UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXII CICLO DEL DOTTORATO DI RICERCA IN SCIENZE DELLA RIPRODUZIONE E DELLO SVILUPPO

HIGH-THROUGHPUT DATA ANALYSIS OF HEARING PHENOTYPES ON 9000 SUBJECTS FROM TEN COHORTS AND IN 200.000 INDIVIDUALS FROM UK BIOBANK

Settore scientifico-disciplinare: **MED/03 GENETICA MEDICA**

DOTTORANDO
MARCO BRUMAT

COORDINATORE
PROF. PAOLO GASPARINI

SUPERVISORE DI TESI
DR. GIORGIA GIROTTO

ANNO ACCADEMICO 2018/2019

TABLE OF CONTENTS

ABSTRACT

English.

Normal hearing function (NHF) and its decline with age (Age-Related Hearing Loss, or ARHL) are influenced by both environmental and genetic factors, but which exactly the involved genes are is still unclear. With the ageing of the world population, and the heavy socioeconomic impact of ARHL, the relevance of a detailed knowledge of the genetic inner workings behind these traits is rising.
This thesis has two aims: 1) the identification of new genes playing a role in NHF and in ARHL by using GWAS methodology on cohorts with a different phenotyping approach to hearing (pure-tone audiometry and speech-in-noise test); 2) the assessment of the possibility of cross-replication of the results.

NHF was described with audiometric data using single hearing thresholds or their average value across specific frequencies [Pure-Tone Averages (PTA of low frequencies (PTAL): 250, 500, 1KHz; PTA of medium frequencies (PTAM): 500, 1K, 2KHz; PTA of high frequencies (PTAH): 4K, 8KHz)]; ARHL was represented in subjects aged 50+ with a case-control definition based on PTAH.
Using speech-in-noise test data, after the calculation of the speech reception threshold (SRT), NHF was represented by the full variation of SRT (phenotype A), and ARHL by the extremes of SRT distribution in subjects aged 50+ (phenotype B).
Gene discovery on NHF and ARHL with audiometric data was performed through GWAS meta-analyses on ten cohorts belonging to the G-EAR consortium and coming from Italy, Northern Europe, North America, Caucasus and Central Asia. Overall 9000 subjects aged 18+ with detailed clinical characterisation and full audiometric data were available. A subset of 5745 individuals was selected for ARHL meta-analyses.
The ~200,000 volunteers sampled by UK Biobank were used for gene discovery on ARHL using speech-in-noise data. Due to technical limitations, phenotype A was not tested genome-wide.
Replication of the results on NHF and ARHL from G-EAR analyses was sought in UK Biobank in phenotypes A and B respectively. Replication of the results from phenotype B was sought in G-EAR's analyses on ARHL.

Strongly suggestive association signals ($p < 10^{-6}$ and same direction of effect in all involved cohorts) with NHF traits were detected in G-EAR analyses near or within *CADM2* (250Hz), *SLC7A2*, *CALB1*, *LRRC4C, CAAP1*, *PLXDC2* (1KHz), and *ROBO2* (PTAL). For *CADM2*, *CALB1*, *LRRC4C* and *ROBO2* in particular, literature provides strong evidence supporting their involvement in the hearing system. Furthermore, a genome-wide significant association ($p < 5x10^{-8}$) was found between PTAM and HLA genes. With regards to ARHL, association signals were detected with *CDH13* and *CTIF* genes, previously linked in literature to the NHF.
In the GWAS performed on phenotype B in UK Biobank data, a genome-wide significant signal was detected with *D2HGDH* gene and a suggestive association with *PDE1A* gene.

As the cross-replication of signals between studies was not successful, the replication of the results will be sought in independent cohorts with identical phenotypes. *In vitro*/*in vivo* studies will be needed to clarify the role of the identified genes in the modulation of the hearing function and in age-related hearing loss.

Italiano.

La funzione uditiva (Normal Hearing Function - NHF) e il suo declino dovuto all'età (la presbiacusia, o Age-Related Hearing Loss - ARHL) sono influenzati sia da fattori ambientali che da fattori genetici, ma quali siano esattamente i geni coinvolti è una questione ancora aperta. Con l'invecchiamento della popolazione mondiale, e il pesante impatto socioeconomico dell'ARHL, sta diventando sempre più importante acquisire una conoscenza dettagliata delle basi genetiche che stanno dietro a questi tratti.
Questa tesi ha due obiettivi: 1) l'identificazione di nuovi geni coinvolti nella NHF e nell'ARHL mediante l'uso della metodologia GWAS su coorti in cui è stato usato un diverso approccio alla raccolta del fenotipo uditivo (audiometria e test speech-in-noise); 2) la valutazione della possibilità di una replica incrociata dei risultati ottenuti.

La NHF è stata descritta a partire dai dati audiometrici facendo uso delle singole soglie uditive o del loro valore medio su specifiche frequenze [Medie di toni puri (Pure Tone Averages) alle basse, medie ed alte frequenze (PTAL: 250, 500, 1KHz; PTAM: 500, 1K, 2KHz; PTAH: 4K, 8KHz)]; l'ARHL è stata rappresentata nei soggetti di almeno 50 anni di età mediante una definizione basata sul loro valore di PTAH.

Facendo uso dei dati provenienti dal test speech-in-noise, dopo aver calcolato la soglia di ricezione del parlato (speech reception threshold – SRT), la NHF è stata rappresentata con la variazione totale dell'SRT (fenotipo A), e l'ARHL con gli estremi della distribuzione dell'SRT nei soggetti di almeno 50 anni di età (fenotipo B).

L'individuazione di geni per la NHF e l'ARHL mediante dati audiometrici è stata effettuata mediante una meta-analisi GWAS su dieci coorti facenti parte del consorzio G-EAR e provenienti da Italia, Nord Europa, Nord America, Caucaso e Asia Centrale. Per lo scopo erano disponibili complessivamente circa 9000 soggetti di almeno 18 anni di età con una dettagliata caratterizzazione clinica e audiometrica. Un sottoinsieme di 5745 individui è stato utilizzato nelle meta-analisi sull'ARHL.
I circa 200.000 volontari campionati da UK Biobank sono stati usati per l'individuazione di geni per l'ARHL mediante dati provenienti dal test speech-in-noise. A causa di limitazioni tecniche, il fenotipo A non è stato impiegato.
La replica dei risultati su NHF e ARHL provenienti dalle analisi su G-EAR è stata ricercata in UK Biobank nei fenotipi A e B rispettivamente. La replica dei risultati ottenuti usando il fenotipo B è stata ricercata nelle analisi effettuate in G-EAR sull'ARHL.

Segnali di associazione fortemente indicativi ($p < 1x10^{-6}$ e la stessa direzione degli effetti in tutte le coorti) con i tratti della NHF sono stati individuati nelle analisi G-EAR vicino o entro *CADM2* (250Hz), *SLC7A2*, *CALB1*, *LRRC4C, CAAP1*, *PLXDC2* (1KHz), e *ROBO2* (PTAL). Per *CADM2*, *CALB1*, *LRRC4C* and *ROBO2* in particolare, la letteratura fornisce una robusta evidenza a supporto del loro coinvolgimento nel sistema uditivo. In aggiunta, un'associazione significativa genome-wide ($p <5x10^{-8}$) è stata individuata tra PTAM e geni della regione dell'HLA. Per quanto riguarda l'ARHL, segnali di associazione sono stati individuati con i geni *CDH13* e *CTIF*, già collegati in letteratura alla NHF.
Nel GWAS sul fenotipo B eseguito sui dati di UK Biobank, un'associazione significativa genome-wide è emersa con il gene *D2HGDH* mentre un'associazione indicative con *PDE1A*.

Dal momento che la replica incrociata dei segnali ottenuti non ha avuto successo, la replica dei risultati sarà ricercata in coorti indipendenti dotate degli stessi fenotipi. Ulteriori studi *In vitro*/*in vivo* saranno necessari per chiarire il ruolo dei geni identificati nella modulazione della funzione uditiva e nella presbiacusia.

INTRODUCTION

"Blindness cuts us off from things, but deafness cuts us off from people."

Helen Adams Keller (1880-1968), activist and educator

The auditory system.

Hearing, the ability to detect sound, is one of the most complex mechanisms of perception in humans. It has played a fundamental role in our evolution, and it has developed to the point of becoming a cornerstone of our social communication and integration (Fuchs J.C. et al, 2015).
At the basis of this crucial ability lies an intricate system, the *auditory system* (Figure 1), composed of three main sections - outer, middle, and inner ear – that work together to convert mechanical soundwaves into electric impulses for the brain to process (Dror AA. et al, 2010).
Soundwaves from the environment reach the outer ear and are captured by the auricle, which directs them through the ear canal up to a thin diaphragm at its end, known as the eardrum or the *tympanum*. Just beyond it lie three small bones – *malleus*, *incus* and *stapes* – connected in sequence one to the other, which, with the eardrum, make up the middle ear. Vibrations of the eardrum set the three bones into motion, and the impulses are then propagated by the stapes to an oval window on the side of a fluid-filled coiled structure, the *cochlea*, located in the inner ear. In the cochlea, the *organ of Corti* translates the mechanical soundwaves into electrical neural impulses by means of hair cells. The sound vibrations make their stereocilia (actin-rich hair-like projections on the upper surface of these cells) move, and this movement opens ion channels that modulate potential within the cell. Neurotransmitters are then released to the synaptic junctions between hair cells and neural fibres (Steel K.P. et al, 2001): this creates a neural spike that moves along the auditory nerve fibre to the temporal lobe of the brain, where it is processed.

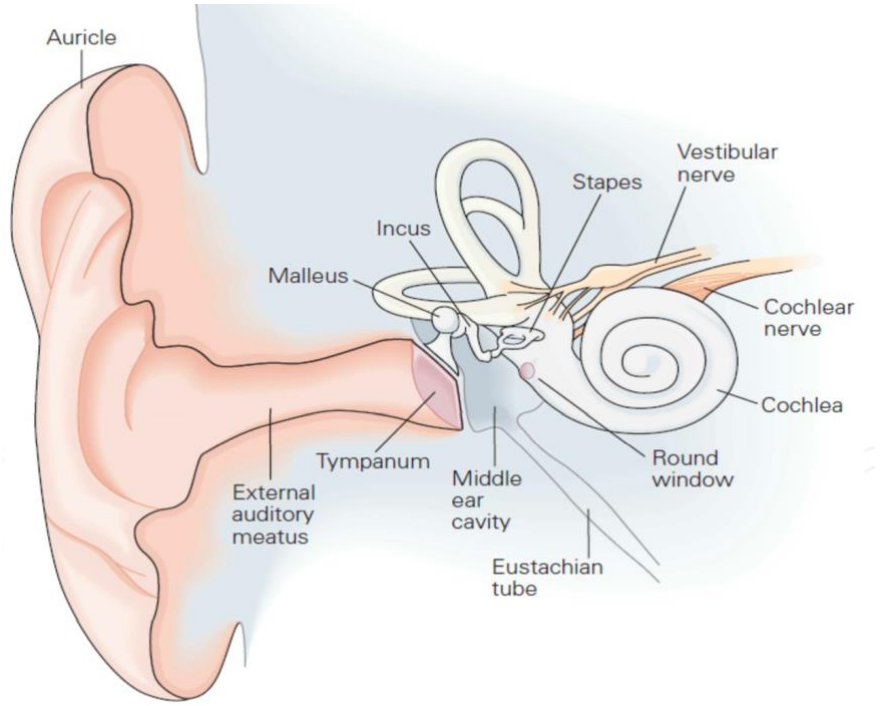


Figure 1 - Schematic representation of the human auditory system. Source: Kandel E. et al, 2013.

Malfunctions and damages to one or more of the elements which constitute this complex mechanism can lead to hearing impairment (or *hearing loss*): for example, complications at birth, a prolonged exposure to noise or the general ageing of the body are amongst the many factors that have been identified as involved in reduced hearing ability.

Also genetic factors play a fundamental role, with more than 140 loci and 80 genes reported as involved in hereditary hearing impairment to date[1].

Furthermore, for some forms of hearing loss the aetiology is multifactorial (or *complex*): the impairment is given by an intricate interplay between environmental and genetic factors. This is the case, for example, of *noise-induced* hearing loss (Lavinsky J. et al, 2016) and *age-related* hearing loss (Gates G.A. et al, 2005; Tu N.C. et al, 2018). As of today, still little is known about the genetic factors underlying complex forms of hearing loss.

[1] http://hereditaryhearingloss.org/

Hearing ability.

Hearing ability is measured in decibels (dB). This quantity describes how intense a sound must be at a given frequency before it can be detected; it is known as the *hearing threshold* at that frequency for the tested ear.

The hearing ability of a subject is considered *normal* if the threshold of both ears is lower than 25 dB. Table 1 details the different degrees of hearing loss.

Degree of hearing loss	Hearing threshold (dB range)
Normal hearing	-10 – 25
Mild	26 – 40
Moderate	41 – 55
Moderately severe	56 – 70
Severe	71 – 90
Profound	91+

Table 1 - Degree of hearing loss based on the hearing threshold. From Clark J.G., 1981.

Hearing loss is considered *disabling* when the threshold exceeds 40db in adults and 30dB in children[2]. In 2019, around 466 million people (of which 34 million children) suffer from disabling hearing loss, and projections estimate this number will grow to over 900 million by 2050. Unaddressed hearing loss costs 750 billion dollars per year in healthcare[3].

Evaluation of the hearing ability.

Hearing ability can be evaluated through a variety of tests, the most common of which is the *pure-tone audiometry*.

Briefly, in this test sound is presented at a specific frequency at a time, selected from a set of six standard frequencies (250Hz, 500Hz, 1KHz, 2KHz, 4KHz and 8KHz) and its intensity is gradually increased starting from -10 decibels. Subjects are instructed to inform the tester (e.g. by raising a

[2] https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
[3] *ibidem*.

hand, pressing a button, etc.) whenever they can hear the sound and the corresponding hearing threshold is then recorded.

Moreover, this threshold is called *air conduction hearing threshold* if sound is presented through earphones or through loudspeakers in a booth, in order to test the ability of a subject to perceive sound from the environment. It is called *bone conduction hearing threshold* if the outer and middle ear are bypassed and sound is directly delivered to the inner ear using a bone conductor placed on the mastoid process.

The recorded hearing thresholds are usually represented in a diagram, called *audiogram* (Figure 2).

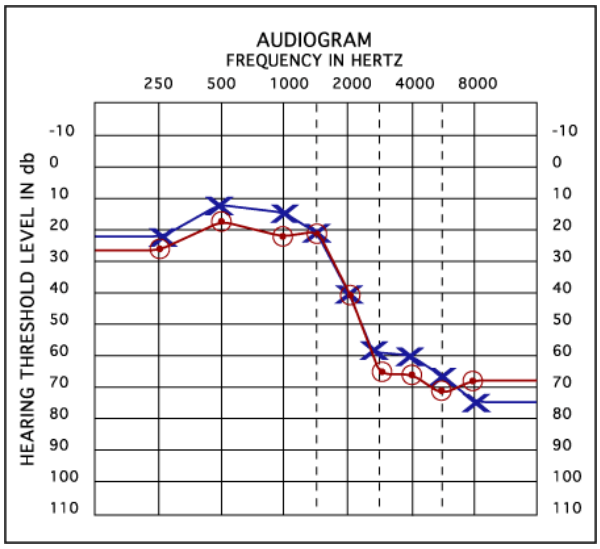


Figure 2 - Example of an audiogram. Source: mynoise.net.

In this diagram, frequency is reported in Hertz (Hz) on the horizontal axis; the hearing thresholds are reported in dB on the vertical axis. Usually the vertical axis is reversed to reflect the concept of hearing *loss*: the lower the point lies on the diagram, the greater is the loss at that frequency for the ear of the subject. Blue X's and red circles represent hearing thresholds for left and right ear respectively.

As a final remark, hearing thresholds can also be studied in aggregated forms, the Pure Tone Averages (PTAs), that summarise the hearing ability. For example, the PTA at 250Hz and 500Hz is calculated by taking the mean of the hearing thresholds recorded for those two frequencies.

Age-related hearing loss.

Age-related hearing loss (ARHL; also known as *presbycusis*) is a multifactorial disease that leads to a gradual bilaterally symmetrical hearing loss in elderly people. High frequencies are the first to be affected, and the disease slowly progresses to involve the medium and the low frequencies as well. ARHL is already one of the most common diseases in the elderly and its prevalence is expected to grow even further with the progressive ageing of the world population.

The quality of life can be heavily affected by ARHL: understanding spoken words becomes increasingly difficult – specially in noisy environments (Wolber L. et al, 2012) – and this difficulty in communication can lead to social isolation (Dalton DS. et al, 2003), loneliness (Sung YK et al., 2016), depression (Brewster KK. et al., 2018), and even dementia (Lin FR. et al., 2014). In addition, the diffusion of ARHL has an important impact in terms of economic burden on healthcare, with 3.10 billion dollars in excess medical expenditures per year in the US only (Foley DM. et al, 2014).

As ARHL is a multifactorial disease, both environmental and genetic causes are involved. While several of the former have been identified (one of the most important being the prolonged exposure to noise (Fransen E. et al, 2008)), still little is known about the latter (Roth TN., 2015), despite all the efforts. The heritability of ARHL is estimated in the range of 36-70% (Nagtegaal A.P. et al., 2019), but only a handful of genes have been confidently associated to ARHL to date (Tawfik KO. et al., 2019).

The study of multifactorial traits and diseases through genome-wide association studies (GWAS).

The *genome-wide association study* (*GWAS*) is an experimental design created to detect statistical association between a trait or disease of interest and variants (mainly SNPs – Single Nucleotide Polymorphisms) along the genome (Visscher P.M. et al, 2017); it has been a popular tool for the last years (Visscher P.M. et al, 2017; Buniello A. et al, 2019), leading to the successful identification of many new genetic variants linked to multifactorial traits and diseases (e.g. Yengo L. et al., 2018; Zeggini E. et al., 2011), such as the hearing function (e.g. Girotto G. et al., 2011; Vuckovic D. et al., 2015; Nagtegaal AP. et al., 2019).

The GWAS' design has three main requirements. First, GWAS requires a large number of genetic markers that cover the full genome. SNP arrays can capture genotype information of several hundred thousand (up to a couple of millions) SNPs, and the number of markers available for association can be expanded *in silico* to tens of millions using *imputation*, a method that exploits the linkage disequilibrium information in a reference population to statistically infer genotypes of variants that have not been directly genotyped (Li Y. et al, 2009; Marchini J. et al, 2010; Das S. et al, 2018).

Second, GWAS requires a large number of samples, adequately phenotyped for the trait of interest and potential confounding factors, in order to reach a statistical power sufficient to detect association. Genotyping of a large number of samples has been made possible by SNP arrays becoming increasingly cheap, while the difficulty of phenotyping a large number of samples depends heavily on the investigated trait.

A common way to address the requirement of a large number of samples is to apply meta-analysis techniques to combine the association results from different populations, thus obtaining a larger conjoined sample size (Zeggini E. et al, 2009; Evangelou E. et al, 2013).

Third, GWAS requires statistically powerful analytic methods for association capable of handling the potential biases in the data. One of the most reliable methods performs linear mixed model regression (LMM) in order to account for population stratification and relatedness (Bolker BM. et al, 2009), while allowing adjustment for other covariates, such as age and sex.

Finally, GWAS involves an extreme number of statistical tests, one for each available variant. For this reason, the number of potential spurious associations increases and it is important to adjust the p-value significance threshold (usually set at 5%) accordingly. Assuming that approximately 1,000,000 of the tests performed in a GWAS are actually independent and applying Bonferroni correction, significance threshold for GWAS (called *genome-wide significance threshold*) is set at $5x10^{-8}$ (Risch N. et al., 1996).

The study of hearing traits through genome-wide association studies (GWAS).

The addressal of the sample size issue.

While pure-tone audiometry has been the standard evaluation of the hearing ability for decades (Johnson EW., 1970), a complete audiological and otological visit is a process that takes a considerable amount of clinical time, limiting the number of subjects that can be sampled by a single research group.

For this reason, in the perspective of studying hearing traits by means of genome-wide association studies, the large sample size required by this study design is an issue that needs to be addressed.

A first solution lies in the foundation of *consortia*, the joining of forces between research groups that share similar interests and that agree about performing GWAS analyses according to a common research plan, the results of which will subsequently be put together by means of meta-analysis. In this light, in 2010 the University of Trieste launched the *G-EAR Consortium* project, aimed at the investigation of genetic factors that underlie the normal hearing function and age-related hearing loss by GWAS meta-analysis on well characterised audiometric cohorts.

A second solution is to perform GWAS on genotype and phenotype data collected by large epidemiological studies such as the UK Biobank project, one of the largest biobanks as of today (Bycroft C. et al, 2018).

For the reasons explained above, these large biobanks rely on different methods of assessment of the hearing ability, selected for ease and speed of administration, and cost-effectiveness.

Self-reporting of hearing impairment.

The most common of such methods consists in interviewing subjects about their perception of their own hearing ability. This interview is carried out by means of direct questions such as "*Do you ever have difficulties with your hearing?*", used in the aforementioned UK Biobank project, and "*Do you have a hearing problem now*?", used in Framingham Heart Study, a prospective cohort study based in Framingham, Massachusetts (USA) (Tsao CV. et al, 2015; Dawber TR. et al, 1963; Dawber TR. et al, 1951). Possible answers may be binary ("*yes*/*no*") or graded ("*No, I always hear everything / Yes, sometimes I do not hear what is being said / Yes, I regularly do not hear what is being said / Yes, I*

almost never hear what is being said"). The derived phenotype is usually referred to as *self-reported* (or *subjective*) *hearing impairment*.

This method has been described in recent studies (B.C. Oosterloo et al., 2020) as having a ~90% accuracy of detection of hearing loss in adults.

The speech-in-noise test.

