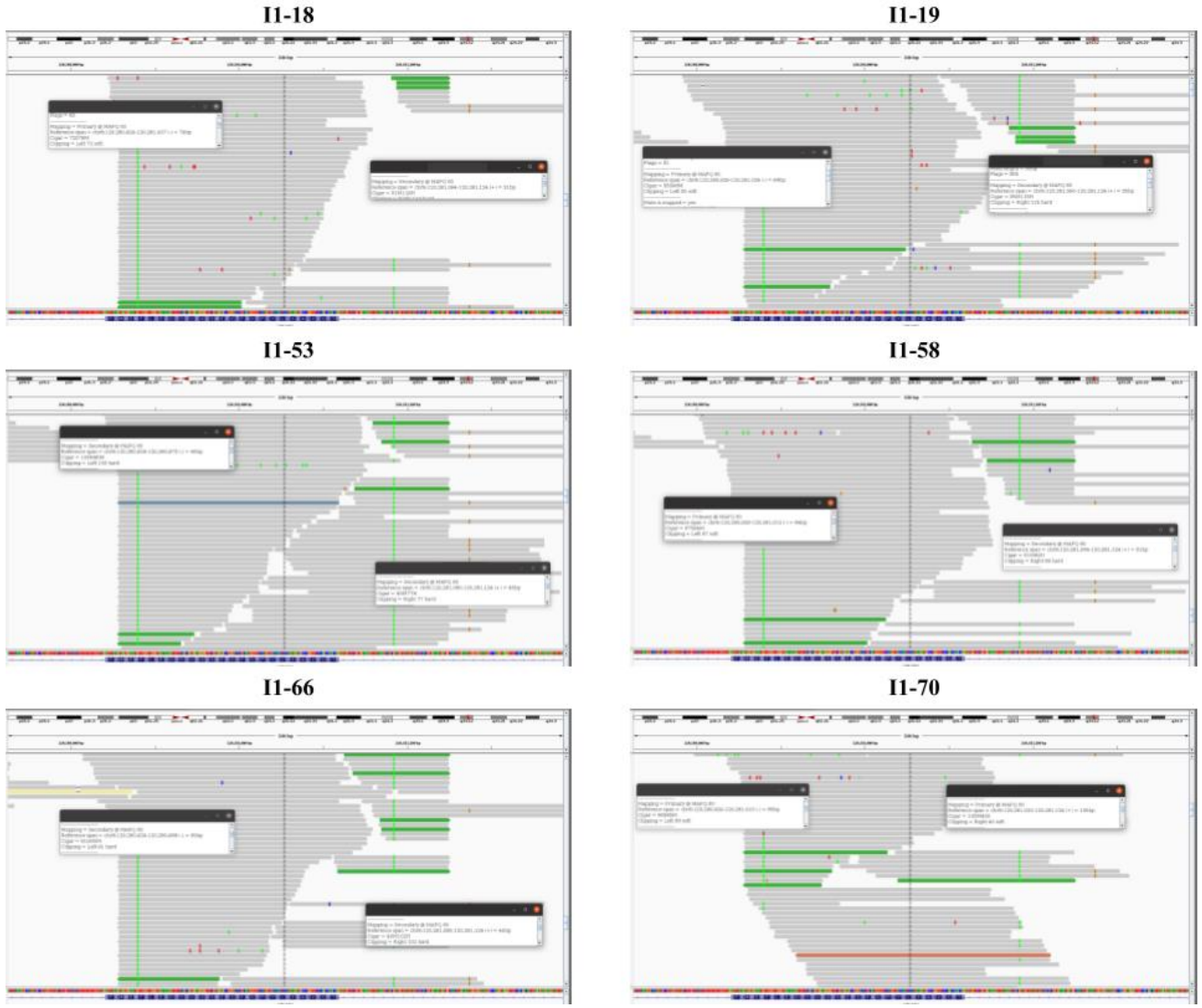

Figure S3 - Validation IGV COL14A1 gene.

Figure S4 - Validation IGV GMCL1 gene.

