GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank

Helena R.R. Wells,[1,2] Maxim B. Freidin,[1] Fatin N. Zainul Abidin,[2,3] Antony Payton,[4] Piers Dawes,[5] Kevin J. Munro,[5,6] Cynthia C. Morton,[5,6,7] David R. Moore,[5,8] *Sally J. Dawson,[2,9] **Frances M.K. Williams,[1,9]

[1]Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, SE1 7EH, UK

[2]UCL Ear Institute, University College London, WC1X 8EE, UK

[3]Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, WC1E 7JE, UK.

[4]Division of Informatics, Imaging & Data Sciences, The University of Manchester, M13 9PT, UK

[5]Manchester Centre for Audiology and Deafness, The University of Manchester, M13 9PL, UK

[6]Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, M13 9WL, UK

[7]Departments of Obstetrics and Gynecology and of Pathology, Brigham and Women's Hospital, Harvard Medical School, MA 02115, USA

[8]Cincinnati Children's Hospital Medical Centre, Department of Otolaryngology, University of Cincinnati College of Medicine, OH 45229, USA

[9]These authors contributed equally to this work.

Corresponding authors: * sally.dawson@ucl.ac.uk and ** frances.williams@kcl.ac.uk

Abstract

Age-related hearing impairment (ARHI) is the most common sensory impairment in the ageing population; a third of individuals are affected by disabling hearing loss by the age of 65. It causes social isolation, depression and has recently been identified as a risk factor for dementia. The genetic risk factors and underlying pathology of ARHI are largely unknown, meaning that targets for new therapies remain elusive, yet heritability estimates range between 35% and 55%. We performed genome-wide association studies (GWAS) for two self-reported hearing phenotypes, using over 250,000 UK Biobank (UKBB) volunteers aged between 40-69 years. Forty-four independent genome-wide significant loci (P<5E-08) were identified, considerably increasing the number of established trait loci. Thirty-four loci are novel associations with hearing loss of any form, and only 1 of the 10 known hearing loci has a previously reported association with an ARHI-related trait. Gene sets from these loci are enriched in auditory processes such as synaptic activities, nervous system processes, inner ear morphology and cognition, while genetic correlation analysis revealed strong positive correlations with multiple personality and psychological traits for the first time. Immunohistochemistry for protein localisation in adult mouse cochlea implicate metabolic, sensory and neuronal functions for *NID2*, *CLRN2* and *ARHGEF28.* These results provide insight into the genetic landscape underlying ARHI, opening up novel therapeutic targets for further investigation. In a wider context, our study also highlights the viability of using self-report phenotypes for genetic discovery in very large samples when deep phenotyping is unavailable.

Introduction

ARHI is characterised by a non-syndromic bilateral, sensorineural hearing loss that progresses with increasing age and is an established risk factor for depression[1–3] and dementia[4–7] . Hearing loss was ranked fourth in the latest study into the Global Burden of Diseases[8], yet hearing amplification devices are the only treatment option currently available for ARHI, representing an area of unmet clinical need. The economic burden of the condition is increasing in response to the growing size of the ageing population; a recent review highlighted the excess medical costs due to hearing impairment as $3.3-$12.8 billion in the USA alone[9].

ARHI is expected to be a highly genetically heterogeneous trait given that over 150 genetic loci have been identified in non-syndromic congenital hearing loss alone (https://hereditaryhearingloss.org/). Previous GWAS of ARHI related phenotypes have only identified a small number of promising candidate genes, though there has been poor replication of findings to date, possibly reflecting varied phenotyping methods and limited sample sizes[10–19]. Alternatively, it might suggest that the genetic contribution to ARHI has been overestimated. To date, five genomic loci have been significantly associated with ARHL in previous GWAS[15,16,18] only two of which were replicated in independent population samples. Understanding the genetic aetiology of ARHI sheds light on the pathophysiology of the condition and ultimately facilitates the development of novel therapeutic or preventative interventions.

Here, we conducted two large hearing GWAS with sample sizes of over 250,000 individuals using the self-reported hearing difficulty and hearing aid use of UKBB participants and refined our results using a combination of conditional analysis and replication analysis. In addition, we performed *in silico* functional annotation and *in vivo* expression analysis (see Figure 1 for study design) to understand the role of gene variants in the biological mechanisms of ARHI.

Our aim was to identify the genetic components of adult hearing impairment in the UK population and provide insight into the pathology of ARHI and associated traits.

Subjects and methods

Participants

The sample used for this study consisted of individuals who participated in the UKBB study. The UKBB is a national resource, initially set up to study lifestyle and genetic factors affecting ageing traits with the aim of understanding and improving healthy ageing at a population level. Over 500,000 volunteers attended 23 assessment centres across the UK between 2007-2013 where they donated samples for genotyping, completed lifestyle questionnaires and have standard measurements taken. The UKBB resource is described extensively elsewhere[20].

The cohort used for discovery association analysis consisted of UKBB participants with 'White British' ancestry. The UKBB sample classification 'White British' is derived from both principal component (PC) analysis and self-declared ethnicity[21]. Samples with excess heterozygosity, excess relatedness and sex discrepancies were identified and removed prior to analysis, resulting in samples sizes of n =250,389 and n = 253,918 for hearing difficulty (*HDiff*) and hearing aid (*HAid*) use respectively.

For replication analysis, we used the UKBB ethnic group 'Caucasians' (white non-British Europeans). To assign participants into discrete ancestry clusters, we used the 1st and 2nd PC vectors provided by UKBB. A k-means clustering algorithm was applied to generate clusters for each PC. We then combined cluster indices for the PCs (1.1, 1.2, …, 5.5), compared them against self-reported ancestry and assigned the ancestry group accordingly. If contradictory, the pairwise clusters took precedence over the self-report grouping.

Two samples were used for replication analysis, the English Longitudinal Study of Ageing (ELSA) and TwinsUK. They were selected as they comprise predominantly Northern European ancestry samples and include relevant questionnaire data. ELSA is a longitudinal study, consisting of around 12,000 respondents from the Health Survey for England. Eight waves of data collection have been completed since 2002[22]. TwinsUK is the largest adult twin registry in the UK and comprises over 13,000 healthy twin volunteers aged 16-98. Collection of data and biological materials commenced in 1992 and is ongoing. During study participation, twins regularly complete health and lifestyle questionnaires and attend for clinical evaluation[23].

Phenotype definitions

Two phenotypes were derived for this study; a phenotype representing self-reported hearing difficulty (*HDiff*) and a phenotype representing self-reported hearing aid use (*HAid*). Participants in the UKBB study completed a touchscreen questionnaire during their visit to the assessment centre, which included questions regarding hearing status. Participants were assigned case/control status based on their responses to questionnaire measures regarding hearing difficulty and hearing aid use. *HDiff* cases responded "Yes" to both of the questions "Do you have any difficulty with your hearing" and "Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?" *HDiff* controls were selected if their response to both of these questions was "No". Participants with any other combination of responses were removed. In addition, *HDiff* controls aged <50 were removed from analysis, as were any controls that responded "Yes" to the question "Do you use a hearing aid most of the time?" *HAid* cases responded "Yes" to "Do you use a hearing aid most of the time?" and controls responded "No". Details of how the UKBB phenotype was derived are displayed in Figure S1.

If participants answered the questionnaire twice, i.e. attended an assessment centre for a repeat visit, the answer at the second time point was used in analysis, in order to increase the mean age of the sample. To reduce the likelihood of including congenital forms of deafness, participants who selected 'I am completely deaf' in the UKBB questionnaire were excluded from analysis. Note that a further, objective measure of hearing, the speech reception threshold using the 'Digits in Noise' (DIN) protocol, was obtained from 160,955 of the UK Biobank participants.[24,25] Preliminary heritability assessment of the DIN did not yield clear heritability or association with age and therefore it was not considered suitable for the present study.

Questionnaire responses for the ELSA and TwinsUK replication samples were derived to obtain comparable phenotypes to the UKBB phenotype (Figure S1). For the ELSA sample, case/control phenotypes were derived from responses to questionnaire measures collected during study Wave 7. The *HDiff* phenotype was derived using responses from two questions; "Do you ever have any difficulties

with your hearing?" and "Do you find it difficult to follow a conversation if there is background noise, such as TV, radio or children playing (using a hearing aid as usual)?" Cases were defined as participants who responded "Yes" to both questions, and controls who responded "No" to both questions. As in the UKBB analysis, controls who reported hearing aid use or age <50 were removed, as were any cochlear implant users in the case or control samples. The *HAid* phenotype was derived using the question "Whether ever wears a hearing aid"; cases responded "Yes most of the time", or "Yes some of the time" while controls responded "No". During ELSA data processing, age was capped at 90 years, and thus individuals aged > 90 were reported to be 90 years of age. Resulting ELSA samples sizes for association analysis of these traits were *HDiff* = 3545 and *HAid* = 4482.

The TwinsUK phenotypes were likewise derived from responses to questions. *HDiff* cases responded either "Yes, diagnosed by doctor or health professional" or "Yes, not diagnosed by health professional" to the question "Do you suffer from hearing loss?" while controls responded "No". *HAid* cases responded or indicated "Yes" to either of "Do you wear a hearing aid?" and 'Wearing a hearing aid'. *HAid* controls responded "No". As TwinsUK is a longitudinal study, a number of participants gave responses to the same questions on multiple occasions. The most recent response was included in analysis, unless the latest response indicated that hearing had improved. In this scenario, the participant was excluded. TwinsUK recruits adult twins aged over 18 years. Twins aged <40 were removed from analysis so that the lower age limit was comparable to the UKBB cohort. In order to retain the size of the TwinsUK sample and thus optimise power, controls aged <50 were not removed (as in the discovery *HDiff* UKBB analysis). Resulting Twins UK samples sizes for association analysis of these traits were *HDiff* = 3636 and *HAid* = 3435.

Genotyping and imputation

The ~500,000 samples in UKBB were genotyped on one of two arrays; 50,000 samples were genotyped on the Affymetrix UK BiLEVE Axiom array while the remaining ~450,000 were genotyped on the Affymetrix UK Biobank Axiom® array. The two arrays shared 95% coverage resulting in >800,000 genotyped SNPs. Imputation was carried out centrally by UKBB, primarily using the HRC reference panel and IMPUTE2[26]. SNPs which do not feature on this panel were imputed with the UK 10K and

1000G panel. Analysis in this study was conducted with version 3 of the UKBB imputed data with 487,409 samples imputed and available for analysis following UKBB centrally performed QC filters.

ELSA samples were genotyped at UCL Genomics in two batches using the Illumina HumanOmni 2.5M platform. Imputation was carried out centrally by ELSA with IMPUTE2, using the 1000 Genomes phase I data set[27] (see Web Resources).

Genotyping of TwinsUK was conducted with a combination of Illumina arrays; HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M. The imputation reference was 1000G Phase3 v5 (GRCh37).

Association analysis

Discovery association analysis was performed using a linear mixed-effects model approach to test for association between imputed SNP dosages and the two traits. BOLT-LMM v.2[28] was used for the association analysis, which corrects for population stratification and within-sample relatedness. In addition, the analysis was adjusted for age, sex, UKBB genotyping platform and UKBB PCs1-10. For quality control, SNPs were filtered based on two thresholds: (1) minor allele frequency (MAF) $\geq$ 0.01; and (2) INFO score > 0.7. By implementing a MAF cutoff of 0.01, we reduced the likelihood of including participants with forms of congenital deafness, as we only detected variants that occur at least in 1/100 participants, a higher frequency of variants than the frequency of congenital deafness. Individuals with < 98% genotype call rate were removed.

Conditional and joint analysis

Conditional and joint SNP analysis was performed to identify independent signals within highly associated regions, using GCTA-COJO[29]. This analysis requires the linkage disequilibrium reference sample, which was obtained by random selection of 10,000 individuals from the UKBB cohort with White British ancestry. The reference sample size was selected to maximise power based on previous data simulations[29]. The distance assumed for complete linkage equilibrium was 10Mb and a cut off value of R^2=0.9 was used to check for collinearity between the selected SNPs and those to be tested. Alleles with a frequency difference >0.2 between the reference sample and GWAS sample were

excluded. Independent SNPs identified with GCTA-COJO were mapped to the nearest protein coding gene using variant effect predictor (VEP)[30], genome build GRCh37. VEP was used to establish whether the SNP was in an exonic, intronic or intergenic region, and also the functional consequence of the variant at that position.

LD Score regression

Univariate linkage disequilibrium score regression (LDSC) was used to calculate whether inflated test statistics were likely due to the polygenic nature of the trait or confounding bias, by analysing the relationship between test statistic and LD[31].

We also performed genetic correlation analysis between the *HDiff* trait and 765 traits available for correlation analysis on LD hub[31,32]. To filter our results, we calculated a conservative significance threshold with a multiple-test correction (0.05/764, p = 6.5E-5) and selected those with a correlation (rg) >0.3 or < -0.3 to report in this study and grouped the remaining traits into five categories.

Heritability estimates

SNP heritability estimates for the two traits were calculated with BOLT-LMM (h2g). As *HDiff* and *HAid* are both qualitative traits, these estimates were recalculated to the liability scale. Sample and population prevalence were specified as the case prevalence in the analysed sample; *HDiff* at 0.35 and *HAid* at 0.052. SNP heritability was also calculated using a region-based approach with Heritability Estimation from Summary Statistics (HESS)[33]. SNP heritability was partitioned into 1702 approximately independent loci (see Web Resources). The EUR 1000G reference panel was used for LD estimation (see Web Resources).

Replication association analysis

The lead SNPs for each locus identified with conditional analysis (Table 1) were tested for association with *HDiff* and *HAid* phenotypes in each of the three cohorts UKBB (white non-British Europeans), TwinsUK and ELSA. The UKBB white non-British sample was examined using the same protocol as the White British dataset described above, under the linear mixed-effects models method with BOLT-LMM adjusting for age, sex, UKBB PCs 1-10 and genotyping platform. As BOLT-LMM is unsuitable

for analysis of samples with N<5000, alternative software was used for the TwinsUK and ELSA association analysis. The TwinsUK sample was analysed using a linear mixed-effects model regression adjusting for age and sex with GEMMA[34], the most suitable software for twins as it can control for family structure. For the ELSA sample, one of each pair of related individuals was excluded from analysis during central quality control checks at ELSA (relatedness was estimated in PLINK 1.9[35]), and PLINK2 logistic regression was used to test for association in the ELSA sample, adjusting for age and sex.

For SNPs significantly associated with hearing difficulty in the discovery, a fixed-effect inverse-variance weighted meta-analysis was conducted using METAL[36] version 2011-03-25 with the three samples: White non-British UKBB, ELSA and TwinsUK. BOLT-LMM does not report analysed sample size per SNP, so to obtain the weight of the UKBB replication sample per SNP, sample size was calculated from PLINK linear regression.

A power calculation[37] was performed for each independent locus analysed in the replication analysis ($p<0.05$ for both traits and $p<0.0012$ for *HDiff* and $p<0.00714$ for *HAid*, Table S1 and S2).

Gene prioritization, pathway and tissue enrichment analysis

Summary statistics from the UKBB *HDiff* trait were input for Functional Mapping and Annotation of Genome-wide Association Studies (FUMA[38]). Firstly, the SNP2GENE function was used to identify lead SNPs from *HDiff* analysis that reached a suggestive level of significance and were independent at $r^2<0.1$. These identified lead SNPs were used to perform gene set enrichment analysis with ToppGene Suite[39].