Another method is the *speech-in-noise test*. As the name implies, the test evaluates the ability to recognise speech in a noisy environment. This approach, although different from pure-tone audiometry, makes for a good alternative assessment of the hearing ability (Smits C. et al, 2013). The several developed versions all follow a general scheme:

- Subjects are asked to recognise some speech (one or more words, digits, etc.) played with a varying level of background noise, for several rounds. Left and right ear are tested separately. The noise level varies between rounds and the difficulty in speech recognition changes accordingly. For each round, the signal-to-noise ratio is recorded;
- At the end of the test, played speech and reported answers are compared to calculate the *speech reception threshold* (SRT), defined as the signal-to-noise ratio at which half of the presented speech can be recognised correctly. The higher the SRT value, the worse the hearing impairment of the subject.

The exact number of rounds, the rules according to which the background noise changes, the composition of the speech and the means of administration vary amongst the different versions (e.g. Wolber LE. et al., 2012; Potgieter JM. et al., 2015). For example, in one of these - the *digit triplet test* (DTT) – the presented speech is limited to a series of triplets of digits. This particular version is the one included in the hearing assessment sessions of the UK Biobank project.

It has been found that PTAs and SRT show a strong correlation (R=0.8) (Koole A. et al., 2016), but it is still unclear if (and, if so, to which extent) the genetic factors that can be detected using these two approaches overlap.

An overview of the studies to date.

GWAS methodology has already been applied successfully to the study of hearing, leading to the identification of new candidate genes.

The first GWAS study was focused on ARHL (Friedman R.A. et al, 2009) and was performed on individuals of European ancestry. A significant association with glutamate receptor *GRM7* was detected, which was later corroborated by a study carried out on the Finnish population of the Saami (Van Laer L. et al, 2010) and replicated in European American subjects (Newman D.L. et al, 2012).

Normal hearing function was investigated in a GWAS meta-analysis by G-EAR Consortium, which made use of isolated European populations (Girotto G. et al, 2011). This study highlighted a series of suggestive associations, including *DCLK1*, *PTPRD*, *CMIP* and another glutamate receptor, *GRM8*. A different study on the normal hearing function (Wolber LE. et al, 2014) identified *SIK3*, member of the salt-inducible kinase family, a family known to be involved in the inner ear (Degerman E. et al, 2011). Later, *PCDH20* and *SLC28A3* were linked to the modulation of the normal hearing function (Vuckovic D. et al, 2015).

A subsequent meta-analysis (Hoffmann T.J. et al, 2016) identified a candidate gene for ARHL in *ISG20*, and hypothesised the involvement in this disease of *TRIOBP*, *EYA4* and *ILDR1*, three genes known to be related to other forms of hearing loss (Riazuddin S. et al, 2006; Shahin H. et al, 2006; Wayne S. et al, 2001; Borck G. et al, 2011).

Finally, a recent investigation (Nagtegaal A.P. et al, 2019), which studied ARHL as a quantitative trait, listed five novel associated loci and strengthened the possible involvement of *ILDR1* in ARHL.

All the aforementioned studies relied on pure-tone audiometry data, with the vast majority of them being meta-analyses of smaller association studies in order to increase the overall number of analysed subjects. It was not until recently that the large sample size obtained in large biobanks using the alternative assessment methods was exploited to the end of gene discovery.

Self-reported hearing impairment data from UK Biobank was used in the largest association study to date (Wells H.R.R. et al, 2019), which identified *NID2*, *CLRN2* and *ARHGEF28* as having a potential role in ARHL.

With regards to the SRT or other phenotypes derived from *speech-in-noise* test data, to the best of our knowledge, no genome-wide association study has been published using UK Biobank or other resources as of today.

Aim of the thesis.

Thanks to the availability at the same time of a large collection of genotyped samples with deep audiometric characterisation belonging to University of Trieste's G-EAR Consortium and the complete data from the screening carried out with the DTT test on UK Biobank cohort, this thesis can pursue two aims:

1) the identification of new genes potentially playing a role in the normal hearing function and in age-related hearing loss;
2) assessment of the possibility of replication of the results obtained from G-EAR Consortium data in those obtained from UK Biobank, and vice versa.

With regards to the first aim, the identification of new genes is carried out by using GWAS methodology on cohorts of subjects whose hearing ability was evaluated through pure-tone audiometry and through DTT test. In order to do this, the normal hearing function and the effects of age-related hearing loss on hearing are described through the data available from the two different approaches at phenotyping, and the derived traits are then used in the GWAS approach on the respective cohorts; as concerns the second aim, summary statistics of association from the two studies are compared.

During this research, G-EAR consortium data has been extended with an important contribution from Framingham Heart Study and Rotterdam Study. This contribution was coordinated by the author of the thesis.

With regards to UK Biobank data, the research herein described has been conducted using the UK Biobank Resource under Application Number 28489.

MATERIALS and METHODS

Discovery of genes related to normal hearing function and age-related hearing loss through GWAS meta-analysis on G-EAR Consortium cohorts.

Involved cohorts.

The latest meta-analysis from G-EAR Consortium involves ten cohorts, belonging to the studies described below. In-depth details about recruitment and sampling procedures for each cohort can be found in the articles cited in each description. To avoid cumbersome repetitions in the next pages, a short id code is assigned to each cohort.

Italian populations from the INGI project

The INGI (Italian Network of Genetic Isolates) project is a collaboration between research institutions in Italy aimed at reconstructing the molecular bases of complex traits by investigating genetically isolated Italian populations. Studies were conducted referring to a common operational protocol. In each population, genotype samples were collected, alongside a detailed anamnesis, more than 120 biochemical parameters and 400 phenotypes, including anthropometric measures, lifestyle habits, diseases and pure-tone audiometry (Girotto G. et al, 2011). Data from three of these populations were available: INGI Friuli-Venezia Giulia, INGI Val Borbera and INGI Carlantino.

The INGI Friuli-Venezia Giulia cohort (FVG) is a collection of samples coming from six small villages (Clauzetto, Erto, Illegio, Resia, San Martino del Carso and Sauris) located in north-eastern Italy, in the Friuli-Venezia Giulia region (Esko T. et al, 2013).

The INGI Val Borbera cohort (VBI) consists of samples collected from seven small villages in the Borbera Valley, a geographically isolated valley located within the Apennine Mountains, in the southern part of Piedmont, in north-western Italy. The valley is inhabited by approximately 3000 descendants from the original population (Traglia M. et al, 2009).

The INGI Carlantino cohort (CAR) consists of samples collected in a small village of the same name, situated in the northern part of Puglia region, in southern Italy (Esko T. et al, 2013).

Silk Road (SR)

Silk Road cohort is composed of samples collected in 2010 by the Marco Polo Scientific Expedition (promoted by the University of Trieste) from small rural communities located in Caucasus and Central Asia, along the historical trade route known as Silk Road. Each participant underwent a careful clinical (psychological, neurological, cardiological, …) examination and an audiometric test; data on lifestyle and food habits was also collected (Girotto G. et al, 2011; Mezzavilla M. et al, 2014).

Milan Presbycusis Cohort (PM)

Milan presbycusis cohort is a set of samples from outbred elderly people, collected by the University of Milan from 2011 onwards, specifically aimed at the study of ARHL. Involved subjects are self-identified white people, coming from Milan and the Lombardy region, in Italy; pure-tone audiometry and a detailed anamnesis were collected.

Age-related Hearing Impairment Study Antwerp (AWP)

The Age-Related Hearing Impairment Study is a population-based study involving inhabitants of a residential village of Antwerp, in Belgium, invited through population registries (Van Eyken E. et al, 2006). Subjects underwent clinical examination, otoscopy, audiometry and completed a detailed questionnaire on medical history and exposure to environmental risk factors.

Rotterdam study (RS)

Rotterdam Study is a prospective, population-based cohort study among subjects living mainly in the Ommoord district in the city of Rotterdam in The Netherlands. Started in 1990, the study is aimed at the investigation of cardiovascular, endocrine, hepatic, neurological, ophthalmic, psychiatric, dermatological, otolaryngological, locomotor, and respiratory diseases. During the years it has had several recruitment extensions (referred to as RS 1, 2, 3, …), expanding at the same time the range of collected phenotypes (Ikram MA. et al., 2017). In 2011 hearing assessment by means of pure-tone audiometry was implemented in the study protocol (Rigters SC. et al, 2018).

Framingham Heart Study (FHS)

Framingham Heart Study is a long-term prospective cohort study started in 1948 aimed at the identification of cardiovascular disease (CVD) and its risk factors (Tsao CV. et al, 2015; Dawber TR. et al, 1963; Dawber TR. et al, 1951). As of today, the study spans three generations of volunteers,

living in the town of Framingham, in Massachusetts (USA). Nearly all FHS participants are self-identified white (of European descent). Besides CVD, a wide range of other phenotypes was collected, including hearing (specific details can be found online[4]), which are publically available at the dbGaP website[5].

All studies received the approval of the ethical committee of the respective promoting institutions; every participant signed an informed consent.

Genotyping and quality control.

DNA was extracted from blood or saliva samples, and genotyped with different genotyping platforms (Table 2). Standard quality control was performed in each cohort separately, applying filters that took into account minor allele frequency, sample and genotype call rate and Hardy-Weinberg equilibrium. Further inclusion/exclusion criteria were applied in the specific studies (Table 2). Genotypes were referred to the forward strand and reported with the coordinates of the 1000 Genomes Project build 37 (The 1000 Genomes Project Consortium et al, 2015).

Imputation.

After quality control, SHAPEIT v2 (O'Connell J. et al, 2014) was used for the phasing step on FVG, VBI, CAR, SR and PM data; IMPUTE v2 (Howie B. et al, 2012) was used for the imputation phase. Data from the INGI cohorts was imputed to the Italian Genome Reference Panel (IGRP) v.1. (Cocca M. et al, 2019). Data from PM was imputed to the 1000 Genomes phase 3 reference panel (The 1000 Genomes Project Consortium et al, 2015).

The Imputation Server of the University of Michigan (Das S. et al, 2016) was used for the phasing and imputation of genotypes of RS, FHS and SR: RS and FHS were imputed to the 1000 Genomes phase 3 reference panel, while SR to the HRC (Haplotype Reference Consortium) reference panel (McCarthy S. et al, 2016).

SHAPEIT v2 and IMPUTE v2 were used for haplotype estimation and imputation of AWP data to 1000 Genomes phase v1v3 reference panel.

[4] http://www.framinghamheartstudy.org/researchers/description-data/index.php

[5] http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v2.p1

Cohort	FVG	VBI	CAR	SR	PM	AWP	RS1	RS2	RS3	FHS
N (in analyses)	1470	425	261	656	179	1102	684	1173	2155	1100
Men; Women	634; 836	180; 245	113; 148	249; 407	76; 103	555; 547	280; 404	530; 643	941; 1214	406; 694
Age range	18-91 (mean=51.9)	25-90 (mean=58.7)	18-89 (mean=53.4)	18-82 (mean=39.8)	51-67 (mean=59.6)	53-67 (mean=60.9)	77-101 (mean=82.9)	66-99 (mean=75.4)	52-95 (mean=62.3)	32-86 (mean=59.2)
Genotyping platform	Illumina 370K, Illumina 700K, MegaEX	Illumina 370K, Illumina 700K	Illumina 370K	Illumina 700K	Illumina 700K	Illumina CNV370 quad chip, Illumina HumanOmniExpress BeadChip	Illumina 550K/660Q	Illumina 550K/660Q	Illumina 550K/660Q	Affymetrix 500K mapping array + Affymetrix 50K supplemental array
Genotypes called with	Illumina BeadStudio	Illumina BeadStudio	Illumina BeadStudio	Illumina BeadStudio	Illumina BeadStudio	Illumina BeadStudio	Illumina GenomeStudio	Illumina GenomeStudio	Illumina GenomeStudio	Affymetrix BRLMM
Sample quality control (inclusion)	Sample call rate >= 0.95, gender check	Sample call rate >= 0.95, gender check	Sample call rate >= 0.95, gender check	Sample call rate >= 0.95, gender check	Sample call rate >= 0.95, gender check	Sample call rate >= 0.95, gender check	Sample call rate >= 0.95, gender check, excess heterozygosity, European ancestry	Sample call rate >= 0.95, gender check, excess heterozygosity, European ancestry	Sample call rate >= 0.95, gender check, excess heterozygosity, European ancestry	Sample call rate >= 0.95; autosomal heterozygosity < ± 5SD, non-excessive Mendelian errors
Genotype quality control (inclusion)	Call rate >= 0.99; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.99; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.99; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.99; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.99; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.95; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.95; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.95; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.95; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01	Call rate >= 0.97; Hardy-Weinberg P > $1x10^{-6}$; MAF >= 0.01

Table 2 – Genotyping and quality control.

Investigated traits.

In all cohorts, air conduction thresholds at 250Hz, 500Hz, 1KHz, 2Khz, 4KHz and 8KHz were collected for both ears for every participant.

The normal hearing function was described through nine quantitative traits - the six thresholds listed above and three pure-tone averages, at low, medium and high frequencies (PTAL, PTAM and PTAH, respectively) – and ARHL with a case-control phenotype.

These ten traits were considered for investigation through genome-wide association studies.

Normal hearing function.

Subjects aged less than 18 were excluded, as well as the ones affected by clear monogenic forms of hearing loss, or with a history of exposure to noise. The three PTAs were calculated: PTAL, by taking the mean value of the thresholds at 250Hz, 500Hz and 1KHz; PTAM, by taking the mean value of those at 500Hz, 1KHz and 2KHz; and for PTAH those at 4KHz and 8KHz.

To avoid non-genetic variation, thresholds and PTA values referring to the *best ear* were chosen. Here, *best ear* is defined as the one with the lowest threshold (or the lowest PTA* value) between the two. The "*best ear"* phenotype was subsequently adjusted for sex and age. All nine traits showed a skewed distribution (see e.g. Figure 3) and were subsequently normalised through rank-normalisation. As a last step, outliers (subjects with normalised values more than 3 standard deviations away from the mean) were removed.

ARHL case-control phenotype.

For the ARHL case-control phenotype, the same exclusion criteria as the ones used for the normal hearing function were applied to subjects aged 50 or older; these subjects were then divided in cases and controls with respect to ARHL as follows:

- *CASES*: subjects with the PTAH value of the best ear >= 40;
- *CONTROLS*: subjects with the PTAH value of the best ear <= 25,

as previously defined in literature (Vuckovic D. et al, 2018).

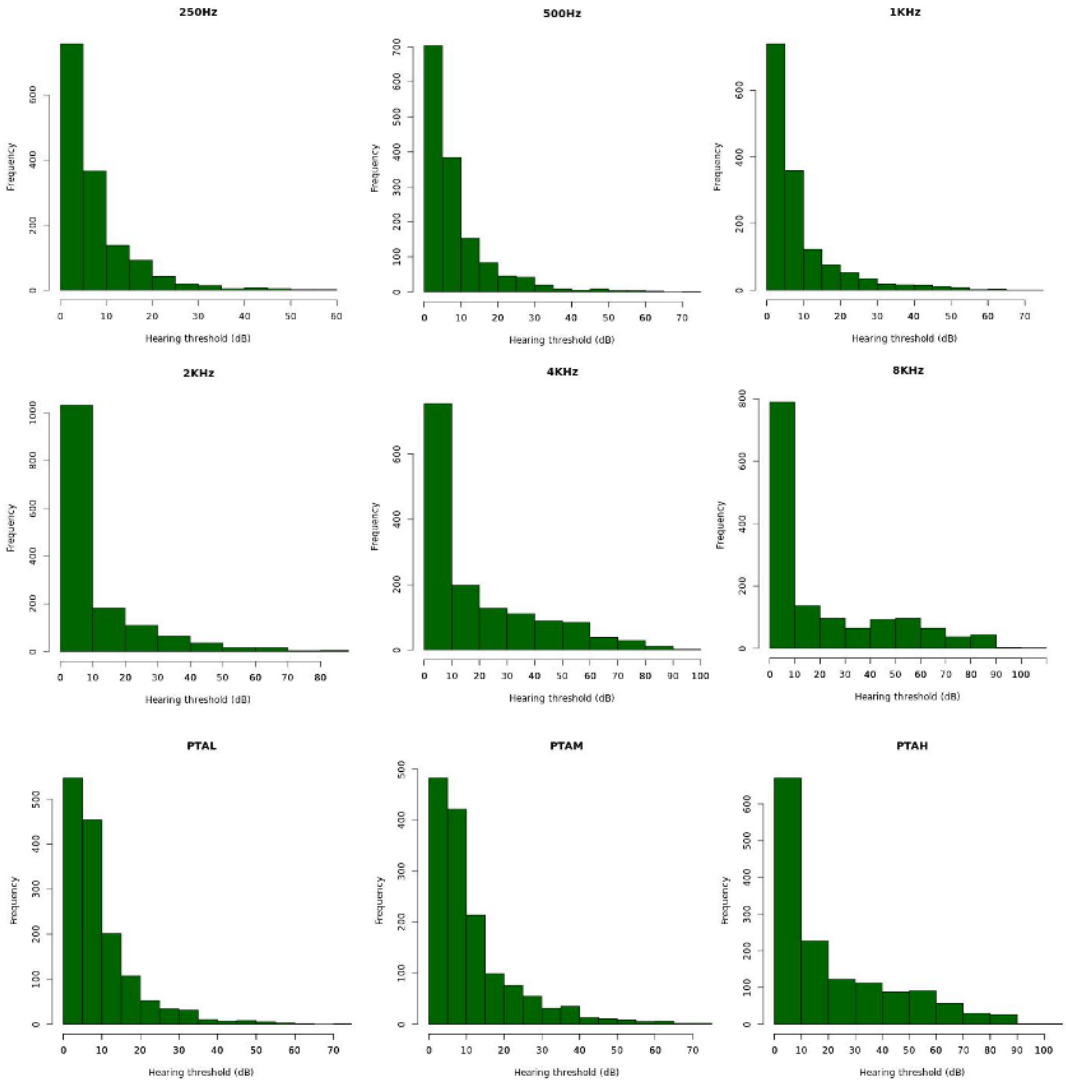


Figure 3 - Distributions in FVG cohort of the nine traits describing the normal hearing function.

Association analyses.

Genome-wide association analyses were carried out separately in all involved cohorts. The additive genetic model was assumed; genotype-phenotype association was assessed with Wald's test.
For the nine quantitative traits, correction for sex and age was already part of the construction of the input phenotype, so no further adjustment was needed; association with ARHL phenotype was adjusted for sex and age at this stage.

For FVG, VBI, CAR, SR, AWP, PM association analyses were performed in R (R Core Team, 2018) using linear mixed model regression, which allowed to account for genomic kinship. On genotyped variants, analyses used GRAMMAR-gamma method (as implemented in GenABEL (Svisheva GR. et

al, 2012)), while MixABEL (Aulchenko YS. et al, 2007) was used on the imputed ones. With regards to the ARHL phenotype, the effect sizes resulting from linear mixed model regression were converted to odds ratios with LMOR (Lloyd-Jones L. et al, 2018).

In RS1, RS2, RS3 and FHS association studies were carried out in rvtests v20171009 (Zhan X. et al, 2016).

Quality control of association analyses.

Summary statistics for the ten association analyses were collected for all cohorts, and checked for integrity. Variants with minor allele frequency (MAF) < 1% and imputation quality score (INFO) < 0.4 were excluded. Quantile-quantile plots were produced and lambda coefficients calculated (see Supplementary Table 5) for all analyses in order to check for inflation due to population stratification.

Meta-analysis.

METAL (Willer CJ. et al, 2010) was used to perform an inverse-variance based meta-analyses of the summary statistics from the single studies. Genomic control was performed in METAL during the input phase of the meta-analysis. As PM is a cohort aimed specifically at the study of ARHL, data from PM was used for the ARHL meta-analysis only.

After meta-analysis, variants present in less than three of the meta-analysed populations were removed, to ensure a minimum conjoined sample size greater than 1000. Inflation factor was recalculated (Supplementary Table 6) and quantile-quantile plots produced (Supplementary Figure 1 and Supplementary Figure 2) to check for inflation at this stage. Results were inspected with Manhattan Harvester, a tool designed for the identification of peak structures in Manhattan plots (Haller T. et al, 2019); further refinement, when needed, was performed visually.

Genome-wide significance threshold was set as $5x10^{-8}$. For normal hearing function traits, a variant was considered as suggestively associated if $p < 10^{-6}$, while for ARHL, in order to compensate for the loss of power due to the use of linear regression on a binary trait in FVG, VBI, CAR, SR, AWP and PM, a variant was considered as suggestively associated if $p < 1x10^{-5}$ and it was present in all populations with the same direction of effect.

Annotation of lead variants.

ANNOVAR (Wang K. et al, 2010) was used to map the lead variants of the detected peaks to the nearest genes, and to determine whether the variants were contained in an intronic, intergenic or exonic region.

Conditional analysis.

As genotype-level data was not directly available for some cohorts involved in the meta-analyses (i.e. RS1, RS2, RS3, FHS), function COJO (Yang Y. et al, 2012) of the software GCTA (Yang Y. et al, 2011) was used to assess whether multiple independent signals were present within the associated loci obtained from meta-analyses. For each locus, a region with a width of 4 megabases and centered on the lead variant was considered.

Genotype data of the European super-population of the 1000 Genomes Project (The 1000 Genomes Project Consortium et al, 2015) was selected as the linkage disequilibrium reference sample required by COJO. Variants with effect alleles showing frequency difference greater than 0.2 between the meta-analysis sample and the reference sample were removed.

Expression in mouse and zebrafish ear.

Public databases (i.e. gEAR (https://umgear.org) and SHIELD (Shen J. et al., 2015)) were consulted to assess the expression in mouse and zebrafish ear of the genes close to the lead variants.

Replication in UK Biobank DTT data of the signals detected by G-EAR Consortium meta-analysis; discovery of genes in UK Biobank DTT data.

