Unravelling the genetic relationships between auditory processing and speech and language

Lidiya Talbot

Degree awarded by

Oxford Brookes University

This thesis is submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy

January 2021

Acknowledgements

I would first like to thank my supervisor, Dr Dianne Newbury, for inspiring me and supporting me not only through my PhD journey, but also through my MSc at Oxford Brookes. Thank you, Dianne, for always being there for me, no matter how busy you were, thank you for your training in the lab (even if I did not spend long in there) and on the computer (where I spent most/all of my time), and for always trying to understand my vague explanations with very little detail. Thank you for encouraging me to attend so many training opportunities, conferences, and science festivals (I really enjoyed all of them), and for always finding the best in me.

Massive thanks to Dr Hayley Mountford, my Post-Doctoral Researcher, for all your endless help with bioinformatics and lab work (do you remember those tissue cultures experiments at the start?), for reading my numerous drafts and for encouraging me with all my outreach/public engagement ideas. You are the Best! Special thanks to Dr Fabiola Ceroni, for inspiring me and for always thinking of me, for keeping me sane with our long chats in the corridor and lots of coffee breaks. Thanks to Professor Nicky Ragge who so readily offered me a position with her incredible team; Dorine, Yesim, Richard and Fabiola, I am so happy to be back working with you after my MSc project.

A big thank you to my parents, who supported me and believed in me and who encouraged me to keep going even when I found it almost impossible. You have really been amazing, thank you for the weekend grandparent duties and for the “Baba and Dyado” summer of 2020, it has been invaluable. Thank you to my husband, David Talbot, who was always there for me during my writing struggles and who built me a desk so I could work from the bedroom. You have been a Super Daddy, thinking of all those toddler activities to entertain little Ivy while I was hiding in the bedroom trying my best to concentrate. We have been through so much during my PhD, purchasing a house, becoming parents, getting married and muddling through a pandemic, I hope one day I can support you in your career development, too (just please do not do a PhD!). Lastly, thank you, Ivy, for coming into my life and distracting me in the most wonderful way!

Abstract

Auditory processing disorder is a common developmental disorder affecting about 10% of children. It is characterised by poor perception of speech sounds, especially in background noise environments, despite normal hearing sensitivity, which can lead to poor performance in school with a negative impact on education and everyday life. Previous studies have shown that auditory processing skills have a substantial genetic component, however, it is not clear which genes or molecular mechanisms are involved. In this thesis three different genetic approaches are applied (monogenic, common disease-common variant and common disease-rare variant) to assess the effect of candidate genes on neurodevelopmental measures, including hearing and language phenotypes, in a population cohort (ALSPAC) of more than 14,000 children. To complement these analyses, a reverse phenotype to genotype approach is used, focussing on a surrogate measure of auditory processing difficulties in ALSPAC children, to identify potential high impact coding variants that may explain these difficulties.

Given previous work, these genetic investigations focus upon candidate genes related to Usher syndrome, a recessive disorder leading to hearing and vision loss resulting from dysfunctional neurosensory cells in the inner ear and retina (hair cells and photoreceptor cells respectively). Analyses indicate that there is no one single risk variant, but a complex mix of variation across Usher genes (such as *USH2A, PCDH15, CLRN1,* and *ADGRV1*) might explain some of the APD risk. The phenotype to genotype analysis across coding regions further shows that rare pathogenic variants with large effect in other genes (such as *GRHL3, DIAPH1, FAT4* and *IFT88*) can contribute to risk of APD in simplex cases.

These results provide insights into the genetic landscape underlying APD and offer candidate genes and variants for further investigation and validation. Furthermore, the results highlight allelic heterogeneity where multiple variants present in the same Usher gene (*USH2A*) can display different, but related hearing phenotypes. In a wider context, this study also highlights the viability of using related/surrogate phenotypes for genetic discovery in a large sample when deep phenotyping of APD is unavailable.

Publications arising from this work

Perrino, P.A., **Talbot, L.** ,Kirkland, R., Hill, A., Rendall, A. R., Mountford, H. S., Taylor, J., WGS500 Consortium, Buscarello, A. N., Lahiri, N., Saggar, A., Fitch, H. R. and Newbury D.F. (2020) 'Multi-level evidence of an allelic hierarchy of *USH2A* variants in hearing, auditory processing and speech/language outcomes', *Commun Biol,* 3(1), pp. 180. doi: 10.1038/s42003-020-0885-5

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAA	American Academy of Audiology
ABR	Auditory Brainstem Response
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
ADHD	Attention deficit hyperactivity disorder
ALC	Ankle link complex
ALSPAC	Avon Longitudinal Study of Parents and Children
APD	Auditory Processing Disorder
AR	Autosomal recessive
ARHL	Age-related hearing loss
ASD	Autism spectrum disorder
ASHA	American Speech-Language and Hearing Association
bp	Base pairs
BSA	British Society of Audiology
BZ	Burden-Zeggini test
CANS	Central Auditory Nervous System
CCC	Children's Communication Checklist
CI	Confidence interval
CNS	Central Nervous System
CNV	Copy number variant
dB	Decibels
dB HL	Decibels hearing level
dB SPL	Decibels sound pressure level
daPa	Dekapascals
DPOAE	Distortion Product Otoacoustic Emissions
DLD	Developmental Language Disorders
gnomAD	Genome Aggregation database
Hz	Hertz
HC	Hearing cells
HL	Hearing loss
HHL	Hidden hearing loss
GWAS	Genome-wide association studies

IHC	Inner hair cells
LD	Linkage Disequilibrium
MAF	Minor allele frequency
MET	Mechanoelectrical transducer channels
NGS	Next Generation Sequencing
NWR	Nonword repetition
OAE	Otoacoustic Emissions
OHC	Outer hair cells
OME	Otitis media with effusion
OMIM	Online Mendelian Inheritance in Man
RP	Retinitis pigmentosa
RR	Relative risk
SLD	Speech and language disorders
SNHL	Sensorineural hearing loss
sAPD	Suspected auditory processing disorder
SEM	Scanning electron microscope
SEN	Special education needs
SGN	spiral ganglion neurons
SNP	Single nucleotide polymorphism
USH	Usher Syndrome
UTLD	Upper tip link density region
Vcf	Variant call format
WES	Whole exome sequencing
WGS	Whole genome sequencing
WISC	Wechsler Intelligence Scale for Children

1. Introduction

Our understanding of auditory perception can be informed by the investigation of individuals in whom this process is disrupted. For example, in auditory processing disorder (APD). This term describes a deficit in the perception of speech sounds, especially in background noise, despite normal hearing and auditory function (Section 1.2).

The work described in this thesis was based on preliminary findings in a discovery family affected by an apparently autosomal dominant form of APD. The discovery family is briefly described in Section 2.1.1 and further described by Perrino *et al.* (Perrino *et al.,* 2020). The preliminary findings resulted from whole genome sequencing analysis (completed before the start of this work), which identified a stop-gain heterozygous variant in *USH2A*, which co-segregated with the disorder in the family (Perrino *et al.,* 2020). Perrino *et al.,* further showed that heterozygous *Ush2a* knockout mice (*Ush2a*$^{+/-}$) had a distinctive low-frequency (15Hz) hearing loss while their hearing at high-frequency (40Hz) remained intact (Perrino *et al.,* 2020). Even after allowance for this hearing loss, *Ush2a*$^{+/-}$ mice were impaired on complex pitch discrimination tasks, which involved the detection of a deviant target sound embedded in background distractors (Perrino *et al.,* 2020). Furthermore, analysis of ultrasonic vocalisations showed that *Ush2a*$^{+/-}$ mice vocalised at a significantly higher pitch and produced calls that were shorter and louder than wildtype mice (Perrino *et al.,* 2020). In contrast, the complete knockout mice (*Ush2a*$^{-/-}$) had high-frequency hearing loss (as expected) but showed superior performance on pitch discrimination tasks and did not have altered vocalisations (Perrino *et al.,* 2020). These animal studies provided a direct link between heterozygous *Ush2a* knockout and auditory perception and warranted further analyses.

This thesis therefore investigates the wider effects of *USH2A* variation (and the variation in other Usher syndrome genes) upon hearing, auditory processing and language in a large human population cohort. The sections that follow introduce the background concepts that later link to the presented data.

1.1. The human auditory system

1.1.1. Central auditory anatomy for information processing

The human auditory system includes the ear (peripheral auditory organ) and regions within the brainstem, midbrain, the thalamus and auditory cortex (forming the central auditory system), that receive signals propagated from the inner ear (Figure 1.1). The sound waves picked up by the ear are transduced into electrical signals and transmitted via the cochlear (auditory) nerve to the cochlear nucleus within the brainstem, where different sound cues are detected (Figure 1.1).

Most of the information is then transmitted through crossing fibres into the superior olivary complex (important in sound localisation) and from there up through the contralateral side of the brainstem and the inferior colliculus within the midbrain (important in integrating auditory input with other sensory inputs) (Figure 1.1). The information then passes through the medial geniculate nucleus in the thalamus (important in primary auditory information transmission) and reaches the auditory cortex where processing takes place. Processing includes complex tasks such as resolving multiple concurrent sound sources, attending to sounds in noisy environments, recognising auditory objects or contexts and performing higher-order auditory tasks such as language (Figure 1.1). This ascending pathway of auditory information is also known as the bottom-up or afferent route. Reciprocal connections from the cortex to subcortical structures all the way to the cochlea also exist, forming the top-down (descending or efferent) route (Figure 1.1). Independent tracks from the auditory cortex to the thalamus (medial geniculate nucleus), midbrain (inferior colliculus) and brainstem (superior olive) have been identified (Winer, 2006) (Figure 1.1). The efferent networks have been proposed to fine-tune afferent signal encoding and control gain in the system so important information can be extracted easily (Robinson and McAlpine, 2009). While the processing of auditory information is thought to take place along the central auditory pathway, transduction in hearing happens in the periphery, i.e. the ear, as explained below.

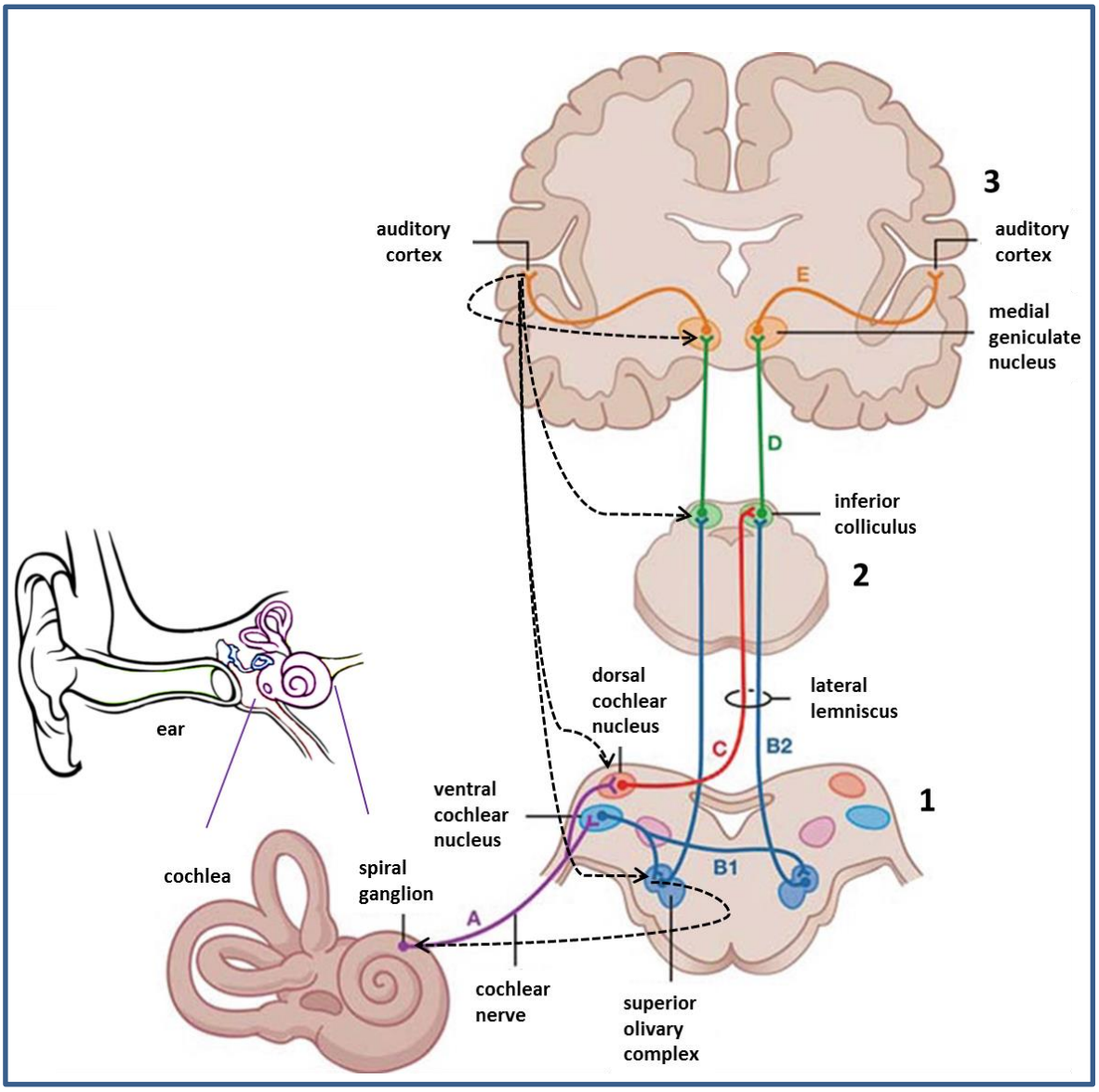

Figure 1.1. Diagram of the bilateral ascending auditory pathway from the cochlea and descending pathway from the auditory cortex (adapted from Hall, 2011).
The ascending pathway includes structures in the brainstem (1), midbrain (2) and auditory cortex (3). Unilateral auditory pathway involves: cochlear nerve (A), ventral cochlear nucleus to contralateral and ipsilateral superior olivary complex (B1), superior olivary complex to inferior colliculus (B2), dorsal cochlear nucleus to contralateral inferior colliculus (C), inferior colliculus to medial geniculate nucleus (D) within thalamus, and medial geniculate nucleus to auditory cortex (E). Descending pathway is illustrated in dotted black lines from auditory cortex back down to each of the auditory pathway structures (E back to A).

1.1.2. Peripheral auditory system and the process of hearing

The process of hearing starts with sound waves entering the outer (external) ear and travelling through the ear canal until they reach the ear drum, causing it to vibrate (Figure 1.2a). These vibrations are amplified within the middle ear and transmitted to the cochlea within the inner ear. The cochlea is a snail-like structure, filled with fluid with each turn consisting of three sections: scala vestibuli, scala tympani and scala media (Figure 1.2a). The middle section, scala media, contains the sensory auditory organ, the organ of Corti (Figure 1.2b-c). The organ of Corti comprises of a single row of inner hair cells (IHC) and three rows of outer hair cells (OHC) (Figure 1.2c-d). The hair cells form tight connections with supporting cells, which are in turn connected at their basal surface to an extracellular matrix, called the basilar membrane (Figure 1.2c). The apical surface of each hair cell contains the mechanically sensitive organelle, the hair bundle, which consists of dozens of "hairs", called stereocilia (Figure 1.2c-d). Another extracellular matrix, called the tectorial membrane, is attached to the stereociliary bundles of OHCs and covers the apical surface of the organ of Corti (Figure 1.2c). In contrast to the OHC bundle, the IHC bundle is free standing in the subtectorial space (Figure 1.2c). As the sound vibrations reach the organ of Corti, they form a travelling wave which causes displacement of the basilar membrane. This in turn leads to deflection (bending) of the stereocilia bundle and opening of mechanically gated cation channels (Figure 1.2d). The influx of calcium and potassium ions into the hair cell induces depolarisation and release of neurotransmitter glutamate at the basal end of the hair cell (Hudspeth, 2014). Glutamate stimulates the cochlear (auditory) nerve, which transmits the signal to the auditory cortex.

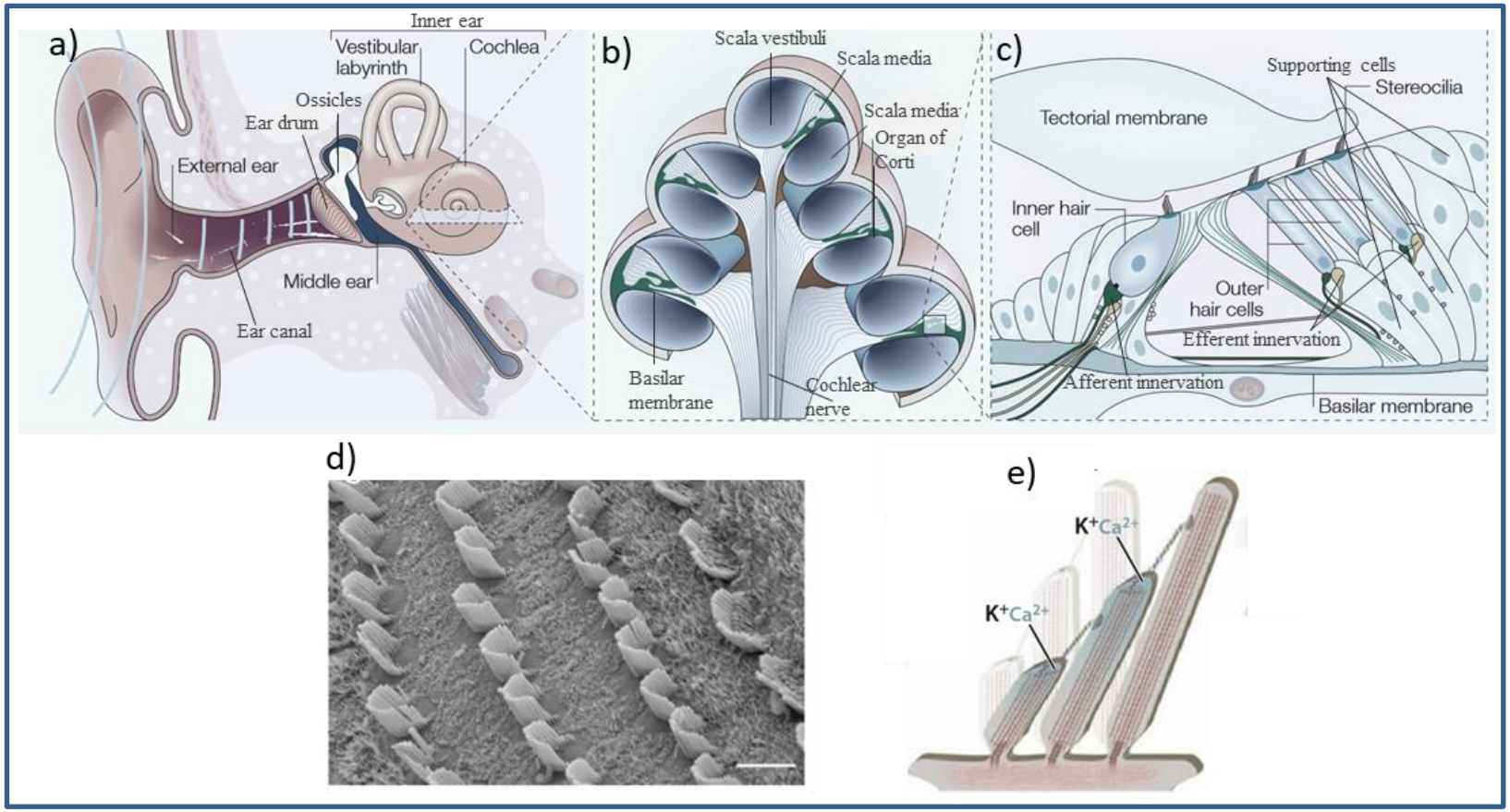

Figure 1.2. The mammalian ear and the structures involved in sound processing (adapted from (Frolenkov *et al.*, 2004).
a) External ear with the inner ear auditory organ; b) sections of the cochlea; c) the sensory cells of the organ of Corti: inner (IHC) and outer (OHC) hair cells with afferent terminals (in green) that form the ascending pathway and efferent terminals (in yellow) which exit the descending pathway ; d) a scanning electron microscopy picture of hairs cells with three rows of OHCs (left) and one row of IHCs (right); e) the mechanosensitive organelle of IHCs, the hair bundle, with Ca2+ and K+ influx through the mechanoelectrical transduction channels of the stereocilia upon hair bundle deflection.

There are approximately 16,000 sensory hair cells in the human cochlea (Schwander, Kachar and Müller, 2010). Due to variation in width and thickness of the basilar membrane and variation in height of the stereocilia, hair cells at different positions along the cochlear duct are tuned to different frequencies, described as tonotopic organisation (Mann and Kelley, 2011) (Figure 1.3a). Hair cells at the base of the duct (nearest to the tympanic middle ear) are sensitive to high frequencies, while those at the apical end of the coiled cochlea are sensitive to low-frequencies (Hudspeth, 1997) (Figure 1.3a). The same orderly map of sound frequency is functionally represented throughout the auditory system all the way to the auditory cortex (Figure 1.3b).

The IHC and OHC are specialised to do different things. The IHC (approximately 3,500 in the human cochlea) are the primary receptor cells, innervated by dendrites of the auditory nerve and so they form the majority of the afferent links to the central auditory system (from the ear to the brain). As such they are responsible for the transduction of sound waves into electrical signals and so can be perceived as the cells that do the "hearing" process. The OHC (11,000 in human cochlea) are the target of efferent neural pathway from the brainstem through the superior olivary complex's olivocochlear bundle (OCB). The OCB contacts OHCs directly and the IHCs indirectly via the afferent fibres beneath them (type I afferent fibres) (Figure 1.2c). Therefore, the OCB exerts fine-tuning right at the periphery. The OHCs are primarily involved in sound amplification and are considered as "amplifiers", interpreting auditory feedback to balance input between ears and facilitate auditory perception (Murakoshi, Suzuki and Wada, 2015). Both types of hair cells are exposed to a variety of external and internal ototoxic factors, such as loud noise, ototoxic drugs, ageing and genetic defects, which can cause damage. As mammalian hair cells do not regenerate, the damage can build up over time and lead to decreased hearing ability and even hearing loss, known as peripheral (sensorineural) hearing loss. If the IHC are damaged, that is expected to directly disrupt hearing through the afferent pathway. If the OHC are damaged, that will affect the ability to amplify important signals, controlled through the efferent pathway, which is essential for understanding complex information such as speech in noisy environments. Yet, the afferent and efferent pathways are connected and processes in one can influence the output in the other. For example, a study where IHCs in chinchillas were preferentially damaged with carboplatin (an anti-cancerous drug), showed that in spite of the fact that the chemical damage was purely peripheral (mainly as a result of >80% IHCs loss), a cascade of neuroplastic changes in the central auditory pathway could compensate for the reduced neural output from a damaged cochlea (Salvi *et al.*, 2016). Salvi *et al.* reported that the chinchilla's hearing in noise was affected, but it was largely preserved in quiet. This was because the activity from the few remaining intact IHCs was progressively amplified through the central auditory system so weak signals become comfortably loud (Salvi *et al.*, 2016). The difficulty hearing in noise was proposed to arise from the central auditory system not being able to compensate for the reduced neural output in the

cochlea (by turning up its gain and decreasing its inhibitory effect which worked for weak sounds in quiet, but not in adverse conditions with competing noise) (Salvi *et al.*, 2016). Indeed, the efferent control of OHCs has been shown to be important in speech discrimination in noise (Winslow and Sachs, 1987; Kawase and Liberman, 1993) and sound localisation (Andeol *et al.*, 2011) and as such is an important pathway studied in disorders of auditory processing (Reynard, Veuillet and Thai-Van, 2020; Boothalingam *et al.*, 2015). Furthermore, the function of OHCs is thought to also directly contribute to hearing in noise by fine tuning the motion of the basilar membrane (Parker, 2020). As such OHCs dysfunction has been shown to correlate with better performance in quiet, but poor performance in noise and is regarded as an important cause for hidden hearing loss (HHL- explained in Section 1.2) (Hoben *et al.*, 2017; Parker, 2020).

Auditory nerve (AN) fibers transmit signals from IHCs to the brainstem targets in the cochlear nucleus. The multiple innervation of single IHCs, whereby each IHC is contacted by 10-30 AN fibers, is important in auditory processing and understanding speech in noise because the AN fibers differ in spontaneous discharge rate (SR) and threshold to acoustic stimuli (Liberman, 1978). The low-SR fibers (<20 sp/s) have higher thresholds and increased dynamic range of the auditory periphery compared the high-SR fibers (>20 sp/s) (Liberman, 1978). The low-SR fibers are thought to be important for hearing in noisy environments because of their resistance to masking by continuous background noise (or because of their ability to follow the quick change of the amplitude of acoustic signals) (Costalupes, Young and Gibson, 1984). In contrast, the high-SR fibers are responsible for the sensitivity to quiet sounds and are saturated by high-level background noise. Kujawa and Liberman showed that acoustic overexposure to intense sound (8 to 16 kHz octave band at 100 dB SPL for 2h) can cause a permanent damage of auditory nerve fibers in mice, but without damaging cochlear hair cells and despite a complete recovery of cochlear thresholds (Kujawa and Liberman, 2009), as measured by ABR. It was further confirmed this was likely due to the selective loss of low-SR fibers following exposure to neuropathic noise (4 to 8 kHz octave band at 106 dB SPL for 2h) in guinea pigs, which could explain the recovery of ABR thresholds despite significant noise-induced neuropathy (Furman, Kujawa and Liberman, 2013). It was therefore suggested that selective loss of low-SR high threshold AN fibers may be another contributor to problems of hearing in noise in humans and therefore hidden hearing loss (Liberman *et al.*, 2016) and APD (Section 1.2).

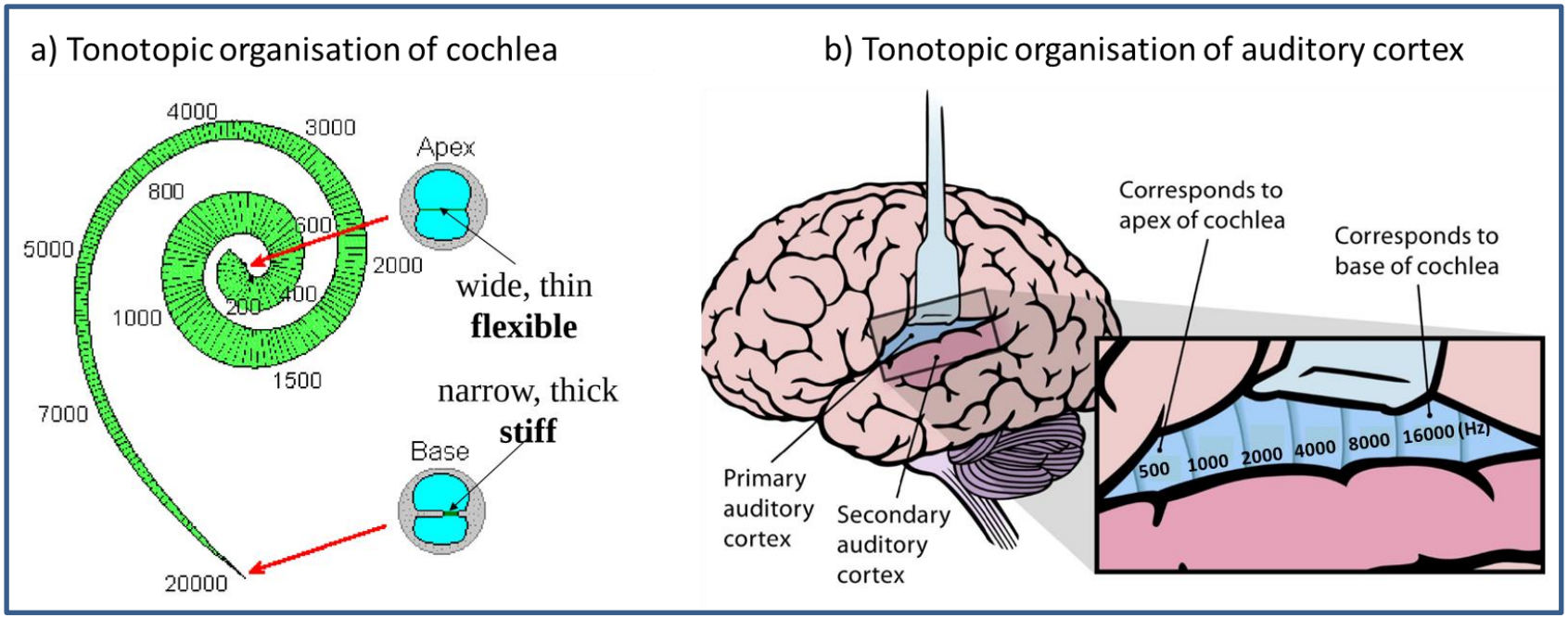

Figure 1.3. Tonotopic organisation in the auditory system from cochlea to auditory cortex (adapted from Purves *et al.,* 2001).

1.1.3. Evaluating hearing

Amongst the tests used to assess hearing and diagnose hearing loss, the ones I will review and later refer to in this thesis are pure tone audiometry, Otoacoustic Emissions (OAE) and Auditory Brainstem Response (ABR).

Pure tone audiometry (also referred to as threshold audiometry) determines the quietest (softest) sound that a person can hear at different frequencies from 250 to 8000Hz, which span most of the human audible range. It is also called air conduction testing as the sound goes through the outer and middle ear to reach the inner ear. This is usually the first-line test to evaluate hearing deficits. The results are recorded on an audiogram (Figure 1.4). Sound frequency (ranging from low to high pitch) is measured in Hertz (Hz) and is recorded on the audiogram's horizontal axis (Figure 1.4). Sound intensity (hearing level) is measured in decibels hearing level (dB HL) and is recorded on the vertical axis (Figure 1.4). The quietest sound at each tested frequency is called air conduction threshold and is recorded as a point on the audiogram for the left and right ear separately. Any points that are heard at 20dB or quieter are considered to be within the normal range. Pure tone audiometry is a key measure in the evaluation of hearing disorders as it is able to provide information regarding the type, degree and configuration of hearing loss, however, it is considered to offer limited insight into auditory function and auditory processing in real world settings (such as speech, music, noisy environments) (Musiek *et al.*, 2017).

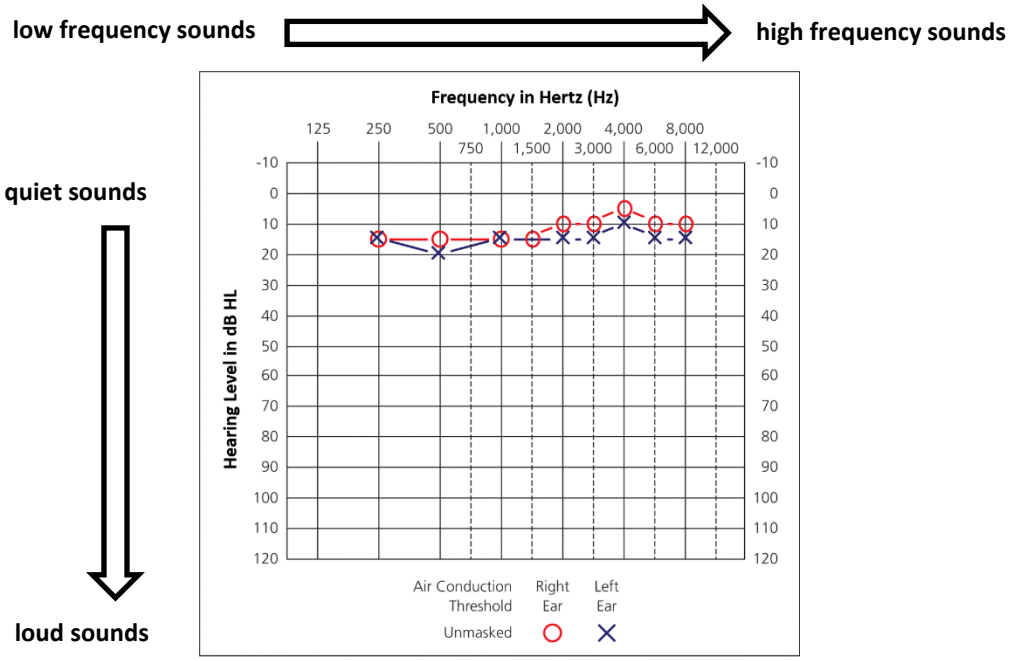

Figure 1.4. Audiogram showing hearing with normal hearing thresholds for both right and left ear (ANSI- American National Standards Institute).

Additional electrophysiological tests, such as OAE and ABR, are used to test the cochlea and the brain pathways (such as auditory nerve and brainstem pathways) and therefore provide

diagnostic insight into specific disorders of the auditory system. Both tests are particularly useful for assessment of infants and young children (Stanton *et al.*, 2005). The otoacoustic emissions are generated in the cochlea and are measurable in the ear canal, so they are an objective measure of the cochlea functioning. Distortion Product Otoacoustic Emissions (DPOAE) are distorted sounds generated by the OHCs in the cochlea in response to two tones that are close in frequency. The presence of a DPOAE response is typically an indication that the OHCs specifically are functioning properly, while the IHCs abnormal function does not seem to affect the DPOAE input/output (Salvi *et al.*, 2016). DPOAE is a common hearing test used in mouse models to provide additional information about OHCs function.

ABR measures the electrical activity of the auditory nerve pathway from the inner ear to the brainstem that is recorded by electrodes. The ABR response is displayed as a waveform with five major peaks used to assess brainstem function at different levels of the auditory pathway. Through the different peaks ABR shows the activity of the auditory nerve and the activity of neurons in successive nuclei of the auditory hindbrain within the brainstem. The time between the peaks (latencies) and the amplitude of the peaks are measured and compared to normative data. The reliability, sensitivity, non-invasive nature and ease of application has made ABR a method of first choice to assess hearing impairment in mouse models (Zheng, Johnson and Erway, 1999).

1.2. Disorders of the auditory system: Hidden Hearing Loss and Auditory Processing Disorder

The audiogram as the standard method to evaluate hearing (Figure 1.4) is sensitive at identifying hearing loss but is unable to detect “hidden” cochlear impairments such as cochlear synaptopathy (loss of synaptic contact between auditory nerve and IHCs) or hair cell dysfunction (Liberman *et al.*, 2016; Chen, 2018). Cochlear impairment with normal audiometric thresholds is defined as hidden hearing loss (HHL) (Schaette and McAlpine, 2011) and is thought to result from noise-induced synaptopathy, cochlear demyelination or possibly hair cell dysfunction (Kujawa and Liberman, 2009; Wan and Corfas, 2017; Hoben *et al.*, 2017). These forms of hidden cochlear impairment do not result in overt hearing loss (Chen, 2018), but they may increase the susceptibility of the cochlea to further damage and weaken or disturb central auditory processing such as sound discrimination (D *et al.*, 2020). Due to this secondary impact of HHL on central auditory pathway functioning, the term auditory processing disorder is sometimes preferred in order to include those individuals with peripheral auditory impairment (Iliadou *et al.*, 2017).

1.2.1. APD definition and prevalence

Auditory Processing Disorder (APD) is a highly heterogeneous neurodevelopmental disorder. It is included in the International Classification of Diseases by the World Health Organization, 10th edition (ICD-10) as H93.25: "a disorder characterised by impairment of the auditory processing, resulting in deficiencies in the recognition and interpretation of sounds by the brain". According to the British Society of Audiology (BSA), APD is considered as "poor perception of speech and non-speech sounds" (BSA, 2018). Affected individuals report listening difficulties, despite in most cases having a normal hearing sensitivity (as detected on an audiogram- Figure 1.4) (Bamiou, Musiek and Luxon, 2001). The BSA proposes three types of APD: developmental, acquired and secondary (BSA, 2018) (Table 1.1), and highlights the international focus towards developmental APD, because of its unknown aetiology and its potential impact on learning difficulties, which can in turn affect school performance.

Table 1.1. APD types as recommended by the British Society of Audiology

Developmental APD	There is usually no known cause other than a family history of developmental communication and related disorders, it is present in childhood and may continue into adulthood.
Acquired APD	Associated with ageing or a known neurological event (brain lesion, trauma, stroke, infection).
Secondary APD	Occurs together with, or as a result of either short term (for example glue ear) or permanent hearing impairment.

The hallmark of APD is difficulty understanding speech when in noisy environments. Affected individuals require frequent repetition of information, they are easily distracted and struggle to follow instructions, and they find it difficult to interact in noisy group activities (Moore *et al.*, 2013; Jerger and Musiek, 2000). These difficulties can consequently impact on school achievements and social skills (Moore *et al.*, 2013). This collection of symptoms can be traced back to a different auditory deficit (Table 1.2), but they are also present in co-occurring neurodevelopmental disorders such as attention deficit hyperactivity disorder (ADHD), language impairment, autism spectrum disorder (ASD) and dyslexia (Witton, 2010; Ferguson *et al.*, 2011; Moore and Hunter, 2013; Dawes *et al.*, 2008; Dawes and Bishop, 2010). Effective processing of sounds is important for the development of language and this together with the likely cognitive elements of APD may explain some of the developmental overlaps. Several causal models have been proposed to explain the co-occurrence between APD and language disorders. The earliest theory was developed by Tallal and supports the idea that language disorders are caused by APD due to a deficit in rapid auditory temporal processing (processing sounds when closely spaced in time) (Tallal, 2004). Later studies concluded the opposite relationship: deficits in auditory processing are a consequence of language impairment with processing of sounds within the brain being affected by poor language skills (Shafer, Schwartz and Martin, 2011; Bishop, Hardiman and

Barry, 2012). An alternative model argues that a number of genetic and environmental factors interact with other risk factors to determine whether an individual will develop a language disorder (Bishop, 2006). According to this theory, auditory processing deficits may constitute one of these risk factors and thus contribute to language difficulties in the absence of a one-to-one relationship between the two. Most recently, data from a correlation study investigating the links between language and auditory processing tasks in children with mild to moderate sensorineural hearing loss, ruled out the extreme models of direct relationships (Halliday, Tuomainen and Rosen, 2017). Instead, the results suggested that deficits on auditory processing tasks which required higher level cognitive ability were associated with language difficulties, but were not sufficient for causality, implying that the relationships were not as straight-forward as hypothesised before and were likely influenced by other factors (genetic and environment).

APD is relatively common with estimates ranging from between 0.5%-1% in the general population to 10% when APD occurs in combination with other neurodevelopmental disorders (Brewer *et al.*, 2016; Hind *et al.*, 2011). APD impacts on both school performance and everyday life, predominantly through a reduced ability to listen and respond to speech and sounds appropriately (Moore *et al.*, 2013).

Table 1.2. Auditory deficits and their implications on day-to-day performance and behaviour, characteristic of APD (adapted from (Bamiou, Musiek and Luxon, 2001).

Auditory deficits in:	**Resulting symptoms and behaviours:**
Sound localisation (ability to pinpoint the source and location of a sound)	Poor performance in confusing environments with competing speech and noise
Auditory pattern recognition (ability to determine similarities and differences in patterns on sounds)	Difficulties following oral instructions
Auditory discrimination (ability to distinguish between different sounds)	Language, reading, and spelling difficulties
Temporal processing (the ability to process multiple auditory stimuli in their order of occurrence)	
Processing degraded auditory signals (ability to perceive a signal in which some of the information is missing)	Difficulties with rapid speech and with filling in missing portion of the auditory signal to recognise the whole message
Processing the auditory signal when embedded in competing acoustic signals (ability to perceive speech and other sounds when another signal is present)	Difficulties in background noise
	Inattention
	Distractibility
	Academic difficulties

1.2.2. Diagnosis of APD and hypotheses about its origin

There is no single test that can diagnose APD. Instead, the diagnosis is typically based on evidence from a multi- and inter-disciplinary team. These include concerns from parents and teachers, reports from medical professionals (such as GPs, speech and language therapists, audiologists, psychologists), medical and family history, observation of speech quality, examination of

peripheral auditory function and cochlear feedback pathway, and results from auditory processing (AP) specific tests (forming an APD test battery). Although individual elements of the APD battery are standardised, no agreement has been reached on the exact tests or the number of tests that should comprise the battery, resulting in no "gold standard" (Moore *et al.*, 2013; Dillon *et al.*, 2012). APD evaluation is described as a dynamic assessment, where specific AP tests are selected based on the individual's age, observational data, medical history and their problems and presentations, including any other issues and co-occurring conditions (Campbell *et al.,* 2019). Four broad areas of auditory processing are generally measured in the APD battery, including dichotic processing, temporal processing, perception of monoaural low redundancy speech and binaural interaction (Campbell *et al.,* 2019) (Table 1.3).

Table 1.3. Main auditory processing areas tested through the APD battery with examples of tests used.

Main auditory processing areas tested:	Examples of tests
Dichotic processing (measures the ability to process information when a different stimulus is presented to each ear simultaneously)	Dichotic Digit Test (DDT) Competing Sentence Test
Temporal processing (assesses the ability to process nonverbal auditory signals and to recognise order or patterns)	Frequency Pattern Test (FPT) Duration Pattern test (DPT)
Perception of monoaural low redundancy speech (tests whether each ear can independently recognise distorted words)	Filtered Speech Test
Binaural interaction (tests integration of information using both ears)	Masking Level Difference (MLD)

Concerns have been expressed regarding the suitability of the tests included in the APD battery (Dillon *et al.*, 2012; Moore, 2018), as the tests were initially developed in the 1970s and 1980s for the detection of brain tumours and cortical lesions in adults (Berlin, 1976; Musiek and Geurkink, 1980). These AP tests typically carry a high cognitive load and therefore also measure language, attention and memory skills indirectly. As a result, a child with a language development problem might score poorly on the AP tests because of their poor language skills and not because of their poor auditory processing skills. Furthermore, the recommendations regarding what threshold to use to interpret APD-battery tests to reach a diagnosis also differ between the three auditory bodies: American Speech-Language and Hearing Association (ASHA), American Academy of Audiology (AAA) and BSA. The most commonly used diagnostic criteria is performance at or below 2 standard deviations of the mean on at least two tests from the APD battery, or 3 standard deviations below the mean on one AP test (ASHA, 2005). The AAA recommends that these are met for at least one ear (AAA, 2010), while BSA suggests that at least one of the diagnostic tests should be non-speech stimuli so it does not depend on language skills (BSA, 2013). The lack of consensus has been highlighted by Wilson and Arnott (2013), who compared the diagnostic rates of APD in 150 children by using recommendations from BSA, AAA and ASHA and selected

researchers in the APD field (Wilson and Arnott, 2013). The authors found that 7.3% of the children studied were diagnosed with APD under the strictest criteria compared to 96% using the most lenient criteria (when using the same traditional APD tests) (Wilson and Arnott, 2013). This further emphasised the heterogeneity of the disorder and the need for any diagnosis to be qualified by an explicit statement of the criteria used.

The conventional hypothesis about the origin of APD is that it results from impaired bottom-up sensory processing within the central auditory system, which may involve lesions in CANS or functional impairment of basic auditory processing (Cacace and McFarland, 2005). The cochlea at the periphery is usually not included in this hypothesis, however, it should not be overlooked as considerable encoding of sound stimuli is carried out in the cochlea and transmitted via the auditory nerve and so synaptic damage preferentially involving low-SR auditory neurons after noise damage, is thought to explain problems with understanding speech in background noise (Furman, Kujawa and Liberman, 2013). Moreover, impaired hair cell function and resulting changes in cochlear compression can influence spectral and temporal tuning to some extent independently of pure tone sensitivity (Oxenham and Bacon, 2003). Other researchers argue that APD is a multi-modal deficit or even entirely cognitive, incorporating higher order functions, such as cognition, attention and language, exerting non-specific effect on perception (Musiek, Bellis and Chermak, 2005; Moore *et al.*, 2010).

1.2.3. Genetic studies

Twin studies have shown that both speech and non-speech based auditory processing skills have a substantial genetic component (Morell *et al.*, 2007; Brewer *et al.*, 2016). Identifying candidate genes specific to APD has proven difficult not only due to the diagnostic controversies (highlighted in Section 1.2.2), but also due to its frequent co-occurrence with other neurodevelopmental conditions, making it challenging to disentangle genetic relationships. Nevertheless, mouse models of disrupted genes, known to play a role in other neurodevelopmental disorders such as dyslexia, Developmental Language Disorders (DLD) and ASD, have given an insight into candidate genes for auditory processing deficits. For example, Guidi *et al*. identified an impairment in behavioural gap-in-noise detection task in double knock-out of the dyslexia susceptibility mouse gene *Kiaa0319* and its homologous gene *Kiaa0319L*, indicating a deficit in the auditory system (Guidi *et al.*, 2017). Felix *et al*., studied the contribution of *Chrna7* to auditory processing in knock-out α_7-nAChR mice, showing delays of evoked ABR responses, impaired forward masking and impaired gap detection (Felix *et al.*, 2019). Human *CHRNA7* has been further associated with reading and language skills and SLI in genome-wide association studies (GWAS) (Pettigrew *et al.*, 2015; Simpson *et al.*, 2015; Gialluisi *et al.*, 2016).

In two studies *Cntnap2* mutant mice showed reduced vocalisations and an impaired temporal auditory processing (Peñagarikano *et al.*, 2011; Truong *et al.*, 2015). In humans homozygous mutations of *CNTNAP2* lead to severe disease, characterised by profound intellectual disability, epilepsy, language difficulties and autistic traits (Strauss *et al.*, 2006; Rodenas-Cuadrado *et al.*, 2016), while multiple studies have identified other variants to be associated with increased risk for ASD and language-related disorders (Vernes *et al.*, 2008; Alarcón *et al.*, 2008; Arking *et al.*, 2008). Scott *et al*. showed that a new *Cntnap2* knock-out rat model had typical hearing thresholds but reduced auditory evoked neural responsivity and slowed signal transmission in juvenile animals (Scott *et al.*, 2018). By adulthood the disruptions in auditory signal processing mostly disappeared, indicating a delayed maturation of auditory processing pathways (Scott *et al.*, 2018). However, disruptions in brainstem-mediated auditory evoked behaviour persisted in adulthood, suggested that early developmental disruptions in sensory processing can cause permanent defects in circuitries responsible for auditory reactivity (Scott *et al.*, 2018).

A further insight into potential genetic regions and candidates for APD can also be gained from family genetic studies, with only one study in the literature and no GWAS or sequencing studies of APD available. Addis *et al.* investigated auditory processing as a core deficit of language impairment by performing genome-wide linkage analysis on a three-generation German family (NE family) (Addis *et al.*, 2010). The results suggested that a gene within the central region of chromosome 12 is likely linked to the auditory processing difficulties (Addis *et al.*, 2010). Although a potential "causal" variant was not identified with the techniques available at the time, the study proposed that APD could follow a simple inheritance pattern, where a damaging variation within one gene can have a direct influence on auditory processing difficulties.

1.3. Disorders of syndromic hearing loss and the example of Usher Syndrome

1.3.1. Definition and prevalence

Usher syndrome (USH) is an autosomal recessive genetic disorder that causes hearing and vision loss and occasional balance problems (Keats and Corey, 1999). Although considered collectively rare with an estimated prevalence of between 4 and 17 in 100,000 people worldwide (Kimberling *et al.*, 2010; Boughman, Vernon and Shaver, 1983), it is the most common cause of combined hearing and vision loss, accounting for approximately 50% of all deaf-blindness cases (Keats and Corey, 1999).

1.3.2. Usher Syndrome Types

Based on the clinical symptoms, relating to severity and age of onset, and the presence or absence of balance problems, USH patients are classified into clinical types. Types I to III are the more well described ones, with both a newer type IV and atypical cases also reported (Table 1.4).

1.3.2.1. Usher Syndrome Type I

Usher syndrome type I (USH1) is the most severe of the three USH types. It is characterised by severe to profound deafness from birth affecting both ears, known as congenital bilateral sensorineural hearing loss (SNHL- affecting the inner ear) (Lentz and Keats, 1993b) (Table 1.4). Affected children who receive cochlear implants at an early age (within the first 2 years of life) can usually communicate using speech and lip-reading (Lentz and Keats, 1993b). Poor balance from birth, known as vestibular areflexia, is associated with the deafness and causes delays in sitting and walking (Lentz and Keats, 1993b) (Table 1.4). Loss of side vision and night blindness are early signs of retinitis pigmentosa (RP) which can manifest before the age of 10 (Lentz and Keats, 1993b) (Table 1.4). RP is a progressive, symmetric degeneration of the retina, affecting both eyes, which starts at the periphery, mainly disturbing the function of the photoreceptive cells active in dark-adapted state (rod cells).

1.3.2.2. Usher Syndrome Type II

Usher syndrome type II (USH2) is the most common form of the disorder, representing around half of all cases (Reiners *et al.*, 2006). It is less severe than type I and causes a moderate bilateral SNHL from birth and RP that may not become apparent until adulthood (Lentz and Keats, 1993a) (Table 1.4). The hearing loss is mild to moderate in the low frequencies (for the sounds that are low-pitched) and severe to profound in the higher frequencies (sounds that are high pitched) (Lentz and Keats, 1993a) (Table 1.4). The rate and degree of hearing loss is variable among and within families with progressive PR and visual symptoms diagnosed around the third decade of life (Blanco-Kelly *et al.*, 2015) (Table 1.4). Balance is not affected and so walking and sitting is developed at typical age (Table 1.4). Children affected by USH2 benefit from conventional hearing aids and often have close to normal speech acquisition. With progression of hearing loss, cochlear implants have shown to increase speech intelligibility, quality of life and communication in later life (Hartel *et al.*, 2017).

1.3.2.3. Usher Syndrome Type III

Usher syndrome type III (USH3) is the mildest and rarest form in the UK, but is particularly prevalent in Finland and among Ashkenazi Jewish people (Pakarinen *et al.*, 1995; Ness *et al.*, 2003). Children usually have normal hearing and vision at birth and develop SNHL after

development of speech (usually detected in the first decade of life) and RP in adulthood (Keats and Corey, 1999) (Table 1.4). Hearing loss is of progressive nature with high frequencies affected first and more severely around the first decade of life and patients still showing some residual hearing at the lower frequencies at 40 years or older (Sadeghi *et al.*, 2005). Balance problems are experienced by some with most patients reporting a normal age of independent walking (Sadeghi *et al.*, 2005). Hearing aids can be of benefit early in the course of disease while cochlear implants have been found beneficial later on in life as hearing loss progresses (Pietola *et al.*, 2012)

1.3.2.4. Usher Syndrome Type IV and atypical Usher Syndrome

Although clinically divided into three types, atypical cases having incompatible phenotypes with the three established types are also described in the literature. For example, Khateb *et al.* described a family with an atypical form of USH (later designated as USH4), inherited in the typical autosomal recessive manner, but presenting with a distinctive retinal degeneration phenotype (resulting in ring scotoma) and moderate to severe SNHL, both with a late onset (around the age of 40), without vestibular involvement (Khateb *et al.*, 2018) (Table 1.4). The cause was a homozygous mutation in *ARSG* (Table 1.4) (further described in Section 1.3.3.2), not previously associated with Usher Syndrome or any other disorder.

Variants in well characterised USH genes (*MYO7A, USH2A, CDH23 and SARS*) have also been associated with atypical clinical presentations of milder symptoms. In two consecutive studies Liu *et al*. described two siblings with bilateral progressive hearing loss and mild RP and another four unrelated individuals with progressive SNHL, RP, and variable vestibular function (vestibular dysfunction was in one individual only), symptoms most closely related to USH3 (Liu *et al.*, 1998; Liu *et al.*, 1999). The siblings in the first study were compound heterozygous for pathogenic variants in *MYO7A*, typically related to syndromic USH1B (Liu *et al.*, 1998). The four individuals in the second study had a pathogenic variant in *USH2A* (homozygous in three individuals and heterozygous in one individual), typically associated with USH2 (Liu *et al.*, 1999) and not USH3 as was expected by the mild symptoms in the two studies. Atypically milder phenotypes have also been described in cases with pathogenic variants in *CDH23* (homozygous or compound heterozygous splice-site) and *SANS* (homozygous deletion and homozygous missense), expected to cause severe USH1 symptoms, but presenting with no obvious vestibular dysfunction, mild retinal symptoms and moderate to severe hearing loss, resembling an USH2 diagnosis (Bashir, Fatima and Naz, 2010; Kalay *et al.*, 2005; Bork *et al.*, 2001; Valero *et al.*, 2019). These atypical cases underlie the genetic and phenotypic heterogeneity observed in USH.

Table 1.4. Usher Syndrome types, symptoms, genes and proteins.

USH type (phenotype MIM number)		**Subtype**	**Gene name** (gene MIM number)	**Protein name**	**Protein Function**	**Hearing loss**	**Retinitis pigmentosa**	**Vestibular function**	**References**
USH1	(# 276900)	USH1B	*MYO7A* (* 276903)	myosin VIIa	actin-based motor protein	Congenital, severe to profound	Prepubertal onset; average age of diagnosis in second decade	Bilateral areflexia; motor development may be delayed	(Weil *et al.*, 1995)
	(# 276904)	USH1C	*USH1C* (* 605242)	harmonin	PDZ scaffold protein				(Verpy *et al.*, 2000)
	(# 601067)	USH1D	*CDH23* (* 605516)	cadherin 23	cell adhesion				(Bolz *et al.*, 2001)
	(# 602083)	USH1F	*PCDH15* (* 605514)	protocadherin 15	cell adhesion				(Ahmed *et al.*, 2001)
	(# 606943)	USH1G	*USH1G/SANS* (* 606943)	SANS	scaffold protein				(Weil *et al.*, 2003)
	(# 614869)	USH1J	*CIB2* (* 605564)	CIB2	calcium and integrin binding				(Riazuddin *et al.*, 2012)
USH2	(# 276901)	USH2A	*USH2A* (* 608400)	usherin	Cell adhesion & signalling	Congenital moderate to severe; high frequencies most affected	Onset in second decade; average age of diagnosis in third decade	Normal	(Eudy *et al.*, 1998)
	(# 605472)	USH2C	*ADGRV1/GPR98* (* 602851)	GPR98/VLGR1	adhesion G-protein coupled receptor VI, signalling				(Weston *et al.*, 2004)
	(# 601067)	USH2D	*WHRN/ DFNB31* (* 607928)	whirlin	PDZ scaffold protein				(Ebermann *et al.*, 2007)
USH3	(# 276902)	USH3A	*CLRN1* (* 606397)	clarin-1	scaffolding and cellular trafficking	Post-lingual and progressive, variable	Variable onset, typically in second decade	Variable	(Adato *et al.*, 2002)
	(# 614504)	Proposed USH3B	*HARS* (* 142810)	histidyl tRNA synthetase	synthesis of histidyl-transfer RNA	Post-lingual	Variable	Normal	(Puffenberger *et al.*, 2012)
	(# 612674)	Clinical USH3	*ABHD12* (* 613599)	alpha/beta hydrolase domain containing 12	catalyses the hydrolysis of 2-arachidonoyl glycerol	Late onset	Late onset	Normal	(Li *et al.*, 2019)
USH4	(# 618144)	-	*ARSG* (* 610008)	Arylsulfatase G	hydrolyse sulfatase esters, located in lysosome	Progressive moderate to severe	Late onset retinal degeneration	Normal	(Khateb *et al.*, 2018; Abad-Morales *et al.*, 2020)
Digenic cases of USH2 (# 605472)		USH2C digenic	*PDZD7* (* 612971)	PDZ domain-containing protein 7	ciliary protein	Typical for USH2	Earlier onset of RP	Normal	(Ebermann *et al.*, 2010)
Atypical		-	*CEP250* (* 609689)	CEP2	centrosome cohesion during interphase	Progressive	Mild RP	Normal	(Khateb *et al.*, 2014; Fuster-Garcia *et al.*, 2018)
Atypical		-	*C2orf71/PCARE* (* 613425)	Photoreceptor cilium actin regulator	photoreceptor cell maintenance and vision	Normal	AR RP	Normal	(Khateb *et al.*, 2014)

1.3.3. Implicated genes

Usher Syndrome is a not only clinically, but also genetically heterogeneous and to date 16 genes have been associated with it (Table 1.4). The numbering system in the Usher Syndrome classification (USH1, USH2 and USH3) corresponds to the associated severity of the clinical presentation. Each type is split into subtypes, labelled with letters, which indicate the molecular subtype (Table 1.4).

Ten of the USH genes are classed as causative of typical Usher syndrome: *MYO7A*, *USH1C*, *CDH23*, *PCDH15*, *USH1G*, *CIB2* causing USH1; *USH2A*, *ADGRV1*, *WHRN* causing USH2 and *CLRN1* causing USH3 (Table 1.4). The remaining genes can be classed as USH-related or atypical as the symptoms of affected individuals do not fully match the three typical clinical types of USH (explained in Section 1.3.2) or the genes have only been detected in a very small number of individuals: *PDZD7*, *CEP250*, *C2orf72*, *ARSG*, *HARS* and *ABHD12* (Table 1.4). To date all pathogenic variants within the causative genes of typical USH have been reported as homozygous or compound heterozygous (inherited in an autosomal recessive model).

1.3.3.1. Usher Syndrome causative genes

1.3.3.1.1. USH1 causative genes

USH1B (# 276900), caused by mutations in *MYO7A,* is the most common type of USH1, accounting for 55-75% of USH1 cases and about 21% of all USH cases (Le Quesne Stabej *et al.*, 2012; Roux *et al.*, 2006) (Figure 1.5). *MYO7A* also harbours mutations causing both recessive (DFNB2) and dominant (DFNA11) nonsyndromic deafness and atypical USH (as discussed in Section 1.3.2.4) (Liu *et al.*, 1998; Liu *et al.*, 1997c; Liu *et al.*, 1997a; Weil *et al.*, 1997). Pathogenic *MYO7A* variants appear to be spread across the whole gene sequence, with a cluster forming at the motor head protein domain and another cluster in the tail (Liu *et al.*, 1998). With the exception of the DFNA11-causing mutation c.2662_2670del (p.Lys888_Lys890del), affecting the coiled-coil region, there appears to be no obvious correlation between mutations and the resulting phenotype. Heterozygous mutations in the coiled-coil region, responsible for the dimerisation of the protein, have been proposed to result in dominant hearing loss because of a dominant-negative effect (Liu *et al.*, 1998), with affected individuals showing less severe post-lingual hearing loss compared to USH2B and DFNB2 (Liu *et al.*, 1997b). The pathogenic variants c.93C>A (p.Cys31Ter) is a major recurring *MYO7A* variant causing USH1B, which has been described as a founder mutation, accounting for 33% of all USH1 cases in Denmark (Dad *et al.*, 2016).

Pathogenic variants in *CDH23*, *PCDH15* and *USH1C* are responsible for USH1D (# 601067), USH1F (# 602083) and USH1C (# 276904) respectively (Oshima *et al.*, 2008; Jouret *et al.*, 2019; Roux *et*

al., 2006) (Figure 1.5). Mutations within the three genes are also implicated in nonsyndromic autosomal recessive deafness (DFNB12 for *CDH23*, DFNB23 for *PCDH15* and DFNB18A for *USH1C*) (Bolz *et al.*, 2001; Bork *et al.*, 2001; Ahmed *et al.*, 2001; Ahmed *et al.*, 2003; Verpy *et al.*, 2000). A genotype-phenotype correlation has been shown for *CDH23* and *PCDH15*: null variants (stop-gain, some splice site and frameshift) which result in truncating USH1 proteins are responsible for the occurrence of USH1, whereas missense and some splice site mutations that do not truncate the protein lead to milder phenotypes associated with DFNB (Doucette *et al.*, 2009; Astuto *et al.*, 2002; Schultz *et al.*, 2011). This has led to the hypothesis that the residual function of USH1 proteins in DFNB "spares" retinal and vestibular function in patients. *CDH23* pathogenic variants have been found across the gene, affecting all protein domains, except for the signal domain, with clustering in the region encoding the EC domain (Astuto *et al.*, 2002). The most frequent *CDH23* pathogenic variant is the splice-site c.336+1G>A, which has been exclusively observed in Swedish USH1D individuals and has thus been suggested as a founder mutation (Astuto *et al.*, 2002; Oshima *et al.*, 2008). The most common *PCDH15* variant is c.733C>T (p.Arg245Ter), which is specific to the Ashkenazi Jews and is the predominant cause of USH1 in that population (Ben-Yosef *et al.*, 2003). Pathogenic variants in *USH1C* are a relatively rare cause of USH, with the exception of the Acadian population, where *USH1C* mutations represent the most common cause of USH1 with c.238dup (p.Arg80ProfsTer69) exhibiting highest prevalence, suggesting a founder effect (Verpy *et al.*, 2000; Zwaenepoel *et al.*, 2001; Ouyang *et al.*, 2005).

Pathogenic variants in *USH1G* (also known as *SANS*), responsible for USH1G (# 606943), are the rarest cause of USH1 (Bonnet *et al.*, 2011; Jouret *et al.*, 2019; Le Quesne Stabej *et al.*, 2012) (Figure 1.5). The nonsense variant c.113G>A (p.Trp38Ter) has been reported in six USH1G cases in three independent USH genetic screens, making it the most frequent *USH1G* mutation (Ouyang *et al.*, 2005; Bonnet *et al.*, 2011; Bujakowska *et al.*, 2014).

CIB2 is the causative gene of USH1J (# 614869). To date, a total number of 13 pathogenic variants (including a CNV, missense, frameshift and splicing variants) have been reported in *CIB2*. Of those, only one variant (c.192G>C, p.Glu64Asp) is linked to USH1J while the rest are associated with autosomal recessive hearing loss (DFNB48) (Booth *et al.*, 2018; Riazuddin *et al.*, 2012). The missense variant c.192G>C was identified in homozygous state in four individuals from a single consanguineous Pakistani family diagnosed with USH1J (Riazuddin *et al.*, 2012). The role of *CIB2* variants as disease causing in USH1J has been challenged by Booth *et al*., who reported four more novel *CIB2* pathogenic variants (three were loss of function and one was a missense) in families from diverse origins, all causing DFNB48 (Booth *et al.*, 2018), providing evidence to disqualify *CIB2* as an USH-causing gene. Yet, because of the private pathogenic variant causing USH1J in the

Pakistani family, *CIB2* is included in some genetic studies when investigating Usher syndrome's molecular basis (Aparisi *et al.*, 2014).

1.3.3.1.2. USH2 causative genes

Pathogenic variants within *USH2A* can explain the symptoms of up to 85% of USH2 patients and about 50% of all USH cases (Jouret *et al.*, 2019) (Figure 1.5), making it the most common cause of Usher Syndrome. *USH2A* pathogenic variants cause USH2A (# 276901). Mutations are spread throughout the gene with missense variants being the most common type (Baux *et al.*, 2014; Lenassi *et al.*, 2015). The majority of disease-causing *USH2A* variants are extremely rare, except the ancestral pathogenic variant c.2299delG (p.Glu767SerfsTer21), which accounts for 15-31% of USH2 cases in European patients (Dreyer *et al.*, 2008; Garcia-Garcia *et al.*, 2011; Le Quesne Stabej *et al.*, 2012). The *USH2A* variant c.2276 G>T, (p.Cys759Phe), also commonly found in patient cohorts, has been mainly associated with eye phenotypes (Rivolta *et al.*, 2000; Bernal *et al.*, 2003), proposing the existence of *USH2A* alleles that are retinal-disease specific. In a more recent study, investigating 186 individuals with recessive retinal disease and no childhood hearing complaints, the allelic hierarchy was evidenced by six *USH2A* alleles that were associated with nonsyndromic retinal disease only (defined as "retinal disease-specific") and were proposed to be likely phenotypically dominant to USH2 alleles (so the presence of at least one retinal disease specific *USH2A* allele would likely result in preservation of hearing) (Lenassi *et al.*, 2015).

