Genetics of age-related hearing loss

Helena RR Wells
Tracey A Newman
Frances MK Williams

1Department of Twin Research and Genetic Epidemiology
King’s College London
London
SE1 7EH

2CES, Medicine
B85,M55, Life Sciences
University of Southampton, UK
SO17 1BJ

Correspondence to
Prof FMK Williams
Dept of Twin Research
King’s College London
St Thomas’ Campus
London SE17EH
0207188 6743
frances.williams@kcl.ac.uk

Abstract
Age-related hearing loss has recently been confirmed as a common complex trait, that is, it is heritable with many genetic variants each contributing a small amount of risk, as well as environmental determinants. Historically, attempts to identify the genetic variants underlying ARHL have been of limited success, relying on the selection of candidate genes based on limited knowledge of the pathophysiology of the condition, and linkage studies in samples comprising related individuals. More recently genome-wide association studies have been performed, but these require very large samples having consistent and reliable phenotyping for hearing loss, and early attempts suffered from lack of reliable replication of their findings. Replicated variants shown associated with ARHL include those lying in genes GRM7, ISG20, TRIOBP, ILDR1 and EYA4. The availability of large biobanks and the development of collaborative consortia has led to a breakthrough over the last couple of years, and many new genetic variants associated with ARHL are becoming available, through the analysis publicly available bioresources and electronic health records. These findings, along with immunohistochemistry and mouse models of hearing loss look set to help disentangle the genetic architecture of ARHL, and highlight the need for standardisation of phenotyping methods to facilitate data sharing and collaboration across research networks.

Significance statement
The role of genetics in influencing ARHL is now clear: multiple genetic variants are involved, each contributing a small effect. The recent advances in genetics come from scale: in keeping with many other traits, there is a recognised trade-off between accurate phenotyping and sample size. Studies show that questionnaires about hearing difficulty, applied to very large samples comprising tens or hundreds of thousands of individuals, can reveal new variants and hint at novel pathways in ARHL. Many such datasets exist around the world and progress is sure to be rapid. The variants and pathways identified will need further investigation so that therapeutic benefits can be drawn.
Introduction

Age related hearing loss (ARHL) is considered by many to be a common complex trait; development of the condition being dependent on a number of genetic and environmental factors. It is a neurodegenerative condition that varies between individuals in regard to onset, pathology and progression. ARHL affects a considerable proportion of the population, and is the most common sensory impairment in the elderly population with a third of people aged >65 years of age experiencing debilitating hearing loss (www.who.int/pbd/deafness/estimates/en/). ARHL is a form of sensorineural hearing loss, yet little is known about the pathology of the condition. It is understood that in the aging cochlea, synaptopathy develops between the sensory hair cells and cochlea nerve fibers1, and in conjunction with sensory hair cell damage, there is a gradual loss of functional sensory and neuronal cells2,3. Currently there are no cure or prevention strategies in place, and the only treatment options are hearing aids and assistive devices.

Understanding the genetic components of ARHL will elude to the pathology of the condition with the potential to translating the findings into clinical applications. ARHL is understood to be genetically heterogeneous as indicated by genetic association studies to date4. In addition, the 150 genomic loci that have been linked to non-syndromic hereditary hearing loss (https://hereditaryhearingloss.org) highlight the very many genetic mechanisms that may lead to deafness. Under the common trait-common variant hypothesis, likely thousands of genetic risk loci contribute to the development and progression of ARHL and this genetic complexity reflects the complexity of the anatomical, and in turn cellular and molecular makeup of the auditory system. Heritability estimates for ARHL are moderate, with estimates ranging from 30-70%5–10, with variability likely to be due to the different phenotyping methods, limited sample sizes and differing sample demographics - particularly age - of ARHL genetic studies4,11–20.

Methods to uncover genetic components of common complex traits have advanced rapidly from the early days of candidate gene studies to genetic association studies (GWAS) and the examination of linked traits to extract further information. GWAS requires very large samples to be adequately powered but several ARHL risk loci can be identified from new bioresources. ARHL genetic research to date has focused mainly on molecular work with mouse models, differing mouse strains21 and most recently the study of selective genetic alteration using CRISPR/Cas22. This review will cover work to date conducted in human samples in the pursuit of identifying and validating genetic risk variants associated with ARHL.

