Investigating the Characteristics of Genes and Variants Associated with Self-Reported Hearing Loss in Older Adults in the UK Biobank

Morag A. Lewis[1], Bradley A. Schulte[2], Judy R. Dubno[2], Karen P. Steel[1].

[1]Wolfson Centre for Age-Related Diseases, King`s College London, SE1 1UL, UK

[2]The Medical University of South Carolina, SC, USA

Abstract

Age-related hearing loss is a common, heterogeneous disease with a strong genetic component, but to date most of the genes underlying human adult-onset hearing loss remain unknown. Most genetic studies have focussed on very rare variants (such as family studies and patient cohort screens) or very common variants (genome-wide association studies). Here we present a study based on exome sequencing and self-reported hearing loss in the UK Biobank, a large-scale biomedical database. We have carried out variant load analyses and compared the resulting gene lists to a manually-curated list of nearly 700 genes known to be involved in hearing in humans and/or mice. Our results suggest that relatively common as well as rare variants with a high predicted impact contribute to age-related hearing loss, and that the genetic contributions to age-related hearing loss may differ between the sexes. Our manually-curated list of deafness genes is a useful resource for candidate gene prioritisation in hearing loss.

Introduction

Hearing loss is one of the most common sensory deficits in the human population and has a strong genetic component. However, the auditory system is a complex system with many interacting parts, which offers many routes to loss of function. Accordingly, although over 150 genes have been identified as contributing to non-syndromic human hearing loss [1], the majority of genes involved in hearing remain unknown. Moreover, most of the genes identified so far are those where mutations result in early-onset, severe hearing loss. While age-related hearing loss (ARHL) is very common, it is also very heterogeneous, and the associated landscape of genetic variation remains unclear, both at the gene and at the variant level. Even when analysing rare variants in known deafness genes, a wide mutational spectrum can be observed, with a range of allele frequencies and predicted impacts which differ on a gene-by-gene basis [2].

As early as 1997 it was noted that single gene mutations can lead to early postnatal or adult-onset progressive hearing loss [3]. This remains the case 25 years later; 45 out of the 51 known human autosomal dominant deafness genes result in progressive hearing loss when mutated [1]. These mutations are rare, high-impact variants which have been identified through family studies and candidate gene screening of patient cohorts, for example [4-7]. However, most such variants are ultra-rare or even private [5], and while they fully explain the hearing loss seen in the affected individual or family, they can't explain all the ARHL seen in the population. On the other end of the scale, looking at common variants, very large genome wide association studies (GWAS) have recently uncovered several new loci [8, 9], but because GWAS work by identifying markers linked to disease loci, they cannot detect recent mutations or those which are not widespread throughout the population. A recent GWAS on hearing loss, which reports both common and rare variant association analyses, found that the rare variant association signals were mostly independent of the common variant associations nearby, confirming that it is important to include consideration of rare variants in the genetic landscape of ARHL [10].

Alternative approaches are therefore required to identify novel variants and genes associated with age-related hearing loss. Here we have investigated variants associated with self-reported hearing loss in the 200,000 UK Biobank participants with available exome sequence data. We have assessed both rare variants (minor allele frequency (MAF) < 0.005) and variants with an intermediate frequency ("intermediate") (MAF < 0.1, which includes the rare variants) and compared the resulting gene lists to a much larger list of known deafness genes that we have curated and present here, based on work in mice as well as in humans. Our results suggest that both rare and intermediate variants with a high predicted impact contribute to age-related hearing loss, and also that the genetic contributions to age-related hearing loss may differ between men and women.

Methods

UK Biobank participant selection. UK Biobank is a large-scale biomedical database and research resource containing genetic, lifestyle and health information from half a million UK participants, aged between 40-69 years in 2006-2010, who were recruited from across the UK. Participants have consented to provide their data to approved researchers who are undertaking health-related research that is in the public interest. We aimed to select participants ≥55 years of age who had exome sequencing data available (200,619 exomes available in total, September 2020) and could be classified as having normal hearing or hearing loss, based on their self-report of hearing difficulty, hearing difficulty in noise, or use of a hearing aid. If people reported no hearing difficulties or hearing aid use at any assessment and had been asked about their hearing at least once when they were ≥55, we included them in the "normal hearing" group. If people reported consistent or worsening hearing loss, or that they had at any point been a hearing aid user and had been asked at least once about their hearing when they were ≥55, we included them in the "hearing loss" group. We completely excluded participants who reported otologic disorders (eg Meniere's disease). We excluded people who reported high levels of noise exposure or moderate/severe tinnitus from the normal hearing group (Supplementary Figure 1). This resulted in a total of 48731 people in the normal hearing group (18235 men and 30496 women), and 45581 people in the hearing loss group (24237 men and 21344 women). Overall, 106307 participants with exomes were excluded based on the above criteria. We did not filter by self-reported ethnicity. The vast majority (96%) of participants described themselves as "British", "Irish", "White", or "any other White background", or some combination thereof (hereafter referred to as White). In most of the broad ethnic groupings (Supplementary Table 1), more women than men volunteered (Supplementary Figure 2A). We compared participants included in this study to the entire UK Biobank, to the UK Biobank participants who had had exome sequencing, and to the data from the UK census 2011 [11], and found that while the proportion of self-reported minority ethnicities was smaller in the UK Biobank than in the 2011 census [11], it was smaller still in the participants included in this study (Supplementary Figure 2B). However, the distribution of self-reported ethnicities in the participants with exome sequencing reflected that of the entire Biobank (Supplementary Figure 2B). The "healthy volunteer" effect, meaning that participants tend to be healthier in terms of lifestyle and health conditions, has been previously noted in the UK Biobank when compared to the UK 2011 census data, as has the greater proportion of people reporting their ethnicity as White [11]. It is not clear why the subset of the UK Biobank selected for this study, on the basis of their answers to questions about hearing and related issues, has an even greater proportion of participants who report their ethnicity as White.

Variant annotation and filtering. UK Biobank variant calls were made available following processing, variant calling and joint genotyping [12, 13], but without any filters applied at the sample or variant level. We annotated the variants using the Ensembl Variant Effect Predictor [14], including data from

ReMM, which provides a measure of pathogenicity for regulatory variants [15], SpliceAI, which scores variants based on their predicted effect on splicing [16], Sutr, which provides annotations for 5'UTR variants, including a predicted effect on translation efficiency [17], and the deleteriousness predictor CADD [18]. Minor allele frequencies were obtained from gnomAD (African, admixed American, Ashkenazi Jewish, East Asian, Finnish, Non-Finnish European, Other) [19], the 1000 Genomes project (African, admixed American, East Asian, European, South Asian) [20], TopMed (not divided by population) [21] or ESP6500 (African, European) [22], and the maximum reported minor allele frequency (MAF) was used. Variants were then filtered based on the overall quality of the variant call (QUAL, minimum 20) and the read depth (DP, minimum 10) and genotype quality (GQ, minimum 10) of individual calls. Variants with more than 10% of calls missing were also excluded, as were those which were not in Hardy-Weinberg equilibrium and those which had a high private allele frequency within the UKBB cohort (defined as the recorded minor allele frequency + 0.4) [23].

Variant classification filters. We selected variants based on their minor allele frequency and a combination of pathogenicity and consequence filters (Table 1). We defined rare variants as those with MAF < 0.005 (0.5%), and intermediate variants as those with MAF < 0.1 (10%), and we did not look at variants with MAF ≥ 0.1. We defined two levels of impact upon a gene product, low impact and high impact. High impact variants were those in coding regions, intronic splice sites or mature miRNAs with a CADD score > 25 or a SpliceAI score > 0.5, and those in 5'UTRs with a Sutr score > 1. Low impact variants were all those in coding regions, intronic splice sites, mature miRNAs and 5'UTR regions, those in 3'UTR regions, and any variants with other classifications (eg regulatory region variants) which had a ReMM score > 0.95 (see Supplementary Table 2 for the exact variant classification terms and filters). These are inclusive lists, so the list of intermediate variants includes the rare variants, and the list of low impact variants includes the high impact variants.

Regression outlier analysis. For each analysis, we summed the total number of variants in each gene in participants from the normal hearing group and compared them to the total number of variants in that gene from the hearing loss group using a linear regression. We did the comparisons across both men and women, and also did them separately. Outlier genes with a high variant load were identified as those with residuals greater than 6x the inter-quartile distance from the upper and lower quartiles [24], and hypergeometric distribution tests were carried out using R. gProfiler [25] was used to carry out a GO enrichment analysis of the outlier gene lists.

Compilation of the list of deafness genes. We compiled a manually curated list of known deafness genes in humans and mice, including all genes listed in the Hereditary Hearing Loss Homepage [1], and genes which, when mutated, result in altered hearing thresholds in mutant mice, reported by the International Mouse Phenotyping Consortium (www.mousephenotype.org [26, 27]; average thresholds were individually checked for shifts >10dB with small standard deviations). We also included mouse and human deafness genes described in the literature (for example [28, 29]; for full reference list see Supplementary Table 3). There were 118 genes shown to underlie hearing in mice and humans, 67 human deafness genes (with 66 mouse orthologues) and 506 mouse deafness genes (with 535 human orthologues) (Figure 1, Supplementary Table 3).

Compilation of the list of highly variable genes. Some genes are often reported as having a high number of variants in multiple exome sequencing projects. This can be because they are large genes (eg *TTN*), or because they belong to groups of paralogues such as olfactory receptors, which are sufficiently similar to make correct alignment difficult, resulting in incorrect variant calls. Two such lists were compiled by Adams et al [30], and Fuentes Fajardo et al, [31], and consist of genes which contributed many variant calls to multiple exomes as well as human leukocyte antigen (HLA), taste receptor (TAS), olfactory receptor and mucin family genes. Additionally, some genes have been

identified as prone to recurrent false positive calls, which are variants that did not validate with further genotyping and were not heritable [32]. We combined all three lists, resulting in 1213 genes in total (Supplementary Table 4).

Gene expression analysis using the gEAR. To assess the expression of lists of genes of interest in the inner ear, including the list of known mouse and human deafness genes, we used single cell RNAseq data from the mouse inner ear, accessed via the gEAR portal (https://umgear.org/) [33]. We chose datasets from mice aged between embryonic day (E) 16 and postnatal day (P) 35. The datasets we used came from E16 cochlea, P1 cochlea, P7 cochlea [34], P15 cochlea [35], P20 inner ear [36], P28-35 cochlea [37], P30 stria vascularis [38] and P17-33 spiral ganglion neurons [39]. Expression levels were normalised to *Hprt* expression; where *Hprt* was not present in the dataset, or had an expression level of 0, we did not use the data. We then summarised the data, taking the maximum level per cell type without accounting for age. Because the expression levels ranged from 0 to 279.2 (*Ocm*, in outer hair cells), we transformed the data such that levels between 100 and 1000 were scaled to between 3 and 4, levels between 10 and 100 were scaled to between 2 and 3, and levels between 1 and 10 were scaled to between 1 and 2. We used R to plot a heatmap of the genes that showed the most variability between cell types, suggestive of specific expression patterns rather than non-specific expression (n=720, variance across datasets > 0.15), and to cluster cell types and genes. We further defined gene clusters first based on the R dendrograms and then on the gene expression levels within specific cell types or groups of cell types.

Results

After filtering, in the normal hearing group there were 18235 men (average age=62.37 years) and 30496 women (average age=62.20 years) (48731 people in total, overall average age 62.29). In the hearing loss group, there were 24237 men (average age=63.60 years) and 21344 women (average age=62.96 years) (45581 people in total, overall average age 63.28). It is notable that while the overall group sizes are similar, there are many more women than men in the normal hearing group than in the hearing loss group, reflecting the better hearing that women have later in life [40], although the average age of the participants (62-63 years) is later than the average onset of menopause, after which women's hearing tends to decline rapidly [41]. When we plotted the distribution of each broad ethnic grouping within each category (Supplementary Table 1 and Supplementary Figure 2C,D), we found that there were many more Black men and women in the normal hearing group than in the hearing loss group (especially Black women) (Supplementary Figure 2E). Similar results have been noted in previous studies [42]. The distribution of other self-reported ethnicities were broadly similar across the sex-separated groups, but it is notable that the largest difference in self-reported hearing phenotype between the sexes is in the White ethnic grouping (Supplementary Figure 2C, D, E).

Outlier analysis of rare, high impact variants

We first asked which genes had a high load of rare variants with a high predicted impact (defined in Table 1) in people in the hearing loss group (n=45581) compared to those in the normal hearing group (n=48731). We found 14 genes with a high variant load in the hearing loss group and 13 genes with a high variant load in the normal hearing group (Figure 2, Supplementary Table 5A).