UK Biobank's DTT test data was used both for the replication of signals associated to the normal hearing function and ARHL in G-EAR analyses and for gene discovery, following these three steps:

1. Two different phenotypes were obtained: the first (***phenotype A***) was defined to include the whole variation of the ability of speech recognition in noisy environments measured by the test, in order to represent the variation of the normal hearing function; the second

(***phenotype B***) was defined to consider the extremes of speech-recognition ability in the middle-aged and older, in order to describe a case-control situation for potential age-related hearing loss.

2. Association between genotype and phenotypes A and B was performed to replicate the lead SNPs from the loci identified in the G-EAR meta-analyses on normal hearing function and ARHL respectively.
3. A genome-wide association study aimed at gene discovery was performed using phenotype B.

Cohort description.

The UK Biobank project is a long-term prospective epidemiological resource in the United Kingdom. By means of assessment centres located throughout the country, from 2006 onwards, UK Biobank project collected biological samples and a wide range of phenotypes from approximately 500,000 volunteers with different ethnic backgrounds and aged between 40 and 69 at the time of recruitment (Bycroft C. et al, 2018).

During the visits, participants were asked questions on socio-demographic factors, lifestyle and their medical history and completed a set of physical measures; the evaluation of the hearing ability was included at a later stage as an enhancement to the original visit formula and was carried out on approximately 200,000 participants (ibidem).

Samples of blood, urine and saliva were collected to perform different assays, including genetic analyses. All participants signed an informed consent (ibidem).

The data release available in 2018 and used for this analysis is the final result of a huge effort made by UK Biobank's centres. Recruitment, genotype and phenotype collection, genotyping and the subsequent data imputation and quality control phases cannot be adequately summarised here: a detailed description is available in the UK Biobank flagship paper (see: Bycroft C. et al, 2018).

The Digit Triplet Test.

As stated in the introduction, UK Biobank employed a specialised version of the *speech-in-noise* test, the Digit Triplet Test (DTT), in which the presented speech is limited to triplets of digits. The

particular implemented version involved 15 rounds per ear, and limited the signal-to-noise levels in the range +8 (minimum) to -12 (maximum).
A detailed description of the implementation is presented in the operation manual[6].

Definition of phenotypes A and B.

The definition of phenotypes A and B follows similar steps; differences between the two will be highlighted when necessary.
For 195,004 subjects, the complete history of the DTT assessment (including the two follow-up sessions, for the subset of subjects who took part) was available, alongside other variables of interest for the present study (Supplementary Table 7). A summary of the variables included in the analyses is presented in Supplementary Table 8.

Phenotypes A and B were defined through subsequent selection phases.

- For the volunteers who attended multiple hearing assessment sessions, only data pertaining to the first one was used for the construction of the phenotypes.
- Using the provided information about the genetic kinship between subjects, a subset was selected to include no first or second degree relatives.
- The Speech Reception threshold (SRT) value was estimated by UK Biobank only for the subjects who completed all 15 rounds for a given ear. In order to have an estimation of the SRT value for both ears, only subjects who completed all 15 rounds of the DTT for both ears were selected; in addition, this filtering excludes individuals who chose to be tested for one ear only, for reasons that may include severe monolateral hearing loss.
- Subjects wearing cochlear implants and subjects affected by severe tinnitus were removed from the analyses to avoid bias in speech recognition in both ways.
- The triplets played during the 15 rounds of the test and the relative triplets entered by the subjects in response were compared, and the total number of correct answers for left and right ear were computed. As extreme differences in performance between the ears may

[6] Hearing 'Speech-in-Noise' Test Version 1.3 (http://biobank.ndph.ox.ac.uk/showcase/showcase/docs/Hearing.pdf)

reasonably hint at forms of monolateral hearing loss or potentially unbalanced DTT test sessions (Figure 4), subjects with a difference in correct answers greater than mean + 3 standard deviations or lower than mean – 3 standard deviations were removed.

- Only subjects with “white British” ancestry (defined in UK Biobank flagship paper (Bycroft C. et al, 2018)) were considered, as previously reported in literature on association studies using UK Biobank data (Wells H.R.R. et al, 2019).
- Individuals flagged by UK Biobank analysts during quality control (Bycroft C. et al, 2018) as showing unusually high heterozygosity or with a missing rate greater than 5% were excluded.
- For phenotype A, no age cut-off was applied. For phenotype B, only subjects aged 50 or older were considered, as from this age the difficulty of speech recognition in noise starts increasing exponentially (Moore DR. et al, 2014).
- To avoid non-genetic variation, the SRT of the best ear, defined in this context as the ear with the lowest SRT value (Moore DR. et al, 2014), was considered.

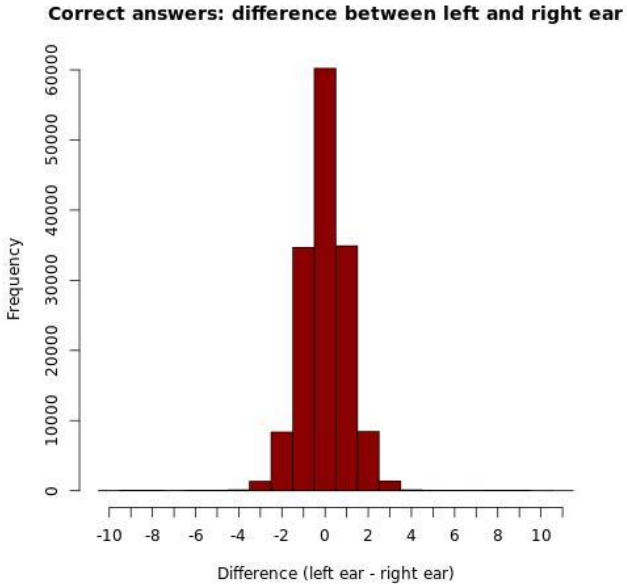


Figure 4 – Distribution of the difference in the number of correct answers: left ear minus right ear

Phenotype A showed a skewed distribution (Figure 5), and was normalised using the quantile-normalisation function of Snptest v2.5.4 (Marchini J. et al, 2010) before association.

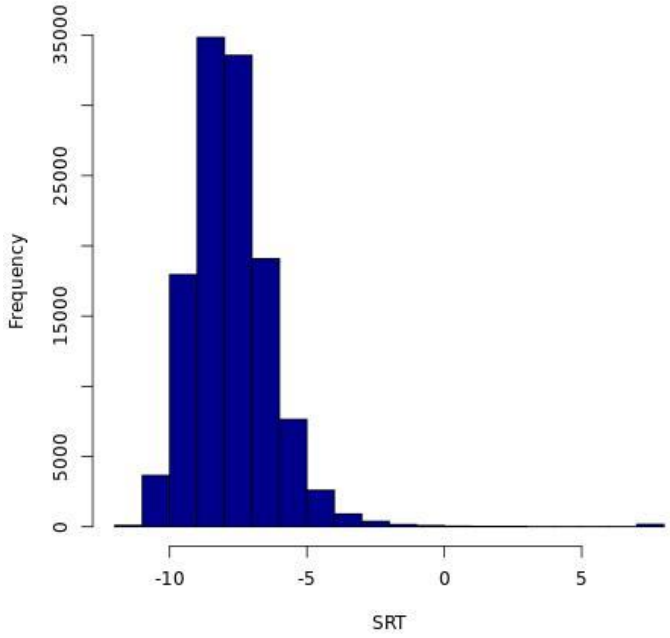


Figure 5 - Distribution of phenotype A.

With regards to phenotype B, subjects in the resulting dataset were divided into "*best hearing subjects*" and "*worst hearing subjects*" by taking those with SRT lower or equal than the first quartile of the SRT distribution and greater or equal to the third quartile respectively. The resulting binary phenotype ("*best hearing subjects*" coded with 0 and "*worst hearing subjects*" with 1) was used in the association analysis.

Association analysis.

Snptest v2.5.4 (Marchini J. et al, 2010) was used to perform association analysis on:

- the lead SNPs identified by the meta-analyses of G-EAR Consortium data. Association was tested with phenotype A for SNPs highlighted by the meta-analysis on normal hearing function traits, and with phenotype B for those coming from the meta-analysis on ARHL;
- the full set of autosomal variants from UK Biobank (phenotype B only).

The additive genetic model was assumed. For both phenotypes, association was adjusted for sex, age and the first 10 principal components. Other confounding factors that were used as covariates were the presence/absence of tinnitus, environmental risk factors (work conducted in a noisy place and the frequency of exposure to loud music) and choices made during the DTT (volume selected by the subjects).

Both after replication analyses and after the genome-wide association study, minor allele frequency (<1%) and imputation quality score (INFO <0.4) exclusion filters were applied.

With regards to the genome-wide association study results, a quantile-quantile plot was produced (Supplementary Figure 3) and the lambda coefficient calculated in order to check for inflation due to population stratification.

Annotation of lead variants.

ANNOVAR (Wang K. et al, 2010) was used to map the lead variants of the detected peaks to the nearest genes, and to determine whether the variants were contained in an intronic, intergenic or exonic region.

Assessment of the replication in UK Biobank DTT data of the signals detected by G-EAR Consortium meta-analyses.

Summary statistics of the lead SNPs of the loci highlighted by G-EAR meta-analyses on the normal hearing function were compared to the summary statistics obtained by the association analyses with phenotype A on UK Biobank data to look for significant associations in both datasets with the same direction of effect with regard to the same effect allele (in both G-EAR meta-analyses and UK Biobank GWAS, a positive effect represents a worse hearing ability); the same was done for ARHL meta-analyses in G-EAR and phenotype B in UK Biobank.

Nominal significance threshold in UK Biobank data was Bonferroni corrected for the number of variants tested for replication. Although a nominal level of significance is commonly accepted as a threshold for replication studies, this is referred to replication studies that involve identical phenotypes. As the two approaches to phenotyping here compared are different, previous studies

in literature (e.g. Nagtegaal AP. et al, 2019) in similar contexts opted for the more conservative approach that was adopted in this research.

Assessment of the replication in G-EAR Consortium meta-analyses on ARHL phenotype of the signals detected in UK Biobank DTT data.

As the genome-wide analysis on UK Biobank data was performed on phenotype B only, the replication was sought in the ARHL meta-analyses of G-EAR consortium, looking for significant associations in both datasets with the same direction of effect with regard to the same effect allele. Nominal significance threshold in G-EAR data was Bonferroni corrected for the number of variants tested for replication, for the reasons explained in the previous section.

Assessment of the replication in UK Biobank DTT data of the signals detected in UK Biobank using self-reported hearing loss data.

As an addition to the gene discovery performed with phenotype B, the possibility of shared findings with the recent association study published by Wells HRR. et al. (Wells HRR. et al, 2019) on ARHL using self-reported hearing impairment data in UK Biobank was evaluated. The replication in phenotype B of the 41 independent SNPs identified in the study was assessed, following the same procedure used for the evaluation of replication with G-EAR data. In both analyses on phenotype B and self-reported hearing loss, a positive direction of effect with regards to the effect allele implies a worse hearing ability.

RESULTS

Discovery of genes related to normal hearing function and age-related hearing loss through GWAS meta-analysis on G-EAR Consortium cohorts.

The normal hearing function was described using the single thresholds at the six frequencies collected during pure-tone audiometry (250Hz, 500Hz, 1KHz, 2KHz, 4KHz and 8KHz) and through the calculation of the pure-tone averages at low (PTAL), medium (PTAM) and high (PTAH) frequencies, which were analysed as quantitative traits. Combining all participating cohorts, the G-EAR study reached an overall sample size of about 9000 subjects, with the exact number varying slightly between traits (Supplementary Table 1). More details about subjects and phenotypes can be found in Supplementary Table 2 and Supplementary Table 3. Slight inflation was present in some cohorts (Supplementary Table 5) before genomic control; after meta-analysis and the removal of variants present in less than three populations, no inflation was detected (Supplementary Table 6).
Age-related hearing loss was investigated as a case-control phenotype instead, considering subjects aged 50 or older, and comparing those with a good perception of high frequency sounds with those with clinically relevant impairment. For this phenotype, the study reached an overall sample size of 5745 individuals (Supplementary Table 4). RS1 was not included: in this cohort, the division of the subjects in cases and controls according to the proposed definition was extremely unbalanced (only 8 controls in 652 subjects) and therefore the corresponding association study was not performed. After the removal of variants present in less than three populations, no inflation was detected (Supplementary Table 6).

The meta-analyses on the normal hearing function in G-EAR cohort succeeded in the identification of eight loci suggestively associated to hearing traits, plus two reaching genome-wide significance; two more loci were detected as suggestively associated with the ARHL case-control phenotype. These loci are detailed here below, divided by trait. A complete list of lead SNPs from each locus is reported in Table 3 and Table 4, alongside the relative association statistics.

Normal hearing function - quantitative traits.

250Hz.

A locus with p-values reaching the suggestive threshold was detected on chromosome 3 (Figure 6), in 3p12.1 (Figure 7). The lead SNP, rs7623099 ($p=3.68x10^{-7}$), lies in a region flanking *CADM2* gene, at approximately 102Kb. This gene is expressed in the mouse inner ear (gEAR, SHIELD). GCTA-COJO results showed that a single independent association is present in the locus.

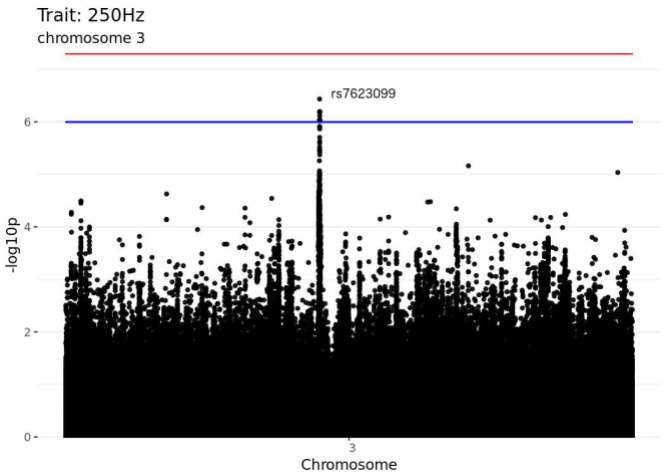


Figure 6 – Manhattan plot of the meta-analysis results on normal hearing function at 250Hz, chromosome 3. The lead SNP, rs7623099, is highlighted. The position of the variants is reported on the x-axis, while the y-axis shows –log10 of the p-values obtained from the meta-analysis. Red and blue horizontal lines mark the –log10 of genome-wide significance ($5x10^{-8}$) and suggestiveness ($1x10^{-6}$) p-value thresholds respectively.

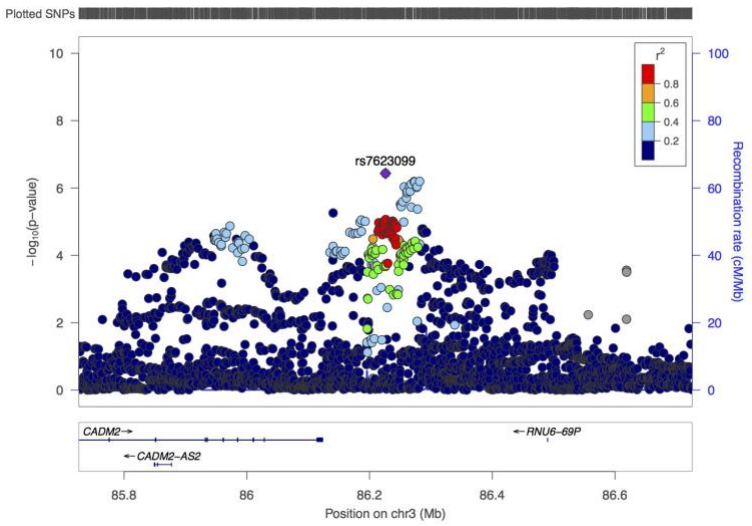


Figure 7 – Regional plot of the locus detected in 3p12.1 as associated with 250Hz. The lead SNP, rs7623099, is highlighted. The position of the variants is reported on the x-axis, alongside CADM2 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

1KHz.

Meta-analyses on 1KHz yielded five loci with suggestive p-values on chromosomes 8, 9, 10 and 11 (Figure 8).

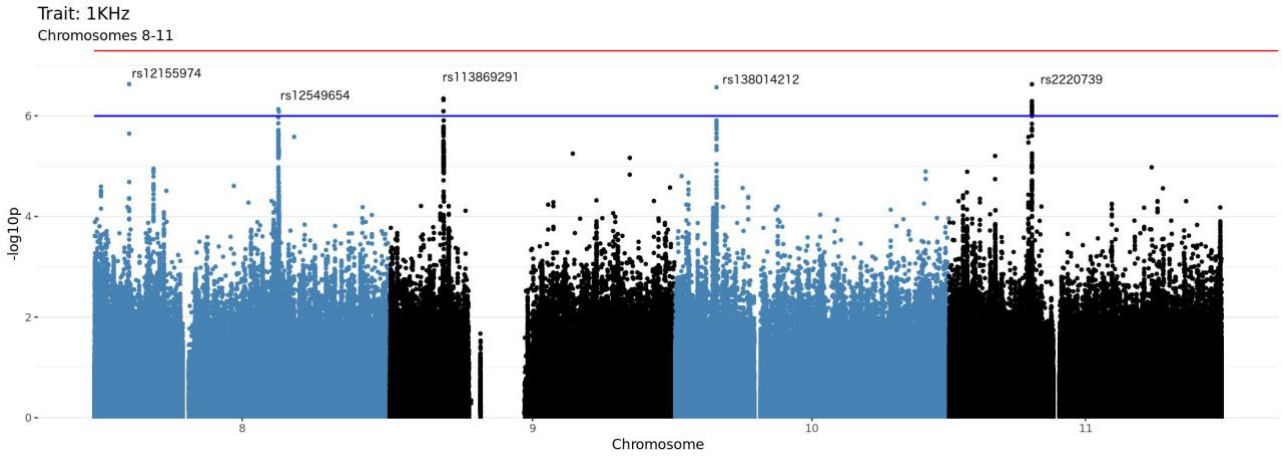


Figure 8 - Manhattan plot of the meta-analysis results on normal hearing function at 1KHz, chromosomes 8 to 11. The lead SNPs, rs12155974, rs12549654, rs113869291, rs138014212 and rs2220739 are highlighted. The position of the variants is reported on the x-axis, while the y-axis shows –log10 of the p-values obtained from the meta-analysis. Red and blue horizontal lines mark the –log10 of genome-wide significance ($5x10^{-8}$) and suggestiveness ($1x10^{-6}$) p-value thresholds respectively.

Two different loci were identified on chromosome 8. The first lies in 8p22, in an intergenic region between *MTMR7* and *SLC7A2* (Figure 9); the lead SNP, rs12155974, is the only SNP in the locus reaching a suggestive p-value (p= $2.35x10^{-7}$).

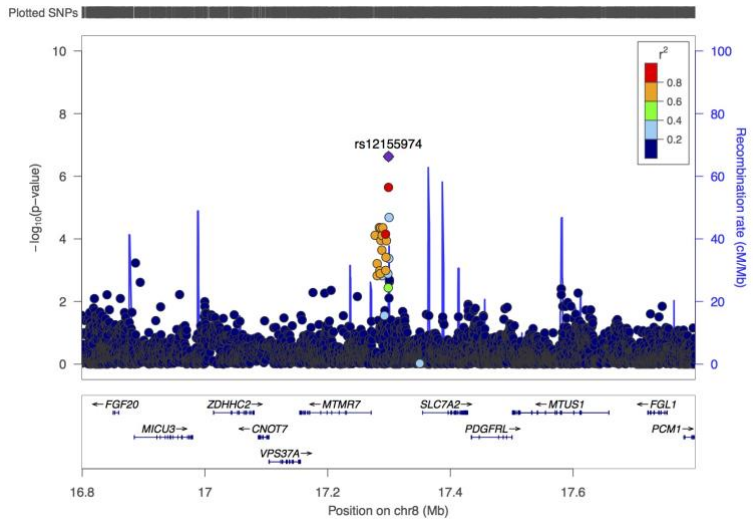


Figure 9 - Regional plot of the locus detected in 8p22 as associated with 1KHz. The lead SNP, rs12155974, is highlighted. The position of the variants is reported on the x-axis, alongside MTMR7, SLC7A2 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

The second locus is found in 8q21.3 (lead SNP: rs12549654, p= 7.39×10^{-7}), at approximately 9Kb from *CALB1* (Figure 10). This gene has a peculiar expression profile in the mouse inner ear, as it is expressed only in the cochlear hair cells (gEAR, SHIELD).

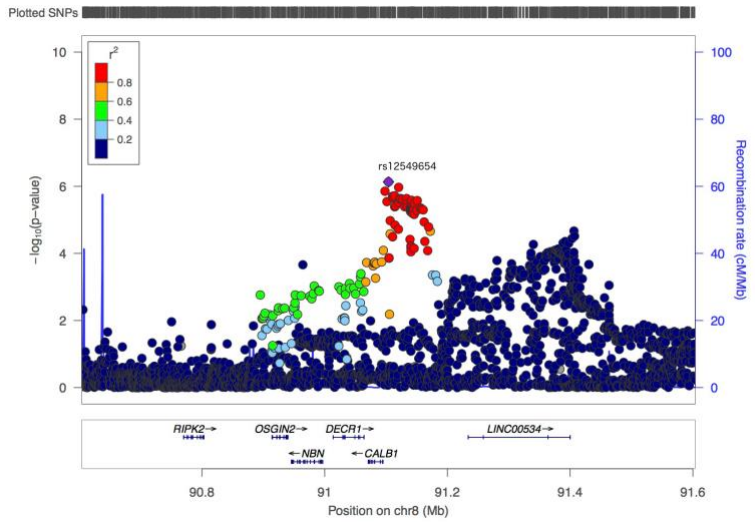


Figure 10 - Regional plot of the locus detected in 8q21.3 as associated with 1KHz. The lead SNP, rs12549654, is highlighted. The position of the variants is reported on the x-axis, alongside CALB1 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

On chromosome 9, the reported locus falls in an intergenic region in 9p21.2 (lead SNP: rs113869291, p= 4.54×10^{-7}); the closest gene is *CAAP1*, 348Kb away (Figure 11).