Secondly, gene-set analysis was performed with MAGMA[40] using all of the SNPs that were analysed in the *HDiff* discovery association analysis. Here, SNPs were mapped to protein coding genes that were within 10kb of the SNP location. The effects of multiple SNPs were combined to calculate the significance of an association of a gene with *HDiff*. These significantly associated genes were analysed for differential expression in 30 general tissue types.

Mouse tissue preparation

Adult mouse cochleae were collected at p28-p30 from C57BL/6 mice, bred in an in-house facility. Mice were euthanised according to Schedule 1 procedures as described in United Kingdom legislation outlined in the Animals (Scientific Procedures) Act 1986. Dissected inner ears were fixed in 4% paraformaldehyde diluted in PBS for 1 hour at room temperature before being washed several times in PBS. They were then decalcified in 10% EDTA overnight at 4^{o}C, before being separated from the vestibular system. Cochlea were mounted in 4% low-melting point agarose and sectioned on a Vibratome (1000 plus system, Intracel) at 200-µm intervals.

Immunofluorescence

Antibodies used to identify protein localisation in the organ of Corti were: anti-nidogen-2 (NID2) at a 1:750 dilution (Ab14513, Abcam), anti-clarin-2 (CLRN2) at 1:1000 (HPA042407, Atlas Antibodies) and anti-rho guanine nucleotide exchange factor 28 (ARHGEF28) at 1:1000 (HPA037602, Atlas Antibodies). All were detected using of an isotype-specific Alexa Fluor 488 goat anti-rabbit secondary antibody (Santa Cruz Biotechnology). Antibodies were diluted in a goat blocking solution (4% triton, 8% goat serum, 1g BSA, 50ml PHEM buffer) and sections were stained with primary antibodies overnight at 4^{o}C. Following PBS washes, sections were incubated with the secondary antibody at 1:1000 in darkness at room temperature for 2 hours. Phalloidin-Atto 647N to f-actin (Sigma-Aldrich, Gillingham, UK) and DAPI were added to the secondary antibody incubations at 1:1000 to stain hair cell stereocilia and DNA respectively. Samples were imaged using a Zeiss LSM 880 Airyscan 20x objective.

Results

Phenotype Definition

UKBB participants were categorised using a case-control design based on responses to questions regarding hearing difficulty (*HDiff*, n=498,281) and hearing aid use (*HAid*, n=316,629), (see methods and Figure S1 for details). Following quality control filters and selection of white British participants (described in methods), the final sample sizes used for association analyses were n=250,389 (87,056 cases and 163,333 controls) for *HDiff* , and n=253,918 (13,178 cases and 240,740 controls) for *HAid*.

Genome-wide Association analysis

A linear mixed-effects model was used to test for association between 9,740,198 SNPs and the two hearing traits, using BOLT-LMM v.2[28], which corrects for population stratification and within sample relatedness. The studies identified 2,080 and 240 SNPs at genome-wide significance (P<5E-08) for *HDiff* and *HAid* analysis, respectively (Figure 2 and Figure S2). Conditional and joint analysis using GCTA-COJO[29] identified that these SNPs represent 41 and seven independent loci associated with *HDiff* and *HAid*, respectively, including four loci common to both traits, thus resulting in 44 independent loci overall.

SNP heritability estimates for the two traits calculated with BOLT-LMM (h2g) were 0.117 ± 0.001 for *HDiff* and 0.029, ±.001 for *HAid*. Estimates recalculated to the liability scale are 0.19 and 0.13 for *HDiff* and *HAid* respectively. SNP heritability was also calculated using a region-based approach (Figure S3); one local heritability estimate was significant (Bonferroni corrected p<0.05 for 1702 loci) for *HDiff*, Chr5 base positions 71240456- 73759326, p= 1.16E-05. No regions were significant in *HAid* (Figure S3). The Variant Effect Predictor (VEP)[30] was used to map independent lead SNPs to the nearest protein coding genes, using the GRCh37 genomic reference. Of 41 independent SNPs associated with *HDiff*, six variants lie in exons, four of which result in missense mutations in *EYA4* (MIM: 603550), *CDH23* (MIM: 605516), *KLHDC7B* and *TRIOBP* (MIM: 609761), 21 SNPs lie within introns and 14 are intergenic (Table 1). Six of the independent SNPs associated with *HAid* reside in intronic regions and 1 is intergenic. Two highly significant independent associations with *HDiff* are found within 100 kb of

the *ARHGEF28* (MIM: 612790) gene, this locus is also highly associated with the *HAid* phenotype. Other gene loci common to both traits are *NID2* (MIM: 605399), *CTBP2* (MIM: 602619) and *EYA4* (see Figure S4 for locus plots).

Replication analysis

Replication was attempted for the lead SNPs (41 *HDiff* and 7 *HAid*) by meta-analysing three independent samples; the remaining Caucasians in the UKBB cohort (white, non-British Europeans), TwinsUK, and the English Longitudinal Study of Ageing (ELSA), totalling *HDiff* n = 30,765 and *HAid* n = 35,004. Two intronic SNPs, rs759016271 in *ZNF318* (MIM: 617512) and rs1566129 in *NID2*, reached significance in the *HDiff* replication analysis (Bonferroni correction 0.05/41=0.0012, P<0.0012), and a further intronic SNP, rs4597943 in *ARHGEF28* replicated in *HAid* analysis at the significance threshold (0.05/7=0.00714, P<0.00714). An additional 14 SNPs reached nominal significance and the power to detect each variant is also shown (Tables S1 and S2).

We investigated whether any of the candidate genes identified in adult hearing in previously published genetic association studies were replicated within the discovery White British sample (Table 2) and found only two previous variant associations located in close proximity to *ISG20* (MIM: 604533) and within *TRIOBP*, which were identified in a GWAS performed with data from electronic health records[18].

ISG20 is a novel association with hearing, but mutations in *TRIOBP* cause a form of autosomal recessive non-syndromic deafness, Deafness, autosomal recessive 28 (DFNB28, [MIM 609823]) [41,42]. No other lead variants from previous ARHI genetic studies were replicated at nominal level in our analysis, including the first reported ARHI associated gene variant in *GRM7*[11,13] (MIM: 604101).

Gene prioritization, pathway and tissue enrichment analysis

Functional gene annotation was undertaken with genes mapped from SNPs associated at a suggestive level (P<1E-05) in the *HDiff* association analysis (Figure 3). Genes were significantly enriched in a number of processes required for auditory function: synaptic activities, trans-synaptic signalling, nervous system processes, modulation of chemical synaptic transmission, positive dendritic spine morphogenesis, and inner ear morphology as well as cognition, learning or memory. These genes were

also significantly enriched with mouse phenotype ontologies, mostly relating to inner ear abnormalities and abnormal auditory brainstem response, and were significant at FDR 0.05 (Figure 3). As well as suggesting pathogenic pathways, this finding demonstrates the shared genetic pathology in mouse and human auditory systems, supporting the use of mouse models to study human auditory function.

In silico tissue-specific gene expression analysis undertaken with MAGMA[40] indicates a significant association between *HDiff* genes and transcription levels of genes in the brain (P = 5.4E-04; Figure S5).

LD Score Regression

The LD score regression intercepts for the two analyses were 1.032 for *HDiff* and 1.03 for *HAid*. The ratio (intercept-1)/(mean(χ^2)-1) for *HDiff* was 8% and represents the proportion of inflation in the χ^2 statistic that the intercept attributes to alternative explanations than polygenicity. The ratio for *HAid* was 5%. We also performed genetic correlation analysis between *HDiff* and the 764 available traits on LDSC[31]. After removing the 3 traits that were used to create the *HDiff* and *HAid* phenotypes used in this study, 153 traits were significantly correlated with *HDiff*. Here we have highlighted 41 traits (significant correlation with *HDiff* and an rg< -0.3 or rg> 0.3) and grouped these into five categories; 'Hearing', 'Low mood/depression', 'Pain', 'Breathing difficulties' and 'Health report/Subjective wellbeing' (Table S3). Regarding 'Hearing' traits, a strong positive correlation was observed between *HDiff* and self-reported tinnitus 'Now or most of the time' (rg = 0.6, SE 0.056, p=1.40E-26). Traits in the Low mood/depression group included 'frequency of tiredness/lethargy in last 2 weeks' rg = 0.41, SE 0.029, p = 2.79E-45), neuroticism score (rg = 0.31, SE 0.026, p = 1.94E-34), miserableness (rg = 0.33, SE 0.027, p = 2.69E-33) and whether 'seen doctor (GP) for nerves, anxiety, tension or depression' (rg = 0.35 SE 0.03, p = 5.90E-31).

Protein localisation of putative novel hearing genes

We investigated expression of three putative novel hearing genes *NID2, ARHGEF28* and *CLRN2* in adult mouse cochlea using immunohistochemistry based on a number of factors. The lead SNP in *NID2* in both *HDiff* and *HAid* is located in intron 5 and was also replicated in the *HDiff* meta-analysis. Two

independent lead SNPs were identified at the *ARHGEF28* locus in the *HDiff* analysis, along with a third SNP in the *HAid* analysis which replicated in the meta-analysis. The lead independent SNP at the *CLRN2* locus in the *HDiff* analysis is within 2kb of *CLRN2*, although several other genes are within 100kb. Because *CLRN1* (MIM: 606397), a paralog of *CLRN2*, is expressed in hair cells and mutations in *CLRN1* cause autosomal recessive Usher syndrome Type-3 (USH3, [MIM 276902]) with progressive sensorineural hearing loss,[43,44] we investigated whether clarin-2 is also expressed in the inner ear.

Immunostaining for nidogen-2, a basement membrane component encoded by *NID2*, was most prominent in the epithelial lining of the inner spiral sulcus between the tectorial membrane and the inner hair cell (Figure 4 and Figure S6), as well as localizing to nerve fibres and blood vessel basement membranes. Similar to clarin-1, clarin-2 immunostaining localised to the inner and outer hair cells, the primary sensory cells of sound detection (Figure 4 and Figure S6). *ARHGEF28* encodes Rho Guanine Nucleotide Exchange Factor 28, for which immunostaining was observed in both hair cells and the spiral ganglion neuron cell bodies and axons (Figure 4 and Figure S6).

Discussion

Using data from UKBB we have performed a GWASwith a sample size of over 250,000 individuals, identifying over 2,320 genome-wide significant SNPs, representing 44 independent associations with self-reported adult hearing loss in participants aged 40-69 years. Only one of these loci, *TRIOBP* has previously been associated with ARHI. Thirty-four of the loci are novel to hearing function while ten loci have previously been linked to some form of hearing loss either in mouse models or humans (Table 1). Two independent associations were found within 100 kb of the *ARHGEF28* locus. The findings represent a more than 400-fold increase in the number of reported genome-wide significant SNP associations and a nine-fold increase in independent genomic loci for ARHI-related phenotypes and provide a major breakthrough in revealing the genetic architecture of ARHI.

Pure tone audiometry, the gold standard measure of hearing, requires an audiologist, a quiet environment and significant clinical time and is therefore challenging to collect in the large samples needed for GWAS. Due to this limitation previous GWAS in well characterised audiometric cohorts have included fewer than 5,000 cases. The most recent study, by Hoffmann et al. (2016)[18] utilised ARHI related diagnoses in electronic health records to perform a larger study with 6,527 cases and 45,882 controls and identified 2 genome wide significant SNPs. Our study is an order of magnitude larger than previous reported work using 87,056(*HDiff*) and 13,178 (*HAid*) cases with total sample sizes over 250,000 for each trait. The increased sample size is achieved by utilising self–reported hearing data and self-report hearing aid use rather than using objective audiometric measures. The benefit of the increased power of this study is highlighted by the ability to detect a large number of associations by using self-report measures, as was previously demonstrated with another sensory trait[45].

Hearing aid users in the UK will have received a diagnosis of hearing impairment following a pure-tone audiometry test, and so this phenotype provides a good surrogate of abnormal pure tone audiometry. *HDiff* is a more subjective measure than would be provided by pure-tone audiometry and may well be influenced by psycho-social elements as well as hearing ability. However, it is known that a pure tone-audiometry test does not diagnose all hearing loss including ‘hidden hearing loss’ (a key symptom of which is difficulty hearing with background noise present), meaning that *HDiff* may be more

representative of "real world" hearing impairment. Therefore, different forms of hearing loss may be defined by the two phenotypes, which may have subtle differences in pathology and genetic risk variants. In addition, within the UKBB sample hearing aid use is correlated with socio-economic status and education level[46], providing evidence that environmental factors are associated with and may influence the UKBB phenotypes. These factors were not included as covariates for association analysis as the interaction with hearing loss is not currently fully understood.

A long-standing question in ARHI has been whether the susceptibility genes for ARHI would be variants in known congenital deafness genes or completely novel genes. Approximately one quarter of the genetic loci identified in the UKBB GWAS are known hearing loss genes while the rest are novel. For four of the 10 established hearing loci, this study provides the first link to hearing pathology in humans; *BAIAP2L2*[47] (MIM: 617536), *TUB*[48] (MIM: 601197), *SYNJ2*[49] (MIM: 609410), and *SPTBN1*[50] (MIM: 182790), as they have previously only been linked to hearing function in animal models. In addition, a number of these 10 loci have existing links to early onset, congenital deafness. *CDH23* which encodes cadherin-23, is a component of the stereocilia tip link, deflection of which results in opening of the mechano-transduction necessary for sound detection[51]. It has long been proposed as a candidate human ARHI gene since an exon skipping mutation of *cdh23* in the commonly used C57BL/6 strain of mice causes an accelerated age related hearing loss[52]. However, previous mutations in humans have been associated with a form of early onset recessive hearing loss, Deafness, autosomal recessive 12 (DFNB12, [MIM 601386]) and with Usher's syndrome type ID (USHID, [MIM 601067]) which causes early onset deafness and blindness[53]. This study provides a link between this locus and common adult hearing impairment, suggestive of multiple variants being present at the same loci but displaying different types of hearing phenotypes.

Four significant gene loci are common to both analyses; *EYA4*, *NID2*, *ARHGEF28* and *CTBP2*. Variants within *EYA4* have been reported in autosomal dominant non-syndromic hearing loss[54–56], Deafness autosomal dominant 10 (DFNA10, [MIM 601316]), while *NID2* and *ARHGEF28* are new associations with hearing impairment. *CTBP2*, though not previously linked to genetic risk of ARHI,

encodes C-terminal Binding Protein 2 a critical protein component of the inner ear hair cell pre-synaptic ribbon[57]. The remaining loci were distinct between the two studies.

SNP-based heritability recalculated to the liability scale for *HDiff* (19%) and *HAid* (13%) are at the mid-lower end of previous heritability estimates for ARHL[58–60]. This may be because this method only accounts for the additive genetic effects of the SNPs that were included in our analysis[61]. This sample is however larger than previous samples used for a heritability estimate of common adult hearing loss and the phenotype measures do differ to those used for previous estimates.

Genetic correlation analysis revealed a link between genetic data for common adult hearing loss and depression-associated traits and pain related traits . Studies have reported numerous epidemiological links between these depressive symptoms and hearing loss[1,2,62], but there is no evidence of a common genetic aetiology. This correlation and significance does not account for genetic confounding, meaning that intermediate factors could contribute or be causal of the correlations reported. Due to the nature of these significantly correlated traits, the correlation may be confounded by general wellbeing. In addition, sampling bias may be present due to variability between studies such as trait and covariate categorisation and inclusion[32].