ADGRV1 (also known as *GPR98*, *VLGR1* and previously- *MASS1*) pathogenic variants cause USH2C (# 605472). *ADGRV1* mutations are the second most common cause of USH2, accounting for about 6% of USH2 cases and about 5% of all USH cases (Le Quesne Stabej *et al.*, 2012; Jouret *et al.*, 2019) (Figure 1.5). To date about 106 *ADGRV1* variants are classed as presumed pathogenic (LOVD-USHBase, accessed 4 July 2020). It needs to be noted that from the reported pathogenic *ADGRV1* variants, not all are causative of USH with studies suggesting association of *ADGRV1* to febrile seizures and epilepsy (Nakayama *et al.*, 2002; Coll *et al.*, 2017). Reported *ADGRV1* mutations causing USH1C appear to be spread along the whole sequence with a cluster emerging in the terminal end and mutations predominantly resulting in a truncated protein product (Besnard *et al.*, 2012). In contrast to USH2A cases, no major recurrent mutations have been identified for USH2C (Besnard *et al.*, 2012).

WHRN (also known as *DFNB31*) was first linked to nonsyndromic deafness in two families (Mburu *et al.*, 2003) with its involvement in USH was later demonstrated by Ebermann *et al.* in a family with typical USH2 symptoms (Ebermann *et al.*, 2007). To date only four pathogenic variants have been linked to USH2D, all located towards the start of the gene (between exon 1 and exon 3)

(Ebermann *et al.*, 2007; Besnard *et al.*, 2012), showing that *WHRN* pathogenic variants account for a very small proportion of USH2 cases (Figure 1.5).

Some genotype-phenotype correlations have been suggested for USH2 genes. In a study analysing the audiological findings in 100 USH2 patients, Abadie *et al.* found that as a group the *ADGRV1*-caused USH2 cases had a higher proportion of severe hearing loss (40%), compared to *USH2A* cases (16%) (Abadie *et al.*, 2012). Although this was not statistically significant, it suggests that *ADGRV1* pathogenic variants are likely to result in more severe hearing symptoms than *USH2A* pathogenic variants in patients with USH2.

1.3.3.1.3. USH3 causative genes

CLRN1 pathogenic variants are the main cause of USH3, explaining about 2% of all USH cases (Figure 1.5) (Jouret *et al.*, 2019; Joensuu *et al.*, 2001). The missense variant c.144T>G (p.Asn48Lys) is relatively common and detected in Ashkenazi Jewish population (Fields *et al.*, 2002). *CLRN1* pathogenic variants are also implicated in nonsyndromic recessive retinitis pigmentosa 61 (Khan *et al.*, 2011) with less severe (missense) *CLRN1* variants, representing hypomorphic mutations and preserving hearing.

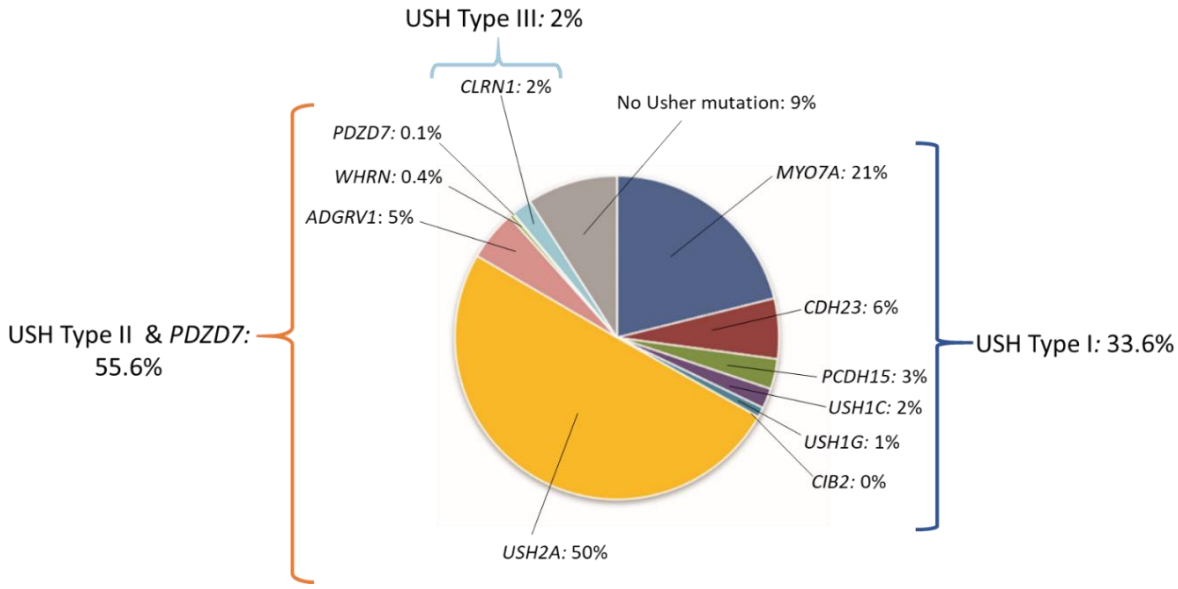

Figure 1.5. Frequency of causative mutations in patients presenting with both SNHL and RP suggesting Usher Syndrome (adapted from (Jouret *et al.*, 2019).

1.3.3.2. Usher Syndrome atypical and rare genes

PDZD7, encoding PDZ domain-containing 7 scaffolding protein, was first proposed as a prime candidate for USH syndrome due to its sequence similarity with the causative USH genes *USH1C* and *WHRN* (Schneider *et al.*, 2009). Next, Ebermann *et al*. reported the first clinical cases diagnosed with USH2 to carry pathogenic variants in *PDZD7* (Ebermann *et al.*, 2010). In one family the authors reported a heterozygous frameshift mutation in *PDZD7* detected in the presence of a homozygous truncating *USH2A* mutation in an USH2 affected individual with more severe retinal phenotype than their USH2 sister, suggesting that *PDZD7* has a modifying effect on retinal disease (Ebermann *et al.*, 2010). In another family a heterozygous truncating *PDZD7* mutation was found in and USH2 affected individual with a heterozygous frameshift mutation in *ADGRV1*, indicating digenic inheritance (instead of the typical monogenic inheritance in USH) (Ebermann *et al.*, 2010). The role of *PDZD7* in USH is further supported by evidence that *PDZD7*, *USH2A*, *ADGRV1* and *WHRN* interact *in vivo* and *in vitro* to form an USH2 complex (Zou *et al.*, 2014; Zou *et al.*, 2015; Chen *et al.*, 2014).

HARS codes for histidyl tRNA synthetase which charges tRNA molecules with histidine amino acids for protein translation (Antonellis and Green, 2008). The gene was proposed as a novel USH3 player as a result of an exome sequencing study of Old Order Amish and Mennonite (Plain) children, descendants of Swiss immigrants (Puffenberger *et al.*, 2012). A homozygous missense pathogenic *HARS* variant was identified in two individuals, diagnosed with USH3 and presenting with progressive sensorineural hearing loss, retinitis pigmentosa and episodic psychosis (Puffenberger *et al.*, 2012). Different *HARS* compound heterozygous missense pathogenic variants were identified in another Swiss individual, diagnosed with Usher Syndrome (unspecified type) in a later study (Tiwari *et al.*, 2016). The *HARS* variant identified by Puffenberger *et al*. was shown, *in vitro* to cause a reduction in the thermal stability of the protein (Abbott *et al.*, 2017). The role of *HARS* in the inner ear and retina has not been elucidated, therefore, the role of *HARS* as a typical USH3 causing gene requires further investigation.

CEP250 encodes centrosomal protein 250 (CEP250 or CNAP1), which is a member of CEP family of centrosome-associated proteins (Kumar *et al.*, 2013). *C2orf7*1, also known as *PCARE*, codes for photoreceptor cilium actin regulator protein PCARE and is a known retinitis pigmentosa causing gene (Collin *et al.*, 2010; Nishimura *et al.*, 2010). Homozygous stop-gain mutations in both *CEP250* and *C2orf7*1 (double homozygotes) were identified in three siblings from a consanguineous family of Iranian Jewish origin, who had hearing loss and an early onset severe PR (Khateb *et al.*, 2014). Three other siblings from the same family presented with a milder retinal phenotype and were found to be homozygous for the *CEP250* mutation, but heterozygous for the *C2orf7*1 mutation. It was thus concluded that the severe retinal involvement in the double homozygotes indicated a

potential additive effect on retinal burden, caused by nonsense mutations in the 2 genes occurring within the same cell (and possibly having an impact on the same photoreceptor region). In a later study Kubota *et al.* identified compound heterozygous mutations in *CEP250* in a Japanese family with cone-rod dystrophy and SNHL while Fuster-Garcia *et al.* detected two novel stop-gain *CEP250* mutations in a compound heterozygous state in an individual with symptoms similar to the Japanese family (Fuster-Garcia *et al.*, 2018; Kubota *et al.*, 2018). Interactions with other USH proteins and how CEP250 and PCARE function in the inner ear and eye are areas still open to investigation.

ABHD12 (α/β-hydrolase domain containing 12) is a membrane-embedded serine hydrolase, which catalyses the hydrolysis of 2-arachidonoyl glycerol and is a known causal gene for polyneuropathy, hearing loss, ataxia, retinitis pigmentosa and early onset cataract (PHARC). Stop-gain homozygous *ABHD12* mutations were identified in two siblings from a consanguineous Lebanese family and later in two siblings from a Chinese family, all clinically diagnosed with USH3 (Eisenberger *et al.*, 2012; Li *et al.*, 2019). A homozygous splice site mutation was detected in two other cases, where the individuals were originally thought to be affected by Usher Syndrome, but associated PHARC symptoms of dysfunctional central and peripheral nervous system made a PHARC diagnosis more plausible (Yoshimura *et al.*, 2015). All three authors emphasised that Usher Syndrome and PHARC have overlapping symptoms and phenotypic variability. *ABHD12* is known to be expressed in the retina but its expression in the inner ear would need to be investigated to support the role of *ABHD12* in USH3.

ARSG (arylsulfatase G) encodes a member of a class of enzymes called sulfatases, responsible for hydrolysing ester- sulphate bonds, and playing a role in a variety of biochemical processes (Ferrante *et al.*, 2002). Khateb *et al*. described 5 patients from three Yemenite Jewish Families with an atypical form of Usher syndrome, later designated as USH4 and characterised by distinctive retinal degeneration (involving a ring scotoma) and usually late-onset of progressive sensorineural hearing loss without vestibular involvement (Khateb *et al.*, 2018). The authors reported a homozygous *ARSG* missense mutation which segregated with the disorder in the families (Khateb *et al.*, 2018). A different homozygous *ARSG* missense mutation, a homozygous frameshift deletion and two compound heterozygous missense mutations, were recently described in three further cases, presenting with clinical features closely resembling the USH4 phenotype described by Khateb *et al*. (Abad-Morales *et al.*, 2020; Peter *et al.*, 2020). Khateb *et al.* and Peter *et al.* further demonstrated that the *ARSG* mutations in their studies abolished enzyme function and the gene was expressed in the human retina (Khateb *et al.*, 2018), while Girotto et al. showed *Arsg* expression at the top of hair cells in mouse cochlea (Girotto *et al.*, 2014). Knock-out homozygous *Arsg* mice show retinal degeneration and behavioural dysfunction suggesting systematic effects (Kruszewski *et al.*, 2016). Ingham *et al.* showed that knockdown of *Arsg* in mice

did not cause any detectable auditory dysfunction (Ingham *et al.*, 2020). However, that could be because hearing was tested in mice aged 14 weeks old (Ingham *et al.,* 2020) while Khateb *et al.* found that the onset of hearing loss was relatively late in the reported patients (Khateb *et al.*, 2018). Possible interactions with the other USH proteins have not been yet explored.

1.3.4. Role of Usher proteins in the inner ear

The ten USH causative genes code for proteins with a wide range of functions: actin-binding molecular motors (myosin VIIa encoded by *MYO7A*), cell adhesion proteins (cadherin 23, protocadherin 15 and usherin encoded by *CDH23, PCDH15* and *USH2A* respectively), scaffolding proteins (harmonin, SANS and whirlin encoded by *USH1C, USH1G* and *WHRN* respectively), an adhesion G-coupled receptor (ADGRV1/GPR98 encoded by *ADGRV1*), calcium and integrin binding protein (CIB2 encoded by *CIB2*) and transmembrane protein involved in scaffolding and cellular trafficking (clarin-1 encoded by *CLRN1*). These proteins form complexes and function cooperatively in both the inner ear hair cells and retinal photoreceptor cells. As this thesis focuses on hearing and audition, only the role of Usher proteins within the cochlea will be discussed. Their role in the retina is reviewed elsewhere (Kremer *et al.*, 2006; Cosgrove and Zallocchi, 2014). Most of the USH proteins (except CIB2 and clarin-1) are directly involved in the formation of links that hold stereocilia together and therefore play key roles in the morphogenesis of the hair bundle as summarised in the following sections.

1.3.4.1. Morphogenesis of the hair bundle within the inner ear

The inner ear stereocilia on the IHC and OHCs are organised in bundles, called stereociliary bundles which are important for transducing mechanical sound stimuli into electrical signals and essentially for sound perception. At the onset of hair bundle formation (as defined in the chick by Tilney *et al.*), the apical surface of each hair cell is covered with microvilli, which consist of parallel actin filaments held together by a set of actin-bundling proteins (Figure 1.6a) (Tilney, Tilney and DeRosier, 1992). The microvilli grow in length and form stereocilia of similar length, while one single kinocilium (primary cilium) is in the centre of the surface (Figure 1.6b). The kinocilium then moves to the periphery (lateral edge) of the hair cell, dictating the orientation of the hair bundle (i.e. its planar polarity) (Figure 1.6c). Next, the stereocilia closest to the kinocilium start to grow longer (Figure 1.6c), which is followed by elongation of stereocilia in adjacent rows, forming a staircase pattern of height-ranked rows (Figure 1.6d). The stereocilia then stop growing in length but grow in width and then again in length until they reach their final length (Figure 1.6d). The processes of elongation and widening of stereocilia are separated in chick but occur in parallel in mammals. Finally, excess microvilli on the apical surface are reabsorbed (Figure 1.6e). Upon maturation the stereocilia in each bundle are arranged into three rows of increasing length and the kinocilium (which during early development is positioned next to the longest stereocilia), is lost during the early stages of postnatal development and before the onset of hearing in mammalian cochlea (Figure 1.6f). The graded height across the rows of stereocilia (Figure 1.6e) is central to their transduction ability. This is because tension on the sloping tip links connecting the tops of stereocilia (discussed below) controls the opening probability of the mechanoelectrical

transducer (MET) channels which are located at the tips of the short- and middle-row stereocilia (Figure 1.6g) (Gillespie and Müller, 2009; Beurg *et al.*, 2009). Bending of the bundle towards the tallest row increases tension on tip links and channel conductance whereas deflection the opposite way (away from the tallest row) decreases channel conductance (Figure 1.6g), which is known as directional sensitivity (Shotwell, Jacobs and Hudspeth, 1981). This directional sensitivity is established along the epithelial plane and as a result the apical cytoskeleton in individual hair cells is planar polarised leading to the typical V-shaped stereocilia bundle and the off-centre position of the kinocilium (Figure 1.6h). Neighbouring hair cells also orient their hair bundles in the same direction, to presumably respond in a coherent manner to shared local stimuli (Figure 1.6h). This cell-to cell communication property is known as planar cell-polarity (PCP). Planar polarity mechanisms thus act at two levels (single-cell and intercellular) to provide the specific formation required for the hair bundles to function correctly in mechanotransduction and sound perception.

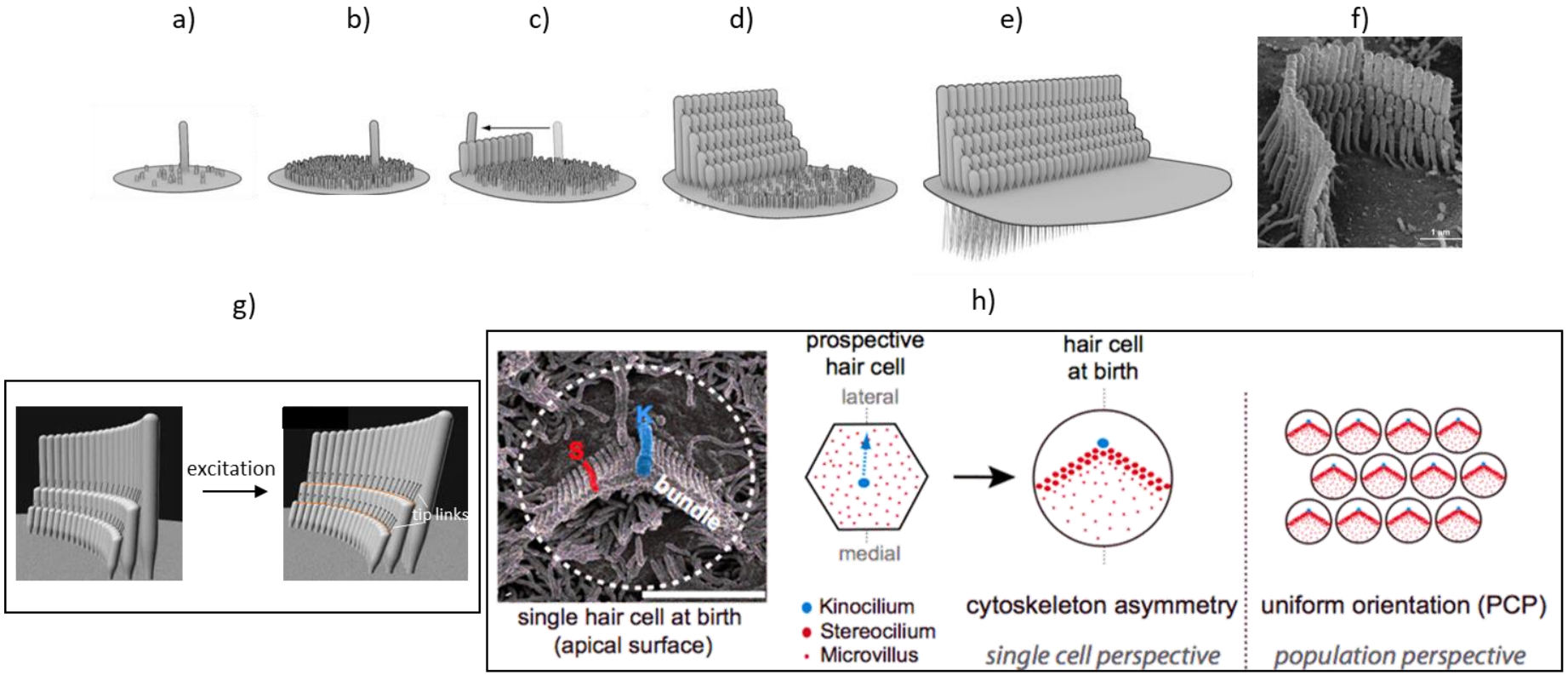

Figure 1.6. Hair bundle development, mechanotransduction and planar polarity (adapted from (Schwander, Kachar and Müller, 2010; Frolenkov *et al.*, 2004) and https://www.tarchini-lab.org/

a) at the onset of stereocilia formation the apical hair cell surface contains microvilli and one kinocilium; b) the microvilli grow in length; c) the kinocilium moves to the lateral edge of the hair cell; d) microvilli grow in width and reach their final length; e) excess microvilli are reabsorbed; f) scanning electron microscopy of mature OHC bundle that shows the staircase organisation of the three stereocilia rows (scale bar: 1µm); g) deflection of hair bundles in the direction of the longest stereocilia leads to the opening of transduction channels at the lower ends of tip links (highlighted in orange); h) model of planar polarity mechanisms in non-mammalian cells.

1.3.4.2. Stereocilia links and Usher proteins in development

During the stereocilia morphogenesis and growth and in adulthood, the stereocilia stay together through fibrous inter-stereociliar links and links with the kinocilium. By keeping the stereocilia and kinocilium together in a bundle, these links maintain the stereociliary cohesion, which is essential for bundle development, maintenance and function (Goodyear *et al.*, 2005; Michalski *et al.*, 2007). Developing cochlear hair cells in mice show kinociliary links, transient lateral links and ankle links (during embryonic and early postnatal development), whereas functionally mature cochlear cells contain tip links and horizontal top connectors (Goodyear *et al.*, 2005) (Figure 1.7).

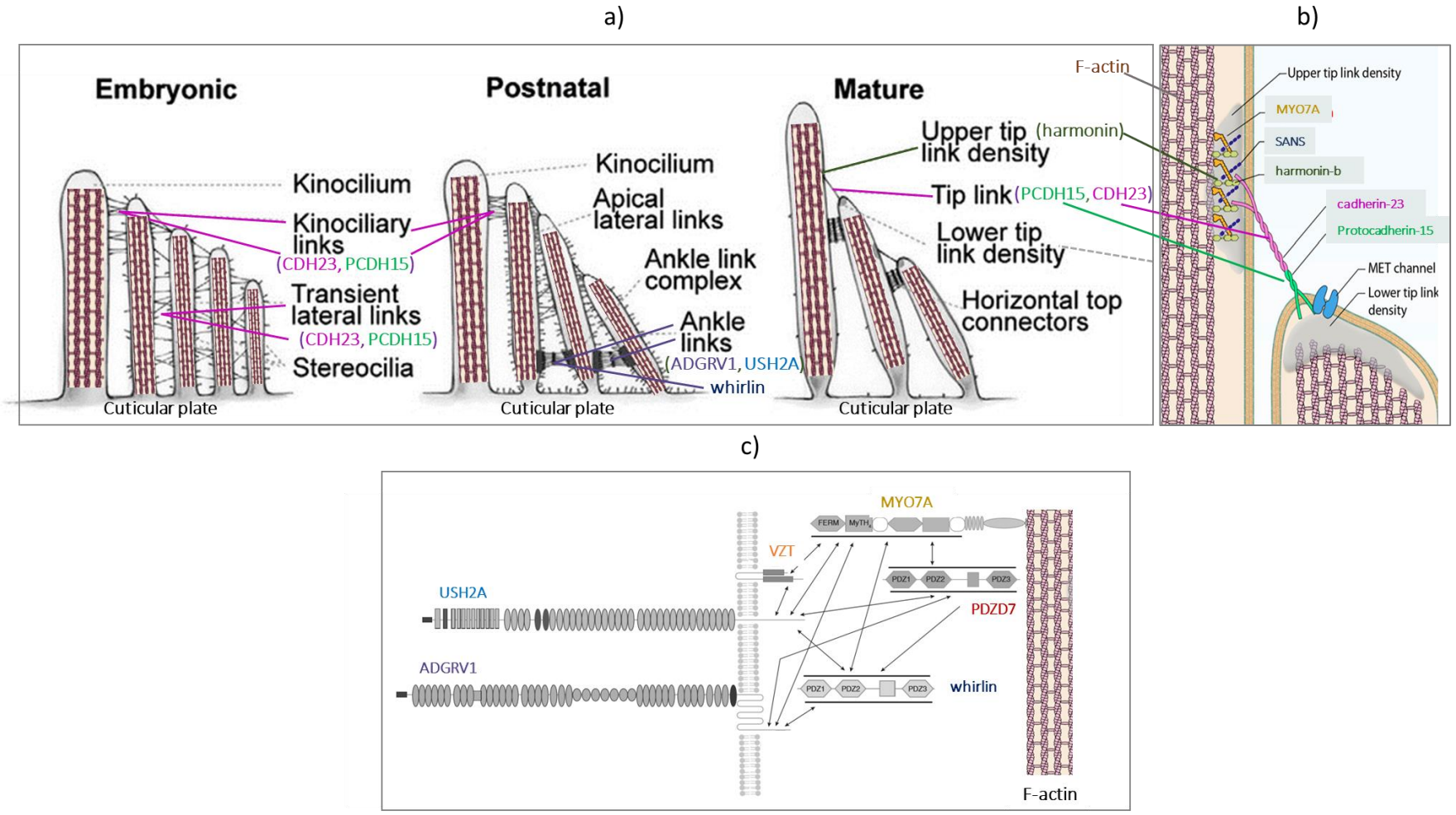

Figure 1.7. Stereociliary links in the hair bundle through development (a), molecular components of the tip links and mechanotransduction complex (b) and components of the ankle link complex (c) (adapted from (Mathur and Yang, 2015; Schwander, Kachar and Müller, 2010; Michalski *et al.*, 2007; Richardson and Petit, 2019; El-Amraoui and Petit, 2014).

1.3.4.2.1. Kinociliary and transient lateral links

The kinociliary links connect the longest row of stereocilia to the kinocilium (Figure 1.7a) and are present only transiently in mammal cochlea, disappearing with the kinocilium shortly after birth. As the kinocilium is crucial for establishing the hair bundle polarity, the kinociliary links are thought to be essential for ensuring the kinocilium and the stereociliary bundle are coordinated within the hair bundle during early development. Transient lateral links also form during the early stages of development, recorded as early as embryonic day (E) 17.5 (Goodyear *et al.*, 2005), and are the first links to interconnect the stereocilia along their entire length (Figure 1.7a). These lateral links become progressively restricted towards the distal end of the stereocilia forming apical lateral links in postnatal development (Figure 1.7a). The apical lateral links are also thought to become integrated into the links at the very tip of the stereocilia (tip links) in mature bundles (Goodyear *et al.*, 2005; Richardson and Petit, 2019) (Figure 1.7a). The USH1 proteins CDH23 and PCDH15 have been shown to be expressed in developing stereocilia and are both localised to kinociliary and transient links (Goodyear *et al.*, 2010; Michel *et al.*, 2005) (Figure 1.7a).

1.3.4.2.2. Ankle links

At postnatal day (P) 2, thin filaments called ankle links become concentrated at the bottom of the stereocilia (where they insert into the cuticular plate) (Goodyear *et al.*, 2005) (Figure 1.7a). In the mouse these links are transient and are only detected until P12, disappearing before the onset of hearing (Goodyear *et al.*, 2005). The USH2 proteins ADGRV1, usherin, and whirlin are all expressed at the base of the inner ear hair cells, detected in embryonic development (as early as E17 for ADGRV1 and E18 for usherin) and just after birth (as early as P1 for whirlin) (Michalski *et al.*, 2007; Zou *et al.*, 2014; Delprat *et al.*, 2005). Because of the localisation of ADGRV1 and usherin and because they both have long extracellular regions, both proteins were suggested as components of the ankle link filaments (Figure 1.7c) (Adato *et al.*, 2005a; McGee *et al.*, 2006). Together with whirlin and PDZD7 (USH2 modifier), all USH2 proteins have been shown to interact with one another to form a multiprotein complex, known as USH2 or ankle link complex (ALC) (Figure 1.7c), which is located at the cytoplasmic region of the ankle links (Michalski *et al.*, 2007; Zou *et al.*, 2015). Another protein, vezatin (VZT), known to interact with myosin VIIa and usherin (Michalski *et al.*, 2007), is also expressed around the base of mouse stereocilia and colocalises with ankle links (Figure 1.7c) (Küssel-Andermann *et al.*, 2000).

1.3.4.2.3. Horizontal top connectors

Horizontal top connectors, also known as top links or side-to-side links first appear at P9 and continue to develop at P12 (when the ankle links disappear), attaining their mature appearance between P12 and P14 and becoming highly organised by P19 (Goodyear *et al.*, 2005). In their

mature state, the horizontal connectors are zipper-like structures (due to the presence of a central density region), located near the tops of the stereocilia (Figure 1.7a) and connecting them within and across rows. The central density region has been observed in top connectors within cochlear OHCs, but not in IHCs (Goodyear *et al.*, 2005; Tsuprun *et al.*, 2003), showing that the structure of the links varies according to type of hair cells which is likely linked to their different function. In mice, a protein called stereocilin (STRC) has been shown to be associated with these links within OHCs, but has not been detected in IHCs (Verpy *et al.*, 2011), and is thus thought to play a key role in the central density region. STRC is also linked to the tectorial membrane attachment crowns (Verpy *et al.*, 2011), present at the tips of the tallest stereocilia of OHCs to couple them to the tectorial membrane above.

1.3.4.2.4. Tip links

Tip links, which connect the tips of the stereocilia between adjacent rows appear on the hair bundles of both OHCs and IHCs at the same time as ankle links (around P2), together with the tectorial membrane attachment crown and are still present in the mature OHC bundle at postnatal day P19 (Figure 1.7a) (Goodyear *et al.*, 2005). The tip links play a major role in transmitting force to the MET channels (Figure 1.7b), which convert sound waves into an electrical signal (Kazmierczak *et al.*, 2007; Michalski and Petit, 2015). Although the IHCs and OHCs have distinct roles (signal transmission for IHCs and frequency dependent amplification for OHCs), they both perform mechano-electrical transduction. Increasing evidence has shown that cadherin 23 and protocadherin 15 make up the upper and lower part of the tip link respectively (Figure 1.7b) (Siemens *et al.*, 2004; Ahmed *et al.*, 2006). Cadherin 23 binds to the upper tip link density region (UTLD) while protocadherin 15 binds to the lower tip link density (LTLD) region (Figure 1.7b), which are protein-dense plaques that underlie the stereocilia membrane at each end of the tip link (Kachar *et al.*, 2000). Protocadherin 15 has also been suggested to interact with TMC1 and TMC2, which are core components of the MET complex (Maeda *et al.*, 2014; Pan *et al.*, 2013). The three USH1 proteins myosin VIIA, SANS and harmonin localise to the UTLD of mature cochlear cells (Figure 1.7b) and form a complex responsible for anchoring cadherin23 to the actin cytoskeleton of stereocilia (Yan *et al.*, 2010; Boeda *et al.*, 2002). Out of the three proteins, it is harmonin which binds to cadherin 23 within the upper part of the tip links (Figure 1.7b) (Boeda *et al.*, 2002), bridging cadherin 23 to the cytoskeletal actin core. Myosin VIIA is expressed along the stereocilia and the cuticular plate (Weil *et al.*, 1995; Hasson *et al.*, 1995) and as a motor protein is known to transport components of the actin assembling machinery to the tips of the stereocilia to regulate length, so it is also likely involved in the transport of other USH molecules (Rhodes *et al.*, 2004). SANS is highly concentrated below the cuticular plate (Adato *et al.*, 2005b) and has been shown to bind harmonin and myosin VIIA (Figure 1.7b), and is required for localisation of harmonin to stereocilia tips (Lefèvre *et al.*, 2008). The molecular composition of the LTLD region is

much less well characterised, however, it is thought that the region contains elastin filaments, important in adaptation to occur following the mechanical deflection of stereocilia in the direction of the tallest stereocilia (Figure 1.7b) (Eatock, 2000). Other proteins such as whirlin, myosin XVa (essential for stereocilia elongation), actin regulatory protein epsin8 and myosin IIIa might also take part in anchoring protocadherin 15 (Delprat *et al.*, 2005; Manor *et al.*, 2011; Schneider *et al.*, 2006).

1.3.4.2.5. Other USH proteins that may interact with stereocilia link proteins and are important hair bundle proteins

The USH1 player calcium and integrin-binding protein-2 (CIB2) is a newly discovered member of the USH protein family. It has been shown to bind to the components of the hair cell mechanotransduction complex, TMC1 and TMC2 (Giese *et al.*, 2017b). It has thus been suggested to play an essential role in mechanotransduction (Wang *et al.*, 2017b) by being involved in limiting the growth of transducing shorter row stereocilia in mammalian auditory hair cells (Giese *et al.*, 2017b). Regulation of stereocilia length is likely to occur at their tips where CIB2 may bind to whirlin and be part of the myosin XVa/whirlin stereocilia elongation complex (Belyantseva *et al.*, 2005). The above findings together with results from Michel *et al.,* propose that unlike the other five known USH1 proteins, functional CIB2 is not required to ensure the early cohesion and shaping of the growing auditory hair bundle but is necessary for the terminal differentiation and maturation stages (Michel *et al.*, 2017).

Similar to CIB2, the USH3 protein clarin-1 has not been detected to play a direct role in stereocilia link formation. However, there is compelling evidence of a link between CLRN1, actin and other USH proteins (such as myosin VIIa), supporting the role of clarin-1 as an essential hair bundle protein (Tian *et al.*, 2009; Adato *et al.*, 1999), required for the development or maintenance of the normal shape of the bundle (Geng *et al.*, 2012). Similarly to the USH2 proteins, clarin-1 is detected at the base of both IHCs and OHCs during embryonic development (Zallocchi *et al.*, 2009; Geng *et al.*, 2009). Postnatally clarin-1 shows expression at the apical regions of hair cells, which disappears from the OHCs at P10, while a weak signal remains at the base of IHCs (Zallocchi *et al.*, 2009). Contrary to previous reports suggesting a possible role for clarin-1 in the sensory synapses of inner hair cells (Adato *et al.*, 2002; Geng *et al.*, 2009), mutant mice findings from Geng *et al.* demonstrate that *Clrn1* is not required for the development or maintenance of ribbon synapse and is not essential in the hair cell presynaptic function (Geng *et al.*, 2012).

1.3.4.3. Usher proteins isoforms

Many of the USH proteins are expressed as multiple isoforms (Figure 1.8) (as reviewed in (El-Amraoui and Petit, 2005; Kremer *et al.*, 2006). The protein isoforms are often expressed in specific

tissues and sub-cellular compartments and even in a tonotopic manner along the cochlea (as recently suggested for myosin VIIa) (Adato *et al.*, 2005a; Li *et al.*, 2020), which adds to the complexity of deciphering the function of the USH proteins. Harmonin is expressed in at least three isoform subclasses (a,b and c) that differ by the number of the PDZ domains and the presence or absence of a second coiled-coil domain (Figure 1.8) (Verpy *et al.*, 2000; Boeda *et al.*, 2002). Cadherin 23 has at least three isoforms that all contain a cytoplasmic domain, but differ at the opposite C-terminus (Figure 1.8) (Lagziel *et al.*, 2005). Three protocadherin 15 splice isoform classes (Pcdh15-CD1, Pcdh15-CD2 and Pcdh15-CD3), which differ in the C-terminal part of their cytoplasmic domain (Figure 1.8) have been shown to be present in the hair bundles of developing cochlear hair cells (Ahmed *et al.*, 2006). Two whirlin isoforms (a longer-FL-whirlin and a shorter- C-whirlin depending on the presence or absence of the PDZ domains 1 and 2 in the N-terminus of the protein) are detected in the inner ear (Figure 1.8) (Ebrahim *et al.*, 2016) . The mouse *Clrn1* is known to give rise to three protein isoform, of which isoform 1 and isoform 2 both contain four transmembrane domains (Figure 1.8), and isoform 3 contains only two (not shown) (Adato *et al.*, 2002).

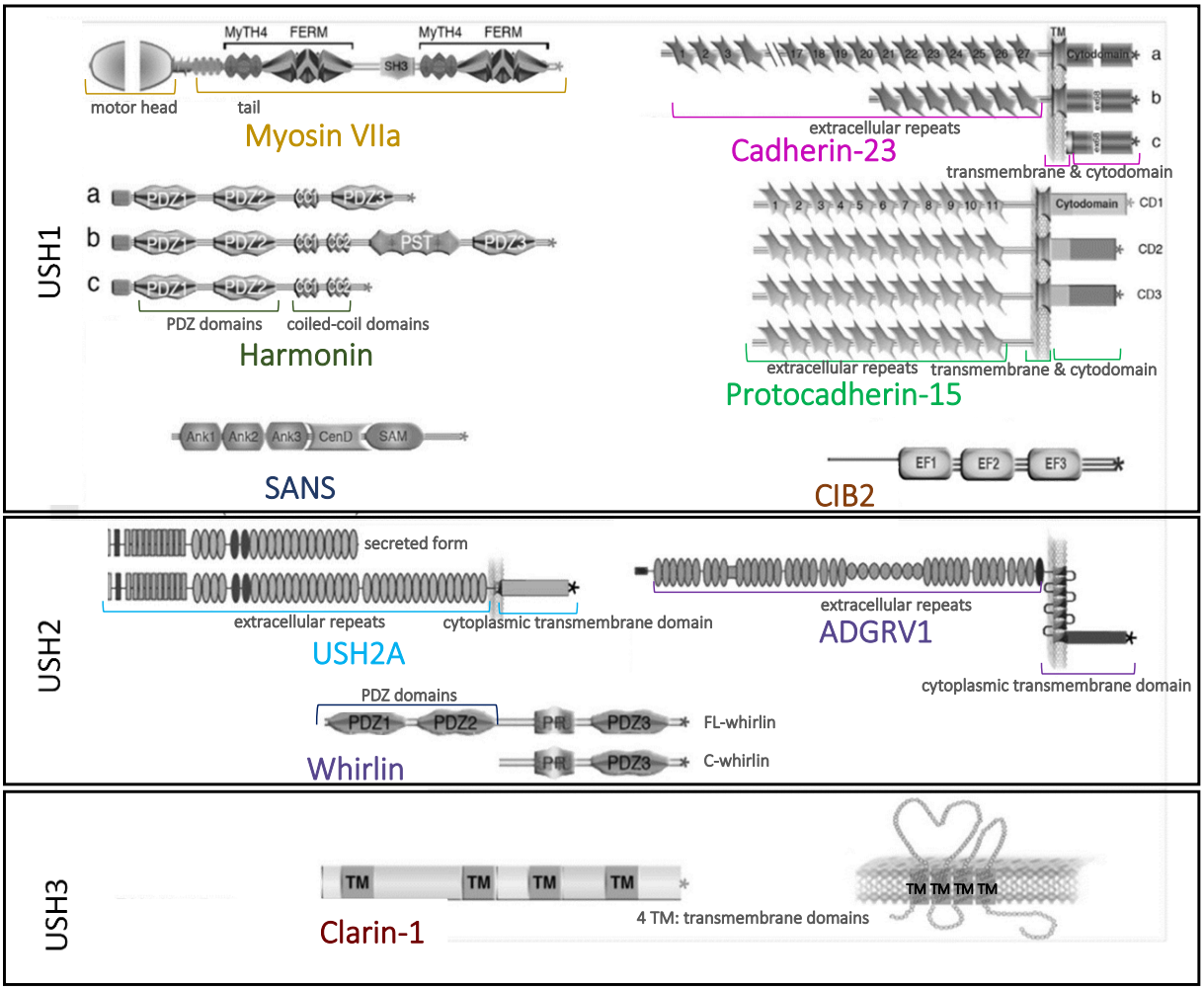

Figure 1.8. USH protein isoforms with predicted domain structures. Many of the proteins exist in multiple isoforms: three sub-classes of harmonin isoforms (a,b and c), three sub-classes of cadherin-23 (a,b and c), three main classes of protocadherin-15 (CD1-CD3), a secreted and a transmembrane form of usherin and two isoforms of whirlin in inner ear (adapted from (El-Amraoui and Petit, 2005; Géléoc and El-Amraoui, 2020).

1.3.5. Pathological insights from mouse models of Usher syndrome

1.3.5.1. Stereocilia defects resulting from dysfunctional Usher proteins

USH1 mouse models of disrupted *Myo7a, Ush1g*, *Chd23*, *Pcdh15* and *Ush1c* genes share hair bundle morphological defects including fragmentation and misorientation in embryonic development and abnormal staircase architecture postnatally (Lefèvre *et al.*, 2008), whereas mouse models for *Cib2* show similar bundle disruptions only postnatally (Table 1.5 & Figure 1.9) (Wang *et al.*, 2017b; Giese *et al.*, 2017b). The kinocilium in *Myo7a*, *Cdh23*, *Pcdh15*, *Ush1g* and *Ush1c* mutated mice appears mispositioned and often misplaced (Lefèvre *et al.*, 2008; Di Palma *et al.*, 2001). While the mutated USH1 stereocilia in OHCs form clumps and display abnormal length leading to loss of staircase architecture (Table 1.5 & Figure 1.9b), the IHCs stereocilia show to be affected differently, with abnormally shaped tips and reduction to complete loss of tip links

(Lefèvre *et al.*, 2008; Alagramam *et al.*, 2011). The stereocilia defects appear to be more severe as the mice grow older with loss of stereocilia observed by P15-P20 in *Ush1c, Myo7a* and *Pcdh15* mutated mice and at P30 for some *Cib2*$^{-/}$ mice, which show loss of OHCs and fusion of IHCs stereocilia (Figure 1.9a) (Lefèvre *et al.*, 2008; Self *et al.*, 1998; Washington *et al.*, 2005; Wang *et al.*, 2017b). It is important to also note that the severity of the mutations for some USH1 models corresponds to the severity of cochlear pathology and hair cell development. For example, mice with in-frame deletions and missense mutations (as in *Pcdh15*$^{av^J}$, *Pcdh15*^{av2J} and *Myo7a*^{sh1}, *Myo7a*^{6j}) have shown less disorganised stereocilia within the organ of Corti compared to same gene presumptive null alleles (*Pcdh15*^{av3J} and *Myo7a*^{4626SB}) (Mburu *et al.*, 1997; Kros *et al.*, 2002; Pawlowski *et al.*, 2006), suggesting a genotype-phenotype correlation.

Table 1.5. USH1 proteins expression, inner ear hair cell localisation and defects in mouse mutants

USH type	Gene: Protein	Protein Function	Expression: Embryonic	Expression: Mature cells	Hair cell localisation	Knock-out mouse mutants	OHCs defects	IHCs defects	Cochlear dysfunction	References
USH1	*MYO7A:* myosin VIIa	Molecular motor protein	IHCs and OHCs' emerging protrusions; tips of differentiated stereocilia and surrounding microvilli along the whole cochlea	UTLD in tallest and middle stereocilia row	Hair cell cytoplasm and hair bundle, UTLD	*Shaker*: *Myo7a*$^{sh1/sh1}$ *Myo7a*$^{816B/816SB}$ *Myo7a*$^{4626SB/4626SB}$	Stereocilia form clumps; many have abnormal length	Medium stereociliary tips not acquiring their typical prolate shape and tip links becoming reduced in number	No CM, CAP or SP responses up to the max intensities used	(Mburu *et al.*, 1997; Lefèvre *et al.*, 2008; Holme and Steel, 2002; Kros *et al.*, 2002; Boeda *et al.*, 2002; Self *et al.*, 1998; Tian *et al.*, 2010; Alagramam *et al.*, 2011; Haywood-Watson *et al.*, 2006; Washington *et al.*, 2005; Di Palma *et al.*, 2001; Raphael *et al.*, 2001; Pawlowski *et al.*, 2006; Caberlotto *et al.*, 2011; Kikkawa *et al.*, 2003; Wang *et al.*, 2017b; Giese *et al.*, 2017a)
	USH1C: harmonin	Scaffold protein			Upper tip link and synapse	*Null: Ush1c*$^{-/-}$	Stereocilia form clumps; stereocilia shorter in small and medium rows compared to wild type		No detectable ABR response to click stimuli at 8, 16, 24 and 32 kHz tone bursts at max intensities used	
	CDH23: cadherin 23	Cell adhesion (upper tip link)		No longer detected	Transient lateral link, kinociliary link, tip link	*Waltzer:* *Cdh23*$^{v/v}$ *Cdh23*$^{v2J/v2J}$		Same as above with tip links becoming completely absent		
	PCDH15: protocadherin 15	Cell adhesion (lower tip link)		Tips of short and middle stereocilia row and apico-lateral region of tallest stereocilia		*Ames Waltzer:* *Pcdh15*$^{av6J/av6J}$ *Pcdh15*$^{avJ/avJ}$ *Pcdh15*$^{av2J/av2J}$ *Pcdh15*$^{av3J/av3J}$ *Pcdh15*$^{av5J/av5J}$			
	USH1G: SANS	Scaffold protein	Apical region of cochlear hair bundle	UTLD in tallest and middle stereocilia row	Upper tip link and stereociliary tip	*Jackson shaker:* *Ush1g*$^{js/js}$ *Null: Ush1g*$^{-/-}$		Same as above with tip links becoming reduced in number	No detectable ABR (same as above). No CM detected. Defective DPOAE for high frequencies, and normal for low frequencies.	
	CIB2: CIB2	mediating intracellular calcium signalling	No data	Stereocilia tips from shortest row	Stereocilia, near tip	*Null: Cib2* $^{-/-}$	Disorganised stereocilia with medium row overgrown and tallest row retracted	Staircase architecture maintained (middle and tallest row overgrown); kinocilium failing to retract	No detectable ABR (same as above). DPOAE show threshold elevations. MET currents abolished in OHCs and IHCs.	

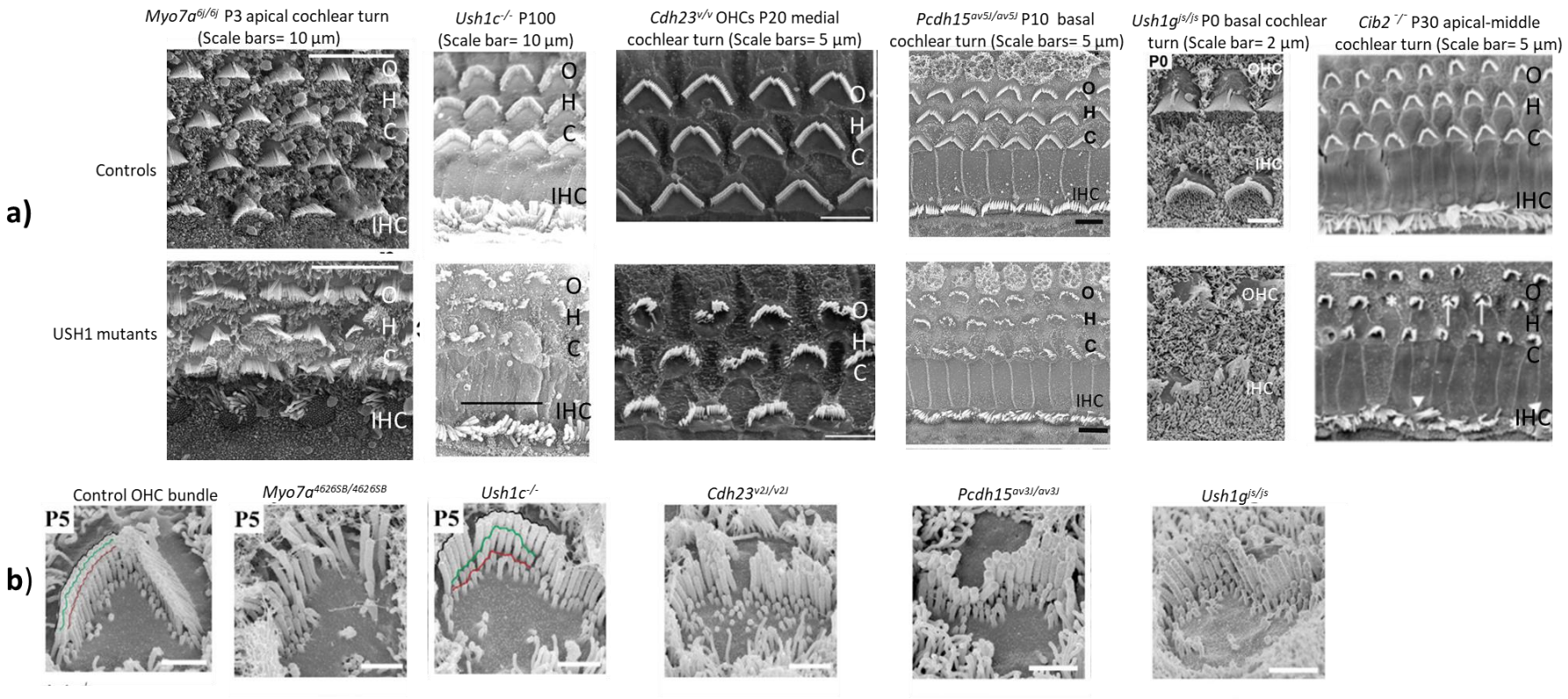

Figure 1.9. SEM images of cochlear hair bundles from USH1 mice of different genotypes and ages (images adapted from (Kros *et al.*, 2002; Tian *et al.*, 2010; Holme and Steel, 2002; Washington *et al.*, 2005; Wang *et al.*, 2017b; Lefèvre *et al.*, 2008)
a) low-magnification SEM images of OHCs and IHCs of USH1 knock-out mice in bottom row (except *Cdh23*$^{v/v}$ where only OHCs bundles imaged), showing disordered arrangement of stereocilia on OHCs and only moderate disorganisation of IHCs stereocilia compared to controls; b) high magnification SEM images of OHC bundles of P5 USH1 mutant mice (view from the end of the apical cochlear turn) showing abnormal height of many stereocilia of the medium row and frequent absence of small stereocilia in mutant hair bundles compared to controls (Scale bar = 1µm)

While the disrupted function of USH1 proteins affects stereocilia early in embryonic development along the whole of the cochlea, knock-out mouse models of the USH2 genes show structural defects affecting the hair cell bundles in post-natal development, preferentially at the basal half of the cochlea (sensitive to high frequencies) (Table 1.6 & Figure 1.10b). *Adgrv1* knock-out mice show the most severe USH2 stereociliary defects, that are also detected the earliest: disorganised OHC stereocilia bundles, stereocilia tilted at their base and forming U-shape (instead of a V-shape) plus significant loss of OHCs by 2 months of age (Table 1.6 & Figure 1.10a) (McGee *et al.*, 2006; Michalski *et al.*, 2007; Zou *et al.*, 2015). *Ush2a* $^{-/-}$ mice show an intermediate phenotype of severity with distorted OHC bundle shape and loss of OHCs by 4 months of age (Table 1.6 & Figure 1.10a-b) (Liu *et al.*, 2007; Zou *et al.*, 2017) and *Whrn* (also known as *Dfnb31*) mutated mice have least severely affected stereociliary bundle with various U-shaped OHC hair bundles observed at P15, leading to degeneration by P99-P102 (Table 1.6 & Figure 1.10a). The IHCs dysmorphology, on the other hand, shows to be less evident and only visible at high magnification (Zou *et al.*, 2015). While according to Lui *et al.* the IHCs in *Ush2a* $^{-/-}$ mice were present throughout the cochlear spiral and appeared to be normal (Figure 1.10b) (Liu *et al.*, 2007), Zou *et al.* showed that at high scanning electron microscope (SEM) magnification the IHCs bundle showed an additional ectopic (outside the bundle) stereocilia at the neural edge of the IHC apex (Figure 1.10a) (Zou *et al.*, 2015). This ectopic stereociliary phenotype was also present in *Adgrv1*$^{-/-}$ but not in *Dfnb31*$^{neo/neo}$ mice (Figure 1.10a). Both *Adgrv1*$^{-/-}$ and *Dfnb31*$^{neo/neo}$ mice had thicker IHC stereocilia than what was observed in wild-type mice, with IHCs lost by 2 months for Vlgr1/del7TM and by P99-P102 for *Dfnb31*$^{wi/wi}$ (Figure 1.10a) . The presence of ankle links in mutated USH2 mice has also been investigated, showing missing ankle links in *Adgrv1*$^{-/-}$(McGee *et al.*, 2006; Michalski *et al.*, 2007) while they were present in Dfnb31$^{neo/neo}$ (Zou *et al.*, 2015) and not studied in *Ush2a* $^{-/-}$mice, suggesting that ADGRV1 is essential for the formation of the ankle links while whirlin is not and usherin is unknown.

Consistent with the expression of clarin-1, *Clrn1*$^{-/-}$ knock-out mice also show bundle morphology defects in cochlear hair cells, that become more obvious by P2 and primarily affect OHCs along all turns of the cochlea while IHCs appear normal or mildly affected (Geng *et al.*, 2009) (Table 1.6 & Figure 1.10c). Profound hearing loss is detected at P21 by which time a loss of OHCs is also observed, while IHCs appear to be missing at both basal and apical turn of the cochlea around P30 (Geng *et al.*, 2009). SEM in *Clrn1*$^{-/-}$ at P3-P4 shows that the disruption of hair bundle integrity is characterised by splits in the V-shaped bundle and loss of some of the tall stereocilia (Geng *et al.*, 2012) (Figure 1.10c). As tip links, the bundle orientation and the graded heights of the stereocilia in the hair bundles of *Clrn1*$^{-/-}$ mice are not affected (Geng *et al.*, 2012), it was concluded that *Clrn1* is not required for the formation of tip links or the polarised orientation of the hair bundle or the development of graded heights of stereocilia. Instead, the disruption of the hair bundle structure

and loss of stereocilia in *Clrn1*$^{-/-}$ mice suggests that *Clrn1* might be required to maintain the structure of the bundle after it is formed (Geng *et al.*, 2012).

Table 1.6. USH2 and USH3 proteins expression, inner ear hair cell localisation and defects in mouse mutants

USH type	Gene: Protein	Protein Function	Expression:		Hair cell localisation	Knock-out mouse mutants	OHCs defects	IHCs defects	Cochlear dysfunction	References
			Embryonic	Mature cells						
USH2	*USH2A:* usherin	Cell adhesion & signalling	At the base of inner ear hair cells. The proteins form the ALC complex, located at the cytoplasmic region of the ankle links	only transient with ankle links	Ankle link and synapse	*Ush2a* $^{-/-}$	Distorted bundle shape and mislocalised kinocilia, stereocilia still connected along their entire length and usually did not tilt. Loss of basal-coil hair cells by 4 months of age while apico-middle turn of cochlea not affected	IHCs present throughout cochlear spiral, showing normal stereociliary thickness	ABR thresholds consistently elevated at high frequencies. Elevated DPOAE thresholds for f2 frequencies >15 kHz from 4 months of age	(Holme *et al.*, 2002; Delprat *et al.*, 2005; McGee *et al.*, 2006; Liu *et al.*, 2007; Michalski *et al.*, 2007; Manor *et al.*, 2011; Zou *et al.*, 2015; Ebrahim *et al.*, 2016; Zou *et al.*, 2017)
	ADRGV1: ADGRV1	Cell adhesion & signalling			Ankle link and synapse	*Adgrv1*$^{-/-}$ Vlgr1/del7TM	Disorganised stereocilia bundles with stereocilia tilted at their base and forming U-shape; significant loss of OHCs from basal half of cochlea by 2 months of age	Loss of IHCs at the basal half of the cochlea by 2 months of age, middle and apical turn less severely affected	ABR thresholds for click stimuli and pure-tone elevated across entire frequency range tested (2-32 kHz); Absence of DPOAEs at all f2 frequencies tested (2-20 kHz)	
	WHRN/ DFNB31: whirlin	PDZ scaffold protein			Ankle link, stereociliary tip and synapse	*Dfnb31*$^{wi/wi}$ *Dfnb31*$^{neo/neo}$	Various U-shaped hair bundles observed at P15, leading to degeneration by P99-P102	IHC stereocilia appear shorter, leading to degeneration of IHCs by P99-P102, more severe in basal regions compared to apical turn of cochlea	ABR thresholds elevated across all frequencies (4-45.2 kHz), not as elevated as Ush2a and Adgrv1 mutants	
USH3	*CLRN1:* clarin-1	Cell adhesion	At the base of both IHCs and OHCs.	At apical aspects of hair cells postnatally, with expression in OHCs lost at P10	Hair bundle, atypical cytoplasm and synapse	*Clrn1*$^{-/-}$	Disorganised stereocilia with splits in the V-shaped bundle and loss of some of the tall stereocilia. Loss of OHCs is observed at P21	Stereocilia in IHCs appeared normal or mildly affected. By P30 almost all IHCs are lost throughout most of the cochlea	Elevated ABR thresholds across entire frequency range tested (8-32 KHz) and prolonged peak and interpeak latencies. No detectable DPOAEs above the noise floor at P21	(Geng *et al.*, 2009; Zallocchi *et al.*, 2009; Geng *et al.*, 2012)

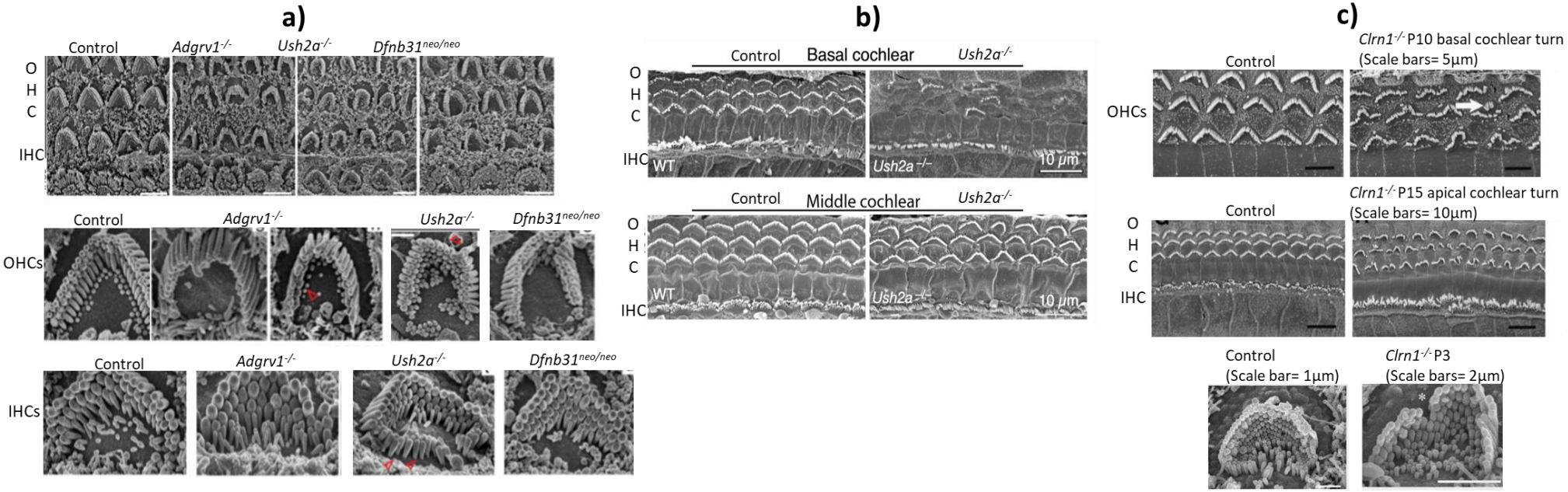

Figure 1.10. SEM images of cochlear hair bundles from USH2 and USH3 mice of different genotypes and ages (adapted from (Zou *et al.*, 2015; Liu *et al.*, 2007; Geng *et al.*, 2009; Geng *et al.*, 2012)

a) Low and high magnification SEM images of USH2 knock-out mice, showing abnormal hair cell bundles (top row); USH2 knock-out mice OHC bundles show: stereocilia tilted at stereociliary bases and missing stereocilia in the shortest stereociliary row (*Adgrv1*$^{-/-}$), distorted shapes and mislocalised kinocilia (*Ush2a*) and a blunt U-rather than a sharp V-shape (*Dfnb31*) (middle row); USH2 knock-out mice IHC bundles show thicker and ectopic stereocilia (*Adgrv1*), ectopic stereocilia, but no difference in thickness (*Ush2a*) and thick stereocilia (*Dfnb31*) (Zou et al., 2015).

b) Low magnification SEM showing *Ush2a* stereocilia affected only in the basal cochlear turn (top) compared to the middle-apical turn (bottom)

c)Low and high magnification SEM images of *Clrn1* mouse mutants at different stages at both the basal and apical cochlear turn showing disorganised stereocilia bundles in OHCs (with splits in the bundles and loss of some of the tallest stereocilia and normal looking bundles in IHCs (top two rows).

1.3.5.2. Cochlear dysfunction and hearing impairment resulting from dysfunctional Usher proteins in knock-out mice

As a result of the severe stereocilia defects in USH1 knock-out mice, affecting both IHCs and OHCs throughout the cochlear turns, the USH1 mutant mice show severe cochlear dysfunction and severe hearing impairment. Absence of ABR response to click stimuli at 8, 16, 24 or 32 kHz tone bursts demonstrated that the cochlea in *Ush1g*, *Chd23*, *Pcdh15* and *Ush1c* and *Cib2* knock-out mice was non-functional and the mice were profoundly deaf at all frequencies at all time points (Raphael *et al.*, 2001; Washington *et al.*, 2005; Haywood-Watson *et al.*, 2006; Di Palma *et al.*, 2001; Tian *et al.*, 2010; Kikkawa *et al.*, 2003; Caberlotto *et al.*, 2011; Wang *et al.*, 2017b). It has also been suggested that no ABR response at the maximum intensity of above 90dB SPL indicates that both OHCs and IHCs were impaired as even a complete loss of OHCs function cannot account for more than a 60bB threshold elevation (Dallos and Harris, 1978).

DPOAE measurements further demonstrated that OHC function of *Ush1g*$^{-/-}$ mice was defective at high frequencies, but was partly preserved at low frequencies in mature mutant mice before the complete loss of OHCs at P15 (Caberlotto *et al.*, 2011). DPOAE measurements are also available for *Cib2*$^{-/-}$ mice, indicating significant threshold elevation compared to controls, suggesting OHC function deficits in the Cib2 knock-out mice (Wang *et al.*, 2017b). Additional electrophysiology experiments, testing MET currents in *Cib2* knock-outs, indicated a complete lack of MET current in OHCs and in the more intact IHCs at P7 compared to controls, suggesting that disrupted *Cib2* function abolishes MET currents in auditory hair cells (Wang *et al.*, 2017b; Giese *et al.*, 2017b).

Compared to the hair cell defects along the whole cochlea in the USH1 knock-out mice, the USH2 mutated mice are preferentially affected at the basal half of the cochlea, which consequently leads to elevated thresholds at high frequency. ABR thresholds of all three knock-out USH2 mouse models at P30 (when the mouse cochlea has matured) showed to be elevated in the frequency range 4-45.2 kHz relative to wild type controls (Zou *et al.*, 2015), with *Ush2a* thresholds being consistently elevated at high frequencies (Liu *et al.*, 2007). *Adgrv1*$^{-/-}$ mice were most severely affected, while *Dfnb31*$^{neo/neo}$ showed the mildest hearing loss and *Ush2a* $^{-/-}$ mice were in-between, correlating with the degree of ALC disruptions and stereociliary bundle defects. DPOAE signal testing of *Ush2a* $^{-/-}$ mice showed conflicting results between the two studies (Liu *et al.*, 2007; Zou *et al.*, 2015). While Lui *et al*. showed elevated thresholds for f2 frequencies only above 15 KHz, but not at low frequencies, Zou *et al*. showed that the thresholds were elevated for all frequencies. This discrepancy might be a result of different genetic backgrounds of the *Ush2a* $^{-/-}$ mice used where potential modifier genes (such as *Cdh23*^{ahl} allele) could rescue the *Ush2a* $^{-/-}$ inner ear phenotype and thus result in more variable and less severe phenotype reported by Liu *et al.*

By P21, when OHCs are lost from the cochlea, ABR thresholds in *Clrn1*$^{-/-}$ mice are significantly elevated (85-95 dB peSPL), with a delay in latency for all four peaks (at 8, 16 and 32 kHz) compared to wild type mice (Geng *et al.*, 2009). In addition, the interpeak latencies P1-P2 and P1-P3 show to be prolonged in mutant mice compared to controls and by P30 no hearing function is detected (Geng *et al.*, 2009), indicating that the mice are deaf around that age, coinciding with the loss of IHCs. These results suggest that *Clrn1*$^{-/-}$ mice have some auditory function at young age, which deteriorates rapidly after P21 and demonstrates hair cell function deficiency, and a neural deficit, necessary for normal sensory transduction. At P21 the mutant mice produced no detectable DPOAE above the noise floor, indicating lack of OHC function (Geng *et al.*, 2009).

1.4. Current genetic strategies to study complex phenotypes

1.4.1. Monogenic disorders and gene identification methods

Monogenic disorders are traditionally defined as resulting from a pathogenic variant/s within a single gene where the variants are both necessary and sufficient to produce the clinical phenotype and to cause the disease (Peltonen and McKusick, 2001) (Figure 1.11a). Large families represent a particularly useful tool for gene identification, because they have multiple affected and unaffected members whose clinical symptoms could be related to a single variant in their genome. Many hearing disorders such as nonsyndromic and syndromic hearing loss are inherited under a monogenic model following a dominant (DFNA loci), recessive (DFNB loci) or X-linked inheritance pattern.

Linkage mapping is a method to identify genetic regions underlying monogenic diseases, relying upon linkage disequilibrium (LD) of genetic markers through large pedigrees (Dawn Teare and Barrett, 2005). Two genetic features are said to be in LD when they are located close enough to not be separated by the recombination process and to always be inherited together. By using genetic markers spaced across the genome and studying their segregation through pedigrees, it is possible to identify chromosome regions that are linked to the disease (where the affected individuals are more similar than expected by chance). The biggest limitation of family linkage is the low level of resolution that it offers. Since linkage studies only narrow down the chromosome location of contributory variants, they usually result in the identification of a large chromosome region rather than a specific variant. Current gene mapping and variant identification methods rely on whole exome and whole genome sequencing techniques (WES and WGS), which offer better resolution and therefore allow detection of near complete variation. These sequencing methods have become relatively cheap and have the advantage of being unbiased regarding the set of genes analysed, allowing parallel examination of most/all of the genes in the human

genome. They also allow identification at the level of novel and very rare variants, which was not possible before.