Defining ARHL in the general population

ARHL studies to date have suffered to some extent from inconsistency in the measures used to classify the condition. Traits may be quantitative or qualitative but valid genetic analysis of complex traits relies on defining and ascertaining a robust phenotype which is reproducible over time. In a quantitative trait accurate and reproducible measurement allows trait severity to be determined. In qualitative traits individuals are assigned case or control status according to predetermined cut-offs with respect to the trait of interest. Pure tone audiometry (PTA), signal to noise ratio (SNR) testing and self-report measures of hearing have all been used to study ARHL and are discussed below.

PTA is considered the ‘gold standard’ measure for hearing assessment. The test is presented as a set of pure-tone pulses at different frequencies and intensities and provides detailed analysis of an individual’s hearing thresholds at each frequency measured, independently for each ear, with results presented as an audiogram. The individual is asked to indicate, usually by pressing a hand-held button, when the sound can be heard. The aim is to determine their ability or loudness threshold needed to perceive sound over a range of different standardised frequencies. While among the most sensitive assessments of hearing ability, PTA is not practical for large-scale population studies as it requires specialist equipment and trained staff and is therefore expensive. In addition, the test does not encompass a measure for distinguishing and interpreting sounds presented alongside background noise; this is a common symptom of ARHL. Studies of PTA show that they do not detect this form of ‘hidden hearing loss’2,23. There is currently no consensus on ARHL definition based on PTA thresholds and researchers have used a range of data reduction methods such as...
principle component analysis and z-scores to generate normally distributed data for genetic association study$^{4,13,15,17,19}$.

SNR testing measures an individual’s ability to detect and process sounds, typically triplet digits, at different frequencies with varying levels of background noise. The test produces a speech reception threshold (SRT) for each individual, indicating the level at which they correctly interpret 50% of the triplet digits presented. The SNR test has been shown to be a useful surrogate for a PTA measure in the study of ARHL$^{24}$ and is likely a more realistic measure of how ARHL impacts an individual’s ability to hear in daily life due to the incorporation of background noise. It is also amenable to genetic studies with large cohorts as it can be completed online or via telephone without additional expensive equipment or trained staff present. Currently, however, there is a lack of large of cohort studies having both SNR and genetic data available and testing protocols vary between studies - standardisation of research methods is required urgently$^{20,24,25}$. Increasingly for complex traits, self-report measures have been found to be useful for large-scale genetic association studies and this is a practical method for collecting phenotype data on a suitably large sample, for example by the American company 23andMe. Although self-reported hearing ability is to some extent a subjective measure, genetic associations with ARHL have been found using this approach$^{20}$(see GWAS section below). Self-report hearing measures include simple questions regarding an individual’s hearing ability, hearing ability under different conditions such as with background noise, use of hearing amplification devices such as hearing aids, or any doctor’s diagnoses of hearing-related problems (which usually implies a PTA has been performed) and hearing-related symptoms such as tinnitus.

Linkage analysis

Linkage analysis is a traditional approach used to identify chromosomal regions that harbor genetic disease markers. The technique makes use of the laws of genetic recombination; markers in close proximity on chromosomal regions are more likely to be inherited together. Family pedigrees are therefore used to identify chromosomal regions that cosegregate with a disease or trait of interest. Logarithm of odds scores are calculated to assess the probability that the observation is due to linkage rather than chance. A linkage analysis study on a group of ARHL cases presenting with high frequency hearing loss was conducted with a number of microsatellite regions on DFNA5. The DFNA5 gene had been identified in non-syndromic sensorineural, progressive hearing loss in multiple families of different genetic backgrounds$^{26–28}$ and was therefore a strong candidate gene for ARHL, but here no evidence of linkage was observed$^{29}$. In 2003 a genome-wide linkage study for ARHL was performed with a sample of 1789 individuals from the Framingham study, derived from 328 extended pedigrees$^{30}$. At the time, heritability of ARHL had been estimated as 40% in a classical twin study from the Danish twin register$^6$, while the linkage study provided broadly similar results with heritability estimates of 31% and 38% for age-adjusted low- and medium pure tone average frequencies respectively. Six chromosomal regions were identified having suggestive evidence of linkage, four of which contained genes known to be involved in forms of deafness, three of which are implicated in Usher syndrome. While this study may indicate the role of congenital deafness genes in adult, progressive hearing loss, no statistically significant findings were reported.