To assess the potential relevance of these high variant load gene lists to hearing loss, we looked for genes associated with deafness in humans and/or mice by comparing genes in the two outlier lists with our own manually-curated list of known deafness genes (using the human orthologues of deafness genes known only in mice where possible, resulting in 720 genes in total) (Supplementary

Table 3). Although many of these known deafness genes have only been linked to early-onset, severe hearing impairment, they are still good candidates for involvement in ARHL, since different variants can result in very different phenotypes. For example, different variants in *TMC1* have been shown to result in either prelingual profound hearing loss or postlingual progressive hearing loss [43, 44], and several recent large-scale studies looking at adult-onset hearing loss have found multiple missense variants in Mendelian deafness genes with milder effects than previously reported [5, 10, 45]. We found two such genes in the high variant load in normal hearing list (*DMD*, *OTOG*), and one, *MYH14*, in the high variant load in hearing loss list. We tested the significance of these overlaps using a hypergeometric distribution, and found that there was not a significant enrichment of known deafness genes in either high variant load list (Table 2).

We then carried out the same analyses on participants separated by sex. The numbers of outlier genes in each case were similar (Table 2), but the gene lists differed almost completely between men and women. Only one gene, *TNRC6A*, was present in high variant load for men and women in hearing loss lists (Supplementary Table 5). There were no significant overlaps with the known deafness gene list (Table 2).

Outlier analysis of intermediate, low impact variants

Because not many outlier genes were identified from the rare, high impact variant analysis, we repeated our analyses on the intermediate, low-impact variants (defined in Table 1), but while we found more genes overall, both with a high variant load in hearing loss (81) and with a high variant load in normal hearing (167) (Supplementary Table 5B), we did not find any known deafness genes in the high variant load in hearing loss list. We repeated the sex-separated analyses, and carried out the significance tests, and again found no significant enrichment of known deafness genes (Table 2, Supplementary Table 5B).

Outlier analyses of rare, low impact variants

We had anticipated that relaxing the restrictions on allele frequency and impact would result in detection of more outlier genes associated with deafness, not fewer, so we then tried varying the combinations of allele frequency and impact filters. We first asked what the outliers were if we looked at rare variants with a low impact, and we found very few genes were identified, both in the combined and the sex-separated analyses. There was no significant overlap with known deafness genes (Table 2, Supplementary Table 5C).

Outlier analyses of intermediate, high impact variants

However, when we carried out the outlier analysis on intermediate variants with a high impact, we found 307 genes with a high variant load in normal hearing, including 21 known deafness genes, and 206 genes with a high variant load in hearing loss, including 18 deafness genes (Table 2, Supplementary Table 5D). In the sex-separated analyses, we found similarly high numbers, and our enrichment analyses found that the overlap of known deafness genes with the high variant load lists was significant in five of the six comparisons (combined sexes, men and women, with high variant loads in normal hearing and in hearing loss) (Table 2). Similar to our first analysis, the genes in the high variant load in hearing loss lists differed considerably between men and women, with only 40 genes in common (16% and 18% in men and women respectively), five of which (*GJB2*, *SLC26A5*, *MYH14*, *CLIC5* and *COL9A3*) are known deafness genes [46-50] (Figure 3). This suggests that the genetic contribution to hearing loss may differ between men and women.

Do highly variable genes contribute to the outlier lists?

It is possible that these outlier genes are simply highly variable genes, and do not reflect the self-reported hearing loss phenotype, and so to investigate this, we compared the high variant load lists to our list of highly variable genes (Supplementary Table 4), genes which are often reported to have a high number of variants in sequencing projects. Eleven of the 24 lists had more highly variable genes in than would be expected by chance (Table 2). Seven of these eleven lists were from the analyses of the low impact variants, suggesting that relaxing the restriction on variant impact is likely to result in detecting naturally variable genes as outliers rather than genes linked to the phenotype under study. Four of the high impact variant lists were enriched with highly variable genes; the outlier list from rare, high impact variants in men with hearing loss, and the outlier lists from intermediate, high impact variants in both sexes, men, and women with normal hearing (Table 2). The enrichment of these outlier lists with highly variable genes suggests that identifying them will be useful in de-prioritising such genes for further analysis.

Characteristics of the variants in the high variant load lists

The large numbers of outlier genes in people with normal hearing was unexpected, so we asked if there might be different types of variants common in hearing loss compared with normal hearing. We investigated the characteristics of the variants in the high variant load lists, taking the most deleterious consequence for each variant in each gene (defined in order in Supplementary Table 2). Variant counts were normalised per person and per gene. When looking at rare, high-impact variants, there were markedly more in-frame insertions in the outlier genes in hearing loss than in the outlier genes in normal hearing (Supplementary Figure 3A). The contributions of different kinds of rare, low-impact variants to the outlier genes varies more than for the other three groups (Supplementary Figure 3B), but it is likely due to the very low numbers of outlier genes in this analysis (Table 2). In the analyses of intermediate variants, the proportions of variant impacts contributing to each total is much more consistent (Supplementary Figure 3C,D). In all analyses, missense variants made up a large proportion of the total variant counts per person per gene.

Gene ontology enrichment analysis

In order to look for any clues to pathological mechanisms, we carried out a gene ontology (GO) enrichment analysis using gProfiler [25] on the high variant load lists from the high impact, intermediate variant analysis (Supplementary Table 5D). We restricted the output to GO terms with between 5 and 200 genes, since terms with more genes than that are overly general, and those with fewer genes are too specific. We found 35 GO terms enriched in the lists, including multiple terms specific to hearing (eg GO:0007605; sensory perception of sound) (Supplementary Table 6). The largest list of GO terms came from the genes with a high variant load in men with normal hearing (Supplementary Table 6), mostly because of a set of genes identified as being involved in stereocilium structure and function. There were 7 genes annotated with the term “stereocilium bundle”; *USH1C*, *USH2A*, *MYO3A*, *TMC2*, *TSPEAR*, *PDZD7* and *PKHD1L1*. Most are known human deafness genes, but *PKHD1L1* and *TMC2* have only been identified as mouse deafness genes to date [51, 52]. Far fewer specific GO terms were identified from the genes with a high variant load in women with hearing loss or with normal hearing (Supplementary Table 6), even though there was a similar number of known deafness genes in the lists (Table 2). The term “sensory perception of sound” was identified as enriched in the hearing loss outlier genes in women and in all participants (Supplementary Table 6). Genes annotated with this term which had a high variant load in women with hearing loss were *MYO3B*, *MYH14*, *CDH23*, *CLIC5*, *CHRNA10*, *FBXO11*, *TMC1*, *GJB2*, *NAV2*, *SLC26A5* and *MYO6*. Genes annotated with this term which had a high variant load in all participants with hearing loss were *MYO3B, COL11A1*, *MYH14*, *CDH23*, *CLIC5*, *TMC1*, *GJB2*, *LOXHD1*, *SLC26A5* and *MYO6*. Most of these are known human deafness genes, but to date, *FBXO11* has only been

identified in the mouse, not in humans, and *MYO3B* and *CHRNA10* are not in our compiled list of known deafness genes (they are included in the GO term annotation through orthologous similarity rather than published evidence). In summary, the GO term analysis showed enrichment for terms relating to sensory hair cells or cytoskeletal elements known to be important to hair cell function, and this appears to be driven by the enrichment for known deafness genes in the lists analysed. Many more genes were included in the high variant load lists than were described by GO annotation terms, reflecting the limitations in current GO annotations of many genes.

Using expression analysis to highlight new candidate genes

We carried out an analysis of the expression of candidate genes from our outlier analyses using single-cell RNAseq datasets from the gEAR database of mouse inner ear tissue analyses [33]. We reasoned that if a gene shows strong, specific expression in certain cell types in the inner ear, that suggests a potential functional role for the gene in those cell types, and would make it a good candidate for further investigation. We chose mouse datasets to cover as many inner ear cell types as possible between embryonic day (E) 16 and postnatal day (P) 35. We selected those candidate genes from the high impact, intermediate variant load lists (Supplementary Table 5D) which had a high quality one-to-one mouse orthologue (n=614), and we also plotted data for the genes known to underlie deafness in both mice and humans which were not already included (an additional 98 genes) as useful markers of cochlear cell types. After selecting those genes showing high variance in expression across the cell types, we had 259 genes from the high variant load lists and 79 more from the known deafness gene list. From the resulting heatmap, we annotated clusters of genes according to their expression, defining high expression levels (red) as 2-3, middle expression levels (orange) as 1-2 and low expression levels as 0-1 (yellow) (Supplementary Figure 4, Supplementary Table 7). We linked gene clusters to specific cell types if they showed high or middle expression specific to those cell types, and classified the clusters by known marker genes for specific cell types where these were present (Supplementary Figure 4, Supplementary Table 7). This allowed us to identify good candidate genes from those outlier genes which were not already known deafness genes. For example, there were three genes which appeared to be strongly and specifically expressed in pillar cells, *Col4a4* and *Col4a3*, which are known deafness genes [53], and *Thsd7a*, which is a gene with high variant loads in women and in both sexes combined with self-reported hearing loss (cluster 3D, Supplementary Figure 4). There were multiple clusters of genes strongly expressed in hair cells (clusters 1A, 1B, 1C, 1D, 1F, 1G, 1H, 1I, 2A, 2B, 2C, 2E, 2F, 2G, 3A, 4A, 4B, 4C, Supplementary Figure 4), and candidate genes from this list included *Strip2* and *Gab2*, both of which had high variant loads in men reporting hearing loss, and *Myom1* and *Chrna10*, both of which had high variant loads in women reporting hearing loss. We observed fewer genes strongly and specifically expressed in the spiral ganglion neurons (SGN) and strial cell types, possibly because there were fewer datasets available in the gEAR database for these cell types compared to hair cells, pillar cells and supporting cells, but clusters 2G and 3B did show consistent SGN expression (Supplementary Figure 4). Candidate genes from the SGN clusters include the mouse deafness genes *Ercc6* [54] and *Ccdc92* (IMPC data, www.mousephenotype.org [26, 27]), which have a high variant load in women with normal hearing and men with hearing loss respectively. We also found some genes which clustered in strial cell types (marginal, intermediate and basal cells, clusters 2A,2D, 4A, 4C, Supplementary Figure 4), including *Vegfb* and *Esrra*, both of which are genes with a high variant load in people with hearing loss and in women with hearing loss when analysed separately. All the genes plotted on the heatmap are listed with their classifications in Supplementary Table 7, and a summary of the clusters and their outlier genes is in Table 3. It is notable that outlier genes are found in all the clusters, suggesting that there is no one cell type or cochlear location wholly or largely responsible for age-related hearing loss in men or women (Figure 4, Table 3).

Discussion

Our data suggest that the genetic contributions to age-related hearing loss may differ between men and women. That is, there may be some genetic impacts which have less effect on women than on men, and vice-versa. The differences observed in the prevalence, severity and onset of ARHL in men and women have been widely reported (for example [40, 42, 55-58], reviewed in [59]). Sex differences in complex traits and disease phenotypes may be attributed to environmental factors (in this case, noise exposure would be relevant, and drug exposure, which can affect hearing in a sex-specific manner [60]), and comorbidities which display sex-related variance may also play a role, for example cardiovascular disease [61-63]. Endogenous factors are also likely to contribute, such as hormone differences, epigenetic and regulatory differences and, of course, the different genetics involved in the XX and XY genomes. There are many studies linking estrogen to hearing sensitivity [64-66], and several genes in the estrogen pathway have been linked to hearing loss [67-69]. Indeed, one of the novel candidates identified by our analysis is the estrogen-related receptor alpha, *Esrra*, which is an outlier gene in women with hearing loss, and in all participants with hearing loss (cluster 2A, Supplementary Figure 4, Supplementary Table 7). However, the average age of our participants was 62-63 years old, which is later than the average age of onset of menopause, so it is unlikely that hormones alone account for the observed effect. A sex-protective effect has been observed in other diseases, such that one sex requires a greater number of risk alleles to develop the disease. This was originally described by Carter et al [70, 71], who noted that women are less likely to suffer from pyloric stenosis but more likely to have children affected by the disease, but the phenomenon can apply to either sex. From our data, we found a similar number of genes bore a high load of variants in each sex, but the gene lists themselves were very different. Age-related hearing loss, while referred to as one condition, is actually the end result of a wide range of inner ear pathologies, so it is plausible that different sets of risk alleles contribute to overall hearing loss in different sexes.