The biggest challenge of WES/WGS has been interpreting the data and ranking the large number of variants in a systematic way to identify potentially causal variants. To aid variant interpretation now available are population datasets, such as the Genome Aggregation Dataset (gnomAD) which provides summary statistics of large scale sequencing data (such as population specific variant frequency) and prediction tools that assess the potential damaging effect on protein level (summarised in Section 2.5.2).

Guidelines for interpretation of variants in molecular genetic testing have also been developed by professional genetic associations such as the American College of Medical Genetics and Genomics (ACMG) (Richards *et al.*, 2015). The recommended criteria for classifying pathogenic variants involve collecting multiple lines of evidence from publicly available databases on allele frequency, functional effect, disease presentation and expected inheritance pattern, computational and predictive data and aligning it to the specifics of the investigated disorder and segregation with the phenotype (Richards *et al.*, 2015). Although challenges remain in classifying non-coding and missense variants (due to limited functional evidence), approaches learnt from monogenic disorders can be translated into polygenic and more complex disorders.

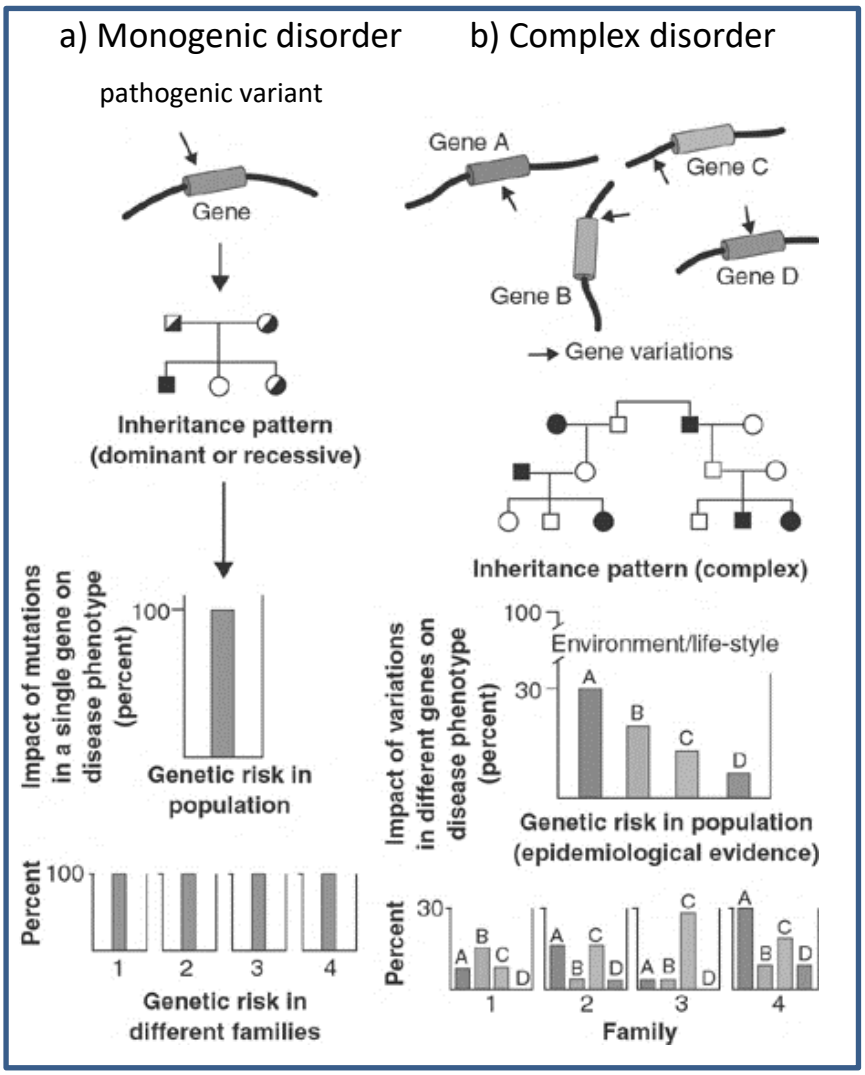

Figure 1.11. Monogenic and complex inheritance patterns and risks (adapted from (Peltonen and McKusick, 2001).
a) In monogenic disorders pathogenic variants in a single gene (within the coding region of the gene) are typically necessary and enough to cause a disease. Pedigrees show Mendelian inheritance patterns. The impact of the pathogenic variant on disease risk is typically 100% (it has a direct effect on disease phenotype) and the risk is the same in different families and within populations.
b) In complex disorders, variants in different genes (within coding or non-coding regions) increase the risk and are typically neither sufficient, nor necessary to explain the disease by themselves. Family trees do not show Mendelian pattern of inheritance and environmental factors act to modify the risk. The effect size of a given variant can differ between individuals and across a population according to their background genetic effects and the environment which they have been exposed to.

1.4.2. Complex disorders and the role of common variants through GWAS (common disease-common variant hypothesis)

Complex genetic disorders are influenced by the combination of risk variants within many genes and environmental factors that can modify that risk (Mitchell, 2012) (Figure 1.11b). For example, complex common disorders such as ARHL, ASD, dyslexia and DLD have an inherited component, described by the term heritability. Heritability shows how much of the variation in a given trait or disorder can be attributed to genetic variation (0% is entirely environmental and 100% - entirely genetic). The heritability estimates for complex disorders typically range from 30% to well over

50%, for example ARHL giving estimates between 35% and 55% and DLD -between 18% and 45% (Gates, Couropmitree and Myers, 1999; Karlsson, Harris and Svartengren, 1997; Wolber *et al.*, 2012; Bishop and Hayiou-Thomas, 2008).

Because of the small effect size of each variant on complex disease, in order to capture risk genome-wide association studies (GWAS) compare frequencies of SNPs between cases and controls in large populations. Higher frequencies in cases compared to controls usually indicate a higher risk for the alleles involved. GWAS are based on the common disease-common variant (CDCV) hypothesis which proposes that if a disease that is heritable is common in the population (a prevalence greater than 1–5%), then the genetic contributors will also be common in the population (Wang *et al.*, 2005). Although for complex diseases such as ARHL and DLD, GWAS have identified variants reaching statistical significance within new candidate genes (*NOP9* and *ROBO2* for DLD and *GRM7*, *PCDH20* and *SLC28A3*, *IGS20* or *ACAN* and *TRIOBP*, and *CLRN2* and *NID2* for ARHL) (Nudel *et al.*, 2014; St Pourcain *et al.*, 2014; Friedman *et al.*, 2009; Vuckovic *et al.*, 2015; Hoffmann *et al.*, 2016; Wells *et al.*, 2019) in the majority of cases findings from GWAS fail to replicate between studies (Tam *et al.*, 2019). Moreover, the variants identified through GWAS are not necessarily functional. This is because they are proxies, which mark the approximate position of the aetiological variant within a small segment of DNA. Therefore, while GWAS provide an informative starting point for functional studies, even high-density SNP association screens will require follow-up studies to enable proof of robust effect (McCarthy and Hirschhorn, 2008).

Table 1.7. Possible frequencies of gene variants in complex traits (adapted from (Cirulli and Goldstein, 2010)

Variant class	Minor allele frequency (MAF)	Implications for analysis
Common	Between 5% and 50%	Detectable in association analysis using GWAS methods
Less common	Between 1% and 5%	Detectable in association analysis using GWAS methods in larger association studies
Rare but not private	Less than 1% but sill polymorphic in one or more major human populations	Detectable in exome/whole genome sequencing studies and co-segregation in families
Very rare and private	Restricted to single families	Difficult to analyse even in co-segregation and whole genome sequencing in affected families

1.4.3. The role of rare coding variants through genome sequencing (common disease-rare variant hypothesis)

Association studies assume the presense of major shared genetic effect, meaning that very rare and private SNPs are overlooked (Table 1.7). Such rare (low-frequency) variants have been proposed as players in susceptibility to complex disease even before the GWAS field was established to capture common risk variants (Pritchard, 2001). This view was based on the theoretical explanation for the importance of rare variants where variants that strongly predispose to disease are likely to be deleterious and therefore kept at low frequencies by

purifying selection (Pritchard, 2001). These rare, usually coding, variants differ from the rare causal variants implicated in Monogenic disorders (described in Section 1.4.1) as they are usually not sufficient to cause disease leading to genetic heterogeneity between individuals, but instead can contribute to risk in complex disorders (with some exceptions like rare variants in *FOXP2* (Lai *et al.*, 2001). While common variants, which have high frequency in the general population (MAF≥ 5%) are expected to have low relative risk, rare highly penetrant variants may confer high risk (as shown in neurodevelopmental disorders- ASD, ADHD, schizophrenia) (Henriksen, Nordgaard and Jansson, 2017; Anney *et al.*, 2012; Yang *et al.*, 2013; Faraone and Larsson, 2019). Therefore, rare risk variants have been considered an important source of variation, leading to the common disease-rare variant hypothesis (Iyengar and Elston, 2007). Within this hypothesis the rare variants are all considered together in a single analysis rather than being restricted to highly penetrant variants. Several statistical methods have been developed to increase statistical power in rare variant association studies by combining information across multiple rare variants within a specified genomic functional unit such as within a gene (gene-based methods that provide multivariate analysis such as burden tests) or across many genes (such as polygenic risk scores) (for review of methods see (Lin *et al.*, 2018).

The role of rare risk variants in neurodevelopmental disoders is now well-recognised. A rare variant burden for protein-truncating variants was reported in a WES study in individuals with complex neurodevelopmental disorders (such as ASD, ADHD, bipolar disorder or intellectual disability) (Ganna *et al.*, 2018), suggesting a widespread genetic effect among those disorders. Another recent large scale exome sequencing study of ASD reported 102 genes in ASD risk (of which 30 were novel) and a greater burden for *de novo* protein-truncating varians over *de novo* missense variants, with all exome *de novo* variants in the autosomes in total explaining 1.92% of the variance in ASD (Satterstrom *et al.*, 2020).

1.4.4. Alternative mechanisms and combinations

It is becoming apparent that rare and common variants at single bases of DNA sequence (SNPs) are unlikely to fully account for the heritability of complex genetic disorders. An alternative mechanism such as copy number variants (CNVs representing deletions and duplications ranging from 50bp to 5Mb in size) has been identified in many neurodevelopmental disorders such as ASD, ADHD, schizophrenia and DLDs (Simpson *et al.*, 2015; Kalnak *et al.*, 2018; Zarrei *et al.*, 2019; Shearer *et al.*, 2014).

Increasing evidence also points to a complex interplay between genes and the environment in neurodevelopmental disorders (Tran and Miyake, 2017). Epigenetic mechanisms (affecting gene expression, without changing the DNA sequence) such as DNA methylation and histone modification can act at this interface (Banik *et al.*, 2017). We need to keep in mind, that although

each environmental risk factor (being driven through epigenetic or other mechanisms) might have a very small effect, when it occurs in the context of genetic background with specific risk variants, the overall combination may result in what's referred to as a "perfect storm", leading to disruption of normal neurodevelopment.

The emerging picture of complex genetic neurodevelopmental disorders, including ASD, ADHD, dyslexia, schizophrenia and DLDs, is that they can arise for many different reasons, each involving different combination of underlying risk. In most cases, we now expect risk models to involve many factors that act together: hundreds of genetic variants together with CNVs, gene x gene interactions, epigenetic modifications and environmental influences. The application of diverse technologies and multi-national collaborative effort is thus required to explore the role of each factor when studying a complex disease in order to better explain not only the genetic contributions, but also our understanding of the biological mechanisms and interactions involved in its aetiology.

1.5. Aims of thesis

The overall aim of this thesis is to comprehensively identify and examine the overlapping genetic mechanisms between hearing, auditory processing, and emergent language.

The specific aims are:

1. To examine the effect of heterozygous pathogenic variants in USH causative genes (as candidate APD susceptibility genes) on hearing, auditory processing, and language under a Mendelian model.

2. To examine the effect of common and rare risk USH variants on hearing, auditory processing, and language abilities under a complex model.

3. To discover new pathways implicated in APD by identifying pathogenic variants with big effect on a genome-wide level in individuals that show difficulties discriminating words in noise.

2. Subjects and Methods

2.1. Subjects and ALSPAC population cohort

2.1.1. Discovery Family

This thesis came about because of a large family affected by APD. The discovery family consisted of 12 members; eight of which (including I.1 and all his seven descendants) (Figure 2.1) were affected by expressive language disorder with acute auditory processing difficulties and disfluent speech (dysarthria), indicating an autosomal dominant inheritance pattern. The affected individuals showed difficulties characteristic of APD such as processing speech and following instructions particularly in presence of background noise. The discovery family is further described by Perrino *et al* (Perrino *et al.*, 2020).

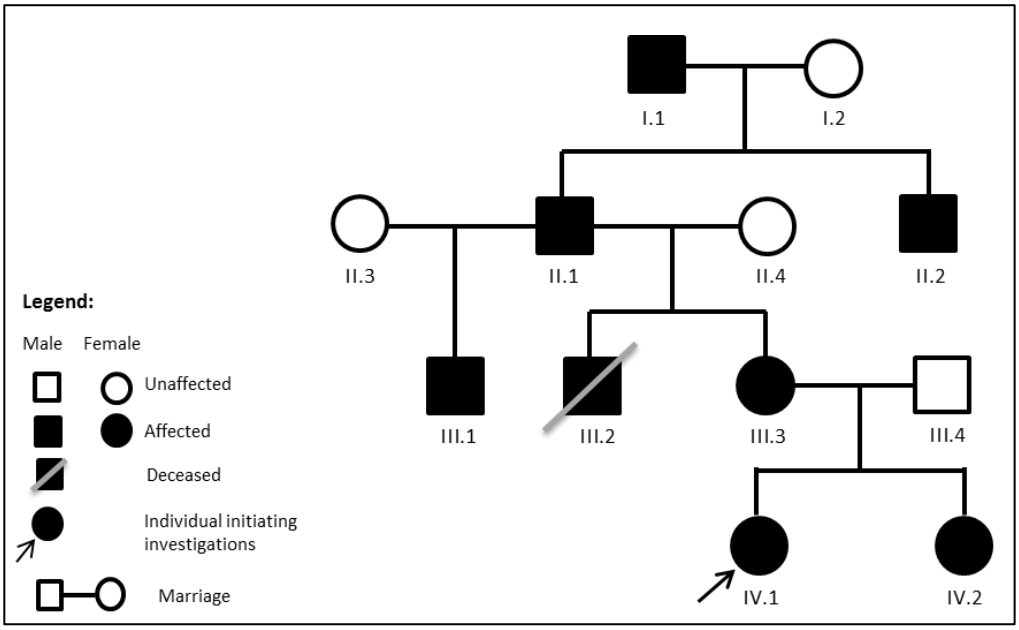

Figure 2.1. Pedigree of the four-generational APD family

2.1.2. ALSPAC population Cohort

Following candidate gene identification in the discovery family (Perrino *et al.*, 2020), genetic investigations were performed in the Avon Longitudinal Study of Parents and Children (ALSPAC). ALSPAC is a population-based birth study of the children born to 14,541 mothers in the Avon area (Boyd *et al.*, 2013). Three datasets generated from the ALSPAC study were used in this thesis. They will be referred to as ALSPAC phenotype dataset, ALSPAC genotype dataset and ALSPAC UK10K dataset. The ALSPAC phenotype dataset provides a wide range of neurodevelopment phenotype measures of all participating children including language, memory, hearing and neuropsychiatric measures (Table S1). The ALSPAC genotype dataset contains genetic data of

common variation, providing 500,527 SNPs genotyped in 8,365 children. The ALSPAC UK10K dataset contains low read depth whole genome sequence data for 1,867 children as part of the UK10K project, which aimed to study the contribution of genes to phenotype traits and disease (Walter *et al.*, 2015) (https://www.uk10k.org/studies/cohorts.html). The ALSPAC cohort profile and study protocol have been described elsewhere (Boyd *et al.*, 2013; Golding, Pembrey and Jones, 2001) and further information is available on the ALSPAC website (http://www.bris.ac.uk/alspac).

2.1.3. Ethical approval

Ethical approval for the discovery family was provided by University of London & St George's University Hospitals. All members gave informed consent/assent of investigation. Access to the ALSPAC and UK10K data for the use of this study was agreed as part of project B2341 by the ALSPAC Ethics and Law Committee and an ethical approval for secondary data analysis was granted by Oxford Brookes University (DREC Reference: 1216_29).

2.2. ALSPAC phenotype dataset

The ALSPAC phenotype dataset contained 684 measures available for this this study (413 qualitative and 271 quantitative) (Table S1), for full list of measures see http://www.bristol.ac.uk/alspac/researchers/our-data/). Performance on all these measures was considered for rare Mendelian analyses utilising the ALSPAC UK10K dataset in Results Chapters 1 and 4. Subsets of the ALSPAC phenotype dataset were also created and are detailed below.

2.2.1. Core phenotypes for complex analyses

Eight core phenotypes were selected from the ALSPAC phenotype dataset for association analyses of common variants on the ALSPAC genotype dataset (Results Chapter 2) and gene-based association analyses on the ALSPAC UK10K dataset (Results Chapter 3). These core measures consisted of three measures of hearing and five measures of language development and are summarised below.

2.2.1.1. Hearing measures

Otitis Media with Effusion status (OME @ 7 years) (ALSPAC variable f7hs062)

Bilateral otitis media with effusion (OME) status ("glue ear") was chosen for analysis because it is a known risk factor for secondary APD (Khavarghazalani *et al.*, 2016) and as such is expected to show some associations with signs and symptoms of APD. OME status was documented on the basis of tympanometry results which were obtained by audiologists and trained staff as part of the ALSPAC study when children were 7 years old. In short, the two ears were tested one by one, with the right ear tested first. A Kamplex AT2 tympanometer was used and the probe was placed

at the entrance of the ear canal, measuring the eardrum mobility and middle ear pressure. A graph, called tympanogram, was produced, indicating how effectively sound is transmitted into the middle ear and thus how well the middle ear functions. Tympanogram tracings were classed as type AA (indicating bilateral normal middle ear function with middle ear pressure of +100 to -100 daPa), Type C1 (indicating bilateral slight eustachian tube dysfunction but no OME with middle ear pressure of -101 to -200 daPa), Type C2/B (indicating unilateral eustachian tube dysfunction with middle ear pressure < 200 daPa (Type C2) or unilateral OME with flat trace (Type B); Type BB (indicating bilateral OME with flat trace). Data for this measure were available for 5,410 genotyped children, with a distribution of 74% of children having Type AA, 11.9%- Type C1, 8.4%- Unilateral C2/B and 5.7%- Bilateral C2/B (Figure 2.2).

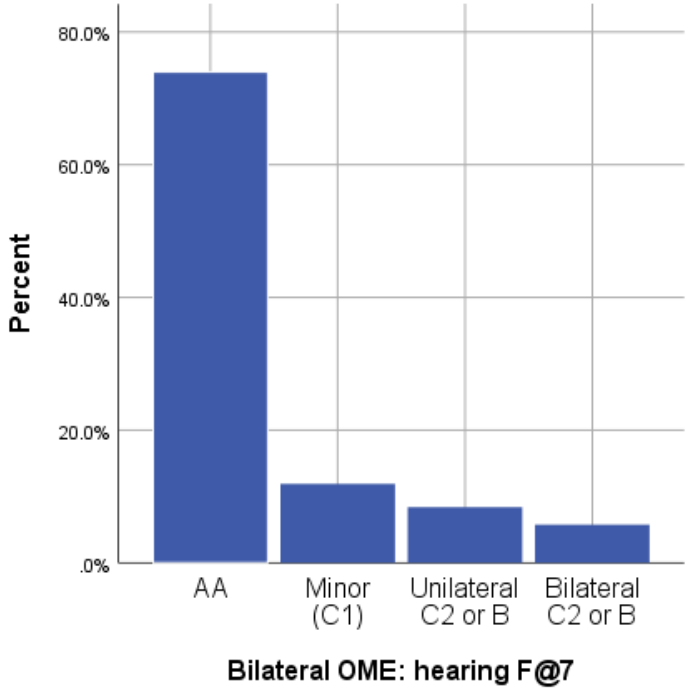

Figure 2.2. Distribution of OME status across the ALSPAC genotype core cohort (N= 5,410 genotyped children).

Low-frequency hearing thresholds at 0.5 kHz frequency (LowFreq_min)

Hearing thresholds at low frequencies (0.5 kHz) were chosen for analysis based on the deficits shown by *Ush2a* heterozygous knock-out mice by Perrino *et al.* (Perrino *et al.*, 2020). Audiometry was performed as per British Society of Audiologists (BSA) standards. Air conduction was performed using either a GSI 61 clinical audiometer or a Kamplex AD12 audiometer. All hearing tests were carried out by audiologists and trained staff in a room with minimal external noise (not exceeding 35 dB) as part of the ALSPAC study. Low frequency hearing thresholds were not directly available as a separate score within ALSPAC and so scores at 0.5kHz were derived through calculations (see Calculations Box 2.1). In short, lowFreq_min hearing thresholds were defined as the minimum air conduction threshold in the better performing ear at 0.5kHz. The measure was

derived from ALSPAC variables f7hs017, f7hs018, f7hs027 and f7hs028 (Calculations Box 2.1), which indicate average thresholds across different frequencies (0.5 kHz, 1kHz, 2kHz, 4kHz, 8kHz and 16kHz in each ear. This measure was available for 4,440 genotyped children with thresholds ranging from -10 dB HL (better performance) to +40 dB HL (poorer performance), with a mean= 10.36 dB HL and SD= 5.7 (Figure 2.3).

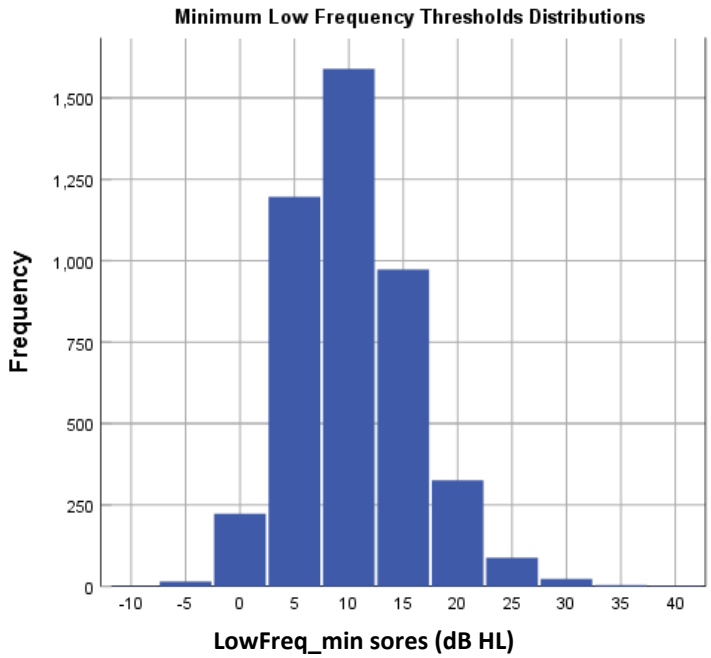

Figure 2.3. Distribution of Low frequency_min hearing scores across the ALSPAC genotype core cohort (N= 4,440 genotyped children).

Calculations Box 2.1.

f7hs017 and f7hs018 are composites for Right ear:
f7hs017 = (f7hs010+f7hs011+f7hs012+f7hs013)/4
f7hs018 = (f7hs011+f7hs012+f7hs013)/3
where f7hs010 is air cond hearing threshold level (dB HL) r ear 500 Hz: hearing F @ 7
where f7hs011 is air cond hearing threshold level (dB HL) r ear 1 kHz: hearing F @ 7
where f7hs012 is air cond hearing threshold level (dB HL) r ear 2 kHz: hearing F @ 7
where f7hs013 is air cond hearing threshold level (dB HL) r ear 4 kHz: hearing F @ 7
f7hs027 and f7hs028 are composites for Left ear:
f7hs027 = (f7hs020+f7hs021+f7hs022+f7hs023)/4
f7hs028 = (f7hs021+f7hs022+f7hs023)/3
where f7hs020 is air cond hearing threshold level (dB HL) l ear 500 Hz: hearing F @ 7
where f7hs021 is air cond hearing threshold level (dB HL) l ear 1 kHz: hearing F @ 7
where f7hs022 is air cond hearing threshold level (dB HL) l ear 2 kHz: hearing F @ 7
where f7hs023 is air cond hearing threshold level (dB HL) l ear 4 kHz: hearing F @ 7
from these, measures of low frequency thresholds were derived:
lowfreqR = (f7hs017*4) -(f7hs018*3)
lowfreqL = (f7hs027*4) -(f7hs028*3)

Mid-frequency hearing thresholds (MidFreq_min)

Mid-frequency hearing thresholds measure was chosen for analysis to cover frequencies in the middle ranges between low and high. The thresholds were defined as the minimum air conduction thresholds in the better performing ear at mid-range hearing frequencies, which were averaged across 1, 2 and 4 kHz (ALSPAC variables f7hs018 and f7hs028 for right and left ear respectively). This measure was available for 4,520 genotyped children with scores ranging from 0 dB HL (better performance) to +40 dB HL (poorer performance), with a mean of 5.55 dB HL and SD of 4.4 (Figure 2.4).

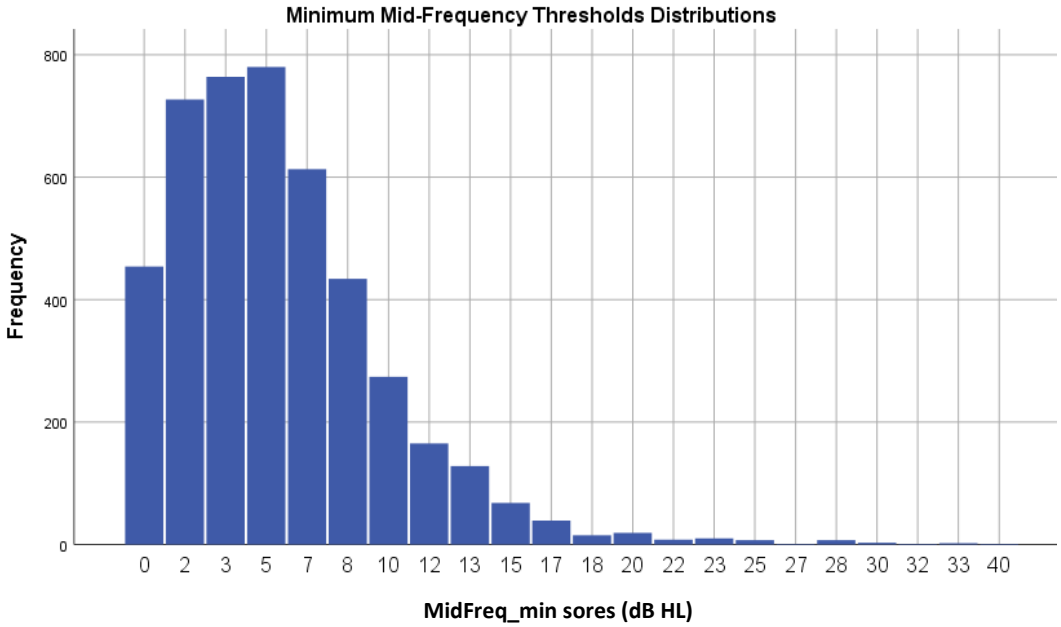

Figure 2.4. Distribution of Mid frequency_min hearing scores across the ALSPAC genotype core cohort (N= 4,520 genotyped children).

2.2.1.2. Language measures

Early communication score (comm @ 18 months) (ALSPAC variable kd654)

The communication score at 18 months (1.5 years of age) was chosen for analysis as a marker of very early language as it starts to develop through signs of communication. It represents a sum of items from a list of 14 tasks, which incorporate measures of hearing (child reacts to ringing bells and people speaking), vocabulary (child copies what you say, child understands and says words and can name colours) and grammar (child combines words and makes negative statements). Data were available for 6,344 genotyped children with scores ranging from 1 (child has done the task 1-2 times) to 28 (child can do well all the tasks) (mean= 15.96, SD= 4.6) (Figure 2.5).

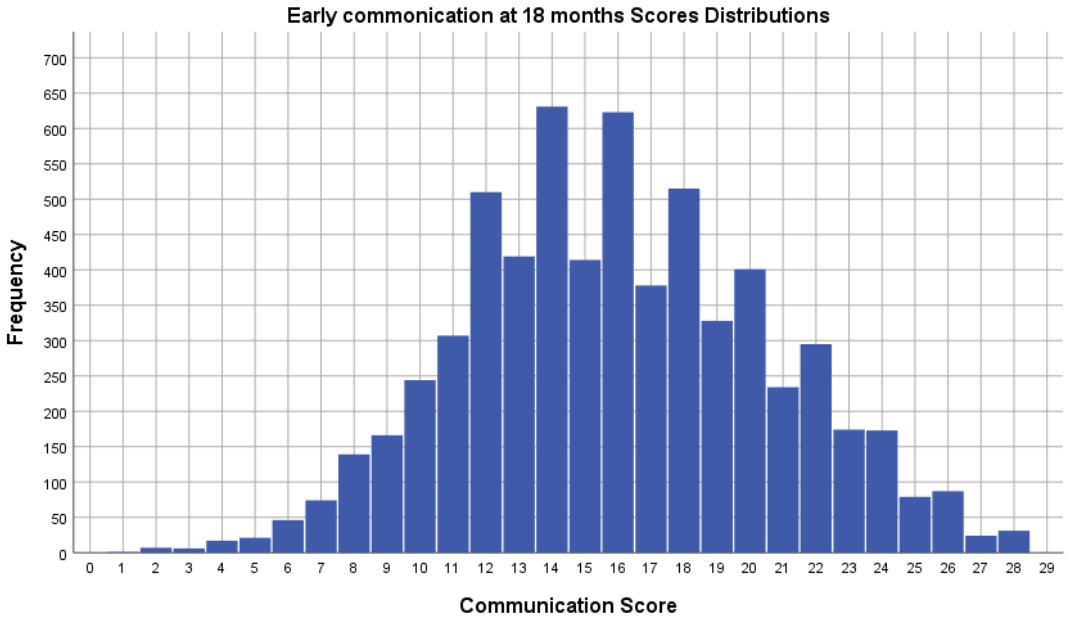

Figure 2.5. Distribution of early communication scores across the ALSPAC genotype core cohort (N= 6,344 genotyped children).

Early vocabulary (vocab @ 38 months) (ALSPAC variable kg865)

The vocabulary score taken at 38 months (3 years of age) was chosen for analysis as it represents an early marker of expressive language. The measure represents a sum of items that a child could use and/or understand, from a list of 123 words (ALSPAC variable kg865). The scores are derived from a parental questionnaire. Data were available for 6,165 genotyped children with scores ranging from 0 (child did not understand or use any of the 123 words) to 246 (child could use and understand all of the 123 words) (mean= 229.8, SD= 29.4) (Figure 2.6). The ceiling effect observed shows that at 3 years of age, most children can use and understand all the words tested.

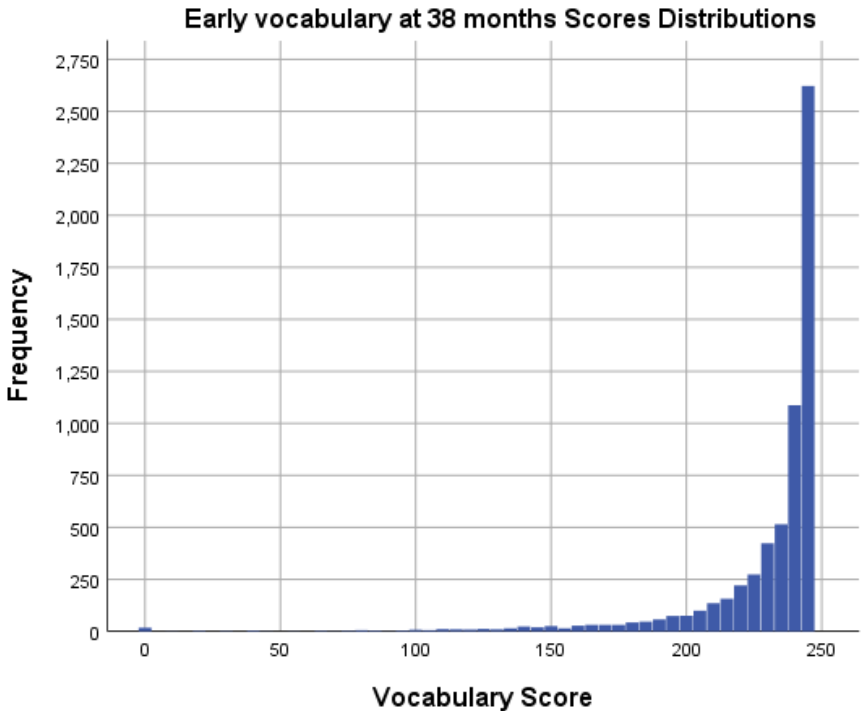

Figure 2.6. Distribution of early vocabulary scores across the ALSPAC genotype core cohort (N= 6,165 genotyped children)

Verbal Intelligence Quotient (VIQ @ 8 years) (ALSAC variable f8ws110)

Verbal IQ was chosen for analysis as a cognitive ability marker of verbal comprehension at school age (8Y). It was measured using the Wechsler Intelligence Scale for Children Version III (WISC III), which is the most widely used individual ability test worldwide for children between the ages of 6 and 16 (Wechsler *et al.,* 1992). For the ALSPAC study, a short form of the measure was employed to reduce the length of the session. Tests were administered by psychology professionals. The Verbal IQ scores were finally derived from the sum of five verbal subtests: information (assessing child's knowledge), similarities (where similarities between things must be explained), arithmetic (mental arithmetic questions), vocabulary (ascertaining child's understanding of the meaning of different words) and comprehension (questions about different situations), which were scaled for age, using look-up tables in the WISC manual. Data were available for 5,218 children and scores ranged from 50 (very low scores classed as mild intellectual disability) to 155 (high scores classed as "gifted") with a mean of 108.25 and SD of 16.5 (Figure 2.7).

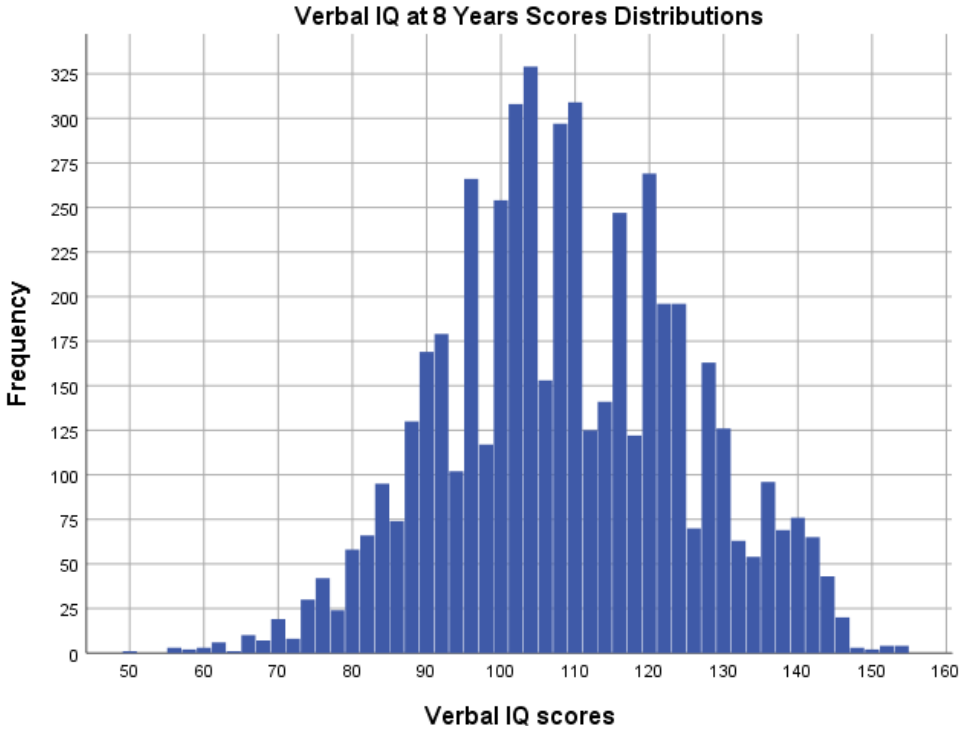

Figure 2.7. Distribution of Verbal IQ scores across the ALSPAC genotype core cohort (N= 5,218 genotyped children)

Nonword repetition (NWR @ 8 years) (ALSPAC variable f8sl105)

Nonword repetition (NWR) was chosen for analysis as it has been shown to be an accurate biomarker of speech and language difficulties (Bishop, Adams and Norbury, 2006) (Newbury *et al.*, 2009). For the ALSPAC study, an adaptation of the Nonword Repetition Test by (Gathercole *et al.*, 1994) was used to assess short term memory. The ALSPAC test was completed in clinic and consisted of 12 nonsense words, split into four of 3, 4 and 5 syllables and conforming to English

rules for sound combinations. The child had to listen to each word via an audio cassette recorder and repeat each item. The repetition attempt was correct if there was no phonological deviation from the target form. The total number of correctly repeated items (including 3, 4 and 5 syllables) was used for analysis. The test was completed at 8 years of age and data were available for 5,229 children, who scored from 0 (no nonsense words repeated correctly) to 12 (all nonsense words repeated correctly) (mean= 7.3, SD= 2.5) (Figure 2.8)

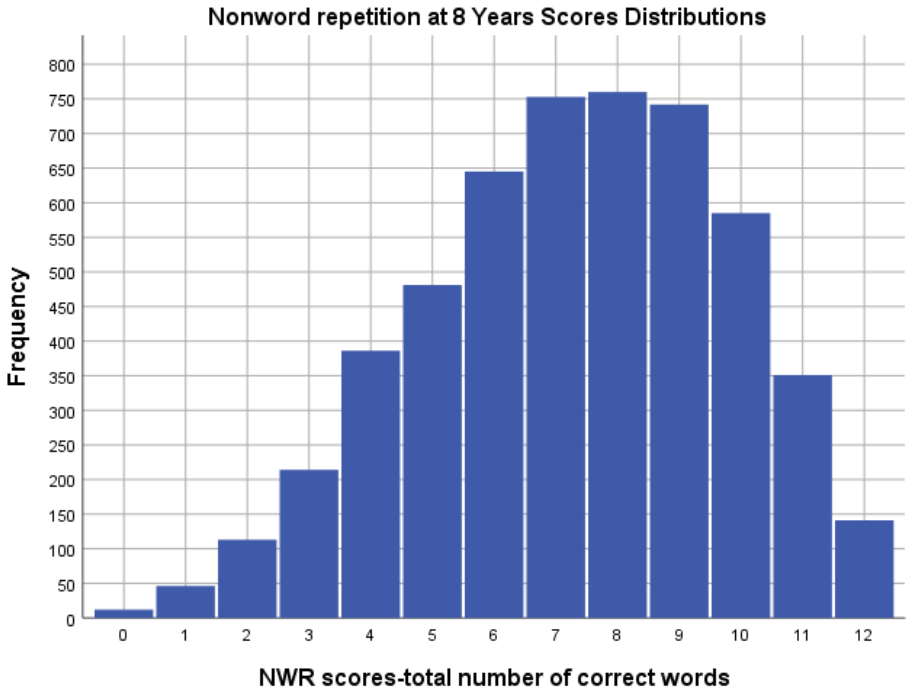

Figure 2.8. Distribution of Nonword repetition scores across the ALSPAC genotype core cohort (N= 5,229 genotyped children)

Developmental Language Disorder Status (DLD)

DLD status was assigned to children based on their performance on language tests, covering comprehension, verbal fluency and syntax as defined in previous publications (Newbury *et al.*, 2009). In short, "case" status was assigned to children who performed at least 1SD below mean on WOLD comprehension (ALSPAC variable f8sl040) OR had CCC verbal fluency AND syntax (ALSPAC variables ku503b and ku504b respectively) ≥ 1SD below mean with no evidence for ASD or hearing impairment. Typically developing "controls" were selected to perform above expected levels across all three selected measures (WOLD comprehension, CCC verbal fluency and CCC syntax) and had nonverbal (performance) IQ > 80 with no documented neurodevelopmental disorders or special education needs. The ALSPAC genotype core cohort included 2,114 controls and 731 cases from a total of 2,845 genotyped individuals (Figure 2.9).

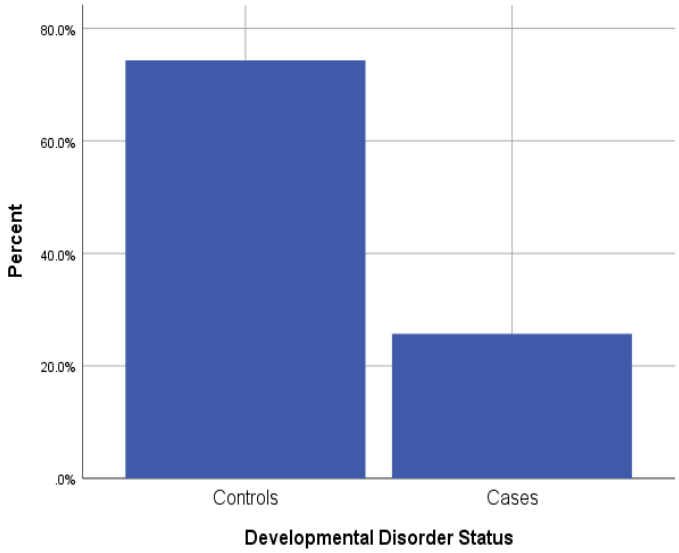

Figure 2.9. Distribution of DLD status across the ALSPAC genotype core cohort (N= 2,845 genotyped children)

2.2.2. Core phenotypes for selecting a suspected APD group

The performance of ALSPAC UK10K individuals on two further measures was compared to select a group of children with features of APD (Results Chapter 4). As the most frequently reported characteristic of individuals with APD is problems with understanding speech in challenging listening conditions, despite normal hearing sensitivity (Bamiou, Musiek and Luxon, 2001; Chermak, 2002; Keith, 1999; Vanniasegaram, Cohen and Rosen, 2004), the ALSPAC phenotype database was scanned for hearing measures that captured speech identification in quiet and in noise as they together, can reveal potential auditory deficits with understanding words only when in noisy environments, but not in quiet. The only hearing measure available in quiet and in noise conditions was the word discrimination threshold (cf573 in quiet and cf577 in noise), obtained as part of the Institute of Hearing Research (IHR)- McCormick Automated Toy Test (ATT) hearing measures taken at 61 months (5 years old). The ATT is the most commonly used test of speech recognition in preschool children within the UK and it measures the minimum sound level at which a child can identify words presented in quiet (Ousey *et al.*, 1989; Palmer, Sheppard and Marshall, 1991; Summerfield *et al.*, 1994).

2.2.2.1. Word discrimination threshold in Quiet (ALSPAC variable cf573) and word discrimination threshold in Noise (ALSPAC variable cf577) @ 61 months

The word discrimination threshold provides a direct measure of the ease with which a child can identify speech in quiet versus noise and is a surrogate measure of auditory sensitivity (Ousey *et al.*, 1989). The test involves children attempting to identify which of 7 pairs of toys is requested by a pre-recorded message from an audio speaker at varying volume levels. The objects are in pairs with similar sounds (cup/duck, tree/key, man/lam, fork/horse, spoon/shoe, cow/house and

plate/plane), so that an error most commonly arises when a toy is confused with its pair member (Summerfield *et al.*, 1994). For the word discrimination in noise an output of a pink noise generator is mixed with the stimulus word (Summerfield *et al.*, 1994). An effort is made to ensure that children are familiar with all of the objects before the test is started and, in the event, that some were unknown, the pairs were removed, and the remaining pairs included. The volume of the instructions increased and decreased 6 times in order to find the level at which the child could hear. The two measures were available for 180 ALSPAC children with scores ranging from 15 (better performance) to 34 (poorer performance) with a mean of 24.37 and SD of 4.48 for word discrimination in quiet (Figure 2.10) and scores from 53 (better performance) to 68 (poorer performance) with a mean of 58.03 and SD of 2.32 for word discrimination in noise (Figure 2.11).

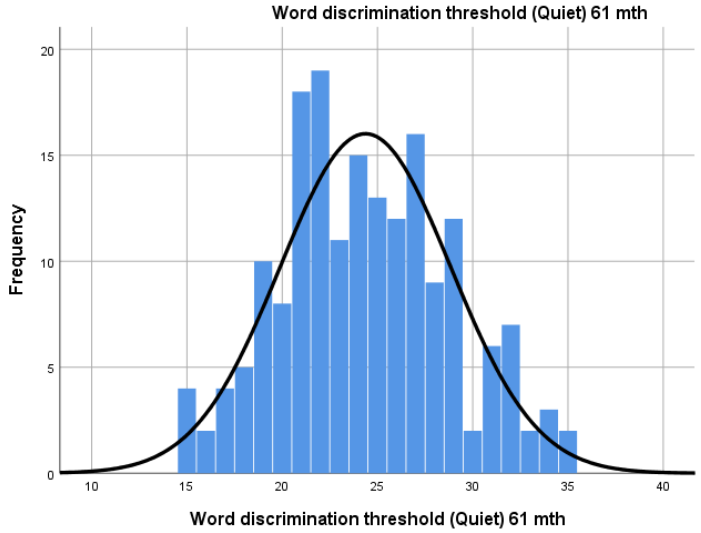

Figure 2.10. Distribution of Word discrimination threshold (Quiet) at 61 months across 180 ALSPAC children with available scores (Mean= 24.37, SD= 4.483)

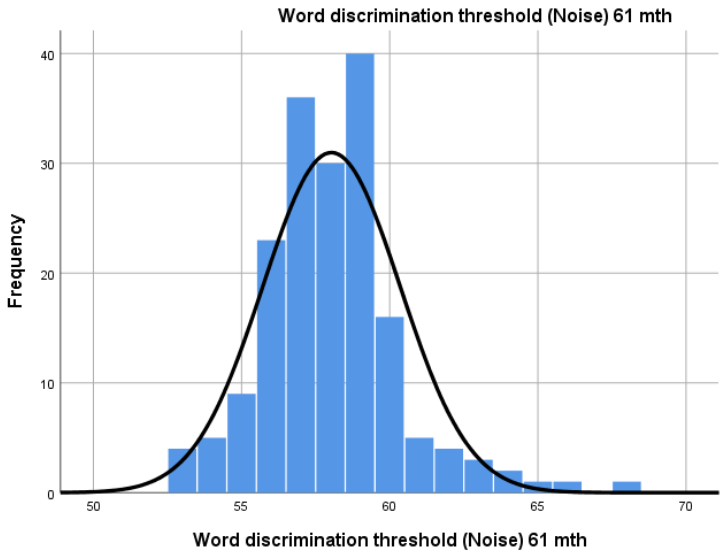

Figure 2.11. Distribution of Word discrimination threshold (Noise) at 61 months across 180 ALSPAC children with available scores (Mean= 58.03, SD= 2.318)

2.3. ALSPAC genetic datasets and core cohorts

2.3.1. ALSPAC genotype dataset generation

The ALSPAC children had been genotyped previous to this work, using the Illumina Human Hap 550-quad chip (Illumina Inc., San Diego, CA), following an already described protocol (Boyd *et al.*, 2013). After quality control based on gender mismatches; minimal or excessive heterozygosity; disproportionate level of individual missingness (> 3%); evidence of cryptic relatedness (> 10% IBD) and selection for European descent, 8,365 children were kept within the study. Quality control measures removed all rare SNPs with a minor allele frequency of < 1% (because rare SNPs are not informative in a gene association study), a call rate of < 95% and the ones which violated the Hardy-Weinberg equilibrium ($p< 5x10^{-7}$), resulting in 500,527 SNPs. These quality control steps were performed by ALSPAC prior to data release.

2.3.2. ALSPAC genotype core cohort selection

As this thesis focuses upon auditory processing and speech and language difficulties, the ALSPAC genotype dataset (N=8,365) was filtered to exclude confounding factors (Table 2.1). Children were excluded if they were a twin, not of British ethnicity, who were bilingual, born earlier than 32 weeks gestation, and having a birth weight < 1500g, who were missing a large amount of data and had an overt pathology that could confound language development (non-verbal IQ <65) and hearing loss (hearing thresholds above 40dB) (Table 2.1). These filtering criteria resulted in a selected cohort of 7,141 children (3,615M and 3,526F), which will be referred to as ALSPAC genotype core cohort and is used in Results Chapter 2. The number of individuals provides 80% power to detect a variant that explains at least 0.38% of the traits' variance at a Bonferroni-corrected alpha level of 1.13×10^{-05} (Figure S1).

Table 2.1. Exclusion criteria to generate ALSPAC genotype core cohort of 7,141 individuals

Exclusion criteria used ALSPAC code	**Reason for exclusion**	**No. inds excluded**	**% exclusions**	**No. inds remaining**
				15,444
no gtyp data	Not informative	7,165	46.39%	8,279
Twins MZ010>1 (kept A if genotyped) [1]	The genetic effect is falsely over-represented	51	0.62%	8,228
English not main language[2] f8sl200 not=1	May confound language assessment scores as bilingual children can be slower to develop language (Hoff *et al.*, 2012)	7	0.09%	8,221
child uses other language regularly[2] f8sl201=1	May confound language assessment scores (same as above)	104	1.27%	8,117
English not mother tongue[2] cf442=2	May confound language assessment scores (same as above)	9	0.11%	8,108
more than one language spoken at home[2] cf443>1	May confound language assessment scores (same as above)	17	0.21%	8,091
English not first language[2] plascc24=2	May confound language assessment scores (same as above)	7	0.09%	8,084
ethnicity not British[3] plasca20 not=1	Confounding ethnic factor	4	0.05%	8,080
Premature bestgest[4] < 32	May confound overall development as premature children are at risk of reduced cognitive scores (Bhutta *et al.*, 2002) (Zimmerman, 2018) SLI to DLD criteria development references to add	39	0.48%	8,041
low birth weight[4] KZ030<1500	May confound overall development (Zimmerman, 2018)	6	0.07%	8,035
premature AND low birth weight[4] KZ030<2000 AND bestgest < 34	May confound overall development (Zimmerman, 2018)	22	0.27%	8,013
no phenotypes more than 114 of 116 core phenotypes data missing	Missing data above acceptable threshold, does not allow informative conclusion	67	0.84%	7,946
Missing data from 16 essential hearing & language phenotypes[5]	Missing data above acceptable threshold, does not allow informative conclusion	663	8.34%	7,283
PIQ (performance IQ)[6] <65 F8ws111<65	May confound language assessment scores-add reference	78	1.07%	7,205
Individuals presenting with moderate to severe hearing loss[7] (hearing thresholds >40 dB HL)	May confound auditory processes- add reference	64	0.89%	7,141
Total		**8,303**	**60.79%**	**7,141**

1. In multiple pregnancies only 1 genotyped baby was kept in core cohort; **2**. Children, whose main language at home was different from English and whose mother tongue and/or first language was not English, who used other languages regularly and spoke more than one language at home, were all excluded; **3**. Children who were not of white ethnicity were excluded; **4**. Children who were born before 32 weeks were excluded, together with children born weighing less than 1.5kg and children born before 34 weeks with birth weight less than 2kg were all excluded; **5**. Hearing & language phenotypes included: six developmental problems that required special arrangement at school, reading and spelling @ 7Y, WOLD comprehension, nonword rep, Verbal IQ and Performance IQ @ 8Y, Vocab score @ 38 mths, high frequency hearing loss, bilateral OME, low and mid-frequency derived scores; **6**. Children with Performance IQ scoring less than 65 were excluded; **7**. Children with hearing thresholds >40 dB HL, representing moderate to severe hearing loss were excluded.

2.3.3. ALSPAC UK10K dataset generation

The ALSPAC UK10K sequence dataset was generated previous to this study and full details are available within the supplementary data of other studies (Taylor *et al.*, 2015; Timpson *et al.*, 2014). In short, low-read-depth whole genome sequencing was performed on 2,040 individuals using Illumina HiSeq platform (low read depth at an average 7x depth). Sequencing reads that passed quality control were aligned to the GRCh37 human reference, using BWA (v0.5.9- r16) (Li and Durbin, 2010). This resulted in 1,976 sequenced samples that went through the variant calling procedure, generated using samtools/bcftools (version 0.1.18-r579) (Danecek *et al.*, 2011). Following a standard filtering pipeline and quality control steps, 1,867 samples remained for analysis.

2.3.4. ALSPAC UK10K core cohort selection

The ALSPAC UK10K dataset was filtered on the same criteria as ALSPAC genotype dataset (Section 2.3.2), but individuals with hearing loss and neurodevelopmental difficulties (last two filtering steps in Table 2.1) were left in. This produced a selected cohort of 1,681 individuals (806M: 875F), which will be referred to as ALSPAC UK10K core cohort and is used in Results Chapters 1 and 3.

2.3.5. ALSPAC suspected APD cohort selection

This ALSPAC UK10K core cohort was further filtered to produce the suspected APD (sAPD) cohort: individuals who were not subjected to the word discrimination hearing tests and those who had missing data for the two core phenotypes of word discrimination threshold in quiet (cf573) and in noise (cf577) (Section 2.2.2) were excluded. Because the two word discrimination thresholds (cf573 and cf577) were tested only in a sub-cohort of the ALSPAC UK10K cohort, called Children in Focus, the number of available children dropped significantly, resulting in 180 individuals remaining (Table 2.2). To exclude the confounding effect of hearing loss, all individuals who presented with moderate to severe hearing loss were excluded (those who had hearing thresholds f7hs017, f7hs018, f7hs028, f7hs028 >40 Db), resulting in 177 individuals remaining (Table 2.2). Lastly, only children whose word discrimination thresholds in quiet were considered typical within a normative range (cf573 ≤29 which excludes individuals with scores 1SD above the mean) and whose word discrimination thresholds in noise were elevated (cf577 ≥61 which includes individuals with scores 1SD above the mean) were included, resulting in 13 individuals in the sAPD cohort (7%) (Table 2.2). The sAPD cohort is used in Result Chapter 4.

Table 2.2. Exclusion criteria to generate APD suspected core cohort

Exclusion criteria used ALSPAC code	N children remaining
ALSPAC UK10K core cohort	1681
Individuals with available cf573 and cf577 scores	180
Individuals presenting with moderate to severe hearing loss (hearing thresholds f7hs017, f7hs018, f7hs028, f7hs028 >40 Db)	177
cf577 ≥61 and cf573 ≤29	13

2.4. Genetic analyses

The genetic investigations on the discovery family (2.4.1) were completed prior to the current PhD project and therefore the PhD candidate was not involved in the analyses. All the follow-up analyses, including the ALSPAC association analyses of common variants (2.4.2), the ALSPAC UK10K gene-based association analyses (2.4.3) and the sAPD cohort coding variant analysis (2.4.4) were completed by the PhD candidate herself and are further explained below.

2.4.1. Discovery family genetic investigations

All genetic analyses on the discovery family were performed by the Newbury group at the Wellcome Centre for Human Genetics, Oxford and formed the basis of the current PhD project. Full details are provided by Perrino *et al.* (Perrino *et al.*, 2020). In short, seven members of the discovery family (five affected individuals: II.1, II.2, III.3, IV.1 and IV.2, and two unaffected individuals, II.4, III.4, Figure 2.1) were genotyped to allow identification of shared chromosome segments between affected individuals and to identify copy number variants (CNVs). Two individuals (II.2 and IV.1, Figure 2.1) underwent whole genome sequencing, enabling the identification of possibly pathogenic variants. Candidate variants were further validated by Sanger sequencing using BigDye (v3.1). Data from these analyses enabled the identification of a set of candidate genes that set the foundation of this thesis.

2.4.2. ALSPAC association analyses of common variants

The ALSPAC genotype data were filtered for the purpose of Results Chapter 2 to include variants with a minor allele frequency > 5% within the 11 Usher genes of interest , which were pruned within PLINK (Section 2.5.1) to obtain a pairwise tagging SNP set with R2< 0.8. (Table 2.3). The UCSC genome browser was used to position the Usher genes to the human genome based on the NCBI RefSeq genes track (GRCh37 assembly).

Table 2.3. Usher genes of interest- chromosome positions and length of studied regions

chr	gene	transcript -/+ 10Kb	length of region covered (bp)	No SNPs after pruning
1	*USH2A*	chr1:215,786,236-216,606,738	820,502	129
3	*CLRN1*	chr3:150,633,950-150,700,786	66,836	8
5	*ADGRV1*	chr5:89,844,617-90,470,033	625,416	66
5	*HARS*	chr5: 140,043,490- 140,081,312	37,822	Gene too small-0 SNPs*
9	*WHRN*	chr9:117,154,360-117,277,736	123,376	21
10	*PCDH15*	chr10:55,552,533-56,571,051	1,018,518	135
10	*CDH23*	chr10:73,146,691-73,585,704	439,013	127
11	*MYO7A*	chr11:76,829,310-76,936,286	106,976	22
11	*USH1C*	chr11:17,505,442-17,575,963	70,521	34
15	*CIB2*	chr15:78,386,948-78,433,877	46,929	11
17	*USH1G*	chr17:72,902,176-72,929,351	27,175	Gene too small- 0 SNPs*

0 SNPs* as genes are too small and not spanned by any common SNPs (MAF>0.05) after pruning using Illumina Human Hap 550-quad chip assay.

The eight core phenotypes of hearing and language (Section 2.2.1) were tested for association with the pruned common variants within the USH genes (Table 2.3) using a linear model of regression for quantitative traits and a logistic model for discrete traits within PLINK (additive model). Bonferroni correction for multiple testing was used to calculate the significance threshold using the formula: 0.05/number of tested traits/number of tested pruned SNPs. Regional association plots were generated with LocusZoom (http://locuszoom.org).

Gene-environment (GxE) interactions were modelled within PLINK at the gene level to further investigate the relationships between Usher genes and speech and language outcomes. Within this model, the ALSPAC genotype core cohort was analysed by adding low frequency hearing as an interaction factor within a linear regression model (Y= b0 + b1.ADD + b2COV1 + b3.ADD.COV1 + e) (interaction model)(Purcell *et al.*, 2007). As an output, p-values were reported for each term. ADDxCOV p-values were used to detect interaction as they represent the interaction between the SNP and covariate factor and show if together they exert a stronger effect on the phenotype than expected through the linear addition of their individual effect. These analyses were performed on the five language outcome measures (early communication, early vocabulary, VIQ, NWR and DLD status).

2.4.3. ALSPAC UK10K gene-based association analyses

The sequence data from ALSPAC UK10K includes all variants across the genome. This data was filtered to include only the 11 genes of interest and only SNPs, resulting in 33,452 variants for analysis (Figure 2.12). The resulting variants had an allele count of at least 1 in the sample set (meaning each variant was carried by at least 1 individual), with a minimum base quality of 20, a minimum mean depth of 3 across samples and Hardy Weinberg equilibrium p value of > $1x10^{-5}$. The transition-transversion ratio of the variants was 2.3.

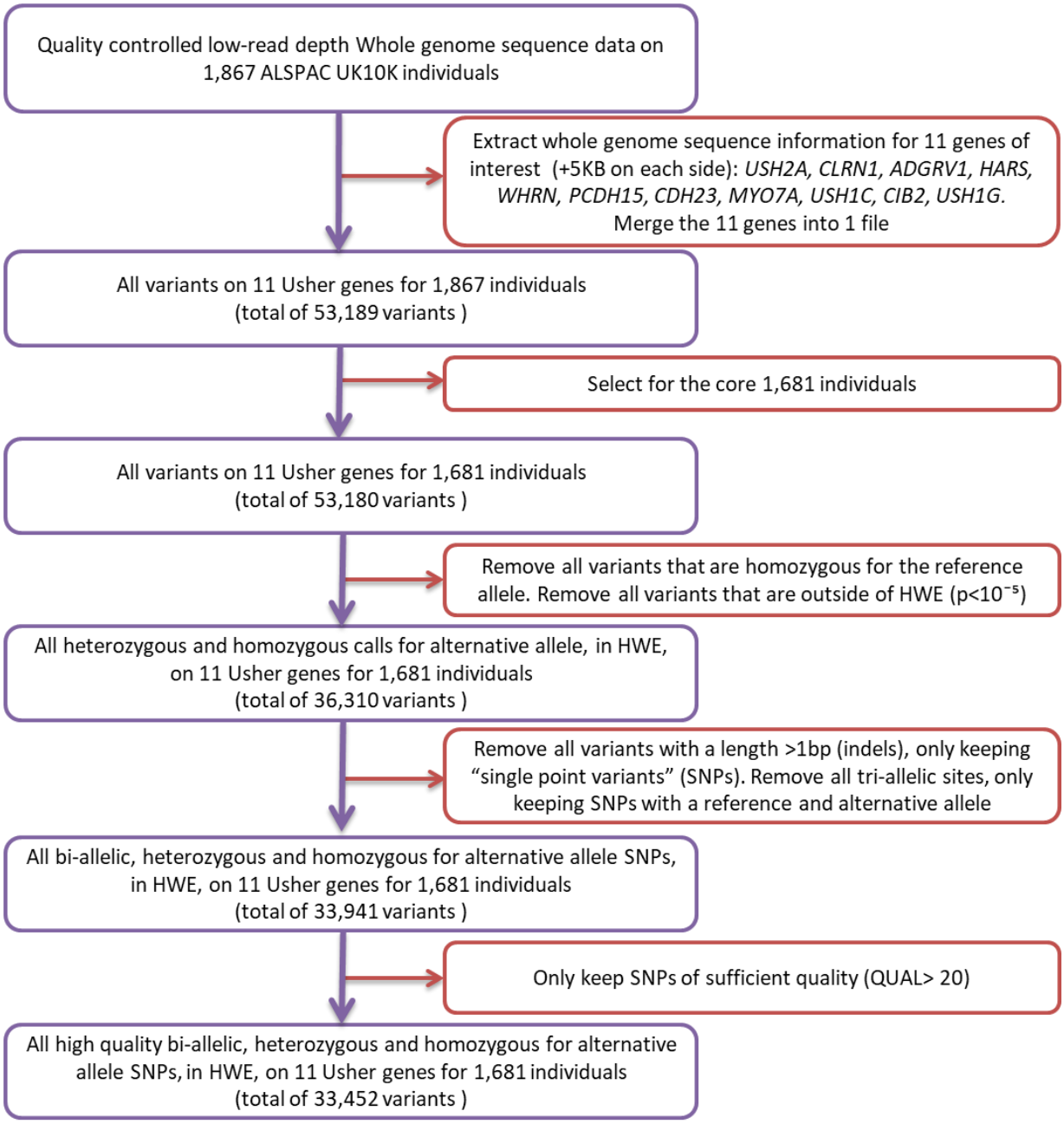

Figure 2.12. A flow diagram showing the filtering steps of ALSPAC UK10K core cohort data. Purple boxes show each filter step and red boxes describe the inclusion/exclusion criteria involved in each step.

The eight core phenotypes of hearing and language (Section 2.2.1) were tested for association with rare variants (MAF≤1%) within the 11 Usher genes of interest (Figure 2.12) under Burden-Zeggini model for gene-based associations within RVTESTS (Section 2.5.1). Analysing the eight core phenotypes for all rare variants yielded a Bonferroni significance threshold of P= 0.05/8 phenotypes/11 Usher genes= 5.68×10^{-4} at an alpha level of 0.05.