Linkage analysis has also been performed using the self-report hearing measures of a cohort of elderly male twin US veterans from World War II and the Korean war. Four hundred markers were studied, and a region of suggestive linkage was identified on chromosome 3, in the same region that DFNA18 is located. This is the first report suggestive of DFNA18 having a role in general hearing impairment in the aging population$^{31}$. A third linkage study was performed in 2008 and although no significant associations were identified, variation in a region on chromosome 8 was correlated with a concave audiogram shape$^{32}$.

In keeping with many other common complex traits, linkage analysis had limited success in identifying ARHL susceptibility loci owing to its inherent limitations. The power to detect linkage via pedigree analysis relies on the presence of highly penetrant, rare variants such as those in Mendelian or oligogenic disorders$^{33}$.
Common, polygenic traits such as ARHL are caused by multiple, lower penetrance variants. The main limitation of linkage analysis however is its limited power of resolution, as the regions denoted by the widely spaced markers - usually microsatellites - are large, and may contain hundreds of possible candidate genes.

**Candidate gene studies**

A candidate gene approach may be used to identify a causal SNP or genome variant when there is prior knowledge of a gene, biological mechanism or pathway associated with the phenotype of interest. This method is hypothesis driven and so it is susceptible to false positive results. When genotyping costs were very high in the early 2000s this was the most cost-effective approach. Variants in genes known to cause monogenic forms of hearing loss also provided candidate genes for ARHL association studies. As costs fell and genome-wide association studies became more widely available, their advantage for being hypothesis-free was fully recognised, with the realisation across many common complex traits that considerable numbers of candidate gene studies had provided false results and their results could not be reproduced even using very large samples in GWAS.

As noted earlier, DFNA2 was considered a strong candidate gene for ARHL. Using the same data as for the linkage analysis study, a case-control association study was conducted, and again no evidence of association was found. A candidate gene study that used samples from 7 different populations tested for association between ARHL and GSTT1, GSTM1 and NAT2, as polymorphisms in these genes were understood to have a role in defence against reactive oxygen species, a putative pathogenic mechanism in ARHL. In the combined European samples a significant association with NAT2*6A was observed while in the Finnish subsample associations with GSTT1 and GSTM1 were observed. In 2008 a more comprehensive candidate gene study of ARHL was performed using 768 SNPs spanning 70 genes. GRHL2 was found to be associated with ARHL; one SNP was significantly associated and surrounding SNPs showed suggested association, with concordant direction of effect in all nine contributing cohorts. One SNP reached the multiple-testing p value threshold of association; the 70 genes were selected from monogenic forms of hearing loss genes in both mice and men.

The most comprehensive, large-scale candidate gene study on normal hearing was conducted using a multistep approach. Nineteen candidate genes previously identified in published ARHL association analyses but not genome-wide significant were examined for gene expression in mouse cochlea tissue. Of 12 genes showing cochlea expression, 9 were replicated in independent samples drawn from the Caucasus and Central Asia. A relationship was also identified between audiometric patterns and 7 of the replicated genes (CDH13, GRM8, ANK2, SLC16A6, ARSG, RIMBP2 and DCLK1) which also showed association with audiometric pattern.