We observed the highest number of known deafness genes in the outlier analysis of intermediate variants with a high predicted impact. There were 1,804,420 rare variants with high predicted impact, and 1,827,068 intermediate variants with high predicted impact, so altering the allele frequency filter from <0.005 to <0.1 only added 22,648 variants (1.2% of the total), and yet this changed the total number of outlier genes linked to hearing loss from 14 to 206 (both sexes) (Table 2), and the number of known deafness genes in that list increased from 1 to 18. This suggests that relatively common variants (MAF < 0.1) with a high predicted impact do contribute to ARHL, which correlates with the findings of another recent UK Biobank study [10], which reports that 16.8% of SNP heritability is contributed by "low-frequency variants" (0.001 < MAF ≤ 0.05). This is lower than our definition of intermediate variants (MAF < 0.1), but still higher than the standard cutoff of 0.001 recommended for autosomal dominant hearing loss [72]. Since ARHL is a complex disease rather than a Mendelian one, it is unsurprising that a different approach is needed when filtering for causative variants.

Previous studies have concentrated on human hearing loss genes [2, 5], but we have compiled a larger list of nearly 700 genes based on human and mouse studies, and from this we have identified multiple candidate genes among our outliers, including *FSCN2*, *SYNJ2*, *FBXO11*, *NAV2*, *TMC2*, *ERCC6*, *CCDC92* and *PKHD1L1*. Of the 185 known human deafness genes, 118 are also mouse deafness genes (Supplementary Table 3), suggesting that mouse deafness genes are indeed good candidate human deafness genes. This is supported by the report from Praveen et al [10], who identified rare variant gene burdens in the mouse deafness genes *KLHDC7B*, *FSCN2* and *SYNJ2*, the latter two of which were also identified in our analyses (Table 3, Supplementary Table 7).

We took two approaches to explore the outlier gene lists, GO analysis and expression analysis. The GO analyses largely reiterated the comparisons with the deafness gene list (Supplementary Table 6). The lack of GO annotations linking the genes bearing a high variant load in hearing loss in the sex-separated analyses suggests that more pathways underlying hearing loss remain to be discovered and annotated. Our approach using the gEAR expression data is not limited by annotation, but is restricted to genes which have a high-quality one-to-one mouse orthologue, of which there were 614 (out of 1002 outlier genes in total). It is also subject to ascertainment bias due to the relative lack of data on inner ear cell types which are not hair cells or supporting cells. We obtained 6 datasets from hair cells and supporting cells at different stages from E16 to P35, but only 2 datasets from cochlear lateral wall cell types at 2 adult ages (P20 and P30), and only one dataset from SGNs (P17-33). This means that any gene expressed during development in the lateral wall or SGNs, but not expressed in adult stages, will have been missed out of our heatmap. Additionally, most of the known deafness genes which we plotted on the heatmap are hair cell or supporting cell genes, and this may have biased the clustering. This may be why there are more outlier genes assigned to clusters with expression in hair cells (Figure 4, most notably clusters 1A, 1G, 1H, 2C and 2F). Despite that, we did observe several clusters with expression in the lateral wall and spiral ganglion (Figure 4, Supplementary Figure 4). We have identified multiple potential candidate genes based on their presence in the outlier gene lists and their expression in specific cell types within the cochlea (Table 3), such as *THSD7A*, which is expressed in pillar cells, and *VEGFB*, a gene with expression in the stria and the organ of Corti, both of which have a high variant load in hearing loss (women and both sexes).

The nature of the regression analysis means that we detected outlier genes associated with normal hearing as well as hearing loss, and our enrichment analyses of the intermediate, high impact outlier gene lists showed almost all of them were significantly enriched in deafness genes (outliers with a high variant load in men with hearing loss was the only exception) (Table 2). This suggests that the high variant loads are driven by the association with the self-reported hearing phenotype, not just statistical noise, sequencing error and the natural genetic variability observed in some genes, particularly large genes like *TTN* and *USH2A*. This includes the high variant loads associated with normal hearing as well as those associated with hearing loss. It is possible that there may be protective variants in some of these genes, for example, variants which result in protection against noise trauma or ototoxic drug exposure, or which simply improve the maintenance of the inner ear machinery. Such a variant has recently been reported in *GJB6* in mice; homozygotes for the deleterious Ala88Val mutation displayed better hearing at older ages, better neural output from the inner ear, and reduced hair cell loss [73]. This is not the only precedent for deleterious mutations having a beneficial impact on a phenotype, and such mutations may be attractive targets for drug development. For example, Akbari et al (2021) recently reported that multiple rare protein-truncating variants in the gene *GPR75* were associated with protection from obesity, and mice lacking the orthologue, *Gpr75*, were resistant to weight gain on a high fat diet [74]. Similarly, rare deleterious variants in *B4GALT1* have been linked to decreased coronary artery disease via reduction of fibrinogen and low-density lipoprotein cholesterol [75]. Further investigation of the genes and variants linked to normal hearing and sex differences in hearing loss is needed.

Previously-reported GWAS of the UK Biobank which also used the self-reported hearing phenotype identified multiple overlapping loci; 71 in total, 19 of which were shared between all three studies (Figure 5) [8, 9, 45]. Four of those 19 were also identified in our study (*CHMP4C*, *NID2*, *SYNJ2* and *CDH23*). Five further genes from our outlier lists were shared between a subset of the GWAS lists; *TMPRSS9*, *LOXHD1*, and *TUB* were identified by Ivarsdottir *et al* and Kalra *et al*, and *SLC26A5* and *FSCN2* by Ivarsdottir *et al* alone. Those genes identified by multiple studies are obvious candidates

for involvement in ARHL, but the differences in loci identified using the GWAS approach suggest there are many more to investigate, and the results of our exome sequencing analysis support that as well as suggesting further candidates (Table 3). It has recently been observed that rare variants do not account for the GWAS hits of common markers [10], so it is unsurprising to find different variants and different genes associated with ARHL in GWAS compared with exome/genome sequence analysis studies.

The biggest limitation of this study is the lack of measured hearing loss (such as an audiogram) and detailed auditory phenotyping. Self-reported hearing loss has been shown to be sufficiently informative for general hearing capacity [9, 76], but there is more to hearing loss than just an average threshold shift. Being able to compare specific subtypes of age-related hearing loss (for example, using a classification system such as the one described in [77], [78], and [79]), offers the potential to link genes with a high variant load to specific inner ear pathologies, an important step for stratifying patient populations and developing therapeutics.

Acknowledgements

This research has been conducted using data from UK Biobank, a major biomedical database (www.ukbiobank.ac.uk, project ID 49593). The authors are grateful to everyone involved in the UK Biobank, especially the participants, without whom there would be no data. We thank Maria Lachgar-Ruiz for helpful comments on the manuscript text.

Data availability

All UK Biobank data, including the exome sequence data and questionnaire data used in this study, is publicly available to registered researchers through the UK Biobank data access protocol. Further information about registration may be found at http://www.ukbiobank.ac.uk/register-apply/. The mouse single cell RNAseq data used in this study is publicly available from the gEAR database at https://umgear.org.

Funding

This study was supported by the National Institutes of Health/National Institute on Deafness and Other Communication Disorders (NIH/NIDCD, grant number P50 DC 000422) and the National Institute for Health Research (NIHR) Biomedical Research Centre, King's College London.

References

1. Van Camp G, Smith RJH. **Hereditary Hearing Loss Homepage** [https://hereditaryhearingloss.org/] Accessed December 2021.
2. Azaiez H, Booth KT, Ephraim SS, Crone B, Black-Ziegelbein EA, Marini RJ, Shearer AE, Sloan-Heggen CM, Kolbe D, Casavant T *et al*: **Genomic Landscape and Mutational Signatures of Deafness-Associated Genes**. *Am J Hum Genet* 2018, **103**(4):484-497.
3. Van Camp G, Willems PJ, Smith RJ: **Nonsyndromic hearing impairment: unparalleled heterogeneity**. *Am J Hum Genet* 1997, **60**(4):758-764.
4. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T *et al*: **Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss**. *Nat Genet* 2009, **41**(5):609-613.
5. Boucher S, Tai FWJ, Delmaghani S, Lelli A, Singh-Estivalet A, Dupont T, Niasme-Grare M, Michel V, Wolff N, Bahloul A *et al*: **Ultrarare heterozygous pathogenic variants of genes causing dominant forms of early-onset deafness underlie severe presbycusis**. *Proc Natl Acad Sci U S A* 2020, **117**(49):31278-31289.

6. Walsh T, Walsh V, Vreugde S, Hertzano R, Shahin H, Haika S, Lee MK, Kanaan M, King MC, Avraham KB: **From flies' eyes to our ears: mutations in a human class III myosin cause progressive nonsyndromic hearing loss DFNB30**. *Proc Natl Acad Sci U S A* 2002, **99**(11):7518-7523.
7. Baek JI, Oh SK, Kim DB, Choi SY, Kim UK, Lee KY, Lee SH: **Targeted massive parallel sequencing: the effective detection of novel causative mutations associated with hearing loss in small families**. *Orphanet J Rare Dis* 2012, **7**:60.
8. Kalra G, Milon B, Casella AM, Herb BR, Humphries E, Song Y, Rose KP, Hertzano R, Ament SA: **Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty**. *PLoS Genet* 2020, **16**(9):e1009025.
9. Wells HRR, Freidin MB, Zainul Abidin FN, Payton A, Dawes P, Munro KJ, Morton CC, Moore DR, Dawson SJ, Williams FMK: **GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank**. *Am J Hum Genet* 2019, **105**(4):788-802.
10. Praveen K, Dobbyn L, Gurski L, Ayer AH, Staples J, Mishra S, Bai Y, Kaufman A, Moscati A, Benner C *et al*: **Population-scale analysis of common and rare genetic variation associated with hearing loss in adults**. *medRxiv* 2021:2021.2009.2027.21264091.
11. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, Collins R, Allen NE: **Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population**. *Am J Epidemiol* 2017, **186**(9):1026-1034.
12. Szustakowski JD, Balasubramanian S, Kvikstad E, Khalid S, Bronson PG, Sasson A, Wong E, Liu D, Wade Davis J, Haefliger C *et al*: **Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank**. *Nat Genet* 2021, **53**(7):942-948.
13. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, Gonzaga-Jauregui C, Khalid S, Ye B, Banerjee N *et al*: **Exome sequencing and characterization of 49,960 individuals in the UK Biobank**. *Nature* 2020, **586**(7831):749-756.
14. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F: **The Ensembl Variant Effect Predictor**. *Genome Biol* 2016, **17**(1):122.
15. Smedley D, Schubach M, Jacobsen JOB, Kohler S, Zemojtel T, Spielmann M, Jager M, Hochheiser H, Washington NL, McMurry JA *et al*: **A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease**. *Am J Hum Genet* 2016, **99**(3):595-606.
16. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB *et al*: **Predicting Splicing from Primary Sequence with Deep Learning**. *Cell* 2019, **176**(3):535-548 e524.
17. Pajusalu S. **5utr** [https://github.com/leklab/5utr] 2021.
18. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M: **CADD: predicting the deleteriousness of variants throughout the human genome**. *Nucleic Acids Res* 2019, **47**(D1):D886-D894.
19. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP *et al*: **The mutational constraint spectrum quantified from variation in 141,456 humans**. *Nature* 2020, **581**(7809):434-443.
20. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA *et al*: **A global reference for human genetic variation**. *Nature* 2015, **526**(7571):68-74.
21. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM *et al*: **Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program**. *Nature* 2021, **590**(7845):290-299.
22. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J *et al*: **Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants**. *Nature* 2013, **493**(7431):216-220.