Gene set enrichment analysis using ToppGene contributed further evidence that this trait is highly polygenic and has a heterogeneous pathology. Genes and terms associated with *HDiff* were enriched for abnormal inner ear abnormality, synaptic function, cognition and abnormal neuron morphology. These findings indicate that genetic factors have a role in the common dysfunction of a range of mechanisms within the auditory system. Connecting pathogenic genetic variants to distinct mechanisms within the auditory system will enable us to better understand the biological relevance of an individual's genetic risk for hearing loss, and also how individual subsystems are dysregulated in the ageing population. With the knowledge that certain gene groups have a role in particular cell or tissue types, studies can be targeted to a likely pathogenic location.

Analysis using ToppGene also revealed *HDiff* gene sets overlap with those involved in 'cognition', providing evidence of a possible genetic role to the link between hearing loss and cognition phenotypes.

However, the finding from MAGMA analysis of a significant association between *HDiff* genes and transcription levels of genes in brain tissue does not necessarily confirm this finding or that these genes cause pathology in higher auditory systems with age. The results from these expression studies may simply reflect the fact that sensory cells of the inner ear are of neural origin, or that a substantial amount of neuronal tissue expression data is available in comparison to the limited datasets derived from cochlear tissue.

Suggestive level associations between *HDiff* and mouse ontologies is a promising finding for the field. Phenotype driven screens in mice have been critically important in identifying deafness genes and are the model organism of choice for the study of hearing genetics, primarily due to the similarity of audiology systems and convergent evolution. However, there is less evidence that similar genetic predisposition to auditory ageing is comparable between mice and humans. These results, together with the identification of *CDH23* as a human ARHI loci, supports the use of mouse models in genetic studies of ARHI.

While Nidogen-2 is an established component of the basement membrane protein complex, anti-NID2 staining was not observed in the basilar membrane of the organ of Corti. It was however observed staining the blood vessel basement membranes, as has been noted in other tissues previously[63], providing evidence of antibody specificity in our sample. Anti-NID2 staining was most prominent in a restricted region of the epithelial lining between the inner spiral sulcus and spiral limbus suggesting it may have a role in maintaining the structure of the epithelia here in close proximity to the attachment site of the tectorial membrane.

Similar to clarin-1[64], clarin-2 immunostaining localised to the inner and outer hair cells, the primary sensory cells of sound detection, suggesting it may have a similar role in hearing (Figure 4,Figure S5). During preparation of the manuscript recent evidence from concurrent studies suggests that mutations in *clrn2* can cause progressive hearing loss in mouse and zebrafish and that mutations in *CLRN2* underlie a recessive form of congenital hearing loss in an Iranian family[64,65]. In addition to observing anti-ARHGEF28 staining in the sensory cells, anti-ARHGEF28 stains the lining of the nerve fibres in the SGN region. Previous reports demonstrate a role for *ARHGEF28* in regulation of neurofilaments[66,67]

and axon growth and branching[67]. It has also been implicated in the pathogenesis of motor neuron disease through formation of neurofilament and *ARHGEF28* aggregates[68], perhaps implying a role in neuronal function or maintenance.

Our study should be received in the context of its limitations; first, there is currently a lack of adequately powered studies with which to replicate our results. Despite meta-analysing three cohorts, the replication sample remains an order of magnitude smaller than the discovery set. It is therefore unsurprising that only three genes *ZNF318, NID2* and *ARHGEF28* were replicated. This replication rate likely reflects sample heterogeneity as well as 'winner's curse'[70] . This heterogeneity includes varied questionnaire measures used to derive the phenotypes in the three replication samples and the different demographics of the individual cohorts. The lack of samples with consistent phenotype measures is a key limitation in genetic hearing research. One of the three cohorts included in our replication meta-analysis is a subset of the UKBB cohort and therefore is not a truly independent sample from the discovery cohort. However, the identification of known hearing genes, gene annotation analysis and the results of *in vivo* expression provide support and putative mechanisms for involvement of these genes in hearing loss.

Second, we could not confirm the age of onset of hearing difficulty or hearing aid prescription, making an accurate diagnosis of ARHI a challenge. Some of the associations, for example, may be driven by individuals with congenital hearing impairment due to highly penetrant variants. We reduced the likelihood of this by implementing a minor allele frequencycut-off of 0.01 (i.e., higher than the known frequency of variants involved in congenital deafness) and by excluding participants who selected 'I am completely deaf' in the UKBB questionnaire.

In summary, we have conducted a GWAS on adult hearing using over 250,000 samples and have identified 44 associated independent loci. Several genes identified are known to have a role in congenital deafness, are already known to have an important role in hearing or have been identified in mouse models. One locus has been associated with human ARHI traits previously and 34 of the 44 loci identified have not previously been associated with hearing loss phenotypes in humans or mice. Interestingly, these genes have a range of biological functions and are suggestive of multiple pathogenic

pathways underlying ARHI and are not restricted to only one of sensory or neural or metabolic forms hearing loss. For three genes we demonstrate different localised cell specific expression within the mouse adult cochlea. Importantly, this study demonstrates that self-reported hearing loss in adults is suitable for use in association studies using large cohorts such as the UKBB. Our results present a framework for further study into the auditory pathways influenced by the genomic loci identified and provide therapeutic opportunities in ARHI and possibly dementia.

Description of Supplemental Data

Supplemental Data includes Six figures and three tables.

Declaration of interests

David R. Moore is a Scientific Advisor for hearX Ltd., Otonomy Inc. The other authors declare no competing interests.

Acknowledgements

The research was carried out using the UK Biobank Resource under application number 11516. HRRW is funded by a PhD Studentship Grant, S44, from Action on Hearing Loss. The study was also supported by funding from NIHR UCLH BRC Deafness and Hearing Problems Theme, a grant from MED_EL and the NIHR Manchester Biomedical Research Centre. The English Longitudinal Study of Ageing is jointly run by University College London, Institute for Fiscal Studies, University of Manchester and National Centre for Social Research. Genetic analyses have been carried out by UCL Genomics and funded by the Economic and Social Research Council and the National Institute on Ageing. All ELSA GWAS data have been deposited in the European Genome-phenome Archive (Database: EGAS00001001036). Data governance was provided by the METADAC data access committee, funded by ESRC, Wellcome, and MRC. (2015-2018: Grant Number MR/N01104X/1 2018-2020: Grant Number ES/S008349/1). TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility, and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. We would like to thank all the participants of UK Biobank, English Longitudinal Study of Ageing and TwinsUK.

Availability of data

Data that support the findings of this study are publicly available upon successful application from the UK Biobank, the English Longitudinal Study of Ageing and TwinsUK. All ELSA GWAS data have been deposited in the European Genome-phenome Archive (Database: EGAS00001001036).

Derived data from the UK Biobank data fields and the discovery GWAS summary statistics that support the findings of this study, will be made available as part of the UK Biobank Returns Catalogue following the publication of this manuscript, see Web Resources.

Web Resources

ELSA GWAS QC & Imputation Analysis

https://www.ucl.ac.uk/drupal/site_iehc/sites/iehc/files/elsa_gwas_qc_imputation_analysis.pdf

Region-based heritability estimates genome partition map

https://bitbucket.org/nygcresearch/ldetect-data/downloads/

Region-based heritability reference panel, EUR 1000G

http://www.internationalgenome.org/data/

UK Biobank datashowcase

http://biobank.ctsu.ox.ac.uk/crystal/

References

1. Brewster, K.K., Ciarleglio, A., Brown, P.J., Chen, C., Kim, H.O., Roose, S.P., Golub, J.S., and Rutherford, B.R. (2018). Age-Related Hearing Loss and Its Association with Depression in Later Life. Am. J. Geriatr. Psychiatry *26*, 788–796.

2. Rutherford, B.R., Brewster, K., Golub, J.S., Kim, A.H., and Roose, S.P. (2018). Sensation and psychiatry: Linking age-related hearing loss to late-life depression and cognitive decline. Am. J. Psychiatry *175*, 215–224.

3. Han, J.H., Lee, H.J., Jung, J., and Park, E.C. (2018). Effects of self-reported hearing or vision impairment on depressive symptoms: a population-based longitudinal study. Epidemiol. Psychiatr. Sci. *28*, 343–355.

4. Jayakody, D.M.P., Friedland, P.L., Martins, R.N., and Sohrabi, H.R. (2018). Impact of aging on the auditory system and related cognitive functions: A narrative review. Front. Neurosci. *12*, 125.

5. Wei, J., Hu, Y., Zhang, L., Hao, Q., Yang, R., Lu, H., Zhang, X., and Chandrasekar, E.K. (2017). Hearing impairment, mild cognitive impairment, and dementia: A meta-analysis of cohort studies. Dement. Geriatr. Cogn. Dis. Extra *7*, 440–452.

6. Gurgel, R.K., Ward, P.D., Schwartz, S., Norton, M.C., Foster, N.L., and Tschanz, J.T. (2014). Relationship of Hearing Loss and Dementia. Otol. Neurotol. *35*, 775–781.

7. Lin, F.R., Metter, E.J., O'Brien, R.J., Resnick, S.M., Zonderman, A.B., and Ferrucci, L. (2011). Hearing loss and incident dementia. Arch. Neurol. *68*, 214–220.

8. Vos, T., Barber, R.M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F., Davis, A., Degenhardt, L., Dicker, D., et al. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013.

9. Huddle, M.G., Goman, A.M., Kernizan, F.C., Foley, D.M., Price, C., Frick, K.D., and Lin, F.R. (2017). The economic impact of adult hearing loss: A systematic review. JAMA Otolaryngol. - Head Neck Surg. *143*, 1040–1048.

10. Fransen, E., Bonneux, S., Corneveaux, J.J., Schrauwen, I., Di Berardino, F., White, C.H., Ohmen, J.D., Van De Heyning, P., Ambrosetti, U., Huentelman, M.J., et al. (2015). Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. Eur. J. Hum. Genet. *23*, 110–115.

11. Newman, D.L., Fisher, L.M., Ohmen, J., Parody, R., Fong, C.T., Frisina, S.T., Mapes, F., Eddins, D.A., Frisina, R.D., Frisina, R.D., et al. (2012). GRM7 variants associated with age-related hearing loss based on auditory perception. Hear. Res. *294*, 125–132.

12. Van Laer, L., Huyghe, J.R., Hannula, S., Van Eyken, E., Stephan, D.A., Mäki-Torkko, E., Aikio, P., Fransen, E., Lysholm-Bernacchi, A., Sorri, M., et al. (2010). A genome-wide association study for age-related hearing impairment in the Saami. Eur. J. Hum. Genet. *18*, 685–693.

13. Luo, H., Yang, T., Jin, X., Pang, X., Li, J., Chai, Y., Li, L., Zhang, Y., Zhang, L., Zhang, Z., et al. (2013). Association of GRM7 Variants with Different Phenotype Patterns of Age-Related Hearing Impairment in an Elderly Male Han Chinese Population. PLoS One *8*, e77153.

14. Friedman, R.A., Van Laer, L., Huentelman, M.J., Sheth, S.S., Van Eyken, E., Corneveaux, J.J., Tembe, W.D., Halperin, R.F., Thorburn, A.Q., Thys, S., et al. (2009). GRM7 variants confer susceptibility to age-related hearing impairment. Hum. Mol. Genet. *18*, 785–796.

15. Wolber, L.E., Girotto, G., Buniello, A., Vuckovic, D., Pirastu, N., Lorente-Cánovas, B., Rudan, I., Hayward, C., Polasek, O., Ciullo, M., et al. (2014). Salt-inducible kinase 3, SIK3, is a new gene

associated with hearing. Hum. Mol. Genet. *23*, 6407–6418.

16. Vuckovic, D., Dawson, S., Scheffer, D.I., Rantanen, T., Morgan, A., Di Stazio, M., Vozzi, D., Nutile, T., Concas, M.P., Biino, G., et al. (2015). Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss. Hum. Mol. Genet. *24*, 5655–5664.

17. Girotto, G., Pirastu, N., Sorice, R., Biino, G., Campbell, H., d'Adamo, A.P., Hastie, N.D., Nutile, T., Polasek, O., Portas, L., 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*, 369–374.

18. Hoffmann, T.J., Keats, B.J., Yoshikawa, N., Schaefer, C., Risch, N., and Lustig, L.R. (2016). A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records. PLoS Genet. *12*, e1006371.

19. Nolan, L.S., Maier, H., Hermans-Borgmeyer, I., Girotto, G., Ecob, R., Pirastu, N., Cadge, B.A., Hübner, C., Gasparini, P., Strachan, D.P., et al. (2013). Estrogen-related receptor gamma and hearing function: evidence of a role in humans and mice. Neurobiol. Aging *34*, 2077.e1-2077.e9.

20. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med. *12*, e1001779.

21. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants Supplementary Material. BioRxiv 166298.

22. Steptoe, A., Breeze, E., Banks, J., and Nazroo, J. (2013). Cohort profile: The English Longitudinal Study of Ageing. Int. J. Epidemiol. *42*, 1640–1648.

23. Moayyeri, A., Hammond, C.J., Hart, D.J., and Spector, T.D. (2013). The UK adult twin registry (twinsUK resource). Twin Res. Hum. Genet. *16*, 144–149.

24. Dawes, P., Fortnum, H., Moore, D.R., Emsley, R., Norman, P., Cruickshanks, K., Davis, A., Edmondson-Jones, M., McCormack, A., Lutman, M., et al. (2014). Hearing in middle age: a population snapshot of 40–69 year olds in the UK. Ear Hear. *35*, e44–e51.

25. Moore, D.R., Edmondson-Jones, M., Dawes, P., Fortnum, H., McCormack, A., Pierzycki, R.H., and Munro, K.J. (2014). Relation between speech-in-noise threshold, hearing loss and cognition from 40-69 years of age. PLoS One *9*, e107720.

26. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.

27. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.

28. Loh, P.R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. *47*, 284–290.

29. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375.

30. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

31. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani,

G., Tansey, K., Laurin, C., Pourcain, B.S., et al. (2017). LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics *33*, 272–279.

32. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

33. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am. J. Hum. Genet. *99*, 139–153.

34. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824.

35. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

36. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.

37. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. *101*, 5–22.

38. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826.

39. Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. *37*, W305–W311.

40. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput. Biol. *11*, e1004219.

41. Shahin, H., Walsh, T., Sobe, T., Abu Sa'ed, J., Abu Rayan, A., Lynch, E.D., Lee, M.K., Avraham, K.B., King, M.-C., and Kanaan, M. (2005). 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*, 144–152.

42. Riazuddin, S., Khan, S.N., Ahmed, Z.M., Ghosh, M., Caution, K., Nazli, S., Kabra, M., Zafar, A.U., Chen, K., Naz, S., et al. (2005). Mutations in TRIOBP, Which Encodes a Putative Cytoskeletal-Organizing Protein, Are Associated with Nonsyndromic Recessive Deafness. Am. J. Hum. Genet. *78*, 137–143.

43. Fields, R.R., Zhou, G., Huang, D., Davis, J.R., Möller, C., Jacobson, S.G., Kimberling, W.J., and Sumegi, J. (2002). Usher Syndrome Type III: Revised Genomic Structure of the USH3 Gene and Identification of Novel Mutations. Am. J. Hum. Genet. *71*, 607–617.