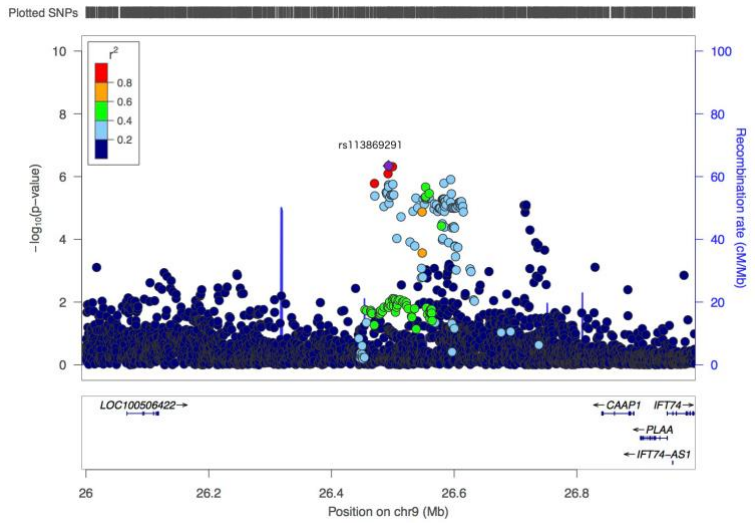


Figure 11 - Regional plot of the locus detected in 9p21.2 as associated with 1KHz. The lead SNP, rs113869291, is highlighted. The position of the variants is reported on the x-axis, alongside CAAP1 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

Regarding chromosome 10, a locus was identified in 10p12.31 in an intronic region of *PLXDC2* gene (Figure 12). A single SNP in the locus, rs138014212, showed a suggestive p-value (p= 2.71×10^{-7}).

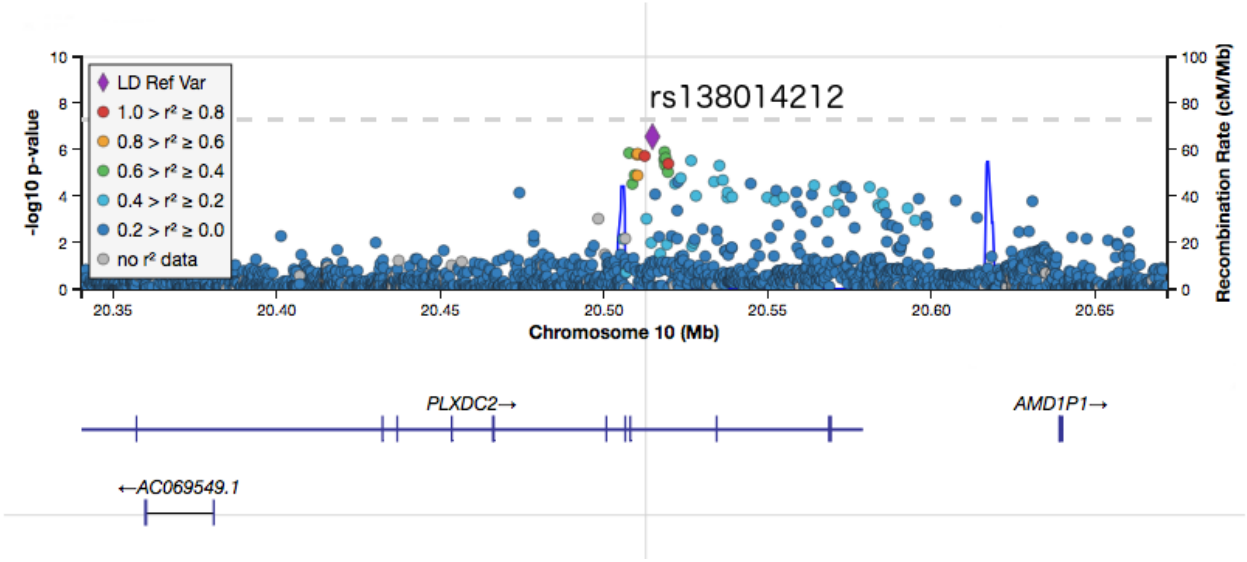


Figure 12 - Regional plot of the locus detected in 10p12.31 as associated with 1KHz. The lead SNP, rs138014212, is highlighted. The position of the variants is reported on the x-axis, alongside PLXDC2 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

One final locus was detected on chromosome 11, in 11p12, in an intronic region of *LRRC4C* gene (lead SNP: rs2220739, p= 2.36×10^{-7}) (Figure 13). *LRRC4C* is expressed in the mouse inner ear (gEAR, SHIELD).

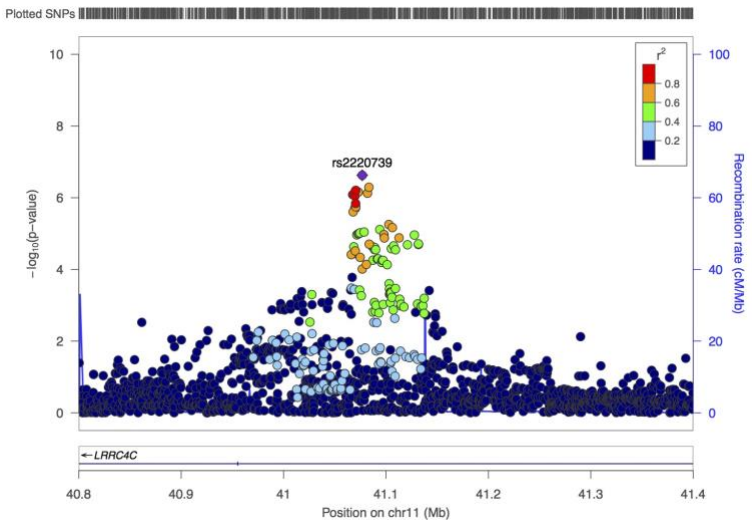


Figure 13 - Regional plot of the locus detected in 11p12 as associated with 1KHz. The lead SNP, rs2220739, is highlighted. The position of the variants is reported on the x-axis, alongside LRRC4C and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

PTAL.

Two loci suggestively associated with pure-tone averages at low frequencies were detected on chromosome 3 (Figure 14).

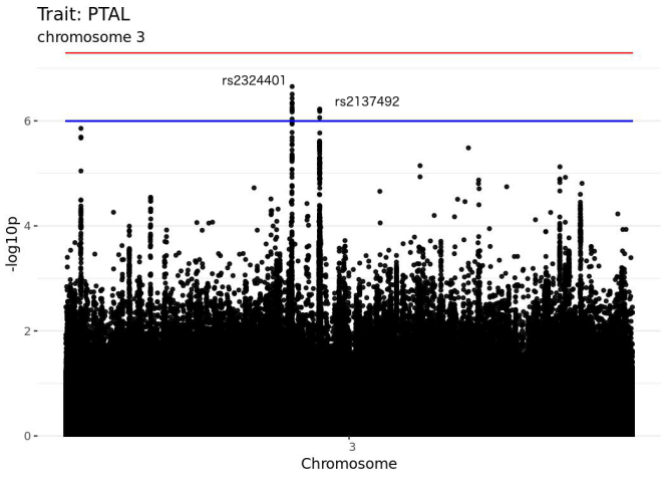


Figure 14 - Manhattan plot of the meta-analysis results on normal hearing function, PTAL, chromosome 3. The lead SNPs, rs2324401 and rs2137492, are highlighted. The position of the variants is reported on the x-axis, while the y-axis shows –log10 of the p-values obtained from the meta-analysis. Red and blue horizontal lines mark the –log10 of genome-wide significance ($5x10^{-8}$) and suggestiveness ($1x10^{-6}$) p-value thresholds respectively.

The first locus reaching suggestive p-values (lead SNP: rs2324401, $p=2.22x10^{-7}$) lies in 3p12.3, between *ZNF717* and *ROBO2* (Figure 15). *ROBO2* is highly expressed in the hair cells of both mouse and zebrafish inner ear (gEAR).

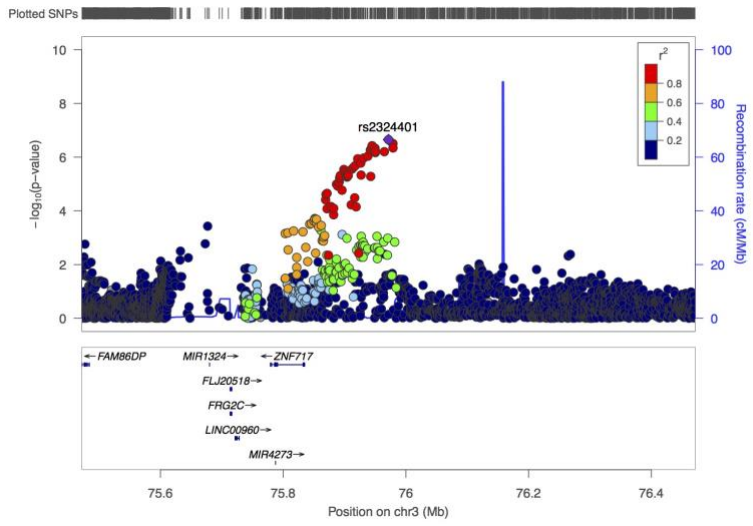


Figure 15 - Regional plot of the locus detected in 3p12.3 as associated with PTAL. The lead SNP, rs2324401, is highlighted. The position of the variants is reported on the x-axis, alongside ZNF717 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

A second peak falls in 3p12.1, 143Kb from *CADM2* (lead SNP: rs2137492, p=5.97x10^{-7}) (Figure 16). This gene, expressed in the mouse inner ear (gEAR, SHIELD), emerged also in association analyses on trait 250Hz. In this case also, GCTA-COJO analyses show that a single independent association signal is present in this region.

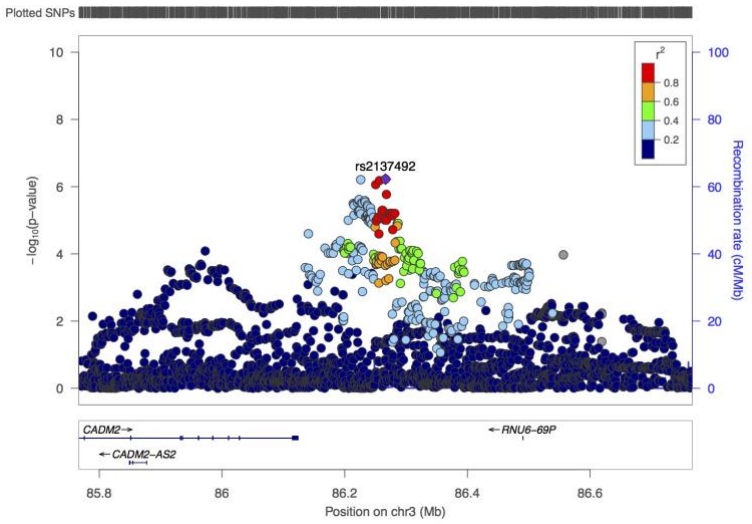


Figure 16 - Regional plot of the locus detected in 3p12.1 as associated with PTAL. The lead SNP, rs2137492, is highlighted. The position of the variants is reported on the x-axis, alongside CADM2 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

PTAM.

Results of the meta-analysis on pure-tone averages at medium traits highlighted two distinct peaks on chromosome 6, in 6p21.32, in the HLA region (Figure 17).

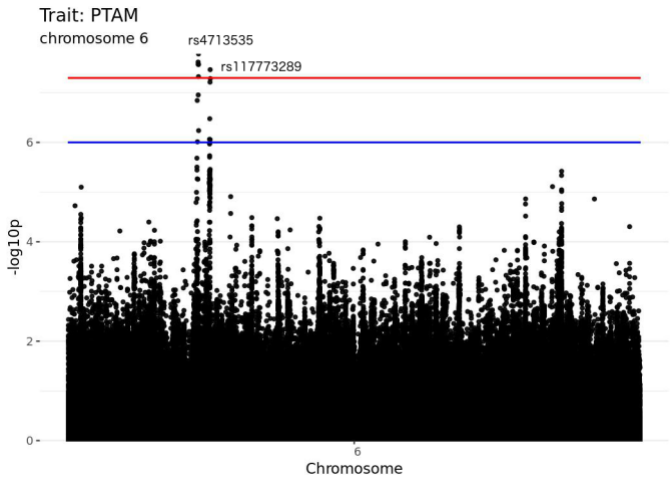


Figure 17 - Manhattan plot of the meta-analysis results on normal hearing function, PTAM, chromosome 6. The lead SNPs, rs4713535 and rs117773289, are highlighted. The position of the variants is reported on the x-axis, while the y-axis shows –log10 of the p-values

obtained from the meta-analysis. Red and blue horizontal lines mark the –log10 of genome-wide significance ($5x10^{-8}$) and suggestiveness ($1x10^{-6}$) p-value thresholds respectively.

The first peak (lead SNP: rs4713535) lies between *HLA-DRA* and *HLA-DRB5*; the second (lead SNP: rs117773289) between *HLA-DQB1* and *HLA-DQA2* (Figure 18). Both lead SNPs reach genome-wide significance ($p=1.66x10^{-8}$ and $p=3.44x10^{-8}$, respectively). GTCA-COJO results showed that the two signals are not independent.

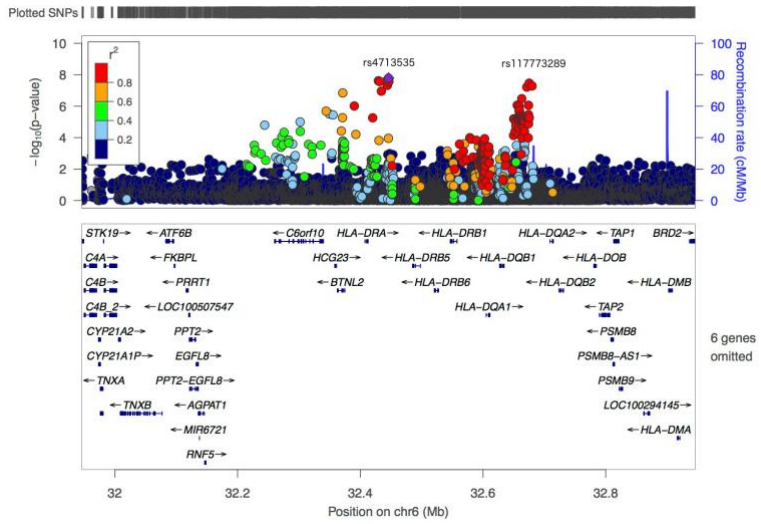


Figure 18 - Regional plot of the locus detected in 6p21.32 as associated with PTAM. The lead SNPs, rs4713535 and rs117773289, are highlighted. The position of the variants is reported on the x-axis. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

500Hz, 2KHz, 4KHz, 8KHz, PTAH.

For the remaining traits referring to the normal hearing function, the analyses did not yield any suggestive locus ($p < 10^{-6}$).

ARHL - qualitative trait.

Two loci on chromosomes 16 and 18 emerged from the meta-analyses as suggestively associated to the ARHL phenotype. The first locus was detected on chromosome 16, in 16q23.3, in an intronic region of *CDH13* (lead SNP: rs825251, $p=1.75x10^{-6}$) (Figure 19).

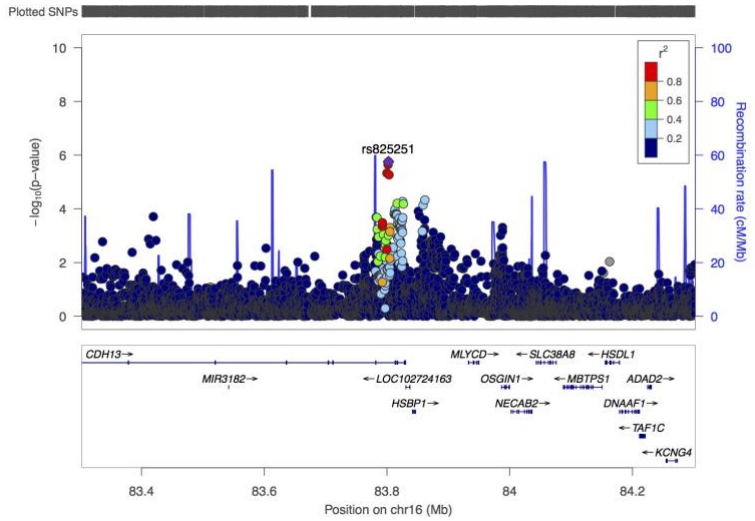


Figure 19 - Regional plot of the locus detected in 16q23.3 as associated with ARHL. The lead SNPs, rs825251, is highlighted. The position of the variants is reported on the x-axis. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

Another locus was identified on chromosome 18, in 18q21.1, in an intronic region of *CTIF* gene (lead SNP: rs55658959, p=4.20x10^{-6}) (Figure 20).

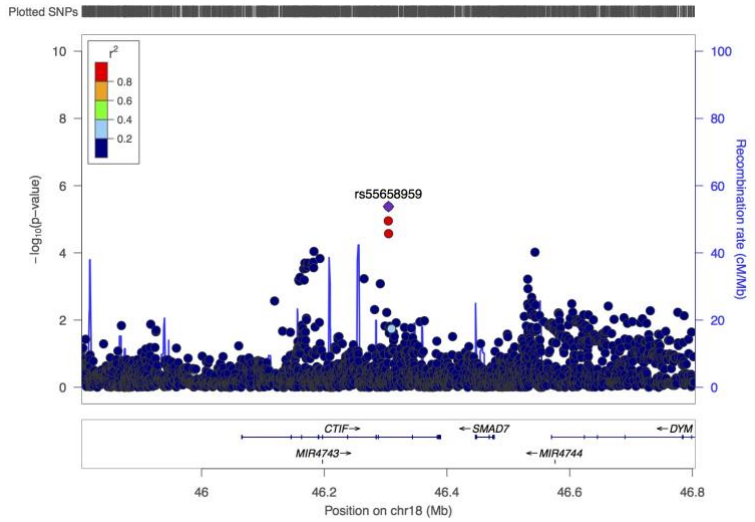


Figure 20 - Regional plot of the locus detected in 18q21.1 as associated with ARHL. The lead SNPs, rs55658959, is highlighted. The position of the variants is reported on the x-axis. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

Trait	SNP	Chr	Position (b37)	Band	EA	EAF	Beta	SE	p	Direction	Functional annotation	Nearest gene(s)	Distance (base pairs)	N	Npop
250Hz	rs7623099	3	86226002	3p12.1	A	0.2483	0.026	0.0051	3.68×10^{-7}	++++?++++	intergenic	*CADM2*	102,423	7,883	8
1KHz	rs12155974	8	17299124	8p22	C	0.0887	0.0495	0.0096	2.35×10^{-7}	++++?++++	intergenic	*MTMR7; SLC7A2*	28,264; 55,473	7,932	8
1KHz	rs12549654	8	91103947	8q21.3	T	0.0232	-0.0763	0.0154	7.39×10^{-7}	--+------	intergenic	*CALB1*	8,840	8,991	9
1KHz	rs113869291	9	26492959	9p21.2	G	0.0591	0.0475	0.0094	4.54×10^{-7}	+++-+++++	intergenic	*CAAP1*	347,724	8,991	9
1KHz	rs138014212	10	20514742	10p12.31	AT	0.1436	-0.0396	0.0077	2.71×10^{-7}	---??----	intronic	*PLXDC2*	0	7,285	7
1KHz	rs2220739	11	41076806	11p12	A	0.611	-0.0247	0.0048	2.36×10^{-7}	-??------	intronic	*LRRC4C*	0	8,309	7
PTAL	rs2324401	3	75971624	3p12.3	A	0.3143	-0.0193	0.0037	2.22×10^{-7}	---------	intergenic	*ZNF717; ROBO2*	136,924; 1,117,626	9,000	9
PTAL	rs2137492	3	86266408	3p12.1	G	0.5942	-0.0192	0.0038	5.97×10^{-7}	-??-?----	intergenic	*CADM2*	142,829	7,231	6
PTAM	rs4713535	6	32445957	6p21.32	T	0.0395	0.0736	0.013	1.66×10^{-8}	+++++???+	intergenic	*HLA-DRA; HLA-DRB5*	33,134; 39,173	4,977	6
PTAM	rs117773289	6	32675292	6p21.32	G	0.0283	0.0835	0.0151	3.44×10^{-8}	++?++???+	intergenic	*HLA-DQB1; HLA-DQA2*	40,858; 33,876	4,723	5

Table 3 - Table detailing the lead SNPs of the loci identified in the meta-analyses on normal hearing function traits. Contents of the columns: Trait: Analysed trait; SNP: rs id of the lead SNP; Chr: chromosome; Position (b37): position of the lead SNP, in base pairs (build 37); Band: cytogenetic band; EA: effect allele; EAF: frequency of the effect allele in the conjoined populations in the meta-analysis; Beta: effect from meta-analysis; SE: standard error of Beta; p: p-value of the association; Direction: string summarising the direction of the effect in the involved populations, in the fixed order FVG VBI CAR SR AWP RS1 RS2 RS3 FHS. "+" and "-" represent positive or negative effects respectively, a question mark shows that the specific variant was unavailable or had been excluded prior to meta-analysis; Functional annotation: functional annotation; Nearest gene(s): nearest gene; Distance: distance between variant and closest gene; N: overall number of subjects in the meta-analysis of the variant; Npop: number of populations involved in the meta-analysis of the variant.