The gene GRM7 contained the first variant to be identified by association analysis in 2009, in a population sample comprising cases and controls collected in 8 centres in 6 European countries (n=1700 total). The top associated variants (Finnish and non-Finnish samples considered separately as the former is an isolated population) were taken to independent replication in 138 samples and the GRM7 variant was replicated. This gene has been the focus of a number of subsequent candidate gene studies. The first study aiming to replicate this finding was performed on a European-American population (687 individuals from the Rochester cohort), and used a range of hearing measures in an attempt to define better the different pathologies contributing to the ARHL phenotype. Six phenotypes were defined and tested and three variants from GRM7 examined: rs11928865, Haplotype 6 and Haplotype 7. The study confirmed an association between variation in GRM7, PTA frequencies and SRT scores. This study incorporated age and sex into z-score phenotype measures and stratified by sex. A second study assessed whether GRM7 variants were associated with ARHL risk in Han Chinese. Here individuals were assigned to phenotype groups...
based on the shape of their audiograms from PTA measures utilising K-cluster analysis, following Schuknecht’s model of the four sub-pathologies contributing to ARHL. Significant associations were identified for rs11928865, a SNP in intron 1 of GRM7 for the audiogram phenotypes ‘abrupt loss’ and ‘sloping shape’, but not for ‘flat shape’ and ‘8kHz dip’. This work was among the first to indicate the presence of different genetic factors underlying the observed audiogram shapes.

The latest study to assess GRM7 risk variants in an ARHL population sample tested 4 SNPs in elderly Taiwanese and found one associated in a dominant pattern. Two studies have analysed GRM7 risk variants in relation to traits linked to ARHL; noise-induced hearing loss in a Han Chinese sample and tinnitus and ARHL in a Portuguese sample. Furthermore, a number of subsequent GWAS reported summary statistics for GRM7 SNPs included in their analysis. Although promising, the replicated associations of GRM7 require further validation. First, associated SNPs are located in different regions of the GRM7 gene and were identified using differing phenotype measures in diverse ethnic samples. Secondly, the initial GWAS study did not report a significant association and the latest high powered association studies using very large samples have failed to support an association with GRM7. There is clearly a complexity to the interpretation of findings in different study designs and evidence from other traits shows that candidate gene association studies have provided many false positive results, as demonstrated by well powered GWAS using very large samples.

Until recently the lack of evidence supporting pathogenic mechanisms in ARHL has been the main limitation to the selection of suitable candidate genes. The recent development of very large population samples, and the evidence that simple questions about hearing may provide suitable phenotype for association study (Stacey et al in submission) are leading to a rapid expansion in studies of ARHL. In addition, animal models such as the mouse can be used for extensive, functional annotation of findings from population studies and are discussed in detail elsewhere.

**Genome-wide association analysis (GWAS)**

Initial GWAS studies on ARHL used samples comprising hundreds of individuals, while recent studies feature samples of tens- or hundreds of thousands. Increased genome array densities and imputation reference panels also contribute to improved resolution of association findings. Recently published studies attest to the highly polygenic nature of the trait.

In 2009 the first GWAS on ARHL was published with 1,692 participants, allocating participants to case-control status based on their hearing Z-scores derived from PTA frequency thresholds. The study used pooled samples and while it did not report any genome-wide significant associations, 23 lead variants having suggestive association were examined in an independent sample. Gene expression analysis and immunohistochemistry of the lead gene finding showed its product metabotropic glutamate receptor 7 was shown in mice cochlea and also in a single human cochlea sample. This study revealed GRM7 as the most promising gene association at the time. The following year, a group published an association analysis but used ARHL as a quantitative trait, using principle components derived from PTA thresholds. Again no significant associations were identified in the sample (n=352), but the top-ranking association was reported as a SNP in gene IQGAP2. A SNP upstream of GRM7 was reported as 7th-ranked association with PC1, but no further replication was performed. A second GWAS to use a quantitative phenotype approach was published in 2011 and also used principle components derived from PTA thresholds. This study had a larger sample size (n=3,417) and its authors performed meta-analysis of samples from across Europe. Four candidate genes were highlighted DCLK1, PTPRD, CIMP and GRM8. Replication was not conducted in the original study but the association with GRM8 was later supported in a 2014 GWAS.