44. Adato, A., Vreugde, S., Joensuu, T., Avidan, N., Hamalainen, R., Belenkiy, O., Olender, T., Bonne-Tamir, B., Ben-Asher, E., Espinos, C., et al. (2002). USH3A transcripts encode clarin-1, a four-transmembrane-domain protein with a possible role in sensory synapses. Eur. J. Hum. Genet. *10*, 339–350.

45. Kiefer, A.K., Tung, J.Y., Do, C.B., Hinds, D.A., Mountain, J.L., Francke, U., and Eriksson, N. (2013). Genome-Wide Analysis Points to Roles for Extracellular Matrix Remodeling, the Visual Cycle, and Neuronal Development in Myopia. PLoS Genet. *9*, e1003299.

46. Sawyer, C.S., Armitage, C.J., Munro, K.J., Singh, G., and Dawes, P.D. (2019). Correlates of Hearing Aid Use in UK Adults. Ear Hear. 1.

47. Bowl, M.R., Simon, M.M., Ingham, N.J., Greenaway, S., Santos, L., Cater, H., Taylor, S., Mason, J., Kurbatova, N., Pearson, S., et al. (2017). A large scale hearing loss screen reveals an extensive

unexplored genetic landscape for auditory dysfunction. Nat. Commun. *8*, 886.

48. Ohlemiller, K.K., Hughes, R.M., Mosinger-Ogilvie, J., Speck, J.D., Grosof, D.H., and Silverman, M.S. (1995). Cochlear and retinal degeneration in the Tubby mouse. Neuroreport *6*, 845–849.

49. Manji, S.S.M., Williams, L.H., Miller, K.A., Ooms, L.M., Bahlo, M., Mitchell, C.A., and Dahl, H.H.M. (2011). A mutation in synaptojanin 2 causes progressive hearing loss in the ENU-mutagenised mouse strain mozart. PLoS One *6*, e17607.

50. Liu, Y., Qi, J., Chen, X., Tang, M., Chu, C., Zhu, W., Li, H., Tian, C., Yang, G., Zhong, C., et al. (2019). Critical role of spectrin in hearing development and deafness. Sci. Adv. *5*, eaav7803.

51. Richardson, G.P., and Petit, C. (2019). Hair-Bundle Links: Genetics as the Gateway to Function. Cold Spring Harb. Perspect. Med. a033142.

52. Noben-Trauth, K., Zheng, Q.Y., and Johnson, K.R. (2003). Association of cadherin 23 with polygenic inheritance and genetic modification of sensorineural hearing loss. Nat. Genet. *35*, 21–23.

53. Schultz, J.M., Bhatti, R., Madeo, A.C., Turriff, A., Muskett, J.A., Zalewski, C.K., King, K.A., Ahmed, Z.M., Riazuddin, S., Ahmad, N., et al. (2011). Allelic hierarchy of CDH23 mutations causing non-syndromic deafness DFNB12 or Usher syndrome USH1D in compound heterozygotes. J. Med. Genet. *48*, 767–775.

54. Schönberger, J., Wang, L., Shin, J.T., Sang, D.K., Depreux, F.F.S., Zhu, H., Zon, L., Pizard, A., Kim, J.B., MacRae, C.A., et al. (2005). Mutation in the transcriptional coactivator EYA4 causes dilated cardiomyopathy and sensorineural hearing loss. Nat. Genet. *37*, 418–422.

55. Makishima, T., Madeo, A.C., Brewer, C.C., Zalewski, C.K., Butman, J.A., Sachdev, V., Arai, A.E., Holbrook, B.M., Rosing, D.R., and Griffith, A.J. (2007). Nonsyndromic hearing loss DFNA10 and a novel mutation of EYA4: Evidence for correlation of normal cardiac phenotype with truncating mutations of the Eya domain. Am. J. Med. Genet. Part A *143*, 1592–1598.

56. Pfister, M., Tóth, T., Thiele, H., Haack, B., Blin, N., Zenner, H.-P., Sziklai, I., Nürnberg, P., and Kupka, S. (2002). A 4-bp insertion in the eya-homologous region (eyaHR) of EYA4 causes hearing impairment in a Hungarian family linked to DFNA10. Mol. Med. *8*, 607–611.

57. Sheets, L., Trapani, J.G., Mo, W., Obholzer, N., and Nicolson, T. (2011). Ribeye is required for presynaptic Ca V 1.3a channel localization and afferent innervation of sensory hair cells . Development *138*, 1309–1319.

58. Bedin, E., Franzè, A., Zadro, C., Persico, M.G., Ciullo, M., Hladnik, U., Dolcetta, D., Grasso, D.L., Riccardi, P., Nutile, T., et al. (2009). Age-related hearing loss in four Italian genetic isolates: An epidemiological study. Int. J. Audiol. *48*, 465–472.

59. Bogo, R., Farah, A., Johnson, A.C., Karlsson, K.K., Pedersen, N.L., Svartengren, M., and Skjönsberg, Å. (2015). The Role of Genetic Factors for Hearing Deterioration Across 20 Years: A Twin Study. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. *70*, 647–653.

60. Gates, G.A., Couropmitrcc, N.N., and Alycrs, R.H. (1999). Genetic associations in age-related hearing thresholds. Arch. Otolaryngol. - Head Neck Surg. *125*, 654–659.

61. Zaitlen, N., and Kraft, P. (2012). Heritability in the genome-wide association era. Hum. Genet. *131*, 1655–1664.

62. Keidser, G., Seeto, M., Rudner, M., Hygge, S., and Rönnberg, J. (2015). On the relationship between functional hearing and depression. Int. J. Audiol. *54*, 653–664.

63. Kohfeldt, E., Sasaki, T., Göhring, W., and Timpl, R. (1998). Nidogen-2: A new basement membrane protein with diverse binding properties. J. Mol. Biol. *282*, 99–109.

64. Dunbar, L., Patni, P., Aguilar, C., Mburu, P. Corns, L., Wells, H.R.R., Delmaghani, S., Parker, A.

Johnson, S., Chessum, L., et al. (2019). Clarin-2 is Required for Maintenance of Mechanotransducing Stereocilia and Essential for Hearing. In ARO Midwinter Meeting Abstract, p. https://cdn.ymaws.com/www.aro.org/resource/resmgr/.

65. Gopal, S.R., Vona, B., Maroofian, R., Azaiez, H., Booth, K.T., Clancy, K., Mazaheri N., S., and G., Sedaghat, A., Stepanyan, R., et al. (2019). Mutation in the Clarin-2 Gene Cause Hearing Loss in Human and a Zebrafish Model Reveals the Likely Cause of that Hearing Loss. In ARO Midwinter Meeting Abstract, p. https://cdn.ymaws.com/www.aro.org/resource/resmgr/.

66. Cañete-Soler, R., Wu, J., Zhai, J., Shamim, M., and Schlaepfer, W.W. (2001). p190RhoGEF Binds to a Destabilizing Element in the 3′ Untranslated Region of Light Neurofilament Subunit mRNA and Alters the Stability of the Transcript. J. Biol. Chem. *276*, 32046–32050.

67. Volkening, K., Leystra-Lantz, C., and Strong, M.J. (2010). Human low molecular weight neurofilament (NFL) mRNA interacts with a predicted p190RhoGEF homologue (RGNEF) in humans. Amyotroph. Lateral Scler. *11*, 97–103.

68. Rico, B., Beggs, H.E., Schahin-Reed, D., Kimes, N., Schmidt, A., and Reichardt, L.F. (2004). Control of axonal branching and synapse formation by focal adhesion kinase. Nat. Neurosci. *7*, 1059–1069.

69. Droppelmann, C.A., Wang, J., Campos-Melo, D., Keller, B., Volkening, K., Hegele, R.A., and Strong, M.J. (2013). Detection of a novel frameshift mutation and regions with homozygosis within ARHGEF28 gene in familial amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Front. Degener. *14*, 444–451.

70. Kraft, P., Zeggini, E., and Ioannidis, J.P.A. (2010). Replication in Genome-Wide Association Studies. Stat. Sci. *24*, 561–573.

Figure Titles and Legends

Figure 1. Workflow schematic for discovery and validation of associated loci. N, sample size; QC, quality control; PC, principal components; MAF, minor allele frequency; INFO, quality metric, combination of imputation score and dosage confidence

Figure 2. Manhattan plots displaying GWAS results for (A) Hearing difficulty, and (B) Hearing aid use phenotypes. The Manhattan plots display the P values of all SNPs tested in discovery analysis. The threshold for genome wide significance ($p<5x10-8$) is indicated by a red dotted line. Loci that reached genome-wide significance in both phenotypes are annotated with gene symbol.

Figure 3. Heatmap of the enriched functional terms for genes mapped to lead SNP at suggestive level (*HDiff* analysis), using ToppGene Suite. Genes and functional terms were grouped using clustering of the strength of the enrichment of genes for respective functional terms. Functional terms include GO Biological Process, GO Molecular Function, GO Cellular Component, Mouse Phenotype, Pathway, and Disease.

Figure 4. Cochlear expression of three putative hearing genes identified in *HDiff* and *HAid* GWAS. (A,B,C,D) Locus zoom plots of associated loci, generated with *HDiff* summary statistics. Four associated loci are plotted which have lead SNPs in or in proximity to ARHGEF28 (A,B), NID2 (C), and CLRN2 (D). Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. Colouring of remaining SNPs is based on linkage disequilibrium (LD) with the lead SNP. The genes within the region are annotated, and the direction of the transcripts is shown by arrows. Two independent regions were identified within the ARHGEF28 locus; both are shown. (E,F,G,H) Immunofluorescence images of adult mouse cochlea, spiral ganglion neurons (E) and organ of Corti (F-H). Vibratome sections stained with the three proteins of interest in mouse inner ear; DAPI (blue) and Phalloidin (magenta) were also used for staining of actin and nuclei respectively. (E) Anti-ARHGEF28 staining is observed in the neuronal cell bodies and axons. (F) Anti-ARHGEF28 (green) is mainly observed in outer and inner hair cells. (G) Anti-NID2 (green) staining is observed lining blood vessels and the epithelial lining of the inner spiral sulcus. (H) Anti-CLRN2 (green) staining is observed in outer and inner hair cells, in addition to the stria vascularis. The scale bar in image (G) represents 100µm. The scale is consistent for all images in this figure.

Tables

Table 1. Results output from BOLT-LMM and GCTA-COJO. Chr., chromosome; SNP, single nucleotide polymorphism; EA, effect allele; EAF, frequency of effect allele in BOLT-LMM; INFO, quality metric, combination of imputation score and dosage confidence; β, effect size from BOLT-LMM approximation to infinitesimal mixed model; SE, standard error of the effect size; *p*-value, infinitesimal mixed-effects model association test p-value; *p*J-value, p-value from a joint analysis of all the selected SNPs; Nearest Gene, protein-coding gene in closest proximity to SNP; Distance to gene (bp), distance in base pairs between SNP and nearest gene, a value of 0 indicates the SNP lies within the gene; Other genes within 100kb, list of genes within 100kb of the SNP. [a] denotes genes previously linked to hearing phenotypes in mice or humans. Two SNPs reached genome-wide significance in the *HAid* analysis that are in close proximity to HLA-DQA1 on Chr 6 (Figure 2) but were not present in conditional analysis results.

Table 1. Independent SNPs significantly associated (P<5 × 10−8) with the two phenotypes regarding hearing ability in the UK Biobank discovery sample.