23. Maffucci P, Bigio B, Rapaport F, Cobat A, Borghesi A, Lopez M, Patin E, Bolze A, Shang L, Bendavid M *et al*: **Blacklisting variants common in private cohorts but not in public databases optimizes human exome analysis**. *Proc Natl Acad Sci U S A* 2019, **116**(3):950-959.
24. Vuckovic D, Mezzavilla M, Cocca M, Morgan A, Brumat M, Catamo E, Concas MP, Biino G, Franze A, Ambrosetti U *et al*: **Whole-genome sequencing reveals new insights into age-related hearing loss: cumulative effects, pleiotropy and the role of selection**. *Eur J Hum Genet* 2018, **26**(8):1167-1179.
25. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J: **g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)**. *Nucleic Acids Res* 2019, **47**(W1):W191-W198.
26. Birling MC, Yoshiki A, Adams DJ, Ayabe S, Beaudet AL, Bottomley J, Bradley A, Brown SDM, Burger A, Bushell W *et al*: **A resource of targeted mutant mouse lines for 5,061 genes**. *Nat Genet* 2021, **53**(4):416-419.
27. Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H *et al*: **High-throughput discovery of novel developmental phenotypes**. *Nature* 2016, **537**(7621):508-514.
28. Ingham NJ, Pearson SA, Vancollie VE, Rook V, Lewis MA, Chen J, Buniello A, Martelletti E, Preite L, Lam CC *et al*: **Mouse screen reveals multiple new genes underlying mouse and human hearing loss**. *PLoS Biol* 2019, **17**(4):e3000194.
29. Ohlemiller KK, Jones SM, Johnson KR: **Application of Mouse Models to Research in Hearing and Balance**. *J Assoc Res Otolaryngol* 2016, **17**(6):493-523.
30. Adams DR, Sincan M, Fuentes Fajardo K, Mullikin JC, Pierson TM, Toro C, Boerkoel CF, Tifft CJ, Gahl WA, Markello TC: **Analysis of DNA sequence variants detected by high-throughput sequencing**. *Hum Mutat* 2012, **33**(4):599-608.
31. Fuentes Fajardo KV, Adams D, Program NCS, Mason CE, Sincan M, Tifft C, Toro C, Boerkoel CF, Gahl W, Markello T: **Detecting false-positive signals in exome sequencing**. *Hum Mutat* 2012, **33**(4):609-613.
32. Field MA, Burgio G, Chuah A, Al Shekaili J, Hassan B, Al Sukaiti N, Foote SJ, Cook MC, Andrews TD: **Recurrent miscalling of missense variation from short-read genome sequence data**. *BMC Genomics* 2019, **20**(Suppl 8):546.
33. Orvis J, Gottfried B, Kancherla J, Adkins RS, Song Y, Dror AA, Olley D, Rose K, Chrysostomou E, Kelly MC *et al*: **gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration**. *Nat Methods* 2021, **18**(8):843-844.
34. Kolla L, Kelly MC, Mann ZF, Anaya-Rocha A, Ellis K, Lemons A, Palermo AT, So KS, Mays JC, Orvis J *et al*: **Characterization of the development of the mouse cochlear epithelium at the single cell level**. *Nat Commun* 2020, **11**(1):2389.
35. Ranum PT, Goodwin AT, Yoshimura H, Kolbe DL, Walls WD, Koh JY, He DZZ, Smith RJH: **Insights into the Biology of Hearing and Deafness Revealed by Single-Cell RNA Sequencing**. *Cell Rep* 2019, **26**(11):3160-3171 e3163.
36. Xue N, Song L, Song Q, Santos-Sacchi J, Wu H, Navaratnam D: **Genes related to SNPs identified by Genome-wide association studies of age-related hearing loss show restriction to specific cell types in the adult mouse cochlea**. *Hear Res* 2021, **410**:108347.
37. Liu H, Chen L, Giffen KP, Stringham ST, Li Y, Judge PD, Beisel KW, He DZZ: **Cell-Specific Transcriptome Analysis Shows That Adult Pillar and Deiters' Cells Express Genes Encoding Machinery for Specializations of Cochlear Hair Cells**. *Front Mol Neurosci* 2018, **11**:356.
38. Korrapati S, Taukulis I, Olszewski R, Pyle M, Gu S, Singh R, Griffiths C, Martin D, Boger E, Morell RJ *et al*: **Single Cell and Single Nucleus RNA-Seq Reveal Cellular Heterogeneity and Homeostatic Regulatory Networks in Adult Mouse Stria Vascularis**. *Front Mol Neurosci* 2019, **12**:316.

39. Petitpre C, Wu H, Sharma A, Tokarska A, Fontanet P, Wang Y, Helmbacher F, Yackle K, Silberberg G, Hadjab S *et al*: **Neuronal heterogeneity and stereotyped connectivity in the auditory afferent system**. *Nat Commun* 2018, **9**(1):3691.
40. Cruickshanks KJ, Wiley TL, Tweed TS, Klein BE, Klein R, Mares-Perlman JA, Nondahl DM: **Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study**. *Am J Epidemiol* 1998, **148**(9):879-886.
41. Hederstierna C, Hultcrantz M, Collins A, Rosenhall U: **The menopause triggers hearing decline in healthy women**. *Hear Res* 2010, **259**(1-2):31-35.
42. Helzner EP, Cauley JA, Pratt SR, Wisniewski SR, Zmuda JM, Talbott EO, de Rekeneire N, Harris TB, Rubin SM, Simonsick EM *et al*: **Race and sex differences in age-related hearing loss: the Health, Aging and Body Composition Study**. *J Am Geriatr Soc* 2005, **53**(12):2119-2127.
43. Kitajiri S, Makishima T, Friedman TB, Griffith AJ: **A novel mutation at the DFNA36 hearing loss locus reveals a critical function and potential genotype-phenotype correlation for amino acid-572 of TMC1**. *Clin Genet* 2007, **71**(2):148-152.
44. Kurima K, Peters LM, Yang Y, Riazuddin S, Ahmed ZM, Naz S, Arnaud D, Drury S, Mo J, Makishima T *et al*: **Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function**. *Nat Genet* 2002, **30**(3):277-284.
45. Ivarsdottir EV, Holm H, Benonisdottir S, Olafsdottir T, Sveinbjornsson G, Thorleifsson G, Eggertsson HP, Halldorsson GH, Hjorleifsson KE, Melsted P *et al*: **The genetic architecture of age-related hearing impairment revealed by genome-wide association analysis**. *Commun Biol* 2021, **4**(1):706.
46. Donaudy F, Snoeckx R, Pfister M, Zenner HP, Blin N, Di Stazio M, Ferrara A, Lanzara C, Ficarella R, Declau F *et al*: **Nonmuscle myosin heavy-chain gene MYH14 is expressed in cochlea and mutated in patients affected by autosomal dominant hearing impairment (DFNA4)**. *Am J Hum Genet* 2004, **74**(4):770-776.
47. Faletra F, D'Adamo AP, Bruno I, Athanasakis E, Biskup S, Esposito L, Gasparini P: **Autosomal recessive Stickler syndrome due to a loss of function mutation in the COL9A3 gene**. *Am J Med Genet A* 2014, **164A**(1):42-47.
48. Kelsell DP, Dunlop J, Stevens HP, Lench NJ, Liang JN, Parry G, Mueller RF, Leigh IM: **Connexin 26 mutations in hereditary non-syndromic sensorineural deafness**. *Nature* 1997, **387**(6628):80-83.
49. Liu XZ, Ouyang XM, Xia XJ, Zheng J, Pandya A, Li F, Du LL, Welch KO, Petit C, Smith RJ *et al*: **Prestin, a cochlear motor protein, is defective in non-syndromic hearing loss**. *Hum Mol Genet* 2003, **12**(10):1155-1162.
50. Seco CZ, Oonk AM, Dominguez-Ruiz M, Draaisma JM, Gandia M, Oostrik J, Neveling K, Kunst HP, Hoefsloot LH, del Castillo I *et al*: **Progressive hearing loss and vestibular dysfunction caused by a homozygous nonsense mutation in CLIC5**. *Eur J Hum Genet* 2015, **23**(2):189-194.
51. Wu X, Ivanchenko MV, Al Jandal H, Cicconet M, Indzhykulian AA, Corey DP: **PKHD1L1 is a coat protein of hair-cell stereocilia and is required for normal hearing**. *Nat Commun* 2019, **10**(1):3801.
52. Kawashima Y, Geleoc GS, Kurima K, Labay V, Lelli A, Asai Y, Makishima T, Wu DK, Della Santina CC, Holt JR *et al*: **Mechanotransduction in mouse inner ear hair cells requires transmembrane channel-like genes**. *J Clin Invest* 2011, **121**(12):4796-4809.
53. Mochizuki T, Lemmink HH, Mariyama M, Antignac C, Gubler MC, Pirson Y, Verellen-Dumoulin C, Chan B, Schroder CH, Smeets HJ *et al*: **Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome**. *Nat Genet* 1994, **8**(1):77-81.
54. Nagtegaal AP, Rainey RN, van der Pluijm I, Brandt RM, van der Horst GT, Borst JG, Segil N: **Cockayne syndrome group B (Csb) and group a (Csa) deficiencies predispose to hearing loss and cochlear hair cell degeneration in mice**. *J Neurosci* 2015, **35**(10):4280-4286.

55. Dubno JR, Lee FS, Matthews LJ, Ahlstrom JB, Horwitz AR, Mills JH: **Longitudinal changes in speech recognition in older persons**. *J Acoust Soc Am* 2008, **123**(1):462-475.
56. Dubno JR, Lee FS, Matthews LJ, Mills JH: **Age-related and gender-related changes in monaural speech recognition**. *J Speech Lang Hear Res* 1997, **40**(2):444-452.
57. Lee FS, Matthews LJ, Dubno JR, Mills JH: **Longitudinal study of pure-tone thresholds in older persons**. *Ear Hear* 2005, **26**(1):1-11.
58. Pearson JD, Morrell CH, Gordon-Salant S, Brant LJ, Metter EJ, Klein LL, Fozard JL: **Gender differences in a longitudinal study of age-associated hearing loss**. *J Acoust Soc Am* 1995, **97**(2):1196-1205.
59. Nolan LS: **Age-related hearing loss: Why we need to think about sex as a biological variable**. *J Neurosci Res* 2020, **98**(9):1705-1720.
60. Mills JH, Matthews LJ, Lee FS, Dubno JR, Schulte BA, Weber PC: **Gender-specific effects of drugs on hearing levels of older persons**. *Ann N Y Acad Sci* 1999, **884**:381-388.
61. Gates GA, Cobb JL, D'Agostino RB, Wolf PA: **The relation of hearing in the elderly to the presence of cardiovascular disease and cardiovascular risk factors**. *Arch Otolaryngol Head Neck Surg* 1993, **119**(2):156-161.
62. Tan HE, Lan NSR, Knuiman MW, Divitini ML, Swanepoel DW, Hunter M, Brennan-Jones CG, Hung J, Eikelboom RH, Santa Maria PL: **Associations between cardiovascular disease and its risk factors with hearing loss-A cross-sectional analysis**. *Clin Otolaryngol* 2018, **43**(1):172-181.
63. Regitz-Zagrosek V, Kararigas G: **Mechanistic Pathways of Sex Differences in Cardiovascular Disease**. *Physiol Rev* 2017, **97**(1):1-37.
64. Al-Mana D, Ceranic B, Djahanbakhch O, Luxon LM: **Alteration in auditory function during the ovarian cycle**. *Hear Res* 2010, **268**(1-2):114-122.
65. Coleman JR, Campbell D, Cooper WA, Welsh MG, Moyer J: **Auditory brainstem responses after ovariectomy and estrogen replacement in rat**. *Hear Res* 1994, **80**(2):209-215.
66. Souza DDS, Luckwu B, Andrade WTL, Pessoa LSF, Nascimento JAD, Rosa M: **Variation in the Hearing Threshold in Women during the Menstrual Cycle**. *Int Arch Otorhinolaryngol* 2017, **21**(4):323-328.
67. Buniello A, Ingham NJ, Lewis MA, Huma AC, Martinez-Vega R, Varela-Nieto I, Vizcay-Barrena G, Fleck RA, Houston O, Bardhan T *et al*: **Wbp2 is required for normal glutamatergic synapses in the cochlea and is crucial for hearing**. *EMBO Mol Med* 2016, **8**(3):191-207.
68. Konig O, Ruttiger L, Muller M, Zimmermann U, Erdmann B, Kalbacher H, Gross M, Knipper M: **Estrogen and the inner ear: megalin knockout mice suffer progressive hearing loss**. *FASEB J* 2008, **22**(2):410-417.
69. Nolan LS, Maier H, Hermans-Borgmeyer I, Girotto G, Ecob R, Pirastu N, Cadge BA, Hubner C, Gasparini P, Strachan DP *et al*: **Estrogen-related receptor gamma and hearing function: evidence of a role in humans and mice**. *Neurobiol Aging* 2013, **34**(8):2077 e2071-2079.
70. Carter CO: **The inheritance of congenital pyloric stenosis**. *Br Med Bull* 1961, **17**:251-254.
71. Carter CO, Evans KA: **Inheritance of congenital pyloric stenosis**. *J Med Genet* 1969, **6**(3):233-254.
72. Oza AM, DiStefano MT, Hemphill SE, Cushman BJ, Grant AR, Siegert RK, Shen J, Chapin A, Boczek NJ, Schimmenti LA *et al*: **Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss**. *Hum Mutat* 2018, **39**(11):1593-1613.
73. Kelly JJ, Abitbol JM, Hulme S, Press ER, Laird DW, Allman BL: **The connexin 30 A88V mutant reduces cochlear gap junction expression and confers long-term protection against hearing loss**. *J Cell Sci* 2019, **132**(2).
74. Akbari P, Gilani A, Sosina O, Kosmicki JA, Khrimian L, Fang YY, Persaud T, Garcia V, Sun D, Li A *et al*: **Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity**. *Science* 2021, **373**(6550).