Trait	SNP	Chr	Position (b37)	Band	EA	EAF	logOR	SE_logOR	p	Direction	Functional annotation	Nearest gene(s)	Distance (base pairs)	N	Npop
ARHL	rs825251	16	83802440	16q23.3	A	0.6239	0.1823	0.0381	1.75×10^{-6}	+++++++++	intronic	*CDH13*	0	5,745	9
ARHL	rs55658959	18	46304662	18q21.1	G	0.1177	-0.2924	0.0635	4.20×10^{-6}	---------	intronic	*CTIF*	0	5,745	9

Table 4 - Table detailing the lead SNPs of the loci identified in the meta-analyses on ARHL phenotype. Contents of the columns: Trait: Analysed trait; SNP: rs id of the lead SNP; Chr: chromosome; Position (b37): position of the lead SNP, in base pairs (build 37); Band: cytogenetic band; EA: effect allele; EAF: frequency of the effect allele in the conjoined populations in the meta-analysis; logOR: log-odds from the meta-analysis; SE: standard error of the logOR; p: p-value of the association; Direction: string summarising the direction of the effect in the involved populations, in the fixed order FVG VBI CAR SR AWP PM RS2 RS3 FHS. "+" and "-" represent positive or negative effects respectively, a question mark shows that the specific variant was unavailable or had been excluded prior to meta-analysis; Functional annotation: functional annotation; Nearest gene(s): nearest gene; Distance: distance between variant and closest gene; N: overall number of subjects in the meta-analysis of the variant; Npop: number of populations involved in the meta-analysis of the variant.

Replication in UK Biobank DTT data of the signals detected by G-EAR Consortium meta-analysis; discovery of genes in UK Biobank DTT data.

Phenotypes A and B.

The preliminary step in both the replication of signals from G-EAR Consortium and the gene discovery in UK Biobank data was the construction of phenotypes A - which includes the whole variation of the hearing ability as recorded by the DTT - and B – which compares the extremes of the recognition ability in the middle-aged and older, in a potential case-control situation for age-related hearing loss.

After the application of the inclusion/exclusion criteria detailed in the *Materials and Methods* chapter, ***phenotype A*** involved a total of 121,385 subjects, whose age and SRT ranges can be found in Table 5. A distribution plot was presented in Figure 5.

N	121,385 (56,390 M; 64,995 F)
Age range	40-70 (mean=57.17)
SRT range	-11.5 - 8 (mean= -7.48)

Table 5 - Phenotype A.

Phenotype A was then quantile-normalised before association.

Phenotype B involved 54,385 subjects divided in “best hearing” and “worst hearing” as in Table 6.

	Best hearing subjects	**Worst hearing subjects**
N (tot=54,385)	26,638 (12,482 M; 14,156 F)	27,747 (13,043 M; 14,704 F)
Age range	50-70 (mean=58.99)	50-70 (mean=61.86)
SRT range	-11.25 - -8.5 (mean=-8.97)	-6.5 - 8 (mean=-5.55)

Table 6 - Phenotype B.

Assessment of the replication in UK Biobank DTT data of the signals detected by G-EAR Consortium meta-analyses.

Of the ten lead SNPs from G-EAR meta-analyses on the normal hearing function, only rs12549654 was associated with phenotype A at a nominal level (p=0.0476), and showed the same direction of effect. The same direction of effect was observed with rs825251, but both lead SNPs from ARHL analyses failed to reach nominal significance in the corresponding association in UK Biobank with phenotype B (Table 7).

After Bonferroni correction (new thresholds: p=0.05/10=0.005 and p=0.05/2=0.025 for associations run on phenotypes A and B respectively), no SNP from G-EAR analyses reached the new significance threshold in UK Biobank.

Trait	SNP	Chr	Position (b37)	EA	Effect	SE	p	Nearest gene(s)	UKBB pheno	UKBB p	UKBB effect	UKBB se
250Hz	rs7623099	3	86226002	A	0.026	0.0051	3.68×10^{-7}	*CADM2*	A	0.0381	-0.0087	0.0042
1KHz	rs12155974	8	17299124	C	0.0495	0.0096	2.35×10^{-7}	*MTMR7; SLC7A2*	A	0.5849	0.0047	0.0087
1KHz	rs12549654	8	91103947	T	-0.0763	0.0154	7.39×10^{-7}	*CALB1*	A	0.0476	-0.0260	0.0131
1KHz	rs113869291	9	26492959	G	0.0475	0.0094	4.54×10^{-7}	*CAAP1*	A	0.5802	-0.004	0.0076
1KHz	rs138014212	10	20514742	AT	-0.0396	0.0077	2.71×10^{-7}	*PLXDC2*	A	0.2047	0.0079	0.0063
1KHz	rs2220739	11	41076806	A	-0.0247	0.0048	2.36×10^{-7}	*LRRC4C*	A	0.0774	0.007	0.0039
PTAL	rs2324401	3	75971624	A	-0.0193	0.0037	2.22×10^{-7}	*ZNF717; ROBO2*	A	0.6613	0.002	0.0043
PTAL	rs2137492	3	86266408	G	-0.0192	0.0038	5.97×10^{-7}	*CADM2*	A	0.2466	0.004	0.0039
PTAM	rs4713535	6	32445957	T	0.0736	0.013	1.66×10^{-8}	*HLA-DRA; HLA-DRB5*	A	0.5773	-0.006	0.0115
PTAM	rs117773289	6	32675292	G	0.0835	0.0151	3.44×10^{-8}	*HLA-DQB1; HLA-DQA2*	A	0.8749	-0.002	0.0140
ARHL	rs825251	16	83802440	A	0.1823	0.0381	1.75×10^{-6}	*CDH13*	B	0.8869	0.0034	0.0242
ARHL	rs55658959	18	46304662	G	-0.2924	0.0635	4.20×10^{-6}	*CTIF*	B	0.6639	0.0175	0.0403

Table 7 – Contents of the columns: Trait: Analysed trait in G-EAR meta-analyses; SNP: rs id of the lead SNP; Chr: chromosome; Position (b37): position of the lead SNP, in base pairs (build 37); EA: effect allele; Effect: beta from the meta-analysis on NHF traits (250Hz, 1KHz, PTAL and PTAM), logOR from meta-analysis on ARHL; SE: standard error of the effect; p: p-value of the association in G-EAR; Nearest gene(s): nearest gene; UKBB pheno: phenotype used in association study on UK Biobank data; UKBB p: p-value of the association in UK Biobank; UKBB effect: beta from linear mixed model if UKBB pheno=A, logOR from logistic regression if UKBB pheno=B; UKBB se: standard error of UKBB effect.

Gene discovery on UK Biobank DTT data.

Running a GWAS analysis on phenotype A was ultimately too computationally intensive for the available resources, and was therefore not performed; investigation was limited to phenotype B. Genotype-phenotype association was tested on 93,879,879 autosomal variants; after filtering, 10,504,478 remained. No inflation was detected (lambda=0.989; see also Supplementary Figure 3).

As highlighted in the Manhattan plot below (Figure 21), three loci reached suggestive ($p < 1x10^{-6}$) or genome-wide significant ($p < 5x10^{-8}$) p-values, two of them on chromosome 2 and the last on chromosome 19. Lead SNPs from each locus and their association statistics can be found in Table 8.

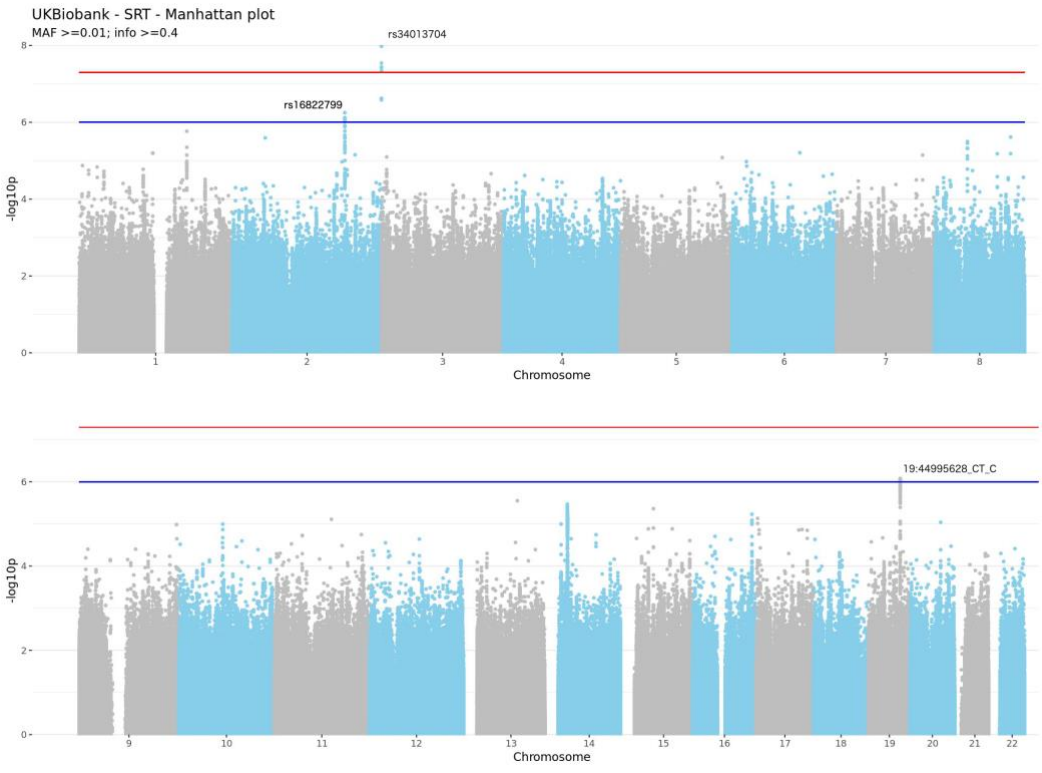


Figure 21 - Manhattan plot of the results of the GWAS performed on phenotype B. The position of the variants is reported on the x-axis, while the y-axis shows –log10 of the p-values obtained from the meta-analysis. Red and blue horizontal lines mark the –log10 of genome-wide significance ($5x10^{-8}$) and suggestiveness ($1x10^{-6}$) p-value thresholds respectively.

A locus with genome-wide significant p-values was identified in 2q37.3, in an intron region within the *D2HGDH* gene (Figure 22). The lead SNP in this locus is rs34013704 ($p=9.8x10^{-9}$).

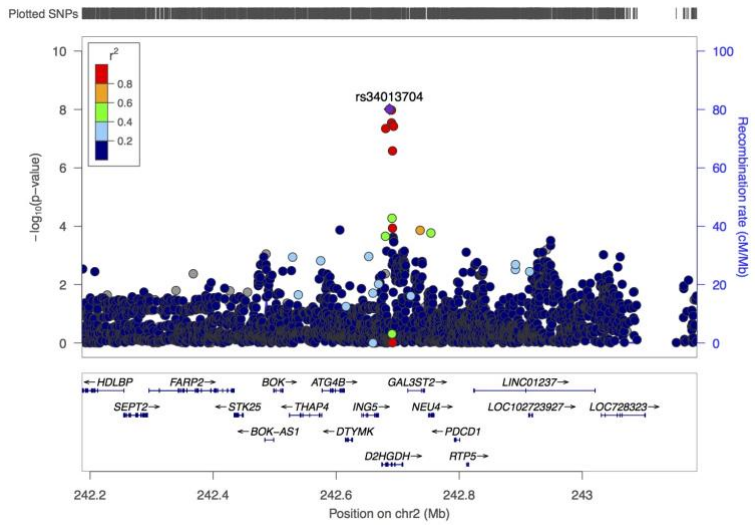


Figure 22 - Regional plot for the identified locus in 2q37.3. The lead SNP of the region, rs34013704, is highlighted. The position of the variants is reported on the x-axis, alongside D2HGDH and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

A locus reaching the suggestive threshold was identified on chromosome 2, in 2q32.1, within an intron region of *PDE1A* gene. The lead SNP in this locus is rs16822799 ($p=5.62x10^{-7}$) (Figure 23).

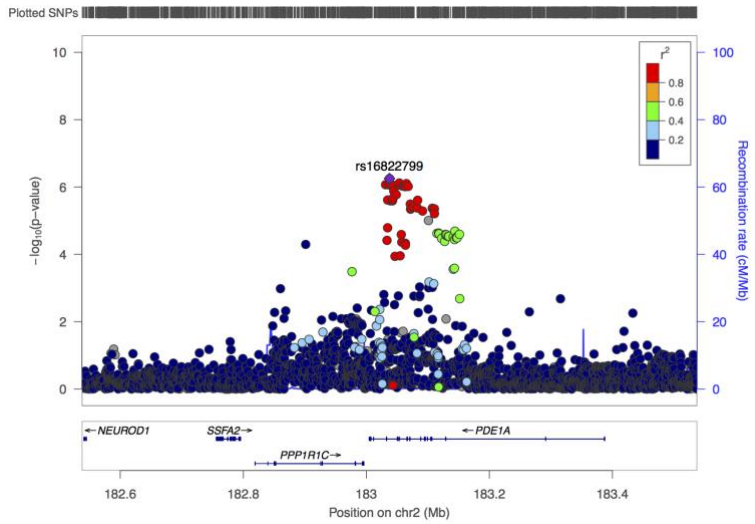


Figure 23 - Regional plot for the identified locus in 2q32.1. The lead SNP of the region, rs16822799, is highlighted. The position of the variants is reported on the x-axis, alongside PDE1A and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis. SNPs showing strong linkage disequilibrium with rs16822799 are highlighted in red.

Finally, a third locus, suggestively associated, was detected on chromosome 19, in 19q13.31. The lead SNP 19:44995628_CT_C ($p=8.38x10^{-7}$) falls within an intergenic region between *ZNF229* and *ZNF180* (Figure 24).

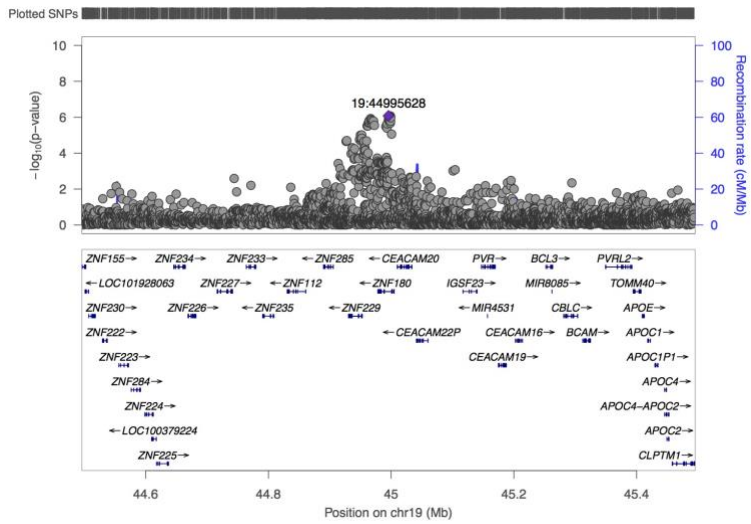


Figure 24 - Regional plot for the identified locus in 19q13.31. The lead SNP of the region, 19:44995628_CT_C, is highlighted. The position of the variants is reported on the x-axis, alongside ZNF180, ZNF229 and the other genes in the region. The y-axis shows –log10 of the p-values obtained from the meta-analysis.

SNP	Chr	Position (b37)	EA	EAF	logOR	SE	p	INFO	Nearest gene(s)	Distance (base pairs)
rs16822799	2	183037670	A	0.183	0.1487	0.0297	5.62×10^{-7}	0.999	*PDE1A*	0
rs34013704	2	242686629	A	0.0654	-0.2701	0.0471	9.80×10^{-9}	0.982	*D2HGDH*	0
19:44995628_CT_C	19	44995628	C	0.226	0.1360	0.0276	8.38×10^{-7}	0.986	*ZNF229; ZNF180*	11015; 14883

Table 8 - Table detailing the lead SNPs of the three identified loci. Contents of the columns: SNP: the identifier of the variant (rs id, when available); Chr: chromosome; Position: position of the variant, in base pairs, referred to build 37; EA: effect allele; EAF: frequency of the effect allele; logOR: logarithm of the odds ratio; SE: standard error of logOR; p=p-value obtained from the association study; INFO: imputation quality score of the variant; Nearest gene(s): nearest gene(s); Distance: distance of the variant from the nearest gene(s), in base pairs.

Assessment of the replication in G-EAR Consortium meta-analyses on ARHL phenotype of the signals detected in UK Biobank DTT data.

Of the three lead SNPs highlighted by the GWAS on phenotype B in UK Biobank data, two - rs16822799 and rs34013704 – showed the same direction of effect with regards to the effect allele.

None was significant in G-EAR ARHL analyses at the Bonferroni corrected threshold (p=0.05/3=0.017) (Table 9).

Assessment of the replication in UK Biobank DTT data of the signals detected in UK Biobank using self-reported hearing loss data.

After Bonferroni correction, the new significance threshold was set at p=0.05/41=~0.0012. Of the 41 reported SNPs reported in Wells HRR et al (2019), none was replicated in phenotype B: 15 showed the same direction of effect, but none of them was significant according to the corrected threshold (Table 10).

SNP	Chr	Position (b37)	EA	logOR	SE	p	Nearest gene(s)	G-EAR trait	G-EAR p	G-EAR logOR	G-EAR se
rs16822799	2	183037670	A	0.1487	0.0297	5.62×10^{-7}	*PDE1A*	ARHL	0.7024	0.0186	0.0486
rs34013704	2	242686629	A	-0.2701	0.0471	9.80×10^{-9}	*D2HGDH*	ARHL	0.9173	-0.0147	0.1414
19:44995628_CT_C	19	44995628	C	0.136	0.0276	8.38×10^{-7}	*ZNF229; ZNF180*	ARHL	0.3024	-0.0852	0.0826

Table 9 - Contents of the columns: SNP: lead SNP from Phenotype B analyses; Chr: chromosome; Position (b37): position of the lead SNP, in base pairs (build 37); EA: effect allele; logOR: logOR from logistic regression; SE: standard error of logOR; p: p-value of the association in UK Biobank; Nearest gene(s): nearest gene; G-EAR trait: corresponding trait in G-EAR analyses; G-EAR p: p-value of G-EAR meta-analysis; G-EAR logOR: logOR from logistic regression from G-EAR meta-analyses; G-EAR se: standard error of G-EAR logOR.

SNP	Chr	EA	EAF	Effect	SE	p	SRT Effect	SRT SE	SRT p	Nearest Gene(s)	Distance (base pairs)
rs36062310	22	G	0.96	0.0315	0.003	1.90×10^{-22}	-0.041344331	0.057527735	0.472333	KLHDC7B	0
rs6453022	5	C	0.5	0.0126	0.001	1.70×10^{-21}	-0.006788665	0.023104213	0.768889	ARHGEF28	0
rs759016271	6	AGTAGTCCACTTTTCTTCTTTGCCTG	0.39	0.0127	0.001	6.10×10^{-21}	-0.000813973	0.023169719	0.971975	ZNF318	0
rs6890164	5	A	0.51	0.0119	0.001	3.30×10^{-19}	0.003776664	0.023072321	0.869977	ARHGEF28	6177
rs7951935	11	G	0.62	0.0114	0.001	7.80×10^{-17}	0.037443994	0.023614923	0.112828	TYR	1472
rs35186928	6	G	0.62	0.0109	0.001	1.70×10^{-15}	0.000222276	0.023948584	0.992595	HLA-DQA1	13352
rs9493627	6	G	0.68	0.0104	0.001	1.40×10^{-13}	-0.005448535	0.024457742	0.823712	EYA4	0
rs132929	22	G	0.59	0.0098	0.001	2.20×10^{-13}	-0.003636733	0.023352149	0.876242	BAIAP2L2	0
rs5756795	22	T	0.54	0.0092	0.001	5.10×10^{-12}	-0.03557065	0.023136344	0.124186	TRIOBP	0
rs1566129	14	T	0.41	0.0091	0.001	1.40×10^{-11}	-0.039739756	0.023331655	0.0885204	NID2	0
rs35414371	4	T	0.87	0.0131	0.002	1.60×10^{-11}	-0.009214592	0.034060559	0.786749	CLRN2	1965
3:182069497_TA_T	3	TA	0.84	0.0118	0.002	4.10×10^{-11}	-0.044710035	0.031602036	0.157131	ATP11B	441791
rs12225399	11	G	0.65	0.009	0.001	8.60×10^{-11}	0.025278764	0.024186575	0.29595	PHLDB1	0
rs55635402	11	A	0.81	0.0105	0.002	2.90×10^{-10}	-0.013794648	0.029257187	0.637286	TUB	0
rs62033400	16	A	0.61	0.0085	0.001	2.90×10^{-10}	-0.0312142	0.023552562	0.185072	FTO	0
rs13277721	8	G	0.49	0.0083	0.001	3.30×10^{-10}	-0.015672883	0.02306996	0.496908	AGO2	0
rs62188635	2	C	0.45	0.0083	0.001	4.70×10^{-10}	-0.018332977	0.023104038	0.427488	KLF7	50519
rs2236401	6	C	0.49	0.0081	0.001	9.30×10^{-10}	-0.018857555	0.023087402	0.414048	SYNJ2	0
rs4947828	7	T	0.23	0.0096	0.002	1.00×10^{-9}	-0.028725176	0.027673033	0.299261	GRB10	0
rs6597883	10	T	0.84	0.0111	0.002	1.00×10^{-9}	0.014509209	0.031238661	0.642316	CTBP2	0

rs34442808	5	T	0.49	0.008	0.001	1.30×10^{-9}	-0.023030795	0.023030835	0.317311	MCTP1, SLF1	0
rs835267	10	A	0.53	0.008	0.001	1.60×10^{-9}	-0.050270222	0.022852509	0.0278231	EXOC6	0
rs4948502	10	T	0.57	0.0081	0.001	1.70×10^{-9}	0.008455808	0.023434566	0.71823	ARID5B	0
rs10824108	10	G	0.42	0.0079	0.001	3.00×10^{-9}	-0.043630759	0.023312528	0.0612675	ADK	0
rs12027345	1	G	0.57	0.0079	0.001	3.60×10^{-9}	0.005852993	0.023291321	0.801586	MAST2	12668
rs217289	6	G	0.56	0.0078	0.001	4.90×10^{-9}	0.0124895	0.023233058	0.59087	SNAP91	0
rs13093972	3	A	0.55	0.0078	0.001	5.50×10^{-9}	0.028609335	0.023188082	0.217279	ZBTB20	121137
rs62015206	15	C	0.41	0.0078	0.001	7.70×10^{-9}	-0.062078499	0.023608132	0.0085502	MAPK6	15613
rs10475169	5	A	0.88	0.0117	0.002	9.30×10^{-9}	-0.023651494	0.035130723	0.500794	IRX2	190445
rs17671352	17	T	0.38	0.0078	0.001	1.00×10^{-8}	0.025159405	0.023583991	0.286062	ACADVL	0
rs7525101	1	C	0.56	0.0075	0.001	1.50×10^{-8}	-0.027551409	0.02308572	0.232697	LMX1A	61973
rs12938775	17	G	0.5	0.0075	0.001	1.60×10^{-8}	-0.01482715	0.023030837	0.519708	PAFAH1B1	0
rs76837345	8	A	0.93	0.0146	0.003	1.90×10^{-8}	0.024643969	0.045074045	0.584554	CHMP4C	0
rs9366417	6	G	0.26	0.0085	0.002	2.10×10^{-8}	0.006263724	0.026087518	0.81025	SOX4	291019
rs3890736	8	G	0.63	0.0077	0.001	2.20×10^{-8}	-0.020189813	0.023890263	0.398051	GFRA2	15676
rs143282422	10	G	0.99	0.0349	0.006	2.40×10^{-8}	-0.104442316	0.110044578	0.342573	CDH23	0
rs9691831	7	A	0.42	0.0074	0.001	3.10×10^{-8}	-0.021760163	0.023264072	0.349605	TMEM213	0
rs141403654	11	A	0.98	0.0313	0.006	3.50×10^{-8}	0.10442584	0.100195977	0.297312	AGBL2	0
rs4611552	18	T	0.78	0.0089	0.002	3.60×10^{-8}	0.003301843	0.028061458	0.906333	CCDC68	9362
rs12552	13	A	0.44	0.0073	0.001	4.80×10^{-8}	0.004938528	0.023267879	0.831915	OLFM4	0
rs10927035	1	C	0.35	0.0075	0.001	4.90×10^{-8}	-0.022179747	0.023681298	0.348968	AKT3	0

Table 10 - Contents of the columns: SNP: lead SNP from Wells HRR et al (2019); Chr: chromosome; EA: effect allele; EAF: frequency of the effect allele; Effect: effect size from Wells HRR et al; SE: standard error of Effect; p: p-value of the association in Wells HRR et al; SRT Effect: logOR from association with Phenotype B; SRT SE: standard error of SRT Effect; SRT p: p-value from association with Phenotype B; Nearest gene(s): nearest gene; Distance: distance of the variant from the nearest gene(s), in base pairs.