As gene-based association analyses collapse the effects of many variants into one, it is difficult to distinguish the group of variants that are most strongly contributing to any observed association. Further in-depth analyses were therefore performed to isolate possible groups of rare driver variants in genes with evidence of association. The in-depth analyses included grouping rare

variants by function (1) (into rare intronic, regulatory and exonic and if applicable further sub-grouping the exonic variants into missense, stop-gained, splice-site and synonymous) and by location (2) (Figure. 2.13). For grouping variants by function (1), gene variant annotations and functional effect predictions were performed through SnpEff (Section 2.5.1.) and were based on the canonical gene transcript. Gene variants were grouped into exonic (located within exons), intronic (located within introns) and regulatory (located within 5- and 3-prime untranslated regions). Variants within the former two groups could be further divided into variant type such as missense, stop, frameshift, splice-site, synonymous for exonic (which were not filtered out because of their possible effect on splicing), and introns towards the 5-prime end and towards the 3-prime end for intronic variants. Finally, to group gene variants by position (2), the Haploview programme was used (Barrett *et al.*, 2005), which generates LD information and haplotype blocks, so variants within the same LD block were grouped together. Due to the memory limitation of Haploview to analyse genes that contain more than 1,000 variants, larger genes (*USH2A* and *ADGRV1*) were manually grouped into 5' and 3' -ends.

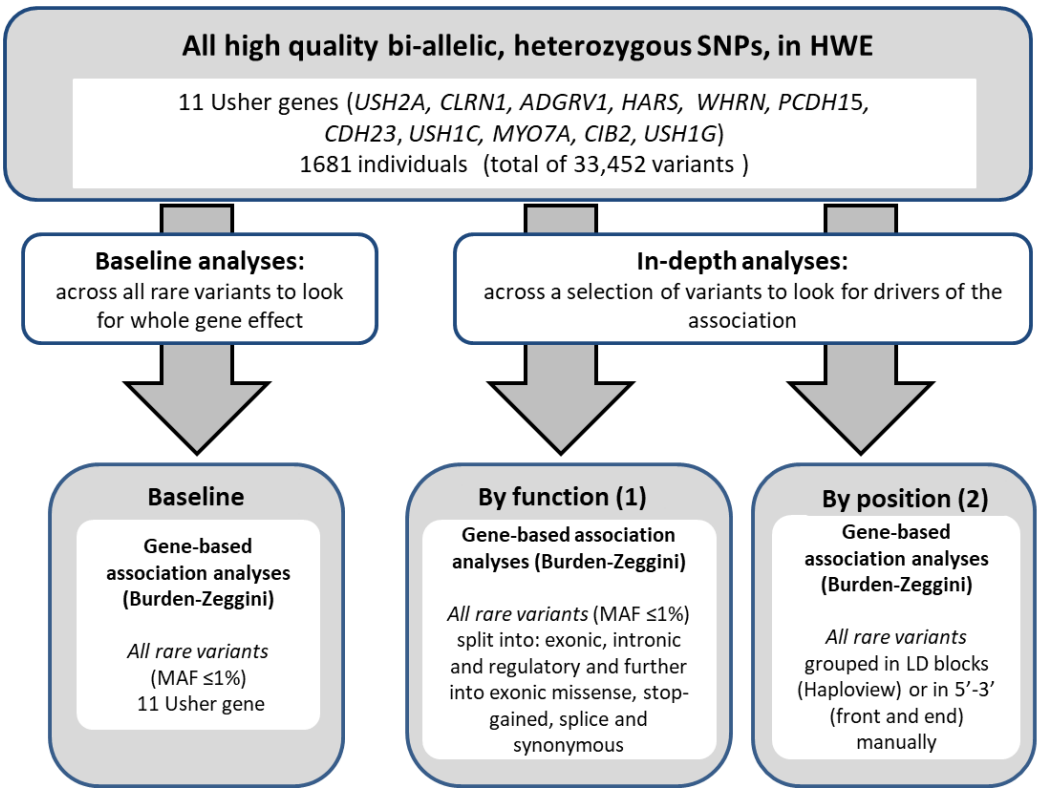

Figure 2.13. A flow diagram for gene-based associations in the investigated Usher genes, divided into baseline and in-depth analyses.

2.4.4. ALSPAC UK10K data- sAPD cohort coding variant analysis

The ALSPAC UK10K sequence data included all variants across the 13 sAPD individuals. In order to identify whether this set of data included potentially pathogenic variants in individuals with sAPD, the following pipeline was used: all variants which were homozygous for the reference allele were removed (using VCFtools- Section 2.5.1) and the potential functional relevance of the remaining variants was annotated using SnpEff (v.4.3) (Section 2.5.1) (Figure 2.14). The variants were then filtered to only include sites within exons or potential splicing site regions (coding variants) (Figure 2.14). These coding variants included missense, splice, frameshift, stop loss/gain, start loss, exon loss and transcription factor ablation site. To identify novel/very rare variants that were most likely to be deleterious, very stringent criteria were used where variants were excluded if they were reported in the gnomAD Non-Finnish European population (Figure 2.14) (including 55,860 individuals which allows for the detection of variants with an expected population frequency of 8.95×10^{-6}). The pathogenicity of the novel remaining variants (Table S7) was then ranked according to ACMG guidelines (Richards *et al.*, 2015; Abou Tayoun *et al.*, 2018) and only pathogenic or likely pathogenic variants were prioritised for follow-up (Figure 2.14). As an addition to supporting evidence PP4 (gene involved in a disorder with affected hearing), a list of known deafness genes in humans and/or mice was used as compiled by Lewis *et al.* (Table S8) (Lewis *et al.*, 2018).

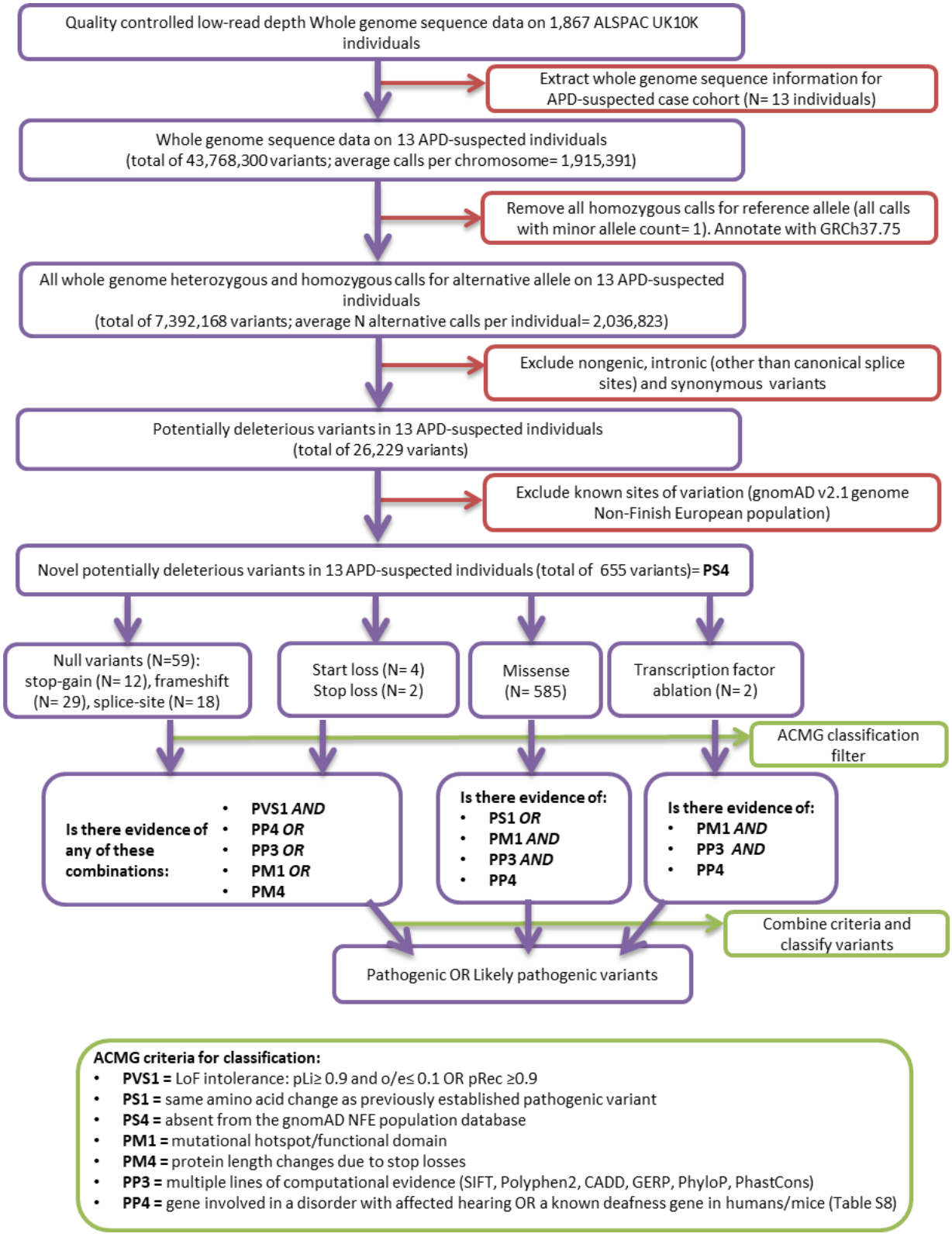

Figure 2.14. A flow diagram showing the filtering of the UK10K whole genome data for sAPD individuals.

2.5. Tools for data analysis and interpretation

2.5.1. Bioinformatics packages

PLINK is an open-source package, developed to facilitate the analysis of genome-wide association studies and can effectively analyse very large datasets for hundreds of thousands of SNPs across large sample sizes (Purcell *et al.*, 2007). Therefore, PLINK (v1.90) was selected for association analyses of common variants in this work and all analyses were run within a Linux environment.

RVTESTS is a software package which allows gene-based collapsing analysis to be performed (Zhan *et al.*, 2016), in this way identifying whether qualifying rare variants across the 11 candidate Usher genes were, in aggregate, associated with a hearing and/or language phenotype. The Burden-Zeggini (BZ) model was employed upon all variants with MAF≤1% to test for increased burden (MAF was based on the frequency of each variant in the 1,681 ALSPAC UK10K core cohort). BZ works by aggregating counts of rare variants into 1 variable for each individual and tests whether the variable is associated with phenotype variation on a single phenotype basis (Morris and Zeggini, 2010). As this approach works under the assumption that a large proportion of the variants are causal and have the same direction of effect, it is a powerful method to detect a group of rare variants with a deleterious effect.

VCFtools is a programme package designed for working with variant call format (vcf) files (Danecek *et al.*, 2011), generated from sequencing studies, such as the ALSPAC UK10K datasets. A range of VCFtools commands were used to summarise and filter out data, to compare and merge files, to run calculations and summarise variants.

SnpEff (v.4.3) is a genetic variant annotation and functional prediction toolbox, designed to work with vcf files (Cingolani *et al.*, 2012). It was used for simple annotations of variants within ALSPAC genotype and ALSPAC UK10K datasets in order to locate each variant within the human genome and to predict its effect.

wAnnovar is a web-based rapid and efficient tool that uses the ANNOVAR software to annotate functional consequences of genetic variation (http://wannovar.wglab.org/) (Chang and Wang, 2012). It was used to link population allele frequency information for investigated variants and to predict functional effect to the protein.

2.5.2. *In Silico* tools

2.5.2.1. ClinVar and gnomAD

Pathogenicity in Results Chapter 1 was defined by a clinical significance of 4 or 5 of reported variants in ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/). Each identified Usher pathogenic mutation from the ALSPAC UK10K cohort was annotated with genome information

retrieved from the UCSC Genome Browser (http://genome.ucsc.edu/), which included genome position, identification number, transcript details, reference and alternative allele, functional effect, DNA and protein change. ClinVar database was used to interpret the clinical significance of each mutation and gnomAD (v2.1) browser (http://gnomad.broadinstitute.org/) was used to retrieve information on the allele frequency of each mutation.

2.5.2.2. Constraint scores

For loss of function gene classification in Results Chapter 4, the following constraint scores were taken into account: pLI, pRec and o/e ratios (all accessed through the gnomAD browser at http://gnomad.broadinstitute.org/). The probability of being loss-of-function (LoF) intolerant (pLI) separates genes of sufficient length into LoF intolerant (pLI ≥ 0.9) or LoF tolerant (pLI ≤ 0.1) categories. pRec shows the probability of being intolerant to two loss of function variants within the same gene (recessive inheritance). pRec≥ 0.9 is the cut-off for highly intolerant genes. The o/e LoF ratio is a score showing a ratio of observed/expected individuals with that particular variant. The scale of o/e is the opposite of pLI, where low o/e values are indicative of strong intolerance. A gene with LoF o/e score of 0.1 will be interpreted as a gene where only 10% of the expected LoF variants were observed and therefore is likely under selection against LoF variants.

2.5.2.3. Putative functional effect, splicing defect prediction and evolutionary conservation

Putative functional effects of associated variants in Results Chapter 4 were evaluated using Sorting Intolerant From Tolerant (SIFT), Polymorphism Phenotyping v2 (PolyPhen2), Combined Annotation Dependent Depletion (CADD) score and Genomic Evolutionary Rate Profiling (GERP) score, while splicing-altering single nucleotide variants were predicted using ada- and rf-scores, and evolutionary conservation of the affected nucleotide was assessed across 100 vertebrate species using the phastCons and phyloP conservation scores from wAnnovar.

To assign pathogenicity, SIFT and PolyPhen-2 algorithms consider protein sequences, motifs and structures and thus can only be used for coding changes and missense mutations in particular (Ng and Henikoff, 2001; Adzhubei *et al.*, 2010). SIFT scores range between 0 and 1, where amino acid substitutions are classified as “deleterious” for scores ≤ 0.05 and “tolerated” when > 0.05. PolyPhen-2 provides scores for two models: HumDiv, more appropriate for identification of rare alleles at loci involved in complex phenotypes, and HumVar, more appropriate for distinguishing mutations with drastic effect in Mendelian disorders. For both models, PolyPhen-2 scores range between 0 and 1, where scores ≥ 0.95 predict a “deleterious” effect. Functional effects are evaluated as “benign”, “possibly damaging” and “probably damaging”, based on pairs of false positive rate thresholds.

The CADD score integrates many diverse annotations into a single, quantitative score by contrasting variants that survived natural selection with simulated mutations (Kircher *et al.*, 2014). CADD measures deleteriousness of single nucleotide variants, as well as insertions/deletions variants in the human genome. CADD scores strongly correlate with allelic diversity, pathogenicity of both coding and non-coding variants, and experimentally measured regulatory effects, and also highly rank causal variants within individual genome sequences. PHRED-like (-10*log10(rank/total)) scaled CADD score ranks a variant relative to all possible substitutions of the human genome ($8.6x10^9$). A scaled CADD score of greater of equal 10 indicates that these are predicted to be the 10% most deleterious substitutions that you can do to the human genome, whereas scores of greater or equal 30 are predicted to be the 0.1% most deleterious possible substitutions in the human genome.

The GERP score is defined as the reduction in the number of substitutions in the multi-species sequence alignment compared to the neutral expectation (Davydov *et al.*, 2010). For example, a GERP score of 4 would mean there are 4 fewer substitutions at a particular site than what is expected based on the neutral rate of evolution across the phylogeny. As such, the GERP score is a measure of sequence conservation across multiple species. Thus, positive scores represent a substitution deficit (which would be expected for sites under selective constraint), while negative scores represent a substitution surplus.

Splicing ada and rf scores are splicing-change predictions for splicing consensus SNPs (scSNVs) across the human genome, based on adaptive boosting and random forest ensemble learning methods, respectively (Jian, Boerwinkle and Liu, 2014). A variant is considered to be splice altering when both predictions are taken into account and both are higher than the optimum cut off value of 0.6 (Jian, Boerwinkle and Liu, 2014).

PhastCons is based on a hidden Markov model (a statistical model of sequence evolution) which shows the probability of each nucleotide belonging to a conserved element, considering flanking alignment columns (Siepel *et al.*, 2005). Scores range between 0 and 1, where 0 shows divergent bases and 1 conserved. PhyloP measures conservation at individual alignment columns and disregards the effect of their neighbours (Pollard *et al.*, 2010). Scores range between -14 and +6, where negative values represent nucleotide bases that have undergone accelerated evolution and positive values represent conserved bases.

2.5.3. Statistical analyses tools

The performance of Usher pathogenic carriers (N= 17 individuals) (Section 3.2.1 for description) and sAPD group (N= 13 individuals) on neurodevelopmental measures (including language, reading, cognition, educational support, neurodevelopmental disorders and hearing) was

compared to control groups. Those included individuals that were known non-carriers of Usher heterozygous pathogenic variants (N= 1,664 non-carriers not found to carry a heterozygous pathogenic variant in the 11 USH genes) and individuals not suspected for APD according to the same selection criteria as the sAPD group (N= 163 unsuspected individuals). Descriptive statistics to show score distributions of quantitative and discrete measures for the two groups (Usher carriers vs non-carriers and sAPD group vs unsuspected group) was performed in IBM SPSS, version 25. The expected normal range between 5^{th} and 95^{th} percentile was computed for each tested quantitative measure in the control groups so individuals from the Usher carriers and the sAPD groups performing outside the normal range could be identified. Effect sizes were computed as standardised mean difference (Cohen's d) with 95% confidence interval in SPSS (Cohen, 1998). The effect size for the tested discrete measures was calculated as Relative Risk (RR) with 95% confidence interval using a web-based effect-size calculator (https://campbellcollaboration.org/research-resources/effect-size-calculator.html).

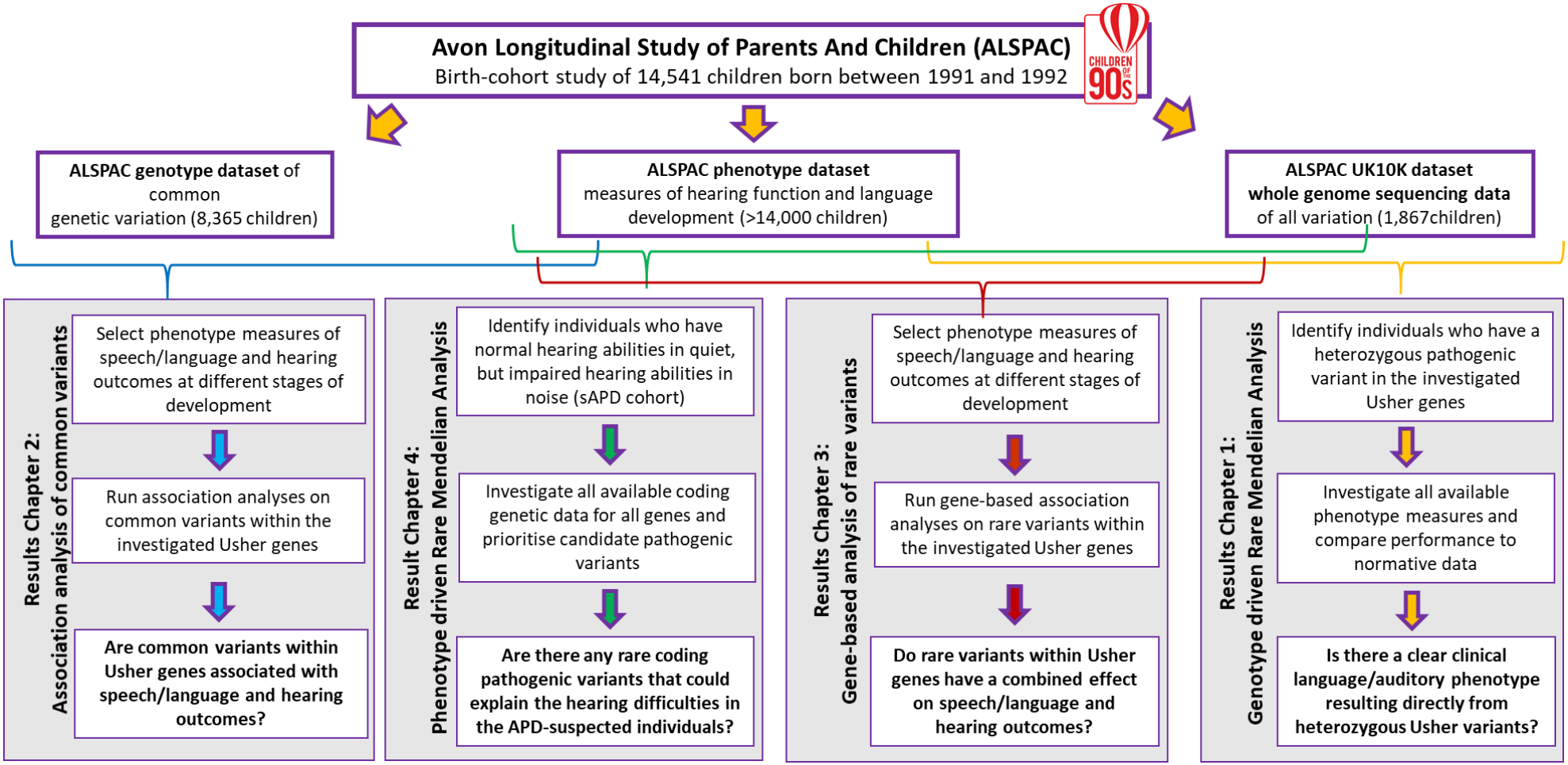

Figure 2.15. Workflow of Result Chapters 1-4.

3. Results Chapter 1: Genotype-driven rare Mendelian analysis

3.1. Rationale

Preliminary genetic investigations of a discovery family affected by a severe language disorder with AP difficulties (Figure 2.1) showed a stop-gain heterozygous pathogenic variant in the *USH2A* gene as further described by (Perrino *et al.*, 2020). The apparent AD inheritance pattern in the family (Figure 2.1) matched with the observed stop-gain heterozygous *USH2A* pathogenic variant, which co-segregated in the family (Perrino *et al.*, 2020). At a population level, Perrino *et al*. showed that these *USH2A* variants increased the risk of delayed language milestones, but they alone did not result in a discernible carrier phenotype (Perrino *et al.*, 2020). Given these findings, and the known role of usherin in stereocilia development in Usher syndrome (Section 1.3.4.), we hypothesised that pathogenic heterozygous variants in other USH causing genes (*MYO7A*, *CDH23*, *PCDH15*, *USH1C, USH1G, CIB2, USH2A*, *ADGRV1, WHRN* and *CLRN1*) have a similar subtle effect on developmental profiles in other carrier individuals (Hypothesis 1: H1). The atypical USH gene *HARS* was added to the hypothesis as a novel USH3B player.

To test H1, a genotype-driven Mendelian approach was utilised where individuals who carried variants previously identified as pathogenic in Usher syndrome were identified from the ALSPAC UK10K core cohort (N= 1,681) (Section 2.3.4). Individuals found to carry a pathogenic USH variant were compared to non-carrier controls on a range of neurodevelopmental measures to identify potential clinical phenotypes arising from carrying the variant (Figure 2.15).

3.2. Results

3.2.1. Pathogenic Usher carriers

Seventeen UK10K individuals (14 M: 3F, from 1,681 individuals with sequence and phenotypic data available, 1.01% detection rate) were identified as carriers of known USH gene variants that were designated as "pathogenic" in ClinVar. These consisted of eight distinct variants which were always detected in a heterozygous form in *USH2A* (relating to USH2 syndrome) and *MYO7A*, *USH1C* or *CDH23* (relating to USH1 syndrome) (Table 3.1). No second hit was observed in either *USH2A*, *MYO7A*, *USH1C* or *CDH23*, indicating that all 17 individuals were carriers. None of the 17 individuals had a genetic diagnosis of Usher syndrome. This is not surprising as the number of individuals tested here (1,681) is not large enough to detect USH cases with reported population frequency of 4- 17 in 100,000 individuals (meaning that 1 case will be detected in a population of 5,882 individuals) (Kimberling *et al.*, 2010). The variant found in the discovery family (rs765476745) was not present in the UK10K samples (Table 3.1). No pathogenic variants were found in the rest of the USH genes tested (*ADGRV1*, *WHRN*, *CLRN1*, *PCDH15*, *USH1G*, *CIB2* and *HARS*) across the UK10K sample. It needs to be noted that the 1,681 individuals in the ALSPAC UK10K core cohort were not excluded on basis of hearing or neurodevelopmental impairment (Section 2.3.4) and as such they represent an entire population spectrum that will include individuals with deficits in hearing, language and cognition.

Table 3.1. Heterozygous pathogenic mutations identified across *USH2A, MYO7A, USH1C and CDH23* in individuals from ALSPAC UK10K core cohort

Usher gene	N carriers	Genome Location[1] [hg19]	SNP ID[2]	Transcript variant ID[3]	Ref[4]	Alt[5]	QUAL score[6]	Clinical significance (ClinVar)[7]	MAF (gnomAD browser)[8]	Functional effect[9]	DNA change[10]	Amino Acid change[11]
USH2A	1	chr1:215956104	rs111033264	NM_206933.2	A	G	24.8	pathogenic: USH2A	0.0000003	missense	c.10561T>C	p.Trp3521Arg
	5	chr1:215963510	rs148660051	NM_206933.2	C	T	247	pathogenic: USH2A and AR retinitis pigmentosa	0.000004	missense	c.10073G>A	p.Cys3358Tyr
	1	chr1:216019240	rs397518041	NM_206933.2	C	T	46.4	pathogenic: USH2A	0.00002	stop-gained	c.8981G>A	p.Trp2994*
	6	chr1:216420436	rs80338903	NM_206933.2	C	(-)	967	pathogenic: USH2A and AR retinitis pigmentosa	0.0007	frameshift	c.2299delG	p.Glu767Serfs
	1	chr1:216497582	rs121912600	NM_206933.2	C	A	171	pathogenic: USH2A	0.00004	missense	c.1256G>T	p.Cys419Phe
MYO7A	1	chr11:76867949	rs121965080	NM_000260.3	C	T	114	pathogenic: USH1B	Not reported	missense	c.634C>T	p.Arg212Cys
USH1C	1	chr11:17552955	rs397515359	NM_005709.3	C	CG	15.5	pathogenic: USH1C	0.0003	frameshift	c.238dupC	p.Arg80fs
CDH23	1	chr10:73492049	rs121908351	NM_022124.5	G	A	84.2	pathogenic/likely pathogenic: USH1 and AR non-syndromic deafness	0.000	missense	c.4036G>A	p.Asp1346Asn

1- Genome location according to GRCh37/hg19 assembly.
2- Single nucleotide polymorphism identification number.
3- Transcript variant identification number according to NCBI RNA reference sequence collection (RefSeq).
4- Reference allele on forward (+ strand) of the human genome.
5- Alternative allele on forward (+ strand) of the human genome; (-) represents a deleted base.
6- Phred-scaled quality score: assertion made in Alt i.e. give -10log_10 prob (call in Alt is wrong). High QUAL scores indicate high confidence calls. QUAL> 15 is acceptable.
7- Clinical significance value as recommended by the American College of Medical Genetics and Genomics for variants interpreted for Mendelian disorders.
8- Minor allele frequency: frequency of the second most common allele as reported on gnomAD browser (http://gnomad.broadinstitute.org/).
9- Predicted functional effect of variant on RefSeq transcript.
10- Coding DNA position where the alteration has taken place; (>) represents substitution, (del)- deletion and (dup)- duplication.
11- Protein consequence showing the exact amino acid where the change has occurred; (*) represents a stop codon and (fs)- frameshift.
MAF=0.000 (rs121908351) = variant not reported

3.2.2. Cohort characteristics as a group

Performance on 684 individual measures of neurodevelopment (including measures of hearing and language) (Table S1) was compared between the 17 USH carriers and 1,664 non-carriers in order to characterise the carrier individuals at a broad neurodevelopmental level and to find areas where they showed difficulties (performed below expected).

Characterising the USH carriers at a broad neurodevelopmental level showed that, as a group, the USH carriers performed below expected on a measure of early vocabulary at 3 years (ALSPAC code kg865) (Cohen's d= 0.52, 95% CI= 0.03- 1.01) and word combination at 3 years (ALSPAC code kg868) (Cohen's d= 0.84, 95% CI= 0.34- 1.33) compared to non-carriers (Table 3.2a) (Figure 3.1a and 3.1b). Although the effect size of the difference between the means of the USH carrier group and the non-carrier group is considered as medium to large for vocabulary and word discrimination scores respectively, the 95% CI for both measures spans the value of 0 (95% CI= 0.03- 1.01 and 95% CI= 0.34- 1.33 respectively), which indicates lower confidence (<95%) in the effect sizes, possibly a result of the small sample size available.

Analysis of available discrete measures showed that parents/carers of USH carriers were three times more likely to be concerned about their child's speech at 2.5 years of age (RR= 3.34, 95% CI= 1.41- 7.95) and reported higher incidence of stuttering for carrier children at 8 years (RR= 2.31, 95% CI= 0.82- 6.54) (Table 3.3) (Figure 3.1c and 3.1f). Children who were USH carriers were also twice as likely to have problems with talking at 3 years of age, compared to non-carriers (RR= 2.02, 95% CI= 0.98- 4.17) (Table 3.3) (Figure 3.1d). Teachers expressed complaints towards the USH carrier children more often than expected (RR= 2.47, 95% CI= 1.48- 4.13) (Table 3.3) (Figure 3.1e).

Table 3.2a. Quantitative measures of language, reading and cognition in carriers of USH pathogenic variants compared to non-carriers in ALSPAC UK10K core cohort (lower scores represent poor performance)

Measure	Age	Range of scores (carriers)	Mean score (carriers)	N non-carriers	Range of scores(non-carriers)	Mean score (non-carriers)	SD (non-carriers)	5th percentile	N carriers below 5th percentile
Vocabulary score	3 years	64 - 246	219.31	1598	0 - 246	232.40	24.83	186	2 of 16
Plurals score	3 years	7 - 12	10.47	1591	1 - 12	10.32	2.02	6	0 of 15
Past tense score	3 years	11 - 42	35.47	1581	0 - 42	34.23	9.44	13	1 of 16
Word combination score	3 years	0 - 26	19.13	1590	0 - 26	22.79	4.31	15	4 of 16
Language score	3 years	223 - 324	296.00	1568	94 - 326	300.90	31.49	236.9	2 of 15
Reading score	7 years	14 - 45	28.59	1555	0 - 50	30.10	8.74	15	1 of 17
Spelling score	7 years	3 - 15	7.29	1544	0 - 15	8.36	4.27	2	0 of 17
Nonword Repetition (NWR)	8 years	4 - 10	6.76	1555	0 - 12	7.48	2.46	3	0 of 17
WOLD comprehension	8 years	5 - 13	8.18	1554	2 - 14	7.76	1.90	5	0 of 17
WISC - Verbal IQ	8 years	93 - 130	115.18	1548	54 - 155	112.02	16.75	86	0 of 17
WISC - Performance IQ	8 years	82 - 139	103.29	1547	46 - 145	103.56	16.85	76	0 of 17
WISC - Total IQ	8 years	90 - 135	111.06	1542	46 - 148	109.22	16.21	82	0 of 17

Table 3.2b. Quantitative measures of language, reading and cognition in carriers of USH pathogenic variants compared to non-carriers in ALSPAC UK10K core cohort (higher scores represent poor performance)

Measure	Age	Range of scores (carriers)	Mean score (carriers)	N non-carriers	Range of scores(non-carriers)	Mean score (non-carriers)	SD (non-carriers)	95th percentile	N carriers above 95th percentile
Air conduction Right average 0.5, 1, 2, 4 kHz	7 years	1.25 - 12.5	7.43	1444	-3.75 - 78.75	8.82	6.75	18.75	0 of 17
Air conduction Left average 0.5, 1, 2, 4 kHz	7 years	0 - 16.25	7.35	1443	-8.75 - 66.25	8.64	7.03	20	0 of 17
Low_frequency_min	7 years	5 - 15	10.88	1444	-10 - 50	10.53	6.62	20	0 of 17
Mid_frequency_min	7 years	-1.67 - 11.67	4.12	1489	-8.33 - 71.67	5.41	5.84	15	0 of 17

Table 3.3. Discrete measures of educational support, neurodevelopmental disorders and hearing in carriers of USH pathogenic variants compared to non-carriers in ALSPAC UK10K core cohort

Measure	Age	N affected carriers	Freq in carriers	N affected non-carriers	Freq in non-carriers
Carer worried about child's speech	2.5 years	4 of 16	0.25	118 of 1577	0.075
Child has problems with talking	3 years	5 of 15	0.33	262 of 1589	0.16
OME/abnormal middle ear pressure (< -100 daPa)	7 years	6 of 17	0.35	421 of 1507	0.278
Hearing Impairment	7 years	0 of 17	0	114 of 1489	0.076
Teacher's rating of child is average/below average	7 years	2 of 10	0.2	307 of 857	0.35
Child received complaints from the teacher*	7.5 years	8 of 17	0.47	297 of 1559	0.19
Child currently has uncontrollable tics or twitches	7.5 years	1 of 16	0.63	34 of 1543	0.02
Child has learning difficulties requiring special arrangements at school	7.5 years	0 of 16	0	55 of 1545	0.036
Child has speech problems requiring special arrangements at school	7.5 years	0 of 16	0	13 of 1545	0.008
Child has hearing problems requiring special arrangements at school	7.5 years	0 of 16	0	30 of 1545	0.019
Child has eyesight problems requiring special arrangements at school	7.5 years	0 of 16	0	13 of 1545	0.008
Child has physical problems requiring special arrangements at school	7.5 years	1 of 17	0.059	11 of 1545	0.007
Child has reading difficulties requiring special arrangements at school	7.5 years	0 of 17	0	70 of 1545	0.045
Child has emotional/behavioural problems requiring special arrangements at school	7.5 years	1 of 17	0.059	21 of 1545	0.014
DAWBA DSM-IV clinical diagnosis - Any ADHD disorder	7.5 years	0 of 17	0	20 of 1566	0.013
Child has ever had speech/language therapy	7.5 years	3 of 14	0.214	156 of 1523	0.102
Child stutters/stumbles when speaks	8 years	3 of 17	0.176	119 of 1557	0.07
Mother told child has Dyslexia	9 years	1 of 14	0.07	71 of 1538	0.05
B6b: Mother told child has Dyspraxia	9 years	1 of 14	0.07	28 of 1510	0.02
B6e: Mother told child has Dyscalculia	9 years	0 of 13	0	8 of 1494	0.005

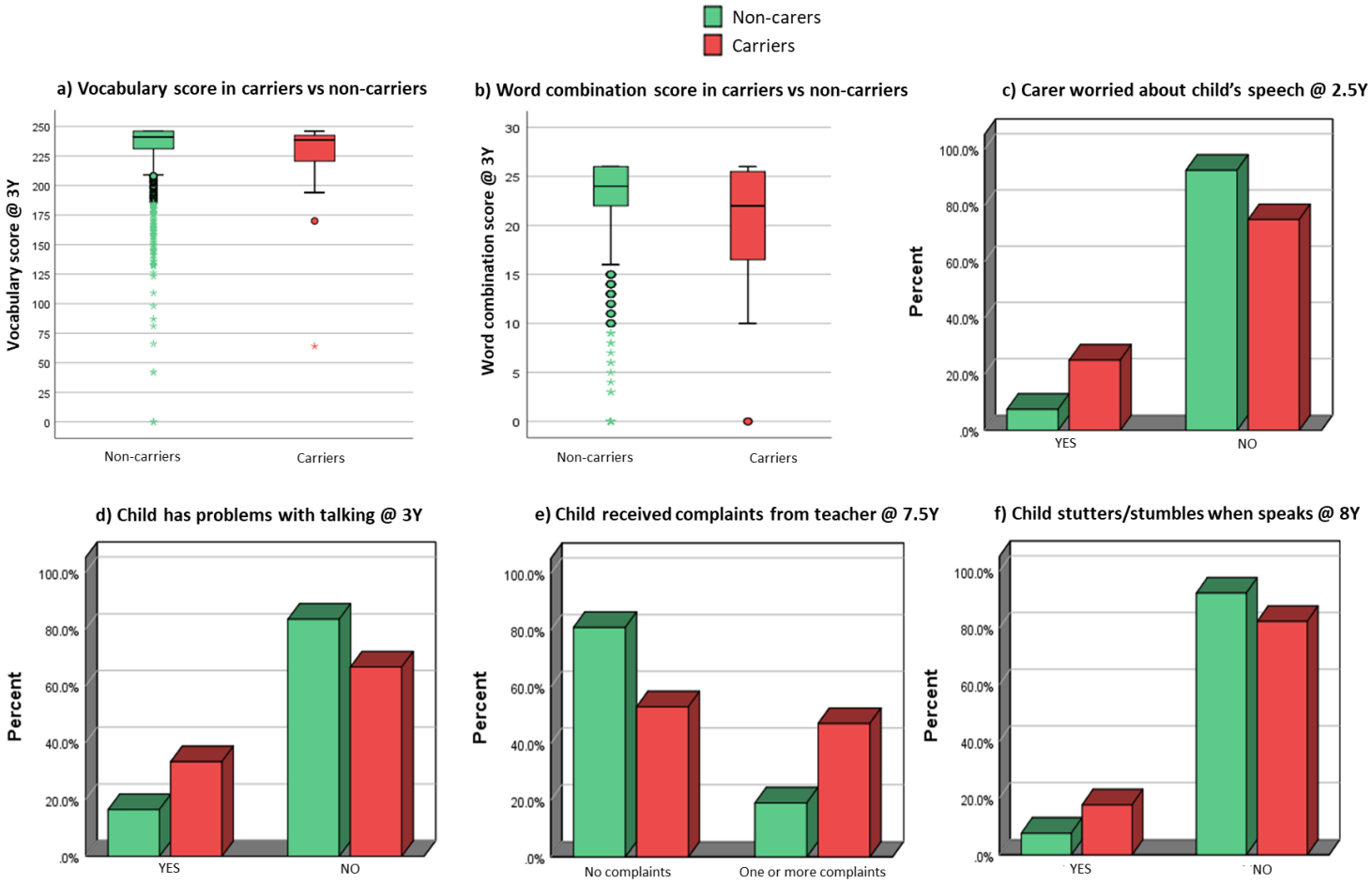

Figure 3.1. Differences in the performance for carriers of USH pathogenic variants compared to non-carriers on a range of language and neurodevelopmental measures.

3.2.3. Individual Usher carriers' characteristics

In order to further characterise and compare the 17 USH carriers on an individual, rather than group level, neurodevelopmental ALSPAC measures where at least 1/17 carriers showed deficits were combined into composite scores across different time points allowing any trends of a recurrent difficulty to be revealed (Table S2-S3). This analysis showed that 11/17 USH carriers (65%) had recurrent ear problems that included earache, ear discharge, red and sore ears, pulling on ears, ear infections and use of ear drops between 6 months and 7.5 years of age (Table 3.4). Furthermore, 10/17 carriers (59%) showed a recurrent hearing problem/concern which could be attributed to an early sign of difficulties processing sounds, concerns over hearing abilities from carers/health professionals, to a referral for a hearing assessment and abnormal middle ear pressure measurements between 15 months and 9 years of age (Table 3.4). It needs to be noted, however, that none of the 17 individuals suffered from overt hearing loss or had a hearing impairment. Six of sixteen carriers (38%) indicated a recurrent problem with speech/talking based on carers expressing concerns about speech development, stuttering/stumbling, use of babbling noise, wrong order of words or had attended speech therapy (Table 3.4). Moreover, 6/17 carriers (35%) showed recurrent signs of speech and language difficulties, defined by vocabulary size, use of grammar, combination of words and intelligibility between the age of 1.5 and 3 years (early scores), reading, spelling and NWR between the age of 7 and 9 years and pragmatic skills at 9.5 years (measured by Children's Communication Checklist- CCC) (Table 3.4). Finally, 11/17 carriers (65%) showed persistent poor scores in a composite measure designated as general development (Table 3.4). The general development measure included early fine and gross motor coordination skills, understanding and communication skills (between 6 months and 1.5 years), IQ scores at 8 years (as defined by Wechsler Intelligence Scale for Children- WISC), problems at school between 7.5 and 13 years (defined by teacher's reports on child's ability, medical diagnosis and special education needs requiring school action) and friendship score at 8 years old. Three of the seventeen USH carriers (17%), of which one had dyslexia and one dyspraxia, were classed as children with special education needs or had received school action (Table 3.4). None of the carriers showed deficits on verbal and non-verbal intelligence (minimum score was low average), with total IQ scores ranging from average to exceptional (Table 3.4 and S2).

To further investigate any genotype-phenotype correlations resulting from the type of USH variant carried by each of the 17 individuals, the USH carriers were grouped into two groups: LoF group (one carrier of *USH2A* stop-gain, six carriers of *USH2A* frameshift and one carrier of *USH1C* frameshift variant) and missense group (seven carriers of *USH2A,* 1 carrier of *MYO7A* and 1 carrier of *CDH23* variant) (Table 3.4). This revealed that 5/8 (55%) LoF USH variant carriers had recurrent problems leading to both ear-related and hearing problems, compared to 3/9 (33%) missense USH

variant carriers (Table 3.4). Further 2/8 (25%) LoF USH variant carriers showed recurrent problems in both areas of talking abilities (reported by parent/carer) and speech and language skills (clinically recorded), while none of the missense USH carriers showed recurrent deficits in both areas (Table 3.4). In contrast 4/8 (50%) LoF USH variant carriers showed recurrent deficits in overall development compared to 7/9 (77%) missense carriers (Table 3.4).

Table 3.4. Heat map of individual deficits (in stop-gain and frameshift- USH pathogenic carriers) across different time points of development.

Ind	Genotype	Ear problems (6m-7.5Y)	Hearing problems/concerns (15m-9Y)	Talking problems/concerns (1.5Y-8Y)	Speech and language (15m- 9.5Y)	General development (6m-13Y)
2081	***USH2A* stop-gain**	ear discharge & pus (1.5 & 3.5Y)	hearing worse during a cold (3.5-4.5Y)	avoids eye contact 3Y|sometimes silent 3Y|stumbles/repeats words (3-8Y)	low spelling score and slow reading speed 9Y	teacher's rating negative 7Y|dyslexia 7.5Y tics/twitches 7.5Y|school action 11-13Y
21131	***USH2A* frame-shift**	earache (1.5-7.5Y)|pulls/scratches ears 1.5Y|red/sore ears 1.5Y|ear infection 4.5Y	hearing worse during a cold (3.5-4.5Y)	avoids eye contact 3Y|stutters/stumbles 3Y	Typical	Typical
18459		earache (2.5Y-3.5Y)|pulls/scratches ears 3.5Y	abnormal tympanometry 7Y	worried about speech 3.5Y|talking and speech problems 3.5Y|stutters/stumbles 3Y	low plurals, tenses, and word combination scores (2-3Y)	Typical
6526		earache (1.5-3.5Y)|pulls/scratches ears when poorly (1.5-3.5Y)|red/sore ears when poorly 1.5Y	hearing worse during a cold 3.5Y|hearing referral 7Y	No speech/talking problem	low vocab and word combination score (2-3Y)	poor social, fine motor and comm skills (6m-1.5Y)| low total dev (6m-1.5Y)|less positive friendships
598		earache (3-5Y)|ear infection (3-4.5Y)	hearing problems 8Y|hearing referral 9Y	talking problems 3Y|speech therapy 7.5Y	low tenses and intelligibility scores (2-3Y)	poor gross motor coordination 6m| PIQ low average
2029		earache 6m	sus hearing problem 15m|abnormal tympanometry 7Y	husky voice 3Y|words in wrong order 3Y|stumbles/repeats words 3Y	Typical	Typical
14200		No ear problems	hearing referral 9Y	husky voice 3Y|words in wrong order 3Y|stumbles/repeats words 3Y	low word combination score 3Y	teacher complaints 7.5Y|PIQ low average
14242	***USH1C* frame-shift**	earache 7.5Y|ear discharge 1.5Y|ear infection 7.5Y	sometimes head turns towards sounds 15m|sus hearing problem 3Y|hearing referral 7Y	worried about speech (2.5-7.5Y)|words in wrong order 3Y|speech therapy 7.5Y	low intelligibility score 3Y|low reading score 9Y|low word comprehension score 8Y	poor communication 1.5Y

red cells= persistent problem, reported on two or more time points OR two different problems reported at one or more timepoints; orange cells= isolated problem reported at one time point OR two problems that are related/happen at the same time point; green cells= no problems reported.

Speech and language composite score includes early scores (15m-3Y)| reading/ spelling/ NWR (7Y-9Y) & pragmatic skills (CCC 9.5Y); **Developme**nt composite score includes early scores (6m-1.5Y) | IQ (WISC 8Y) | teacher reports & SEN (7.5-13Y) | Friendship score (8Y).

Table 3.4 (cont). Heat map of individual deficits (in missense USH pathogenic carriers) across development.

Inds	Geno-type	**Ear problems** (6m-7.5Y)	**Hearing problems/concerns** (15m-9Y)	**Talking problems/concerns** (1.5Y-8Y)	**Speech and language** (15m- 9.5Y)	**General development** (6m-13Y)
3896	***USH2A*** **mis-sense**	earache (2.5-7Y)|red/sore ears when poorly (2.5Y & 3.5Y)|ear infection 4.5Y	No hearing problems	avoids eye contact & words in wrong order 3Y|stumbles/repeats words (3-8Y)	Typical	Typical
5032		earache 1.5Y	sus hearing problem 15m|hearing worse during a cold 2.5Y|abnormal tympanometry 7Y	stutters/stumbles 3Y	Typical	teacher complaints 7.5Y|attention/ activity problem 7.5Y| less positive friendships
6546		earache & pus (1.5-3.5Y)|pulls/scratches ears 1.5Y	hearing worse during a cold 3.5Y|abnormal tympanometry 7Y	worried about speech 2.5Y	Typical	teacher complaints 7.5Y
12321		earache & ear discharge/pus/ mucous (1.5-3.5Y)|pulls/scratches ears when poorly (1.5-2.5Y)|red/sore ears (1.5-3.5Y)	OME 9Y	No speech/talking problem	low reading score 7Y	teacher complaints 7.5Y|school action 11-13Y
16203		Ear drops 15m	hearing referral 7Y	words in wrong order & stutters/stumbles 3Y	low vocabulary, plurals and language scores 3Y| poor pragmatic composite 9.5 Y	poor communication 6m|low total dev 6m| teacher complaints 7.5Y
17275		No ear problems	No hearing problems	Missing	Typical	poor fine and gross motor skills 6m|poor communication 1.5Y|low total dev score (6m-1.5Y)| less positive friendships
19476		earache & pulls/scratches ears (1.5-2.5Y)|red/sore ears when poorly (1.5- 3.5Y)	abnormal tympanometry 7Y|OME 9Y	worried about speech 2.5Y	low past tenses, word combination and language scores 3Y| poor pragmatic composite	poor gross motor coordination 6m| teacher complaints 7.5Y
17155	***MYO7A*** **mis-sense**	earache & ear discharge (1.5-7.5y)|pulls/scratches ears (1.5-3Y)|ear infection (4.5-7Y)	sus hearing problem 3Y |hearing worse during a cold 2.5Y-4.5Y|hearing referral 7Y	worried about speech (1.5-7.5Y)|never babbled & talking problems 3Y	very poor pragmatic composite	poor understand score 15m|teacher complaints 7.5Y|behavioural difficulties 7.5Y|dyspraxia| recognised SEN 7Y|PIQ low average
16009	***CDH23*** **mis-sense**	pulls/scratches ears & red/sore ears 1.5Y	abnormal tympanometry 7Y	worried about speech (2.5-7.5Y)|babbling noises/difficulty while talking 3Y|speech therapy 7.5Y	Typical	poor gross motor coordination 1.5Y|teacher complaints 7.5Y

3.3. Discussion

Identifying USH pathogenic carriers from the ALSPAC UK10K core cohort resulted in detecting 14 individuals carrying *USH2A* pathogenic variants, and a further 3 individuals carrying pathogenic *MYO7A*, *USH1C* and *CDH23* variants (Table 3.1). Comparing developmental profiles between USH carriers and non-carriers showed that as a group the 17 USH carriers had expressive language delays at an early age (3 years) and a higher incidence of stuttering plus increased complaints from teachers in the classroom at the age of 7.5- 8 years (Figure 3.1). Further analysis of individual profiles of USH carriers across different ages indicated common trends amongst the carriers towards recurring subtle difficulties with hearing, language and overall development which in the majority of cases did not lead to a recognised clinical problem (Table 3.4).

3.3.1. Carriers of pathogenic USH gene variants show poor early expressive language as a sign of a delay rather than a deficit

The vocabulary and word combination scores at 3 years of age are early markers of expressive language. The vocabulary measure is a sum of items that a child could use and/or understand, from a list of 123 words (ALSPAC variable kg865) while the word combination measure is a sum of 13 sets of items that a child could combine correctly (ALSPAC variable kg868). The scores for both measures were derived from a parental questionnaire and range from 0 (child did not understand or use any of the 123 words) to 246 (child could use and understand all of the 123 words) for vocabulary, and from 0 (child did not combine correctly any of the 13 items) to 26 (child could combine all 13 items correctly) for word combination. Although as a group the USH pathogenic carriers showed slightly smaller vocabulary size and fewer word combinations than expected, later language measures (Verbal IQ and NWR) were found to be comparable with the non-carrier group (Table 3.2a), suggesting a trend towards a delay in early expressive language.

The measures “carer worried about child’s speech” and “child has problems with talking” (ALSPAC variables kf550 and kg904) are subjective scores that examine signs of speech development in children aged 3 from the perspective of the parent/carer. A greater proportion of the parents/carers looking after children who were pathogenic USH carriers answered “yes, worried about speech” and “yes, has problems with talking” to the two questions, indicating a possible problem with speech/talking at age of 3. However, by 7 years of age, those USH carrier children who had shown a speech/talking difficulty at the age of 3, were reported clear of it and none of them had speech problems requiring special arrangements at school (Table 3.3). This demonstrates that, although there is a trend, the early detected talking and speech problems are likely to represent a speech delay rather than a deficit.

The questionnaires on the "child's speech including stuttering/stumbling" are also aimed at identifying speech problems, but at a later age, at 8 years old, and so can be predictive of speech disorders at school age. Although USH carriers showed a higher risk of stuttering/stammering, the children showing these risks were different individuals from those who showed concerns over speech and talking at 3 years of age. Nevertheless, if we consider the USH carriers as a group, rather than as individuals, the group results show an increased overall risk for speech difficulties. The higher frequency of USH carriers exhibiting this problem could therefore be an additional indicator of a correlation between heterozygous changes and possible speech difficulties, in this case- at school age.

3.3.2. High teacher complaints as a sign of classroom difficulties for USH carriers

The teacher complaint score (kr468b) is a derived sum of complaints from teachers toward students at the age of 7.5 years including reasons such as restlessness or overactivity; poor concentration or being easily distracted; acting without thinking, frequently butting in, or not waiting their turn. A score of 0 denotes no complaints at all and a score of 6- maximum complaints (the higher the score, the higher the number and severity of teacher complaints). Because this is an overall score that could result from one or more reasons, from this measure alone, it is not possible to conclude whether the child received a high complaint score because they were overactive and restless in class, because they showed poor concentration or because they showed impatience (or maybe all three together), which could be signs of neurodevelopmental difficulties in ASD, ADHD, APD, DLD. Therefore, the higher likelihood of higher teacher complaints score for children carriers of USH pathogenic variants can be interpreted as a sign of classroom difficulties for carriers, but is not a phenotype showing a deficit in a specific area of development. It needs to be noted that the USH carriers all had typical hearing sensitivity (as measured by pure tone audiometry- Table 3.2b).

3.3.3. USH carriers show trends for subtle problems with hearing

The identification of subtle difficulties in hearing amongst USH carriers, revealed through analysis of composite neurodevelopmental scores, is not surprising because of the role of Usher related genes in inner ear hair cell development. In fact, it has been long recognised that carriers of genes for deafness exhibit some subclinical abnormalities in pure-tone sensitivity (Anderson and Wedenberg, 1968), which has later been shown to be the case for USH1 and USH2 subtype carriers (Wagenaar *et al.*, 1995; van Aarem *et al.*, 1995). While homozygous pathogenic variants in Usher genes are in general expected to cause Usher syndrome with overt hearing loss (Le Quesne

Stabej *et al.*, 2012), pathogenic heterozygous variants in the same genes can lead to subtle, but persistent changes in auditory processes which over time can build up and result in an increased risk to APD. For example, a heterozygous *Cdh23* mouse knock-out (heterozygous for a presumed null allele of *Cdh23*) showed hearing loss at both low and high frequencies at 5-6 weeks of age, with high frequency component worsening with age (Holme and Steel, 2004). This suggests that *Cdh23* heterozygous pathogenic changes, and not only homozygous, can lead to an auditory phenotype that persists and gets worse over time. A similar model of effect that builds up has been proposed for chronic otitis media with effusion, which is a known risk factor for secondary APD (Khavarghazalani *et al.*, 2016), likely because hearing gets disrupted during an important developmental period.

Hearing and ear-related problems showed to be enriched amongst USH carriers of LoF pathogenic variants (such as stop-gain and frameshift, which may result in truncating the USH proteins), compared to carriers of missense USH variants who showed to suffer less frequently of such problems. While genotype-phenotype correlations have been previously described for USH genes in relation to severity of hearing loss (Bolz *et al.*, 2001; Bork *et al.*, 2001; Astuto *et al.*, 2002; Doucette *et al.*, 2009; Schultz *et al.*, 2011), in the context of increased risk for APD, it is possible that the more severe heterozygous USH variants are more likely to affect hearing, which can in turn affect auditory processes as explained above.

3.4. Conclusion

Results Chapter 1 was based on the hypothesis that heterozygous variants in Usher causing genes have a subtle effect on developmental profiles, similar to the subtle effect of *USH2A* heterozygous pathogenic variants on early language and audition, shown by Perrino *et al.* (Perrino *et al.*, 2020). This hypothesis is confirmed by results indicating that heterozygous *USH2A*, *MYO7A*, *CDH23* and *USH1C* pathogenic variants are associated with subtle problems in hearing and delays in early language milestones, but there is no one measure that could be considered as a "clinical marker of deficit" of USH carriers. Instead, these subtle difficulties might put Usher carriers at a higher risk of developing APD and having delayed language development.

The finding of no one clear phenotype resulting directly from heterozygous pathogenic Usher gene variants is in line with heterogeneous presentations in other disorders. For example, heterogeneity is seen even in cases where variants explain majority of risk (*SATB1* in relation to neurodevelopmental disorder) (den Hoed *et al.*, 2020) and may be due to the genetic background of the individual and the effect of a particular variant on the protein function. Moreover, heterogeneity in neurodevelopmental conditions is thought to be influenced by the small effect size of individual variants which act in complex genetic mechanisms leading to different behavioural phenotypes between individuals (Girirajan *et al.*, 2012; Hemati *et al.*, 2018). This suggests that, as with other neurodevelopmental conditions, the Usher pathogenic variants alone are not causative of APD and do not result in a discernible carrier phenotype as would be expected in a monogenic model. Instead, they might form part of a genetic risk within a complex genetic model. Given these findings, the effect of Usher genes variation on neurodevelopmental ALSPAC measures (such as common variants that form part of complex disorders) was examined in Results Chapter 2.

4. Results Chapter 2: Association analyses of common variants

4.1. Rationale

Results Chapter 1 showed that rare pathogenic heterozygous variants in known Usher syndrome genes *USH2A*, *MYO7A*, *USH1C* and *CDH23* increased the risk of subtle hearing problems and delayed language milestones in some cases but did not directly lead to a distinct carrier phenotype (as would be expected in a monogenic model). This led to Hypothesis 2 (H2) where under the present investigations of effect on hearing, auditory processing and language abilities, we hypothesised that common variants in Usher genes form part of a complex genetic model. According to H2, the effect of each individual common variant is small and not sufficient to fully explain a particular phenotype, but can contribute to susceptibility or altered ability (in line with complex genetic disorders discussed in Section 1.4.2).

To test H2, SNPs tagging common variants within 10 typical USH genes (*MYO7A*, *CDH23*, *PCDH15*, *USH1C, USH1G, CIB2, USH2A*, *ADGRV1, WHRN, CLRN1*) and one atypical USH gene (*HARS*) were analysed for allelic association in the ALSPAC genotype core cohort (N= 7,141) (Section 2.3.2) using PLINK. Three measures of hearing: low-frequency hearing (low freq_min), mid-frequency hearing (mid freq_min) and otitis media with effusion status (OME) (Section 2.2.1.1), and five measures of language: early communication skills (comm), early vocabulary size (vocab), nonword repetition (NWR), Verbal IQ (VIQ), developmental language disorder (DLD) status (Section 2.2.1.2), were assessed for direct association (Section 2.4.2) (Figure 2.15).

Results from Perrino *et al*. showing that heterozygous disruptions of mouse *Ush2a* led to altered low-frequency thresholds, which were further associated with disrupted mouse vocalisations (Perrino *et al.*, 2020), warranted a further investigation into the relationship between low-frequency hearing abilities and language within an interactive model on a population level. For this purpose, GxE interaction modelling was applied to the ALSPAC genotype core cohort (Section 2.4.2), in which low-frequency hearing was included as an interaction factor to the complex model (thus investigating the combined effect of common variants and low-frequency hearing abilities on language).

4.2. Results

Across the eleven USH genes analysed, two genes (*USH2A* and *CLRN1*) contained at least one significant SNP associated with hearing or language outcomes and one further gene (*PCDH15*) contained a cluster of suggestively associated SNPs with hearing and language outcomes (see Table S4 for significance value for all SNPs per gene). Three more genes (*CDH23*, *ADGRV1* and *USH1C*) indicated sporadic association to hearing and/or language outcomes, which were represented by single SNPs, while three other genes (*CIB2, WHRN* and *MYO7A)* showed no association (Table S4). *USH1G* and *HARS* common variants were not analysed as the genes were not covered by the genotype assay. The sections below review each of the significant genes one by one according to SNPs reaching the Bonferroni corrected p-value (Table 4.1).

Table 4.1. Usher genes with P-values of significance for association analyses

Gene	Length of region (Kb)	N pruned SNPs	N tested traits	Corrected P-value of significance (Bonferroni)*
USH2A	820	129	8	4.84E-05
CLRN1	66	8	8	7.81E-04
ADGRV1	625	66	8	9.47E-05
HARS	37,8	0	8	Gene not covered
WHRN	123	21	8	2.98E-04
PCDH15	1,018	135	8	4.63E-05
CDH23	439	127	8	4.92E-05
MYO7A	106	22	8	2.84E-04
USH1C	70	34	8	1.84E-04
CIB2	46	11	8	5.68E-04
USH1G	27	0	8	Gene not covered

* Bonferroni significance calculated at an alpha level of 0.05

4.2.1. *USH2A* common variants are directly associated with low-frequency hearing outcomes and indirectly associated with early language milestones

A cluster of SNPs located towards the 5'end of *USH2A* (between intron 4 and 13 on NM_206933.2 transcript) showed association specifically with low-frequency hearing thresholds (Table 4.2 & Figure 4.1). The top associated SNP (rs10864237, p= 2.92 x 10^{-5}) explained 0.39% of variance in low-frequency hearing thresholds (ßSE= 0.1), representing a 0.9dB difference between risk (T/T genotype group with low-frequency $thresholds_{mean}$= 10.77 dB) and non-risk individuals (C/C genotype group with low-frequency $thresholds_{mean}$= 9.864 dB) (Table 4.2 & Figure 4.3). This difference in low frequency hearing thresholds was found to be statically significant, $t(2466)= 3.1$, p= 0.0018 (Figure 4.3). These analyses show that common *USH2A* variants exert a small effect on low-frequency hearing thresholds within the typical range. No association was found between common variants in *USH2A* and language outcomes (Table 4.2).

The association between *USH2A* variation and low-frequency hearing abilities together with the results from Perrino *et al.,* showing that heterozygous *Ush2a* disruptions in mice affect their expressive communication abilities, warranted a further investigation into the relationships between *USH2A* and language. In this model, common variants in *USH2A* were assessed for association to language outcomes in ALSPAC, but this time including low-frequency hearing thresholds as an interaction factor (GxE interactions in Section 2.4.2). Significant association was now observed with early vocabulary at 3 years (kg865), represented by the top SNP rs7532570 (Table 4.2), located towards the 5' end of *USH2A* (between intron 4 and 13 on NM_206933.2 transcript) (Figure 4.2b). Within this interactive model rs7532570 had a p-value of 1.16×10^{-4} compared to P= 0.197 in the additive model. Together these data suggest that low frequency hearing can act as a modifying factor and modulate the effects of *USH2A* variants upon language development.

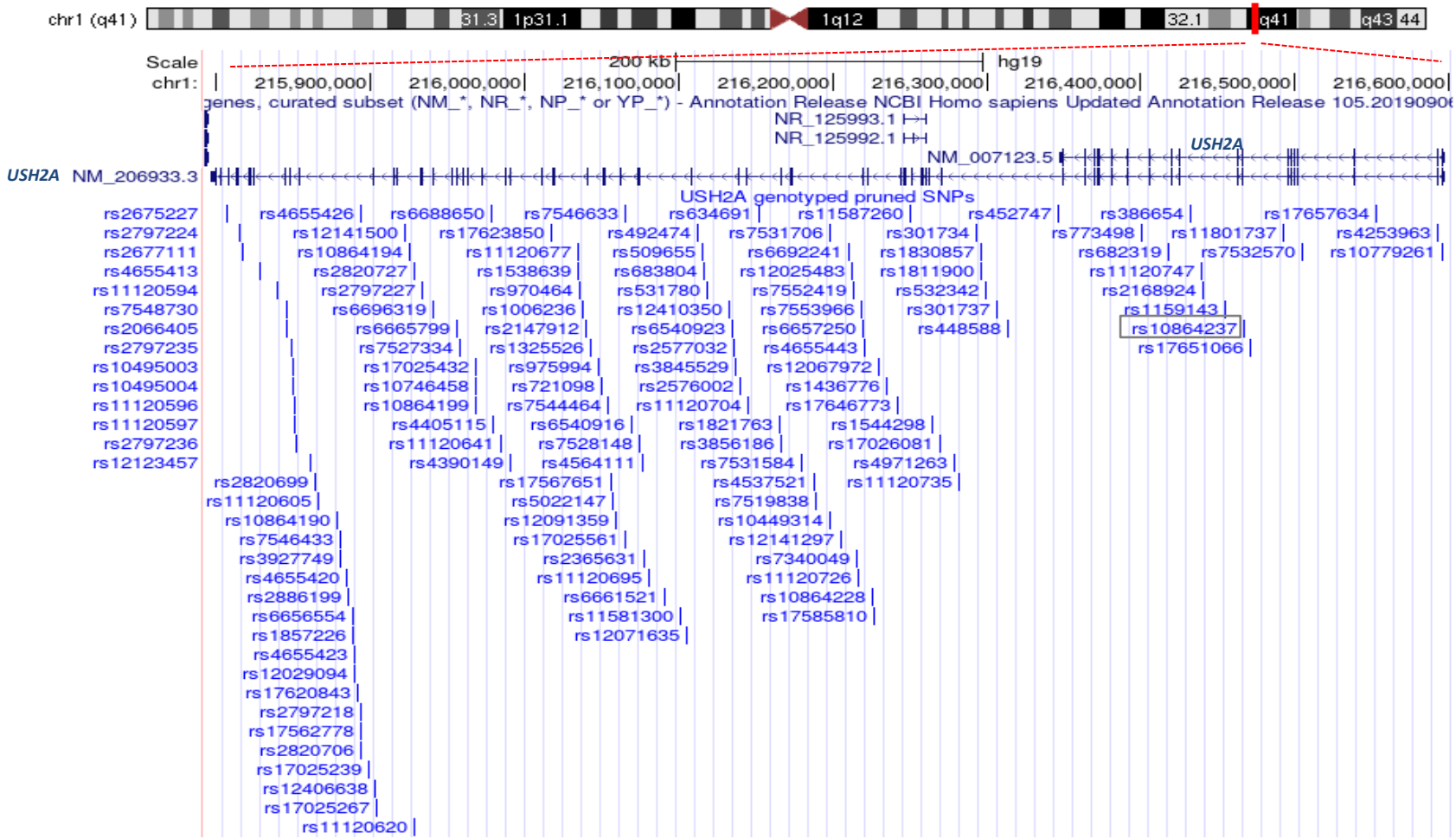

Figure 4.1. UCSC custom track of *USH2A* genotyped SNPs. Annotations show two protein coding *USH2A* RefSeq gene transcripts, located on the reverse DNA strand of chromosome 1. In blue (bottom of figure) are all genotyped SNPs that are of sufficient quality and have been pruned for high LD. Top associated SNP is marked in grey box.