75. Montasser ME, Van Hout CV, Miloscio L, Howard AD, Rosenberg A, Callaway M, Shen B, Li N, Locke AE, Verweij N *et al*: **Genetic and functional evidence links a missense variant in B4GALT1 to lower LDL and fibrinogen**. *Science* 2021, **374**(6572):1221-1227.
76. Cherny SS, Livshits G, Wells HRR, Freidin MB, Malkin I, Dawson SJ, Williams FMK: **Self-reported hearing loss questions provide a good measure for genetic studies: a polygenic risk score analysis from UK Biobank**. *Eur J Hum Genet* 2020, **28**(8):1056-1065.
77. Dubno JR, Eckert MA, Lee FS, Matthews LJ, Schmiedt RA: **Classifying human audiometric phenotypes of age-related hearing loss from animal models**. *J Assoc Res Otolaryngol* 2013, **14**(5):687-701.
78. Vaden KI, Jr., Matthews LJ, Eckert MA, Dubno JR: **Longitudinal Changes in Audiometric Phenotypes of Age-Related Hearing Loss**. *J Assoc Res Otolaryngol* 2017, **18**(2):371-385.
79. Vaden KI, Eckert MA, Matthews LJ, Schmiedt RA, Dubno JR: **Metabolic and Sensory Components of Age-Related Hearing Loss**. *Journal of the Association for Research in Otolaryngology* 2022.

Figures

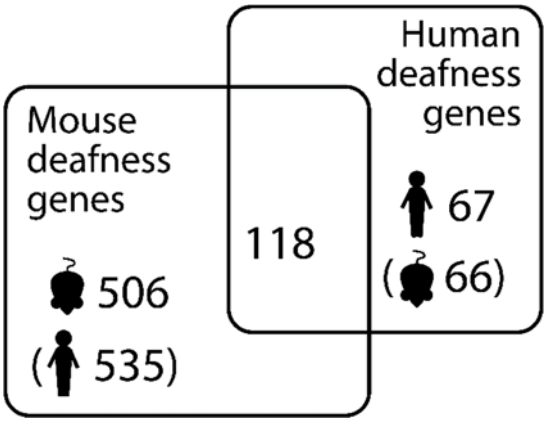

Figure 1. Deafness gene counts in mice and humans. Brackets indicate orthologues (eg there are 66 mouse orthologues of the 67 human deafness genes).

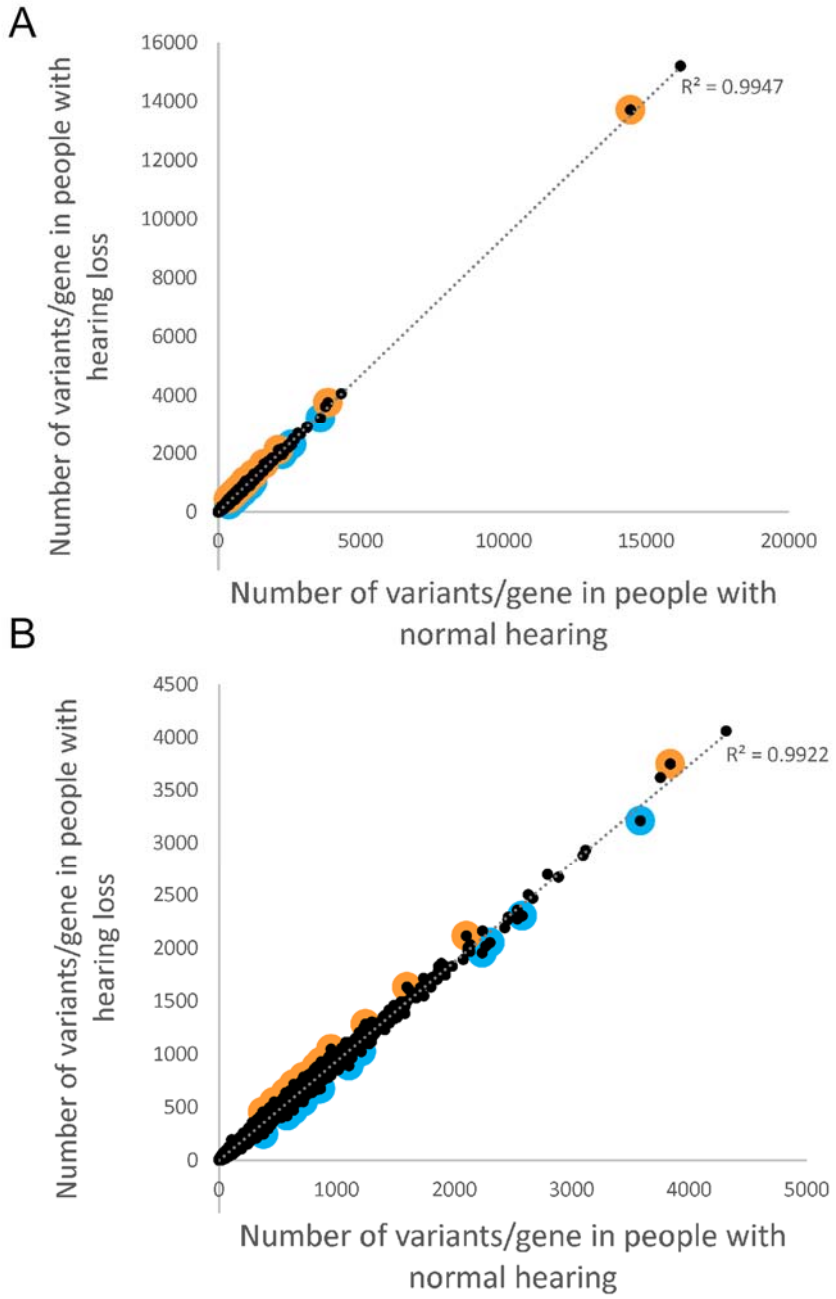

Figure 2. Comparison of variant load per gene for high impact rare variants. Each point represents a gene. Outliers are marked in orange (for higher load in participants with hearing loss) or blue (for higher load in participants with normal hearing). A shows all the data, including *TTN* and *TNRC6A*, genes with a much higher variant count than all the others, and B shows the data without those two genes.

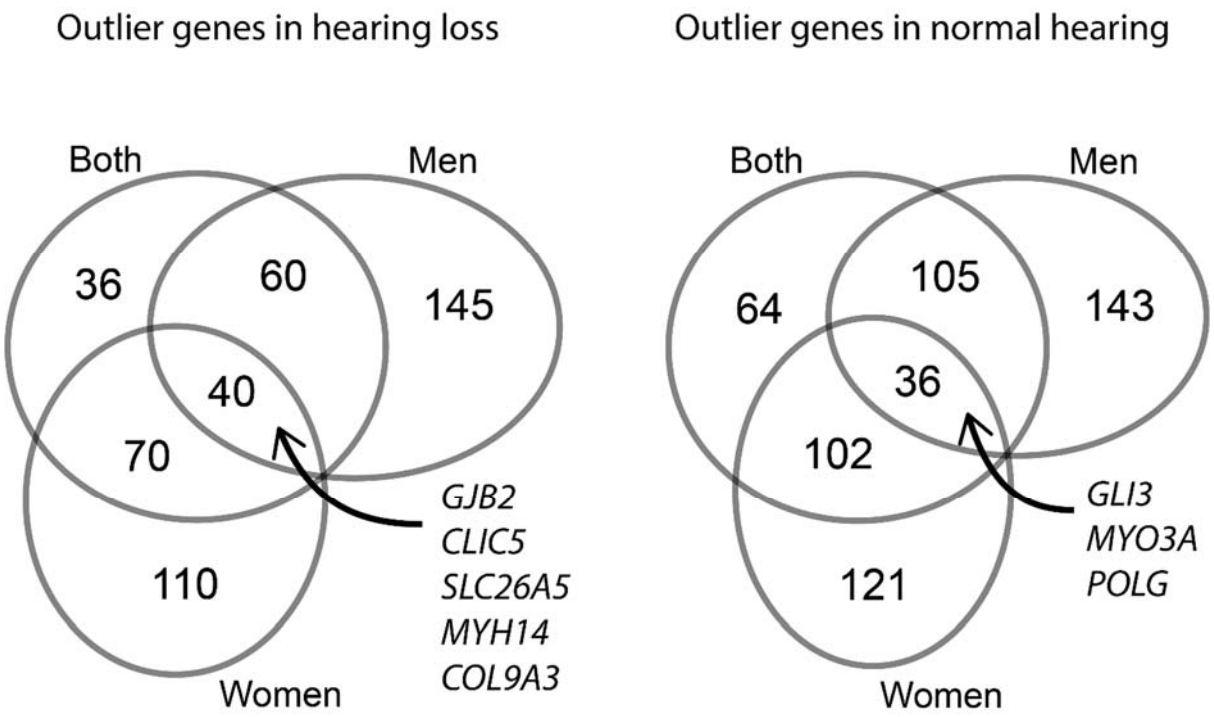

Figure 3. Venn diagrams showing the overlap of the outlier gene lists when looking at only men, only women, or all participants (outliers with intermediate variants with high impact). The known deafness genes in the intersection (5 in the outlier genes in hearing loss, 3 in the outlier genes in normal hearing) are labelled.

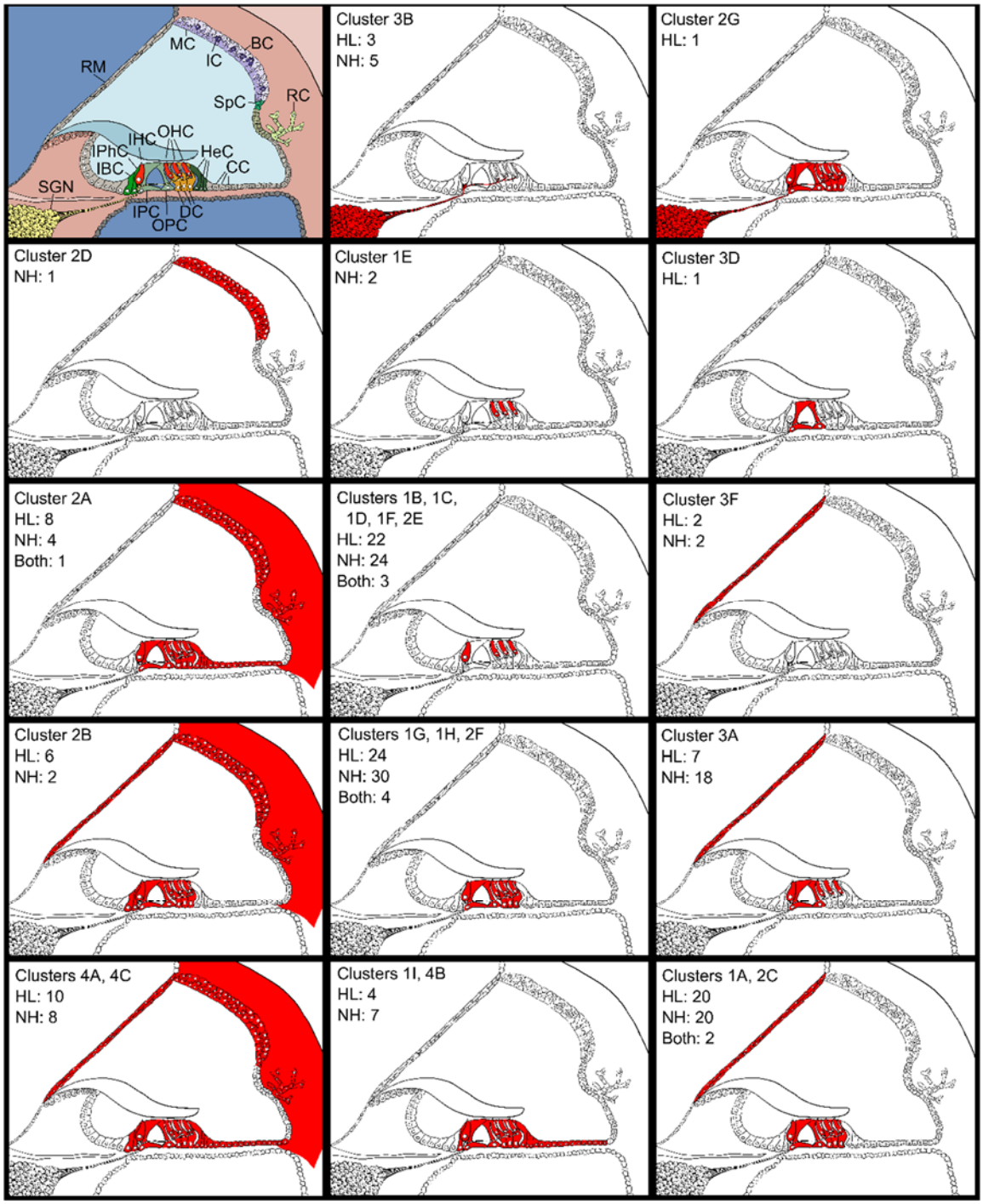

Figure 4. Schematic of the cochlear duct showing cell types (top left) and expression patterns based on the scRNAseq data downloaded from the gEAR database. The numbers show how many outlier genes were present in the cluster; "HL" for the number of outlier genes in hearing loss lists, "NH" for the number of outlier genes in normal hearing lists, and "Both" for where an outlier gene was present in a hearing loss list and a normal hearing list. See Table 3 for clusters and for gene names. RM=Reissner's membrane; MC=marginal cells; IC=intermediate cells; BC=basal cells; RC=root cells; SpC = spindle cells; SGN=spiral ganglion neurons; IBC=inner border cells; IphC=inner phalangeal cell; IHC=inner hair cell; OHC=outer hair cells; HeC=Hensen cells; CC=cells of Claudius; IPC=inner pillar cell; OPC=outer pillar cell; DC=Deiters' cells.