DISCUSSION

In the last years, the search for candidate genes for complex traits through the detection of association signals in GWAS (meta-)analyses and the replication of the association in an independent cohort, with subsequent *in vitro* and *in vivo* validation, has led to several advancements in the comprehension of the mechanisms behind these traits. As highlighted in the introduction, the study of hearing has benefited of this strategy, but despite the successes there is still need for a deeper understanding of the genetic bases of hearing traits and diseases.

In this thesis, I had the opportunity to conduct a large study on hearing traits/diseases and perform gene discovery using in parallel data from two projects, which make use of two different phenotyping approaches towards the study of hearing traits:

- G-EAR Consortium, with overall 9000 individuals coming from different cohorts with deep audiometric characterisation, making this one of the largest meta-analyses performed on pure-tone audiometry data;
- UK Biobank, a large population-based study where the hearing ability of approximately 200,000 volunteers was assessed through Digit Triplet Test, for which, to our knowledge, there is no other GWAS available in literature;

and, as a second objective, assess if the results from one dataset are replicated in the other and vice versa.

In both contexts, the exploration of the results highlighted some genes for which literature and public databases with expression data on hearing cells (i.e. gEAR (https://umgear.org) and SHIELD (Shen J. et al., 2015)) provide interesting evidence supporting their potential involvement in the hearing system. In particular, *CADM2*, *CALB1*, *LRRC4C* and *ROBO2* in the G-EAR Consortium and *PDE1A* in the UK Biobank are the most promising candidates.

The first interesting candidate named *CADM2* has been detected as being associated with low frequency hearing traits (250Hz and PTAL).

CADM2 encodes the cell adhesion molecule 2 (also known as SynCAM2, Igsf4d, and Nectin-like molecule 3), which belongs to the family of nectin-like molecules (Necl), that are $Ca2^{+}$-independent

immunoglobulin superfamily cell adhesion molecules, expressed in most cell types (Mandai K. et al, 2015). Necl family is composed of five members that together with Nectins mediate the formation of cell adherens junctions by forming homo- or hetero-dimers (*ibidem*).

In particular, Necls and Nectins are involved in several biological processes, such as spermatogenesis, axon guidance, synapse formation, and myelination, and in the organogenesis of the eye, tooth, cerebral cortex and inner ear (*ibidem*). Regarding their involvement in the hearing system, it has been shown that the trans-interaction between nectin-1 and -3 mediates the heterotypic adhesion between the hair cells and supporting cells and establishes the checkerboard-like cellular pattern (*ibidem*). Moreover, aside from being essential for the correct development of the inner ear, these molecules might exert an important role also in the maintenance of the hearing function. As an example, it has been demonstrated that *CADM1* is abundantly expressed in mouse inner ear, and it is a binding partner of *PDZD7*, a ciliary protein involved in Usher syndrome (MIM# 605472) and non-syndromic hearing loss (MIM# 618003) (Du H. et al., 2018). So far, *CADM2* has been mainly associated to axon guidance processes. However, considering a) the homology between the different Necs and Necls members, and that b) according to data collected both from public databases (i.e. gEAR and SHIELD) and from recent literature updates (Du H. et al., 2018) the gene is expressed in the mouse inner ear, it is possible to assume that it might play a role in the hearing system and in hearing impairment, probably mediating cell to cell interaction.

Another interesting finding regards *CALB1* gene, which encodes the calbindin protein, a member of the calcium-binding protein superfamily that includes calmodulin and troponin C and has been detected here as associated with 1KHz hearing threshold. *CALB1* contains four active calcium-binding domains plus two modified domains that are unable to bind calcium, and is thought to regulate calcium influx following the activation of glutamate receptors (Cao LQ. et al., 2019). So far, genetic mutations in *CALB1* gene have been observed in patients with Huntington disease (Massouh M. et al, 2008), however, its role in other human diseases still need to be defined. Several literature data support the hypothesis of an involvement of *CALB1* in the hearing system. In fact, interestingly, the gene shows a peculiar expression profile in the mouse inner ear, since it is expressed only in the cochlear hair cells, as shown by the gEAR and SHIELD databases. Moreover, it has been demonstrated that *CALB1* is expressed in the mouse auditory and vestibular systems during embryonic and early postnatal development regulating the development of cochlear hair cells, their afferent nerve fibres, and of the vestibulocochlear ganglion (Buckiová D. et al., 2009). Despite Calb1 knock out mice do not display hearing loss (Airaksinen L. et al., 2000), quite recently it has been

shown that calbindin levels are upregulated in the spiral ganglion neurons of ageing mice affected by age-related hearing loss (i.e. C57BL/6J and 129/SvJ mice) and it is assumed that this protects cells from Ca2+ overload and apoptosis (Fischer N. et al., 2019). Based on this evidence, CALB1 turns out to be an extremely interesting candidate for the regulation of the normal hearing function and the etiopathogenesis of age-related hearing loss.

A suggestive association was identified between *LRRC4C* and the normal hearing function at 1KHz. *LRRC4C* encodes the transmembrane protein NGL-1, which is one member of a highly specialized family of neuronal guidance molecules (Maussion G. et al., 2017). In particular, NGL-1 and its family member protein NGL-2 are thought to promote excitatory synapse development through largely non-overlapping neuronal pathways, even though the exact function of NGL-1 is still unclear (Choi Y. et al., 2019). *LRRC4C* is abundantly expressed in mouse inner ear, in particular in the surrounding cells (gEAR database), and it is reported to bind to two of the three PDZ domains of whirlin (Delprat B. et al., 2005). Whirlin is a protein transported to the tips of stereocilia by myosin XVa and when mutated causes nonsyndromic, congenital, profound hearing loss (Shaikh RS. et al. 2005). It has been proposed that the interaction between whirlin and NGL-1 might be involved in the stabilization of interstereociliar links (Delprat B. et al., 2005), thus it is reasonable to think that NGL-1 might be involved in the maintenance of normal hearing or in complex hearing diseases.

ROBO2 encodes a conserved axon-binding receptor. Robo receptors and their Slit ligands (which are secreted chemorepellent proteins) are conserved from fly to human and play a key role in axon guidance and cell migration (St Pourcain B. et al., 2014). *ROBO2* is highly expressed in the hair cells of both mouse and zebrafish inner ear (gEAR database) and seems to be involved in controlling the spatial positioning of Spiral Ganglion Neuron (SGN) cell bodies in the mouse cochlear development (Wang SZ. et al., 2013). SGNs are the bipolar afferent neurons in the medial surface of the cochlear duct that run in parallel with hair cells in the cochlear sensory epithelium. It has been demonstrated that Slit/Robo signaling plays an essential role in the correct assembly and positioning of SGN cell bodies, which is critical for their afferent axons to navigate to hair cells without errors. In particular, in *Robo2* mutants the neurites of the individually mispositioned neurons travel randomly and appear not to innervate hair cells, leading to an imprecise cochlear innervation (Wang SZ. et al., 2013).

This data, together with the identification of an association of *ROBO2* with the pure-tone average at low frequencies (PTAL), reinforce the hypothesis of the involvement of *ROBO2* in the maintenance of the correct hearing function.

PDE1A encodes the cyclic nucleotide phosphodiesterases (PDE) 1A, that belong to the PDEs family. PDEs are known for playing a role in signal transduction by regulating intracellular cyclic nucleotide concentrations through hydrolysis of cAMP and/or cGMP to their respective nucleoside 5-prime monophosphates. Members of the PDE1 family, such as *PDE1A*, are Ca(2+)/calmodulin (see *CALM1*; 114180)-dependent PDEs (CaM-PDEs) that are activated by calmodulin in the presence of Ca(2+) (Michibata H. et al., 2001; Fidock M. et al., 2002). In mouse model, Pde1a is involved in the renal pathogenesis of autosomal dominant polycystic kidney disease and in the regulation of blood pressure (Wang X. et al., 2017), but its possible role in the inner ear has not been described so far. Nevertheless, it has been demonstrated that various PDEs are expressed in the inner ear, and it has been suggested that they might play a crucial role in the auditory system (Mittal R. et al., 2017).
Furthermore, recently *PDE1C*, another member of the PDEs family, has been identified as causative of progressive, postlingual autosomal dominant nonsyndromic hearing loss in a five-generation Chinese family (Wang L. et al., 2018). Here, we identified an association of *PDE1A* with phenotype B in UK Biobank, that together with literature data, might highlight the possible involvement of this gene in the auditory system, encouraging further study to precisely assess its role in the inner ear.

Across the two analyses, apart from the relevant genes above described, three genome-wide significant loci were detected. In G-EAR Consortium analyses, we identified a genome-wide significant statistical association between PTAM and two loci in the HLA region, located between *HLA-DQB1* and *HLA-DQA2*, and *HLA-DRA* and *HLA-RDB5* respectively. In UK Biobank, a genome-wide significant association was detected between phenotype B and *D2HGDH*.

The human leukocyte antigen (HLA) complex is a gene complex that encodes the major histocompatibility complex proteins in humans, which are cell-surface glycoproteins responsible for the regulation of the immune system. The HLA genes are located on chromosome 6 and are highly polymorphic, ensuring a fine-tuning regulation of the adaptive immune system. They fall into two classes, class I (HLA-ABC) and class II (HLA-DR, DP, DQ), which are structurally and functionally different, in addition to have a distinct distribution among cells (HLA class I molecules are expressed on the surface of nearly all nucleated cells, while HLA class II molecules are expressed on antigen presenting cells) (Aptsiauri N. et al., 2011).
The protein encoded by the *HLA-DQB1* gene binds to the protein produced from *HLA-DQA1*, forming a functional protein complex called an antigen-binding DQαβ heterodimer. This complex displays

foreign peptides to the immune system to trigger the body's immune response (Megiorni F. et al., 2012). Polymorphisms of *HLA-DQB1* have been associated to the predisposition to various diseases, including diabetes, celiac disease, multiple sclerosis, etc. (Corper AL. et al., 2000; Megiorni F. et al., 2012; Oksenberg JR. et al., 2004).

On the other hand, *HLA-DQA2* encodes the alpha chain that binds the protein produced from HLA-DQB1 (Lenormand C. et al., 2012). Unlike most HLA class II genes, *HLA-DQA2* and *HLA-DQB2* exhibit little or no polymorphism and appear to be specifically expressed in Langerhans cells (Berdoz J. et al., 1989; Lenormand C. et al., 2012). Diseases associated with *HLA-DQA2* include for instance Parkinson's disease (Hill-Burns EM. et al., 2011) and Rheumatoid Arthritis (Galligan CL. et al., 2007). The role of HLA class II proteins in the hearing loss has not been deeply investigated yet. Immunogenetic studies in patients with hearing loss have suggested a role for the HLA complex genes in the disease pathogenesis and it has been found that that specific HLA class II gene products may confer susceptibility or resistance to idiopathic progressive sudden hearing loss (Cao MY. et al., 1996; Amor-Dorado JC. et al., 2005). Furthermore, a recent study (Wells HRR. et al, 2019) reported a genome-wide significant association between the self-reported hearing impairment and a locus close to *HLA-DQA1*. In this light, considering the genome-wide significant association here detected and the one reported in literature (Wells HRR et al, 2019), and the possible contribution of immunity in the development of hearing loss, it might be reasonable to think that HLA, together with ageing and environmental factors, might have a predisposing or protecting effect that needs to be further investigated.

As mentioned earlier, the HLA region is highly polymorphic: for this reason, the conventional imputation techniques that have been employed genome-wide in this study have limited reliability, and this must be considered a limitation of this result. Several resources have been specifically designed to address this fact, including reference panels (e.g. Mimori T. et al, 2018; Degenhardt F. et al, 2019) and statistical tools (see Karnes J. H. et al, 2017; Khor S.S. et al, 2018; Dilthey A. et al, 2013), which could be employed in a follow-up investigation of this association signal.

D2HGDH encodes the mitochondrial enzyme D-2-hydroxyglutarate dehydrogenase. Within mitochondria, D-2-hydroxyglutarate dehydrogenase converts the D-2-hydroxyglutarate to 2-ketoglutarate. A series of additional enzymes further process 2-ketoglutarate to produce energy. The gene is associated to D-2-hydroxyglutaric aciduria, inherited with autosomal recessive pattern. Recently, the gene has been found mutated in a patient with D-2-hydroxyglutaric aciduria and

speech delay, but the hearing assessment was normal (Phillips E. et al., 2019); further studies will be needed to investigate the role, if any, of this gene in the hearing system.

In G-EAR Consortium analyses, three further suggestive association signals with the normal hearing function at 1KHz were detected in an intron region of *PLXDC2*, between *MTMR7* and *SLC7A2* and near *CAAP1*. For these genes, no involvement with the hearing system is known from literature.

Little is known so far about the *PLXDC2* gene/protein. It is recognized as a cell surface transmembrane receptor expressed in various tissues (see, e.g., GTEx Portal21) without so far clear arguments that it could be directly involved in mRNA regulation (Thibord F. et al., 2019). Previous association studies have linked *PLXDC2* to the coagulation process (Thibord F. et al., 2019) and open angle glaucoma (Chen M. et al., 2019).
PLXDC2 is a receptor for PEDF (Pigment epithelium-derived factor) factor, also known as serpin F1 (SERPINF1), which is a multifunctional secreted protein with anti-angiogenic, anti-tumorigenic, and neurotrophic functions (Cheng G. et al., 2014).
No involvement of *PLXDC2* with hearing is known to date; however, as this gene is expressed in mouse inner ear, and in the supporting cells in particular, it is still possible that *PLXDC2* may have a role in the auditory system or in hearing diseases, that needs to be investigated by further studies.
MTMR7 encodes the Myotubularin-related protein 7 which a member of the myotubularin (MTM) family (Fletcher AE. et al., 1992). It is specifically expressed in the central nervous system and dephosphorylates the D3 position of phosphatidylinositol 3-phosphate and inositol 1,3-bisphosphate (Sanchez-Juan P. et al., 2012). The gene is expressed in both cochlear and vestibular mouse hair cells (gEAR database), however no association with the hearing function or loss have been described so far.
SLC7A2 encodes a cationic amino acid transporter and a member of the APC (amino acid-polyamine-organocation) family of transporters. It is membrane protein responsible for the cellular uptake of arginine, lysine and ornithine. It has been associated to a variety of phenotypes, including cancer progression (Sun P. et al, 2017), inflammation (Niese KA. et al., 2010), immunity defects (Thompson RW. et al., 2008), etc. However, until now it has never been associated to the hearing system. According to gEAR database the gene is highly expressed in both the mouse and zebrafish inner ear, and considering that several solute carriers have been described for being causative of hearing loss

(e.g. *SLC26A4*, *SLC17A8*, *SLC22A4*, etc.) these results might represent the first association of this gene with the hearing function.

Finally, very little is known about *CAAP*. It has been implicated as a negative regulator of the intrinsic apoptosis pathway by modulating caspase expression and activity (Aslam MA. et al., 2019); no association with hearing function is known so far.

With regards to the G-EAR Consortium analyses on the ARHL phenotype, two suggestive associations were detected with *CDH13* and *CTIF*.

CDH13 encodes an atypical member of the cadherin superfamily, highly expressed in the brain and cardiovascular system, and localized to the surface of the cell membrane via a glycosylphosphatidylinositol (GPI) anchor (Killen AC et al., 2017). It has cell adhesion properties and it is thought to affect cellular behaviour through its binding partners and signalling properties (Mavroconstanti T. et al., 2013). This protein acts as a negative regulator of axon growth during neural differentiation. It also protects vascular endothelial cells from apoptosis due to oxidative stress, and is associated with resistance to atherosclerosis. Moreover, the gene is hypermethylated in many types of cancer (Mavroconstanti T. et al., 2013). This gene has already been identified as suggestively associated with the normal hearing function (Girotto G. et al., 2011), on a set of samples that partially overlaps with the present study; *CDH13* has been shown to be broadly expressed in murine cochlea (Girotto G. et al., 2014). In this light, present results strengthen the relation between this gene and the auditory system.

CTIF encodes a component of the CBP80 (NCBP1; MIM# 600469)/CBP20 (NCBP2; MIM# 605133) translation initiation complex that binds cotranscriptionally to the cap end of nascent mRNA. The CBP80/CBP20 complex is involved in a simultaneous editing and translation step that recognizes premature termination codons (PTCs) in mRNAs and directs PTC-containing mRNAs toward nonsense-mediated decay (NMD). On mRNAs without PTCs, the CBP80/CBP20 complex is replaced with cytoplasmic mRNA cap-binding proteins, including EIF4G (MIM# 600495), and steady-state translation of the mRNAs resumes in the cytoplasm (Kim KM. et al., 2009).

Here we detected an association of *CTIF* with the ARHL phenotype. Interestingly, a previous study identified a correlation between rs12959910 in *CTIF* for the high frequency PTA in the ALSPAC children cohort (Harrison S. et al., 2015), already suggesting an association of this gene with the hearing phenotype. This is in line with the expression of this gene in the cochlear hair cells, as reported by gEAR and SHIELD databases.

Finally, in G-EAR Consortium and UK Biobank analyses two suggestive loci were identified near or within genes coding for zinc-finger proteins (one of the loci associated to PTAL lies near *ZNF717*; a locus associated to phenotype B in UK Biobank lies between *ZNF229* and *ZNF180*).

Zinc-finger proteins (ZNFs) are one of the most abundant groups of proteins and have a wide range of molecular functions. Given the wide variety of zinc-finger domains, ZNFs are able to interact with DNA, RNA, PAR (poly-ADP-ribose) and other proteins. Thus, ZNFs are involved in the regulation of several cellular processes. In fact, ZNFs are implicated in transcriptional regulation, ubiquitin-mediated protein degradation, signal transduction, actin targeting, DNA repair, cell migration, and numerous other processes (Cassandri M. et al., 2017). So far, ZNF proteins have no known role in the hearing system or in hearing impairment.

The second objective of the research was to assess if the results obtained from one dataset are replicated in those obtained from the other, and vice versa.

Despite the strong correlation that has been shown to exist between PTAs and SRT (Koole A. et al., 2016), the lead SNPs were not successfully replicated.

A first hypothesis to explain this fact is that, even though the two assessments techniques have been both developed for the same purpose, the resulting phenotypes are two sides of the same coin, describing in the end different aspects of the hearing ability and therefore associations detected by GWAS technique are different. This may also be the reason behind the lack of common findings between SRT and self-reported hearing loss in the UK Biobank cohort.

On the other hand, if we consider the results from the GWAS performed on Phenotype B, the scarcity of association signals despite the available sample size may hint at the presence of an important confounding factor not accounted for during the analyses on both Phenotype A and B. A complete GWAS on phenotype A will be needed to support or discard this last hypothesis.

Further studies (e.g. LD score regression studies) will be necessary to clarify the interrelations between these two approaches at phenotyping.

CONCLUSIONS
and future perspectives

This study, for the first time, provides a comparison of two of the largest cohorts ever described for the hearing phenotype - G-EAR Consortium and UK Biobank - and of their approach at its collection. In this research, the application of GWA methodology on both cohorts has highlighted promising new genes for the normal hearing function and age-related hearing loss: *CADM2*, *CALB1*, *LRRC4C*, *ROBO2*, *CAAP1*, *SLC7A2*, *PLXDC2*, *CTIF*, *PDE1A* and *D2HDGH* plus a genome-wide association signal detected within the *HLA* region, prompting further investigation of the role of the immune system in hearing impairment.