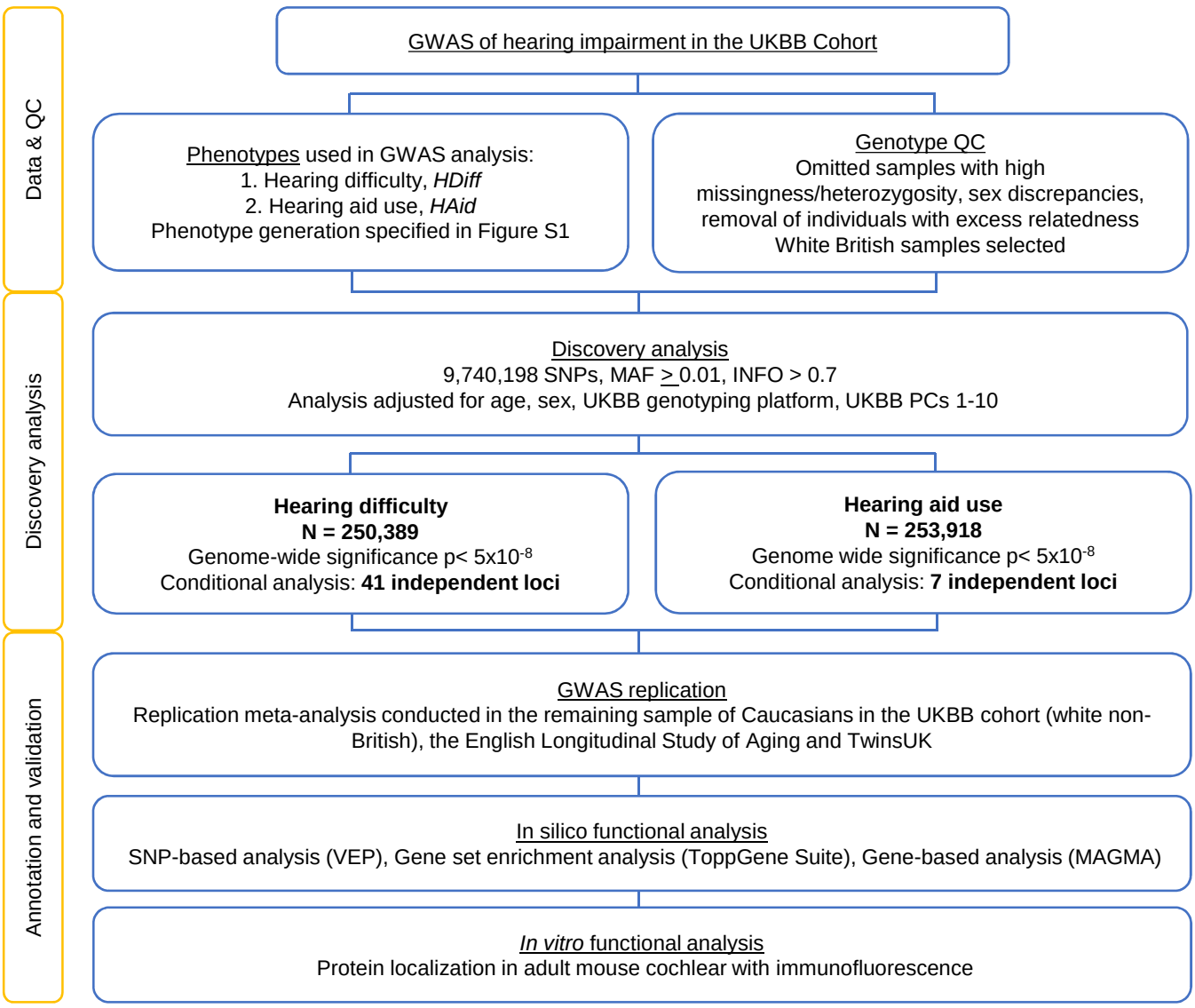
Hearing Difficulty GWAS											
Chr	SNP	EA	EAF	INFO	β	SE	*p*-value	*pJ*-value	Nearest Gene	Distance to gene (bp)	Other genes within 100kb
22	rs36062310	G	0.96	1.000	-0.0315	0.003	1.90E-22	1.92E-22	*KLHDC7B*	0	*SYCE3, ADM2, ARSA[a], CHKB, CPT1B, LMF2, MAPK8IP2, MIOX, NCAPH2, ODF3B, SBF1, SCO2, SYCE3, TYMP*
5	rs6453022	C	0.50	1.000	-0.0126	0.001	1.70E-21	2.07E-12	*ARHGEF28*	0	-
6	rs759016271	AGTAGTC CACTTTT CTTCTTT GCCTG	0.39	0.997	-0.0127	0.001	6.10E-21	6.16E-21	*ZNF318*	0	*CRIP3, SLC22A7, CUL9, DNPH1, TTBK1*
5	rs6890164	A	0.51	0.993	0.0119	0.001	3.30E-19	4.15E-10	*ARHGEF28*	6177	-
11	rs7951935	G	0.62	0.996	-0.0114	0.001	7.80E-17	7.85E-17	*TYR*	1472	*NOX4*
6	rs35186928	G	0.62	0.991	-0.0109	0.001	1.70E-15	1.69E-15	*HLA-DQA1*	13352	*HLA-DRB1, HLA-DRB3, HLA-DRB5, HLA-DRB6*
6	rs9493627	G	0.68	1.000	-0.0104	0.001	1.40E-13	1.41E-13	*EYA4[a]*	0	-
22	rs132929	G	0.59	0.999	-0.0098	0.001	2.20E-13	4.61E-13	*BAIAP2L2[a]*	0	*SLC16A8, PICK1 , PLA2G6, POLR2F*
22	rs5756795	T	0.54	1.000	-0.0092	0.001	5.10E-12	1.09E-11	*TRIOBP[a]*	0	*GALR3, GCAT, GGA1[a], H1F0, LGALS1, NOL12, PDXP, SH3BP1*
14	rs1566129	T	0.41	1.000	0.0091	0.001	1.40E-11	1.37E-11	*NID2*	0	*GNG2, RTRAF*
4	rs35414371	T	0.87	0.998	-0.0131	0.002	1.60E-11	1.64E-11	*CLRN2[a]*	1965	*LAP3, MED28, QDPR*
3	3:182069497_ TA_T	TA	0.84	0.989	-0.0118	0.002	4.10E-11	4.07E-11	*ATP11B*	441791	-
11	rs12225399	G	0.65	0.989	-0.009	0.001	8.60E-11	8.67E-11	*PHLDB1*	0	*ARCN1, IFT46, KMT2A, TMEM25, TREH, TTC36*
11	rs55635402	A	0.81	0.996	0.0105	0.002	2.90E-10	2.94E-10	*TUB[a]*	0	*EIF3F, NLRP10, OR10A3, RIC3*
16	rs62033400	A	0.61	0.999	0.0085	0.001	2.90E-10	2.95E-10	*FTO*	0	*RPGRIP1L*
8	rs13277721	G	0.49	0.992	-0.0083	0.001	3.30E-10	3.35E-10	*AGO2*	0	*PTK2*
2	rs62188635	C	0.45	0.988	0.0083	0.001	4.70E-10	4.72E-10	*KLF7*	50519	-
6	rs2236401	C	0.49	0.997	-0.0081	0.001	9.30E-10	9.38E-10	*SYNJ2[a]*	0	*SERAC1[a], GTF2H5*
7	rs4947828	T	0.23	0.999	-0.0096	0.002	1.00E-09	1.02E-09	*GRB10*	0	-
10	rs6597883	T	0.84	0.989	0.0111	0.002	1.00E-09	1.05E-09	*CTBP2*	0	-
5	rs34442808	T	0.49	0.992	-0.008	0.001	1.30E-09	1.32E-09	*MCTP1, SLF1*	0	-
10	rs835267	A	0.53	0.996	0.008	0.001	1.60E-09	1.58E-09	*EXOC6*	0	*CYP26A1, CYP26C1*
10	rs4948502	T	0.57	0.995	0.0081	0.001	1.70E-09	5.63E-10	*ARID5B*	0	-
10	rs10824108	G	0.42	0.999	-0.0079	0.001	3.00E-09	1.24E-08	*ADK*	0	*AP3M1, VCL*
1	rs12027345	G	0.57	0.995	0.0079	0.001	3.60E-09	3.64E-09	*MAST2*	12668	*GPBP1L1, MAST2, TMEM69, TMA16P2, GPBP1L1*
6	rs217289	G	0.56	0.992	-0.0078	0.001	4.90E-09	4.92E-09	*SNAP91*	0	-
3	rs13093972	A	0.55	0.992	-0.0078	0.001	5.50E-09	5.56E-09	*ZBTB20*	121137	-
15	rs62015206	C	0.41	1.000	-0.0078	0.001	7.70E-09	7.76E-09	*MAPK6*	15613	*BCL2L10, GNB5*
5	rs10475169	A	0.88	1.000	-0.0117	0.002	9.30E-09	9.37E-09	*IRX2*	190445	-
17	rs17671352	T	0.38	0.999	0.0078	0.001	1.00E-08	1.43E-08	*ACADVL*	0	*DVL2[a], DLG4, ASGR1, CLDN7, CTDNEP1, EIF5A, ELP5, GABARAP, GPS2, NEURL4, PHF23, SLC2A4, YBX2*
1	rs7525101	C	0.56	1.000	-0.0075	0.001	1.50E-08	1.45E-08	*LMX1A[a]*	61973	-
17	rs12938775	G	0.50	1.000	0.0075	0.001	1.60E-08	2.25E-08	*PAFAH1B1*	0	*CLUH, RAP1GAP2*
8	rs76837345	A	0.93	0.997	-0.0146	0.003	1.90E-08	1.95E-08	*CHMP4C*	0	*IMPA1, SLC10A5, SNX16, ZFAND1*
6	rs9366417	G	0.26	0.993	0.0085	0.002	2.10E-08	2.12E-08	*SOX4*	291019	-
8	rs3890736	G	0.63	0.993	-0.0077	0.001	2.20E-08	2.22E-08	*GFRA2*	15676	-
10	rs143282422	G	0.99	1.000	-0.0349	0.006	2.40E-08	3.02E-08	*CDH23[a]*	0	*C10orf105*
7	rs9691831	A	0.42	0.995	-0.0074	0.001	3.10E-08	3.11E-08	*TMEM213*	0	*ATP6V0A4[a] , KIAA1549*
11	rs141403654	A	0.98	0.878	-0.0313	0.006	3.50E-08	3.53E-08	*AGBL2*	0	*C1QTNF4, FNBP4, MTCH2 , NUP160*
18	rs4611552	T	0.78	0.995	-0.0089	0.002	3.60E-08	3.56E-08	*CCDC68*	9362	-
13	rs12552	A	0.44	0.994	0.0073	0.001	4.80E-08	4.86E-08	*OLFM4*	0	-
1	rs10927035	C	0.35	0.995	-0.0075	0.001	4.90E-08	4.89E-08	*AKT3*	0	*SDCCAG8*
Hearing Aid GWAS											
Chr	SNP	EA	EAF	INFO	β	SE	*p*-value	*pJ*-value	Nearest Gene	Distance to gene (bp)	Other genes within 100kb

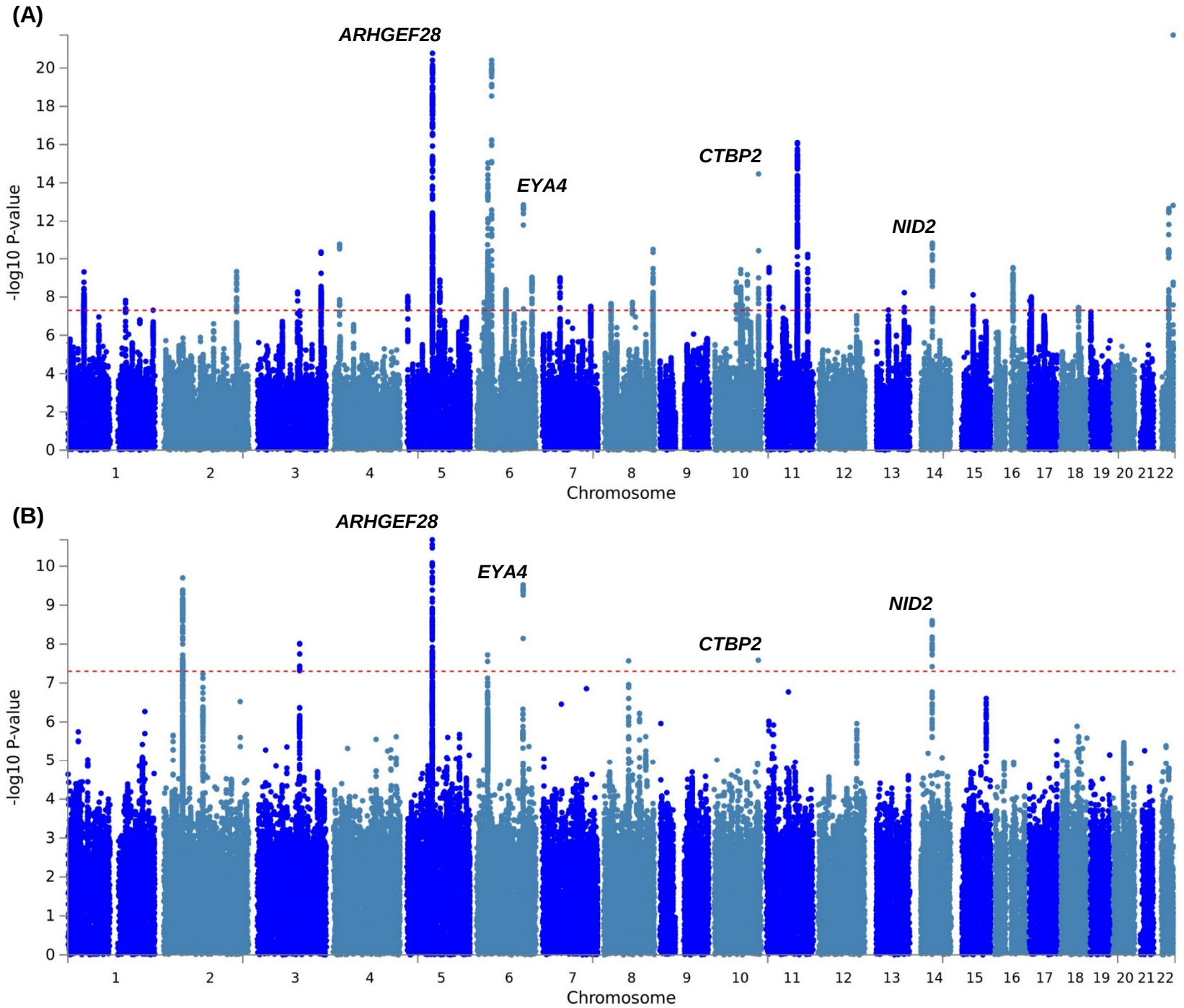
5	rs4597943	G	0.51	0.989	-0.0042	0.001	2.10E-11	2.09E-11	*ARHGEF28*	0	-
2	rs9677089	A	0.75	0.989	-0.0046	0.001	2.00E-10	1.98E-10	*SPTBN1*[a]	0	-
6	rs9321402	G	0.68	0.999	-0.0042	0.001	3.00E-10	3.02E-10	*EYA4*[a]	0	-
14	rs1566129	T	0.41	1.000	0.0037	0.001	2.50E-09	2.53E-09	*NID2*	0	*RTRAF*
3	rs3915060	C	0.27	0.983	0.004	0.001	9.70E-09	9.70E-09	*ILDR1*[a]	0	*CD86, SLC15A2*
10	rs10901863	C	0.73	0.934	-0.004	0.001	2.60E-08	2.65E-08	*CTBP2*	0	-
8	rs7823971	C	0.80	0.991	-0.0043	0.001	2.70E-08	2.68E-08	*RP11-1102P16.1*	0	-

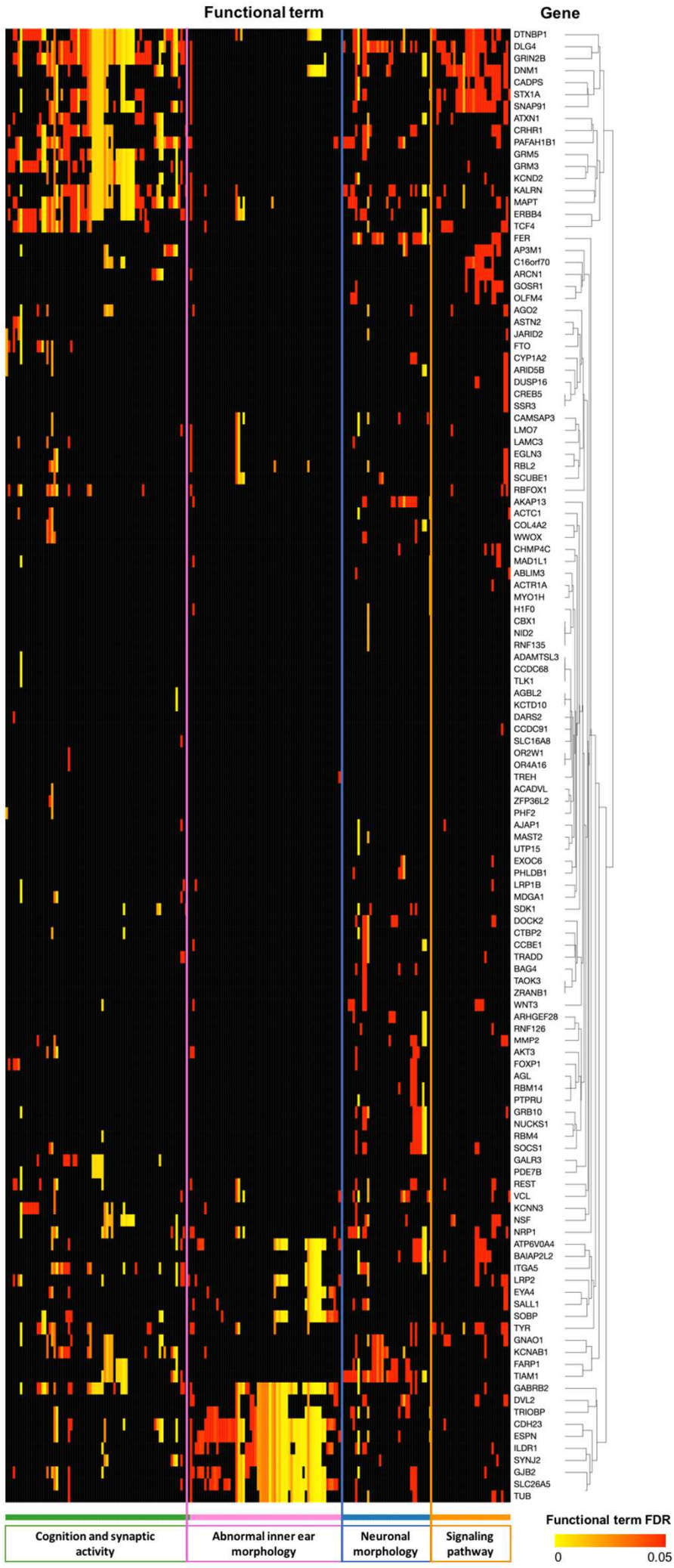
Table 2. Study, publication of previous finding; Gene, gene highlighted in the referenced publication as the lead SNP is either located in the gene region or in close proximity; SNP, single nucleotide polymorphism; CHR, Chromosome; BP, base position; A1, effect allele in analysis; A0, reference allele; INFO, quality metric, combination of imputation score and dosage confidence; UKBB phenotype, phenotype used in this study; A1FREQ, frequency of effect allele in analysis sample; BETA, effect size from BOLT-LMM approximation to infinitesimal mixed model; SE, standard error of the effect size; p-value, infinitesimal mixed model association test p-value. *This study did not analyse SNP rs58389158, but analysed rs5756795 which is in complete LD with this SNP in the British population, and referenced in the previous study.