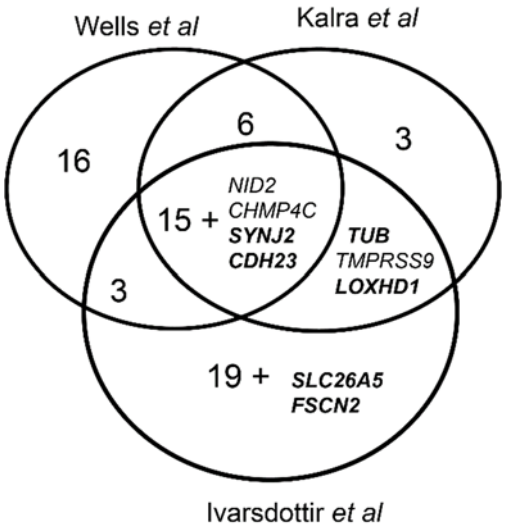

Figure 5. Comparison of gene lists from recent UK Biobank GWAS on self-reported hearing. Labelled genes are those also identified in this study (from the common, high impact variant outlier lists), and are not included in the numbers for those sections. Known deafness genes are in bold. The Wells *et al* and Kalra *et al* analyses used the UK Biobank data only while the Ivarsdottir *et al* analysis included other populations from Iceland.

Tables

Feature	Meaning	Source	Low impact intermediate variants	High impact rare variants	Low impact rare variants	High impact intermediate variants
Consequence	The effect of the mutation on the protein	Ensembl, ReMM	All consequences except for intergenic and intronic variants, variants in transcripts subject to nonsense-mediated decay, and variants up- or downstream of a gene OR Variants with a ReMM score > 0.95	5'UTR variants, splice site mutation, stop gain or loss, start loss, insertion, deletion, duplication, missense variants, and variants in mature miRNAs	All consequences except for intergenic and intronic variants, variants in transcripts subject to nonsense-mediated decay, and variants up- or downstream of a gene OR Variants with a ReMM score > 0.95	5'UTR variants, splice site mutation, stop gain or loss, start loss, insertion, deletion, duplication, missense variants, and variants in mature miRNAs
Minor allele frequency	How rare the alternative allele is in the population	gnomAD, 1000G, TOPMed, ESP6500	<0.1	<0.005	<0.005	<0.1
Pathogenicity	How likely the mutation is to impair protein function	CADD, REVEL, Sutr, SpliceAI	Not assessed	CADD > 25 or REVEL > 0.65 or Sutr > 1 (for 5'UTR variants only) or SpliceAI score > 0.5 (for splice site variants only)	Not assessed	CADD > 25 or REVEL > 0.65 or Sutr > 1 (for 5'UTR variants only) or SpliceAI score > 0.5 (for splice site variants only)
Number of variants			10,138,761	1,804,420	9,846,893	1,827,068

Table 1. Classification criteria for variants by impact and minor allele frequency.

		Variant load in normal hearing (all participants)			Variant load in hearing loss (all participants)		
MAF	**Impact**	**Genes**	**Deafness genes**	**Highly variable genes**	**Genes**	**Deafness genes**	**Highly variable genes**
Rare	High	13	2 (adj.p=1)	3 (adj.p=0.24)	14	1 (adj.p=1)	3 (adj.p=0.30)
Intermediate	Low	167	7 (adj.p=1)	22 (adj.p=3.19×10^{-7})	81	0 (adj.p=1)	10 (adj.p=0.0074)
Rare	Low	22	2 (adj.p=1)	1 (adj.p=1)	3	0 (adj.p=1)	3 (adj.p=0.00210)
Intermediate	High	307	21 (adj.p=0.00029)	30 (adj.p=8.82×10^{-7})	206	18 (adj.p=4.37×10^{-5})	14 (adj.p=0.19)
		Variant load in normal hearing (men)			**Variant load in hearing loss (men)**		
MAF	**Impact**	**Genes**	**Deafness genes**	**Highly variable genes**	**Genes**	**Deafness genes**	**Highly variable genes**
Rare	High	12	2 (adj.p=1)	1 (adj.p=1)	19	2 (adj.p=1)	5 (adj.p=0.0069)
Intermediate	Low	67	5 (adj.p=0.71)	17 (adj.p=4.41×10^{-10})	87	1 (adj.p=1)	10 (adj.p=0.0013)
Rare	Low	2	1 (adj.p=1)	1 (adj.p=1)	1	0 (adj.p=1)	0 (adj.p=1)
Intermediate	High	284	21 (adj.p=8.57×10^{-5})	21 (adj.p=0.0075)	245	9 (adj.p=1)	9 (adj.p=1)
		Variant load in normal hearing (women)			**Variant load in hearing loss (women)**		
MAF	**Impact**	**Genes**	**Deafness genes**	**Highly variable genes**	**Genes**	**Deafness genes**	**Highly variable genes**
Rare	High	10	1 (adj.p=1)	1 (adj.p=1)	13	2 (adj.p=1)	1 (adj.p=1)
Intermediate	Low	71	3 (adj.p=1)	12 (adj.p=5.95×10^{-5})	83	3 (adj.p=1)	7 (adj.p=0.59)
Rare	Low	0	0 (adj.p=1)	0 (adj.p=1)	5	1 (adj.p=1)	3 (adj.p=0.010)
Intermediate	High	259	17 (adj.p=0.0036)	19 (adj.p=0.017)	220	18 (adj.p=0.00011)	1s1 (adj.p=1)

Table 2. The number of genes, known deafness genes and highly variable genes in each high variant load list. The p values for the deafness and highly variable gene counts indicate the probability of finding that many known deafness genes or highly variable genes in the high variant load list, given the overall size of the gene list, the high variant load list, and the numbers of deafness genes and highly variable genes present in the regression analysis. P values were calculated using R and adjusted with a Bonferroni correction. Significant p values ($p<0.05$) are highlighted in yellow.

	Genes in hearing loss outlier lists	Genes in normal hearing outlier lists
Cluster 1A Middle in RM and OC, mid-low in LW, low in SGN and SC Hair cells, Deiters' cells, pillar cells and Reissner's membrane	*Crtc1, Cog1, Nob1, Iqcb1, Rrp36, **Elmo3**, Vwa8, Vps13b, Skiv2l, Noc2l, Ddx23, Aldh4a1, Szt2, Inha, Cdca3*	*Wdr33, Slc39a3, Rpusd3, Cnksr1, Prpf6, Lysmd4, Atr, Sgsh, Baz1a, Sec24d, Pld2, Zfp142, Sgip1*
Cluster 1B High in HC, mid-high in rest of OC, middle in RM, mid-low in LW, low Hair cells (*Synj2*)	*Mink1, Snapc3, **Ube3b**, **Synj2***	*Mink1, Prdm2, 4932438A13Rik, **Mpdz**, Tmem254a, Fggy, **Epn3***
Cluster 1C High in HC, mid-low in rest of OC, low in RM, SGN, SC and LW Hair cells (*Tmie*, *Myo7a*, *Ptprq*)	***Gpr152**, **Slc26a5**, Bmp3, **Tmc1**, Strip2, **Loxhd1***	*Tmprss7, Osbp2*
Cluster 1D Middle in HC, mid-low in rest of OC, low in RM, SGN, SC and LW Hair cells (*Pcdh15*, *Whrn*)	*Pias3, Gab2, Dll3, Ppp6r2, Fer1l6, Rimkla*	*Mn1, **Dvl3**, Nme3, Rgs11, Dll3, **Pkhd1l1**, Fam81b, **Pdzd7**, **Xirp2**, **Myo3a**, Abca7, Obscn, Arhgdig*
Cluster 1E High in OHCs, low in rest of OC, SGN, LW, RM and SC Outer hair cells		***Tub**, Sec16b*
Cluster 1F High in HC, middle in PC, low in DC, SGN, LW, SC and RM Hair cells (*Atoh1*, *Pou4f3*, *Otof*)	***Fscn2**, Chrna10, Lmod3*	*Lmod3*
Cluster 1G Middle in OC, mid-low in LW, low in SGN, SC and RM	*Ddx52, Dxo, Pdcd11, Itsn2, Dlg5, Grwd1, Ptpn13, Cpsf3, Ecd, Polr3c, Rbl2, Klc4, Hgh1, Myom1, Tmprss9, Chmp4c*	*Lmf2, Eif2b3, Grwd1, Nop9, Csrnp1, Mus81, Cpsf3, Ehmt1, Vps51, Cptp, Abcb8, Trpm4, Ccdc22, **Diaph1**, Vps13a, Pgm5*

	Genes in hearing loss outlier lists	Genes in normal hearing outlier lists
Cluster 1H Middle in OC, mid-low in LW and SC, low in SGN and RM Hair cells, pillar cells and Deiters' cells	*Trp53i13, Dnpep, Tgfbr1, E2f4, Cep63, Itgb1bp1, Fzd9, Adamts2, Me3,* ***Col11a1***	*Nsrp1, Sap30bp,* ***Prkra****, Rae1, Zfpl1, Zfp553, P3h3, Zfp428, Cadm1, Tcerg1, Itgb1bp1, Proser2*
Cluster 1I Mid-high in OC, middle in SC, low in SGN, LW and RM Hair cells, Deiters' cells, pillar cells and supporting cells	*Tmem132*	*Dnaja4,* ***Ush1c****,* ***Otog****, Man2b2, Lama5*
Cluster 2A Middle in OC and LW, mid-low in SC, low in RM and SGN Cells of the cochlear duct (*Myh14*)	*Krt10, Yy1,* ***Maz****, Gpr180, Tacc2*	*Nde1, Atp6v0a2, Arhgap17, Tacc2, Lamb2, Itpr2, Vegfb, Esrra,* ***Myh14***
Cluster 2B Middle in OC, RM and LW, mid-low in SC, low in SGN Hair cells, pillar cells and Deiters' cells, Reissner's membrane and	*Zfp740, Numa1, Akap13, Slc25a23, Pdlim5, Pam*	*2610507B11Rik, Tjp1*
Cluster 2C Mid-high in HC, middle in rest of OC and RM, mid-low in LW and Hair cells, pillar cells, Deiters' cells and Reissner's membrane	*Bri3bp, Agrn, Usp19, Kmt2c,* ***Nav2****, Asap1, Madd*	*Igf2bp3, Vsig10, Sf3a2,* ***Rest****, Srp68, Agrn,* ***Nav2****, Apbb1, Peg3*
Cluster 2D High in marginal and intermediate cells, middle in rest Stria vascularis (*Kcnj10*)		***Slc12a2***
Cluster 2E High in HC, mid-high in rest of OC, middle in LW, SC and RM, mid- Hair cells (*Myo6*)	*Nxf1, Fkbp2, Nbeal1,* ***Myo6****, Taf6l, Ccs*	*Ap2m1, Vamp2, Sra1,* ***Brd2***