The lack of replication of the signals between the two cohorts, despite the effort at defining in the single cohorts corresponding phenotypes as close as possible, points out the importance of a uniform and detailed approach when collecting complex phenotypes such as the hearing ability.

Screening tests as the DTT, while suitable for a general assessment, miss to consider the fine details of the hearing phenotype, and thus signals detected by analyses on the more detailed audiometric data may be "lost in translation" when considering only the general picture of the hearing ability given by the DTT and its SRT value.

In this light, the signals identified during this research remain a first step, and will therefore be subjected to further investigation. An important follow-up to this research will be the replication of these results in independent cohorts, following this time a more traditional approach: cohorts with identical and deeply characterised phenotypes have already been contacted for this purpose.

Furthermore, *in vitro* and *in vivo* studies, including studies of expression in murine and zebrafish cochlea, will be required, and will be a crucial point to elucidate the potential role of the identified genes in the modulation of the hearing function and in the etiopathogenesis of age-related hearing loss.

In conclusion, these findings provide a significant contribution to the understanding of the genetic bases underlying complex hearing traits and diseases, within the long-term perspective of the development of personalised therapeutic approaches, preventive strategies and diagnostic screenings for hearing impairment.

CONTRIBUTIONS

G-EAR Consortium - Recruitment, DNA/saliva and phenotype collection, genotyping, genotype quality control and the subsequent data imputation on the specified reference panels were performed for each cohort by the respective institutions, and the final results were made available for this project, with the exception of Silk Road Cohort (SR), for which imputation to the HRC reference panel was performed by the author.
GWAS analyses on Rotterdam Study Cohorts (RS1, RS2 and RS3) were performed by Berthe C. Oosterloo (Department of Otorhinolaryngology, Erasmus Medical Center, Rotterdam, The Netherlands). GWAS analyses on Framingham Heart Study cohort (FHS) were performed by dr. Nancy L. Heard-Costa (Department of Neurology, Boston University School of Medicine, Boston, MA, USA; Framingham Heart Study, Framingham, MA, USA).

UK Biobank - Genotype data, imputation and raw phenotype data was obtained from UK Biobank under Application Number 28489; recruitment, genotype and phenotype collection, genotyping and imputation was performed centrally by UK Biobank project and described in Bycroft C. et al. (2018).

Interpretation of the results - Dr. Anna Morgan (Department of Medicine, Surgery and Health Sciences, University of Trieste, Italy) provided an invaluable contribution to the interpretation of the possible biological meaning of the obtained results.

Every aspect of this research not specified above was performed by the author.

ACKNOWLEDGEMENTS

I conducted this research working in prof. Paolo Gasparini's research group. I would like to thank my supervisor, Dr. Giorgia Girotto, and prof. Paolo Gasparini for their scientific advice and support. I would also like to express my gratitude to everyone else in the group, for their help, encouragement, and patience, with special thanks to Dr. Anna Morgan for her contribution.

Moreover, I am very grateful to Berthe C. Oosterloo (Department of Otorhinolaryngology, Erasmus Medical Center, Rotterdam, The Netherlands) and prof. Nancy Heard-Costa (Department of Neurology, Boston University School of Medicine, Boston, MA, USA; Framingham Heart Study, Framingham, MA, USA), for running the set of analyses I used in this thesis on Rotterdam Study and Framingham Heart Study cohorts.

BIBLIOGRAPHY

Airaksinen L. et al. (2000) Lack of calbindin-D28k does not affect hearing level or survival of hair cells in acoustic trauma. ORL J Otorhinolaryngol Relat Spec. **62**(1), 9-12.

Amor-Dorado JC. et al. (2005) Human leukocyte antigen-DQB1 and –DRB1 associations in patients with idiopathic sudden sensorineural hearing loss from a defined population in Northwest Spain. Acta Otolaryngol. **125**(12), 1277-82.

Aptsiauri N. et al. (2011) HLA Class I. In: Schwab M. (eds) Encyclopedia of Cancer. Springer, Berlin, Heidelberg.

Aslam MA. et al. (2019) Towards an understanding of C9orf82 protein/CAAP1 function. PLoS One. **14**(1), e0210526.

Aulchenko YS. et al. (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics. **23**(10), 1294-6.

Berdoz J. et al. (1989) Remarkable sequence conservation of the HLA-DQB2 locus (DX beta) within the highly polymorphic DQ subregion of the human MHC. Immunogenetics. **29**(4), 241-8.

Bolker BM. et al. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol. **24**(3), 127-35.

Borck G. et al. (2011) Loss-of-Function Mutations of ILDR1 Cause Autosomal-Recessive Hearing Impairment DFNB42. Am J Hum Genet. **88**(2), 127-37.

Brewster KK. et al. (2018) Age-Related Hearing Loss and Its Association with Depression in Later Life, Am J Geriatr Psychiatry. **26**(7), 788-796.

Buckiová D. et al. (2009) Calbindin and S100 protein expression in the developing inner ear in mice. J Comp Neurol. **513**(5), 469-82.

Buniello A. et al. (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleid Acids Res. **47**(D1), D1005-D1012.

Bycroft C. et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature. **562**(7726), 203-9.

Cao LQ. et al. (2019) CALB1 enhances the interaction between p53 and MDM2, and inhibits the senescence of ovarian cancer cells. Mol Med Rep. **19**(6), 5097-5104.

Cao MY. et al. (1996) HLA class II-associated genetic susceptibility in idiopathic progressive sensorineural hearing loss. Ann Otol Rhinol Laryngol. **105**(8), 628-33.

Cassandri M. et al. (2017) Zinc-finger proteins in health and disease. Cell Death Discov. 3, 17071.

Chen M. et al. (2019) Association of Gene Polymorphisms With Primary Open Angle Glaucoma: A Systematic Reviewand Meta-Analysis. Invest Ophtalmol Vis Sci. **60**(4), 1105-1121.

Cheng G. et al. (2014) Identification of PLXDC1 and PLXDC2 as the transmembrane receptors for the multifunctional factor PEDF. Elife. **3**, e05401.

Choi Y. et al. (2019) NGL-1/LRRC4C Deletion Moderately Suppresses Hippocampal Excitatory Synapse Development and Function in an Input-Independent Manner. Front Mol Neurosci. **12**, 119, eCollection 2019.

Clark JG. (1981). Uses and abuses of hearing loss classification. ASHA., **23**(7), 493-500.

Cocca M. et al. (2019) A bird's-eye view of Italian genomic variation through Whole Genome Sequencing. Eur J Hum Genet. [epub ahead of print]

Corper AL. et al. (2000) A structural framework for deciphering the link between I-Ag7 and autoimmune diabetes. Science. **288**(5465), 505-11.

Dalton DS., et al. (2003) The impact of hearing loss on quality of life in older adults. Gerontologist **43**(5), 661-8.

Das S. et al. (2016) Next-generation genotype imputation service and methods. Nat Genet. **48**(10), 1284-1287.

Das S et al. (2018) Genotype Imputation from Large Reference Panels. Annu Rev Genomics Hum Genet. **19**:73-96.

Dawber TR. et al. (1951) Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. **41**(3), 279-81.

Dawber TR. et al. (1963) An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci. **107**, 539-56.

Delprat B. et al. (2005) Myosin XVa and whirlin, two deafness gene products required for hair bundle growth, are located at the stereocilia tips and interact directly. Hum Mol Genet. **14**(3), 401-10.

Degerman E. et al. (2011) Identification of new signalling components in the sensory epithelium of human saccule. Front Neurol. **2**, 48, eCollection 2011.

Dror AA., Avraham K.B. (2010) Hearing impairment: a panoply of genes and functions. Neuron., **68**(2), 293-308.

Du H. et al. (2018) Identification of Binding Partners of Deafness-Related Protein PDZD7. Neural Plast. 2062346, eCollection 2018.

Esko T. et al. (2013) Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. Eur J Hum Gen. **21**(6), 659-665.

Evangelou E. et al (2013) Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. **14**(6):379-89.

Fidock M. et al. (2002) Isolation and differential tissue distribution of two human cDNAs encoding PDE1 splice variants. Cell Signal. **14**(1), 53-60.

Fischer N. et al. (2019) Age-Dependent Calcium-Binding Protein Expression in the Spiral Ganglion and Hearing Performance of C57BL/6J and 129/SvJ Mice. ORL J Otorhinolaryngol Relat Spec. **81**(2-3), 138-154.

Fletcher AE. et al. (1992) Survival of hypertensive subjects identified on screening: results for sustained and unsustained diastolic hypertension. Eur Heart J. **13**(12), 1595-601.

Foley DM. et al. (2014) Association between hearing loss and healthcare expenditures in older adults. J Am Geriatr Soc. **62**(6), 1188-9.

Fransen E. et al. (2008) Occupational noise, smoking, and a high body mass index are risk factors for age-related hearing impairment and moderate alcohol consumption is protective: a European population-based multicenter study. J Assoc Res Otolaryngol. **9**(3), 264-76.

Friedman R. A. et al. (2009) GRM7 variants confer susceptibility to age-related hearing impairment. Hum Mol Genet. **18**(4), 785-796.

Frolenkov GI. et al (2004). Genetic insights into the morphogenesis of inner ear hair cells, Nat Rev Genet., **5**(7), 489-98.

Fuchs J.C., Tucker A.S. (2015). Development and Integration of the Ear. Curr. Top. Dev. Biol., **115**, 213-32.

Galligan CL. et al. (2007) Distinctive gene expression signatures in rheumatoid arthritis synovial tissue fibroblast cells: correlates with disease activity. Genes Immunit. **8**(6), 480-91.

Gates G.A. et al. (2005) Presbycusis. Lancet. **366**(9491), 1111-20.

Girotto G. et al. (2011) Frequency of hearing loss in a series of rural communities of five developing countries located along the Silk Road. Audiol Med. **9**(4).

Girotto G. et al. (2011) Hearing function and thresholds: a genome-wide association study in European isolated populations identifies new loci and pathways. J Med Genet. **48**(6), 369-74.

Girotto G. et al. (2014) Expression and Replication Studies to Identify New Candidate Genes Involved in Normal Hearing Function. PLoS One. **9**(1), e85352.

Haller T. et al. (2019) Manhattan Harvester and Cropper: a system for GWAS peak detection. BMC Bioinformatics. **20**(1), 22.

Harrison S. et al., (2015) Association of SNPs in LCP1 and CTIF with hearing in 11 year old children: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort and the G-EAR consortium. BMC Med Genomics. **8**, 48**.**

Hill-Burns EM. et al. (2011) Evidence for more than one Parkinson's disease-associated variant within the HLA region. PLoS ONE. **6**(11), e27109.

Hoffmann T. J. et al. (2016) Study of Age-Related Hearing Impairment Using Electronic Health Records. PLoS Genet. **12**, e1006371.

Howie B. et al. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. **44**(8), 955-9.

Ikram MA. et al. (2017) The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol. **32**(9), 807-850.

Johnson EW. (1970) Tuning forks to audiometers and back again. Laryngoscope **80**(1), 49-68.

Kandel E. et al. (2013) *Principles of neural science*. New York: McGraw-Hill.

Killen AC. et al. (2017) Protective role of Cadherin 13 in interneuron development. Brain Struct Funct. **222**(8), 3567-3585.

Kim KM. et al. (2009) A new MIF4G domain-containing protein, CTIF, directs nuclear cap-binding protein CBP80/20-dependent translation. Genes Dev. **23**(17), 2033-45.

Koole A. et al. (2016) Using the Digits-In-Noise Test to Estimate Age-Related Hearing Loss. Ear Hear. **37**(5), 508-13.

Lang F. et al. (2007) Functional significance of channels and transporters expressed in the inner ear and kidney. Am J Physiol Cell Physiol., **293**(4), C1187-208.

Lavinsky J. et al. (2016) The Genetic Architecture of Noise-Induced Hearing Loss: Evidence for a Gene-by-Environment Interaction. G3 (Bethesda). **6**(10), 3219-3228.

Lenormand C. et al. (2012) HLA-DQA2 and HLA-DQB2 genes are specifically expressed in human Langerhans cells and encode a new HLA class II molecule. J Immunol. **188**(8), 3903-11.

Li Y. et al. (2009) Genotype Imputation. Annu Rev Genomics Hum Genet. **10**, 387-406

Lin FR. et al. (2014) Hearing loss and dementia - who is listening? Aging Ment Health. **18**(6), 671-3.

Lloyd-Jones L. et al. (2018) Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. Genetics. **208**(4), 1397-1408.

Mandai K. et al. (2015) Nectins and Nectin-Like Molecules in Development and Disease. Curr Top Dev Biol. **112**, 197-231.

Marchini J. et al. (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet. **11**(7), 499-511.

Massouh M. et al. (2008) The fate of the large striatal interneurons expressing calretinin in Huntington's disease. Neurosci Res. **62**(4), 216-24.

Maussion G. et al. (2017) Implication of LRRC4C and DPP6 in neurodevelopmental disorders. Am J Med Genet. **173**(2), 395-406.

Mavroconstanti T. et al. (2013) Functional properties of rare missense variants of human CDH13 found in adult attention deficit/hyperactivity disorder (ADHD) patients. PLoS One. **8**(8), e71445.

McCarthy S. et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. **48**(10), 1279-83.

Megiorni F. et al. (2012) HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. J Biomed Sci. **19**, 88.

Mezzavilla M. et al. (2014) Genetic landscape of populations along the Silk Road: admixture and migration patterns. BMC Genet. **15**, 131.

Michibata H. et al. (2001) Human Ca2+/calmodulin-dependent phosphodiesterase PDE1A: novel splice variants, their specific expression, genomic organization, and chromosomal localization. Biochim Biophys Acta. **1517**(2), 278-87.

Mittal R. et al. (2017) Role of Cyclic Nucleotide Phosphodiesterases in Inner Ear and Hearing. Front Physiol. 8, 908, eCollection 2017.

Moore DR. et al. (2014) Relation between Speech-in-Noise Threshold, Hearing Loss and Cognition from 40-69 Years of Age. PLoS One. **9**(9), e107720.

Nagtegaal AP. et al. (2019) Genome-wide association meta-analysis identifies five novel loci for age-related hearing impairment. Sci Rep. **9**(1), 15192.

Newman D. L. et al. (2012) GRM7 variants associated with age-related hearing loss based on auditory perception. Hear. Res., **294**, 125–132.

Niese KA. et al. (2010) The cationic amino acid transporter 2 is induced in inflammatory lung models and regulates lung fibrosis. Respir Res. **11**, 87.

O'Connell J. et al. (2014) A general approach for haplotype phasing through the full spectrum of relatedness. PLoS Genet. **10**(4), e1004234.

Oksenberg JR. et al. (2004) Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. Am J Hum Genet. **74**(1), 160-7.

Phillips E. et al. (2019) D-2-hydroxyglutaric aciduria in a patient with speech delay due to a novel homozygous deletion in the *D2HGDH* gene. Mol Genet Metab Rep. **20**, 100482, eCollection 2019.

Potgieter JM. et al. (2015) Development and validation of a smartphone-based digits-in-noise hearing test in South African English. Int J Audiol. **55**(7), 405-11.

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Riazuddin S. et al. (2006) Mutations in TRIOBP, which encodes a putative cytoskeletal-organizing protein, are associated with nonsyndromic recessive deafness. Am J Hum Genet. **78**(1), 137-43.

Rigters SC. et al. (2018) Progression of Hearing Loss in the Aging Population: Repeated Auditory Measurements in the Rotterdam Study. Audiol Neurootol. **23**(5), 290-297.

Risch N. et al. (1996) The future of genetic studies of complex human diseases. Science **273**(5281), 1516-7.

Roth TN. et al. (2015) Aging of the auditory system. Handb Clin Neurol. **129**, 357-73.

Sanchez-Juan P. et al. (2012) Genome-wide study links MTMR7 gene to variant Creutzfeldt-Jakob risk. Neurobiol Aging. **33**(7), 1487, e21-8.

Shahin H. et al. (2006) Mutations in a Novel Isoform of TRIOBP That Encodes a Filamentous-Actin Binding Protein Are Responsible for DFNB28 Recessive Nonsyndromic Hearing Loss. Am J Hum Genet. **78**(1), 144-72.

Shaikh RS. et al. (2005) A New Locus for Nonsyndromic Deafness *DFNB51* Maps to Chromosome 11p13-p12. Am J Med Genet A. **138**(4), 392-95.

Shen J. et al. (2015) SHIELD: an integrative gene expression database for inner ear research. Database (Oxford), bav071.

Smits C. et al. (2013) The digits-in-noise test: assessing auditory speech recognition abilities in noise. J Acoust Soc Am. **133**(3), 1693-706.

St Pourcain B. et al. (2014) Common variation near ROBO2 is associated with expressive vocabulary in infancy. Nat Commun. **5**, 4831.

Steel K.P., Kros C.J. (2001). A genetic approach to understanding auditory function. Nature Genetics, **27**(2), 143–9.

Sun P. et al. (2017) Genetic variation in SLC7A2 interacts with calcium and magnesium intakes in modulating the risk of colorectal polyps. J Nutr Biochem. **47**, 35-40.

Sung YK. et al. (2016) Association of Hearing Loss and Loneliness in Older Adults. J Aging Health. **28**(6), 979-94.

Svisheva GR. et al. (2012) Rapid variance components-based method for whole-genome association analysis. Nat Genet. **44**(10), 1166-70.

Tawfik KO. et al. (2019) Advances in understanding of presbycusis. J Neurosci Res. doi: 10.1002/jnr.24426. [*Epub ahead of print*].

The 1000 Genomes Project Consortium et al. (2015) A global reference for human genetic variation. Nature. **526**(7571), 68-74.

Thibord F. et al. (2019) A Genome Wide Association Study on plasma FV levels identified *PLXDC2* as a new modifier of the coagulation process. J Thromb Haemost. **17**(11), 1808-1814.

Thompson RW. et al. (2008) Cationic amino acid transporter-2 regulates immunity by modulating arginase activity. PLoS Pathog. **4**(3), e1000023.

Traglia M. et al. (2009) Heritability and demographic analyses in the large isolated population of Val Borbera suggest advantages in mapping complex traits genes. PLoS One. **4**(10), e7554.

Tsao CV. et al. (2015) Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol. **44**(6), 1800-13.

Tu N.C. et al. (2018) Age-related hearing loss: Unraveling the pieces. Laryngoscope Investig Otolaryngol. **3**(2), 68-72.

Van Eyken E. et al. (2006) KCNQ4: A gene for age-related hearing impairment?. Hum Mutat. **27**(10), 1007-16.

Van Laer L. et al. (2003) Nonsyndromic hearing loss. Ear Hear. **24**(4), 275-88.

Van Laer L. et al. (2010) A genome-wide association study for age-related hearing impairment in the Saami. Eur J Hum Gen. **18**, 685-693.

Visscher P.M. et al. (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. **101**(1), 5-22.

Vuckovic D. et al. (2015) Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss. Hum Mol Genet. **24**(19), 5655-64.

Vuckovic D. et al. (2018) Whole-genome sequencing reveals new insights into age-related hearing loss: cumulative effects, pleiotropy and the role of selection. Eur J Hum Genet. **26**(8), 1167-79.

Wang K. et al. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. **38**(16), e164.

Wang L. et al. (2018) A dominant variant in the PDE1C gene is associated with nonsyndromic hearing loss. Hum Genet. **137**(6-7), 437-46.

Wang SZ. et al. (2013) Slit/Robo signaling mediates spatial positioning of spiral ganglion neurons during development of cochlear innervation. J Neurosci. **33**(30), 12242-54.

Wang X. et al. (2017) Generation and phenotypic characterization of Pde1a mutant mice. PLoS ONE. **12**(7), e0181087, eCollection 2017.

Wayne S. et al. (2001) Mutations in the transcriptional activator EYA4 cause late-onset deafness at the DFNA10 locus. Hum Mol Genet. **10**(3), 195-200.

Wells H.R.R. et al. (2019) GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. Am J Hum Genet. **105**(4): 788-802.

Willer CJ. et al. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. **26**(17), 2190-1.

Wolber LE. et al. (2012) Hearing Ability with Age in Northern European Women: A New Web-Based Approach to Genetic Studies. PLoS ONE **7**(4), e35500.

Wolber LE. et al. (2014) Salt-inducible kinase 3, *SIK3*, is a new gene associated with hearing. Hum Mol Genet. **23**(23), 6407–6418.

Yang Y. et al. (2011) GCTA: a tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. **88**(1), 76-82.

Yang Y. et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet **44**(4), 369-375.

Yengo L. et al. (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet. **27**(20), 3641-3649.

Zeggini E. et al (2009) Meta-analysis in genome-wide association studies. Pharmacogenomics. **10**(2): 191-201.

Zeggini E. et al. (2011) Next-generation association studies for complex traits. Nat. Genet. **43**, 287-288.

Zhan X. et al. (2016) RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. **32**(9), 1423-6.

SUPPLEMENTARY MATERIAL

G-EAR Consortium.

Supplementary Table 1 - Subjects available for analyses, divided by trait and cohort.

	250Hz	500Hz	1KHz	2KHz	4KHz	8KHz	PTAL	PTAM	PTAH
FVG	1450	1462	1461	1470	1446	1435	1455	1455	1426
VBI	423	424	422	425	424	421	424	424	420
CAR	258	254	260	257	261	259	258	254	255
SR	639	641	647	656	656	655	649	645	651
AWP	1056	1068	1059	1062	1086	1092	1087	1088	1102
RS1	684	684	684	684	684	684	684	684	684
RS2	1173	1173	1173	1173	1173	1173	1173	1173	1173
RS3	2155	2155	2155	2155	2155	2155	2155	2155	2155
FHS	1101	1100	1130	1131	1122	1129	1115	1111	1131
TOT	**8939**	**8961**	**8991**	**9013**	**9007**	**9003**	**9000**	**8989**	**8997**

Supplementary Table 2 - Distribution of the hearing traits in G-EAR Consortium cohorts (250Hz, 500Hz, 1KHz, 2KHz, 4KHz).