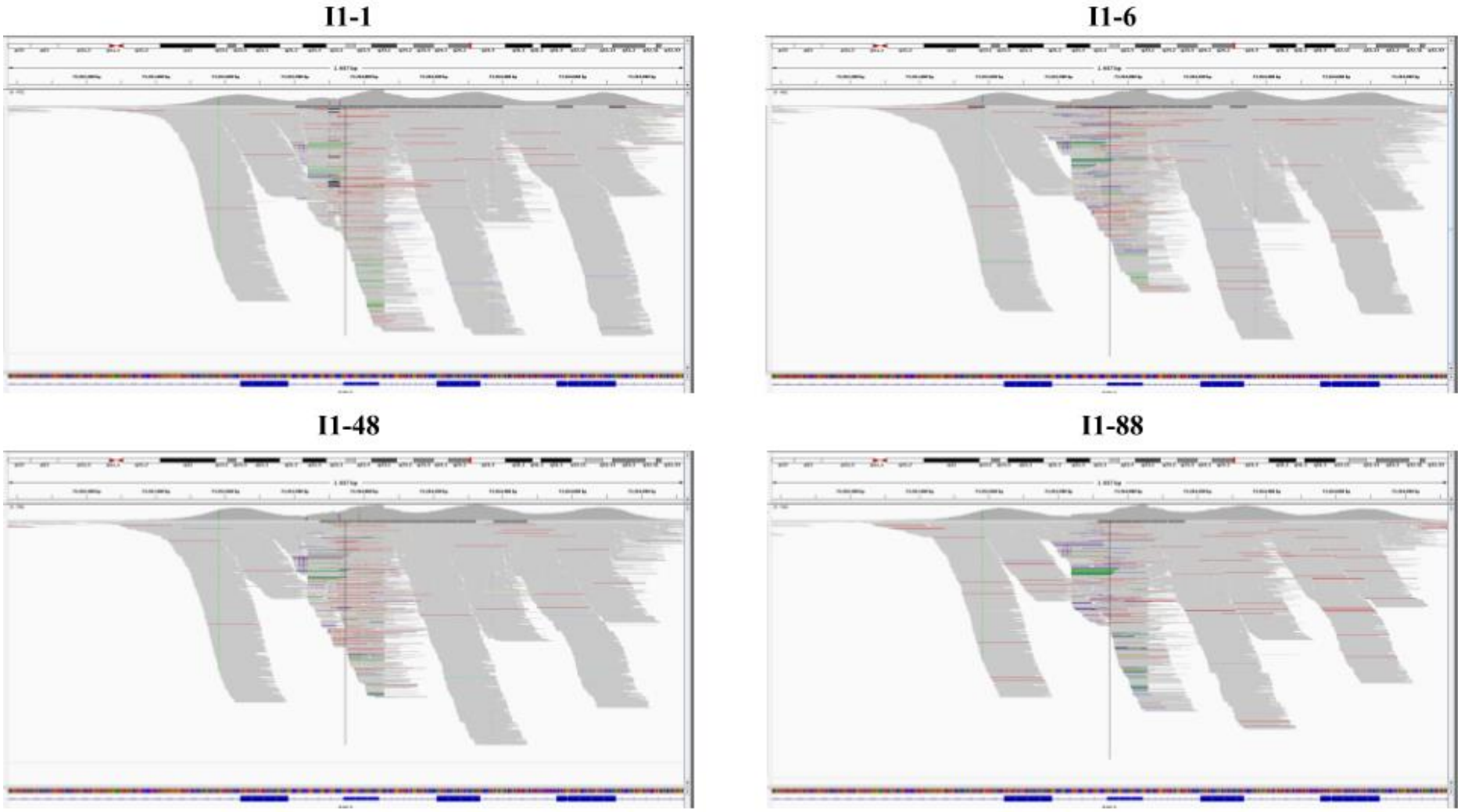

Figure S5 - Validation IGV PAPLN gene.

Figure S6 - Validation IGV PDE6B gene.

Criterios de calidad para obtener el título de Doctor por la Universidad de Granada / Quality criteria to apply for the PhD by the University of Granada

Esta tesis doctoral se ha preparado siguiendo los criterios para obtener el título de Doctor por la Universidad de Granada / *This doctoral thesis has been prepared according to the University of Granada requirements to apply for the PhD degree.*

Publicación de un artículo científico aceptado en revistas relevantes en el campo de conocimiento en el que se enmarca la tesis doctoral, firmado por el doctorando, que incluya parte de los resultados de la tesis / *Publication of a scientific article accepted in relevant journals in the field of knowledge in which the doctoral thesis is framed, signed by the doctoral student, which includes part of the results of the thesis.*

Artículo original que avala la tesis doctoral / *Original article that supports the doctoral thesis:* **Escalera-Balsera, A.**, Parra-Perez, A.M. et al. Rare Deletions or Large Duplications Contribute to Genetic Variation in Patients with Severe Tinnitus and Meniere Disease. *Genes* 2024;15(1),22. doi: 10.3390/genes15010022

Otros artículos originales publicados durante la realización de esta tesis / *Other original articles published during the performing of this doctoral thesis*

- **Escalera-Balsera A**, Roman-Naranjo P, Lopez-Escamez JA. Systematic Review of Sequencing Studies and Gene Expression Profiling in Familial Meniere Disease. *Genes*. 2020;11(12):1414. doi:10.3390/genes11121414
- Schlee W, Schoisswohl S, Staudinger S, (+36 authors), **Escalera-Balsera A** et al. Towards a unification of treatments and interventions for tinnitus patients: The EU research and innovation action UNITI. *Prog Brain Res*. 2021;260:441-451. doi:10.1016/bs.pbr.2020.12.005
- Roman-Naranjo P, Moleon MDC, Aran I, **Escalera-Balsera A** et al. Rare coding variants involving MYO7A and other genes encoding stereocilia link proteins in familial meniere disease. *Hear Res*. 2021;409:108329. doi:10.1016/j.heares.2021.108329

- Flook M*, **Escalera-Balsera A***, Gallego-Martinez A, et al. DNA Methylation Signature in Mononuclear Cells and Proinflammatory Cytokines May Define Molecular Subtypes in Sporadic Meniere Disease. *Biomedicines*. 2021;9(11):1530. doi:10.3390/biomedicines9111530
  * Contributed equally and have the right to list their names first in their CV.