	Genes in hearing loss outlier lists	Genes in normal hearing outlier lists
Cluster 2F High in HC, mid-high in rest of OC, middle in LW and SC, mid-low in Hair cells, Deiters' cells and pillar cells	*Ciapin1, Txndc5*	*Ciapin1, Naa38, Ddrgk1, Capns1, Yif1b, Dhrs1*
Cluster 2G High in OC, middle in SGN and SC, mid-low in LW, low in RM Hair cells, Deiters' cells, pillar cells and spiral ganglion neurons	*Kif21a*	
Cluster 3A Middle in HC and PC and RM, mid-low in DC and LW, low in Hair cells, pillar cells and Reissner's membrane	*Htt, Itfg2, Tcirg1, Atxn7, D230025D16Rik, Tjp3, Dach1*	*Dmxl1, Arid1b, **Polg**, Gramd1a, Abca3, Srr, Sft2d3, **Nectin1**, Rfx7, Lgr4, Dock6, **Gli3**, Hmcn1, Rgs3, Hectd4, Tsr1, **Pjvk**, Ltbp1*
Cluster 3B Middle in SGN, mid-low in OC, LW and RM, low in SC Spiral ganglion neurons	*Tango2, Prune2, Abr*	*Kif1a, Mapk8ip3, **Ccdc92**, Map2, **Ercc6***
Cluster 3C Mid-low in LW and HC, low in rest of OC, SGN, RM and SC	*Manea, Slc13a4*	*Fkbp7, Wfdc1, Fbln7, **Lama2**, Pcolce, Ccdc141*
Cluster 3D High in PC, middle in RM, mid-low in LW, low in rest of OC, SGN, SC Pillar cells	*Thsd7a*	
Cluster 3E Mid-low in OC, LW, and RM, low in SGN and SC	*Mylk*	*Ap1s2, Rusc2, Rin2, **Dmd**, Sema6d, Piezo1, Plec, Fam189b, Slc17a7*

	Genes in hearing loss outlier lists	Genes in normal hearing outlier lists
Cluster 3F High in RM, mid-high in spindle/root, low in rest of LW, Reissner's membrane	*Ap1s3, Atp13a5*	*Frem1,* ***Slc26a4***
Cluster 4A Mid-high in OC, middle in LW, RM and SC, low in SGN Cochlear epithelium (*Sox10*)	*Eif5b, Stk16, Tex261, Tnrc6a, Dguok, Fndc3a, Tcn2, Sdf2l1*	*Fam162a, Nadk, Zfhx3, Lzts2, Mospd3, Kdm6b, Syne2*
Cluster 4B High in OC, mid-high in SC, low in SGN, marginal and intermediate Sensory epithelia (*Six1*)	*Ccdc28b, Dpp4, Igfbp5*	*P3h4,* ***Col11a2***
Cluster 4C High in OC, LW, SC, middle in RM, low in SGN	***Gjb2****, Rps9*	*Srp14*

Table 3. All the heatmap clusters which had outlier genes (from the intermediate, high impact variant analysis, Supplementary Table 5D) are listed with their expression description, their classification (if one had been assigned) and the outlier genes separated by whether they were from hearing loss outlier lists or normal hearing outlier lists. Ten genes are present in both a hearing loss outlier list and a normal hearing outlier list. Known deafness genes are in bold.

Supplementary Figures

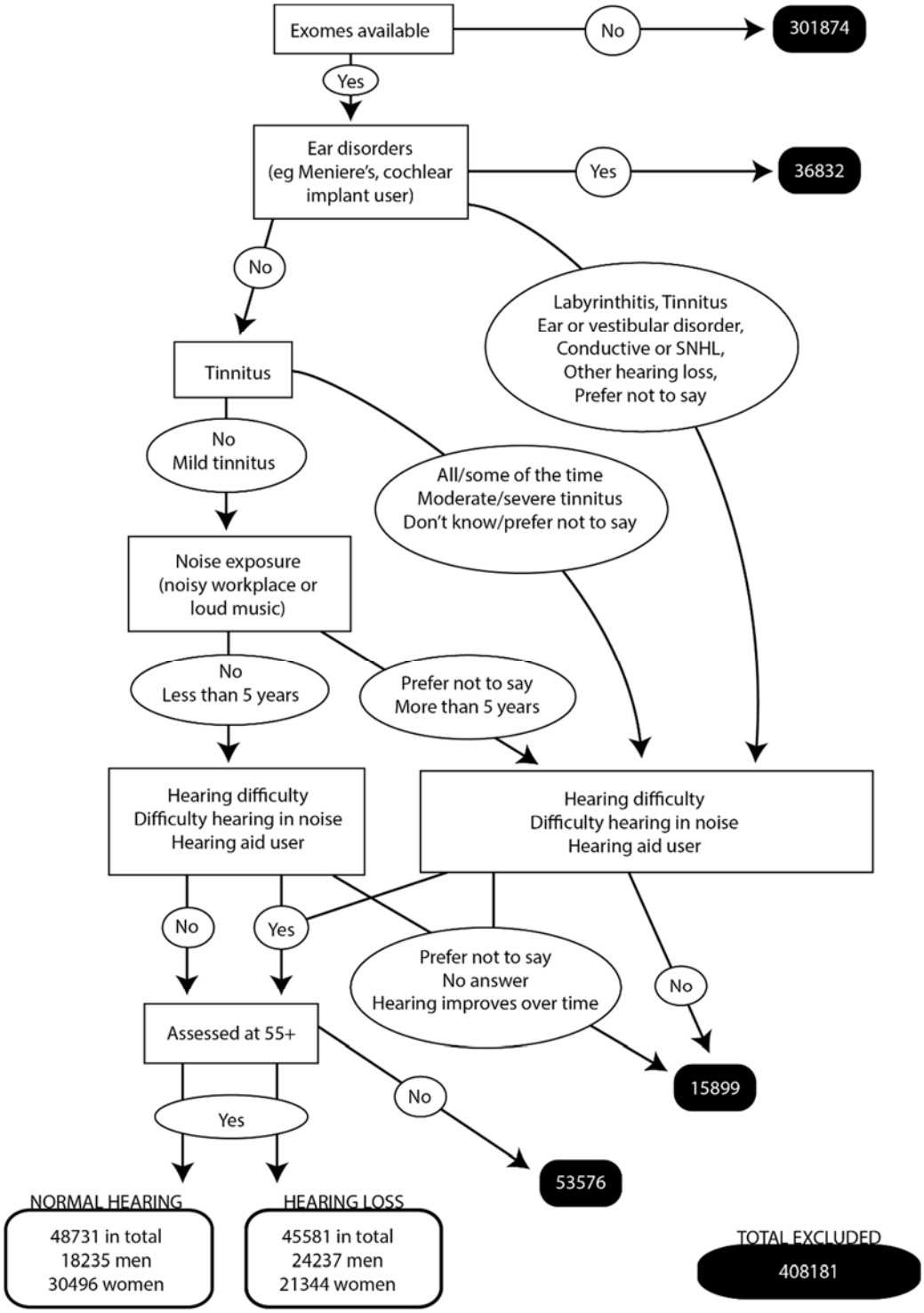

Figure S1. Flow diagram showing the process of categorising participants by availability of exomes, their questionnaire answers, and age when assessed. In black boxes, the number of people excluded at each step.

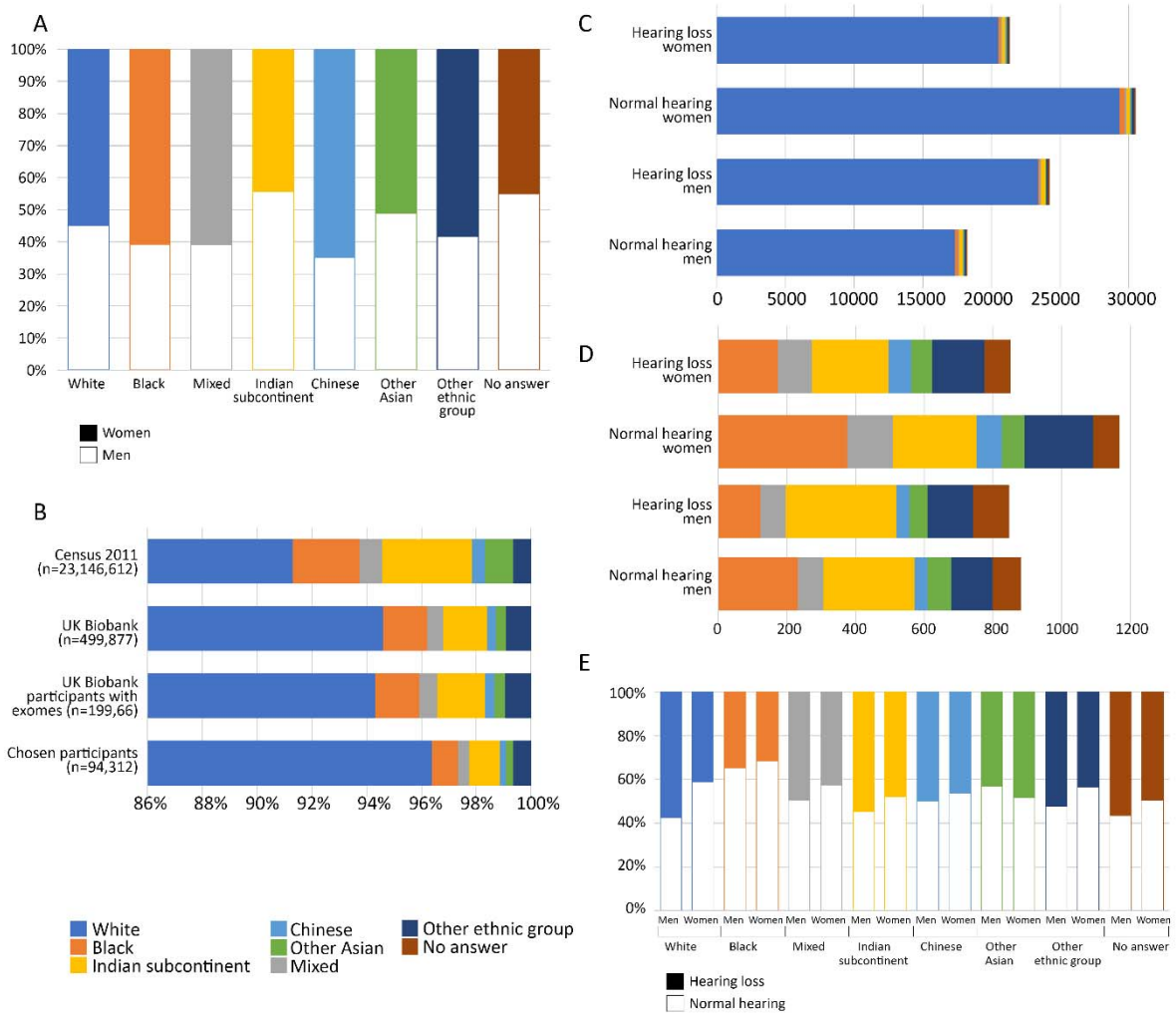

Figure S2. Bar charts showing numbers of people by sex, self-reported ethnicity and self-reported hearing phenotype. Ethnicities were combined to form broad groupings (see Suppl. Table 1). A) Percentage of men (empty bar) and women (filled bar) in each broad ethnic grouping in this study. B) Self-reported ethnicity in the UK census 2011 [11], the entire UK Biobank, UK Biobank participants with exomes, and this study, shown as a percentage of the total. Note that the x-axis starts at 86%, in order to show the minority percentages clearly. C) Self-reported ethnicity of people included in this study. D) Self-reported minority ethnicities of people included in this study. Note the difference in the range of the x-axis between C and D. E) Percentage of men and women classified in the self-reported hearing loss phenotype (filled bar) and classified in the self-reported normal hearing phenotype (empty bar) in each broad ethnic grouping.