POP	N	M; F	Age range	250Hz	500Hz	1KHz	2KHz	4KHz
FVG	1470	M=634; F=836	18-91 (mean=51.9)	0-60 (mean=9.8; sd=8.04)	0-75 (mean=10.58; sd=9.43)	0-75 (mean=10.99; sd=10.57)	0-85 (mean=12.43; sd=13.71)	0-100 (mean=21.3; sd=20.74)
VBI	425	M=180; F=245	25-90 (mean=58.71)	5-75 (mean=17.2; sd=8.46)	5-70 (mean=18.47; sd=9.69)	10-95 (mean=20.52; sd=11.55)	5-90 (mean=22.64; sd=16.3)	5-100 (mean=30.54; sd=21.27)
CAR	261	M=113; F=148	18-89 (mean=53.4)	5-75 (mean=18.66; sd=9.8)	0-90 (mean=16.61; sd=11.41)	0-65 (mean=15.04; sd=12.69)	0-75 (mean=16.83; sd=16.86)	0-130 (mean=26.53; sd=23.99)
SR	656	M=249; F=407	18-82 (mean=39.82)	5-90 (mean=16.74; sd=10.01)	5-90 (mean=18.41; sd=9.74)	0-90 (mean=14.74; sd=10.67)	0-90 (mean=11.54; sd=12.24)	0-90 (mean=16.34; sd=16.72)
AWP	1102	M=555; F=547	53-67 (mean=60.99)	0-45 (mean=13.72; sd=6.5)	5-55 (mean=15.98; sd=7.8)	5-65 (mean=16.39; sd=8.55)	0-70 (mean=18.51; sd=11.31)	5-85 (mean=28.99; sd=16.65)

RS1	684	M=280; F=404	77-101 (mean=82.9)	0-85 (mean=20.8; sd=12.6)	0-80 (mean=22.55; sd=13.7)	0-80 (mean=27.9; sd=14.6)	0-90 (mean=37.8; sd=16.6)	5-125 (mean=55.3; sd=17.7)
RS2	1173	M=530; F=643	66-99 (mean=75.4)	0-95 (mean=16.53; sd=10.67)	0-110 (mean=17.7, sd=11.63)	0-120 (mean=20.67, sd=13.17)	0-130 (mean=27.32, sd=16.19)	5-125 (mean=45.12, sd=19.5)
RS3	2155	M=941; F=1214	52-95 (mean=62.3)	0-85 (mean=10.6, sd=8.04)	0-95 (mean=11.31, sd=8.69)	0-90 (mean=13.41, sd=10.05)	0-90 (men=14.68, sd=13.19)	0-110 (mean=29.02, sd=17.58)
FHS	1131	M=518; F=613)	32-86 (mean=59.25)	0-55 (mean=10.78, sd=7.74)	0-60 (mean=11.99, sd=8.47)	0-80 (mean=11.78, sd=9.93)	0-80 (mean=15.10, sd=13.35)	0-95 (mean=26.77, sd=19.34)

Supplementary Table 3 - Distribution of the hearing traits in G-EAR Consortium cohorts, continued (8KHz, PTAL, PTAM, PTAH).

POP	N	M; F	Age range	8KHz	PTAL	PTAM	PTAH
FVG	1470	M=634; F=836	18-91 (mean=51.9)	0-110 (mean=22.76; sd=25.39)	0-61.67 (mean=10.6; sd=8.51)	0-73.33 (mean=12.53; sd=10.99)	0-102.5 (mean=22.73; sd=22.11)
VBI	425	M=180; F=245	25-90 (mean=58.71)	10-125 (mean=34.94; sd=25.46)	5-58.33 (mean=18.96; sd=8.9)	8.33-70 (mean=20.94; sd=11.37)	2.5-112.5 (mean=33.18; sd=22.69)
CAR	261	M=113; F=148	18-89 (mean=53.4)	0-130 (mean=31.97; sd=25.06)	1.67-61.67 (mean=16.94; sd=10.4)	0-60 (mean=16.71; sd=12.51)	0-92.5 (mean=29.61; sd=22.84)
SR	656	M=249; F=407	18-82 (mean=39.82)	0-90 (mean=21.69; sd=21.61)	3.33-90 (mean=16.81; sd=9.11)	1.67-90 (mean=15.41; sd=9.9)	0-90 (mean=19.95; sd=18.18)
AWP	1102	M=555; F=547	53-67 (mean=60.99)	5-110 (mean=36.34; sd=21.24)	1.67-46.67 (mean=15.37; sd=7.12)	1.67-55 (mean=17.12; sd=8.44)	5-90 (mean=33.12; sd=17.71)
RS1	684	M=280; F=404	77-101 (mean=82.9)	10-110 (mean=72.6; sd=17.6)	3.33-83.33 (mean=24.4; sd=12.58)	1.67-81.67 (mean=30.29; sd=13.48)	7.5-117.5 (mean=64.7-15.84)
RS2	1173	M=530; F=643	66-99 (mean=75.4)	0-110 (mean=60.66, sd=21.27)	0-111.67 (mean=18.87, sd=10.80)	0-123.33 (mean=22.58, sd=12.38)	7.5-117.5 (mean=53.76, sd=18.82)
RS3	2155	M=941; F=1214	52-95 (mean=62.3)	0-110 (mean=35.97, sd=22.44)	0-90 (mean=12.30, sd=7.95)	0-90 (mean=13.87, sd=9.42)	0-107.5 (mean=33.41, sd=18.46)
FHS	1131	M=518; F=613)	32-86 (mean=59.25)	0-95 (mean=33.13, sd=21.01)	0-60 (mean=11.77, sd=8.1)	0-65 (mean=13.32, sd 9.68)	0-95 (mean=30.44, sd=19.07)

Supplementary Table 4 - Subjects available for the analyses on ARHL phenotype, divided by cohort.

ARHL	N	% Cases	Age range
FVG	696	46.4%	50-91 (mean=63.9)
VBI	245	59.6%	50-90 (mean=67)
CAR	127	64.6%	50-89 (mean=67)
SR	154	44.8%	50-82 (mean=58.5)
AWP	874	44.2%	53-67 (mean=61)
PM	179	40.7%	51.3-67.8 (mean=59.7)
RS2	1025	91.6%	66-99 (mean=75.8)
RS3	1653	44.5%	52-95 (mean=62.3)
FHS	792	43.8%	50-86 (mean=61.6)
TOT	5745		

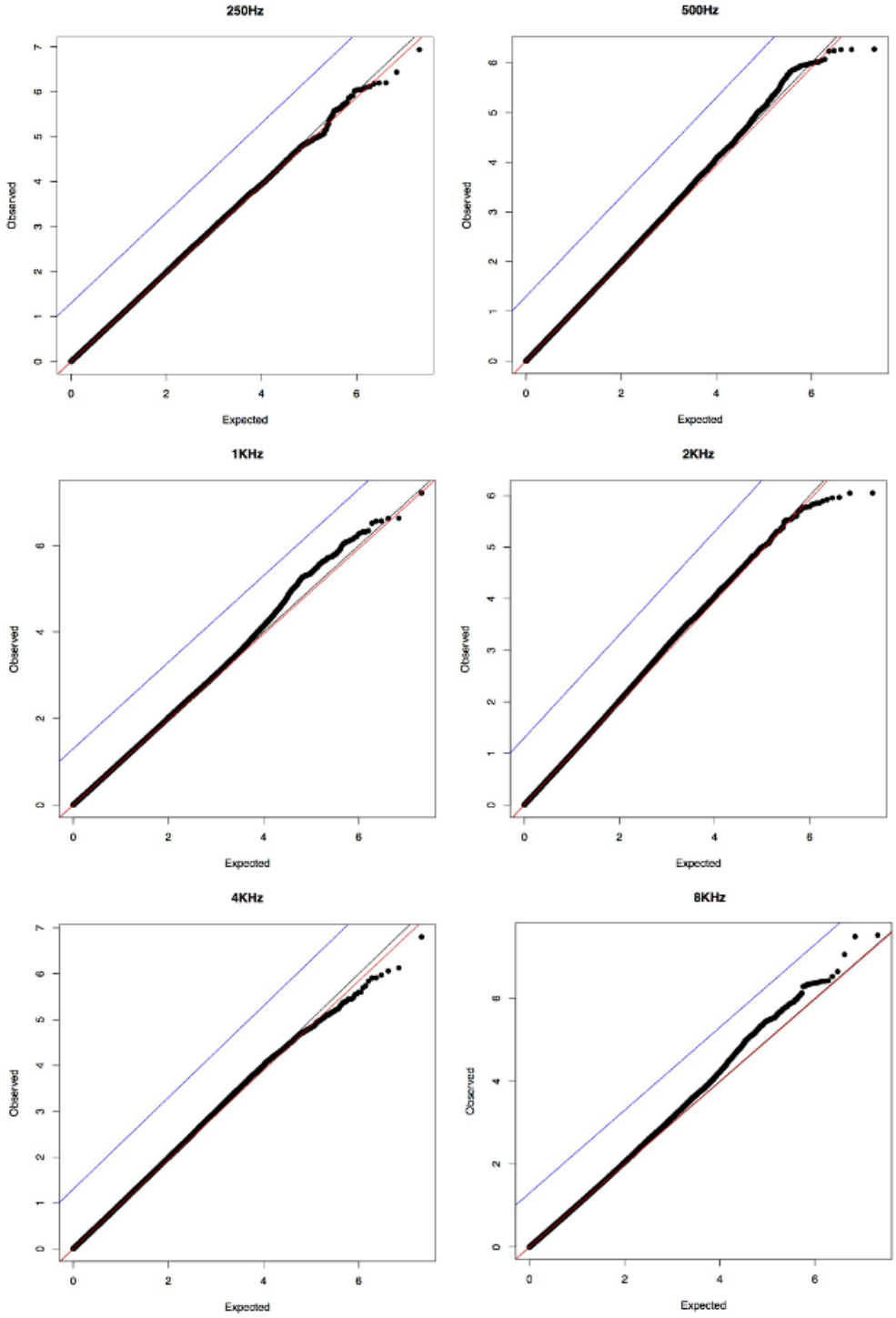
Supplementary Table 5 - Lambda values in single trait analyses (after MAF and imputation quality score filters) divided by trait and cohort.

Trait	FVG	VBI	CAR	SR	AWP	RS1	RS2	RS3	FHS	PM
X250	0.9912	0.9909	0.9466	**1.0662**	0.9870	1.0084	**1.0117**	**1.0185**	1.0006	/
X500	1.0095	**1.0102**	0.9338	1.0046	0.9930	0.9886	**1.0136**	0.9947	1.0019	/
X1K	1.0031	1.0030	0.9995	0.9982	0.9966	**1.0142**	**1.0174**	**1.0215**	**1.0106**	/
X2K	**1.0408**	0.9850	0.9660	0.9698	0.9781	**1.0132**	**1.0149**	**1.0137**	1.0047	/
X4K	0.9816	1.0023	1.0101	1.0051	0.9726	**1.0251**	**1.0244**	**1.0176**	**1.0195**	/
X8K	**1.0514**	0.9996	0.9929	0.9956	1.0000	1.0003	1.0094	1.0102	**1.0393**	/
PTAL	**1.0404**	1.0021	0.9289	0.9864	0.9856	1.0071	**1.0125**	**1.0193**	1.0087	/
PTAM	0.9846	0.9990	0.9936	**1.0124**	0.9819	1.0089	**1.0159**	1.0091	**1.0143**	/
PTAH	0.9802	1.0100	0.9927	0.9880	0.9855	1.0091	**1.0195**	**1.0131**	**1.0207**	/
ARHL	1.0055	**1.0228**	0.9808	0.9933	0.9930	/	1.0106	1.0231	1.0243	1.0085

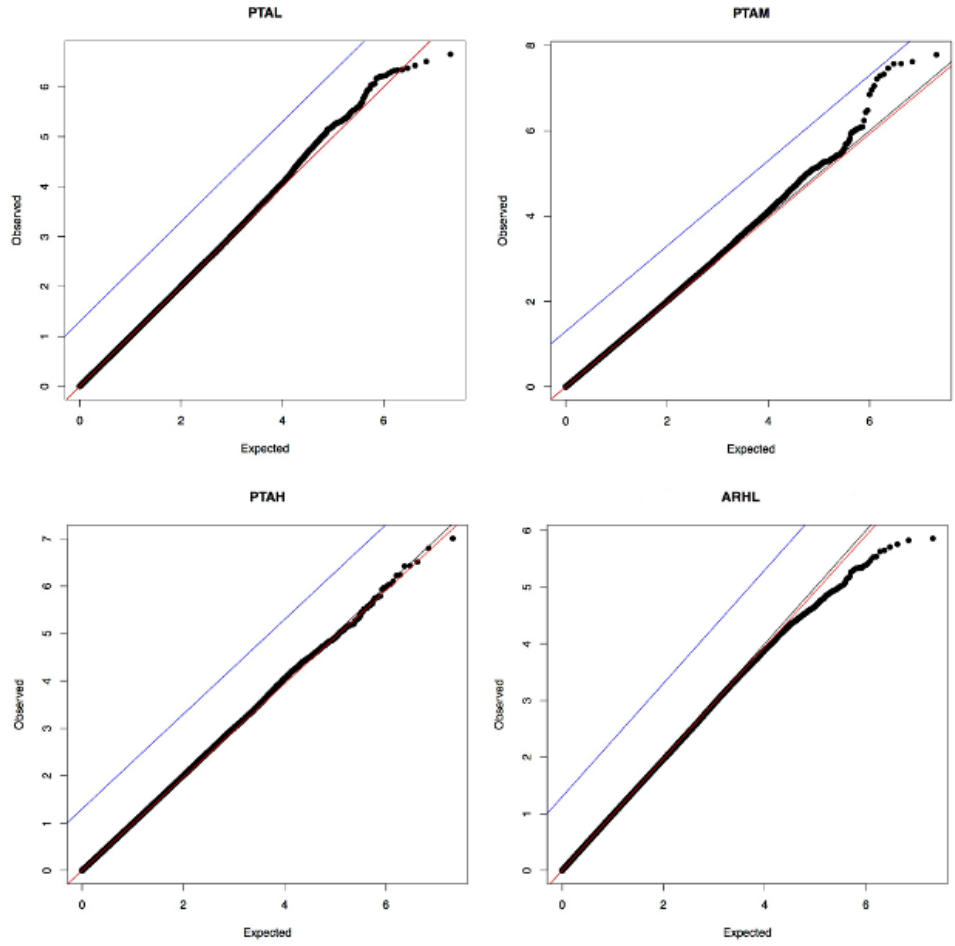
Supplementary Table 6 - Lambda values after meta-analysis, calculated on variants found in at least 3 cohorts.

X250	X500	X1K	X2K	X4K	X8K	PTAL	PTAM	PTAH	ARHL
0.9787	0.9842	0.9879	0.9875	0.9741	0.9963	0.9991	0.9875	0.9852	0.9847

Supplementary Figure 1 - Quantile-quantile plots of the meta-analyses: 250Hz, 500Hz, 1KHz, 2Khz, 4Khz, 8kHz.



Supplementary Figure 2 - Quantile-quantile plots of the meta-analyses: PTAL, PTAM, PTAH, ARHL.



UK Biobank.

Supplementary Table 7 - UK Biobank speech-in-noise test: available information. In-depth details about each variable are available at http://biobank.ndph.ox.ac.uk/showcase/.

id	id code
Sex	Sex
Year of birth	Year of birth
Month of birth	Month of birth
Date of attendance	Date attending assessment centre (assessment session 1,2,3)
Age when attended assessment centre	Age when attended assessment centre
Assessment centre	Which assessment centre was attended (assessment sessions 1,2,3)
Date lost to follow up	Date lost to follow up
Hearing problems	Whether the subject experiences hearing problems (assessment sessions 1,2,3)

Hearing problems with noise	Whether the subject experiences hearing problems with noise (assessment sessions 1,2,3)
Hearing aid user	Whether the subject uses hearing aids (assessment sessions 1,2,3)
Triplet played	Triplet played during assessments (information divided by round, left/right ear and assessment session)
Triplet SNR	Signal to Noise ratio of the round in which the correspondent triplet was played (information divided by round, left/right ear and assessment session)
Triplet entered	Triplet entered by the subject (information divided by round, left/right ear and assessment session)
Triplet correct	Whether the subject identified the played triplet correctly (information divided by round, left/right ear and assessment session)
Time to press first digit	Time taken to press the first digit on the touchscreen in each round (in seconds; information divided by round, left/right ear and assessment session)
Time to press 'next'	Time taken to press the 'next' key on the touchscreen in each round since the triplet played (in seconds; information divided by round, left/right ear and assessment session)
Time to press last digit	Time taken to press the last digit on the touchscreen in each round (in seconds; information divided by round, left/right ear and assessment session)
Keystroke history	Keystroke history (information divided by round, left/right ear and assessment session)
Number of times 'clear' was pressed	Number of times the 'clear' key was pressed to reset the triplet entry during the test (information divided by round, left/right ear and assessment session)
Completion status	Whether the hearing test was not started, abandoned or completed
Triplets attempted	Number of rounds attempted: 0-15 (information divided by ear and assessment session)
Volume level	Volume level set by participant (information divided by round, left/right ear and assessment session)
Duration of hearing test	Total duration of the hearing test (information divided by ear and assessment session)
Cochlear implant	Whether the subject wears a cochlear implant (information divided by assessment session)
Tinnitus	Responses to the question "Do you get or have you had noises (such as ringing or buzzing) in your head or in one or both ears that lasts for more than five minutes at a time?" (information divided by assessment session)
Tinnitus severity	Responses to the question "How much do these noises worry, annoy or upset you when they are at their worst?"
Noisy workplace	Responses to the question "Have you ever worked in a noisy place where you had to shout to be heard?" (information divided by assessment session)

Loud music exposure frequency	Responses to the question "Have you ever listened to music for more than 3 hours per week at a volume which you would need to shout to be heard or, if wearing headphones, someone else would need to shout for you to hear them?" (information divided by assessment session)
SRT estimate	Estimate of the SRT value (information divided by ear and assessment session); SRT is the Speech Reception Threshold, defined as the signal-to-noise ratio at which half of the presented speech can be understood correctly. Estimated for subjects who completed all 15 rounds.
Ethnic background	Ethnic background
Genetic kinship	Genetic kinship to other participants.
Heterozygosity	Heterozygosity across a set of high quality markers (
Missingness	Missing rate of each sample based on a set of high-quality markers (see Resource 530 at http://biobank.ndph.ox.ac.uk for details)
Heterozygosity or missing rate outliers	Samples identified as outliers in heterozygosity or missing rate during quality control.
PC 1-40	Genetic principal components 1 to 40 (see Resource 530 at http://biobank.ndph.ox.ac.uk for details)

Supplementary Table 8 - Summary of the variables as included in the analyses.

Sex	Male	89535
	Female	105469
Age range	Range	40-73
(1st assessment)	Mean	56.56
Noisy workplace	Yes, for more than 5 years	19345
(1st assessment)	Yes, for around 1-5 years	9339
	Yes, for less than a year	8990
	No	126586
	Do not know	1527
	Prefer not to answer	114
	Missing information	29103
Loud music exposure	Yes, for more than 5 years	7197
(1st assessment)	Yes, for around 1-5 years	7803
	Yes, for less than a year	5374
	No	143055
	Do not know	2397

	Prefer not to answer	75
	Missing information	29103
Tinnitus (1st assessment)	Yes, now most or all of the time	10180
	Yes, now a lot of the time	3965
	Yes, now some of the time	14306
	Yes, but not now, but have in the past	18230
	No, never	115692
	Do not know	3444
	Prefer not to answer	84
	Missing information	29103
Tinnitus Severity (1st assessment)	Severely	1567
	Moderately	7636
	Slightly	22212
	Not at all	14889
	Do not know	360
	Prefer not to answer	17
	Missing information	148323
Cochlear implant (1st assessment)	Yes	6
	No	165889
	Prefer not to answer	6
	Missing information	29103
Hearing aid user (1st assessment)	Yes	5154
	No	172599
	Prefer not to answer	93
	Missing information	17158
Volume level (left) (1st assessment)	100%	4625
	70%	11807
	40%	34250
	20%	51240
	10%	62642
	Missing information	30440
Volume level (right) (1st assessment)	100%	4927
	70%	12579
	40%	35531
	20%	51994
	10%	59611
	Missing information	30362

Completion status (left) (1st assessment)	Stopped	163381
	Abandoned	1468
	Error	2
	Missing Information	30153
Completion status (right) (1st assessment)	Stopped	163320
	Abandoned	1625
	Error	3
	Missing Information	30056
Ethnic background	Prefer not to answer	673
	Do not know	71
	White	178
	Mixed	17
	Asian or Asian British	21
	Black or Black British	12
	Chinese	681
	Other ethnic group	2149
	British	167409
	Irish	5596
	Any other white background	6974
	White and Black Caribbean	305
	White and Black African	179
	White and Asian	404
	Any other mixed background	464
	Indian	3596
	Pakistani	750
	Bangladeshi	89
	Any other Asian Background	989
	Caribbean	2679
	African	1696
	Any other Black background	61
	Missing Information	11
Hearing problems with noise (1st assessment)	Yes	70241
	No	120097
	Do not know	4558
	Prefer not to answer	97
	Missing information	11
heterozygosity or missingness outliers	Yes	393
SRT estimate (left)	Min	-11.25

(1st assessment)	Max	8
subjects with completion status (left ear)="stopped"	Mean	-6.587
	Median	-7
	Std	2.096
SRT estimate (right)	Min	-11.5
(1st assessment)	Max	8
subjects with completion status (right ear)="stopped"	Mean	-6.548
	Median	-7
	Std	2.106

Supplementary Figure 3 - Quantile-quantile plot of the genome-wide association study on phenotype B.