Table 4.2. Associations of 5' USH2A common variants with hearing and language phenotypes under an additive and interactive model

				Additive Model								Interactive Model				
				Hearing outcomes			Language outcomes					Language outcomes				
Gene	**SNP**	**BP (hg19) chr 1:**	**A1**	**OME @7Y**	**Low Freq_ min**	**Mid Freq _ min**	**Comm @1.5 Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
3'end	rs682319	216417675	T	1.34E-01	6.88E-03	1.65E-02	3.17E-01	9.30E-02	3.62E-01	8.04E-01	7.39E-01	1.91E-02	**5.87E-04**	1.61E-01	5.03E-02	9.76E-01
	rs386654	216431962	A	9.25E-01	3.90E-01	4.30E-01	5.05E-01	9.38E-01	8.84E-01	8.74E-01	1.98E-01	3.88E-01	6.24E-02	8.12E-01	5.81E-01	8.44E-01
	rs11120747	216438500	G	5.36E-01	5.05E-03	5.81E-01	8.41E-01	2.38E-01	4.97E-01	9.61E-01	1.89E-01	4.79E-01	5.55E-01	2.78E-01	7.56E-02	8.66E-01
	rs2168924	216440105	A	9.98E-02	9.93E-01	4.41E-02	1.98E-01	3.49E-01	3.14E-01	5.37E-02	4.63E-01	1.60E-01	4.54E-01	7.43E-01	7.50E-02	3.75E-01
	rs1159143	216454483	T	2.50E-01	5.13E-03	6.41E-02	6.31E-01	5.38E-01	6.82E-01	7.20E-01	3.20E-01	5.94E-01	9.26E-02	3.12E-01	4.45E-01	6.79E-01
	rs10864237	216466861	C	8.03E-02	**2.92E-05**	1.14E-03	4.03E-01	2.44E-01	9.09E-01	9.88E-01	3.12E-01	3.47E-01	5.25E-01	1.08E-01	4.98E-02	8.79E-01
USH2A	rs17651066	216470121	C	1.29E-01	3.76E-02	2.34E-02	9.42E-01	1.22E-01	4.14E-01	3.15E-01	4.88E-01	5.03E-02	3.18E-03	3.44E-01	1.84E-01	3.28E-01
	rs11801737	216492391	G	7.48E-01	6.39E-04	2.59E-02	3.59E-01	5.39E-01	9.05E-01	7.29E-01	8.23E-01	9.20E-01	4.97E-01	3.68E-01	2.63E-01	6.30E-01
	rs7532570	216504269	G	8.09E-02	1.97E-01	1.16E-01	3.77E-01	1.50E-01	6.86E-01	8.35E-01	8.47E-01	7.64E-03	**1.16E-04**	2.16E-01	1.16E-01	8.15E-01
	rs17657634	216552571	G	2.14E-01	7.81E-01	7.61E-01	5.01E-01	7.28E-01	2.57E-01	3.17E-02	3.41E-01	6.56E-01	2.39E-01	2.96E-01	7.76E-01	6.69E-01
	rs4253963	216592003	T	5.94E-02	1.20E-02	4.30E-02	2.16E-01	1.60E-01	8.49E-01	6.17E-01	7.32E-01	2.47E-01	2.74E-01	1.54E-01	2.28E-01	5.46E-01
	rs10779261	216595306	C	5.60E-02	4.28E-02	1.41E-02	6.52E-01	6.76E-01	3.77E-01	8.25E-01	8.62E-01	4.44E-01	7.54E-01	4.45E-01	2.27E-01	5.30E-01
5' end	rs12723493	216605071	A	6.39E-01	1.53E-02	5.33E-01	7.98E-01	8.27E-01	2.83E-01	6.59E-01	3.95E-01	1.73E-01	3.37E-01	7.35E-01	1.94E-01	4.36E-01

SNPs are shown for the 5' region of association only (chr1:216417675-216605071, hg 19).

A1 is allele 1 (usually minor).

Grey cells indicate p-values <0.01. Bold values indicate that p-value was significant after a Bonferroni correction for multiple testing.

Bonferroni corrected p value for additive model: p= 0.05/129 SNPs/8 traits= 4.84 x 10^{-5}; Bonferroni corrected p value for interactive model: p= 0.05/13 SNPs/5 traits= 7.69 x 10^{-4}

a) Additive Model

b) Interactive Model

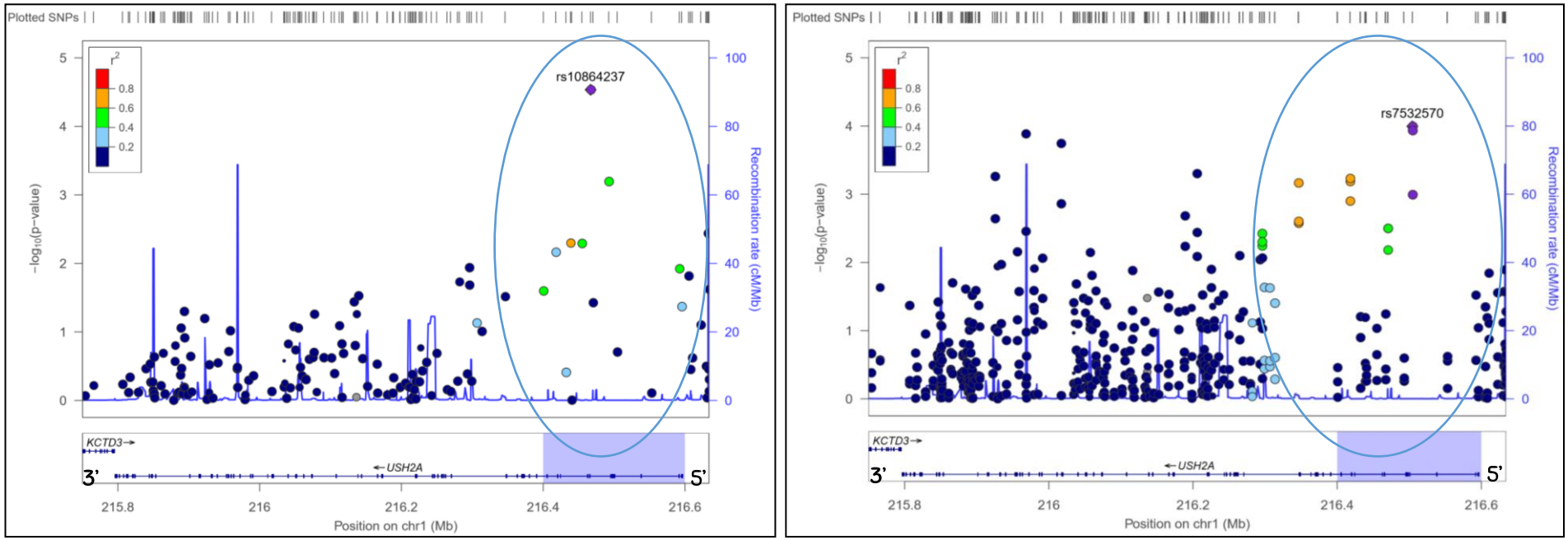

Figure 4.2. Regional *USH2A* plots of association.

a) with low frequency hearing thresholds under additive model and b) with vocabulary score at 38 months under interactive model. The -$\log_{10}$ of p values of the genotyped SNPs are plotted on the y-axis against genomic positions (hg19 assembly) on the x-axis. The top associated SNPs are represented by the purple diamond: a) rs10864237 and b) rs7532570. The colours of all other SNPs are representative of the pairwise r^2 value relative to the top SNP using patterns of Linkage Disequilibrium from the CEU HapMap populations. The top markers in (b) are shown to be in very high LD with each other and for both (a) and (b) they are in moderate LD with a cluster of markers in the 5' end of the gene, which have moderately elevated -log (p-values). Plots were generated using Locus Zoom (http://locuszoom.org/).

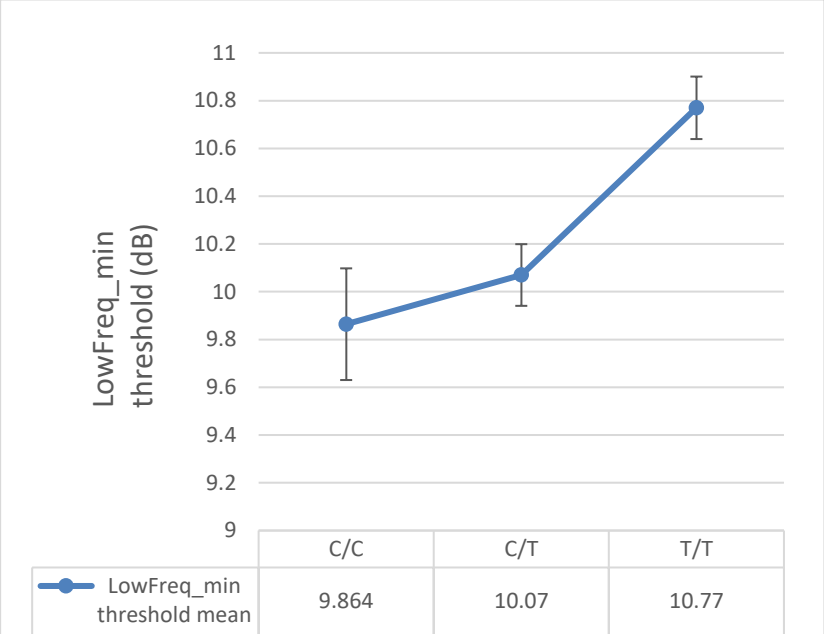

Figure 4.3. Low frequency hearing threshold per genotype. Risk genotype T/T increases the threshold for hearing noises at low-frequency, compared to C/C and C/T genotypes (difference between risk T/T and non-risk C/C genotype was significant (p= 0.0018). Error bars represent standard error of the mean. N individuals with C/C genotype= 479; N individuals with C/T genotype= 1,972; N individuals with T/T genotype= 1,989. Value of beta regression coefficient and standard error of beta is: ß= -0.5, ßSE= 0.1.

4.2.2. *CLRN1* common variants are directly associated with early communication skills

Direct associations under the additive model were observed for a cluster of *CLRN1* SNPs and early communication skills at 1.5 years (kd654) (Table 4.3 & Figure 4.4). The top-associated SNP (rs10935822, p= 2.94 x 10^{-4}) explained 0.21% of variance in early communication scores at 1.5 years (ßSE= 0.1), representing a 0.78 item decrease in the mean communication score for the risk group of individuals (C/C genotype) compared to the non-risk group (T/T genotype) (Table 4.3 & Figure 4.5a). This result suggests that *CLRN1* common variants exert only a small effect on early communication abilities at 1.5 years within the typical range. No associations were observed with the core language measures at later ages: vocabulary at 3 years, nonword repetition at 8 years and Verbal IQ at 8 years (Table 4.3). There were no associations to hearing outcomes tested (Table 4.3).

Based on the significant associations to communication skills at 1.5 years and the lack of association to later language measures at 8 years, associations to other language markers were further explored, just before and just after the 1.5 years milestone (between 6 months and 3 years of age). For this purpose, additional additive model associations were performed on 12 available language phenotypes taken from 6 months to 3 years of age (Table 4.4). These phenotypes captured scores of pre-linguistic communication skills at 6 months, emerging

receptive and expressive language skills between 15 months and 2 years and developed expressive language at 3 years, and were highly positively correlated (Table S5).

These further association analyses under an additive model showed a suggestive cluster of associations between *CLRN1* common SNPs and communication skills at 6 months (kb855a), and between a larger SNP cluster and vocabulary score at 15 months (kc954) (Table 4.4). The SNPs showing highest signals of association were: rs936188 ($p=6.01 \times 10^{-4}$) and rs4680058 ($p=3.66 \times 10^{-3}$). The risk genotypes G/G for communication skills at 6 months and G/G for vocabulary at 15 months showed a decrease in the mean scores of communication and vocabulary of 0.43 and 0.72 less items and words respectively, as compared to the non-risk genotypes of T/T and A/A (Figure 4.5 b-c). One single SNP (rs4680058) showed suggestive associations to vocabulary scores at 15 months and 2 years, communication skills at 1.5 years, use of plurals and tenses at 2 years, and overall grammar score at 2 years, which were all highly positively correlated measures (Table 4.4 & S5 for correlations). No associations were observed with vocabulary, language or grammar scores at 3 years.

Taken together, these results show that the association of *CLRN1* common variants with early language outcomes is only transient (observed up until 2 years) and is most strongly linked to expressive communication skills at 1.5 years.

It needs to be noted that *CLRN1* was not fully covered by the genotype assay: the 5' end was not spanned by common SNPs between chr3: 150,678,233- 150,690,786 (12Mb), including 5'UTR, exon 1 and the 5' end of intron 1 (26.7% of the gene length was not covered) (Figure. 4.4) and so any associations within those regions have been missed.

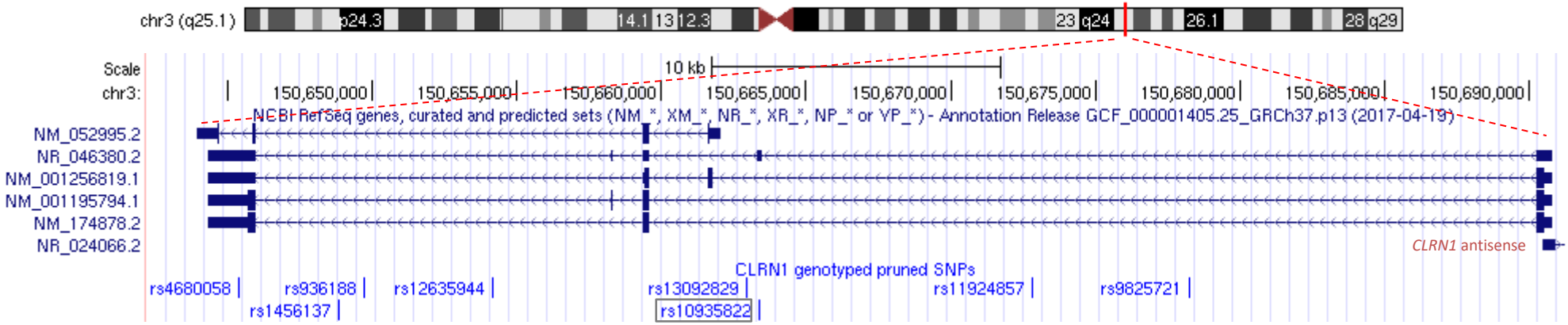

Figure 4.4. UCSC custom track of *CLRN1* genotyped SNPs. Annotations show four protein coding *CLRN1* RefSeq gene transcripts and one non-protein coding transcripts (NR_046380.2), located on the reverse DNA strand of chromosome 3. In blue (bottom of figure) are all genotyped SNPs that are of sufficient quality and have been pruned for high LD. Top associated SNP is marked in grey box.

Table 4.3. Associations of *CLRN1* common variants with hearing and language phenotypes under an additive model

				Additive Model							
				Hearing outcomes			**Language outcomes**				
Gene	**SNP**	**BP (hg19) chr 3:**	**A1**	**OME @7Y**	**Low Freq_ min**	**Mid Freq _min**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
3' end ↑ *CLRN1* ↑ *5'end*	rs4680058	150645351	G	8.21E-01	6.63E-02	1.50E-01	1.27E-03	7.18E-01	9.52E-01	9.96E-01	6.00E-01
	rs1456137	150648811	C	9.40E-01	2.02E-01	1.17E-01	2.12E-02	8.09E-02	5.65E-01	7.30E-01	2.19E-01
	rs936188	150649718	G	6.35E-01	2.35E-01	1.09E-01	9.56E-04	6.28E-01	8.32E-01	7.22E-01	4.15E-01
	rs12635944	150654154	A	9.90E-01	4.40E-01	2.72E-01	4.97E-03	1.02E-01	7.26E-01	5.98E-01	6.76E-01
	rs13092829	150662914	A	6.63E-01	4.57E-01	7.65E-01	5.03E-01	6.08E-01	2.69E-02	4.36E-01	2.81E-01
	rs10935822	150663334	C	8.72E-01	7.65E-01	1.88E-01	**2.94E-04**	8.44E-01	2.37E-01	2.22E-01	3.90E-01
	rs11924857	150672805	G	9.32E-01	4.60E-01	8.30E-02	1.44E-02	2.18E-01	9.07E-01	4.20E-01	6.50E-01
	rs9825721	150678233	G	1.28E-01	9.54E-01	1.78E-01	2.58E-01	3.52E-01	9.82E-01	5.59E-01	9.65E-01

A1 is allele 1 (usually minor).
Grey cells indicate p-values <0.01. Bold values indicate that p-value was significant after a Bonferroni correction for multiple testing.
Bonferroni corrected p value for additive model: p= 0.05/8 SNPs/8 traits= 7.81 x 10^{-4}

Table 4.4. Associations of *CLRN1* common variants with early language phenotypes under an additive model

				Additive Model Language outcomes											
Gene	**SNP**	**BP (hg19) chr 3:**	**A1**	**Comm @6 Mths**	**Understand @15 Mths**	**Vocab @15 Mths**	**Comm @1.5Y**	**Vocab @2Y**	**Plurals @2Y**	**Tenses @2Y**	**Grammar @2Y**	**Vocab @3Y**	**Plurals @3Y**	**Tenses @3Y**	**Word comb @3Y**
3' end	rs4680058	150645351	G	2.91E-02	3.98E-01	3.66E-03	1.27E-03	4.22E-03	3.44E-03	2.74E-03	8.24E-03	7.18E-01	8.25E-01	8.02E-01	3.83E-01
↑	rs1456137	150648811	C	7.19E-03	5.45E-01	1.19E-01	2.12E-02	1.67E-01	2.15E-01	4.44E-01	7.72E-01	8.09E-02	8.41E-01	6.54E-01	3.36E-01
CLRN1	rs936188	150649718	G	6.01E-04	9.93E-01	8.28E-03	9.56E-04	2.06E-01	4.11E-01	4.61E-01	8.21E-01	6.28E-01	7.39E-01	7.96E-01	5.85E-01
	rs12635944	150654154	A	8.62E-03	5.99E-01	5.33E-02	4.97E-03	9.70E-02	5.45E-02	3.05E-01	7.21E-01	1.02E-01	5.66E-01	8.72E-01	2.27E-01
↓	rs13092829	150662914	A	9.88E-01	8.92E-01	4.10E-01	5.03E-01	6.98E-01	3.14E-01	9.86E-01	6.71E-01	6.08E-01	2.16E-01	3.16E-01	2.66E-01
5'end	rs10935822	150663334	C	3.71E-02	9.38E-01	7.38E-03	**2.94E-04**	8.11E-02	1.46E-01	3.02E-01	6.77E-01	8.44E-01	4.44E-01	7.89E-01	4.74E-01
	rs11924857	150672805	G	1.44E-01	7.42E-01	7.77E-02	1.44E-02	3.04E-01	3.15E-01	2.78E-01	8.03E-01	2.18E-01	7.68E-01	1.67E-01	5.75E-02
	rs9825721	150678233	G	1.16E-01	6.21E-01	3.72E-01	2.58E-01	4.03E-01	6.06E-01	5.15E-01	8.19E-01	3.52E-01	1.70E-01	3.08E-01	7.62E-01

A1 is allele 1 (usually minor).

Grey cells indicate p-values <0.01. Bold values indicate that p-value was significant after a Bonferroni correction for multiple testing.

Bonferroni corrected p value for additive model: p= 0.05/8 SNPs/12 traits= 5.21 x 10^{-4}

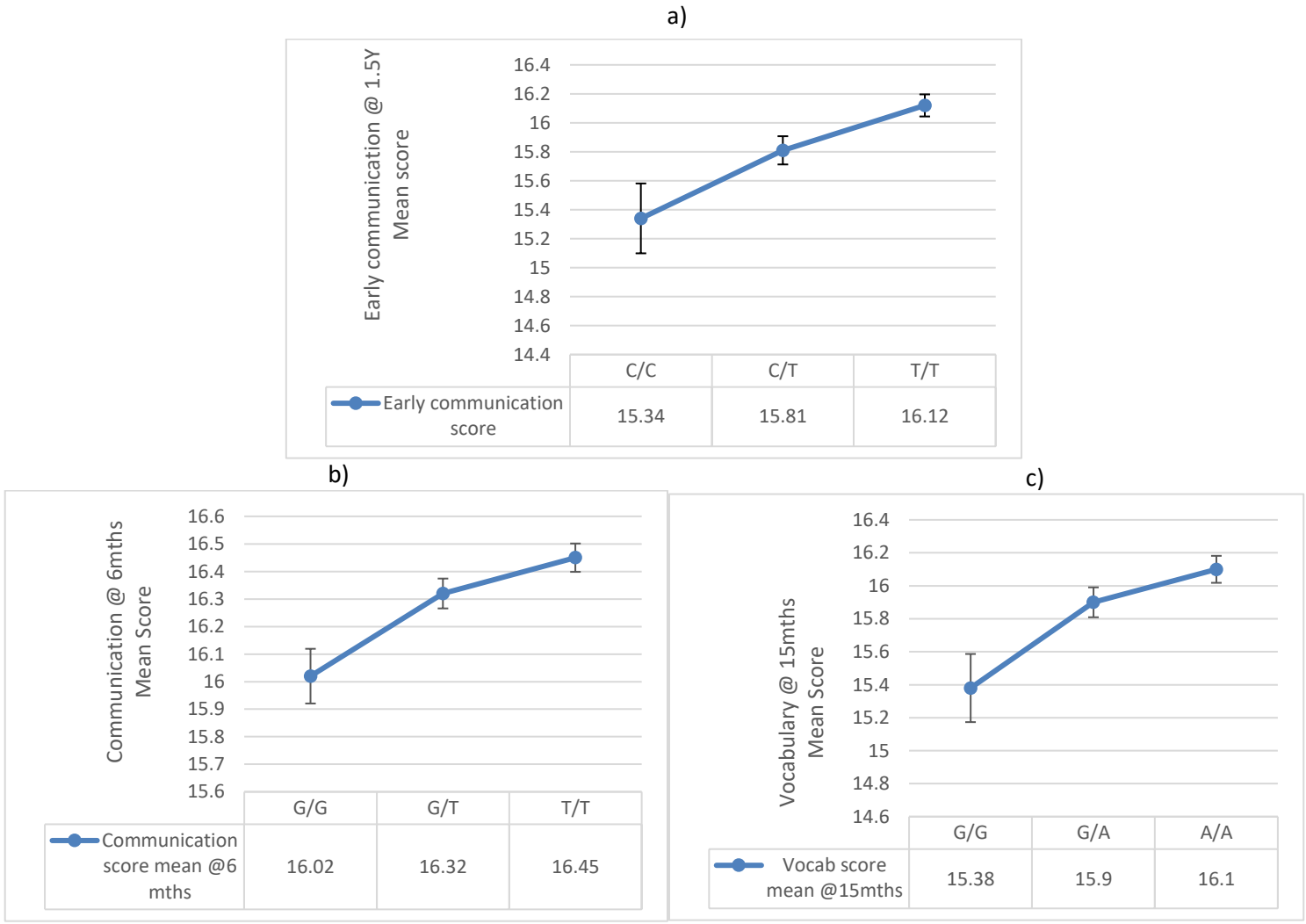

Figure 4.5. Early expressive language scores per genotype. Risk genotypes: C/C gives lower score on early communication test @1.5Y (a), G/G- to early communication test @ 6mths (b) and G/G- to vocabulary test @ 15mths (c). Error bars represent standard error of the mean. Number of available genotyped individuals: a) N individuals with C/C genotype=372; N individuals with C/T genotype= 2,280; N individuals with T/T genotype= 3,689; b) N individuals with G/G genotype=622; N individuals with G/T genotype= 2,689; N individuals with T/T genotype= 3,045; c) N individuals with G/G genotype=517; N individuals with G/A genotype= 2,521; N individuals with A/A genotype= 3,306. Value of beta regression coefficient and standard error of beta is: a) ß= -0.35, ßSE= 0.1; b) ß= -0.18, ßSE= 0.05; c) ß= -2.52, ßSE= 0.9.

4.2.3. A cluster of *PCDH15* common variants are marginally associated with risk of DLD

Under the additive model of direct effect, suggestive association was observed for a cluster of SNPs, located between intron 12 and 16 (NM_001142771.1 transcript) with DLD status (top SNP rs10763086, p= 1.94 x 10^{-3}) (Table 4.5 & Figure 4.6). The top SNP was found to be 1.2 times more common in cases (G/G genotype) compared to controls (A/A genotype) (OR= 1.2). Suggestive direct associations were also observed with low-frequency thresholds, represented by two SNPs and with mid-frequency thresholds, represented by three SNPs (Table 4.5), both located between intron 14 and intron 16 (NM_001142771.1 transcript). The top associated SNPs were rs7904409 (p= 5.24 x 10^{-3}) for low-frequency hearing thresholds and rs11004121 (p= 5.03 x 10^{-4}) for mid-hearing thresholds (Table 4.5). These analyses suggest that common *PCDH15* variants are marginally associated with altered hearing abilities (including low and mid-frequency hearing) and increased risk for DLD.

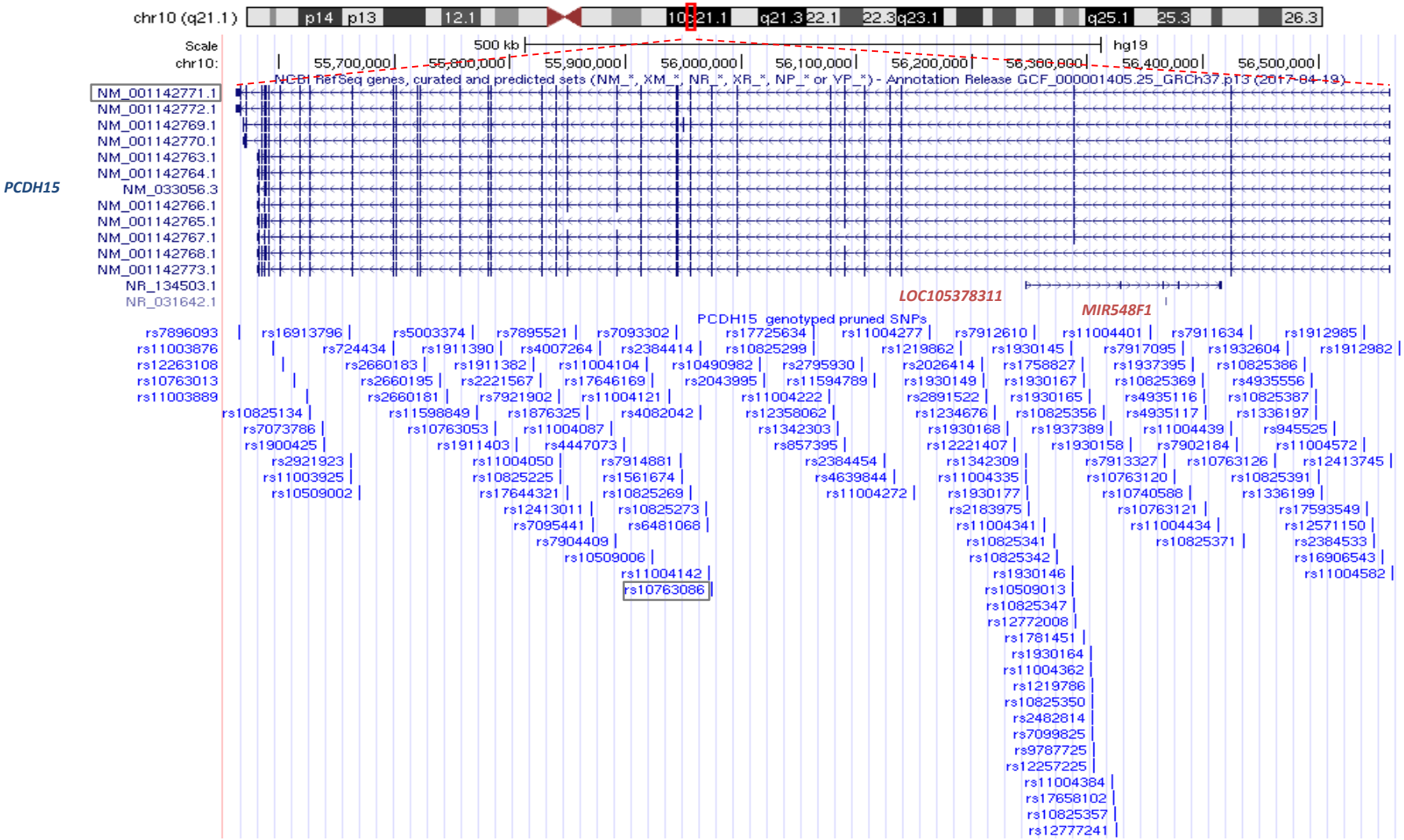

Figure 4.6. UCSC custom track of *PCDH15* genotyped SNPs. Annotations show twelve protein coding *PCDH15* RefSeq gene transcripts, located on the reverse DNA strand of chromosome 10. In blue (bottom of figure) are all genotyped SNPs that are of sufficient quality and have been pruned for high LD. Top associated SNP is marked in grey box.

Table 4.5. Associations of *PCDH15* common variants with hearing and language phenotypes under an additive model

				Additive Model							
				Hearing			Language				
Gene	**SNP**	**BP (hg19) chr 10:**	**A1**	**OME @7Y**	**LowFreq _min**	**MidFreq _min**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
3' end	rs7896093	55564355	T	8.88E-01	6.56E-01	9.01E-01	2.81E-01	5.44E-01	8.93E-01	1.04E-02	1.52E-01
	rs7904409	55890367	G	8.14E-01	5.24E-03	5.18E-04	9.91E-01	2.70E-01	8.69E-01	2.29E-02	2.89E-03
	rs4447073	55898144	T	8.76E-01	3.73E-02	7.19E-02	7.30E-01	7.56E-01	1.39E-01	3.74E-01	8.00E-03
	rs11004104	55918359	T	8.19E-01	1.45E-01	1.96E-01	7.58E-01	7.67E-01	2.32E-01	9.85E-01	6.67E-03
	rs17646169	55922376	T	8.57E-01	7.39E-01	8.97E-01	1.75E-01	9.62E-01	9.25E-01	6.86E-01	3.11E-01
	rs10509006	55922706	G	3.37E-01	4.86E-01	1.93E-02	4.26E-01	8.28E-01	2.49E-01	7.88E-02	2.26E-01
	rs11004121	55937483	C	3.89E-01	5.62E-03	5.03E-04	6.91E-01	2.54E-01	9.55E-01	2.18E-02	5.88E-02
	rs7093302	55943184	T	8.99E-01	1.87E-01	8.14E-02	8.58E-01	2.78E-01	5.15E-01	1.57E-01	9.47E-03
	rs7914881	55947112	C	4.36E-01	2.41E-02	2.36E-03	3.04E-01	6.42E-01	3.10E-01	2.12E-03	3.05E-03
PCDH15	rs1561674	55948714	C	4.20E-01	7.04E-01	5.63E-01	4.70E-01	5.86E-01	1.68E-01	5.73E-01	5.63E-01
	rs10825269	55955610	T	5.49E-01	6.13E-02	5.88E-02	7.23E-01	4.93E-01	1.31E-01	6.54E-02	8.62E-02
	rs2384414	55964008	A	9.13E-01	9.00E-01	5.49E-01	5.63E-01	9.11E-01	2.52E-01	5.46E-01	5.20E-01
	rs4082042	55964451	G	7.71E-01	1.45E-01	9.46E-02	8.76E-01	8.41E-01	5.45E-01	6.88E-01	2.49E-02
	rs10825273	55968685	T	7.19E-01	9.03E-02	3.69E-01	8.32E-01	1.79E-01	9.52E-01	3.62E-01	1.39E-01
	rs6481068	55970721	T	8.54E-01	3.88E-01	7.40E-01	4.18E-01	1.08E-01	7.21E-01	5.07E-01	8.66E-02
	rs11004142	55972031	A	9.09E-01	7.89E-02	1.10E-01	5.63E-01	2.50E-01	3.86E-01	5.88E-01	2.50E-03
	rs10763086	55973889	G	7.33E-01	1.35E-02	3.96E-02	7.74E-01	3.26E-01	6.69E-01	3.18E-01	1.94E-03
5'end	rs1912982	56569551	G	3.94E-02	7.45E-01	6.59E-01	8.12E-02	6.62E-01	9.98E-01	5.76E-02	8.40E-01

A1 is allele 1 (usually minor).

Grey cells indicate p-values <0.01.

Bonferroni corrected p value for additive model: p= 0.05/135 SNPs/8 traits= 4.63 x 10^{-5}

As the DLD status score is a composite measure, based on several aspects of language performance (WOLD Comprehension: f8sl040; CCC Intelligibility and Fluency: ku503b and CCC Syntax: ku504b scores) (Section 2.2.1.2), additional analyses were performed on each individual contributory measure in order to find if one aspect of language was particularly associated with *PCDH15*. Suggestive associations between single variants and all three outcomes were observed (Table S6). The strongest signal of suggestive association was between *PCDH15* variant rs12772008 and CCC Intelligibility and Fluency score represented by $p= 4.17 \times 10^{-3}$ (corrected threshold of significance for multiple testing $P= 1.23 \times 10^{-4}$). These results suggest that the three language measures (WOLD comprehension, CCC Intelligibility and Fluency, and CCC Syntax) within the DLD score all contribute to the suggestive association with *PCDH15* common variants and that there is no individual measure that is driving the association. This further demonstrates that the DLD score as a collective is the most appropriate measure to use in order to capture association with *PCDH15* variants.

The suggestive association between common *PCDH15* variants and DLD status can also be further inspected by testing whether variants within *PCDH15* play a role in wider neurodevelopmental disorders which require special arrangements at school. To do this, common variants in *PCDH15* were assessed for association with ALSPAC measures of problems requiring special school arrangement, dyslexia status and special education needs status. A small cluster of suggestive associations were observed with communication and interaction needs at school (SEN), located within the same cluster of associations with DLD (Table 4.6).The top associated SNP, rs11004121, showed highest signal of association at $p=1.17 \times 10^{-3}$ (corrected threshold of significance for multiple testing $P= 3.47 \times 10^{-4}$).

Taken together, the association analyses suggest that common *PCDH15* variants are directly, but only marginally associated with DLD status, where WOLD Comprehension, CCC Intelligibility and Fluency, and CCC Syntax all contribute to the risk for DLD. Furthermore, common *PCDH15* variants are likely to also increase the risk for requiring special education provision during Key Stage 3 (11-13 years) in the area of communication and interaction.

Table 4.6. Follow-up associations of *PCDH15* common variants (intron 12 to 16) with measures of special school arrangements, dyslexia and SEN statement under additive model.

					Additive Model								
					Problems that require special school arrangement @7.5Y					**Dyslexia @9.5Y**	**Special Education Needs (SEN) statement @11-13Y**		
Gene	**SNP**	**BP (hg19) chr 10:**	**A1**	DLD	Any problems	Learning	Speech	Reading	Emotional/ behavioural	Child had dyslexia	Cognition & learning	Behaviour, emotional & social	Communication & interaction
towards 3' end	rs7904409	55890367	G	2.89E-03	0.8232	0.5495	0.1586	0.5803	0.08859	0.3255	0.3375	0.3869	3.16E-03
	rs4447073	55898144	T	8.00E-03	0.9547	0.4851	0.5225	0.2465	0.4176	0.4086	0.07142	0.2366	0.0936
	rs11004104	55918359	T	6.67E-03	0.913	0.6432	0.2867	0.222	0.1366	0.863	0.16	0.3018	0.2318
	rs17646169	55922376	T	0.311	0.681	0.9652	0.3596	0.6171	0.3221	0.6737	0.5539	2.41E-03	6.57E-03
	rs10509006	55922706	G	0.2256	0.9769	0.4882	0.9385	0.9417	0.9098	0.1893	0.6215	0.7706	0.0876
	rs11004121	55937483	C	0.05884	0.9312	0.8397	0.4774	0.694	0.2624	0.5426	0.9044	0.2973	1.17E-03
	rs7093302	55943184	T	9.47E-03	0.6926	0.5785	0.9358	0.5288	0.06188	0.9063	0.6476	0.1711	0.3591
	rs7914881	55947112	C	3.05E-03	0.6922	0.7804	0.4873	0.775	0.4709	0.8595	0.711	0.3247	0.0255
PCDH15	rs1561674	55948714	C	0.5634	0.7825	0.722	0.04995	0.4446	0.7301	0.358	0.586	0.6514	0.2962
	rs10825269	55955610	T	0.08615	0.3407	0.3761	0.5869	0.9307	0.09019	0.5682	0.09107	0.6501	0.3905
	rs2384414	55964008	A	0.5199	0.1015	0.2312	0.4364	0.7795	0.2573	0.7277	0.1162	0.1782	0.1005
	rs4082042	55964451	G	0.02493	0.4433	0.5369	0.4775	0.9146	0.7541	0.3023	0.3475	0.2242	0.4401
	rs10825273	55968685	T	0.1386	0.1985	0.3469	0.03487	0.265	0.08322	0.3953	0.357	0.8378	0.5767
	rs6481068	55970721	T	0.08662	0.6469	0.7541	0.02744	0.362	0.3908	0.3029	0.4266	0.5861	0.4357
towards 5'end	rs11004142	55972031	A	2.50E-03	0.9417	0.8267	0.1293	0.472	0.3484	0.7963	0.9848	0.1748	0.216
	rs10763086	55973889	G	1.94E-03	0.5454	0.7352	0.2376	0.3365	0.09833	0.9311	0.9704	0.2946	0.316

A1 is allele 1 (usually minor).
Grey cells indicate p-values <0.01.
Bonferroni corrected p value for additive model: p= 0.05/16 SNPs/9 traits= 3.47 x 10^{-4}

4.2.4. *CDH23* common variants show sporadic association across hearing or language measures

Under the additive model of direct effect, suggestive associations were observed for single SNPs and low-frequency hearing thresholds, mid-frequency hearing threshold, communication skills at 1.5 years and Verbal IQ at 8 years (represented by 2 SNPs) (Table 4.7). No clusters of associated SNPs were observed for any one measure, however, the individual SNPs formed a cluster of suggestive associations located between intron 11 and 13 on NM_022124.5 transcript (Figure 4.7) (Table 4.7). The top SNP (rs11819553 with p=3.24 x 10^{-3}) was marginally associated with Verbal IQ at 8 years (Table 4.7 & Figure 4.7). These marginal single variant associations to low- and mid-hearing thresholds, early communication and Verbal IQ at 8 years are not a strong enough evidence of a robust direct link between *CHD23* common variants and hearing/language outcomes in this cohort sample, and therefore do not warrant any further investigations. Larger samples sizes would be needed to replicate these results at a significant level.

Table 4.7. Associations of *CDH23* common variants with hearing and language phenotypes under an additive model (only regions of suggestive associations shown).

				Additive Model							
				Hearing			Language				
Gene	**SNP**	**BP (hg19) chr 10:**	**A1**	**OME @7Y**	**Low Freq _min**	**Mid Freq _min**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
5' end	rs10999801	73150325	A	0.70	0.45	0.77	0.58	0.27	0.42	0.48	0.25
	rs12360332	73152849	T	0.01	0.30	0.0054	0.49	0.75	0.99	0.86	0.65
	rs1868003	73386066	A	0.21	0.60	0.49	0.0084	0.28	0.02	0.17	0.06
	rs4746089	73387685	A	0.60	0.82	0.60	0.19	0.78	0.12	0.46	0.10
	rs2394834	73392998	G	0.84	0.08	0.60	0.11	0.68	0.02	0.46	0.02
	rs1900514	73408326	A	0.58	0.74	0.58	0.04	0.01	0.05	0.60	0.06
	rs956734	73408544	A	0.30	0.62	0.52	0.93	0.18	0.75	0.84	0.54
	rs4746093	73409978	A	0.77	0.0096	0.02	0.81	0.78	0.17	0.20	0.25
	rs10999933	73411591	T	0.87	0.82	0.54	0.27	0.82	0.62	0.17	0.76
CDH23	rs1900515	73413555	A	0.61	0.87	0.48	0.86	0.18	0.19	0.11	0.54
	rs12573587	73413824	T	1.00	0.05	0.50	0.94	0.79	0.0051	0.36	0.04
	rs7896061	73414010	C	0.95	0.18	0.47	0.67	0.89	0.54	0.87	0.13
	rs10762465	73414206	G	0.49	0.22	0.31	0.90	0.12	0.43	0.32	0.12
	rs7087554	73414577	A	0.56	0.87	0.35	0.72	0.72	0.20	0.17	0.58
	rs1665688	73417236	T	0.92	0.05	0.15	0.91	0.33	0.07	0.57	0.07
	rs1665624	73421248	C	0.69	0.32	0.58	0.66	0.45	0.13	0.21	0.15
	rs11819553	73426465	G	0.59	0.24	0.96	0.79	0.81	0.0032	0.58	0.02
3'end	rs1015193	73584426	C	0.37	0.83	0.13	0.40	0.21	0.64	0.49	0.62

A1 is allele 1 (usually minor).
Grey cells indicate p-values <0.01.
Bonferroni corrected p value for additive model: p= 0.05/127 SNPs/8 traits= 4.92 x 10^{-5}

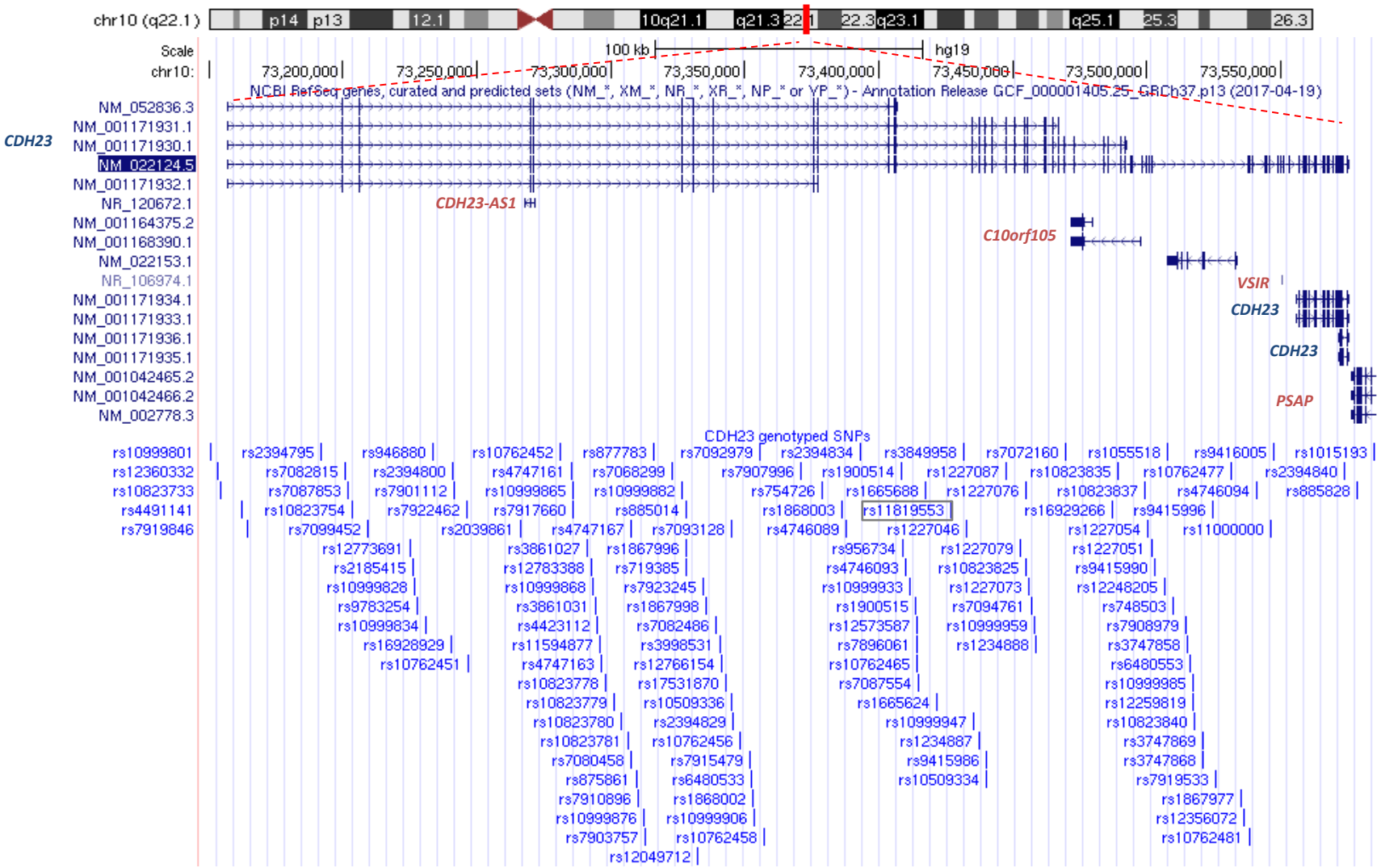

Figure 4.7. UCSC custom track of *CDH23* genotyped SNPs. Annotations show nine protein coding *CDH23* RefSeq gene transcripts, located on the forward DNA strand of chromosome 10. In blue (bottom of figure) are all genotyped SNPs that are of sufficient quality and have been pruned for high LD. Top associated SNP is marked in grey box.

4.2.5. *ADGRV1* and *USH1C* common variants show sporadic association across hearing or language measures

Under the additive model, only suggestive sporadic SNP associations were observed between *ADGRV1* and low-and-mid-frequency hearing thresholds and DLD (Table 4.8), and between *USH1C* and low-and-mid-frequency, vocabulary at 3 years and VIQ at 8 years (Table 4.9). These suggestive associations were spread across both genes and therefore did not form a particular cluster of association (Table 4.8 and Table 4.9). The top SNP for *ADGRV1* (rs2007538, p= 0.0018) was marginally associated with mid-frequency hearing thresholds (Table 4.8), while the top SNP for *USH1C* (rs2237960, p= 0.0006) was marginally associated with VIQ at 8 years. Similar to the *CDH23* results above, these marginal sporadic *ADGRV1* and *USH1C* variant associations within the investigated cohort are not robust and do not warrant any further analysis. Larger samples sizes would be needed to replicate these results at a significant level.

Table 4.8. Associations of 5' *ADGRV1* common variants with hearing and language phenotypes under an additive model (only regions of suggestive associations shown).

				Additive Model							
				Hearing			**Language**				
Gene	**SNP**	**BP (hg19) chr 5:**	**A1**	**OME @7Y**	**Low Freq_ min**	**Mid Freq _ min**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
5' end	rs10514328	89847694	T	0.25	0.46	0.76	0.91	0.18	0.05	0.72	0.77
	rs16869016	90000210	T	0.49	0.0039	0.0064	0.34	0.96	0.29	0.19	0.54
ADGRV1	rs949787	90251205	T	0.04	0.96	0.96	0.79	0.06	0.27	0.17	0.0042
	rs2007538	90374713	C	0.03	0.04	0.0018	0.20	0.43	0.55	0.68	0.69
3'end	rs4537089	90466296	C	1.00	0.97	0.66	0.55	0.27	0.08	0.16	0.82

A1 is allele 1 (usually minor).
Grey cells indicate p-values <0.01.
Bonferroni corrected p value for additive model: p= 0.05/66 SNPs/8 traits= 9.47×10^{-5}

Table 4.9. Associations of *USH1C* common variants with hearing and language phenotypes under an additive model

				Additive Model							
				Hearing			Language				
Gene	**SNP**	**BP (hg19) chr 11:**	**A1**	**OME @7Y**	**Low Freq_ min**	**Mid Freq _ min**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
	rs1076204	17506874	C	0.40	0.58	0.79	0.29	0.02	0.16	0.53	0.64
3' end	rs12278908	17508140	A	0.17	0.08	0.08	0.65	0.55	0.53	0.95	0.14
	rs4757527	17510565	G	0.74	0.81	0.78	0.09	0.0089	0.56	0.73	0.42
	rs17703233	17513147	C	0.62	0.84	0.78	0.02	0.04	0.56	0.70	0.27
	rs2190454	17533635	G	0.39	0.05	0.85	0.28	0.72	0.66	0.20	0.12
	rs2237960	17538048	A	0.80	0.26	0.67	0.43	0.30	0.0006	0.10	0.66
	rs10766408	17539212	T	0.56	0.11	0.37	0.07	0.60	0.17	0.07	0.06
	rs11603262	17540079	T	0.20	0.11	0.28	0.32	0.05	0.14	0.73	0.43
USH1C	rs7129173	17541798	A	0.96	0.54	0.36	0.02	0.19	0.65	0.58	0.32
	rs2041032	17542649	T	0.83	0.28	0.26	0.02	0.23	0.46	0.59	0.03
	rs16934382	17551893	A	0.10	0.29	0.0020	0.78	0.98	0.95	0.93	0.11
	rs2041027	17553105	A	0.90	0.07	0.07	0.02	0.81	0.82	0.92	0.01
	rs7119071	17572402	A	0.75	0.01	0.38	0.39	1.00	0.97	0.31	0.05
	rs10766410	17573329	A	0.76	0.0058	0.21	0.25	0.70	0.91	0.81	0.02
5'end	rs7951242	17573927	T	0.60	0.14	0.76	0.90	0.29	0.35	0.69	0.59
	rs4757543	17574588	T	0.26	0.52	0.93	0.78	0.35	0.18	0.66	0.31

A1 is allele 1 (usually minor).
Grey cells indicate p-values <0.01.
Bonferroni corrected p value for additive model: p= 0.05/34 SNPs/8 traits= 1.84 x 10^{-4}

4.2.6. *CIB2*, *WHRN* and *MYO7A* show no association of their common variants with hearing or language milestones

Association analyses under the additive model showed no significant or suggestive associations between common variants in *CIB2*, *WHRN* and *MYO7A* and any of the hearing or language phenotypes tested (Table 4.10-4.12).

Table 4.10. Associations of *CIB2* common variants with hearing and language phenotypes under an additive model

				Additive Model							
				Hearing			Language				
Gene	**SNP**	**BP (hg19) chr 15:**	**A1**	**OME @7Y**	**Low Freq_ min**	**Mid Freq _min**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
3' end	rs2289524	78390414	C	0.40	0.23	0.49	0.30	0.88	0.97	0.88	0.24
	rs12593575	78390909	T	0.06	0.15	0.50	0.82	0.40	0.93	0.57	0.19
	rs2867922	78391969	A	0.36	0.94	0.90	0.36	0.53	0.91	0.39	0.96
	rs7182113	78392357	A	0.22	0.05	0.19	0.99	0.77	0.83	0.89	0.16
	rs9806257	78395362	C	0.68	0.24	0.48	1.00	0.65	0.62	0.55	0.68
CIB2	rs16953973	78400640	T	0.11	0.64	0.06	0.92	0.40	0.96	0.32	0.74
	rs3784327	78402258	A	0.74	0.65	0.14	0.46	0.33	0.91	0.95	0.38
	rs11856417	78404315	A	0.14	0.97	0.03	0.45	0.54	0.43	0.62	0.32
	rs17478430	78406959	C	0.79	0.48	0.74	0.51	0.73	0.84	0.28	0.84
	rs16969514	78410289	C	0.75	0.76	0.82	0.42	0.89	0.21	0.65	0.92
5'end	rs1542101	78413016	G	0.15	0.79	0.23	0.49	0.35	0.37	0.34	0.88

A1 is allele 1 (usually minor).

Table 4.11. Associations of *WHRN* common variants with hearing and language phenotypes under an additive model

				Additive Model							
				Hearing			Language				
Gene	**SNP**	**BP (hg19) chr 9:**	**A1**	**OME @7Y**	**Low Freq_ min**	**Mid Freq _ min**	**Comm @1.5Y**	**Vocab @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
3' end	rs2274159	117166246	A	0.02	0.67	0.95	0.91	0.55	0.11	0.77	0.35
	rs942519	117169033	A	0.01	0.87	0.80	0.99	0.47	0.08	0.41	0.20
	rs2236388	117169300	A	0.49	0.12	0.38	0.87	0.22	0.66	0.48	0.96
	rs1075559	117174124	G	0.24	0.33	0.91	0.23	0.98	0.17	0.88	0.55
	rs10081699	117182478	G	0.10	0.15	0.51	0.36	0.62	0.06	0.59	0.03
	rs10739411	117183149	G	0.36	0.57	0.84	0.96	0.75	0.68	1.00	0.20
	rs2181928	117185395	A	0.03	0.21	0.48	0.88	0.85	0.18	0.63	0.20
	rs2274162	117187569	T	0.08	0.57	0.94	0.25	0.66	0.74	0.89	0.55
	rs7046973	117193206	G	0.19	0.26	0.45	0.93	0.87	0.31	0.41	0.16
	rs10982218	117200140	C	0.99	0.63	0.35	0.76	0.81	0.51	0.07	0.32
	rs942520	117214909	C	0.96	0.13	0.35	0.96	0.77	0.95	0.66	0.93
WHRN	rs10759707	117222287	A	0.51	0.24	0.27	0.62	0.72	0.77	0.77	0.15
	rs10217748	117225658	T	0.80	0.35	0.72	0.34	0.26	0.47	0.69	0.62
	rs4979407	117228115	C	0.78	0.24	0.05	0.17	0.85	0.84	0.73	0.21
	rs1535971	117229400	T	0.49	0.49	0.98	0.46	0.75	0.76	0.60	0.30
	rs10982234	117230017	C	0.57	0.64	0.45	0.32	0.48	0.53	0.31	0.92
	rs1123056	117236557	T	0.77	0.62	0.67	0.66	0.39	0.50	0.29	0.97
	rs17807115	117256467	G	0.25	0.55	0.15	0.78	0.17	0.83	0.65	0.75
	rs4979415	117261599	A	0.33	0.49	0.76	0.84	0.99	0.30	0.14	0.70
	rs4979418	117264083	G	0.41	0.23	0.20	0.47	0.40	0.69	0.11	0.97
5'end	rs2296262	117265406	T	0.61	0.50	0.23	0.95	0.62	0.88	0.66	0.27

A1 is allele 1 (usually minor).

Table 4.12. Associations of *MYO7A* common variants with hearing and language phenotypes under an additive model

				Additive Model							
				Hearing			Language				
Gene	SNP	BP (hg19) chr 11:	A1	OME @7Y	Low Freq_ min	Mid Freq _min	Comm @1.5Y	Vocab @3Y	VIQ @8Y	NWR @8Y	DLD
5' end	rs2276031	76832011	T	0.82	0.61	0.35	0.96	0.78	0.17	0.28	0.88
	rs1043418	76836350	C	0.58	0.66	0.17	0.57	0.65	0.03	0.03	0.58
	rs7943716	76837499	A	0.54	0.72	0.69	0.15	0.88	0.69	0.94	0.62
	rs10899353	76837552	T	0.44	0.60	0.93	0.86	0.34	0.99	0.04	0.10
	rs7121485	76840709	G	0.60	0.62	0.30	0.89	0.70	0.01	0.19	0.63
	rs7121629	76840798	G	0.06	0.32	0.06	0.67	0.09	0.56	0.95	0.99
	rs948969	76848035	G	0.78	0.86	0.48	0.94	0.98	0.17	0.10	0.80
	rs1052030	76853783	C	0.34	0.85	0.60	0.10	0.49	0.85	0.96	0.10
	rs12279716	76858756	T	0.14	0.90	0.32	0.97	0.90	0.70	0.46	0.81
	rs3737454	76868278	A	0.87	0.14	0.84	0.97	0.49	0.75	0.08	0.83
	rs762667	76868372	C	0.76	0.45	0.34	0.84	0.10	0.43	0.54	0.32
	rs3740763	76873620	T	0.56	0.44	0.99	0.37	0.31	0.86	0.92	0.72
MYO7A	rs4944147	76880518	A	0.72	0.49	0.73	0.83	0.29	0.22	0.71	0.21
	rs4945156	76892179	A	0.30	0.85	0.51	0.66	0.94	0.30	0.19	0.15
	rs3758708	76894463	T	0.20	0.51	0.12	0.23	0.26	0.89	0.36	0.80
	rs12805353	76898797	T	0.21	0.09	0.73	0.73	0.41	0.63	0.31	0.25
	rs3781694	76899265	A	0.07	0.04	0.42	0.41	0.50	0.59	0.41	0.31
	rs948962	76919478	A	0.17	0.02	0.40	0.51	0.66	0.53	0.08	1.00
	rs885442	76920038	T	0.05	0.01	0.54	0.24	0.50	1.00	0.07	0.73
	rs12793189	76921090	A	0.19	0.07	0.73	0.46	0.38	0.39	0.01	0.78
	rs12793619	76921358	A	0.88	0.31	0.32	0.71	0.32	0.19	0.05	0.26
3'end	rs11237123	76922946	A	0.90	0.34	0.49	0.84	0.53	0.24	0.04	0.46

A1 is allele 1 (usually minor).

4.2.7. *USH1G* and *HARS* common variants are not covered by the genotype assay

Due to the small size of *USH1G* (7Kb) and *HARS* (17Kb), plus filtering of low-quality variants and pruning for variants in high LD, no common variants within the genotype assay were left covering these genes.

4.3. Discussion

Across the eleven Usher genes analysed, three genes (*USH2A*, *CLRN1*, *PCDH15*) showed evidence of association to hearing and/or language measures across multiple time points (Table 4.13). Three further genes (*CDH23*, *ADGRV1* and *USH1C*) showed sporadic associations to hearing and language outcomes (Table 4.13), which were represented by single SNPs and therefore no trend was observed. The Usher genes *WHRN*, *MYO7A* and *CIB2* did not show association of their common variants to hearing or language outcomes and *USH1G* and *HARS* common variants were not analysed as the genes were not covered by the genotype assay (Table 4.13).

Table 4.13. Summary results for the effect of common variants in 11 Usher genes on hearing and language phenotypes plus wider effects and protein function

Gene	Association with hearing	Association with early language marker	Association with language outcomes (DLD)	Effect of biallelic mutations on hearing in USH syndrome	Protein Function	Protein Localisation	Cochlea expression	Brain expression (Human Brain Transcriptome Project)
USH2A	low-frequency	indirect	none	congenital & moderate hearing loss	cell adhesion & signalling	stereocilia ankle links	transient (embryonic development in ankle links)	negligible
CLRN1	none	direct	none	post-lingual hearing loss	cell adhesion	hair bundle	embryonic & postnatal hair cells	negligible
PCDH15	suggestive low/ mid-frequency	none	direct	congenital & profound hearing loss	cell adhesion, upper tip link	stereocilia lateral, kinociliary and tip links	transient (embryonic in lateral links) and postnatal in mature hair cells	prenatal (CBC); postnatal (HIP, AMY, MD, NCX, STR)
CDH23	sporadic low/mid-frequency	direct sporadic	none	congenital & profound hearing loss	cell adhesion, lower tip link	stereocilia lateral, kinociliary and tip links	transient (embryonic in lateral links) and postnatal in mature hair cells	postnatal (CBC, MD)
ADGRV1	sporadic low/mid-frequency	none	direct sporadic	congenital & moderate hearing loss	cell adhesion	stereocilia ankle links	transient (embryonic development in ankle links)	early prenatal (STR, AMY, HIP); postnatal (HIP, AMY, NCX, STR)
USH1C	sporadic low/mid-frequency	direct sporadic	none	congenital & profound hearing loss	scaffold protein	stereocilia upper tip link and synapse	embryonic & postnatal hair cells	prenatal (small peak in HIP); postnatal (HIP, STR AND MD)
WHRN	none	none	none	congenital & moderate hearing loss	scaffold protein	stereocilia ankle link, stereocilia tip & synapse	transient (embryonic development in ankle links)	N/A
MYO7A	none	none	none	congenital & profound hearing loss	molecular motor protein	hair cell cytoplasm, hair bundle, upper tip link density	embryonic & postnatal hair cells	N/A
CIB2	none	none	none	congenital & profound hearing loss	intracellular calcium signalling	stereocilia near tip	no data on embryonic development/ in mature hair cells	N/A
USH1G HARS	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Brain expression data is from the Human Brain Transcriptome Project (https://hbatlas.org/pages/hbtd).
CBC= cerebellar cortex; HIP= hippocampus; AMY= amygdala; MD= mediodorsal nucleus of the thalamus; NCX= neocortex; STR= striatum.

4.3.1. *USH2A* common variants and hearing/language abilities

The associations of common *USH2A* variants with hearing abilities suggests that individuals with common risk variants located at the start of the gene (5' end) are susceptible to having difficulties detecting low-frequency sounds. Moreover, individuals carrying *USH2A* risk variants in the presence of altered low-frequency hearing thresholds were more likely to have a limited vocabulary than those who carried only one of these risk variants in isolation. This indicates a directionality of effects which stems from low-frequency hearing thresholds and further shows that hearing can modulate the effects of *USH2A* upon language development (Figure 4.8a). These population-based findings are in line with the results from the discovery family, implicating *USH2A* in impaired auditory processing and language, suggesting that the pathology of APD may stem from subtleties in early hearing abilities (Perrino *et al.*, 2020). As peripheral auditory mechanisms are not typically included in the diagnosis of APD, rather, individuals are only screened for overt hearing loss, therefore subtle differences at low-frequency hearing abilities would be missed. As auditory perception involves the integration of bottom-up, auditory 'sensory' information (from the inner ear up) with top-down, multimodal 'cognitive' information (from the auditory cortex down) (Moore *et al.*, 2010), a subtle change in low-frequency hearing can exert a small, but potentially deleterious effect that disrupts higher order processes and therefore, language. Further studies would need to be performed to investigate the underlying mechanisms.

4.3.2. *CLRN1* common variants and language abilities

The direct association of common *CLRN1* variants with very early expressive communication skills was only transient and not observed with later language markers. This suggests that children who carried *CLRN1* risk variants were more likely to show a slight delay in their communication abilities during pre-lingual and early lingual period compared to children who did not carry the risk variants (Figure 4.8c). Clarin-1 localises to the mouse inner ear hair bundle (Adato *et al.*, 2002) (Table 4.13) with complete loss of *clrn1* in knock-out models resulting in attenuated, but not completely lost hair bundle function (Gopal *et al.*, 2015). Clarin-1 has been therefore proposed as essential for the early stages of hair cell development but not required to maintain the integrity of hair bundles. This is consistent with the effects of recessive pathogenic *CLRN1* variants in USH3 individuals (Gopal *et al.*, 2015; Geng *et al.*, 2012). USH3 individuals are born hearing, but fail to maintain proper hearing function (Plantinga *et al.*, 2005) (Table 4.13). This suggests that they develop functional hair cells but cannot maintain them, leading to progressive loss of hair bundle structure and function over time, resulting in post-lingual hearing loss. This might also explain why no association was found to any of the hearing abilities tested (low and mid-frequency and risk of OME) at 7 years of age when normal/close to normal hearing function is still maintained. This explanation is further supported by the highly variable type and degree of progressive hearing

loss amongst USH3 patients showing asymptomatic to moderate hearing presentation at young age (Plantinga *et al.*, 2005; Abu-Ameerh *et al.*, 2020). The post-lingual effects of *CLRN1* on hearing abilities therefore cannot explain the observed susceptibility to pre-lingual and early lingual communication delays in children with risk *CLRN1* common variants (Figure 4.8c). *CLRN1* has not been shown to be associated with language abilities in the literature and its expression across different regions of the brain is negligible (Table 4.13). Further research is needed to investigate the true effect of common *CLRN1* variants on early language and the molecular mechanism behind it.