In 2013, Nolan et al. studied three cohorts (1958 Birth Cohort; London ARHL cohort; Isolated populations cohort) totaling 6,134 individuals using a quantitative phenotype derived from logistic regression analysis of PTA thresholds. The focus of the study was ESRRG, where an association was identified in females but not
males in two of the three study cohorts\textsuperscript{12} and the link was further supported with work conducted in ESRRG knock out mice. A year later Fransen et al. performed a GWAS and attempted replication of all previous associations, concluding that ARHL is likely a highly polygenic trait. Genes highlighted in the new analysis were \textit{ACVR1B} and \textit{CCBE1} as possible associations with principle components of PTA thresholds (N = 2,161\textsuperscript{4}). At this stage, none of the five ARHL association studies that had been performed had revealed any significant associations with risk variants and there was little replication of suggestive findings between studies.

Later in 2014 two studies published the first genome-wide significant associations between risk loci and ARHL. Firstly, salt-inducible kinase 3 (\textit{SIK3}) was identified in a meta-analysis of samples (overall n=4,939) also principle components from PTA. Although this finding is yet to be replicated in an independent sample, SIK3 expression was observed in the stria, hair cells and spiral ganglion neurons in P0-P5 mice\textsuperscript{17}. The second study identified significant associations near to \textit{SLC28A3} and \textit{PCDH20} and nominal replications were reported\textsuperscript{18}.

A very much larger GWAS published in 2016 used self-reported hearing and SRT data obtained from electronic health records. The study identified (n >6500 ARHL cases and n >45,000 controls) 2 novel SNPs and these were replicated, in genes \textit{ISG20} and \textit{TRIOBP}. As \textit{TRIOBP} has already been associated with other forms of hearing loss, the group then took a candidate gene approach and identified two novel SNPs in \textit{ILDR1} and \textit{EYA4} \textsuperscript{20}. The discovery association analysis was conducted with the non-Hispanic Whites subset of the GERA cohort, and replication of the lead associated SNPs was performed in subsets having different ethnicity and different ARHL phenotypes (SRT and speech discrimination score, SDS and self-report measures).

The latest ARHL GWAS to use phenotypes derived from PTA thresholds was performed by the CHARGE consortium and consisted of a meta-analysis of 6 cohorts (n=9,675) and subsequent attempted replication in a further 5 samples (n=10,910) with additional \textit{in silico} clinical validation using self-report hearing measures obtained in UK Biobank. In an attempt to divide the sample by putative pathologies of biological pathways, four z-score phenotypes were derived from PTA thresholds, each describing a hearing loss of a different frequency range. Four loci reached genome-wide significance in the discovery sample, yet only one locus (\textit{SPIRE2}) was nominally replicated in the five replication cohorts. In the UK Biobank sample significant associations were seen at 5 of the loci that were highlighted in the discovery analysis (\textit{IPP, SPTBN1, ILDR1, TRIL and ISG20}) (Nagtegal et al accepted Scientific Reports 2019).

UK Biobank has been used to perform the largest ARHL GWAS to date, comprising n>250,000 White-British individuals aged 40-69. Two phenotypes were derived from participant responses to questionnaire measures regarding (1) hearing difficulty and hearing difficulty in the presence of background noise, and (2) frequent hearing aid use. The study identified 44 independent genome-wide significant loci. \textit{TRIOBP} was the only previously identified locus that was identified, with 34 of the other loci have no previous links to any form of hearing loss phenotypes. Having assessed all loci previously highlighted in ARHL GWAS prior to the CHARGE study, the only SNPs that were replicated in this sample were variants located in \textit{TRIOBP} and close to \textit{ISG20}\textsuperscript{44}. Within the study various bioinformatic approaches were used including pathway, correlation and expression analysis. The results indicate that genetic risk variants are involved in a number of processes from cochlea to cortex, and positive genetic correlations between hearing loss and psychological traits were identified\textsuperscript{41}.