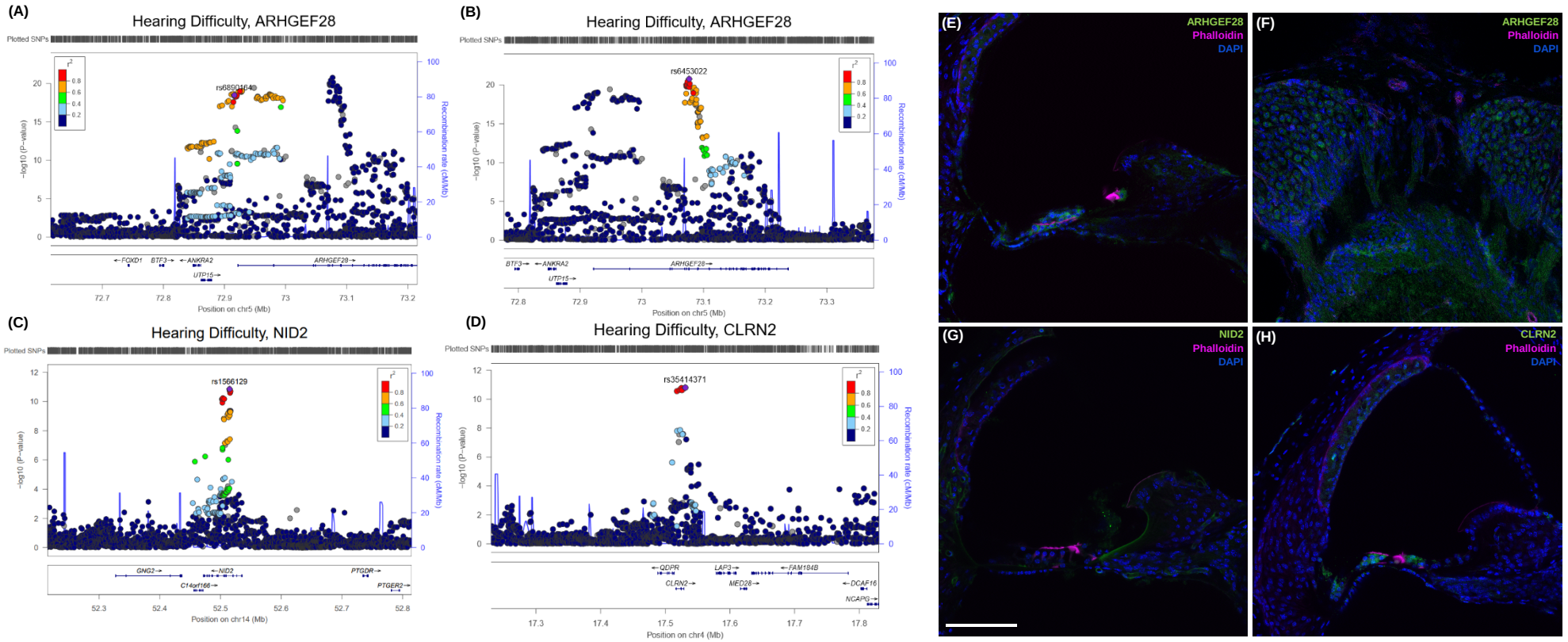
Table 2. Summary statistics from *HDiff* and *HAid* GWAS analysis, at SNPs highlighted in previous adult hearing loss GWAS.												
Variant highlighted in previous study				Summary statistics from *HDiff* and *HAid* analysis in the UKBB cohort								
Citation	Gene	SNP	CHR	BP	A1	A0	INFO	UKBB Phenotype	A1FREQ	BETA	SE	P
Friedman et. al 2009[14]	*GRM7*	rs11928865	3	7155702	T	A	0.989	*HDiff*	0.741	0.0016	0.0015	0.28
								HAid	0.742	-0.0014	0.0007	0.05
Van Laer et al., 2010[12]	*IQGAP2*	rs457717	5	75920972	A	G	0.986	*HDiff*	0.326	0.0013	0.0014	0.34
								HAid	0.325	-0.0006	0.0007	0.37
	GRM7	rs161927	3	7838242	G	A	0.988	*HDiff*	0.134	0.0038	0.0019	0.05
								HAid	0.136	-0.0002	0.0009	0.86
Girotto et al., 2011[17]	*DCLK1*	rs248626	5	141097725	A	G	1.000	*HDiff*	0.251	0.0018	0.0015	0.23
								HAid	0.252	-0.0003	0.0007	0.71
	KCNMB2	rs4603971	3	177902467	G	A	0.992	*HDiff*	0.934	-0.0015	0.0027	0.58
								HAid	0.934	0.0006	0.0012	0.63
	CMIP	rs898967	16	81566780	C	T	0.981	*HDiff*	0.476	0.0010	0.0013	0.45
								HAid	0.476	0.0002	0.0006	0.76
	GRM8	rs2687481	7	125869122	G	T	0.998	*HDiff*	0.811	-0.0018	0.0017	0.28
								HAid	0.810	0.0012	0.0008	0.14
Nolan et al., 2013[19]	*ESSRG*	rs2818964	1	216682448	G	A	0.978	*HDiff*	0.366	-0.0015	0.0014	0.27
								HAid	0.366	0.0004	0.0006	0.55
Wolber et al., 2014[15]	*SIK3*	rs681524	11	116748314	T	C	0.992	*HDiff*	0.927	-0.0010	0.0026	0.71
								HAid	0.928	0.0018	0.0012	0.13
Vuckovic et al., 2015[16]	*PCDH20*	rs78043697	13	62467039	T	C	0.995	*HDiff*	0.928	0.0000	0.0025	1.00
								HAid	0.928	0.0010	0.0012	0.38
	SLC28A3	rs7032430	9	86714002	C	A	0.959	*HDiff*	0.782	-0.0013	0.0016	0.43
								HAid	0.783	-0.0001	0.0008	0.91

Fransen et al., 2015[10]	*ACVR1B*	rs2252518	12	52381026	C	A	0.996	*HDiff*	0.739	-0.0010	0.0015	0.50
								HAid	0.739	0.0001	0.0007	0.85
	CCBE1	rs34175168	18	57180682	G	A	0.990	*HDiff*	0.986	0.0112	0.0056	0.04
								HAid	0.986	-0.0009	0.0026	0.74
Hoffman et al., 2016[18]	*ISG20*	rs4932196	15	89253268	T	C	1.000	*HDiff*	0.809	0.0085	0.0017	4.60E-07
								HAid	0.809	0.0039	0.0008	6.40E-07
	TRIOBP	rs5756795*	22	38122122	T	C	1	*HDiff*	0.539	-0.0092	0.0013	5.10E-12
								HAid	0.538	-0.0027	0.0006	1.60E-05


Data & QC
Discovery analysis
Annotation and validation
GWAS of hearing impairment in the UKBB Cohort
Phenotypes used in GWAS analysis:
1. Hearing difficulty, HDiff
2. Hearing aid use, HAid
Phenotype generation specified in Figure S1
Genotype QC
Omitted samples with high missingness/heterozygosity, sex discrepancies, removal of individuals with excess relatedness
White British samples selected
Discovery analysis
9,740,198 SNPs, MAF ≥ 0.01, INFO > 0.7
Analysis adjusted for age, sex, UKBB genotyping platform, UKBB PCs 1-10
Hearing difficulty
N = 250,389
Genome-wide significance p< 5x10-8
Conditional analysis: 41 independent loci
Hearing aid use
N = 253,918
Genome wide significance p< 5x10-8
Conditional analysis: 7 independent loci
GWAS replication
Replication meta-analysis conducted in the remaining sample of Caucasians in the UKBB cohort (white non-British), the English Longitudinal Study of Aging and TwinsUK
In silico functional analysis
SNP-based analysis (VEP), Gene set enrichment analysis (ToppGene Suite), Gene-based analysis (MAGMA)
In vitro functional analysis
Protein localization in adult mouse cochlear with immunofluorescence


(A)
ARHGEF28
EYA4
CTBP2
NID2
-log10 P-value
Chromosome
(B)
ARHGEF28
EYA4
NID2
CTBP2
-log10 P-value
Chromosome


Functional term
Gene
DTNBP1
DLG4
GRIN2B
DNM1
CADPS
STX1A
SNAP91
ATXN1
CRHR1
PAFAH1B1
GRM5
GRM3
KCND2
KALRN
MAPT
ERBB4
TCF4
FER
AP3M1
C16orf70
ARCN1
GOSR1
OLFM4
AGO2
ASTN2
JARID2
FTO
CYP1A2
ARID5B
DUSP16
CREB5
SSR3
CAMSAP3
LMO7
LAMC3
EGLN3
RBL2
SCUBE1
RBFOX1
AKAP13
ACTC1
COL4A2
WWOX
CHMP4C
MAD1L1
ABLIM3
ACTR1A
MYO1H
H1F0
CBX1
NID2
RNF135
ADAMTSL3
CCDC68
TLK1
AGBL2
KCTD10
DARS2
CCDC91
SLC16A8
OR2W1
OR4A16
TREH
ACADVL
ZFP36L2
PHF2
AJAP1
MAST2
UTP15
EXOC6
PHLDB1
LRP1B
MDGA1
SDK1
DOCK2
CTBP2
CCBE1
TRADD
BAG4
TAOK3
ZRANB1
WNT3
ARHGEF28
RNF126
MMP2
AKT3
FOXP1
AGL
RBM14
PTPRU
GRB10
NUCKS1
RBM4
SOCS1
GALR3
PDE7B
REST
VCL
KCNN3
NSF
NRP1
ATP6V0A4
BAIAP2L2
ITGA5
LRP2
EYA4
SALL1
SOBP
TYR
GNAO1
KCNAB1
FARP1
TIAM1
GABRB2
DVL2
TRIOBP
CDH23
ESPN
ILDR1
SYNJ2
GJB2
SLC26A5
TUB
Cognition and synaptic activity
Abnormal inner ear morphology
Neuronal morphology
Signaling pathway
Functional term FDR
0
0.05


(A)
Hearing Difficulty, ARHGEF28
Plotted SNPs
rs68094164
−log10 (P−value)
Recombination rate (cM/Mb)
FOXD1
BTF3
ANKRA2
ARHGEF28
UTP15
Position on chr5 (Mb)
(B)
Hearing Difficulty, ARHGEF28
Plotted SNPs
rs6453022
−log10 (P−value)
Recombination rate (cM/Mb)
BTF3
ANKRA2
ARHGEF28
UTP15
Position on chr5 (Mb)
(C)
Hearing Difficulty, NID2
Plotted SNPs
rs1566129
−log10 (P−value)
Recombination rate (cM/Mb)
GNG2
NID2
C14orf166
PTGDR
PTGER2
Position on chr14 (Mb)
(D)
Hearing Difficulty, CLRN2
Plotted SNPs
rs35414371
−log10 (P−value)
Recombination rate (cM/Mb)
QDPR
LAP3
FAM184B
CLRN2
MED28
DCAF16
NCAPG
Position on chr4 (Mb)
(E)
ARHGEF28
Phalloidin
DAPI
(F)
ARHGEF28
Phalloidin
DAPI
(G)
NID2
Phalloidin
DAPI
(H)
CLRN2
Phalloidin
DAPI

Supplemental Data

Supplemental Data include six figures and three tables.

Figure S1a. Flow chart describing case-control assignment for the hearing difficulty (*HDiff*) phenotype.

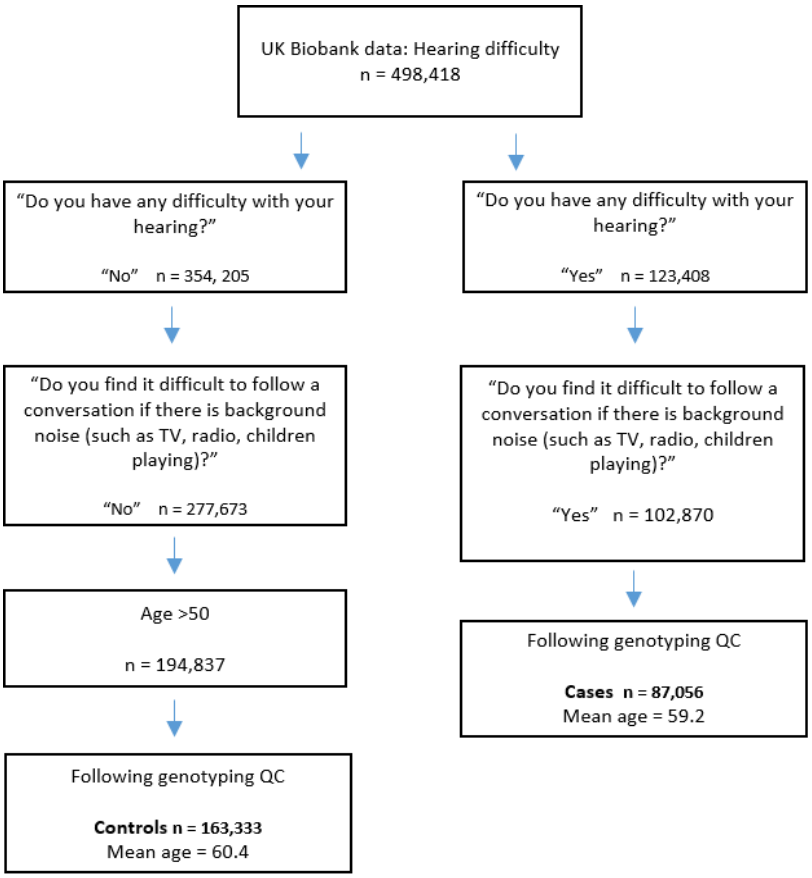


Figure S1a. Flow chart describing case-control assignment for the hearing difficulty (*HDiff*) phenotype. Participants answered questions as part of the UKBB questionnaire administered at UKBB assessment centres. Participants who answered 'Prefer not to answer', 'I am completely deaf' and 'Do not know' were removed from the analysis. Participants were removed from the control group if they answered "Yes" to "Do you use a hearing aid most of the time?" A lower age limit of 50 was implemented for controls to ensure age was consistent between the case and control groups due to the association of aging with the trait.

Figure S1b. Flow chart describing case-control assignment for the hearing aid use (*HAid*) phenotype

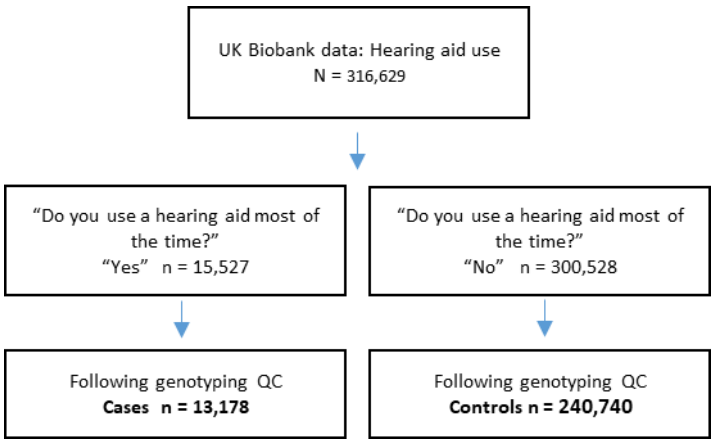


Figure S1b. Flow chart describing case-control assignment for the hearing aid use (*HAid*) phenotype. Participants answered questions as part of the UKBB questionnaire administered at the UKBB assessment centres. No information was collected regarding age at hearing aid prescription or cause of hearing loss.