4.3.3. *PCDH15* common variants and language abilities

Common *PCDH15* variants were found to be directly, but only marginally associated with hearing abilities (to low and mid-frequencies) and DLD risk (Figure 4.8b). Children who carried these risk variants were also more likely to require special help with their communication and interaction skills at school (11- 13 years of age), compared to children who did not carry the risk variants. Protocadherin-15 is a structural protein at the tip link of stereocilia (Alagramam *et al.*, 2011), also expressed in hair cell synapses and spiral ganglion neurons (SGN), suggesting a role in synaptic maturation (Zallocchi *et al.*, 2012). Defective *pcdh15* in knock-out mice show disordered arrangements of stereocilia in hair cells and a reduction in the number of SGN (Washington *et al.*, 2005). Phenotypically USH1F affected patients present with congenital and profound hearing loss (Ahmed *et al.*, 2001). Pathogenic variants within *PCDH15* have also been associated with poor cochlear implantation outcomes, with individuals with bi-allelic *PCDH15* mutations showing poorer auditory receptive ability, speech perception and speech intelligibility compared to controls (Wu *et al.*, 2015). The poorer outcomes were attributed to the preferential expression of *PCDH15* in SGN compared to other parts of the cochlea, with biallelic mutations leading to more severe pathology affecting the auditory neural pathway (Nishio, Takumi and Usami, 2017). Libé-Philippot *et al.* showed that *Pcdh15* is expressed in GABAergic interneurons of the developing auditory cortex (Libé-Philippot *et al.*, 2017). Moreover, *Pcdh15* knock-out mice displayed impaired interneuron development, directly as a result from an intrinsic role of *Pcdh15* in the developing auditory cortex (before the onset of hearing), rather than as a consequence of the peripheral auditory impairment in the knock-out mice (Libé-Philippot *et al.*, 2017). According to the Human Brain Transcriptome Project, *PCDH15* shows low expression in embryonic development across different regions of the brain, which increases and peaks in the cerebral cortex between 19 and 24 prenatal weeks, followed by relatively low expression postnatally (Table 4.13). Based on this, a plausible hypothesis could be that the observed association to DLD risk (which may also lead to an increased need of special help with communication and interaction) in children carrying *PCDH15* risk common variants occurs directly through the auditory neural pathway and possibly as a result

of interneuron deficit in the auditory cortex. Larger samples sizes would, however, be needed to replicate this finding at a significant level and to further study the biology of this relationship.

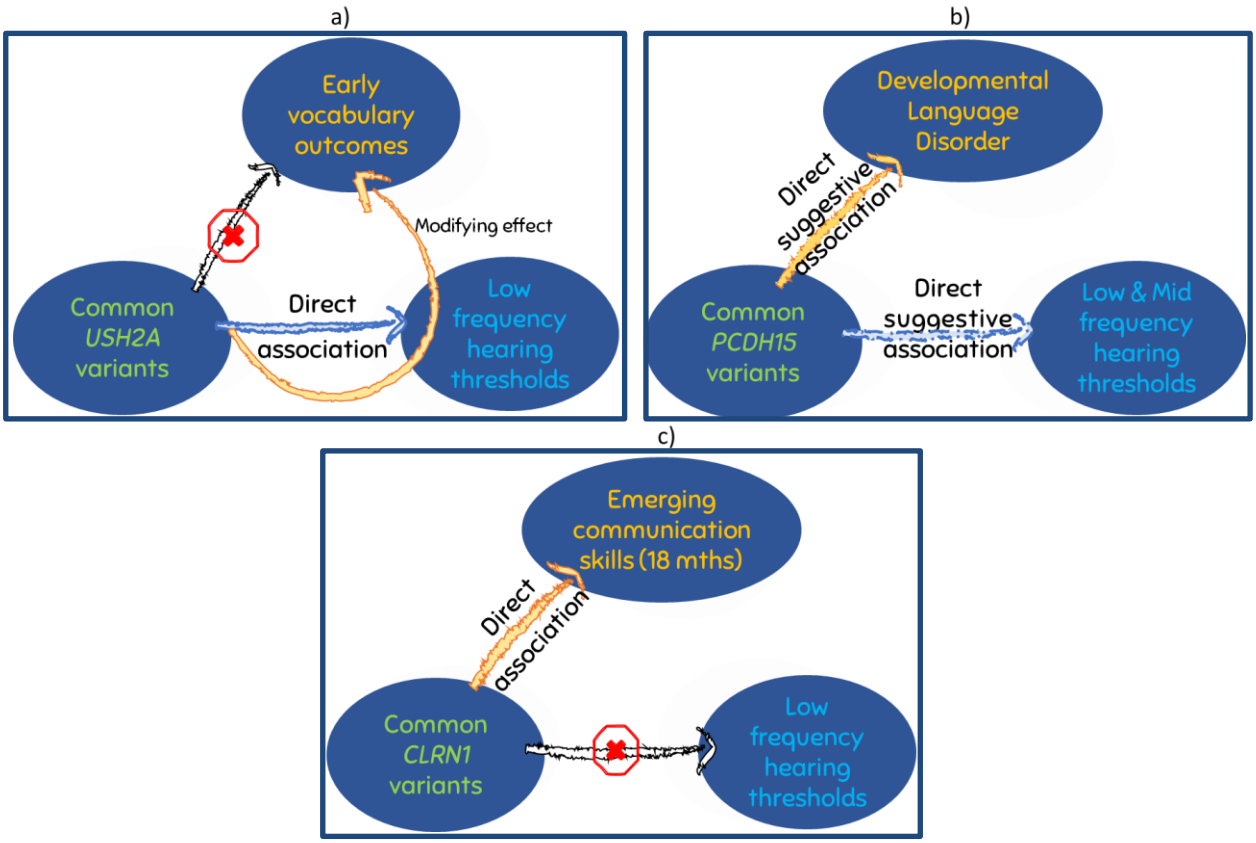

Figure 4.8. USH common variants association models.
a) *USH2A*: altered hearing thresholds combine with existing genetic factors to moderate the risk of early vocabulary outcomes
b) *PCDH15* and c) *CLRN1*: risk variants directly affect language outcome (b) and early communication skills (c).

4.4. Conclusion

In summary, in Results Chapter 2, we hypothesised that common variants in Usher genes form part of a complex model where common risk variants contribute to the susceptibility to altered hearing and/or language ability or skills (H2). This hypothesis is confirmed by the presence of direct and indirect associations between *USH2A*, *CLRN1* and *PCDH15* common SNPs and low-frequency hearing thresholds, early language markers and language outcomes such as DLD. This further supports a risk model for hearing difficulties such as auditory processing, which may indirectly contribute to language difficulties, involving common genetic variants. This is in line with complex models of genetic contribution to other neurodevelopmental and learning/hearing disorders such as language disorders (Gialluisi *et al.*, 2014; Luciano *et al.*, 2013), dyslexia (Gialluisi *et al.*, 2020), ASD (Arking *et al.*, 2008; Grove *et al.*, 2019) and ARHL (Wells *et al.*, 2019).

5. Results Chapter 3: Gene-based analyses of rare variants

5.1. Rationale

Findings from Results Chapter 2 indicated that common variation across some USH genes (*USH2A*, *CLRN1* and *PCDH15*) play a complex role in low-frequency hearing abilities and/or language development. Research has shown that rare and common variants can both play a role in complex disorders (Fritsche *et al.*, 2016; Chen *et al.*, 2017). This led to Hypothesis 3 (H3) where in order to gain a complete picture of the role of Usher gene variation in a complex model, we hypothesised that multiple rare variants across coding USH gene regions have an effect on hearing, auditory processing and/or language abilities. Therefore, H3 considers the combinatory effect of rare risk variants to susceptibility to disease or altered ability in a single multivariate analysis.

To test H3, gene-based analyses (Section 2.4.3) were performed within RVTESTS on the eleven USH candidate genes (*USH2A*, *CLRN1*, *CDH23*, *PCDH15*, *ADGRV1*, *USH1C*, *USH1G*, *MYO7A*, *WHRN*, *CIB2*, *HARS*) using whole genome sequence data from ALSPAC UK10K core cohort (Section 2.3.4). The phenotypes tested for association were the same three measures of hearing (low-frequency hearing, mid-frequency hearing and OME status) and the same five measures of language (early communication skills, early vocabulary size, NWR, VIQ, DLD status) as tested in Results Chapter 2 (Section 2.2.1), allowing direct comparison of the effects of rare versus common variants (Figure 2.15).

5.2. Results

Collectively gene-based association results showed that rare variants within *USH2A* were suggestively associated with mid-frequency hearing abilities, whereas rare variants within *CLRN1* and *ADGRV1* were suggestively associated with markers of language (Table 5.1). The sections below review each of the significant genes one by one.

Table 5.1. USH genes gene-based association testing results using Burden-Zeggini tests

	Rare variants (MAF≤1%) BZ model							
Gene	**OME@ 7Y**	**Low Freq _min**	**Mid Freq _min**	**Comm_ score@ 1.5Y**	**Vocab score @3Y**	**VIQ @8Y**	**NWR @8Y**	**DLD**
USH2A	0.09	0.61	0.0073	0.61	0.90	0.14	0.70	0.25
CLRN1	0.96	0.86	0.33	0.87	0.85	0.0069	0.46	0.12
ADGRV1	0.34	0.10	0.58	0.07	0.13	0.24	0.00095	0.30
HARS	0.20	0.99	0.91	0.81	0.30	0.65	0.72	0.88
WHRN	0.07	0.03	0.02	0.21	0.88	0.91	0.40	0.44
PCDH15	0.05	0.28	0.25	0.67	0.48	0.73	0.75	0.87
CDH23	0.26	0.68	0.36	0.65	0.36	0.10	0.03	0.94
USH1C	0.99	0.44	0.82	0.08	0.85	0.12	0.78	0.01
MYO7A	0.59	0.92	0.78	0.68	0.35	0.83	0.64	0.68
CIB2	0.67	0.62	0.83	0.79	0.87	0.33	0.69	0.37
USH1G	0.05	0.09	0.06	0.74	0.74	0.17	0.30	0.64

P-values for each hearing and language outcome are reported. Grey cells indicate p-values < 0.01.

5.2.1. Rare *USH2A* variants are suggestively associated with altered mid-frequency hearing thresholds

USH2A showed an increased rare variant burden in relation to altered mid-frequency hearing thresholds (gene-based p= 0.0073) (Table 5.2). Sub-grouping rare *USH2A* variants by function and repeating the Burden-Zeggini (BZ) association test showed that the likely rare variant drivers of association reside within *USH2A* exons (p= 0.0005) (Table 5.2). Further dividing *USH2A* rare exonic variants into types (missense, stop-gain, splice-site and synonymous) showed that the suggestive association to mid-frequency hearing threshold was driven by rare missense *USH2A* variants (p= 0.0017) (Table 5.2). No associations were observed with language outcomes (Table 5.2).

Table 5.2. *USH2A* gene-based association testing results using Burden-Zeggini, grouping variants by function in relation to mid-frequency hearing thresholds.

***USH2A* variant subgroups** (Transcript: ENST00000366943.2)		**No. variants**	**Mid Freq _min** (association p-value)
Baseline	rare variants	5361	0.0073
By function	rare intronic	5178	0.02
	rare regulatory	45	0.01
	rare exonic	138	0.0005
	rare missense	87	0.0017
	rare stop-gained	3	0.95
	rare splice-site	5	0.05
	rare synonymous	43	0.10

The number of variants tested within each subgroup and p-values are reported.
Grey cells indicate p-values < 0.01.

5.2.2. Rare *CLRN1* variants are suggestively associated with altered Verbal IQ

Suggestive association was found between rare *CLRN1* variants and Verbal IQ score at 8 years of age (p= 0.0069) (Table 5.3). Sub-grouping rare *CLRN1* variants by function and repeating the BZ analysis showed nominal association with MinP= 0.012 for rare intronic variants (Table 5.3). Next, using Haploview, the *CLRN1* variants were grouped by position, showing two broad blocks of LD; a 3' block (incorporating five smaller blocks in tight LD) and a 5' block (Figure 5.1). The BZ test was repeated with the rare variants residing in the 5' and 3' LD blocks. Again, no association with p< 0.01 was observed with VIQ at 8 years.

Table 5.3. *CLRN1* gene-based association testing results using Burden-Zeggini, grouping variants by function and position in relation to Verbal IQ at 8 years.

***CLRN1* variant subgroups** (Transcript: ENST00000327047)		**No. variants**	**VIQ @8Y (association p-value)**
Baseline	rare variants	344	0.0069
By function	rare intronic	324	0.012
	rare exonic	6	0.046
	rare regulatory	14	0.372
By position	5' LD block rare	118	0.133
	3' LD block rare	178	0.016

The number of variants tested within each subgroup and p-values are reported.
Grey cells indicate p-values < 0.01.

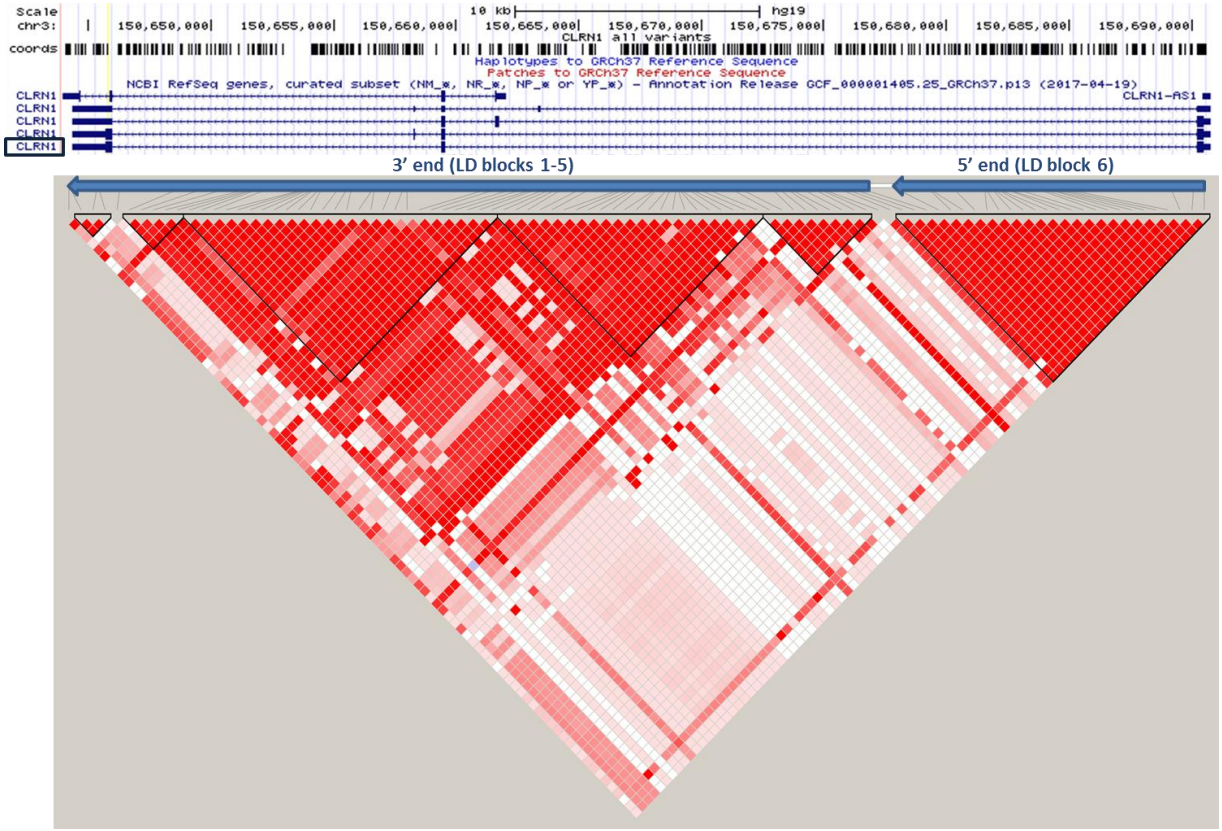

Figure. 5.1. Linkage disequilibrium analysis of *CLRN1* SNPs (plot generated using Haploview). Linkage disequilibrium is displayed by standard colour scheme, with bright red for very strong LD (LOD ≥2, D' =1), pink red (LOD ≥2, D' <1) and white for no LD (LOD < 2, D' <1). Six LD blocks were generated and further grouped into 3-and-5-prime end LD blocks. Canonical *CLRN1* transcript is marked in black box according to UCSC browser.

5.2.3. Rare intronic variants in *ADGRV1* are suggestively associated with altered Nonword repetition score

Suggestive association was observed between rare variants across *AGDRV1* and NWR at 8 years (p= 0.00095) (Table 5.4). Further gene-based analysis showed that this association was driven by rare intronic variants (p=0.00122). As the *ADGRV1* gene is relatively large (90 exons), Haploview software could not generate an LD plot. Instead, the annotated rare variants within the *ADGRV1* introns were split by position manually. Rare variants within the 5' end of the gene formed rare intronic 5' end (including 1759 variants within introns 1 to 82) and the rare variants within the 3' end of the gene formed rare intronic 3' end (including 1814 variants within intron 83 to intron 89) (Figure 5.2). Repeating the BZ model on both subgroups of variants resulted in rare intronic variants located within intron 83 to intron 89 showing a suggestive signal of association (p=0.00387). This group of variants were further divided into rare variants within intron 83 to 85 (N=1205 variants) and rare variants within intron 86 to 89 (N=609 variants). Repeating the BZ model for the 2 groups showed that the signal of increased burden to altered nonword repetition was likely resulting from intron variants between introns 83-85 (towards the middle of the gene). The rare variants within the two largest introns of the gene (intron 83 and intron 85) were separately run through the BZ model and showed that the association signal in *ADGRV1* was most heavily concentrated in rare variants within intron 85 (p= 0.00041).

5.2.4. Rare Variants in *HARS*, *WHRN*, *PCDH15*, *CDH23*, *USH1C*, *MYO7A* *CIB2* and *USH1G* show no associations to the hearing and language outcomes tested

Gene-based analyses showed no significant or suggestive associations of rare *HARS*, *WHRN*, *PCDH15*, *CDH23*, *USH1C*, *MYO7A*, *CIB2* and *USH1G* variants with the tested hearing or language outcomes (Table 5.1).

Table 5.4. *ADGRV1* gene-based association testing results using Burden-Zeggini, grouping variants by frequency, function and position in relation to nonword repetition score at 8 years.

	***ADGRV1* variant subgroups** (Transcript: ENST00000405460)	**No. variants**	**NWR @8Y** (association p-value)
Baseline	rare variants	3718	0.00095
By function	rare intronic	3573	0.00122
	rare exonic	137	0.429
	rare regulatory	8	0.037
By position	rare intronic 5' end (intron 1-82)	1759	0.016
	rare intronic 3’ end (intron 83-89)	1814	0.00378
	rare intronic 3' end (intron 83-85)	1205	0.00399
	rare intronic 3' end (intron 86-89)	609	0.077
	rare intron 83	557	0.460
	rare intron 85	553	**0.00041**

The number of variants tested within each subgroup and p-values are reported. Grey cells indicate p-values < 0.01.

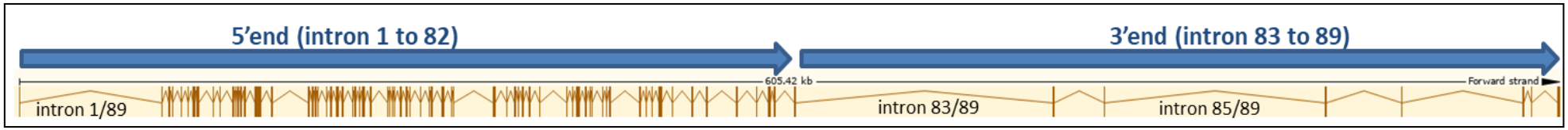

Figure. 5.2. *ADGRV1* gene-based analysis by position (ENST00000405460 Ensembl transcript is made of 90 exons and 89 introns). For gene-based analysis by position, the *ADGRV1* introns were split into 5’ and 3’, covering the start of the gene and the end of the gene.

5.3. Discussion

Complex disorders such as auditory processing disorder and speech/language disorders represent a continuum with high heterogeneity, supported by findings in Results Chapters 1 and 2. Gene-based analyses, applied here, offer further modelling of complex effect as they allow the consideration of the combined effect of rare risk factors across the entire gene. The results revealed that rare missense *USH2A* variants were suggestively associated with mid-frequency hearing abilities, whereas rare *CLRN1* and rare *ADGRV1* variants (concentrated within intron 85/89) showed suggestive associations to VIQ and NWR performance, respectively (Table 5.6). No association was detected between rare variants in *PCDH15*, *CDH23*, *USH1C*, *WHRN*, *MYO7A*, *CIB2*, *USH1G* and *HARS* and the tested hearing and language markers (Table 5.6).

Table 5.6 Summary results for the combined effect of rare variants in 11 Usher genes on hearing and language phenotypes plus wider effects and protein function

Gene	Rare variants & hearing	Rare variants & language	Common variants & hearing	Common variants & language	Effect of biallelic mutations on hearing in USH syndrome	Protein Function	Protein Localisation	Cochlea expression	Brain expression (Human Brain Transcriptome Project)
USH2A	suggestive mid-frequency	none	low-frequency	Indirect (early vocab)	congenital & moderate hearing loss	cell adhesion & signalling	stereocilia ankle links	transient (embryonic development in ankle links)	negligible
CLRN1	none	suggestive (VIQ)	none	direct (early comm)	post-lingual hearing loss	cell adhesion	hair bundle	embryonic & postnatal hair cells	negligible
PCDH15	none	none	suggestive low/ mid-frequency	direct (DLD)	congenital & profound hearing loss	cell adhesion, upper tip link	stereocilia lateral, kinociliary and tip links	transient (embryonic in lateral links) and postnatal in mature hair cells	prenatal (CBC); postnatal (HIP, AMY, MD, NCX, STR)
CDH23	none	none	sporadic low/mid-frequency	direct sporadic	congenital & profound hearing loss	cell adhesion, lower tip link	stereocilia lateral, kinociliary and tip links	transient (embryonic in lateral links) and postnatal in mature hair cells	postnatal (CBC, MD)
ADGRV1	none	suggestive (NWR)	sporadic low/mid-frequency	direct sporadic (DLD)	congenital & moderate hearing loss	cell adhesion	stereocilia ankle links	transient (embryonic development in ankle links)	early prenatal (STR, AMY, HIP); postnatal (HIP, AMY, NCX, STR)
USH1C	none	none	sporadic low/mid-frequency	direct sporadic	congenital & profound hearing loss	scaffold protein	stereocilia upper tip link and synapse	embryonic & postnatal hair cells	prenatal (small peak in HIP); postnatal (HIP, STR AND MD)
WHRN	none	none	none	none	congenital & moderate hearing loss	scaffold protein	stereocilia ankle link, stereocilia tip & synapse	transient (embryonic development in ankle links)	N/A
MYO7A	none	none	none	none	congenital & profound hearing loss	molecular motor protein	hair cell cytoplasm, hair bundle, upper tip link density	embryonic & postnatal hair cells	N/A
CIB2	none	none	none	none	congenital & profound hearing loss	intracellular calcium signalling	stereocilia near tip	no data on embryonic development/ in mature hair cells	N/A
USH1G HARS	none	none	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Brain expression data is from the Human Brain Transcriptome Project (https://hbatlas.org/pages/hbtd).
CBC= cerebellar cortex; HIP= hippocampus; AMY= amygdala; MD= mediodorsal nucleus of the thalamus; NCX= neocortex; STR= striatum.

5.3.1. *USH2A* rare variants and hearing abilities

USH2A is directly associated with hearing. While common variants show strong associations with low-frequency hearing abilities and indirect associations with vocabulary size, rare variants are associated with mid-frequency abilities (perhaps just due to insensitivity of these metrics to detect associations with low-frequency) (Table 5.6). The *USH2A* findings are consistent with emerging evidence that different variant types within the same gene can associate with variable and different outcomes (Lenassi *et al.*, 2015; Molina-Ramírez *et al.*, 2020; Toma *et al.*, 2018). This forms an allelic hierarchy of disease-causing high impact variants and complex risk variants as examined in this study, and represents a shift from Mendelian genetic models. If two rare pathogenic *USH2A* disease-causing variants are inherited in a recessive form, that will lead to the presentation of Usher Syndrome or a nonsyndromic retinopathy (Lenassi *et al.*, 2015). However, if only one pathogenic variant is inherited in a heterozygous form alongside common risk and rare coding *USH2A* variants with marginal effects upon protein function, this is likely to contribute to subtle changes in the processing of low-frequency sounds (opposite frequency end to the high-frequency hearing loss observed in USH2). Such subtle changes would not necessarily be detected in a clinical setting, where the focus would be on USH2-related high-frequency hearing loss. While it is unlikely that such subtle changes in hearing thresholds at low frequencies will directly lead to language disorder, we hypothesise that mild changes in low-level hearing can exert a snow-ball effect that disrupts higher order communicative processes. This model is analogous to the model of persistent OME, which in itself does not lead to a language disorder, but may represent a risk factor if recurrent (Rosenfeld *et al.*, 2016).

5.3.2. *CLRN1* rare variants and language abilities

Both gene-based analyses and single common SNPs analyses of *CLRN1* variation showed association with language markers: significant association between common *CLRN1* variants and early language communication deficits and marginal association of rare variants with Verbal IQ at 8 years (Table 5.6). The finding that different variants (common vs rare) correlated with different language outcomes might reflect the complex nature of association studies and defining phenotypes, which on an individual level do not always reflect aetiology. The absence of any association to hearing measures could be explained by the preserved normal hearing function in childhood even for biallelic *CLRN1* mutations known to cause post-lingual hearing loss in USH3 (see Results Chapter 2). To better understand the involvement of *CLRN1* in language abilities, further genetic analyses using larger databases would need to be performed.

5.3.3. *ADGRV1* rare variants and language abilities

Association results of both common and rare variants within *ADGRV1* suggest that intronic risk variants, located between introns 83 and 86 may be important factors for language abilities.

Gene-based analyses of rare variants showed an increased burden to NWR likely represented by *ADGRV1* rare variants within intron 85. Sporadic marginal association was observed between a common *ADGRV1* SNP (rs2007538) located in intron 83 and mid-frequency hearing abilities and another common SNP (rs949787) in intron 86 and DLD, suggesting the variants are proxies (Table 5.6). Biallelic pathogenic variants in *ADGRV1* cause USH2C with congenital moderate to severe hearing impairment in the first or second decade of life (Weston *et al.*, 2004). Loss of *Adgrv1* in Vlgr1/del7TM mice result in lack of ankle links which disturbs the organisation of the hair bundles leading to profound deafness by 3 weeks of age (McGee *et al.,* 2006), thus recapitulating the symptoms found in USH2C patients. Moreover, a homozygous truncating mutation in the mouse ortholog *Adgrv1* (*Adgrv1*$^{frings/frings}$ mouse), was found to cause audiogenic seizures (Skradski *et al.*, 2001). In humans 5q14.3 microdeletions, incorporating *ADGRV1*, and heterozygous missense *ADGRV1* mutations are identified in patients with febrile seizures and myoclonic epilepsy (Myers *et al.*, 2018; Han *et al.*, 2020). While most reported homozygous mutations associated with USH2C result in frameshift/truncation (Weston *et al.*, 2004; Hilgert *et al.*, 2009; Ebermann *et al.*, 2010; Besnard *et al.*, 2012), vast majority of mutations leading to seizures are heterozygous missense (Myers *et al.*, 2018; Han *et al.*, 2020) suggesting that the complete absence or lack of function of ADGRV1 leads to USH2C while protein dysfunction caused is likely to result in seizures via a different molecular pathway.

ADGRV1 belongs to the family of G-protein coupled receptor, known to be expressed in the central nervous system (CNS) and to play a role in neurodevelopment, with members of the family associated with ADHD (ADGRL3) and schizophrenia (ADGRB3) (McMillan *et al.*, 2002; DeRosse *et al.*, 2008; Arcos-Burgos *et al.*, 2010). The exact mechanisms by which ADGRV1 is involved in epilepsy is not known, but it is suggested it might be through its epilepsy-related repeat (EAR) domain, which is the same as the functional domain found in the *LGI1* gene, known to cause autosomal dominant lateral temporal epilepsy with auditory features (Scheel, Tomiuk and Hofmann, 2002). The EAR domain is thought to be important in attaching to an antiepileptic ligand or in interfering with synaptogenesis or axon guidance for its critical role in developing epilepsy (Scheel, Tomiuk and Hofmann, 2002). All reported individuals harbouring *ADGRV1* pathogenic variants resulting with myoclonic epilepsy also had intellectual disability or developmental delay, which might explain the effect of rare variants on NWR observed here.

5.4. Conclusion

Results Chapter 3 was based on the hypothesis that rare variants in USH genes play a role in a complex model where multiple rare risk variants in combination contribute to the susceptibility to altered hearing and/or language ability or skills (H3). This hypothesis is confirmed by the presence of association (increased burden) between *USH2A*, *CLRN1* and *ADGRV1* rare variants and mid-frequency hearing thresholds or language markers such as Verbal IQ and nonword repetition. This combines with evidence from Results Chapter 2 and demonstrates the important role of common and rare variants within some Usher genes in complex models of hearing, auditory processing and language, suggesting that there is no one single risk variant, but a complex mix of variation across the USH genes might explain some of the APD risk. This combined risk is also in line with wider evidence from neurodevelopmental disorder risk involving the influence of common genetic variants in combination with rare variants (Villanueva *et al.*, 2015; Chen *et al.*, 2017; Satterstrom *et al.*, 2020).

6. Results Chapter 4: Phenotype-driven rare Mendelian analysis

6.1. Rationale

Based on H1 in Results Chapter 1, pathogenic coding variants in the candidate USH genes in a heterozygous form showed an increased risk of delayed language milestones and subtle hearing difficulties, but no clinical distinct carrier phenotype. Additionally, Results Chapter 3 showed that rare variants may be important in some USH genes with regards to language and hearing development. These genotype-driven approaches were, however, restricted to the candidate USH genes and would have missed detecting many other genes that may also play a role. Therefore, in Results Chapter 4, we hypothesise that difficulty discriminating words in noisy environment (as a potential sign of suspected APD- sAPD) in a small number of ALSPAC children can be explained by rare coding pathogenic variants on a genome-wide level using a phenotype to genotype approach. This is the first study that considers the direct effect of rare pathogenic variants on a suspected APD phenotype, which has not been addressed in the literature before. Therefore, this is an exploratory study, presenting preliminary results.

For the purpose of Results Chapter 4, the ALSPAC UK10K core cohort was screened to produce a sAPD cohort of individuals whose word discrimination threshold in quiet was within expected ranges, but whose word discrimination threshold in noise was elevated 1SD above the mean (Section 2.3.5). The performance of identified sAPD individuals on neurodevelopmental ALSPAC measures (available through the ALSPAC phenotype dataset- Section 2.2) was compared to control individuals, allowing the investigation of the wider phenotype that may present together with difficulties recognising speech in noisy environment (Figure 2.15). Potentially deleterious coding variants within each sAPD individual were identified through VCFtools and annotated through SnpEff and wAnnovar (Section 2.5.1) and ranked for pathogenicity following ACMG guidelines (Section 2.4.4).

6.2. Results

6.2.1. sAPD cohort characteristics

Characterising the sAPD individuals at a broad neurodevelopmental level showed that, as a group, sAPD individuals performed below expected on a measure of early vocabulary at 2 and at 3 years (ALSPAC codes ke643 and kg865) (Cohen's d= 0.48; 95% CI= 0.08- 1.05 for vocabulary at 2 years and Cohen's d= 0.65, 95% CI= 0.06- 1.23 for vocabulary at 3 years) (Table 6.1a) (Figure 6.1a). Although the effect size of the difference between the means of the sAPD group and the control group is considered as medium, the 95% CI for both measures spans the value of 0 (95% CI= 0.08- 1.05 and 95% CI= 0.06- 1.23), which indicates lower confidence (<95%) in the effect sizes, likely due to the small sample size.

Analysis of available discrete measures showed that children with sAPD are more likely to suffer from otitis media with effusion/abnormal middle ear pressure very early in life compared to controls (RR= 1.6, 95% CI= 1.08- 2.399 for OME at 1.5Y and RR= 1.49, 95% CI= 0.604- 3.66 for OME at 2.5 Y) (Table 6.2). However, at 7Y the sAPD group showed the opposite results: sAPD individuals were less likely to suffer from otitis media with effusion/abnormal middle ear pressure compared to controls (RR= 0.493, 95% CI= 0.135- 1.807) (Table 6.2 & Figure 6.1b), suggesting that risk of OME was transient and only elevated during early life. Teachers rated the general ability of sAPD children as average or below average more often than expected (RR= 2.1943, 95% CI= 1.06- 4.53) and expressed complaints towards the sAPD children more often than expected (RR= 1.57, 95% CI= 0.55- 4.46) (Table 6.2 and Figure 6.1c-d). sAPD children were much more likely to also have ADHD (RR= 13, 95% CI= 2.00- 84.34) and DLD (RR= 2.06, 95% CI= 0.56- 7.66) (Table 6.2 & Figure 6.1e-f).

Table 6.1a. Quantitative measures of language, reading and cognition in individuals with sAPD compared to controls from ALSPAC dataset (lower scores represent poor performance)

Measure	Age	Range of scores (sAPD)	Mean score (sAPD)	No controls	Range of scores(controls)	Mean score (controls)	SD (controls)	5th percentile	No sAPD below 5th percentile
Vocabulary Score	2 years	35 - 228	136.2308	163	22 - 246	161.23	50.27	79.2	4 of 13
Vocabulary score	3 years	0 - 246	215.75	164	132 - 246	233.28	20.69	195	2 of 12
Plurals score	3 years	5 - 12	9.82	163	5 - 12	10.53	1.74	7	1 of 11
Past tense score	3 years	11 - 42	34	164	3 - 42	34.62	9.43	12	1 of 10
Word combination score	3 years	0 - 26	21.5	162	0 - 26	22.75	4.08	16	1 of 12
Language score	3 years	216 - 324	303.4	161	183 - 326	301.92	28.86	244.4	1 of 13
Reading score	7 years	11 - 43	28.0769	164	6 - 46	30.77	8.05	15.25	1 of 13
Spelling score	7 years	2 - 15	7.2308	164	0 - 15	9.02	3.99	2	0 of 13
Nonword Repetition	8 years	3 - 12	6.77	163	1 - 12	7.91	2.41	4	1 of 13
WOLD comprehension	8 years	4- 10	7.307692	163	2- 14	8.06	1.85	5	1 of 13
WISC - Verbal IQ	8 years	83 - 144	105.6923	161	74 - 151	112.16	16.21	86.2	1 of 13
WISC - Performance IQ	8 years	69 - 123	102.9231	160	55 - 137	102.9	16.74	75.1	1 of 13
WISC - Total IQ	8 years	77 - 137	104.7692	160	68 - 144	108.88	15.51	88	1 of 13

Table 6.1b. Quantitative measures of hearing and friendship in individuals with sAPD compared to controls from ALSPAC dataset (higher scores represent poor performance)

Measure	Age	Range of scores (sAPD)	Mean score (sAPD)	No controls	Range of scores(controls)	Mean score (controls)	SD (controls)	95th percentile	No sAPD above 95th percentile
Air conduction Right average 0.5, 1, 2, 4 kHz	7 years	1.25 - 12.5	6.6346	151	0 - 35	8.48	4.73	16.75	0 of 13
Air conduction Left average 0.5, 1, 2, 4 kHz	7 years	0 - 15	6.9792	148	0 - 42.5	7.99	6.11	17.62	0 of 12
Low_frequency_min	7 years	0 - 20	10	138	0 - 40	10.65	5.66	20	0 of 12
Mid_frequency_min	7 years	0 - 10	3.8892	140	0 - 33.33	4.96	4.26	13.25	0 of 12
Friendship score	8 years	2 - 8	4.42	150	0 - 9	3.41	2.14	7	1 of 13

Table 6.2. Discrete measures of educational support, neurodevelopmental disorders and hearing in individuals with sAPD compared to controls from ALSPAC database

Measure	Age	N affected sAPD	Freq in sAPD	N affected controls	Freq in controls
OME/abnormal middle ear pressure (< -100 daPa)	1.5 years	7 of 13	0.538	55 of 142	0.387
OME/abnormal middle ear pressure (< -100 daPa)	2.5 years	3 of 12	0.25	34 of 150	0.227
Carer worried about child's speech	2.5 years	0 of 13	0	12 of 161	0.075
OME/abnormal middle ear pressure (< -100 daPa)	7 years	2 of 13	0.154	47 of 154	0.305
Hearing Impairment	7 years	0 of 13	0	6 of 154	0.039
Teacher's rating of child is average/below average	7 years	4 of 7	0.57	25 of 96	0.26
Child received complaints from the teacher	7.5 years	3 of 12	0.25	25 of 157	0.15
Child currently has uncontrollable tics or twitches	7.5 years	1 of 12	0.083	0 of 156	0
Child has learning difficulties requiring special arrangements at school	7.5 years	0 of 12	0	3 of 152	0.02
Child has speech problems requiring special arrangements at school	7.5 years	0 of 12	0	0 of 152	0
Child has hearing problems requiring special arrangements at school	7.5 years	0 of 12	0	4 of 152	0.026
Child has eyesight problems requiring special arrangements at school	7.5 years	0 of 12	0	0 of 152	0
Child has physical problems requiring special arrangements at school	7.5 years	0 of 12	0	0 of 152	0
Child has reading difficulties requiring special arrangements at school	7.5 years	0 of 12	0	7 of 152	0.046
Child has emotional/behavioural problems requiring special arrangements at school	7.5 years	1 of 12	0.083	2 of 152	0.013
DAWBA DSM-IV clinical diagnosis - Any ADHD disorder	7.5 years	2 of 12	0.167	2 of 156	0.013
DAWBA DSM-IV clinical diagnosis - Any oppositional-conduct disorder	7.5 years	1 of 12	0.083	4 of 156	0.026
Child has ever had speech/language therapy	7.5 years	1 of 12	0.083	17 of 156	0.109
Child stutters/stumbles when speaks	8 years	1 of 13	0.077	8 of 163	0.049
Mother told child has Dyslexia	9 years	0 of 11	0	7 of 153	0.046
B6b: Mother told child has Dyspraxia	9 years	0 of 11	0	6 of 154	0.039
B6e: Mother told child has Dyscalculia	9 years	0 of 11	0	1 of 152	0.07
DLD status		2 of 8	0.25	12 of 99	0.12

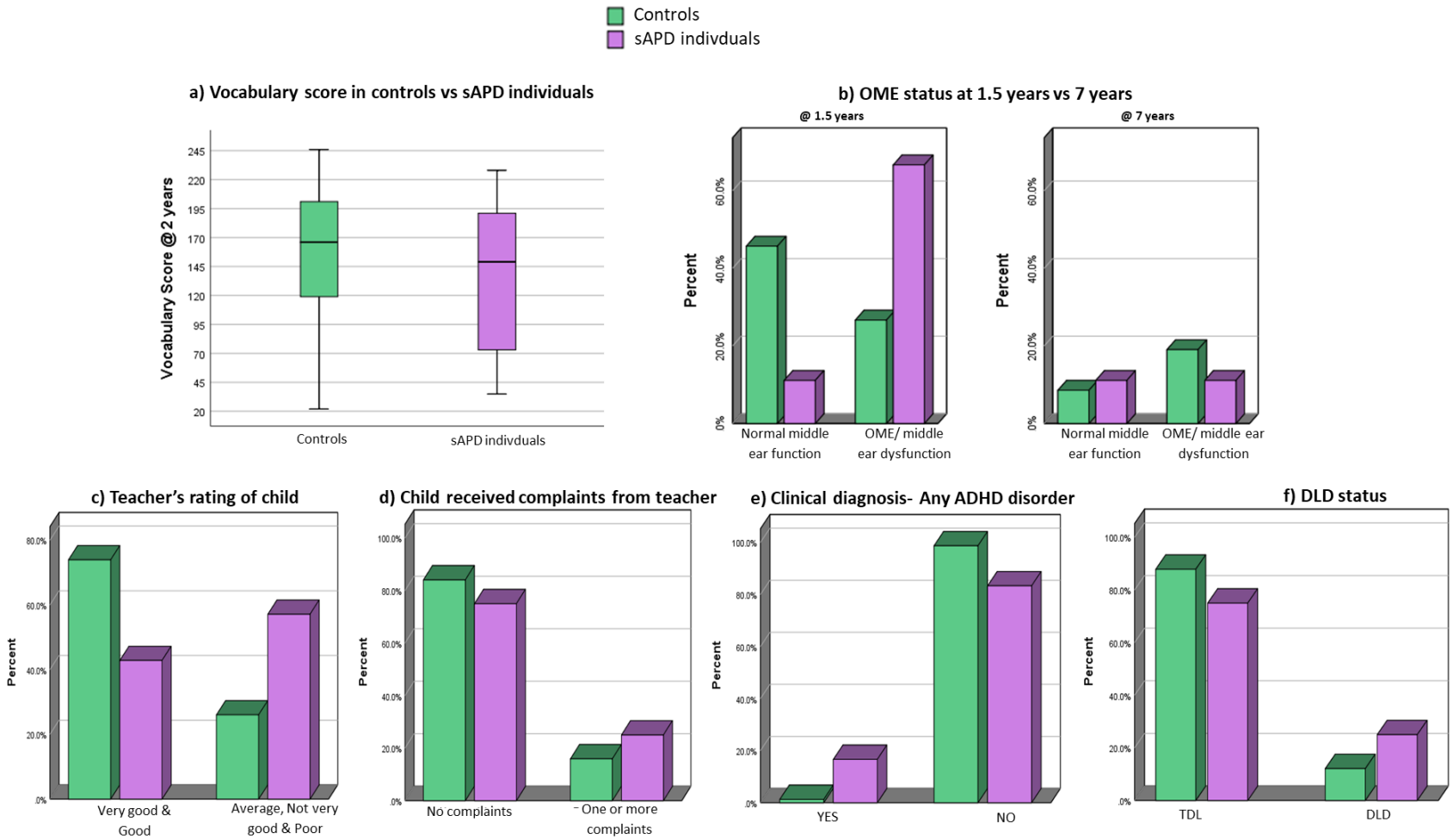

Figure 6.1. Differences in the performance of sAPD group compared to controls on a range of neurodevelopmental measures.

6.2.2. Gene variant analysis

In total 7,392,168 variants were considered in all 13 sAPD individuals (Figure 2.14 in Section 2.4.4). All variants had a quality score over 6.2 (QUAL score) and a minimum mean depth of 4.97 per individual (consistent with the UK10K low sequencing coverage). Ninety one percent of variants were SNPs, 76% of which had MAF > 0.05 and were therefore considered to be common. On average, 2,036,823 genetic variants (which were alternative to the Ref variant) were identified in each of the thirteen sAPD individuals.

Following the filtering exclusion steps (shown in Figure 2.14 in Section 2.4.4), 655 novel potentially deleterious variants were highlighted in the thirteen individuals with sAPD (Table S7). These variants were classed by function into 59 null variants (including stop-gain, frameshift and splice variants), 2 stop-loss and 4 start loss, 585 missense variants and 2 transcription factor ablation sites. Each group of variants was considered separately when applying ACMG classification for pathogenic or likely pathogenic variants. Under the secondary variant search criteria, a further 72 missense variants with MAF ranging from $6.6x10^{-5}$ to 0.047 (according to gnomAD_NFE population) were detected in the four sAPD individuals with primary pathogenic/likely pathogenic variant (individuals 17275, 465, 518, 16005).

6.2.2.1. Primary variants classed as pathogenic/likely pathogenic

Of the null LoF variants, a frameshift variant in *GRHL3* (heterozygous in individual 17275), a splice variant in *FAT4* (heterozygous in individual 465) and a frameshift variant in *IFT88* (heterozygous in individual 518) were classified as pathogenic (*GRHL3* and *FAT4*) and likely pathogenic (*IFT88*), according to the ACMG guidelines (Table 6.3). Of the missense variants, a variant in *DIAPH1* (heterozygous in individual 17275) and a variant in *NAV2* (heterozygous in individual 16005) were classified as likely pathogenic, according to the ACMG guidelines (Table 6.4).

The *GRHL3* frameshift variant (chr1: 24669442, AGACT>A, (hg19), c.1350_1353delTGAC, p.Asp451fs, ENST00000350501) is a null variant (PVS1), it is absent from the gnomAD non-Finnish European (NFE) population database (PS4), and occurs in a gene known to be involved in cleft palate/lip disorder (Van der Woude syndrome 2 MIM: 606713), which has been linked to auditory processing difficulties (Ma, McPherson and Ma, 2016) (PP4) (Table 6.3). A second likely pathogenic heterozygous missense variant was found in the same individual in the deafness gene *DIAPH1* (rs745742167) (Table 6.4).

The *FAT4* splice variant rs762672127 (chr4: 126384823, G>T (hg19), c.11899+1G>T, ENST00000394329) is a null variant (PVS1) absent from the gnomAD NFE population database (PS4), its deleterious effect is supported by multiple lines of computations evidence (PP3) and the gene is involved in a relevant disorder with affected hearing (PP4) (Table 6.3). No secondary

putative pathogenic changes were found in the same gene, however the secondary variant search detected a missense variant in *SLC9A3R1* within the same individual (Section 6.2.2.2).

The *IFT88* frameshift variant (chr13: 21265255, AG>A, (hg19), c.2445delG, p.Ile816fs, ENST00000319980) is in a hot spot for frameshift variants (PM1), it is absent from the gnomAD NFE population database (PS4), and occurs in a gene known to be involved in deafness (PP4) (Table 6.3). No secondary putative pathogenic changes were found in the same gene (or other deafness genes according to the criteria used in secondary variant search).

The *DIAPH1* variant rs745742167 (chr5: 140958708, G>A (hg19), c.880C>T, p.Arg294Cys ENST00000253811) and the *NAV2* variant rs771443047 (chr11: 20065723, G>T (hg19), c.3173G>T, p.Gly1058Val, ENST00000396087) are missense variants absent from the gnomAD NFE population database (PS4) (Table 6.4). Their deleterious effect is supported by multiple lines of computations evidence (SIFT, PolyPhen2, CADD, GERP, PhyloP and PhastCons all showing deleterious scores: PP3) and they occur at well-established functional protein domains (PM1) (Table 6.4). *DIAPH1* is a known deafness gene, with pathogenic variants known to cause Deafness, autosomal dominant 1 (MIM: 124900), while *NAV2* is a deafness candidate gene, with hypomorphic mutant mice showing impaired auditory responses (Peeters *et al.*, 2004).

Table 6.3. Null LoF variants classed as pathogenic/likely pathogenic according to ACMG classification

Ind.	Chr	Pos (hg19)	rs ID	Ref	Alt	Gene	Transcript	Functional annotation	Variant position (exon|DNA|protein)	pLI	pRec	o/e LOF	Splice score (ada|rf)	CADD_ phred	PhyloP	Phast Cons
17275	1	24669442	rs768635791	AGACT	A	*GRHL3*	ENST00000350501 (longest transcript)	frameshift	|11/16| c.1350_1353delTGAC| p.Asp451fs	0.99	0.007	0.107	.	.	.	.
465	4	126384823	rs762672127	G	T	*FAT4*	ENST00000394329 (longest transcript)	splicing	|10/16| c.11899+1G>T	1	1.7E-10	0.121	0.99| 0.94	26.2	9.626	1
518	13	21265255	rs750570861	AG	A	*IFT88*	ENST00000319980 (longest transcript)	frameshift	28/28| c.2445delG| p.Ile816fs	0	0.999	0.507	.	.	.	.

Table 6.4. Missense variants classed as pathogenic/likely pathogenic according to ACMG classification

Ind.	Chr	Pos (hg19)	rs ID	Ref	Alt	Gene	Transcript	Functional annotation	Variant position (exon| DNA|protein)	pLI	o/e LOF	Interpro domain	SIFT	Poly Phen2	CADD_ phred	GERP++ _rs	PhyloP	Phast Cons
17275	5	140958708	rs745742167	G	A	*DIAPH1*	ENST00000253811 (longest transcript)	missense	9/28| c.880C>T| p.Arg294Cys	0.92	0.197	Formin homology 3 (GBD/FH3) domain	0	1	27.5	5.11	4.46	1
16005	11	20065723	rs771443047	G	T	*NAV2*	ENST00000396087 (one of longest transcripts)	missense	14/41| c.3173G>T| p.Gly1058Val	0.99	0.165	Calponin homology domain	0	1	32	5.51	9.76	1

6.2.2.2. Secondary variants classed as pathogenic in Clinvar

The *USH2A* rs148660051 (chr1: 215963510, C>T (hg19), c.10073G>A, p.Cys3358Tyr ENST00000366943) and the *SLC9A3R1* rs35910969 (chr17: 72745313, C>G (hg19), c.328C>G, p.Leu110Val, ENST00000262613) are missense variants reported as pathogenic in Clinvar (Table 6.5). The *USH2A* variant, here identified in a single individual (17275) in a heterozygous state (Table 6.6), has been reported as pathogenic in compound heterozygous state in patients with USH2, atypical Usher and non-syndromic retinitis pigmentosa (McGee *et al.*, 2010; Le Quesne Stabej *et al.*, 2012; Garcia-Garcia *et al.*, 2011) and was also detected in Result Chapter 1 (Table 3.1). This pathogenic variant had a MAF of 0.0057 (gnomAD_NFE), which did not meet the threshold to be included in the primary variant filtering (Figure 2.14). The *SLC9A3R1* variant detected here in a single individual in heterozygous state (Table 6.5) has been previously reported as pathogenic in 2 unrelated patients (1 female and 1 male) with impaired renal phosphate absorption resulting in calcium nephrolithiasis and decreased bone mineral density with autosomal dominant inheritance (Karim *et al.*, 2008). While recessive pathogenic variants in *USH2A* are a well-known cause of USH2 and have been a focus in this thesis, dominant pathogenic missense variants in *SLC9A3R1* have been linked to human ARHL (Girotto *et al.*, 2019), which makes *SLC9A3R1* a potential candidate as secondary player in auditory processing risk.

Table 6.5. Secondary variants reported as pathogenic according to Clinvar

Ind.	Chr	Pos (hg19)	rs ID	Ref	Alt	Gene	Transcript	Functional annotation	Variant position (exon|DNA|protein)	AF (gnomAD NFE)	Protein Domain	SIFT	Poly Phen2	CADD_phred	GERP++_rs	PhyloP	Phast Cons	Clinvar phenotype (inheritance)
17275	1	215963510	rs148660051	C	T	*USH2A*	ENST00000366943	missense	51/73|c.10073G>A|p.Cys3358Tyr	0.0057	Fibronectin type III	0	1	28.8	5.76	7.376	1	Pathogenic USH2 & RP (AR)
465	17	72745313	rs35910969	C	G	*SLC9A3R1*	ENST00000262613	missense	1/6|c.328C>G|p.Leu110Val	0.0241	Between PDZ1 & PDZ2	0.28	0.26	13.54	1.11	-0.095	0.435	Pathogenic Nephrolithiasis/ osteoporosis (AD)

6.3. Discussion

Five pathogenic/ likely pathogenic heterozygous variants within *GRHL3*, *FAT4*, *IFT88*, *DIAPH1* and *NAV2* were identified in four individuals out of 13 with sAPD, yielding a detection rate of 30%. Under the selection criteria to identify very rare pathogenic variants (not reported in gnomAD_NFE population) within candidate genes associated with closely related phenotypes to hearing/sAPD, no such variants were detected in the remaining nine individuals with sAPD. Each individual's neurodevelopmental profile together with identified candidate genes and their potential role in APD are discussed below.

6.3.1. *GRHL3*, *DIAPH1* and *USH2A* in individual 17275

Individual 17275 was found to carry a pathogenic heterozygous frameshift variant (c.1350_1353delTGAC, p.Asp451fs) in the cleft lip/palate gene *GRHL3*, which is classified as extremely intolerant for loss of function variants (pLI and o/e scores) (Table 6.3). The same individuals was also found to harbour a heterozygous likely pathogenic missense variant in the AD deafness gene *DIAPH1* (c.880C>T, p.Arg294Cys) which was predicted to be damaging by six bioinformatics tools (Table 6.4) and another secondary heterozygous missense variant in the AR Usher syndrome gene *USH2A* (c.10073G>A, p.Cys3358Tyr), which has been previously reported as pathogenic in homozygous/compound heterozygous individuals with USH2 and/or RP (Table 6.5).

6.3.1.1. Neurodevelopmental profile of 17275

Individual 17275 showed word discrimination thresholds in quiet to be well within the expected normal range (taken from the 180 ALSPAC individuals with available scores), while his word discrimination thresholds in noise were elevated above the 95th percentile, assigning the sAPD status. Very early on (at six months), individual 17275 showed fine-and-gross motor coordination difficulties (Table 6.6). Between 2 and 4 years, both his expressive and receptive language (measured by vocabulary and grammar scores, number of unintelligible responses, verbal comprehension and VIQ scores) were below expected, indicating early language difficulties or delays (Table 6.6). The individual also suffered from recurrent middle ear infections (otitis media with effusion) between 8 months and 3.5 years of age (Table 6.6), which is a crucial time during which both receptive and expressive language develops (Paul and Roth, 2011). Recurrent OME can result in mild and temporary hearing loss, which might directly impact on developing language skills within those early years and in such a way contribute to the deficits of individual 17275 in early receptive and expressive language. Nevertheless, individual 17275 had normal pure tone audiometry for both ears (mean hearing level at 0.5, 1, 2 and 4 kHz ≤ 20 dB HL) and normal middle ear functioning recorded at 7 years and average VIQ at 8 years (Table 6.6), suggesting the

earlier ear and language problems had cleared. The individual was clear of any overt hearing loss diagnosis or any psychiatric diagnosis at 7 and 7.5 years respectively (Table 6.6).

Table 6.6. Individual 17275 neurodevelopmental profile and proposed genotype

Individual	Neurodevelopmental profile (age)	1st gene (variant)	MIM#; Inheritance; Phenotype	2nd gene (variant)	MIM#; Inheritance; Phenotype	3rd gene (variant)	MIM#; Inheritance; Phenotype
17275	-poor fine and gross motor coordination skills (6mths) -poor expressive and receptive language skills (2-4Y) -low average VIQ and Fullscale IQ; average PIQ (4Y) -average VIQ, PIQ and total IQ (8Y) -poor verbal comprehension on 1 WISC subtest (8Y) -impaired middle ear functioning (1.5-3.5Y) -recurrent middle ear infections (OME) (8m-3.5Y) -normal middle ear functioning and no middle ear infections (7Y) - hearing thresholds in normal range (7Y) -no hearing loss diagnosed (7Y) -no psychiatric clinical diagnosis (ADHD, oppositional/conduct disorder, pervasive developmental disorder, anxiety, phobia, depressive disorder) (7.5)	*GRHL3* (c.1350_1353del TGAC; p.Asp451fs)	606713; AD; cleft lip/ cleft palate	*DIAPH1* (c.880C>T; p.Arg294Cys)	-124900 (AD; Hearing loss, sensorineural (affecting all frequencies) -616632 (AR; Seizures, cortical blindness, microcephaly syndrome)	*USH2A* (c.10073G>A; p.Cys3358Tyr)	-276901 (AR; Usher Syndrome Type 2) -613809 (Retinitis pigmentosa 39)

6.3.1.2. Candidate genes in individual 17275

GRHL3

GRHL3 belongs to the family of three highly conserved grainyhead-like transcription factors (GRHL1-3) which play a crucial regulatory role in epithelial development, maintenance and homeostasis (Bray and Kafatos, 1991). While *GRHL2* has been implicated in AD human age-related hearing impairment and deafness (Peters *et al.*, 2002; Van Laer *et al.*, 2008), mutations in *GRHL3* cause AD syndromic (Van der Woude Syndrome 2) and non-syndromic isolated cleft lip and palate (CL/P) (Peyrard-Janvid *et al.*, 2014; Basha *et al.*, 2018), and neural tube defects such as spina bifida (Lemay *et al.*, 2017). Mouse *Grhl3* is expressed early in development (at E8.5) in non-neural ectoderm adjacent to the neural plate, which undergoes folding to form the neural tube with widespread expression in the surface ectoderm, progressively increasing until E15.5 (Ting *et al.*, 2003). Furthermore, mouse *Grhl3* is expressed in developing brain, localised specifically in the habenula (strongest levels), striatum and posterior lateral ventricles. The habenula is known to regulate locomotor and cognitive functions, including action planning and decision making, and operates to prevent over-stimulation of both serotonergic and dopaminergic systems (Hikosaka, 2010). Behavioural tests on mice with conditionally deleted *Grhl3* in the adult brain (as constitutive loss of Grhl3 causes early post-natal death) showed significant defects in locomotor activity with affected mice taking shorter, quicker and more frequent steps, as an indicator of hyperactivity related behaviours (Dworkin *et al.*, 2017).

A highlight in individual 17275's development is his recurrent OME episodes between 8 months and 3.5 years of age which is likely to have impacted their early language development and hearing. Moreover, OME has been linked to cleft lip/palate with higher incidence amongst individuals with cleft lip/palate (Sheahan *et al.*, 2003; Mangia *et al.*, 2019) and so it is likely that auditory processing difficulties may be secondary. A recent scoping review looking at association of comorbidities with non-syndromic CL/P diagnosis summarised eight papers in the literature that reported auditory processing difficulties in the non-syndromic CL/P population (van Eeden and Stringer, 2020). The emerging theme was that children with non-syndromic CL/P were more likely to score poorly on behavioural tests, language processing with noise distractions and discrimination of non-verbal sounds tests (presentation overlapping with auditory processing difficulties) than age matched non-CL/P controls (van Eeden and Stringer, 2020). The fact that the ALSPAC individual 17275 carried a heterozygous pathogenic variant in *GRHL3,* a gene implicated in CL/P, and is suspected of having auditory processing difficulties (because of elevated word discrimination thresholds in noise) could be suggestive of common aetiological risk factors. It also needs to be noted that there was no available data on CL/P symptoms or diagnosis for individual

17275, so this data is very preliminary and only suggestive until further in-depth investigations of the role of *GRHL3* in auditory processing are completed.

DIAPH1

DIAPH1 encodes the mammalian diaphanous-related formin-1 (mDia1), a protein that plays a role in the regulation of cell morphology and cytoskeletal organization (Al-Maawali *et al.*, 2016). It is widely expressed in embryonic mouse forebrain and brainstem (including the ventricular and subventricular zone progenitor cells) and in the cerebral cortex, basal ganglia, hippocampus, thalamus and external granular layer of cerebellum during postnatal development (Ercan-Sencicek *et al.*, 2015). Neuhaus *et al.* found *Diaph1* expression in the organ of Corti of mouse cochlea, specifically in the pillar cells (contributing to the rigidity of the organ of Corti) and at the base of OHCs and also in neuronal ear structures, including spiral ganglion neurons and the cochlear nerve (Neuhaus *et al.*, 2017). Mutations in *DIAPH1* have been associated with AD SNHL and AR microcephaly and seizures (Lynch *et al.*, 1997; Ercan-Sencicek *et al.*, 2015). There is a phenotype-genotype correlation with heterozygous, presumably truncating *DIAPH1* mutations predominantly located within the Diaphanous-autoregulatory domain (DAD) (5 out of 7 reported mutations) causing AD deafness (in some cases associated with thrombocytopenia) (Lynch *et al.*, 1997; Stritt *et al.*, 2016; Neuhaus *et al.*, 2017; Kang *et al.*, 2017; Baek *et al.*, 2012; Ueyama *et al.*, 2016), while homozygous truncating *DIAPH1* mutations located towards the middle of the gene and affecting the Formin Homology 2 (FH2) domain cause microcephaly (Ercan-Sencicek *et al.*, 2015; Al-Maawali *et al.*, 2016). The FH2 domain is the central catalytic element of formins which nucleates actin filament formation and regulates filament elongation (Higgs and Peterson, 2005). The missense heterozygous c.880C>T (p.Arg294Cys) variant in 17275 individual is predicted to affect the FH3 domain, which has not been linked to causative mutations in either AD deafness or microcephaly before (Figure 6.2). The FH3 domain, which is structurally and functionally least conserved FH domain (Wallar and Alberts, 2003) is believed to mediate the subcellular localisation of mDia proteins (Kato *et al.*, 2001). The DAD domain is highly conserved and shares a mutually exclusive binding site with Rho in the GBD/FH3 region which keeps the protein in inactive confirmation (Rose *et al.*, 2005). A small region of conserved amino acids in the GDB/FH3 domain are considered to be putative DAD binding site. A p.Ala256Asp mutation (which is close proximity to the p.Arg294Cys identified here) was found to significantly reduced the affinity of DAD binding to mDia, but it did not affect Rho binding (Rose *et al.*, 2005). This means the protein will not be able to stay in inactive conformation, which will affect the regulation of its function. However, no mutations with clinical relevance have been observed within the FH3 domain, and therefore their potential effect has not been studied in those circumstances. The p.Arg294Cys mutation changes the highly conserved arginine to cysteine residue, evidenced by the high conservation GERP, PhyloP and PhastCons scores (Table 6.5). The very low population MAF of 0.000008014 (gnomAD)

and the multiple pathogenic *in silico* predictions deem the detected *DIAPH1* variant of particular interest, especially because individual 17275A carries other two pathogenic variant in *GRHL3* and *USH2A*. On the individual's genetic background, it is possible that the *DIAPH1* variant is a hypomorph (causing loss of some of the protein's activity compared to wild type) and as such acting as a risk variant that in combination with the other two pathogenic variants, further increases the overall risk of potential auditory processing difficulties.

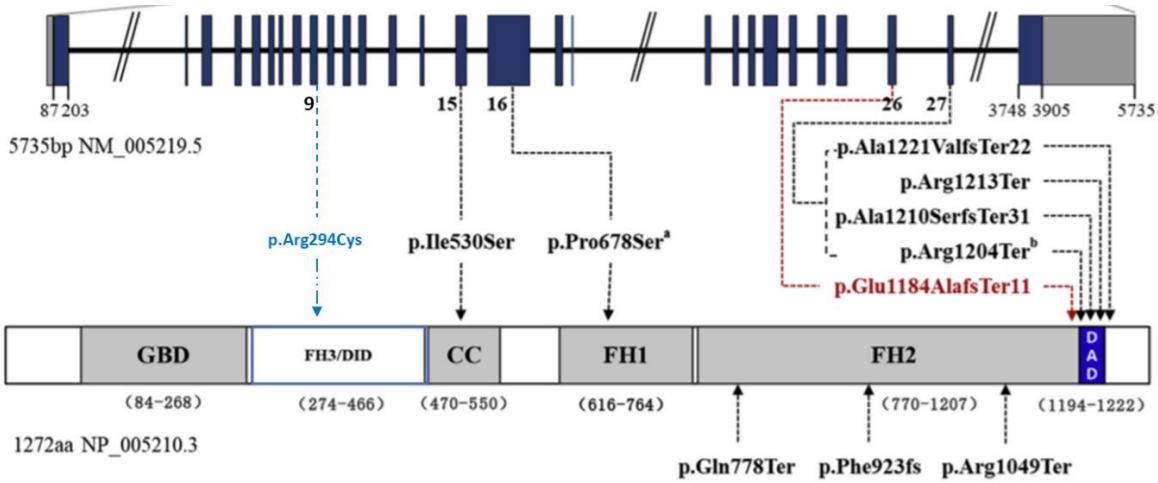

Figure. 6.2. Structure of *DIAPH1* gene (ENST00000253811) with 28 exons and corresponding protein (adapted from (Wu *et al.*, 2020)). The likely pathogenic variant detected here is located in exon 9 of the gene (highlighted in blue). Domains of DIAPH1 protein and locations of variants previously detected and are associated with heterozygous hearing loss are indicated above the protein diagram and the ones associated with homozygous microcephaly are under the protein diagram.

USH2A

USH2A is the gene responsible for AR USH2A and Retinitis Pigmentosa (Section 1.3.3.1). The variant c.10073G>A (p.Cys3358Tyr) detected here has been reported as pathogenic in both USH2 and RP, but also indicated as a RP-specific by multiple studies when in compound het state (McGee *et al.*, 2010; Le Quesne Stabej *et al.*, 2012; Garcia-Garcia *et al.*, 2011), which would challenge its role in hearing. Nevertheless, based on the identification of the same pathogenic variant in USH carriers (Results Chapter 1) and evidence from Results Chapters 2 and 3 showing that *USH2A* variation is likely to play a role in APD risk, the c.10073G>A *USH2A* pathogenic variant is likely to contribute to APD susceptibility which may also be influenced by the identified variants in *GRHL3, DIAPH1* in individual 17275.

There has been no reported functional links or detected interactions between the *GRHL3, DIAPH1* and *USH2A* genes, therefore although the finding of pathogenic variants within those three genes in individual 17275 is of particular interest, their effect (whether combinatory or individually) on sAPD phenotype is unknown. It might be that in combination, the three variants contribute to a susceptibility to auditory processing difficulties, or it might be that they are unrelated and one or

more are incidental findings, not directly involved in auditory processing aetiology (for example *GRHL3* which role in the auditory system is unknown). Further genetic and functional studies of a diagnosed APD cohort and mouse models are required to investigate any overlapping pathways and understand the molecular links, if any, between these three genes.