- Schoisswohl S, Langguth B, Schecklmann M, (+7 authors), **Escalera-Balsera A** et al. Unification of Treatments and Interventions for Tinnitus Patients (UNITI): a study protocol for a multi-center randomized clinical trial. *Trials*. 2021;22:875. doi:10.1186/s13063-021-05835-z

- Roman-Naranjo P, Parra-Perez AM, **Escalera-Balsera A** et al. Defective α-tectorin may involve tectorial membrane in familial Meniere disease. *Clin Transl Med*. 2022;12(6):e829. doi:10.1002/ctm2.829

- Robles-Bolivar P, Bächinger D, Parra-Perez AM, Roman-Naranjo P, **Escalera-Balsera A** et al. A novel nonsense variant in the CENPP gene segregates in a Swiss family with autosomal dominant low-frequency sensorineural hearing loss. *Eur J Hum Genet EJHG*. 2022;30(11):1301-1305. doi:10.1038/s41431-022-01184-w

- Gallego-Martinez A, **Escalera-Balsera A**, Trpchevska N, et al. Using coding and non-coding rare variants to target candidate genes in patients with severe tinnitus. *NPJ Genomic Med*. 2022;7:70. doi:10.1038/s41525-022-00341-w

- Flook M, **Escalera-Balsera A**, Rybakowska P, et al. Single-cell immune profiling of Meniere Disease patients. *Clin Immunol Orlando Fla*. 2023;252:109632. doi:10.1016/j.clim.2023.109632

- Simoes JP, Schoisswohl S, Schlee W, (+6 authors), **Escalera-Balsera A** et al. The statistical analysis plan for the unification of treatments and interventions for tinnitus patients randomized clinical trial (UNITI-RCT). *Trials*. 2023;24(1):472. doi:10.1186/s13063-023-07303-2

Criterios de calidad para obtener la mención internacional / Quality criteria to obtain the international mention

Esta tesis doctoral se ha preparado siguiendo los criterios para obtener la mención internacional junto al título de Doctor por la Universidad de Granada / This doctoral thesis has been prepared according to the University of Granada requirements to apply for an International PhD.

1 - Estancia de al menos tres meses en un centro de investigación de prestigio de un país extranjero / *Stay of at least three months in a prestigious research center in a foreign countr*
Realizada una estancia de tres meses (septiembre 2022 – noviembre 2022) en el grupo de investigación Computational Rare Disease Genomics en el centro de investigación Wellcome Centre for Human Genetics perteneciente a la Universidad de Oxford, dirigido por la Dra. Nicola Whiffin. / *Three-monts stay (September 2022 – November 2022) in the Wellcome Centre for Human Genetics perteneciente at the University of Oxford, directed by Dr. Nicola Whiffin.*

2 – Panel de expertos internacionales / *Panel of international experts*
Se ha obtenido un informe favorable de dos doctores expertos pertenecientes a una institución de educación superior o instituto de investigación no española. / *A favorable report has been obtained from two expert doctors belonging to a non-Spanish higher education institution or research institute.*

3 - Idioma de presentación de la tesis / *Idioma de presentación de la tesis*
Esta tesis ha sido redactada y será defendida en inglés. Siguiendo con los requerimientos de la Universidad de Granada, el Resumen y las Conclusiones están también redactadas en español, y serán defendidas en español / *This thesis has been written and will be defended in English. Following the requirements of the University of Granada, Abstract and Conclusions are also written in Spanish, and will be defended in Spanish.*

4 - Composición del tribunal / *Committee composition*

El tribunal cuenta con al menos un/a experto/a perteneciente a alguna institución de educación superior o centro de investigación no española, con el título de doctor/a, distinto del/de la responsable de la estancia y distinto de las/los expertas/os internacionales. / *The committee has at least one expert belonging to a non-Spanish higher education institutionor research center, with the title of doctor, different from the person responsible for the stay and different from the international experts.*

Grants and Funding

Alba Escalera Balsera work have received the following funding:

- Unification of Treatments and Interventions for Tinnitus Patients (UNITI) Project, from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement Number 848261.
- Proyecto IMPaCT-Data (Exp. IMP/00019), financiado por el Instituto de Salud Carlos III, co-financiado por el Fondo Europeo de Desarrollo Regional (FEDER, "Una manera de hacer Europa").
- EMBO Scientific Exchange Grant Number 9783.
- Acciones de movilidad del CIBERER (Centro de Investigación Biomédica en Red de Enfermedades Raras).