Figure S2a. Q-Q plot of GWAS summary statistics, *HDiff* (left) and *HAid* (right).

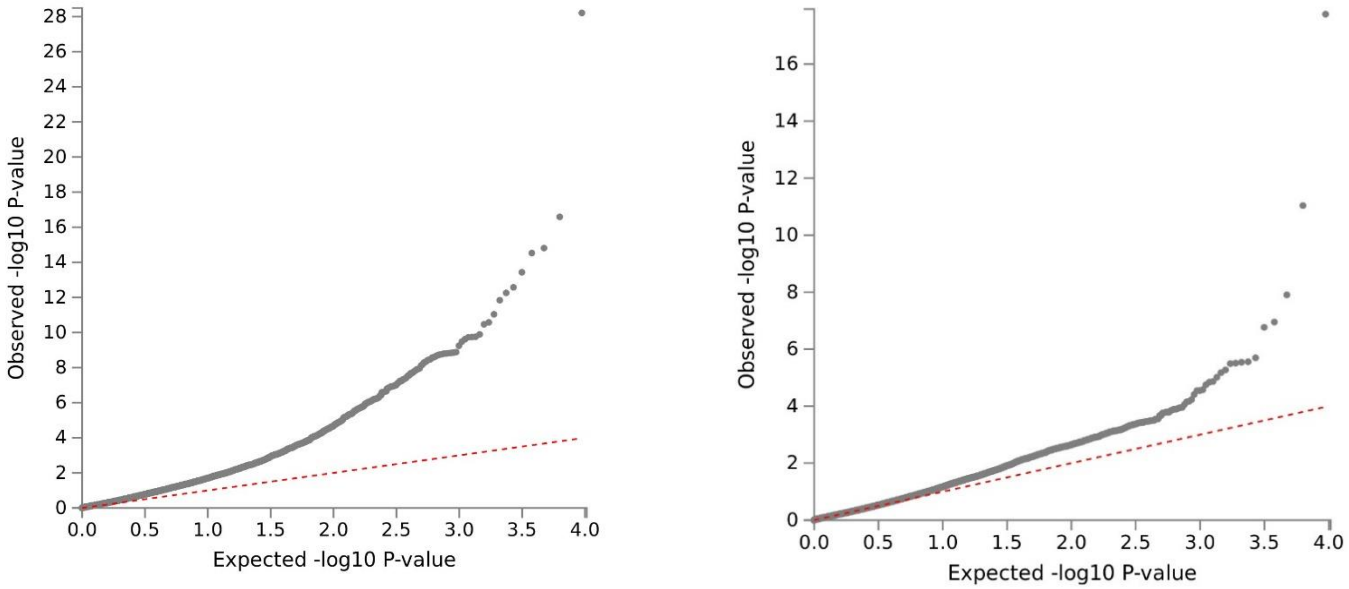


Figure S2a. Q-Q plot of GWAS summary statistics, *HDiff* (left) and *HAid* (right). The LD score regression intercepts for the two analyses were 1.032 for *HDiff* and 1.03 for *HAid*. The ratio (intercept-1)/(mean(χ^2)-1) for *HDiff* was 8% and represents the proportion of inflation in the χ^2 statistic that the intercept attributes to alternative explanations than polygenicity. The ratio for *HAid* was 5%.

Figure S2b. Q-Q plot of the gene-based test computed by MAGMA, *HDiff* (left) and *HAid* (right).

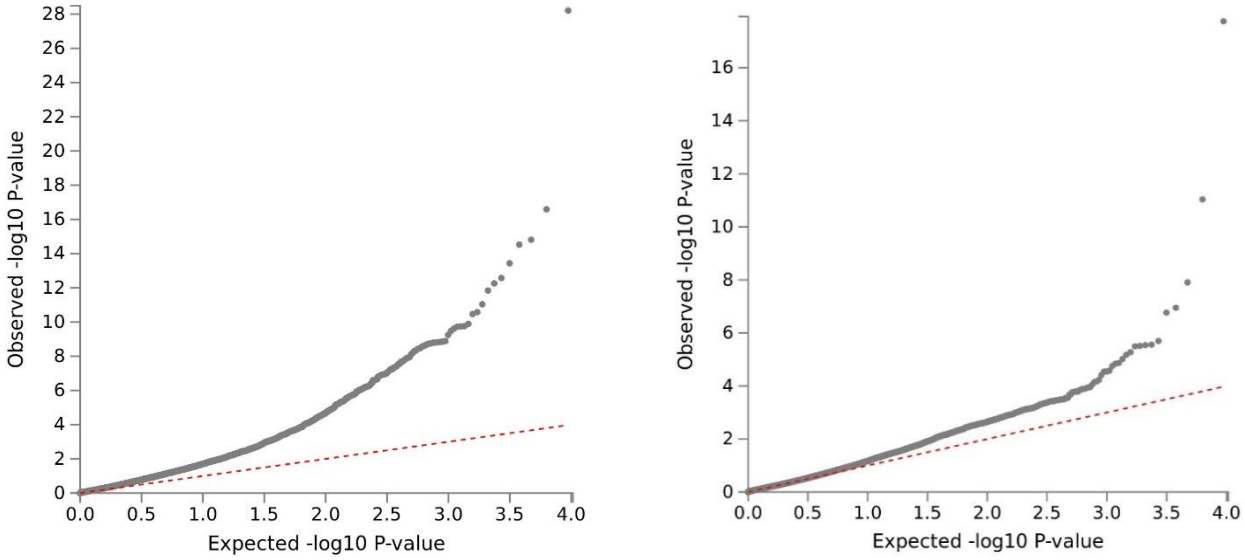


Figure S2b. Q-Q plot of the gene-based test computed by MAGMA, *HDiff* (left) and *HAid* (right).

Figure S3. Manhattan-Style plots of regional heritability across the genome for *HDiff* (a) and *HAid* (b)

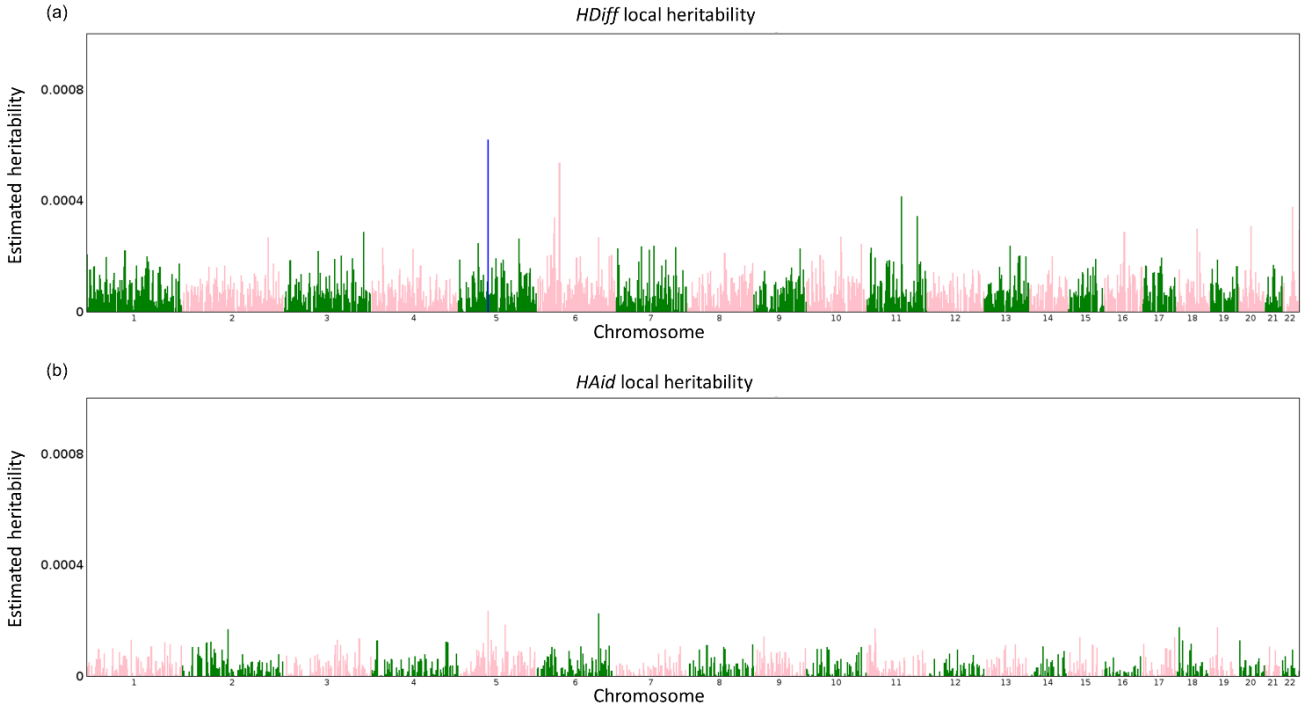


Figure S3. Manhattan-Style plots of Regional heritability across the genome for *HDiff* (a) and *HAid* (b). Estimated local SNP heritability for 1702 loci. Blue denotes an estimate of local_h2g = 0.0006, p<1.1579E-05on Chr5 for *HDiff* in the region between base positions 71240456- 73759326, here represented by 4404 SNPs. HESS total SNP heritability estimates were 0.1, SE=0.005 for *HDiff* and 0.011, SE=0.005 for *HAid*.

Figure S4. Locus plots displaying regions significantly associated in both *HDiff* and *HAid* analysis.

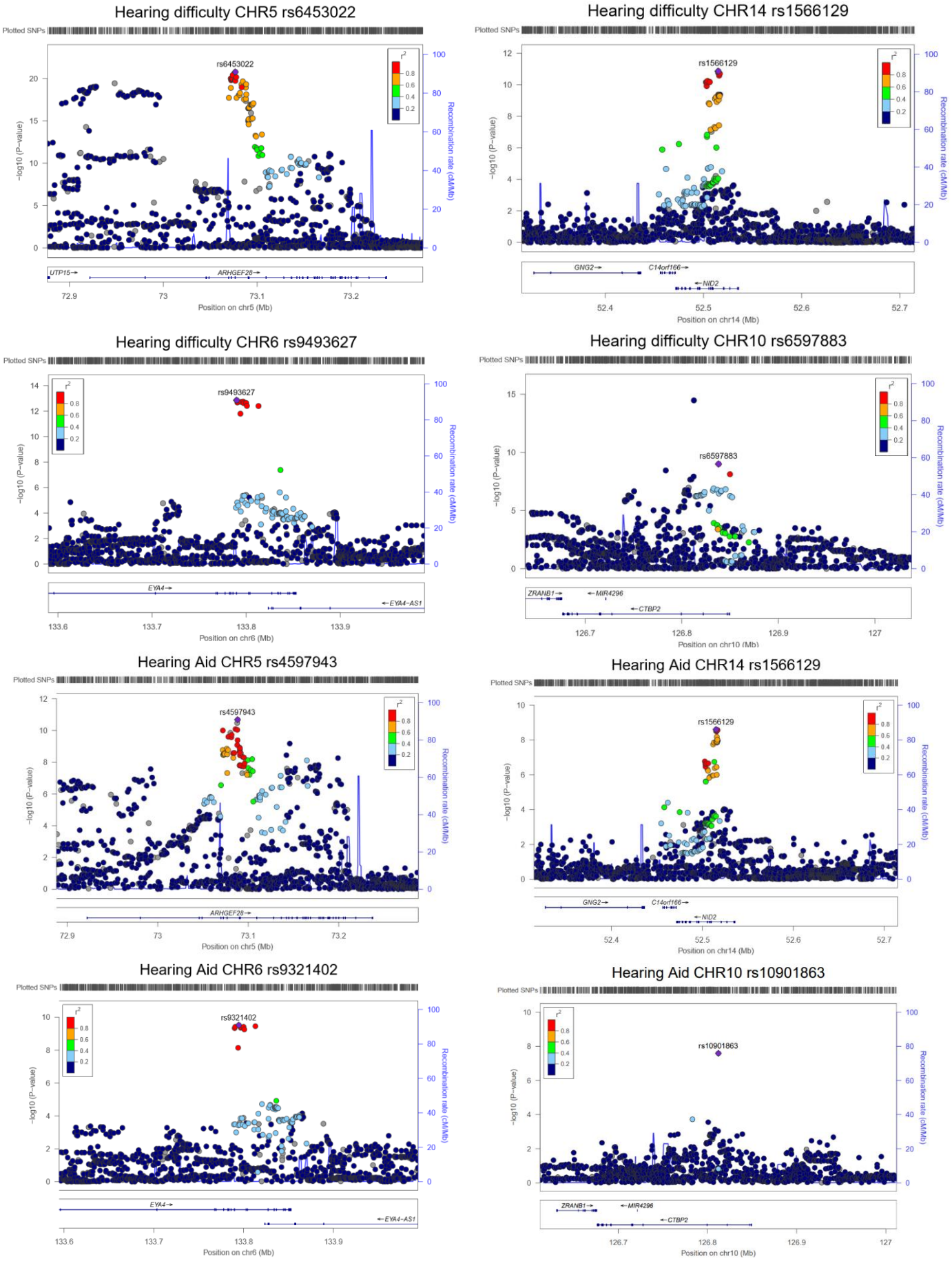


Figure S4. Locus plots displaying regions significantly associated in both *HDiff* and *HAid* analysis. Locus plots generated with *HDiff* summary statistics. Purple indicates lead independent SNP generated from GCTA-COJO conditional analysis. The colouring of remaining SNPs represents the correlation (r^2) to the lead SNP (purple).

Figure S5. MAGMA gene-property analysis for tissue specificity from average expression of 30 general tissue types from GTEx v6.

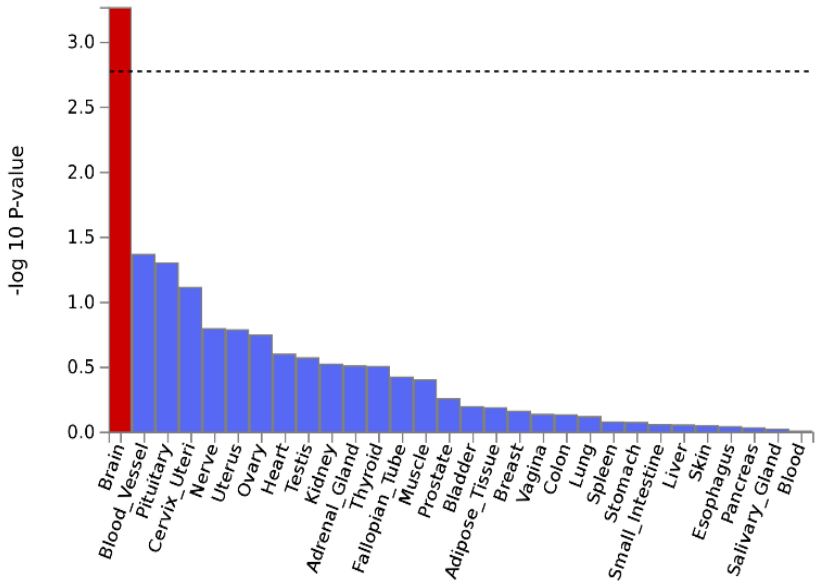


Figure S5. MAGMA gene-property analysis for tissue specificity from average expression of 30 general tissue types from GTEx v6. Relationships between tissue specific gene expression profiles (x-axis) and genetic association of genes (y-axis) is shown. Genetic associations of genes were performed in MAGMA and represent the aggregated effect of all SNPs in a gene. The dotted line indicates the Bonferroni-corrected level of 2.6E-06.