6.3.2. *FAT4* and *SLC9A3R1* in individual 465

Individual 465 was found to have a pathogenic heterozygous splicing variant in the planar polarity gene *FAT4* (c.11899+1G>T), which has a high intolerance score for LoF variants (pLI and o/e ratio) (Table 6.3). A secondary heterozygous missense variant, previously reported as pathogenic in heterozygous state, was also found in *SLC9A3R1*, known to have mutations that might lead to dominant inheritance of ARHL (c.328C>G, p.Leu110Val) (Table 6.5).

6.3.2.1. Neurodevelopmental profile

Individual 465 showed word discrimination thresholds in quiet to be at the higher end of the expected normal range (taken from the 180 ALSPAC individuals with available scores), while his word discrimination thresholds in noise were elevated above the 95th percentile, assigning the sAPD status. Very early on (at six months), individual 465 showed communication and gross motor coordination difficulties (Table 6.7). At 15 months he had early receptive language difficulties and between 2 and 3 years of age he also showed expressive language difficulties (Table 6.7). Later on, at 8 years, individual 465 was reported to stutter/stumble when speaking; he had a low NWR score and an average Verbal IQ, while his Performance IQ was very low (below the expected 5% of the tested population) (Table 6.7). Further assessment of cognitive function, done through the Weschler Intelligence Scale for Children (WISC) confirmed the poor performance of individual 465 on verbal and non-verbal tests: slightly worse performance than controls on vocabulary and comprehension subtests (testing the child's understanding of the meaning of different words and situations) and very poor performance compared to controls on block design and object assembly (testing the ability to copy a specific pattern and to put puzzles together) (Table 6.7). Individual 465 also showed recurrent ear problems between the age of 1.5 and 4 years which consisted of earache, ear discharge (including pus/mucus), sore and red ears and pulling/scratching of ears (Table 6.7). Hearing abilities of individual 465 were reported to get worse during a cold between 3.5 and 4.5 years and a hearing problem was noted by a health visitor at 3 years (Table 6.7). However, no clinical referral was made and the child's pure tone audiometry at 7 years was normal for both ears (mean hearing level at 0.5, 1, 2 and 4 kHz ≤ 20 dB HL) with no hearing impairment and no psychiatric disorder diagnosed (Table 6.7).

Table 6.7. Individual 465 neurodevelopmental profile and proposed genotype

Individual	Neurodevelopmental profile (age)	1st gene (variant)	MIM#; Inheritance; Phenotype	2nd gene (variant)	MIM#; Inheritance; Phenotype
465	-poor communication and gross motor coordination skills (6mths) -poor expressive and receptive language skills (2-3Y) -low Nonword repetition score (8Y) -average VIQ and very low PIQ, below average total IQ (8Y) -poor perceptual reasoning on 2 WISC subtests (8Y) -poor verbal comprehension on 1 WISC subtest (8Y) -stuttering/stumbling when speaking (8Y) -recurrent ear problems (earache, ear discharge, red ears) (1.5-3.5Y) -impaired middle ear functioning and middle ear infection (1.5Y) -hearing worsens during a cold (3.5Y) -health visitor noticed a problem with hearing (3Y) -normal middle ear functioning and no middle ear infections (7Y) - hearing thresholds in normal range (7Y) -no hearing loss diagnosed (7Y) -no psychiatric clinical diagnosis (ADHD, oppositional/conduct disorder, pervasive developmental disorder, anxiety, phobia, depressive disorder) (7.5Y) -no known disorder reported by family (dyslexia, dyspraxia, dysgraphia, dysorthographia, dyscalculia, ASD) (7.5Y) -no recognised difficulties/delays requiring special education (7.5Y)	*FAT4* (c.11899+1G>T)	616006; AR; Hennekam lymphangiectasia-lymphedema syndrome 2 615546; AR; Van Maldergem syndrome 2	*SLC9A3R1* (c.328C>G; p.Leu110Val)	612287, AD Nephrolithiasis/osteoporosis, hypophosphatemic, 2 AD Adult-related hearing loss (Girotto et al. 2019)

6.3.2.2. Candidate genes in individual 465

FAT4, the fat atypical cadherin 4, is the mammalian homolog of Drosophila Fat4, which plays a key role in vertebrate planar polarity (Saburi *et al.*, 2012). Planar polarity (as explained in Section 1.3.4.1) is the organisation of cells within the plane of a tissue, as shown by the orderly arrangements of hair cells in the cochlea. *Fat4*$^{-/-}$ mice exhibit distinctive planar polarity phenotypes, including misorientation of hair cells (predominantly in the third row OHCs) and short and deformed cochleae in inner ear (Mao *et al.*, 2011; Saburi *et al.*, 2012), illustrating the role of the *Fat4* in inner ear development. *Fat4* is expressed in all layers of the developing mouse brain, and in both the neural tube and the intervertebral discs during development (Rock, Schrauth and Gessler, 2005). *Vangl2*, a core "planar cell polarity" gene important in axon turning for innervation of OHCs (Ghimire, Ratzan and Deans, 2018), has been shown to cooperatively interact with *Fat4*, where mutating one copy of *Vangl2* significantly increased the severity of cochlear ducts shortening in *Fat4*$^{-/-}$ cochlea (Saburi *et al.*, 2012). Furthermore, the *Dchs1*, encoding another protocadherin that is a ligand to Fat4, functions with Fat4 to control tissue patterning and is essential in the development of several organs, including ear and cochlea (Mao *et al.*, 2011). Homozygous mutations in the human *FAT4* cause Van Maldergem syndrome (VMS), an AR disorder characterised by intellectual disability, craniofacial, auditory malformations resulting in hearing loss, renal, skeletal and limb malformations (Cappello *et al.*, 2013). Biallelic mutations in the same gene also cause Hennekam lymphangiectasia-lymphedema syndrome-2, which is a distinct disorder with lymphedema with overlapping features (such as intellectual disability, dysmorphic features and hearing loss in some patients) (Alders *et al.*, 2014). Cases with mild manifestations of VMS have also been reported (van der Ven *et al.*, 2017), which might explain the low PIQ in individual 465 and his early ear recurrent discomfort (earache, discharge, red/sore ears), OME at 1.5Y and concerns over a hearing problem with no overt hearing loss as a result of a monoallelic splicing LOF *FAT4* variant. As the expression of *Fat4* in mouse ear has not been studied, its expression in the mouse organ of Corti was examined using the gEAR portal (Hertzano and Orvis, https://umgear.org/), which displays data from the mouse organ of Corti at postnatal day P0 to P7. *Fat4* shows very little expression in the hair cells, but is strongly expressed in non-sensory cells at P0 (Cai *et al.*, 2015). The direct role of *Fat4* in inner ear OHCs' orientation and the *FAT4'*s causative role in human disease associated with hearing loss, makes the gene a relevant candidate for sAPD.

SLC9A3R1 encodes the Na^+/H^+ Exchange Regulatory factor 1 (NHERF1) protein, which belongs to NHERF family of scaffolding proteins. It has been shown that protein-protein interactions involving NHERFs take place in the cochlea and a *Nherf1* KO mouse displays hearing deficits with hair cell anomalies (Kamiya *et al.*, 2014). A recent study by Girotto *et al* confirmed the role of

SLC9A3R1 in the hearing system' function and development and demonstrated its role in human adult-related hearing loss (Girotto *et al.*, 2019). A missense variant in *SLC9A3R1* (c.539G>A) was detected in two unrelated patients with ARHL, characterised by severe to profound high-frequency hearing loss (Girotto *et al.*, 2019). However, the same heterozygous *SLC9A3R1* variant detected here (c.328C>G) was previously reported as pathogenic in 2 unrelated patients with impaired renal phosphate absorption resulting in calcium nephrolithiasis and decreased BMD (Karim *et al.*, 2008) (suggesting this might be an unrelated finding), but hearing abilities of those patients were not tested and therefore are unknown. The deleterious effect of c.328C>G, suggested by SIFT, PolyPhen and CADD is not very strong (Table 6.5) and the amino acid Leu at position 110 of the protein is not very conserved (Table 6.5), suggesting that the effect of the variant on the protein might not be as deleterious as suggested by Karim *et al.* to cause disease. Nevertheless, the involvement of *SLC9A3R1* in auditory processing skills is of interest as the contribution of c.328C>G to hearing abilities might be subtle enough to contribute to such difficulties and would require further investigations.

6.3.3. *NAV2* in individual 16005

Individual 16005 was found to carry a likely pathogenic heterozygous missense variant in the candidate deafness gene *NAV2* (c.3173G>T), which was predicted to be damaging by six bioinformatics tools (Table 6.4).

6.3.3.1. Neurodevelopmental profile

Individual 16005 showed word discrimination thresholds in quiet to be at the higher end of the expected normal range (taken from the 180 ALSPAC individuals with available scores), while his word discrimination thresholds in noise were elevated above the 95th percentile assigning the sAPD status. Between 2 and 3 years of age, individual 16005 showed expressive language difficulties with grammar, plurals and vocabulary scores placed in the expected bottom 5% of the tested population (Table 6.8). He also never babbled and never used gestures to get what he wanted below the age of 3, suggesting a delay in expressive language development. His Verbal IQ measured at 4 years (WPPSI) and later at 8 years (WISC) was of average performance (on a general population level), however WISC information and comprehension scores were below the expected 10% of the tested population, illustrating that some difficulties with verbal comprehension remained (Table 6.8). Individual 16005 also experienced earache before 1.5 and 2.5 years of age and hearing that deteriorated during a cold at 3.5 years but had normal pure tone audiometry at 7 years and normal middle ear functioning recorded between 1.5 and 7 years (Table 6.8).

Table 6.8. Individual 16005 neurodevelopmental profile and proposed genotype

Individual	Neurodevelopmental profile (age)	Gene (variant)	MIM#; Inheritance; Phenotype
16005	-poor expressive language skills (2-3Y) -child never babbled or used gestures to communicate (3Y) -poor receptive language (comprehension) (4Y) -average VIQ, PIQ and Fullscale IQ (4Y) -average VIQ, high average PIQ and average IQ (8Y) -poor verbal comprehension on 1 WISC subtest (8Y) -poor general knowledge on 1 WISC subtest (8Y) -some ear problems (1.5-2.5Y) -hearing worsens during a cold (3.5Y) -normal middle ear functioning and no middle ear infections (1.5-7Y) -normal pure tone air conduction (7Y) -no hearing loss diagnosed (7Y) -no psychiatric clinical diagnosis (ADHD, oppositional/conduct disorder, pervasive developmental disorder, anxiety, phobia, depressive disorder) (7.5Y) -no known disorder reported by family (dyslexia, dyspraxia, dysgraphia, dysorthographia, dyscalculia, ASD) (7.5Y) -no recognised difficulties/delays requiring special education (7.5Y)	*NAV2* (c.3173G>T; p.Gly1058Val)	None Sensory deficits in mice (Peeters et al., 2004) Associated with risk of AD (Wang et al., 2018)

6.3.3.2. Candidate gene in individual 16005

Neuron navigator 2 (Nav2) is the closest human homolog and ortholog of *Caenorhabditis elegans* UNC-53, showing a conserved function in axon elongation and cell migration (Muley et al., 2008; Stringham and Schmidt, 2009). Hypomorphic mutant mice, containing a gene trap which eliminates the expression of the full-length Nav2 transcript, are ataxic and show impaired sense of hearing (with increased startle thresholds compared to wild-type), together with defects in cranial nerve development and cerebellar development (McNeill et al., 2010; Peeters et al., 2004). A recent study looking into the expression of mouse *Nav2* in CNS throughout development showed most abundant expression in the cerebellum, hippocampus, cortex, and thalamus during late embryogenesis and early postnatal life, suggesting a role of NAV2 in CNS development (Pook, Ahrens and Clagett-Dame, 2020). In human disease, *NAV2* has been suggested as a candidate Alzheimer's disease risk gene (Wang et al., 2017a). As the expression of Nav2 in mouse ear has not been studied, its expression in the mouse organ of Corti was examined using the gEAR portal (Hertzano and Orvis, https://umgear.org/), which displays data from the mouse organ of Corti at postnatal day P0 to P7. Nav2 shows to be expressed at high levels in both sensory hair cells and in supporting cells at P0 (Cai et al., 2015), supporting its role in hearing and its candidate role in deafness and possibly sAPD.

6.3.4. *IFT88* in individual 518

Individual 518 was found to carry a heterozygous pathogenic frameshift deletion variant in the planar cell polarity gene *IFT88* (c.2445delG), which is not expected to result in loss of function (pLI=0) but might be sufficient to increase risk (Table 6.3).

6.3.4.1. Neurodevelopmental profile

Individual 518 showed word discrimination thresholds in quiet to be well within the expected normal range (taken from the 180 ALSPAC individuals with available scores), while his word discrimination thresholds in noise were elevated above the 95th percentile, assigning the sAPD status. Individual 518 showed a difficulty with very early communication (at 6 months) and "sometimes talked with words in the wrong order" at 3 years, followed by a poor score on WPSSI visual special reasoning subtest block design at 4 years (Table 6.9). However, his later language scores were within the expected rage with Verbal IQ at 8 years being above average, suggesting that the earlier language problems showed slight delay rather than a deficit. Between 8 months and 3.5 years individual 16005 showed recurrent abnormal middle ear functioning affecting both ears and leading to otitis media with effusion (Table 6.9). Later tests performed at 7 years showed normal hearing sensitivity and no sign of otitis media, suggesting the earlier middle ear malfunctioning had cleared (Table 6.9).

Table 6.9. Individual 518 neurodevelopmental profile and proposed genotype

Individual	Neurodevelopmental profile (age)	Gene (variant)	MIM#; Inheritance; Phenotype
518	-poor communication (6mths) -average VIQ, PIQ and Fullscale IQ (4Y) -poor visual spatial reasoning on 1 WPPSI subtest (4Y) -high average VIQ, average PIQ and high average Total IQ (8Y) -some ear problems (1.5) -sometimes talking with the words in the wrong order (3Y) -impaired middle ear functioning (8mths-3.5Y) -recurrent middle ear infections (OME) (8m-3.5Y) -normal pure tone air conduction (7Y) -no hearing loss diagnosed (7Y) -no psychiatric clinical diagnosis (ADHD, oppositional/conduct disorder, pervasive developmental disorder, anxiety, phobia, depressive disorder) (7.5Y) -no known disorder reported by family (dyslexia, dyspraxia, dysgraphia, dysorthographia, dyscalculia, ASD) (7.5Y)	*IFT88* (c.2445delG| p.Ile816fs)	None Candidate gene for AD craniofacial abnormalities (Tian *et al.*, 2017) AR retinal denegation (Chekuri *et al.*, 2018)

6.3.4.2. Candidate gene in individual 518

IFT88 codes for intraflagellar transport protein 88 (IFT88), which is a core component of IFT complex B and is required for the assembly and maintenance of primary cilia and flagella across species (from unicellular organisms to high-order mammals) (Rosenbaum and Witman, 2002). Ift88 is expressed in kinocilia of the organ of Corti in mouse cochlea (Jones *et al.*, 2008). Null mutations result in embryonic lethality in mice due to severe left-right symmetry defects (Murcia *et al.*, 2000), however, conditional inactivation of Ift88 in cochlea (ablating kinocilia in cochlea) causes stereocilia bundle misorientation and shortening and widening of the cochlea (Jones *et al.*, 2008), indicating the role of Ift88 in planar cell polarity regulation. Jones et al., also showed that Ift88 interacts with *Vangl2*, a core planar cell polarity gene (Jones *et al.*, 2008). Furthermore, functional studies in zebrafish inner ear hair cells have shown the association of Ift88 with USH1 proteins where the reported similarity in of hair cells phenotype in mutated *chd23*, *ush1c*, *myo7a* and *ift88* mice (bent/splayed stereocilia and fewer hair cells forming stereocilia) suggest that the genes might function in same or overlapping developmental processes (Blanco-Sánchez *et al.*, 2014). Chekuri *et al.* prosed *IFT88* as a candidate gene for AR inherited retinal denegation (resulting from two compound heterozygous mutations in two affected sisters) (Chekuri *et al.*, 2018). Interestingly, conditional loss of *Ift88* in the cranial neural crest cells has been shown to lead to craniofacial abnormalities with Tian et al. proposing *IFT88* as a candidate gene for AD craniofacial abnormalities such as cleft lip and palate with variable penetrance (in three affected siblings with heterozygous missense variant) (Tian *et al.*, 2017). The AD inheritance makes the *IFT88* frameshift variant a very plausible candidate for the presentations in individual 465: recurrent early OME which may be linked to cleft lip/palate symptoms, which can impact on hearing and language skills. There have been no reports of *IFT88* pathogenic variants causing hearing phenotypes in humans yet, however based on expression data and mouse models, the role of *IFT88* in hearing processes is strongly suggested.

It needs to be noted, however, that the *IFT88* variant c.2445delG reported here, is a frameshift variant in the last exon of the transcript and so its effect on the amount of expressed protein is questionable and would need further analysis. Moreover, similarly to individual 17275, there was no available data on CL/P symptoms or diagnosis for individual 518, so this data is very preliminary and only suggestive until a molecular overlap between cleft lip/palate, OME and auditory processing is further demonstrated.

6.4. Conclusion

The overall goal of this chapter was to identify pathogenic coding variants with large effect that might be able to explain the poor performance on recognising speech in noisy environment (as a potential sign of suspected APD) in a small number of ALSPAC children. The results here support pathogenic variants in four individuals. Of note, individual 17275 and 425 had pathogenic variants identified in multiple genes: *GRHL3, DIAPH1* and *USH2A* (individual 17275) and *FAT4* and *SLC9A3R* (individual 425). No direct interaction between those genes detected in each individual have been reported, so their overall contribution to APD-related phenotypes would need to be investigated further. The identification of pathogenic variants in candidate genes in four individuals (30% detection rate) supports the conclusion that rare pathogenic variants with large effect can explain some APD risk (in a small number of individuals selected for a specific phenotype). This finding is supported by other studies in neurodevelopmental disorders, such as childhood apraxia of speech (CAS), which has shown *FOXP2* disruptions as sufficient (monogenic) causes in some individuals and families (Fisher *et al.*, 1998), while others show a heterogenous, more complex aetiology (Worthey *et al.*, 2013; Eising *et al.*, 2019).

The genes with pathogenic variants identified here (*GRHL3, DIAPH1, FAT4* and *IFT88*) are all substantiated candidates for sAPD. Two genes are cleft lip/palate genes (*GRHL3 and IFT88*) with non-syndromic cleft lip/palate disorders shown to be linked to recurrent otitis media episodes and worse auditory processing skills, suggesting that the three conditions might have overlapping molecular pathways. Two genes (*FAT4* and *IFT88*) are involved in planar polarity, which is a crucial step of development especially in the inner ear, and thus might be related to auditory processing skills through a reduced ability to hear properly, which is a subtle deficit rather than an overt impairment. Three of the genes (*FAT4*, *GRHL3* and *DIAPH1*) show to be expressed in the brain, which is consistent with the view of APD being a higher order disorder. *NAV2* is a candidate hearing gene and has not yet been detected in human disease, so its role in auditory processing would require more evidence. Furthermore, the findings here demonstrate the potential validity of larger-scale genetic studies in some cases of APD which makes WES/WGS a valid approach that could identify novel genes with high rare penetrance variants that contribute to APD.

7. Overall Discussion

7.1. Summary of findings

Previous research in a large family, affected by a severe language disorder with auditory processing difficulties, showed a heterozygous stop-gain *USH2A* variant (with very strong evidence of pathogenicity) which co-segregated with the disorder (Perrino *et al.*, 2020). Given the role of *USH2A* in syndromic hearing loss (homozygous *USH2A* mutations causing USH2) (Eudy *et al.*, 1998), the location of its protein within stereocilia links in inner ear OHCs (Adato *et al.*, 2005a) and the role of OHCs in selective amplification, facilitating auditory perception (Froud *et al.*, 2015; Murakoshi, Suzuki and Wada, 2015), *USH2A* represented a clear candidate for studying its role in APD. Due to the overlapping and well-studied function of typical Usher syndrome genes in stereocilia development, the current thesis extended the early investigations by Perrino *et al*., hypothesising that pathogenic heterozygous variants in eleven USH causing genes (*MYO7A*, *CDH23*, *PCDH15*, *USH1C, USH1G, CIB2, USH2A*, *ADGRV1, WHRN, CLRN1* and *HARS*) have a subtle effect on developmental profiles in ALSPAC carrier individuals (H1). Hypothesis 2 proposed that common variants in Usher genes have an effect on hearing, auditory processing and/or language abilities as part of a complex genetic model (H2), while hypothesis 3 focussed on multiple rare variants across Usher gene regions and their complex effect on auditory processing, hearing and language (H3). Hypothesis 4 suggested that difficulty discriminating words in noisy environment (as a potential sign of sAPD) in a small number of ALSPAC children can be explained by rare coding pathogenic variants (H4).

Findings from Results Chapter 1 demonstrated the association of heterozygous *USH2A, MYO7A, CDH23* and *USH1C* pathogenic variants with subtle problems in hearing and delays in early language milestones, supporting H1. Moreover, no one individual neurodevelopmental measure could be taken as a "clinical marker of deficit" of Usher carriers. This suggested a more complex model of interaction and susceptibility, which was explored under H2 (Results Chapter 2) and H3 (Results Chapter 3), proposing that USH common and rare variants contribute to altered hearing and/or language abilities. Direct and indirect associations were observed between *USH2A*, *CLRN1* and *PCDH15* common SNPs and low-frequency hearing thresholds, early language markers and language outcomes (DLD) (Results Chapter 2) and between *USH2A*, *CLRN1* and *ADGRV1* rare variants and mid-frequency hearing thresholds or language markers (VIQ and NWR) (Results Chapter 3). These findings demonstrated the important role of common and rare risk variants within Usher genes in complex models of hearing, auditory processing and language. Moreover, they suggested that the combination of heterozygous USH pathogenic variants in a genomic

background of increased risk (through inheritance of other common and rare variants) may lead to increased susceptibility to APD.

Examining the contribution of rare variants on a genome-wide level, in Results Chapter 4, H4 proposed that pathogenic coding variants with large effect can explain poor performance on a surrogate measure of auditory processing difficulties in a small number of APD suspected children. H4 was supported by the identification of pathogenic variants in four of thirteen sAPD individuals highlighting *GRHL3*, *DIAPH1*, *FAT4* and *IFT88* as novel candidate genes for APD. Taken together, this research supports emerging ideas around genetic complexity and indicates a continuous model of complex APD genetic risk that includes multiple interacting factors. Moreover, it provides potential risk APD candidates that can shed light into the molecular pathways underlying difficulty listening in noise and advance our understanding of the pathophysiology of APD. This can further improve detection and diagnosis of APD, which will also lead to more appropriate APD- specific therapy for those affected.

7.2. Relationship between APD and language

The USH carriers showed delays in language milestones, as did the suspected APD (sAPD) group individuals, indicating that a correlate of APD is language delay (supported from both genetic driven and phenotype driven approach in Result Chapters 1 and 4). This adds to the literature which describes overlaps between language and auditory processing based on examining populations with APD/sAPD, speech and language disorders (SLD) or other related problems such as mild to moderate SNHL (Dawes and Bishop, 2009; Ferguson *et al.*, 2011; Sharma, Purdy and Kelly, 2009; Halliday, Tuomainen and Rosen, 2017; Bishop, Hardiman and Barry, 2012). The correlation data between APD and SLD which currently exists is poor at distinguishing causal models (summarised in Section 1.2.1) as it based on behavioural/electrophysiological tests, but not molecular data. This thesis is the first study to examine correlations on a genetic basis and confirm that auditory perception is a building block of communication and language development. The overlaps observed in this work may indicate overlapping gene effects between hearing, auditory processing and language or from auditory processing difficulties, affected directly by the gene variants, having an indirect negative effect on early language development, supporting the risk factor model. This work therefore demonstrates that the genetic architecture of APD is likely to be complex and comparable to the current model of developmental disorders of speech and language with which it overlaps phenotypically. This further suggests that just as DLD, APD is influenced by the combination of many genetic and environmental risk factors (Chen *et al.*, 2017; Gialluisi *et al.*, 2014; Newbury *et al.*, 2009; Villanueva *et al.*, 2015) and yet simplex familial cases of APD with private high penetrance variants are likely to exist, similar to childhood apraxia of speech (Fisher *et al.*, 1998). To further unravel the relationships between APD and speech and

language disorders on a molecular level, further work would need to be undertaken and is discussed in Section 7.6.

7.3. Relationship between APD and hearing

Carriers of pathogenic USH gene variants (largely represented by *USH2A* carriers) showed subtle hearing difficulties that did not lead to an overt hearing impairment, suggesting a subclinical hearing deficit. Under this model, a subclinical impaired function in the peripheral auditory system (affecting the ear) can result in long-term changes that may persist even after normal peripheral function is restored and therefore have a negative impact on auditory processing (Moore and Hunter, 2013). People with normal hearing as measured on the standard audiogram, including those with APD, may thus have very slight, yet functionally significant hearing loss that goes undetected. The hypothesis that auditory processing difficulties/poor listening skills in background noise result from a subclinical hearing deficit has been supported by previous literature (Badri, Siegel and Wright, 2011; Saxena, Allan and Allen, 2015; Hoben *et al.*, 2017). Two of those studies suggested that dysfunctional OHCs and abnormal acoustic reflex (both important auditory system feedback mechanisms) could impact upon speech perception in background noise in children with listening difficulties or suspected APD (Hoben *et al.*, 2017; Saxena, Allan and Allen, 2015). This fits well with the present results supporting the hypothesis that *USH2A* may represent a risk factor for APD: a complete loss of *Ush2a* in mouse models preferentially affects OHCs, leading to overt hearing loss (Liu *et al.*, 2007), while a heterozygous mutation in *USH2A* is hypothesised to preserve some of the OHC function, which may lead to subtle malfunction that could impact hearing in noise, but preserve hearing in quiet. Moreover, from the investigated eleven USH genes as candidates for APD, five (*USH2A, PCDH15, CDH23, ADGRV1* and *USH1C*) showed association to low-and/or-mid-frequency hearing abilities. Such results of subtle differences in hearing abilities among individuals with risk variants in USH genes were expected because of their role in syndromic and/or or non-syndromic deafness. Moreover, they represent an example of allelic hierarchy where multiple variants within the same gene display different types of hearing phenotypes (Lenassi *et al.*, 2015): recessive pathogenic *USH2A* variants are associated with high-frequency deafness and/or RP in USH2 while common risk variants within the same gene are associated with altered abilities at low-frequency hearing.

7.4. Genetic models underlying APD and genetic contributions

The different approaches applied in this thesis to investigate the genetic contributions to APD suggest that APD is a heterogeneous disorder following a complex genetic model, influenced by genetic variants with large effect (Result Chapter 4), and genetic variants with small effect, contributing to risk (Result Chapter 1- 3). This conclusion is in line with other neurodevelopmental

disorders with heterogeneous presentation, combining a complex interplay between genetic factors of different effect size (some large, some small) (Gaugler *et al.*, 2014; Griswold *et al.*, 2015; Niemi *et al.*, 2018; Satterstrom *et al.*, 2020).

Furthermore, this thesis offers examples of deafness genes likely to be involved in the susceptibility to APD phenotypes and potential cellular pathways, with the caveat that this work has only captured preselected genes known to underlie deafness.

From the eleven USH genes, investigated for association to hearing and language phenotypes (Result Chapters 1-3), the strongest contributor for APD susceptibility remains *USH2A* as first explored by Perrino *et al.* Across the three genetic models (rare Mendelian, common risk model and gene-based rare risk model), *USH2A* pathogenic heterozygous variants, common and rare risk variants showed consistent contribution to hearing and/or language abilities. In addition, Perrino *et al*. found that children carriers of pathogenic heterozygous *USH2A* variants had increased low-frequency hearing thresholds (+1.2 dB HL at 500 Hz), which was consistent with other reports of obligate USH carriers from the literature (van Aarem *et al.*, 1995; Wagenaar *et al.*, 1995). Moreover, association analyses including common risk variants illustrated that *USH2A* has a direct influence upon low-frequency hearing abilities and an indirect influence upon language that is, in part, modulated by hearing. In support of the link between subtle hearing abilities and language, an association between mild/moderate SNHL and poorer phonological processing (the use of sounds to process spoken and written language) has been recognised in previous reports (Wake *et al.*, 2006; Briscoe, Bishop and Norbury, 2001). These findings illustrate that even a very mild problem with hearing can lead to a reduction in a particular language skill that together with other risk factors can increase the susceptibility to impairment in other domains of language. The usherin protein functions as a lateral link providing support between stereocilia and *USH2A* is expressed transiently during development (Adato *et al.*, 2005a), suggesting a key role in establishing stereocilia organisation early on. Although *USH2A* is found in both IHC and OHC, its knock-out disproportionally affects OHCs, leading to overt hearing loss specifically at high frequencies (Liu *et al.*, 2007). The low-frequency hearing abilities being affected because of a more complex *USH2A* risk model therefore suggests that *USH2A* risk variants can increase the susceptibility to an APD phenotype through the subtle disruption of feedback mechanisms by OHCs.

Although *USH2A* is primarily expressed in cochlea and retina, but not in the brain, a recent study by Perrino *et al.* provides a molecular explanation to the effects of *USH2A* genetic variation in CNS (Perrino, Newbury and Fitch, 2021). Heterozygous *Ush2a* mice showed an increased right superior olivary complex (SOC) volume compared to a decreased SOC volume in homozygous knock-out *Ush2a* mice (Perrino, Newbury and Fitch, 2021). These findings suggested that altered cochlear development, as a result of usherin malfunction or dysfunction, impacts higher order auditory

processing at both functional and structural level, but with different consequences for heterozygous compared to homozygous subjects.

Findings presented in Results Chapter 4, investigating genetic variants with large effect that could contribute to a sAPD phenotype, offer further insight into potential novel molecular pathways. Two frameshift heterozygous variants, one in *GRHL3* (p.Asp451fs) and another one in *IFT88* (p.Ile816fs) were detected in two individuals with very similar phenotypes of recurrent middle ear infections in early life, early language subtle deficits, typical audiogram and no recognised clinical disorder. Both genes are implicated in cleft lip/palate pathology (Basha *et al.*, 2018; Tian *et al.*, 2017), which is known to be associated with higher incidence of OME (Sheahan *et al.*, 2003), which is also a well-recognised risk factor to secondary APD. *IFT88* is known to be expressed in the organ of Corti and plays a role in planar cell polarity, while the expression of *GRHL3* in inner ear is not known, however, *GRHL3* is expressed in the brain and is known to regulate cognitive function. How and if their pathophysiological pathways overlap with APD would need further investigations but is nevertheless a valuable finding. It should to be noted that no data was available for cleft/lip palate symptoms for these individuals, therefore these conclusions are only suggestive.

A splicing heterozygous pathogenic variant in *FAT4* (c.11899+1G>T) was found in one more sAPD individual presenting with some early expressive and receptive deficits and recurrent ear problems with very low PIQ, average VIQ, no hearing loss and no diagnosed clinical disorder. Homozygous mutations in *FAT4* are a known cause of Van Maldergem syndrome, characterised with intellectual disability, craniofacial and auditory malformations, leading to hearing loss, renal and skeletal malformations (Cappello *et al.*, 2013). However, cases with milder phenotypes and compound heterozygous mutations have been reported (van der Ven *et al.*, 2017), suggesting a phenotypic variability of the syndrome. Although not explored before for *FAT4*, the phenotypic variability could be linked to heterozygous pathogenic variants (for example the *FAT4* c.11899+1G>T variant) which could result in some preservation of the protein function and therefore cause milder symptoms including auditory deficits. This view is similar to the subtle hearing effects seen in *USH2A* heterozygous individuals compared to homozygous *USH2A* pathogenic variants which cause hearing loss in Usher syndrome.

7.5. Limitations

The current work has a number of important limitations that should be considered.

Firstly, due to the small samples size of USH carriers (N= 17) and sAPD cohort (N= 13), no firm conclusions can be drawn as one or two individuals with extreme phenotypes can skew the dataset considerably. A larger sample size from a targeted population is needed to increase the statistical power of the findings. This is also true for the ALSPAC genotype (N= 7,141) and ALSPAC UK10K (N= 1,681) core datasets used for association analyses- much bigger populations are

needed to detect subtle effects of risk variants, which will also improve on the statistical power and replicate any true associations. There are other large developmental cohorts which include genetic and phenotype data to potentially study longitudinal hearing/auditory processing skills in relation to genetic variation. Examples of such cohorts are: the Norwegian Mother, Father and Child Cohort Study (MoBa) (https://www.fhi.no/en/studies/moba/) which includes over 90,000 pregnant women, Born in Bradford cohort (https://borninbradford.nhs.uk/research/documents-data) with over 13,500 children and their parents available, and the Millennium Cohort Study (MCS) (https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/), which follows the lives of around 19,000 young people across the UK. Although these cohorts primarily rely upon questionnaire-based measures and their available hearing measures are not as extensive as the ones within ALSPAC, they may still be suitable to study alongside ALSPAC.

Secondly, the sAPD cohort was selected based on performance on two measures at 5 years of age. Although word discrimination threshold is a speech recognition test that was performed in quiet as well as in noisy conditions, offering selection of individuals who struggle to hear speech in noise (reflecting the most common complaint in individuals with APD), it is not a typical auditory diagnostic assessment included in the APD battery (Campbell *et al.,* 2019). Therefore, the sAPD cohort of 13 children includes individuals with listening difficulties in noise that might reflect a more global deficit (such as cognition) or secondary auditory deficit as a result of episodes of middle ear infections (such as OME) rather than a specific auditory deficit. Moreover, measures utilised to review the broad neurodevelopmental profile for sAPD individuals were available up to the age of 14, with no later milestones to evaluate neurodevelopment in adult life. This further supports the notion that selection of sAPD individuals based on the two measures at an early age (at 5 years) might have in some cases led to inclusion of individuals with subtle hearing/listening difficulties at young age (with secondary APD perhaps), who develop a recognisable hearing loss with an adult onset later in life. It is assuring, however, that all sAPD individuals selected here show some degree of language deficit/delay and have normal hearing sensitivity with no obvious disorders diagnosed. Moreover, some individuals showed poorer PIQ, which fits with a possible auditory processing deficit (either as a developmental APD or part of a subtle language disorder) described in other studies (de Wit *et al.*, 2016).

Thirdly, the analysis of coding regions applied on sAPD individuals only shortlisted the most obvious candidates, which included very rare (novel in gnomAD_NFE population) and strongly pathogenic variants. In this way variants that have a higher allele frequency but are still considered rare (for example MAF ≤ 0.01) and may be relevant, have been missed, which might explain why no candidate variants were detected in the nine remaining sAPD individuals. This would also lead to any other contributing variants to the phenotypes to also remain undetected. Moreover, none of the detected candidate variants were confirmed on Sanger sequencing to

check that they were true positives (because primary DNA samples for the ALSPAC study are not readily available).

Fourth, within the gene-based analyses, the frequency for selecting rare variants (MAF ≤ 0.01) was based on the ALSPAC UK10K cohort variant frequency, rather than a larger population (such as gnomAD) database frequency. The relatively small size of the UK10K cohort (N= 1,867 individuals) compared to gnomAD (N= 76,156 individuals) may have resulted in some rare variants having a falsely elevated frequency and being missed, leading to missed possible gene-based rare variants associations across the USH genes.

Lastly, the effect of CNVs on the ALSPAC tested phenotypes were not explored, further analysis of copy-number variants will be needed to account more completely for the complex genetic contributions to APD.

7.6. Future work and conclusions

7.6.1. *USH2A* effects on language follow-up

Perrino *et al.* showed that heterozygous disruptions of *USH2A* in mice and humans are associated with altered low-frequency hearing abilities (auditory input) and altered early expressive language abilities (Perrino *et al.*, 2020). In a follow-up study, Perrino *et al.* illustrated that altered cochlear development as a result of *Ush2a* heterozygous knock-out mutations can secondarily impact the development of brain regions (specifically SOC volume) and consequently affect auditory processing ability (Perrino, Newbury and Fitch, 2021). However, the exact relationship between altered input and language acquisition is still unclear. It is unclear whether worse language outcomes are a result of an impaired auditory input that affects central mechanisms and thus indirectly affects language (similar to the model of effect on auditory processing ability proposed by Perrino *et al.)* or whether background genetic language risk combines with altered auditory input and modifies the risk. Further large-scale characterisation of *USH2A* variation and behavioural outcomes in a homogenous cohort, such as a large cohort of USH2A carriers, will allow more accurate quantification of disrupted low-frequency abilities in carriers and more precise assessment of language and cognitive abilities across carriers. Moreover, as the mouse strains with *Ush2a* heterozygous disruption used by Perrino *et al.* have a homogenous background which lacks overt risk mutations (Perrino *et al.*, 2020), an additional double vs single risk mouse model would be needed for the investigation of the double-hit model. A good candidate for a double risk model with heterozygous *Ush2a* knock-out is *Cntnap2. CNTNAP2* is a well-characterised gene with robust association to various aspects of language, communication and neurodevelopment in humans (Alarcón *et al.*, 2008; Vernes *et al.*, 2008). Moreover, auditory processing has been proposed as a mediator of CNTNAP2 influence upon expressive language (Scott *et al.*, 2018; Truong *et al.*, 2015). Further functional studies and imaging of the stereocilia in

heterozygous *Ush2a* mice and double risk models with *Cntnap2* will allow the identification of the level at which the input is impaired (obvious structural stereocilia damage will indicate a crucial role).

In addition, gene expression analyses in developmental brain anatomy (for example olivary complex, cochlea and auditory cortex) at different time points in developing mouse models (single and double risk) will allow the investigation of the type of biological processes which underlie the auditory input changes. Such additional investigations of *USH2A* carriers and mouse will inform our understanding of auditory perception processes and the effects of genetic and neurodevelopmental mechanisms upon longer-term brain development and language acquisition.

7.6.2. Association studies follow-up

Future work would also aim to replicate findings from the association analyses (Results Chapter 2-3), implicating common *USH2A* and *PCDH15* variants and rare *ADGRV1* variants in low-frequency abilities and/or language abilities. Larger GWAS or meta-GWAS (incorporating multiple independent GWAS) on cohorts with a range of markers of auditory processing/hearing (for example suprathreshold features of physiological tests like ABR) and language will have better association resolution to capture subtle effects. Following replication, the most promising results would need to be studied at the functional level to characterise molecular consequences of the variants (see Section 7.6.3).

7.6.3. Candidate genes follow-up

In addition to the larger cohorts of relevant populations, the functional effects of the identified predicted pathogenic variants in *GRHL3*, *IFT88* and *FAT4* would need to be investigated in cell or animal models. Mouse mutants would be the first step to study the effect of knocked-out or knocked-down *Grhl3*, *Ift88* and *Fat4* function on more specific auditory and behavioural processes in mice (including auditory brainstem response and prepulse inhibition tasks), similar to the experiments on *Ush2a* in Perrino *et al*. If these genes are confirmed to play a role in auditory perception, then the exact mechanisms can be further studied with the addition of histological, microscopy imaging and gene expression in specific brain regions.

7.6.4. Conclusions

The aim of this thesis was to fully examine the effect of candidate gene variation on auditory processing, hearing, and emergent language skills to better understand the shared mechanisms that underpin these processes. Overall, this research supports the idea of genetic complexity and suggests an overlap between auditory processing, hearing and language on a genetic level. Moreover, it provides an insight into the genetic architecture of APD, proposing several risk genes

(*USH2A*, *GRHL3*, *IFT88* and *FAT4*) that would need to be further examined so we can gain better understanding into the molecular biology of APD. This will lead to a better classification of APD as a disorder, which will further improve the specificity of diagnostic criteria used and the therapeutic input that affected individuals require.

References

Abad-Morales, V.*, et al.* (2020) 'Identification of a novel homozygous ARSG mutation as the second cause of Usher syndrome type 4', *Am J Ophthalmol Case Rep*: © 2020 The Authors., pp. 100736.

Abadie, C.*, et al.* (2012) 'Audiological findings in 100 USH2 patients', *Clin Genet,* 82(5), pp. 433-8. doi: 10.1111/j.1399-0004.2011.01772.x.

Abbott, J. A.*, et al.* (2017) 'The Usher Syndrome Type IIIB Histidyl-tRNA Synthetase Mutation Confers Temperature Sensitivity', *Biochemistry,* 56(28), pp. 3619-3631. doi: 10.1021/acs.biochem.7b00114.

Abou Tayoun, A. N.*, et al.* (2018) 'Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion', *Hum Mutat,* 39(11), pp. 1517-1524. doi: 10.1002/humu.23626.

Abu-Ameerh, M.*, et al.* (2020) 'Extending the spectrum of CLRN1- and ABCA4-associated inherited retinal dystrophies caused by novel and recurrent variants using exome sequencing', *Mol Genet Genomic Med,* 8(3), pp. e1123. doi: 10.1002/mgg3.1123.

Adato, A.*, et al.* (1999) 'Possible interaction between USH1B and USH3 gene products as implied by apparent digenic deafness inheritance', *Am J Hum Genet: Vol. 1*, pp. 261-5.

Adato, A.*, et al.* (2005a) 'Usherin, the defective protein in Usher syndrome type IIA, is likely to be a component of interstereocilia ankle links in the inner ear sensory cells', *Hum Mol Genet,* 14(24), pp. 3921-32. doi: 10.1093/hmg/ddi416.

Adato, A.*, et al.* (2005b) 'Interactions in the network of Usher syndrome type 1 proteins', *Hum Mol Genet,* 14(3), pp. 347-56. doi: 10.1093/hmg/ddi031.

Adato, A.*, et al.* (2002) 'USH3A transcripts encode clarin-1, a four-transmembrane-domain protein with a possible role in sensory synapses', *Eur J Hum Genet,* 10(6), pp. 339-50. doi: 10.1038/sj.ejhg.5200831.

Addis, L.*, et al.* (2010) 'A locus for an auditory processing deficit and language impairment in an extended pedigree maps to 12p13.31-q14.3', *Genes Brain Behav,* 9(6), pp. 545-61. doi: 10.1111/j.1601-183X.2010.00583.x.

Adzhubei, I. A.*, et al.* (2010) 'A method and server for predicting damaging missense mutations', *Nat Methods: Vol. 4*. United States, pp. 248-9.

Ahmed, Z. M.*, et al.* (2006) 'The tip-link antigen, a protein associated with the transduction complex of sensory hair cells, is protocadherin-15', *J Neurosci,* 26(26), pp. 7022-34. doi: 10.1523/jneurosci.1163-06.2006.

Ahmed, Z. M.*, et al.* (2003) 'PCDH15 is expressed in the neurosensory epithelium of the eye and ear and mutant alleles are responsible for both USH1F and DFNB23', *Hum Mol Genet,* 12(24), pp. 3215-23. doi: 10.1093/hmg/ddg358.

Ahmed, Z. M.*, et al.* (2001) 'Mutations of the protocadherin gene PCDH15 cause Usher syndrome type 1F', *Am J Hum Genet,* 69(1), pp. 25-34. doi: 10.1086/321277.

Al-Maawali, A.*, et al.* (2016) 'Novel loss-of-function variants in DIAPH1 associated with syndromic microcephaly, blindness, and early onset seizures', *Am J Med Genet A,* 170a(2), pp. 435-440. doi: 10.1002/ajmg.a.37422.

Alagramam, K. N.*, et al.* (2011) 'Mutations in protocadherin 15 and cadherin 23 affect tip links and mechanotransduction in mammalian sensory hair cells', *PLoS One,* 6(4), pp. e19183. doi: 10.1371/journal.pone.0019183.

Alarcón, M.*, et al.* (2008) 'Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene', *Am J Hum Genet,* 82(1), pp. 150-9. doi: 10.1016/j.ajhg.2007.09.005.

Alders, M.*, et al.* (2014) 'Hennekam syndrome can be caused by FAT4 mutations and be allelic to Van Maldergem syndrome', *Hum Genet,* 133(9), pp. 1161-7. doi: 10.1007/s00439-014-1456-y.

American Academy of Audiology (2010) 'Clinical practice guidelines. Guidelines for the diagnosis, treatment and management of children and adults with central auditory processing

disorder' 1-51. Available online at: www.audiology.org/resources/documentlibrary/Pages/CentralAuditoryProcessingDisorder.aspx

Andeol, G.*, et al.* (2011) 'Auditory Efferents Facilitate Sound Localization in Noise in Humans', *Journal of Neuroscience,* 31(18), pp. 6759-6763. doi: 10.1523/jneurosci.0248-11.2011.

Anderson, H. and Wedenberg, E. (1968) 'Audiometric identification of normal hearing carriers of genes for deafness', *Acta Otolaryngol,* 65(6), pp. 535-54. doi: 10.3109/00016486809119288.

Anney, R.*, et al.* (2012) 'Individual common variants exert weak effects on the risk for autism spectrum disorders', *Hum Mol Genet,* 21(21), pp. 4781-92. doi: 10.1093/hmg/dds301.

Antonellis, A. and Green, E. D. (2008) 'The role of aminoacyl-tRNA synthetases in genetic diseases', *Annu Rev Genomics Hum Genet,* 9, pp. 87-107. doi: 10.1146/annurev.genom.9.081307.164204.

Aparisi, M. J.*, et al.* (2014) 'Targeted next generation sequencing for molecular diagnosis of Usher syndrome', *Orphanet J Rare Dis,* 9, pp. 168. doi: 10.1186/s13023-014-0168-7.

Arcos-Burgos, M.*, et al.* (2010) 'A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication', *Mol Psychiatry,* 15(11), pp. 1053-66. doi: 10.1038/mp.2010.6.

Arking, D. E.*, et al.* (2008) 'A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism', *American Journal of Human Genetics,* 82(1), pp. 160-164. doi: 10.1016/j.ajhg.2007.09.015.

ASHA (American Speech-Language-Hearing Association) (2005) 'Central auditory processing: Working Group on Auditory Processing Disorders'.

Astuto, L. M.*, et al.* (2002) 'CDH23 mutation and phenotype heterogeneity: a profile of 107 diverse families with Usher syndrome and nonsyndromic deafness', *Am J Hum Genet,* 71(2), pp. 262-75. doi: 10.1086/341558.

Badri, R., Siegel, J. H. and Wright, B. A. (2011) 'Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms', *J Acoust Soc Am,* 129(2), pp. 852-63. doi: 10.1121/1.3523476.

Baek, J. I.*, et al.* (2012) 'Targeted massive parallel sequencing: the effective detection of novel causative mutations associated with hearing loss in small families', *Orphanet J Rare Dis,* 7, pp. 60. doi: 10.1186/1750-1172-7-60.

Bamiou, D. E., Musiek, F. E. and Luxon, L. M. (2001) 'Aetiology and clinical presentations of auditory processing disorders - a review', *Archives of Disease in Childhood,* 85(5), pp. 361-365. doi: 10.1136/adc.85.5.361.

Banik, A.*, et al.* (2017) 'Maternal Factors that Induce Epigenetic Changes Contribute to Neurological Disorders in Offspring', *Genes (Basel),* 8(6). doi: 10.3390/genes8060150.

Barrett, J. C.*, et al.* (2005) 'Haploview: analysis and visualization of LD and haplotype maps', *Bioinformatics,* 21(2), pp. 263-5. doi: 10.1093/bioinformatics/bth457.

Basha, M.*, et al.* (2018) 'Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes', *J Med Genet,* 55(7), pp. 449-458. doi: 10.1136/jmedgenet-2017-105110.

Bashir, R., Fatima, A. and Naz, S. (2010) 'A frameshift mutation in SANS results in atypical Usher syndrome', *Clin Genet,* 78(6), pp. 601-3. doi: 10.1111/j.1399-0004.2010.01500.x.

Baux, D.*, et al.* (2014) 'Enrichment of LOVD-USHbases with 152 USH2A genotypes defines an extensive mutational spectrum and highlights missense hotspots', *Hum Mutat,* 35(10), pp. 1179-86. doi: 10.1002/humu.22608.

Belyantseva, I. A.*, et al.* (2005) 'Myosin-XVa is required for tip localization of whirlin and differential elongation of hair-cell stereocilia', *Nat Cell Biol,* 7(2), pp. 148-56. doi: 10.1038/ncb1219.

Ben-Yosef, T.*, et al.* (2003) 'A mutation of PCDH15 among Ashkenazi Jews with the type 1 Usher syndrome', *N Engl J Med,* 348(17), pp. 1664-70. doi: 10.1056/NEJMoa021502.

Berlin, C. I. (1976) 'New developments in evaluating central auditory mechanisms', *Ann Otol Rhinol Laryngol,* 85(6 pt. 1), pp. 833-41. doi: 10.1177/000348947608500613.

Bernal, S.*, et al.* (2003) 'Mutations in USH2A in Spanish patients with autosomal recessive retinitis pigmentosa: high prevalence and phenotypic variation', *J Med Genet,* 40(1), pp. e8. doi: 10.1136/jmg.40.1.e8.

Besnard, T.*, et al.* (2012) 'Non-USH2A mutations in USH2 patients', *Hum Mutat,* 33(3), pp. 504-10. doi: 10.1002/humu.22004.

Beurg, M.*, et al.* (2009) 'Localization of inner hair cell mechanotransducer channels using high-speed calcium imaging', *Nat Neurosci,* 12(5), pp. 553-8. doi: 10.1038/nn.2295.

Bhutta, A. T.*, et al.* (2002) 'Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis', *Jama,* 288(6), pp. 728-37. doi: 10.1001/jama.288.6.728.

Bishop, D. V. (2006) 'What Causes Specific Language Impairment in Children?', *Curr Dir Psychol Sci,* 15(5), pp. 217-221. doi: 10.1111/j.1467-8721.2006.00439.x.

Bishop, D. V., Adams, C. V. and Norbury, C. F. (2006) 'Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins', *Genes Brain Behav,* 5(2), pp. 158-69. doi: 10.1111/j.1601-183X.2005.00148.x.

Bishop, D. V., Hardiman, M. J. and Barry, J. G. (2012) 'Auditory deficit as a consequence rather than endophenotype of specific language impairment: electrophysiological evidence', *PLoS One,* 7(5), pp. e35851. doi: 10.1371/journal.pone.0035851.

Bishop, D. V. and Hayiou-Thomas, M. E. (2008) 'Heritability of specific language impairment depends on diagnostic criteria', *Genes Brain Behav,* 7(3), pp. 365-72. doi: 10.1111/j.1601-183X.2007.00360.x.

Blanco-Kelly, F.*, et al.* (2015) 'Clinical aspects of Usher syndrome and the USH2A gene in a cohort of 433 patients', *JAMA Ophthalmol,* 133(2), pp. 157-64. doi: 10.1001/jamaophthalmol.2014.4498.

Blanco-Sánchez, B.*, et al.* (2014) 'Complexes of Usher proteins preassemble at the endoplasmic reticulum and are required for trafficking and ER homeostasis', *Dis Model Mech,* 7(5), pp. 547-59. doi: 10.1242/dmm.014068.

Boeda, B.*, et al.* (2002) 'Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle', *Embo j,* 21(24), pp. 6689-99. doi: 10.1093/emboj/cdf689.

Bolz, H.*, et al.* (2001) 'Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D', *Nat Genet,* 27(1), pp. 108-12. doi: 10.1038/83667.

Bonnet, C.*, et al.* (2011) 'Complete exon sequencing of all known Usher syndrome genes greatly improves molecular diagnosis', *Orphanet J Rare Dis,* 6, pp. 21. doi: 10.1186/1750-1172-6-21.

Booth, K. T.*, et al.* (2018) 'Variants in CIB2 cause DFNB48 and not USH1J', *Clinical Genetics,* 93(4), pp. 812-821. doi: 10.1111/cge.13170.

Boothalingam, S.*, et al.* (2015) 'Cochlear Delay and Medial Olivocochlear Functioning in Children with Suspected Auditory Processing Disorder', *PLoS One,* 10(8), pp. e0136906. doi: 10.1371/journal.pone.0136906.

Bork, J. M.*, et al.* (2001) 'Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel cadherin-like gene CDH23', *Am J Hum Genet,* 68(1), pp. 26-37. doi: 10.1086/316954.

Boughman, J. A., Vernon, M. and Shaver, K. A. (1983) 'Usher syndrome: definition and estimate of prevalence from two high-risk populations', *J Chronic Dis,* 36(8), pp. 595-603. doi: 10.1016/0021-9681(83)90147-9.

Boyd, A.*, et al.* (2013) 'Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children', *Int J Epidemiol,* 42(1), pp. 111-27. doi: 10.1093/ije/dys064.

Bray, S. J. and Kafatos, F. C. (1991) 'Developmental function of Elf-1: an essential transcription factor during embryogenesis in Drosophila', *Genes Dev,* 5(9), pp. 1672-83. doi: 10.1101/gad.5.9.1672.

Brewer, C. C.*, et al.* (2016) 'Heritability of non-speech auditory processing skills', *European Journal of Human Genetics,* 24(8), pp. 1137-1144. doi: 10.1038/ejhg.2015.277.

Briscoe, J., Bishop, D. V. M. and Norbury, C. F. (2001) 'Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment', *Journal of Child Psychology and Psychiatry and Allied Disciplines,* 42(3), pp. 329-340. doi: 10.1111/1469-7610.00726.

BSA APD (2011) 'Position Statement: Auditory Processing Disorder (APD)' Available online at: www.thebsa.org.uk/images/stories/docs/BSA_APD_PositionPaper_31March11_FINAL.pdf

BSA (2018) 'Position Statement and Practice Guidance- Auditory Processing Disorder (APD)'. Available online at: http://www.thebsa.org.uk/wp-content/uploads/2018/02/Position-Statement-and-Practice-Guidance-APD-2018.pdf

Bujakowska, K. M.*, et al.* (2014) 'Targeted exon sequencing in Usher syndrome type I', *Invest Ophthalmol Vis Sci,* 55(12), pp. 8488-96. doi: 10.1167/iovs.14-15169.

Caberlotto, E.*, et al.* (2011) 'Usher type 1G protein sans is a critical component of the tip-link complex, a structure controlling actin polymerization in stereocilia', *Proc Natl Acad Sci U S A,* 108(14), pp. 5825-30. doi: 10.1073/pnas.1017114108.

Cacace, A. T. and McFarland, D. J. (2005) 'The importance of modality specificity in diagnosing central auditory processing disorder', *Am J Audiol,* 14(2), pp. 112-23. doi: 10.1044/1059-0889(2005/012).

Cai, T.*, et al.* (2015) 'Characterization of the transcriptome of nascent hair cells and identification of direct targets of the Atoh1 transcription factor', *J Neurosci,* 35(14), pp. 5870-83. doi: 10.1523/jneurosci.5083-14.2015.

Campbell N., Grant P., Moore D.R., Rosen S. (2019) 'Auditory Processing Disorder (APD) in Children (Guide) '. Available at: http://www.meshguides.org/guides/node/1432.

Cappello, S.*, et al.* (2013) 'Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development', *Nat Genet,* 45(11), pp. 1300-8. doi: 10.1038/ng.2765.

Chang, X. and Wang, K. (2012) 'wANNOVAR: annotating genetic variants for personal genomes via the web', *J Med Genet,* 49(7), pp. 433-6. doi: 10.1136/jmedgenet-2012-100918.

Chekuri, A.*, et al.* (2018) 'IFT88 mutations identified in individuals with non-syndromic recessive retinal degeneration result in abnormal ciliogenesis', *Hum Genet,* 137(6-7), pp. 447-458. doi: 10.1007/s00439-018-1897-9.

Chen, G. D. (2018) 'Hidden cochlear impairments', *J Otol,* 13(2), pp. 37-43. doi: 10.1016/j.joto.2018.05.001.

Chen, Q.*, et al.* (2014) 'Whirlin and PDZ domain-containing 7 (PDZD7) proteins are both required to form the quaternary protein complex associated with Usher syndrome type 2', *J Biol Chem,* 289(52), pp. 36070-88. doi: 10.1074/jbc.M114.610535.

Chen, X. S.*, et al.* (2017) 'Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment', *Scientific Reports,* 7. doi: 10.1038/srep46105.

Chermak, G. D. (2002) 'Deciphering auditory processing disorders in children', *Otolaryngol Clin North Am,* 35(4), pp. 733-49. doi: 10.1016/s0030-6665(02)00056-7.

Cingolani, P.*, et al.* (2012) 'A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3', *Fly (Austin),* 6(2), pp. 80-92. doi: 10.4161/fly.19695.

Cirulli, E. T. and Goldstein, D. B. (2010) 'Uncovering the roles of rare variants in common disease through whole-genome sequencing', *Nat Rev Genet,* 11(6), pp. 415-25. doi: 10.1038/nrg2779.

Cohen J. (1998) 'Statistical power analysis for the behavioral sciences' 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates.

Coll, M.*, et al.* (2017) 'Targeted next-generation sequencing provides novel clues for associated epilepsy and cardiac conduction disorder/SUDEP', *PLoS One,* 12(12), pp. e0189618. doi: 10.1371/journal.pone.0189618.

Collin, R. W.*, et al.* (2010) 'Mutations in C2ORF71 cause autosomal-recessive retinitis pigmentosa', *Am J Hum Genet,* 86(5), pp. 783-8. doi: 10.1016/j.ajhg.2010.03.016.

Cosgrove, D. and Zallocchi, M. (2014) 'Usher protein functions in hair cells and photoreceptors', *International Journal of Biochemistry & Cell Biology,* 46, pp. 80-89. doi: 10.1016/j.biocel.2013.11.001.

Costalupes, J. A., Young, E. D. and Gibson, D. J. (1984) 'Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat', *J Neurophysiol,* 51(6), pp. 1326-44. doi: 10.1152/jn.1984.51.6.1326.

D, C. K.*, et al.* (2020) 'Hidden Hearing Loss: A Disorder with Multiple Etiologies and Mechanisms', *Cold Spring Harb Perspect Med,* 10(1). doi: 10.1101/cshperspect.a035493.

Dad, S.*, et al.* (2016) 'Usher syndrome in Denmark: mutation spectrum and some clinical observations', *Mol Genet Genomic Med,* 4(5), pp. 527-539. doi: 10.1002/mgg3.228.

Dallos, P. and Harris, D. (1978) 'Properties of auditory nerve responses in absence of outer hair cells', *J Neurophysiol,* 41(2), pp. 365-83. doi: 10.1152/jn.1978.41.2.365.

Danecek, P.*, et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics,* 27(15), pp. 2156-8. doi: 10.1093/bioinformatics/btr330.

Davydov, E. V.*, et al.* (2010) 'Identifying a high fraction of the human genome to be under selective constraint using GERP++', *PLoS Comput Biol,* 6(12), pp. e1001025. doi: 10.1371/journal.pcbi.1001025.

Dawes, P. and Bishop, D. (2009) 'Auditory processing disorder in relation to developmental disorders of language, communication and attention: a review and critique', *Int J Lang Commun Disord,* 44(4), pp. 440-65. doi: 10.1080/13682820902929073.

Dawes, P. and Bishop, D. V. (2010) 'Psychometric profile of children with auditory processing disorder and children with dyslexia', *Arch Dis Child,* 95(6), pp. 432-6. doi: 10.1136/adc.2009.170118.

Dawes, P.*, et al.* (2008) 'Profile and aetiology of children diagnosed with auditory processing disorder (APD)', *International Journal of Pediatric Otorhinolaryngology,* 72(4), pp. 483-489. doi: 10.1016/j.ijporl.2007.12.007.

Dawn Teare, M. and Barrett, J. H. (2005) 'Genetic linkage studies', *Lancet,* 366(9490), pp. 1036-44. doi: 10.1016/s0140-6736(05)67382-5.

de Wit, E.*, et al.* (2016) 'Characteristics of Auditory Processing Disorders: A Systematic Review', *J Speech Lang Hear Res,* 59(2), pp. 384-413. doi: 10.1044/2015_jslhr-h-15-0118.

Delprat, B.*, et al.* (2005) 'Myosin XVa and whirlin, two deafness gene products required for hair bundle growth, are located at the stereocilia tips and interact directly', *Hum Mol Genet,* 14(3), pp. 401-10. doi: 10.1093/hmg/ddi036.

den Hoed, J.*, et al.* (2020) 'Mutation-specific pathophysiological mechanisms define different neurodevelopmental disorders associated with SATB1 dysfunction', *bioRxiv*, pp. 2020.10.23.352278. doi: 10.1101/2020.10.23.352278.

DeRosse, P.*, et al.* (2008) 'The genetics of symptom-based phenotypes: toward a molecular classification of schizophrenia', *Schizophr Bull,* 34(6), pp. 1047-53. doi: 10.1093/schbul/sbn076.

Di Palma, F.*, et al.* (2001) 'Mutations in Cdh23, encoding a new type of cadherin, cause stereocilia disorganization in waltzer, the mouse model for Usher syndrome type 1D', *Nat Genet,* 27(1), pp. 103-7. doi: 10.1038/83660.

Dillon, H.*, et al.* (2012) 'An opinion on the assessment of people who may have an auditory processing disorder', *J Am Acad Audiol,* 23(2), pp. 97-105. doi: 10.3766/jaaa.23.2.4.

Doucette, L.*, et al.* (2009) 'Profound, prelingual nonsyndromic deafness maps to chromosome 10q21 and is caused by a novel missense mutation in the Usher syndrome type IF gene PCDH15', *Eur J Hum Genet,* 17(5), pp. 554-64. doi: 10.1038/ejhg.2008.231.

Dreyer, B.*, et al.* (2008) 'Spectrum of USH2A mutations in Scandinavian patients with Usher syndrome type II', *Hum Mutat,* 29(3), pp. 451. doi: 10.1002/humu.9524.

Dworkin, S.*, et al.* (2017) 'Grainyhead-like 3 (Grhl3) deficiency in brain leads to altered locomotor activity and decreased anxiety-like behaviors in aged mice', *Dev Neurobiol,* 77(6), pp. 775-788. doi: 10.1002/dneu.22469.

Eatock, R. A. (2000) 'Adaptation in hair cells', *Annu Rev Neurosci,* 23, pp. 285-314. doi: 10.1146/annurev.neuro.23.1.285.

Ebermann, I.*, et al.* (2010) 'PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome', *Journal of Clinical Investigation,* 120(6), pp. 1812-1823. doi: 10.1172/jci39715.

Ebermann, I.*, et al.* (2007) 'A novel gene for Usher syndrome type 2: mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss', *Hum Genet,* 121(2), pp. 203-11. doi: 10.1007/s00439-006-0304-0.

Ebrahim, S. *et al.* (2016) 'Alternative Splice Forms Influence Functions of Whirlin in Mechanosensory Hair Cell Stereocilia', *Cell Rep,* 15(5), pp. 935-943. doi: 10.1016/j.celrep.2016.03.081.

Eisenberger, T.*, et al.* (2012) 'Targeted next-generation sequencing identifies a homozygous nonsense mutation in ABHD12, the gene underlying PHARC, in a family clinically diagnosed with Usher syndrome type 3', *Orphanet J Rare Dis,* 7, pp. 59. doi: 10.1186/1750-1172-7-59.

Eising, E.*, et al.* (2019) 'A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development', *Molecular Psychiatry,* 24(7), pp. 1065-1078. doi: 10.1038/s41380-018-0020-x.

El-Amraoui, A. and Petit, C. (2005) 'Usher I syndrome: unravelling the mechanisms that underlie the cohesion of the growing hair bundle in inner ear sensory cells', *J Cell Sci,* 118(Pt 20), pp. 4593-603. doi: 10.1242/jcs.02636.

El-Amraoui, A. and Petit, C. (2014) 'The retinal phenotype of Usher syndrome: pathophysiological insights from animal models', *C R Biol,* 337(3), pp. 167-77. doi: 10.1016/j.crvi.2013.12.004.

Ercan-Sencicek, A. G.*, et al.* (2015) 'Homozygous loss of DIAPH1 is a novel cause of microcephaly in humans', *Eur J Hum Genet,* 23(2), pp. 165-72. doi: 10.1038/ejhg.2014.82.