For the vast majority of complex traits, GWAS findings generally explain only 20-30\% of the trait variance\textsuperscript{45} and ARHL is likely to be no exception. GWAS studies of ARHL have to date not revealed any variants having large effect, and it is likely that the so-called missing heritability lies with a number of other mechanisms such as rare SNPs not currently captured on GWAS arrays, common SNPs with small effect sizes which may be identified with much larger sample sizes, or copy number variations (CNVs). Epigenetic factors and gene-environment interaction are also likely to account for complex traits such as ARHL; epigenetic marks may be inherited and they play an important role in gene regulation which is particularly influential in traits associated with aging.
**Genome sequencing studies**

Pathogenic variants have been identified via sequencing in multiple family pedigrees that have hereditary forms of hearing loss, however the use of these methods on a population-scale for complex traits is in its infancy. Exome and genome sequencing are currently not widely used to identify risk variants for common complex traits. This is mainly due to practical limitations; these methods are relatively expensive and computationally demanding. The two methods permit the identification of rare variants that are not included in GWAS SNP arrays or imputation reference panels. Recent work on height and BMI demonstrates that trait heritability is largely accounted for when genome sequencing is deployed (compared to SNP-based methods), showing that rare variants account for the vast majority of the so-called ‘missing heritability’ 46.

A study using exome sequencing to study adult onset progressive hearing loss aimed to evaluate the method as a diagnostic technique for ARHL in a clinical setting. Thirty patients were selected and placed into one of four subsets; predicted dominant HL, predicted recessive HL, a metabolic audiometric phenotype or a sensory audiometric HL47. Analysis focusing on common variants revealed that all 30 participants had at least 10 deafness gene variants with ≥1 predicted pathogenic mutation, figures that remained when more stringent filters were applied. However, authors emphasized that these putative causal variants must be treated with caution as they could be subject to false pathogenicity reports or incomplete penetrance. As ARHL data become widely available, it will be a natural progression to use sequencing data at a population level to identify risk variants. With genome sequencing, variants that have otherwise been undetected may be identified, such as rare variants in regions of low LD that are not present on genotyping platforms. Currently there are a limited number of large population cohorts with genome sequencing data, and even fewer that also collect hearing phenotypes. For the immediate future, the real value of sequencing in terms of research into ARHL genetic risk, will likely be in the annotation of loci identified by GWAS. With genome sequencing, it is possible to fine map a region of association to locate the pathogenic variant and thus predict its function. Unlike exome sequencing, genome sequencing includes regulatory regions which are where most common variants identified in GWAS lie. A priority before this is possible, therefore, is to standardise methods of hearing data collection and ARHL definition so that existing cohorts can be compared and meta-analysed: only by developing world-wide collaborations can the sort of sample size be achieved necessary to power high quality genetic association studies.

**Conclusion**

An increasing array of genetic approaches as well as population samples having hearing phenotypes has led to the recent identification of a number of pathogenic variants associated with ARHL. Age remains the single greatest risk factor, and a number of environmental risk factors are associated with the incidence and progression of this complex trait which is now known to be highly polygenic. A more detailed genetic dissection of ARHL requires considerable mapping of cohorts with adequately characterised hearing and the relationship between the various phenotypes, including questionnaire-based tools, is currently being explored (Stacey et al in submission). The full extent of the interplay between genetic predisposition to auditory pathway susceptibility to known environmental risk factors, possibly via epigenetic change, is yet to be determined. As with other neurosensory conditions deterioration of the system predates the emergence of overt or notable symptoms, with evidence of a delay of as much as 10 years between the likely onset of pathology and hearing loss. Interventions for ARHL remain limited to hearing-aids and in a small, though growing number of cases, cochlear implants. The identification of traits, particularly those that make individuals susceptible to faster progressing age-related hearing loss, may open up opportunities for interventions to slow or halt progression. In a similar vein, bivariate analyses examining shared genetic factors between ARHL and co-morbidities such as dementia will allow identification of shared pathogenetic pathways and possible targets for therapeutic intervention. A clearer understanding of how ARHL and its risk factors relate to one another will be much needed as part of the drive to reduce the disability associated with hearing loss.

**Authors contributions**

HW produced the first draft of the review, which was modified by TN and FW.


27. Makishima, T. *et al.* Nonsyndromic hearing loss DFNA10 and a novel mutation of EYA4: Evidence for Peer Review.