Figure S6. Immunofluorescence images of adult mouse cochlea Vibratome sections stained with proteins of interest (green), DAPI (blue) and Phalloidin (magenta)

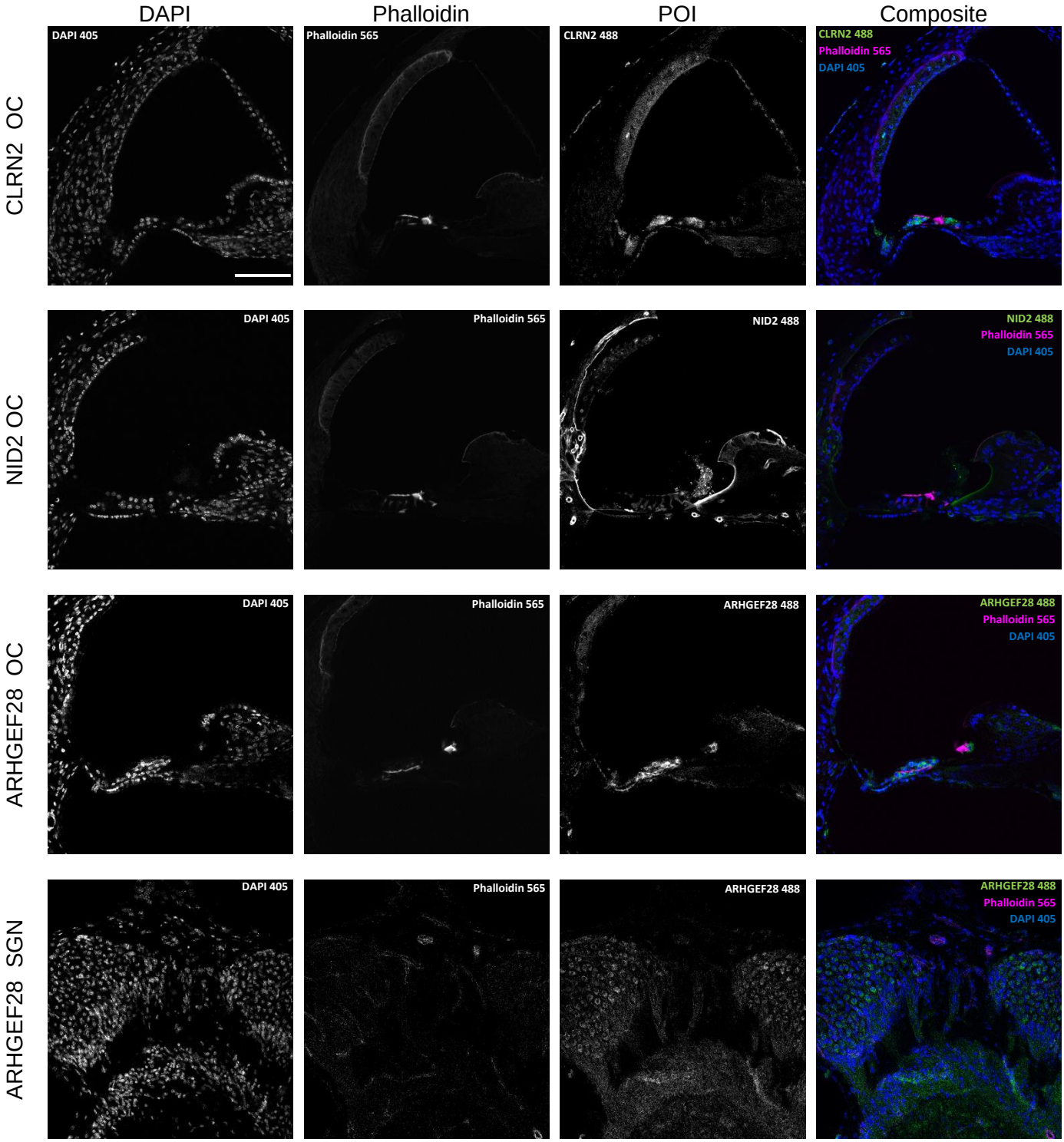


Figure S6. Immunofluorescence images of adult mouse cochlea Vibratome sections stained with proteins of interest (green), DAPI (blue) and Phalloidin (magenta)

Individual channels presented in greyscale, and combined for composite colour images as in Figure 4. Three panels display the organ of Corti (OC) regions (CLRN2 OC, NID2 OC, and ARHGEF28 OC) and one displays the Spiral Ganglion Neurons (SGN) region (ARHGEF28 SGN).

The scale bar in the top left image displays 100µm. This scale is consistent for all images in this figure.

Table S1. Summary statistics for *HDiff* phenotype from the replication meta-analysis of the white non-British UKBB sample, TwinsUK and ELSA

Marker Name	Allele1	Allele2	Weight	Zscore	P-value	Direction	Replication power p<0.05	Replication power p<0.0012
rs759016271	a	agtagtccacttttcttctttgcctg	29866	4.556	5.20E-06	+++	0.900	0.504
rs1566129	t	c	30894	3.534	0.00041	+++	0.661	0.195
rs36062310	a	g	30868	2.974	0.002938	+--	0.928	0.575
rs143282422	a	g	30274	2.652	0.007996	+-+	0.492	0.098
rs12225399	c	g	29802	2.624	0.008688	+++	0.610	0.160
rs6597883	t	c	30274	2.585	0.009748	++-	0.564	0.133
rs62033400	a	g	30851	2.513	0.01198	+++	0.600	0.153
rs7951935	t	g	29802	2.48	0.01312	++-	0.820	0.360
rs141403654	a	t	29802	-2.392	0.01674	---	0.477	0.091
rs35186928	a	g	29866	2.348	0.01885	+-+	0.785	0.314
rs17671352	t	c	30852	2.228	0.02587	++-	0.520	0.110
rs217289	a	g	29866	2.227	0.02597	+-+	0.524	0.112
rs62188635	t	c	30377	-2.189	0.02859	---	0.583	0.144
rs76837345	a	g	30318	-2.158	0.0309	---	0.499	0.101
rs55635402	a	g	29802	2.132	0.03301	+++	0.585	0.145
rs10824108	t	g	30274	1.982	0.04752	+-+	0.541	0.120
rs2236401	t	c	29866	1.873	0.06102	++-	0.561	0.131
rs5756795	t	c	30868	-1.696	0.08981	--+	0.679	0.209
rs7525101	t	c	30389	1.636	0.1018	+++	0.506	0.104
rs10475169	a	c	30356	-1.603	0.1088	-+-	0.516	0.108
rs4948502	t	c	30274	1.574	0.1155	+-+	0.553	0.127
rs9691831	a	g	30717	-1.512	0.1305	-+-	0.492	0.098
rs12938775	a	g	30852	-1.481	0.1385	-+-	0.509	0.105
rs4611552	t	c	30753	-1.434	0.1515	---	0.489	0.096
rs13093972	a	g	30098	-1.413	0.1577	--+	0.525	0.113
rs6453022	a	c	30356	1.327	0.1846	+--	0.912	0.532
rs835267	a	g	30274	1.17	0.242	++-	0.555	0.128
rs35414371	a	t	30856	1.163	0.245	+-+	0.657	0.192
rs9493627	a	g	29866	1.018	0.3086	+++	0.724	0.248
rs12027345	a	g	30389	-0.986	0.3244	--+	0.537	0.119

rs3890736	a	g	30318	-0.706	0.4805	+--	0.495	0.099
rs4947828	t	g	30717	-0.562	0.574	+--	0.571	0.137
rs13277721	a	g	30318	0.527	0.5983	+--	0.590	0.147
rs12552	a	g	30676	0.489	0.6249	+-+	0.480	0.093
rs6890164	a	g	30356	0.298	0.7658	+--	0.877	0.455
3:182069497_TA_T	t	ta	30098	0.298	0.766	++-	0.629	0.172
rs132929	a	g	30868	0.234	0.815	++-	0.731	0.256
rs10927035	t	c	30389	0.219	0.8268	+-+	0.476	0.091
rs34442808	t	ta	30356	0.218	0.8277	-++	0.560	0.131
rs9366417	a	g	29866	0.137	0.8911	+-+	0.490	0.097
rs62015206	t	c	30459	-0.071	0.9437	--+	0.522	0.111

Table S1. Summary statistics for *HDiff* phenotype from the replication meta-analysis of the white non-British UKBB sample, TwinsUK and ELSA. Marker Name, SNP ID; Allele1, the first allele for this marker in the first file where it occurs; Allele 2, the second allele for this marker in the first file where it occurs; Weight, the sum of the individual study weights (N) for this marker; Z-score, the combined z-statistic for the marker; P-value, meta-analysis p-value; Direction, direction of effect for each study ordered: white non-British UKBB sample, TwinsUK, ELSA; Replication power at $p<0.05$, estimated power to identify a replicated association at nominal significance; Replication power $p<0.0012$, estimated power to detect an association at significance threshold 0.05/41 = 0.0021, $p<0.0012$.

Table S2. Summary statistics for the *HAid* phenotype in the replication meta-analysis of white non-British UKBB sample, TwinsUK and ELSA

Marker Name	Allele1	Allele2	Weight	Zscore	P-value	Direction	Replication power p<0.05	Replication power p<0.00714
rs4597943	t	g	34475	2.833	0.004608	+++	0.695	0.413
rs7823971	a	c	34919	-1.886	0.05934	---	0.541	0.265
rs3915060	t	c	34359	-1.664	0.09605	--+	0.560	0.281
rs1566129	t	c	35139	0.784	0.4333	++-	0.602	0.318
rs10901863	t	c	32251	0.638	0.5234	+--	0.510	0.240
rs9321402	a	g	35101	0.58	0.5622	++-	0.649	0.364
rs9677089	a	c	34727	-0.537	0.5915	--+	0.653	0.368

Table S2. Summary statistics for the *HAid* phenotype in the replication meta-analysis of white non-British UKBB sample, TwinsUK and ELSA.
Marker Name, SNP ID; Allele1, the first allele for this marker in the first file where it occurs; Allele 2, the second allele for this marker in the first file where it occurs; Weight, the sum of the individual study weights (N) for this marker; Z-score, the combined z-statistic for the marker; P-value, meta-analysis p-value; Direction, direction of effect for each study ordered: white non-British UKBB sample, TwinsUK, ELSA; Replication power p<0.05, calculated power with replication sample parameters to achieve an association at nominal significance; Replication power p<0.00714, calculated power with replication sample parameters to achieve an association at the significance threshold of (0.05/7 = 0.00714) p<0.00714.

Table S3. Genetic correlation results with *HDiff* phenotype.

Trait	rg	se	p	Group
Wheeze or whistling in the chest in last year	0.3137	0.0306	1.35E-24	Breathing difficulty
Shortness of breath walking on level ground	0.3228	0.0429	5.60E-14	Breathing difficulty
Bring up phlegm/sputum/mucus on most days	0.342	0.0679	4.74E-07	Breathing difficulty
Long-standing illness_ disability or infirmity	0.3763	0.0305	5.48E-35	Health report / subjective wellbeing
Overall health rating	0.3156	0.0266	2.09E-32	Health report / subjective wellbeing
Health satisfaction	0.3405	0.0374	8.47E-20	Health report / subjective wellbeing
Other serious medical condition/disability diagnosed by doctor	0.3203	0.0409	4.80E-15	Health report / subjective wellbeing
Subjective well being	-0.3257	0.0421	1.06E-14	Health report / subjective wellbeing
Other eye problems	0.4311	0.0687	3.59E-10	Health report / subjective wellbeing
Had major operations	0.3196	0.06	9.84E-08	Health report / subjective wellbeing
Illnesses of siblings: None of the above (group 2)	-0.3436	0.0702	9.94E-07	Health report / subjective wellbeing
Former alcohol drinker	0.3136	0.0701	7.56E-06	Health report / subjective wellbeing
Tinnitus: Yes_ now most or all of the time	0.6	0.0562	1.40E-26	Hearing
Loud music exposure frequency	0.3224	0.0583	3.20E-08	Hearing
Frequency of tiredness / lethargy in last 2 weeks	0.4089	0.029	2.79E-45	Low mood /depression
Neuroticism score	0.315	0.0257	1.94E-34	Low mood /depression
Miserableness	0.3283	0.0273	2.69E-33	Low mood /depression
Seen doctor (GP) for nerves_ anxiety_ tension or depression	0.3447	0.0298	5.90E-31	Low mood /depression
Frequency of depressed mood in last 2 weeks	0.3361	0.0318	4.30E-26	Low mood /depression
Guilty feelings	0.3258	0.0309	4.98E-26	Low mood /depression
Loneliness_ isolation	0.3233	0.0333	2.99E-22	Low mood /depression
Frequency of unenthusiasm / disinterest in last 2 weeks	0.3239	0.0348	1.35E-20	Low mood /depression
Frequency of tenseness / restlessness in last 2 weeks	0.3046	0.0336	1.26E-19	Low mood /depression
Ever depressed for a whole week	0.371	0.041	1.45E-19	Low mood /depression
Illness_ injury_ bereavement_ stress in last 2 years: Financial difficulties	0.3098	0.0368	3.70E-17	Low mood /depression
Depressive symptoms	0.3314	0.0409	5.69E-16	Low mood /depression

Ever unenthusiastic/disinterested for a whole week	0.3445	0.0432	1.53E-15	Low mood /depression
Happiness	0.3148	0.0408	1.27E-14	Low mood /depression
Financial situation satisfaction	0.3437	0.0473	3.87E-13	Low mood /depression
Family relationship satisfaction	0.3118	0.0433	6.12E-13	Low mood /depression
Ever highly irritable/argumentative for 2 days	0.3192	0.0457	2.78E-12	Low mood /depression
Insomnia	0.3211	0.0484	3.31E-11	Low mood /depression
Illness_ injury_ bereavement_ stress in last 2 years: Serious illness_ injury or assault to yourself	0.3161	0.0485	7.36E-11	Low mood /depression
Neuroticism	0.3125	0.0724	1.60E-05	Low mood /depression
Chest pain or discomfort	0.3823	0.0341	3.60E-29	Pain
Pain type(s) experienced in last month: None of the above	-0.3225	0.0295	9.62E-28	Pain
Pain type(s) experienced in last month: Neck or shoulder pain	0.3686	0.0367	9.34E-24	Pain
Pain type(s) experienced in last month: Stomach or abdominal pain	0.3409	0.0418	3.73E-16	Pain
Pain type(s) experienced in last month: Hip pain	0.3079	0.0394	5.25E-15	Pain
Mouth/teeth dental problems: Toothache	0.428	0.0699	9.02E-10	Pain
Mouth/teeth dental problems: Painful gums	0.4514	0.0749	1.71E-09	Pain

Table S3. **Genetic correlation results with *HDiff* phenotype**. This table lists traits that had significant correlations with the *HDiff* phenotype and an rg < -0.3 or rg > 0.3. The traits are ordered by group, and by p-value order within each group.