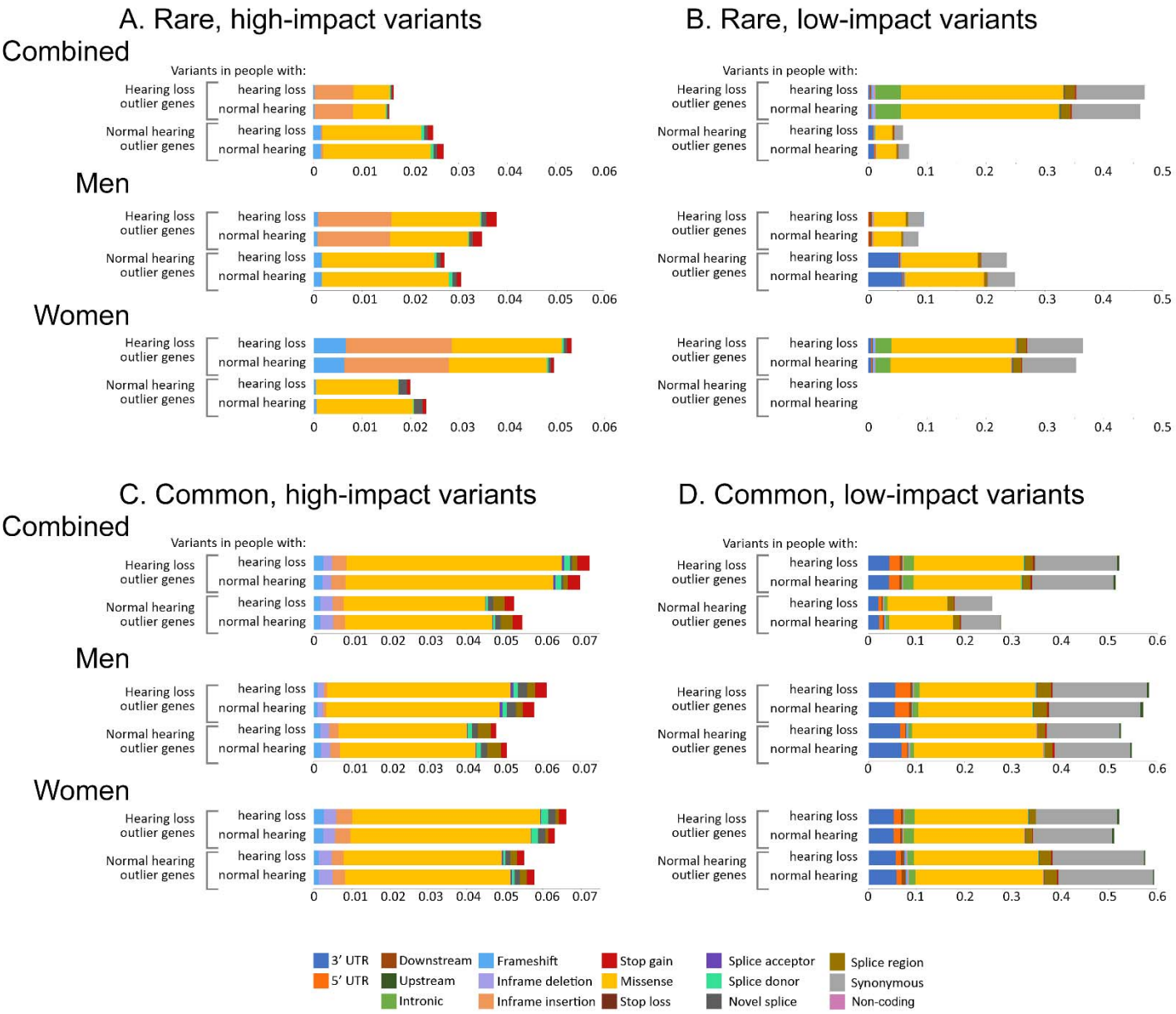

Figure S3. Bar charts showing the number of variants per person, per gene, in the genes of each high variant load list, coloured by maximum impact (see Suppl. Table 2). For each list of outlier genes, we have plotted the variants in people with hearing loss separately from those with normal hearing, to ask whether the variant impacts are different between the phenotype groups. Variant counts which came to less than 0.001 variants per person per gene were excluded (since they would not be visible on the bar charts).

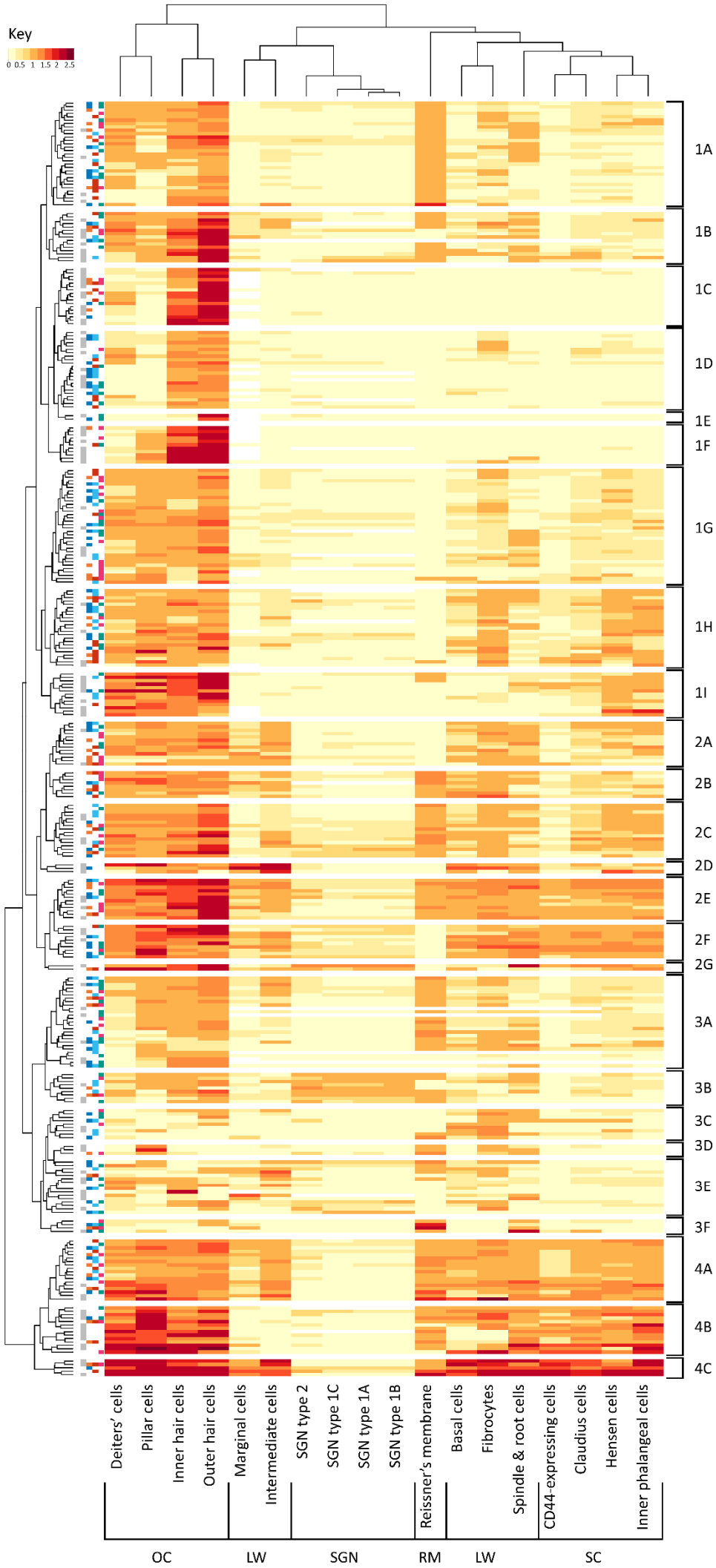
Key
0 0.5 1 1.5 2 2.5
1A
1B
1C
1D
1E
1F
1G
1H
1I
2A
2B
2C
2D
2E
2F
2G
3A
3B
3C
3D
3E
3F
4A
4B
4C
Deiters' cells
Pillar cells
Inner hair cells
Outer hair cells
Marginal cells
Intermediate cells
SGN type 2
SGN type 1C
SGN type 1A
SGN type 1B
Reissner's membrane
Basal cells
Fibrocytes
Spindle & root cells
CD44-expressing cells
Claudius cells
Hensen cells
Inner phalangeal cells
OC
LW
SGN
RM
LW
SC

Figure S4. Heatmap showing maximum expression levels of the high quality one-to-one mouse orthologues of outlier and known deafness genes, across a range of inner ear cell types between E16 and P35. Cell types and genes are clustered according to their expression along the x and y axes respectively. We further defined gene clusters first based on the R dendrograms (taking the first three divisions, resulting in four main clusters numbered 1-4) and then based on the expression within specific cell types (subclusters labelled with letters). To annotate cell types, we grouped Deiters' cells, pillar cells, and inner and outer hair cells (as organ of Corti, OC), all spiral ganglion neuron types (as SGN), marginal, intermediate and basal cells, fibrocytes and spindle and root cells (as lateral wall, LW), and Hensen cells, cells of Claudius, inner phalangeal cells and CD44-expressing cells (CD44) as supporting cells (SC). Expression in Reissner's membrane was left as a single group (RM). The blocks of colour on the left denote deafness genes (grey, first column), outlier genes in normal hearing (blue) and hearing loss (orange) in all participants (second column), outlier genes in normal hearing (cyan) and hearing loss (red) in men (third column) and outlier genes in normal hearing (teal) and hearing loss (magenta) in women (fourth column). The genes are listed in the same order in Supplementary Table 7, along with their classifications and all annotations.

Supplementary Tables

Broad ethnic grouping	Answers assigned to each grouping
White	British, Irish, Any other White background, White, British/Irish, White/British, Other/British, Other/Irish, British/Irish/Any other White background, British/Any other White background, Other/Any other White background
Black	African, Other/African, Caribbean, Other/Caribbean, Any other Black background, Black or Black British
Mixed	White and Black African, White and Asian, Any other mixed background, White and Asian/Bangladeshi, White and Asian/Any other Asian background, Any other White background/Any other mixed background, Mixed, Other/Any other mixed background, British/Any other mixed background, White and Black Caribbean/Caribbean, Chinese/Any other White background/White and Asian, White and Black African/Any other mixed background, Other/White and Black African, White and Black Caribbean/White and Black African, White and Black Caribbean, White and Asian/Any other mixed background
Indian subcontinent	Indian, Indian/Any other Asian background, Bangladeshi, Pakistani, Indian/Pakistani
Chinese	Chinese
Other Asian	Asian or Asian British, Other/Any other Asian background, Any other Asian background, Chinese/Any other Asian background
No answer	Don't know, Prefer not to answer, Blank
Other	Other ethnic group

Supplementary Table 1. Groupings of answers given to the question about ethnicity. Where answers are separated by "/", this indicates multiple different answers given by the same participant, since the questionnaire could be filled in up to 3 times by each person.

Supplementary Table 2. Variant classifications, sourced from http://www.ensembl.org/info/genome/variation/prediction/predicted_data.html, with some modifications to the order.

Supplementary Table 3: Genes reported to underlie hearing loss in humans and/or mice. One of the mouse deafness genes has no known human orthologue (*Akr1c12*), so was not used for comparison in this study.

Supplementary Table 4**.** Genes in which a large number of variants are frequently reported in sequencing projects, compiled from [30-32].

Supplementary Table 5. Outlier gene lists for rare, high impact variants (A), intermediate, low impact variants (B), rare, low impact variants (C) and intermediate, high impact variants (D). Deafness genes are shown in bold, and highly variable genes with a grey background.

Supplementary Table 6. GO terms identified by gProfiler as being enriched in the high variant load lists. The number of genes annotated to the GO term is the full number of genes in the Gene Ontology which are annotated with that GO term. The number of genes in the query is the number of genes in each high variant load list which have an annotation in that GO domain (ie Cell Component, Molecular Function or Biological Process). The intersection is the number of genes in the high variant load list which have that GO term annotation.

Supplementary Table 7. The heatmap genes with their cluster definitions and relevant high variant load lists, arranged in order to match the rows of Supplementary Figure 4. The gene expression patterns which distinguish the clusters are described under “Expression”, summarised by single cell types or groups of cell types (OC: Deiters’ cells, pillar cells, inner and outer hair cells; RM: Reissner’s membrane; SGN: spiral ganglion neurons types 1A, 1B, 1C, 2; LW: marginal, intermediate and basal cells, fibrocytes, spindle and root cells; SC: Hensen cells, cells of Claudius, inner phalangeal cells and CD44-expressing cells). Expression levels over 2 (red) are described as “high”, levels between 1 and 2 (orange) are described as “middle”, and levels between 0 and 1 (yellow) are described as “low”; where expression varies between two classifications, both are listed (eg “mid-high”). Cluster classifications are given where expression was strong in specific cell types, and known marker genes are listed in brackets. Where there was no high or middle expression, clusters have not been classified.

Supplementary Tables 2-7 are included in the accompanying Excel file.