Eudy, J. D.*, et al.* (1998) 'Mutation of a gene encoding a protein with extracellular matrix motifs in usher syndrome type IIa', *Science,* 280(5370), pp. 1753-1757. doi: 10.1126/science.280.5370.1753.

Faraone, S. V. and Larsson, H. (2019) 'Genetics of attention deficit hyperactivity disorder', *Mol Psychiatry,* 24(4), pp. 562-575. doi: 10.1038/s41380-018-0070-0.

Felix, R. A., 2nd*, et al.* (2019) 'Nicotinic acetylcholine receptor subunit α(7)-knockout mice exhibit degraded auditory temporal processing', *J Neurophysiol,* 122(2), pp. 451-465. doi: 10.1152/jn.00170.2019.

Ferguson, M. A.*, et al.* (2011) 'Communication, Listening, Cognitive and Speech Perception Skills in Children With Auditory Processing Disorder (APD) or Specific Language Impairment (SLI)', *Journal of Speech Language and Hearing Research,* 54(1), pp. 211-227. doi: 10.1044/1092-4388(2010/09-0167).

Ferrante, P.*, et al.* (2002) 'Molecular and biochemical characterisation of a novel sulphatase gene: Arylsulfatase G (ARSG)', *Eur J Hum Genet,* 10(12), pp. 813-8. doi: 10.1038/sj.ejhg.5200887.

Fields, R. R.*, et al.* (2002) 'Usher syndrome type III: revised genomic structure of the USH3 gene and identification of novel mutations', *Am J Hum Genet,* 71(3), pp. 607-17. doi: 10.1086/342098.

Fisher, S. E.*, et al.* (1998) 'Localisation of a gene implicated in a severe speech and language disorder', *Nat Genet,* 18(2), pp. 168-70. doi: 10.1038/ng0298-168.

Friedman, R. A. *et al.* (2009) 'GRM7 variants confer susceptibility to age-related hearing impairment', *Hum Mol Genet,* 18(4), pp. 785-96. doi: 10.1093/hmg/ddn402.

Fritsche, L. G.*, et al.* (2016) 'A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants', *Nat Genet,* 48(2), pp. 134-43. doi: 10.1038/ng.3448.

Frolenkov, G. I.*, et al.* (2004) 'Genetic insights into the morphogenesis of inner ear hair cells', *Nat Rev Genet,* 5(7), pp. 489-98. doi: 10.1038/nrg1377.

Froud, K. E.*, et al.* (2015) 'Type II spiral ganglion afferent neurons drive medial olivocochlear reflex suppression of the cochlear amplifier', *Nat Commun,* 6, pp. 7115. doi: 10.1038/ncomms8115.

Furman, A. C., Kujawa, S. G. and Liberman, M. C. (2013) 'Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates', *J Neurophysiol,* 110(3), pp. 577-86. doi: 10.1152/jn.00164.2013.

Fuster-Garcia, C*., et al.* (2018) 'High-throughput sequencing for the molecular diagnosis of Usher syndrome reveals 42 novel mutations and consolidates CEP250 as Usher-like disease causative', *Sci Rep,* 8(1), pp. 17113. doi: 10.1038/s41598-018-35085-0.

Ganna, A*., et al.* (2018) 'Quantifying the Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum', *Am J Hum Genet,* 102(6), pp. 1204-1211. doi: 10.1016/j.ajhg.2018.05.002.

Garcia-Garcia, G*., et al.* (2011) 'Mutational screening of the USH2A gene in Spanish USH patients reveals 23 novel pathogenic mutations', *Orphanet J Rare Dis,* 6, pp. 65. doi: 10.1186/1750-1172-6-65.

Gates, G. A., Couropmitree, N. N. and Myers, R. H. (1999) 'Genetic associations in age-related hearing thresholds', *Arch Otolaryngol Head Neck Surg,* 125(6), pp. 654-9. doi: 10.1001/archotol.125.6.654.

Gathercole, S. E*., et al.* (1994) 'The Children's Test of Nonword Repetition: a test of phonological working memory', *Memory,* 2(2), pp. 103-27. doi: 10.1080/09658219408258940.

Gaugler, T*., et al.* (2014) 'Most genetic risk for autism resides with common variation', *Nat Genet,* 46(8), pp. 881-5. doi: 10.1038/ng.3039.

Géléoc, G. G. S. and El-Amraoui, A. (2020) 'Disease mechanisms and gene therapy for Usher syndrome', *Hear Res,* 394, pp. 107932. doi: 10.1016/j.heares.2020.107932.

Geng, R*., et al.* (2009) 'Usher syndrome IIIA gene clarin-1 is essential for hair cell function and associated neural activation', *Hum Mol Genet,* 18(15), pp. 2748-60. doi: 10.1093/hmg/ddp210.

Geng, R*., et al.* (2012) 'The mechanosensory structure of the hair cell requires clarin-1, a protein encoded by Usher syndrome III causative gene', *J Neurosci,* 32(28), pp. 9485-98. doi: 10.1523/jneurosci.0311-12.2012.

Ghimire, S. R., Ratzan, E. M. and Deans, M. R. (2018) 'A non-autonomous function of the core PCP protein VANGL2 directs peripheral axon turning in the developing cochlea', *Development,* 145(12). doi: 10.1242/dev.159012.

Gialluisi, A*., et al.* (2020) 'Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia', *Mol Psychiatry*. doi: 10.1038/s41380-020-00898-x.

Gialluisi, A*., et al.* (2014) 'Genome-wide screening for DNA variants associated with reading and language traits', *Genes Brain Behav,* 13(7), pp. 686-701. doi: 10.1111/gbb.12158.

Gialluisi, A*., et al.* (2016) 'Investigating the effects of copy number variants on reading and language performance', *J Neurodev Disord,* 8, pp. 17. doi: 10.1186/s11689-016-9147-8.

Giese, A. P. J*., et al.* (2017a) 'CIB2 interacts with TMC1 and TMC2 and is essential for mechanotransduction in auditory hair cells', *Nat Commun,* 8(1), pp. 43. doi: 10.1038/s41467-017-00061-1.

Giese, A. P. J*., et al.* (2017b) 'CIB2 interacts with TMC1 and TMC2 and is essential for mechanotransduction in auditory hair cells', *Nature Communications,* 8. doi: 10.1038/s41467-017-00061-1.

Gillespie, P. G. and Müller, U. (2009) 'Mechanotransduction by hair cells: models, molecules, and mechanisms', *Cell,* 139(1), pp. 33-44. doi: 10.1016/j.cell.2009.09.010.

Girirajan, S*., et al.* (2012) 'Phenotypic heterogeneity of genomic disorders and rare copy-number variants', *N Engl J Med,* 367(14), pp. 1321-31. doi: 10.1056/NEJMoa1200395.

Girotto, G*., et al.* (2019) 'Next Generation Sequencing and Animal Models Reveal SLC9A3R1 as a New Gene Involved in Human Age-Related Hearing Loss', *Front Genet,* 10, pp. 142. doi: 10.3389/fgene.2019.00142.

Girotto, G*., et al.* (2014) 'Expression and replication studies to identify new candidate genes involved in normal hearing function', *PLoS One,* 9(1), pp. e85352. doi: 10.1371/journal.pone.0085352.

Golding, J., Pembrey, M. and Jones, R. (2001) 'ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology', *Paediatr Perinat Epidemiol,* 15(1), pp. 74-87.
Goodyear, R. J.*, et al.* (2010) 'Asymmetric distribution of cadherin 23 and protocadherin 15 in the kinocilial links of avian sensory hair cells', *J Comp Neurol,* 518(21), pp. 4288-97. doi: 10.1002/cne.22456.
Goodyear, R. J.*, et al.* (2005) 'Development and properties of stereociliary link types in hair cells of the mouse cochlea', *J Comp Neurol,* 485(1), pp. 75-85. doi: 10.1002/cne.20513.
Gopal, S. R.*, et al.* (2015) 'Zebrafish Models for the Mechanosensory Hair Cell Dysfunction in Usher Syndrome 3 Reveal That Clarin-1 Is an Essential Hair Bundle Protein', *J Neurosci,* 35(28), pp. 10188-201. doi: 10.1523/jneurosci.1096-15.2015.
Griswold, A. J.*, et al.* (2015) 'Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants', *Mol Autism,* 6, pp. 43. doi: 10.1186/s13229-015-0034-z.
Grove, J.*, et al.* (2019) 'Identification of common genetic risk variants for autism spectrum disorder', *Nat Genet,* 51(3), pp. 431-444. doi: 10.1038/s41588-019-0344-8.
Guidi, L. G.*, et al.* (2017) 'Knockout Mice for Dyslexia Susceptibility Gene Homologs KIAA0319 and KIAA0319L have Unaffected Neuronal Migration but Display Abnormal Auditory Processing', *Cereb Cortex,* 27(12), pp. 5831-5845. doi: 10.1093/cercor/bhx269.
Halliday, L. F., Tuomainen, O. and Rosen, S. (2017) 'Auditory processing deficits are sometimes necessary and sometimes sufficient for language difficulties in children: Evidence from mild to moderate sensorineural hearing loss', *Cognition,* 166, pp. 139-151. doi: 10.1016/j.cognition.2017.04.014.
Han, J. Y.*, et al.* (2020) 'Identification of Missense ADGRV1 Mutation as a Candidate Genetic Cause of Familial Febrile Seizure 4', *Children (Basel): Vol. 9*.
Hartel, B. P.*, et al.* (2017) 'Cochlear Implantation in Patients With Usher Syndrome Type IIa Increases Performance and Quality of Life', *Otol Neurotol,* 38(6), pp. e120-e127. doi: 10.1097/mao.0000000000001441.
Hasson, T.*, et al.* (1995) 'Expression in cochlea and retina of myosin VIIa, the gene product defective in Usher syndrome type 1B', *Proc Natl Acad Sci U S A,* 92(21), pp. 9815-9. doi: 10.1073/pnas.92.21.9815.
Haywood-Watson, R. J., 2nd*, et al.* (2006) 'Ames Waltzer deaf mice have reduced electroretinogram amplitudes and complex alternative splicing of Pcdh15 transcripts', *Invest Ophthalmol Vis Sci,* 47(7), pp. 3074-84. doi: 10.1167/iovs.06-0108.
Hemati, P.*, et al.* (2018) 'Refining the phenotype associated with GNB1 mutations: Clinical data on 18 newly identified patients and review of the literature', *American Journal of Medical Genetics Part A,* 176(11), pp. 2259-2275. doi: 10.1002/ajmg.a.40472.
Henriksen, M. G., Nordgaard, J. and Jansson, L. B. (2017) 'Genetics of Schizophrenia: Overview of Methods, Findings and Limitations', *Front Hum Neurosci,* 11, pp. 322. doi: 10.3389/fnhum.2017.00322.
Hertzano R. and Orvis J. 'gEAR Portal'. Available at: https://umgear.org/ (accessed June 2020)
Higgs, H. N. and Peterson, K. J. (2005) 'Phylogenetic analysis of the formin homology 2 domain', *Mol Biol Cell,* 16(1), pp. 1-13. doi: 10.1091/mbc.e04-07-0565.
Hikosaka, O. (2010) 'The habenula: from stress evasion to value-based decision-making', *Nat Rev Neurosci,* 11(7), pp. 503-13. doi: 10.1038/nrn2866.
Hilgert, N.*, et al.* (2009) 'A large deletion in GPR98 causes type IIC Usher syndrome in male and female members of an Iranian family', *J Med Genet,* 46(4), pp. 272-6. doi: 10.1136/jmg.2008.060947.
Hind, S. E.*, et al.* (2011) 'Prevalence of clinical referrals having hearing thresholds within normal limits', *Int J Audiol,* 50(10), pp. 708-16. doi: 10.3109/14992027.2011.582049.
Hoben, R.*, et al.* (2017) 'Outer Hair Cell and Auditory Nerve Function in Speech Recognition in Quiet and in Background Noise', *Front Neurosci,* 11, pp. 157. doi: 10.3389/fnins.2017.00157.
Hoff, E.*, et al.* (2012) 'Dual language exposure and early bilingual development', *J Child Lang,* 39(1), pp. 1-27. doi: 10.1017/s0305000910000759.

Hoffmann, T. J. *et al.* (2016) 'A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records', *PLoS Genet,* 12(10), pp. e1006371. doi: 10.1371/journal.pgen.1006371.

Holme, R. H. and Steel, K. P. (2002) 'Stereocilia defects in waltzer (Cdh23), shaker1 (Myo7a) and double waltzer/shaker1 mutant mice', *Hear Res,* 169(1-2), pp. 13-23. doi: 10.1016/s0378-5955(02)00334-9.

Holme, R. H. and Steel, K. P. (2004) 'Progressive hearing loss and increased susceptibility to noise-induced hearing loss in mice carrying a Cdh23 but not a Myo7a mutation', *J Assoc Res Otolaryngol,* 5(1), pp. 66-79. doi: 10.1007/s10162-003-4021-2.

Hudspeth, A. J. (1997) 'How hearing happens', *Neuron,* 19(5), pp. 947-950. doi: 10.1016/s0896-6273(00)80385-2.

Hudspeth, A. J. (2014) 'Integrating the active process of hair cells with cochlear function', *Nat Rev Neurosci,* 15(9), pp. 600-14. doi: 10.1038/nrn3786.

Iliadou, V. V.*, et al.* (2017) 'A European Perspective on Auditory Processing Disorder-Current Knowledge and Future Research Focus', *Front Neurol,* 8, pp. 622. doi: 10.3389/fneur.2017.00622.

Ingham, N. J. *et al.* (2020) 'Functional analysis of candidate genes from genome-wide association studies of hearing', *Hear Res,* 387, pp. 107879. doi: 10.1016/j.heares.2019.107879.

Iyengar, S. K. and Elston, R. C. (2007) 'The genetic basis of complex traits: rare variants or "common gene, common disease"?', *Methods Mol Biol,* 376, pp. 71-84. doi: 10.1007/978-1-59745-389-9_6.

Jerger, J. and Musiek, F. (2000) 'Report of the Consensus Conference on the Diagnosis of Auditory Processing Disorders in School-Aged Children', *J Am Acad Audiol,* 11(9), pp. 467-74.

Jian, X., Boerwinkle, E. and Liu, X. (2014) 'In silico prediction of splice-altering single nucleotide variants in the human genome', *Nucleic Acids Res,* 42(22), pp. 13534-44. doi: 10.1093/nar/gku1206.

Joensuu, T.*, et al.* (2001) 'Mutations in a novel gene with transmembrane domains underlie Usher syndrome type 3', *Am J Hum Genet,* 69(4), pp. 673-84. doi: 10.1086/323610.

Jones, C.*, et al.* (2008) 'Ciliary proteins link basal body polarization to planar cell polarity regulation', *Nat Genet,* 40(1), pp. 69-77. doi: 10.1038/ng.2007.54.

Jouret, G.*, et al.* (2019) 'Genetics of Usher Syndrome: New Insights From a Meta-analysis', *Otol Neurotol,* 40(1), pp. 121-129. doi: 10.1097/mao.0000000000002054.

Kachar, B.*, et al.* (2000) 'High-resolution structure of hair-cell tip links', *Proc Natl Acad Sci U S A,* 97(24), pp. 13336-41. doi: 10.1073/pnas.97.24.13336.

Kalay, E.*, et al.* (2005) 'A novel D458V mutation in the SANS PDZ binding motif causes atypical Usher syndrome', *J Mol Med (Berl),* 83(12), pp. 1025-32. doi: 10.1007/s00109-005-0719-4.

Kalnak, N.*, et al.* (2018) 'Enrichment of rare copy number variation in children with developmental language disorder', *Clin Genet,* 94(3-4), pp. 313-320. doi: 10.1111/cge.13389.

Kamiya, K.*, et al.* (2014) 'An unusually powerful mode of low-frequency sound interference due to defective hair bundles of the auditory outer hair cells', *Proc Natl Acad Sci U S A,* 111(25), pp. 9307-12. doi: 10.1073/pnas.1405322111.

Kang, T. H.*, et al.* (2017) 'A novel missense variant in the DIAPH1 gene in a Korean family with autosomal dominant nonsyndromic hearing loss', *Genes Genet Syst,* 91(5), pp. 289-292. doi: 10.1266/ggs.16-00041.

Karim, Z.*, et al.* (2008) 'NHERF1 mutations and responsiveness of renal parathyroid hormone', *N Engl J Med,* 359(11), pp. 1128-35. doi: 10.1056/NEJMoa0802836.

Karlsson, K. K., Harris, J. R. and Svartengren, M. (1997) 'Description and primary results from an audiometric study of male twins', *Ear Hear,* 18(2), pp. 114-20. doi: 10.1097/00003446-199704000-00003.

Kato, T.*, et al.* (2001) 'Localization of a mammalian homolog of diaphanous, mDia1, to the mitotic spindle in HeLa cells', *J Cell Sci,* 114(Pt 4), pp. 775-84.

Kawase, T. and Liberman, M. C. (1993) 'Antimasking effects of the olivocochlear reflex .1. enhancement of compound action-potentials to masked tones', *Journal of Neurophysiology,* 70(6), pp. 2519-2532.

Kazmierczak, P. *et al.* (2007) 'Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells', *Nature,* 449(7158), pp. 87-91. doi: 10.1038/nature06091.
Keats, B. J. and Corey, D. P. (1999) 'The usher syndromes', *Am J Med Genet,* 89(3), pp. 158-66.
Keith, R. W. (1999) 'Clinical Issues in Central Auditory Processing Disorders', *Lang Speech Hear Serv Sch,* 30(4), pp. 339-344. doi: 10.1044/0161-1461.3004.339.
Khan, M. I.*, et al.* (2011) 'CLRN1 mutations cause nonsyndromic retinitis pigmentosa', *Ophthalmology,* 118(7), pp. 1444-8. doi: 10.1016/j.ophtha.2010.10.047.
Khateb, S.*, et al.* (2018) 'A homozygous founder missense variant in arylsulfatase G abolishes its enzymatic activity causing atypical Usher syndrome in humans', *Genet Med,* 20(9), pp. 1004-1012. doi: 10.1038/gim.2017.227.
Khateb, S.*, et al.* (2014) 'A homozygous nonsense CEP250 mutation combined with a heterozygous nonsense C2orf71 mutation is associated with atypical Usher syndrome', *Journal of Medical Genetics,* 51(7), pp. 460-469. doi: 10.1136/jmedgenet-2014-102287.
Khavarghazalani, B.*, et al.* (2016) 'Auditory processing abilities in children with chronic otitis media with effusion', *Acta Otolaryngol,* 136(5), pp. 456-9. doi: 10.3109/00016489.2015.1129552.
Kikkawa, Y.*, et al.* (2003) 'Mutations in a new scaffold protein Sans cause deafness in Jackson shaker mice', *Hum Mol Genet,* 12(5), pp. 453-61. doi: 10.1093/hmg/ddg042.
Kimberling, W. J.*, et al.* (2010) 'Frequency of Usher syndrome in two pediatric populations: Implications for genetic screening of deaf and hard of hearing children', *Genetics in Medicine,* 12(8), pp. 512-516. doi: 10.1097/GIM.0b013e3181e5afb8.
Kircher, M.*, et al.* (2014) 'A general framework for estimating the relative pathogenicity of human genetic variants', *Nat Genet,* 46(3), pp. 310-5. doi: 10.1038/ng.2892.
Kremer, H.*, et al.* (2006) 'Usher syndrome: molecular links of pathogenesis, proteins and pathways', *Hum Mol Genet,* 15 Spec No 2, pp. R262-70. doi: 10.1093/hmg/ddl205.
Kros, C. J.*, et al.* (2002) 'Reduced climbing and increased slipping adaptation in cochlear hair cells of mice with Myo7a mutations', *Nat Neurosci,* 5(1), pp. 41-7. doi: 10.1038/nn784.
Kruszewski, K.*, et al.* (2016) 'Degeneration of Photoreceptor Cells in Arylsulfatase G-Deficient Mice', *Invest Ophthalmol Vis Sci,* 57(3), pp. 1120-31. doi: 10.1167/iovs.15-17645.
Kubota, D.*, et al.* (2018) 'CEP250 mutations associated with mild cone-rod dystrophy and sensorineural hearing loss in a Japanese family', *Ophthalmic Genet,* 39(4), pp. 500-507. doi: 10.1080/13816810.2018.1466338.
Kujawa, S. G. and Liberman, M. C. (2009) 'Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss', *J Neurosci,* 29(45), pp. 14077-85. doi: 10.1523/jneurosci.2845-09.2009.
Kumar, A.*, et al.* (2013) 'CEP proteins: the knights of centrosome dynasty', *Protoplasma,* 250(5), pp. 965-83. doi: 10.1007/s00709-013-0488-9.
Küssel-Andermann, P.*, et al.* (2000) 'Vezatin, a novel transmembrane protein, bridges myosin VIIA to the cadherin-catenins complex', *Embo j,* 19(22), pp. 6020-9. doi: 10.1093/emboj/19.22.6020.
Lagziel, A.*, et al.* (2005) 'Spatiotemporal pattern and isoforms of cadherin 23 in wild type and waltzer mice during inner ear hair cell development', *Dev Biol,* 280(2), pp. 295-306. doi: 10.1016/j.ydbio.2005.01.015.
Lai, C. S.*, et al.* (2001) 'A forkhead-domain gene is mutated in a severe speech and language disorder', *Nature,* 413(6855), pp. 519-23. doi: 10.1038/35097076.
Le Quesne Stabej, P.*, et al.* (2012) 'Comprehensive sequence analysis of nine Usher syndrome genes in the UK National Collaborative Usher Study', *J Med Genet,* 49(1), pp. 27-36. doi: 10.1136/jmedgenet-2011-100468.
Lefèvre, G.*, et al.* (2008) 'A core cochlear phenotype in USH1 mouse mutants implicates fibrous links of the hair bundle in its cohesion, orientation and differential growth', *Development,* 135(8), pp. 1427-37. doi: 10.1242/dev.012922.
Lemay, P.*, et al.* (2017) 'Rare deleterious variants in GRHL3 are associated with human spina bifida', *Hum Mutat,* 38(6), pp. 716-724. doi: 10.1002/humu.23214.

Lenassi, E.*, et al.* (2015) 'A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants', *Eur J Hum Genet,* 23(10), pp. 1318-27. doi: 10.1038/ejhg.2014.283.

Lentz, J. and Keats, B. (1993a) 'Usher Syndrome Type II', in Adam, M.P.*, et al.* (eds.) *GeneReviews(®)*. Seattle (WA): University of Washington, SeattleCopyright © 1993-2020, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.

Lentz, J. and Keats, B. J. B. (1993b) 'Usher Syndrome Type I', in Adam, M.P.*, et al.* (eds.) *GeneReviews(®)*. Seattle (WA): University of Washington, SeattleCopyright © 1993-2020, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.

Lewis, M. A. *et al.* (2018) 'Whole exome sequencing in adult-onset hearing loss reveals a high load of predicted pathogenic variants in known deafness-associated genes and identifies new candidate genes', *BMC Med Genomics,* 11(1), pp. 77. doi: 10.1186/s12920-018-0395-1.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics,* 26(5), pp. 589-95. doi: 10.1093/bioinformatics/btp698.

Li, S.*, et al.* (2020) 'Myosin-VIIa is expressed in multiple isoforms and essential for tensioning the hair cell mechanotransduction complex', *Nat Commun,* 11(1), pp. 2066. doi: 10.1038/s41467-020-15936-z.

Li, T.*, et al.* (2019) 'A novel ABHD12 nonsense variant in Usher syndrome type 3 family with genotype-phenotype spectrum review', *Gene,* 704, pp. 113-120. doi: 10.1016/j.gene.2019.04.008.

Liberman, M. C. (1978) 'Auditory-nerve response from cats raised in a low-noise chamber', *J Acoust Soc Am,* 63(2), pp. 442-55. doi: 10.1121/1.381736.

Liberman, M. C.*, et al.* (2016) 'Toward a Differential Diagnosis of Hidden Hearing Loss in Humans', *PLoS One,* 11(9), pp. e0162726. doi: 10.1371/journal.pone.0162726.

Libé-Philippot, B. *et al.* (2017) 'Auditory cortex interneuron development requires cadherins operating hair-cell mechanoelectrical transduction', *Proc Natl Acad Sci U S A,* 114(30), pp. 7765-7774. doi: 10.1073/pnas.1703408114.

Lin, X. Q.*, et al.* (2018) 'Comparison of common burden tests for genetic association studies of rare variants', *Yi Chuan,* 40(2), pp. 162-169. doi: 10.16288/j.yczz.17-174.

Liu, X.*, et al.* (2007) 'Usherin is required for maintenance of retinal photoreceptors and normal development of cochlear hair cells', *Proc Natl Acad Sci U S A,* 104(11), pp. 4413-8. doi: 10.1073/pnas.0610950104.

Liu, X. Z.*, et al.* (1999) 'A mutation (2314delG) in the Usher syndrome type IIA gene: high prevalence and phenotypic variation', *Am J Hum Genet: Vol. 4*, pp. 1221-5.

Liu, X. Z.*, et al.* (1998) 'Mutations in the myosin VIIA gene cause a wide phenotypic spectrum, including atypical Usher syndrome', *Am J Hum Genet: Vol. 3*, pp. 909-12.

Liu, X. Z.*, et al.* (1997a) 'Mutations in the myosin VIIA gene cause non-syndromic recessive deafness', *Nat Genet,* 16(2), pp. 188-90. doi: 10.1038/ng0697-188.

Liu, X. Z.*, et al.* (1997b) 'Autosomal dominant non-syndromic deafness caused by a mutation in the myosin VIIA gene', *Nat Genet,* 17(3), pp. 268-9. doi: 10.1038/ng1197-268.

Liu, X. Z.*, et al.* (1997c) 'Autosomal dominant non-syndromic deafness caused by a mutation in the myosin VIIA gene', *Nature Genetics,* 17(3), pp. 268-269. doi: 10.1038/ng1197-268.

Luciano, M.*, et al.* (2013) 'A genome-wide association study for reading and language abilities in two population cohorts', *Genes Brain Behav,* 12(6), pp. 645-52. doi: 10.1111/gbb.12053.

Lynch, E. D.*, et al.* (1997) 'Nonsyndromic deafness DFNA1 associated with mutation of a human homolog of the Drosophila gene diaphanous', *Science,* 278(5341), pp. 1315-8.

Ma, X., McPherson, B. and Ma, L. (2016) 'Behavioral Signs of (Central) Auditory Processing Disorder in Children With Nonsyndromic Cleft Lip and/or Palate: A Parental Questionnaire Approach', *Cleft Palate Craniofac J,* 53(2), pp. 147-56. doi: 10.1597/14-057.

Maeda, R.*, et al.* (2014) 'Tip-link protein protocadherin 15 interacts with transmembrane channel-like proteins TMC1 and TMC2', *Proc Natl Acad Sci U S A,* 111(35), pp. 12907-12. doi: 10.1073/pnas.1402152111.

Mangia, L. R. L.*, et al.* (2019) 'Correlation between Type of Clefting and the Incidence of Otitis Media among Children with Lip and/or Palate Clefts', *ORL J Otorhinolaryngol Relat Spec,* 81(5-6), pp. 338-347. doi: 10.1159/000503237.

Mann, Z. F. and Kelley, M. W. (2011) 'Development of tonotopy in the auditory periphery', *Hear Res,* 276(1-2), pp. 2-15. doi: 10.1016/j.heares.2011.01.011.

Manor, U.*, et al.* (2011) 'Regulation of stereocilia length by myosin XVa and whirlin depends on the actin-regulatory protein Eps8', *Curr Biol,* 21(2), pp. 167-72. doi: 10.1016/j.cub.2010.12.046.

Mao, Y.*, et al.* (2011) 'Characterization of a Dchs1 mutant mouse reveals requirements for Dchs1-Fat4 signaling during mammalian development', *Development,* 138(5), pp. 947-57. doi: 10.1242/dev.057166.

Mathur, P. and Yang, J. (2015) 'Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities', *Biochimica Et Biophysica Acta-Molecular Basis of Disease,* 1852(3), pp. 406-420. doi: 10.1016/j.bbadis.2014.11.020.

Mburu, P.*, et al.* (1997) 'Mutation analysis of the mouse myosin VIIA deafness gene', *Genes Funct,* 1(3), pp. 191-203. doi: 10.1046/j.1365-4624.1997.00020.x.

Mburu, P.*, et al.* (2003) 'Defects in whirlin, a PDZ domain molecule involved in stereocilia elongation, cause deafness in the whirler mouse and families with DFNB31', *Nature Genetics,* 34(4), pp. 421-428. doi: 10.1038/ng1208.

McCarthy, M. I. and Hirschhorn, J. N. (2008) 'Genome-wide association studies: potential next steps on a genetic journey', *Hum Mol Genet,* 17(R2), pp. R156-65. doi: 10.1093/hmg/ddn289.

McGee, J.*, et al.* (2006) 'The very large G-protein-coupled receptor Vlgr1: A component of the ankle link complex required for the normal development of auditory hair bundles', *Journal of Neuroscience,* 26(24), pp. 6543-6553. doi: 10.1523/jneurosci.0693-06.2006.

McGee, T. L.*, et al.* (2010) 'Novel mutations in the long isoform of the USH2A gene in patients with Usher syndrome type II or non-syndromic retinitis pigmentosa', *J Med Genet,* 47(7), pp. 499-506. doi: 10.1136/jmg.2009.075143.

McMillan, D. R.*, et al.* (2002) 'Very large G protein-coupled receptor-1, the largest known cell surface protein, is highly expressed in the developing central nervous system', *J Biol Chem,* 277(1), pp. 785-92. doi: 10.1074/jbc.M108929200.

McNeill, E. M.*, et al.* (2010) 'Nav2 is necessary for cranial nerve development and blood pressure regulation', *Neural Development,* 5(1), pp. 6. doi: 10.1186/1749-8104-5-6.

Michalski, N.*, et al.* (2007) 'Molecular characterization of the ankle-link complex in cochlear hair cells and its role in the hair bundle functioning', *J Neurosci,* 27(24), pp. 6478-88. doi: 10.1523/jneurosci.0342-07.2007.

Michalski, N. and Petit, C. (2015) 'Genetics of auditory mechano-electrical transduction', *Pflugers Arch,* 467(1), pp. 49-72. doi: 10.1007/s00424-014-1552-9.

Michel, V.*, et al.* (2017) 'CIB2, defective in isolated deafness, is key for auditory hair cell mechanotransduction and survival', *EMBO Mol Med,* 9(12), pp. 1711-1731. doi: 10.15252/emmm.201708087.

Michel, V.*, et al.* (2005) 'Cadherin 23 is a component of the transient lateral links in the developing hair bundles of cochlear sensory cells', *Dev Biol,* 280(2), pp. 281-94. doi: 10.1016/j.ydbio.2005.01.014.

Mitchell, K. J. (2012) 'What is complex about complex disorders?', *Genome Biology,* 13(1), pp. 237. doi: 10.1186/gb-2012-13-1-237.

Molina-Ramírez, L. P.*, et al.* (2020) 'Establishing Genotype-phenotype Correlation in USH2A-related Disorders to Personalize Audiological Surveillance and Rehabilitation', *Otol Neurotol,* 41(4), pp. 431-437. doi: 10.1097/mao.0000000000002588.

Moore, D. R. (2018) 'Editorial: Auditory Processing Disorder', *Ear Hear,* 39(4), pp. 617-620. doi: 10.1097/aud.0000000000000582.

Moore, D. R.*, et al.* (2010) 'Nature of Auditory Processing Disorder in Children', *Pediatrics,* 126(2), pp. E382-E390. doi: 10.1542/peds.2009-2826.

Moore, D. R. and Hunter, L. L. (2013) 'Auditory processing disorder (APD) in children: A marker of neurodevelopmental syndrome', *Hearing, Balance and Communication,* 11(3), pp. 160-167. doi: 10.3109/21695717.2013.821756.

Moore, D. R.*, et al.* (2013) 'Evolving concepts of developmental auditory processing disorder (APD): A British Society of Audiology APD Special Interest Group 'white paper'', *International Journal of Audiology,* 52(1), pp. 3-13. doi: 10.3109/14992027.2012.723143.

Morell, R. J.*, et al.* (2007) 'A twin study of auditory processing indicates that dichotic listening ability is a strongly heritable trait', *Hum Genet,* 122(1), pp. 103-11. doi: 10.1007/s00439-007-0384-5.

Morris, A. P. and Zeggini, E. (2010) 'An evaluation of statistical approaches to rare variant analysis in genetic association studies', *Genet Epidemiol,* 34(2), pp. 188-93. doi: 10.1002/gepi.20450.

Muley, P. D.*, et al.* (2008) 'The atRA-responsive gene neuron navigator 2 functions in neurite outgrowth and axonal elongation', *Dev Neurobiol,* 68(13), pp. 1441-53. doi: 10.1002/dneu.20670.

Murakoshi, M., Suzuki, S. and Wada, H. (2015) 'All Three Rows of Outer Hair Cells Are Required for Cochlear Amplification', *Biomed Research International*. doi: 10.1155/2015/727434.

Murcia, N. S.*, et al.* (2000) 'The Oak Ridge Polycystic Kidney (orpk) disease gene is required for left-right axis determination', *Development,* 127(11), pp. 2347-55.

Musiek, F. E., Bellis, T. J. and Chermak, G. D. (2005) 'Nonmodularity of the central auditory nervous system: implications for (central) auditory processing disorder', *Am J Audiol,* 14(2), pp. 128-38; discussion 143-50. doi: 10.1044/1059-0889(2005/014).

Musiek, F. E. and Geurkink, N. A. (1980) 'Auditory perceptual problems in children: considerations for the otolaryngologist and audiologist', *Laryngoscope,* 90(6 Pt 1), pp. 962-71. doi: 10.1002/lary.1980.90.6.962.

Musiek, F. E.*, et al.* (2017) 'Perspectives on the Pure-Tone Audiogram', *J Am Acad Audiol,* 28(7), pp. 655-671. doi: 10.3766/jaaa.16061.

Myers, K. A.*, et al.* (2018) 'ADGRV1 is implicated in myoclonic epilepsy', *Epilepsia,* 59(2), pp. 381-388. doi: 10.1111/epi.13980.

Nakayama, J.*, et al.* (2002) 'A nonsense mutation of the MASS1 gene in a family with febrile and afebrile seizures', *Ann Neurol,* 52(5), pp. 654-7. doi: 10.1002/ana.10347.

Ness, S. L.*, et al.* (2003) 'Genetic homogeneity and phenotypic variability among Ashkenazi Jews with Usher syndrome type III', *J Med Genet,* 40(10), pp. 767-72. doi: 10.1136/jmg.40.10.767.

Neuhaus, C.*, et al.* (2017) 'Extension of the clinical and molecular phenotype of DIAPH1-associated autosomal dominant hearing loss (DFNA1)', *Clin Genet,* 91(6), pp. 892-901. doi: 10.1111/cge.12915.

Newbury, D. F.*, et al.* (2009) 'CMIP and ATP2C2 modulate phonological short-term memory in language impairment', *Am J Hum Genet,* 85(2), pp. 264-72. doi: 10.1016/j.ajhg.2009.07.004.

Ng, P. C. and Henikoff, S. (2001) 'Predicting deleterious amino acid substitutions', *Genome Res,* 11(5), pp. 863-74. doi: 10.1101/gr.176601.

Niemi, M. E. K.*, et al.* (2018) 'Common genetic variants contribute to risk of rare severe neurodevelopmental disorders', *Nature,* 562(7726), pp. 268-271. doi: 10.1038/s41586-018-0566-4.

Nishimura, D. Y.*, et al.* (2010) 'Discovery and functional analysis of a retinitis pigmentosa gene, C2ORF71', *Am J Hum Genet,* 86(5), pp. 686-95. doi: 10.1016/j.ajhg.2010.03.005.

Nishio, S. Y., Takumi, Y. and Usami, S. I. (2017) 'Laser-capture micro dissection combined with next-generation sequencing analysis of cell type-specific deafness gene expression in the mouse cochlea', *Hear Res,* 348, pp. 87-97. doi: 10.1016/j.heares.2017.02.017.

Nudel, R.*, et al.* (2014) 'Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment', *Genes Brain Behav,* 13(4), pp. 418-29. doi: 10.1111/gbb.12127.

Oshima, A.*, et al.* (2008) 'Mutation profile of the CDH23 gene in 56 probands with Usher syndrome type I', *Hum Mutat,* 29(6), pp. E37-46. doi: 10.1002/humu.20761.
Ousey, J.*, et al.* (1989) 'The IHR-McCormick Automated Toy Discrimination test--description and initial evaluation', *Br J Audiol,* 23(3), pp. 245-9. doi: 10.3109/03005368909076506.
Ouyang, X. M.*, et al.* (2005) 'Characterization of Usher syndrome type I gene mutations in an Usher syndrome patient population', *Hum Genet,* 116(4), pp. 292-9. doi: 10.1007/s00439-004-1227-2.
Oxenham, A. J. and Bacon, S. P. (2003) 'Cochlear compression: perceptual measures and implications for normal and impaired hearing', *Ear Hear,* 24(5), pp. 352-66. doi: 10.1097/01.aud.0000090470.73934.78.
Pakarinen, L.*, et al.* (1995) 'Usher's syndrome type 3 in Finland', *Laryngoscope,* 105(6), pp. 613-7. doi: 10.1288/00005537-199506000-00010.
Palmer, A. R., Sheppard, S. and Marshall, D. H. (1991) 'Prediction of hearing thresholds in children using an automated toy discrimination test', *Br J Audiol,* 25(5), pp. 351-6. doi: 10.3109/03005369109076609.
Pan, B.*, et al.* (2013) 'TMC1 and TMC2 are components of the mechanotransduction channel in hair cells of the mammalian inner ear', *Neuron,* 79(3), pp. 504-15. doi: 10.1016/j.neuron.2013.06.019.
Parker, M. A. (2020) 'Identifying three otopathologies in humans', *Hear Res,* 398, pp. 108079. doi: 10.1016/j.heares.2020.108079.
Paul, R. and Roth, F. P. (2011) 'Characterizing and predicting outcomes of communication delays in infants and toddlers: implications for clinical practice', *Lang Speech Hear Serv Sch,* 42(3), pp. 331-40. doi: 10.1044/0161-1461(2010/09-0067).
Pawlowski, K. S.*, et al.* (2006) 'Progression of inner ear pathology in Ames waltzer mice and the role of protocadherin 15 in hair cell development', *J Assoc Res Otolaryngol,* 7(2), pp. 83-94. doi: 10.1007/s10162-005-0024-5.
Peeters, P. J.*, et al.* (2004) 'Sensory deficits in mice hypomorphic for a mammalian homologue of unc-53', *Brain Res Dev Brain Res,* 150(2), pp. 89-101. doi: 10.1016/j.devbrainres.2004.03.004.
Peltonen, L. and McKusick, V. A. (2001) 'Genomics and medicine. Dissecting human disease in the postgenomic era', *Science,* 291(5507), pp. 1224-9. doi: 10.1126/science.291.5507.1224.
Perrino, P. A., Newbury, D. F. and Fitch, R. H. (2021) 'Peripheral Anomalies in USH2A Cause Central Auditory Anomalies in a Mouse Model of Usher Syndrome and CAPD', *Genes,* 12(2). doi: 10.3390/genes12020151.
Perrino, P. A.*, et al.* (2020) 'Multi-level evidence of an allelic hierarchy of USH2A variants in hearing, auditory processing and speech/language outcomes', *Commun Biol,* 3(1), pp. 180. doi: 10.1038/s42003-020-0885-5.
Peter, V. G.*, et al.* (2020) 'New clinical and molecular evidence linking mutations in ARSG to Usher syndrome type IV', *Hum Mutat*. doi: 10.1002/humu.24150.
Peters, L. M.*, et al.* (2002) 'Mutation of a transcription factor, TFCP2L3, causes progressive autosomal dominant hearing loss, DFNA28', *Hum Mol Genet,* 11(23), pp. 2877-85. doi: 10.1093/hmg/11.23.2877.
Pettigrew, K. A.*, et al.* (2015) 'Copy Number Variation Screen Identifies a Rare De Novo Deletion at Chromosome 15q13.1-13.3 in a Child with Language Impairment', *PLoS One,* 10(8), pp. e0134997. doi: 10.1371/journal.pone.0134997.
Peyrard-Janvid, M.*, et al.* (2014) 'Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development', *Am J Hum Genet,* 94(1), pp. 23-32. doi: 10.1016/j.ajhg.2013.11.009.
Peñagarikano, O.*, et al.* (2011) 'Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits', *Cell,* 147(1), pp. 235-46. doi: 10.1016/j.cell.2011.08.040.
Pietola, L.*, et al.* (2012) 'Speech recognition and communication outcomes with cochlear implantation in Usher syndrome type 3', *Otol Neurotol,* 33(1), pp. 38-41. doi: 10.1097/MAO.0b013e31823dbc56.

Plantinga, R. F.*, et al.* (2005) 'Serial audiometry and speech recognition findings in Finnish Usher syndrome type III patients', *Audiol Neurootol,* 10(2), pp. 79-89. doi: 10.1159/000083363.
Pollard, K. S.*, et al.* (2010) 'Detection of nonneutral substitution rates on mammalian phylogenies', *Genome Res,* 20(1), pp. 110-21. doi: 10.1101/gr.097857.109.
Pook, C., Ahrens, J. M. and Clagett-Dame, M. (2020) 'Expression pattern of Nav2 in the murine CNS with development', *Gene Expr Patterns,* 35, pp. 119099. doi: 10.1016/j.gep.2020.119099.
Pritchard, J. K. (2001) 'Are rare variants responsible for susceptibility to complex diseases?', *Am J Hum Genet,* 69(1), pp. 124-37. doi: 10.1086/321272.
Puffenberger, E. G.*, et al.* (2012) 'Genetic mapping and exome sequencing identify variants associated with five novel diseases', *PLoS One,* 7(1), pp. e28936. doi: 10.1371/journal.pone.0028936.
Purcell, S.*, et al.* (2007) 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *Am J Hum Genet,* 81(3), pp. 559-75. doi: 10.1086/519795.
Purves D., Augustine G.J. and Fitzpatrick D. (2001) 'The Auditory Cortex' 2nd edn. Sunderland (MA): Sinauer Associates.Available at: https://www.ncbi.nlm.nih.gov/books/NBK10900/
Raphael, Y.*, et al.* (2001) 'Severe vestibular and auditory impairment in three alleles of Ames waltzer (av) mice', *Hear Res,* 151(1-2), pp. 237-249. doi: 10.1016/s0378-5955(00)00233-1.
Reiners, J.*, et al.* (2006) 'Molecular basis of human Usher syndrome: deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease', *Exp Eye Res,* 83(1), pp. 97-119. doi: 10.1016/j.exer.2005.11.010.
Reynard, P., Veuillet, E. and Thai-Van, H. (2020) 'Paradoxical and labile medial olivocochlear functioning as a potential marker of auditory processing disorder in a child with learning disabilities', *Eur Ann Otorhinolaryngol Head Neck Dis*. doi: 10.1016/j.anorl.2020.03.005.
Rhodes, C. R.*, et al.* (2004) 'A Myo7a mutation cosegregates with stereocilia defects and low-frequency hearing impairment', *Mammalian Genome,* 15(9), pp. 686-697. doi: 10.1007/s00335-004-2344-x.
Riazuddin, S.*, et al.* (2012) 'Alterations of the CIB2 calcium- and integrin-binding protein cause Usher syndrome type 1J and nonsyndromic deafness DFNB48', *Nat Genet,* 44(11), pp. 1265-71. doi: 10.1038/ng.2426.
Richards, S.*, et al.* (2015) 'Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology', *Genet Med,* 17(5), pp. 405-24. doi: 10.1038/gim.2015.30.
Richardson, G. P. and Petit, C. (2019) 'Hair-Bundle Links: Genetics as the Gateway to Function', *Cold Spring Harbor perspectives in medicine*. doi: 10.1101/cshperspect.a033142.
Rivolta, C.*, et al.* (2000) 'Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss', *Am J Hum Genet,* 66(6), pp. 1975-8. doi: 10.1086/302926.
Robinson, B. L. and McAlpine, D. (2009) 'Gain control mechanisms in the auditory pathway', *Curr Opin Neurobiol,* 19(4), pp. 402-7. doi: 10.1016/j.conb.2009.07.006.
Rock, R., Schrauth, S. and Gessler, M. (2005) 'Expression of mouse dchs1, fjx1, and fat-j suggests conservation of the planar cell polarity pathway identified in Drosophila', *Dev Dyn,* 234(3), pp. 747-55. doi: 10.1002/dvdy.20515.
Rodenas-Cuadrado, P.*, et al.* (2016) 'Characterisation of CASPR2 deficiency disorder--a syndrome involving autism, epilepsy and language impairment', *BMC Med Genet,* 17, pp. 8. doi: 10.1186/s12881-016-0272-8.
Rose, R.*, et al.* (2005) 'Structural and mechanistic insights into the interaction between Rho and mammalian Dia', *Nature,* 435(7041), pp. 513-8. doi: 10.1038/nature03604.
Rosenbaum, J. L. and Witman, G. B. (2002) 'Intraflagellar transport', *Nat Rev Mol Cell Biol,* 3(11), pp. 813-25. doi: 10.1038/nrm952.
Rosenfeld, R. M.*, et al.* (2016) 'Clinical Practice Guideline: Otitis Media with Effusion Executive Summary (Update)', *Otolaryngol Head Neck Surg,* 154(2), pp. 201-14. doi: 10.1177/0194599815624407.

Roux, A. F.*, et al.* (2006) 'Survey of the frequency of USH1 gene mutations in a cohort of Usher patients shows the importance of cadherin 23 and protocadherin 15 genes and establishes a detection rate of above 90%', *J Med Genet: Vol. 9*, pp. 763-8.

Saburi, S.*, et al.* (2012) 'Functional interactions between Fat family cadherins in tissue morphogenesis and planar polarity', *Development,* 139(10), pp. 1806-20. doi: 10.1242/dev.077461.

Sadeghi, M.*, et al.* (2005) 'Audiological and vestibular features in affected subjects with USH3: a genotype/phenotype correlation', *Int J Audiol,* 44(5), pp. 307-16. doi: 10.1080/14992020500060610.

Salvi, R.*, et al.* (2016) 'Inner Hair Cell Loss Disrupts Hearing and Cochlear Function Leading to Sensory Deprivation and Enhanced Central Auditory Gain', *Front Neurosci,* 10, pp. 621. doi: 10.3389/fnins.2016.00621.

Satterstrom, F. K.*, et al.* (2020) 'Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism', *Cell,* 180(3), pp. 568-584.e23. doi: 10.1016/j.cell.2019.12.036.

Saxena, U., Allan, C. and Allen, P. (2015) 'Crossed and uncrossed acoustic reflex growth functions in normal-hearing adults, typically developing children, and children with suspected auditory processing disorder', *Int J Audiol,* 54(9), pp. 620-6. doi: 10.3109/14992027.2015.1043147.

Schaette, R. and McAlpine, D. (2011) 'Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model', *J Neurosci,* 31(38), pp. 13452-7. doi: 10.1523/jneurosci.2156-11.2011.

Scheel, H., Tomiuk, S. and Hofmann, K. (2002) 'A common protein interaction domain links two recently identified epilepsy genes', *Hum Mol Genet,* 11(15), pp. 1757-62. doi: 10.1093/hmg/11.15.1757.

Schneider, E.*, et al.* (2009) 'Homozygous disruption of PDZD7 by reciprocal translocation in a consanguineous family: a new member of the Usher syndrome protein interactome causing congenital hearing impairment', *Hum Mol Genet,* 18(4), pp. 655-66. doi: 10.1093/hmg/ddn395.

Schneider, M. E.*, et al.* (2006) 'A new compartment at stereocilia tips defined by spatial and temporal patterns of myosin IIIa expression', *J Neurosci,* 26(40), pp. 10243-52. doi: 10.1523/jneurosci.2812-06.2006.

Schultz, J. M.*, et al.* (2011) 'Allelic hierarchy of CDH23 mutations causing non-syndromic deafness DFNB12 or Usher syndrome USH1D in compound heterozygotes', *J Med Genet,* 48(11), pp. 767-75. doi: 10.1136/jmedgenet-2011-100262.

Schwander, M., Kachar, B. and Müller, U. (2010) 'Review series: The cell biology of hearing', *J Cell Biol,* 190(1), pp. 9-20. doi: 10.1083/jcb.201001138.

Scott, K. E.*, et al.* (2018) 'Altered Auditory Processing, Filtering, and Reactivity in the Cntnap2 Knock-Out Rat Model for Neurodevelopmental Disorders', *J Neurosci,* 38(40), pp. 8588-8604. doi: 10.1523/jneurosci.0759-18.2018.

Self, T.*, et al.* (1998) 'Shaker-1 mutations reveal roles for myosin VIIA in both development and function of cochlear hair cells', *Development,* 125(4), pp. 557-66.

Shafer, V. L., Schwartz, R. G. and Martin, B. (2011) 'Evidence of deficient central speech processing in children with specific language impairment: the T-complex', *Clin Neurophysiol,* 122(6), pp. 1137-55. doi: 10.1016/j.clinph.2010.10.046.

Sharma, M., Purdy, S. C. and Kelly, A. S. (2009) 'Comorbidity of auditory processing, language, and reading disorders', *J Speech Lang Hear Res,* 52(3), pp. 706-22. doi: 10.1044/1092-4388(2008/07-0226).

Sheahan, P.*, et al.* (2003) 'Incidence and outcome of middle ear disease in cleft lip and/or cleft palate', *Int J Pediatr Otorhinolaryngol,* 67(7), pp. 785-93. doi: 10.1016/s0165-5876(03)00098-3.

Shearer, A. E.*, et al.* (2014) 'Copy number variants are a common cause of non-syndromic hearing loss', *Genome Med,* 6(5), pp. 37. doi: 10.1186/gm554.

Shotwell, S. L., Jacobs, R. and Hudspeth, A. J. (1981) 'Directional sensitivity of individual vertebrate hair cells to controlled deflection of their hair bundles', *Ann N Y Acad Sci,* 374, pp. 1-10. doi: 10.1111/j.1749-6632.1981.tb30854.x.

Siemens, J.*, et al.* (2004) 'Cadherin 23 is a component of the tip link in hair-cell stereocilia', *Nature,* 428(6986), pp. 950-5. doi: 10.1038/nature02483.

Siepel, A.*, et al.* (2005) 'Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes', *Genome Res,* 15(8), pp. 1034-50. doi: 10.1101/gr.3715005.

Simpson, N. H.*, et al.* (2015) 'Genome-wide analysis identifies a role for common copy number variants in specific language impairment', *Eur J Hum Genet,* 23(10), pp. 1370-7. doi: 10.1038/ejhg.2014.296.

Skradski, S. L.*, et al.* (2001) 'A novel gene causing a mendelian audiogenic mouse epilepsy', *Neuron,* 31(4), pp. 537-44. doi: 10.1016/s0896-6273(01)00397-x.

St Pourcain, B.*, et al.* (2014) 'Common variation near ROBO2 is associated with expressive vocabulary in infancy', *Nat Commun,* 5, pp. 4831. doi: 10.1038/ncomms5831.

Stanton, S. G.*, et al.* (2005) 'Hearing screening outcomes in infants of pregestational diabetic mothers', *Am J Audiol,* 14(1), pp. 86-93. doi: 10.1044/1059-0889(2005/008).

Strauss, K. A.*, et al.* (2006) 'Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2', *N Engl J Med,* 354(13), pp. 1370-7. doi: 10.1056/NEJMoa052773.

Stringham, E. G. and Schmidt, K. L. (2009) 'Navigating the cell', *Cell Adhesion & Migration,* 3(4), pp. 342-346. doi: 10.4161/cam.3.4.9451.

Stritt, S.*, et al.* (2016) 'A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss', *Blood,* 127(23), pp. 2903-14. doi: 10.1182/blood-2015-10-675629.

Summerfield, Q.*, et al.* (1994) 'Clinical evaluation and test-retest reliability of the IHR-McCormick Automated Toy Discrimination Test', *Br J Audiol,* 28(3), pp. 165-79. doi: 10.3109/03005369409086564.

Tallal, P. (2004) 'Improving language and literacy is a matter of time', *Nat Rev Neurosci,* 5(9), pp. 721-8. doi: 10.1038/nrn1499.

Tam, V.*, et al.* (2019) 'Benefits and limitations of genome-wide association studies', *Nat Rev Genet,* 20(8), pp. 467-484. doi: 10.1038/s41576-019-0127-1.

Taylor, P. N.*, et al.* (2015) 'Whole-genome sequence-based analysis of thyroid function', *Nat Commun,* 6, pp. 5681. doi: 10.1038/ncomms6681.

Tian, C.*, et al.* (2010) 'Ush1c gene expression levels in the ear and eye suggest different roles for Ush1c in neurosensory organs in a new Ush1c knockout mouse', *Brain Res,* 1328, pp. 57-70. doi: 10.1016/j.brainres.2010.02.079.

Tian, G.*, et al.* (2009) 'Clarin-1, encoded by the Usher Syndrome III causative gene, forms a membranous microdomain: possible role of clarin-1 in organizing the actin cytoskeleton', *J Biol Chem,* 284(28), pp. 18980-93. doi: 10.1074/jbc.M109.003160.

Tian, H.*, et al.* (2017) 'Intraflagellar transport 88 (IFT88) is crucial for craniofacial development in mice and is a candidate gene for human cleft lip and palate', *Hum Mol Genet,* 26(5), pp. 860-872. doi: 10.1093/hmg/ddx002.

Tilney, L. G., Tilney, M. S. and DeRosier, D. J. (1992) 'Actin filaments, stereocilia, and hair cells: how cells count and measure', *Annu Rev Cell Biol,* 8, pp. 257-74. doi: 10.1146/annurev.cb.08.110192.001353.

Timpson, N. J.*, et al.* (2014) 'A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans', *Nat Commun,* 5, pp. 4871. doi: 10.1038/ncomms5871.

Ting, S. B.*, et al.* (2003) 'Inositol- and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3', *Nat Med,* 9(12), pp. 1513-9. doi: 10.1038/nm961.

Tiwari, A.*, et al.* (2016) 'Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies', *Sci Rep,* 6, pp. 28755. doi: 10.1038/srep28755.

Toma, C.*, et al.* (2018) 'Comprehensive cross-disorder analyses of CNTNAP2 suggest it is unlikely to be a primary risk gene for psychiatric disorders', *PLoS Genet,* 14(12), pp. e1007535. doi: 10.1371/journal.pgen.1007535.

Tran, N. Q. V. and Miyake, K. (2017) 'Neurodevelopmental Disorders and Environmental Toxicants: Epigenetics as an Underlying Mechanism', *Int J Genomics,* 2017, pp. 7526592. doi: 10.1155/2017/7526592.

Truong, D. T.*, et al.* (2015) 'Auditory processing and morphological anomalies in medial geniculate nucleus of Cntnap2 mutant mice', *Behav Neurosci,* 129(6), pp. 731-43. doi: 10.1037/bne0000096.

Tsuprun, V.*, et al.* (2003) 'Structure of the stereocilia side links and morphology of auditory hair bundle in relation to noise exposure in the chinchilla', *J Neurocytol,* 32(9), pp. 1117-28. doi: 10.1023/B:NEUR.0000021906.08847.d2.

Ueyama, T.*, et al.* (2016) 'Constitutive activation of DIA1 (DIAPH1) via C-terminal truncation causes human sensorineural hearing loss', *EMBO Mol Med,* 8(11), pp. 1310-1324. doi: 10.15252/emmm.201606609.

Valero, R.*, et al.* (2019) 'Aberrant Splicing Events Associated to CDH23 Noncanonical Splice Site Mutations in a Proband with Atypical Usher Syndrome 1', *Genes (Basel),* 10(10). doi: 10.3390/genes10100732.

van Aarem, A.*, et al.* (1995) 'The Usher syndrome type 2A: clinical findings in obligate carriers', *Int J Pediatr Otorhinolaryngol,* 31(2-3), pp. 159-74.

van der Ven, A. T.*, et al.* (2017) 'Whole-Exome Sequencing Reveals FAT4 Mutations in a Clinically Unrecognizable Patient with Syndromic CAKUT: A Case Report', *Mol Syndromol,* 8(5), pp. 272-277. doi: 10.1159/000477750.

van Eeden, S. and Stringer, H. (2020) 'Linguistic and auditory processing skills in non-syndromic children with cleft palate: A scoping review', *J Commun Disord,* 87, pp. 106029. doi: 10.1016/j.jcomdis.2020.106029.

Van Laer, L.*, et al.* (2008) 'The grainyhead like 2 gene (GRHL2), alias TFCP2L3, is associated with age-related hearing impairment', *Hum Mol Genet,* 17(2), pp. 159-69. doi: 10.1093/hmg/ddm292.

Vanniasegaram, I., Cohen, M. and Rosen, S. (2004) 'Evaluation of selected auditory tests in school-age children suspected of auditory processing disorders', *Ear Hear,* 25(6), pp. 586-97. doi: 10.1097/01.aud.0000151575.58269.19.

Vernes, S. C.*, et al.* (2008) 'A Functional Genetic Link between Distinct Developmental Language Disorders', *New England Journal of Medicine,* 359(22), pp. 2337-2345. doi: 10.1056/NEJMoa0802828.

Verpy, E.*, et al.* (2011) 'Stereocilin connects outer hair cell stereocilia to one another and to the tectorial membrane', *J Comp Neurol,* 519(2), pp. 194-210. doi: 10.1002/cne.22509.

Verpy, E.*, et al.* (2000) 'A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C', *Nat Genet,* 26(1), pp. 51-5. doi: 10.1038/79171.

Villanueva, P.*, et al.* (2015) 'Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment', *PLoS Genet,* 11(3), pp. e1004925. doi: 10.1371/journal.pgen.1004925.

Vuckovic, D. *et al.* (2015) 'Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss', *Hum Mol Genet,* 24(19), pp. 5655-64. doi: 10.1093/hmg/ddv279.

Wagenaar, M.*, et al.* (1995) 'Clinical findings in obligate carriers of type I Usher syndrome', *Am J Med Genet,* 59(3), pp. 375-9. doi: 10.1002/ajmg.1320590319.

Wake, M.*, et al.* (2006) 'Slight/mild sensorineural hearing loss in children', *Pediatrics,* 118(5), pp. 1842-51. doi: 10.1542/peds.2005-3168.

Wallar, B. J. and Alberts, A. S. (2003) 'The formins: active scaffolds that remodel the cytoskeleton', *Trends Cell Biol,* 13(8), pp. 435-46. doi: 10.1016/s0962-8924(03)00153-3.

Walter, K.*, et al.* (2015) 'The UK10K project identifies rare variants in health and disease', *Nature,* 526(7571), pp. 82-90. doi: 10.1038/nature14962.

Wan, G. and Corfas, G. (2017) 'Transient auditory nerve demyelination as a new mechanism for hidden hearing loss', *Nat Commun,* 8, pp. 14487. doi: 10.1038/ncomms14487.

Wang, K. S.*, et al.* (2017a) 'Family-based association analysis of NAV2 gene with the risk and age at onset of Alzheimer's disease', *J Neuroimmunol,* 310, pp. 60-65. doi: 10.1016/j.jneuroim.2017.06.010.
Wang, W. Y.*, et al.* (2005) 'Genome-wide association studies: theoretical and practical concerns', *Nat Rev Genet,* 6(2), pp. 109-18. doi: 10.1038/nrg1522.
Wang, Y. F.*, et al.* (2017b) 'Loss of CIB2 Causes Profound Hearing Loss and Abolishes Mechanoelectrical Transduction in Mice', *Frontiers in Molecular Neuroscience,* 10. doi: 10.3389/fnmol.2017.00401.
Washington, J. L., 3rd*, et al.* (2005) 'Characterization of a new allele of Ames waltzer generated by ENU mutagenesis', *Hear Res,* 202(1-2), pp. 161-9. doi: 10.1016/j.heares.2004.09.014.
Wechsler, D., Golombok, S., and Rust, J. (1992) 'Weschler Intelligence Scale for Children', 3rd edn London: The Psychological Corporation.
Weil, D.*, et al.* (1995) 'Defective myosin VIIA gene responsible for Usher syndrome type 1B', *Nature,* 374(6517), pp. 60-1. doi: 10.1038/374060a0.
Weil, D.*, et al.* (2003) 'Usher syndrome type I G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with the USH1C protein, harmonin', *Hum Mol Genet,* 12(5), pp. 463-71.
Weil, D.*, et al.* (1997) 'The autosomal recessive isolated deafness, DFNB2, and the Usher 1B syndrome are allelic defects of the myosin-VIIA gene', *Nat Genet,* 16(2), pp. 191-3. doi: 10.1038/ng0697-191.
Wells, H. R. R.*, et al.* (2019) 'GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank', *Am J Hum Genet,* 105(4), pp. 788-802. doi: 10.1016/j.ajhg.2019.09.008.
Weston, M. D.*, et al.* (2004) 'Mutations in the VLGR1 gene implicate G-protein signaling in the pathogenesis of Usher syndrome type II', *Am J Hum Genet,* 74(2), pp. 357-66. doi: 10.1086/381685.
Wilson, W. J. and Arnott, W. (2013) 'Using different criteria to diagnose (central) auditory processing disorder: how big a difference does it make?', *J Speech Lang Hear Res,* 56(1), pp. 63-70. doi: 10.1044/1092-4388(2012/11-0352).
Winer, J. A. (2006) 'Decoding the auditory corticofugal systems', *Hear Res,* 212(1-2), pp. 1-8. doi: 10.1016/j.heares.2005.06.014.
Winslow, R. L. and Sachs, M. B. (1987) 'Effect of electrical-stimulation of the crossed olivocochlear bundle on auditory-nerve response to tones in noise', *Journal of Neurophysiology,* 57(4), pp. 1002-1021.
Witton, C. (2010) 'Childhood auditory processing disorder as a developmental disorder: the case for a multi-professional approach to diagnosis and management', *Int J Audiol,* 49(2), pp. 83-7. doi: 10.3109/14992020903289808.
Wolber, L. E.*, et al.* (2012) 'Hearing ability with age in northern European women: a new web-based approach to genetic studies', *PLoS One,* 7(4), pp. e35500. doi: 10.1371/journal.pone.0035500.
Worthey, E. A.*, et al.* (2013) 'Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech', *J Neurodev Disord,* 5(1), pp. 29. doi: 10.1186/1866-1955-5-29.
Wu, C. C.*, et al.* (2015) 'Identifying Children With Poor Cochlear Implantation Outcomes Using Massively Parallel Sequencing', *Medicine (Baltimore),* 94(27), pp. e1073. doi: 10.1097/md.0000000000001073.
Wu, K.*, et al.* (2020) 'A novel variant in diaphanous homolog 1 (DIAPH1) as the cause of auditory neuropathy in a Chinese family', *Int J Pediatr Otorhinolaryngol,* 133, pp. 109947. doi: 10.1016/j.ijporl.2020.109947.
Yan, J.*, et al.* (2010) 'The structure of the harmonin/sans complex reveals an unexpected interaction mode of the two Usher syndrome proteins', *Proc Natl Acad Sci U S A,* 107(9), pp. 4040-5. doi: 10.1073/pnas.0911385107.
Yang, L.*, et al.* (2013) 'Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common

and rare variants', *Am J Med Genet B Neuropsychiatr Genet,* 162b(5), pp. 419-430. doi: 10.1002/ajmg.b.32169.

Yoshimura, H.*, et al.* (2015) 'Novel ABHD12 mutations in PHARC patients: the differential diagnosis of deaf-blindness', *Ann Otol Rhinol Laryngol,* 124 Suppl 1, pp. 77s-83s. doi: 10.1177/0003489415574513.

Zallocchi, M.*, et al.* (2009) 'Localization and expression of clarin-1, the Clrn1 gene product, in auditory hair cells and photoreceptors', *Hear Res,* 255(1-2), pp. 109-20. doi: 10.1016/j.heares.2009.06.006.

Zallocchi, M.*, et al.* (2012) 'Role for a novel Usher protein complex in hair cell synaptic maturation', *PLoS One,* 7(2), pp. e30573. doi: 10.1371/journal.pone.0030573.

Zarrei, M.*, et al.* (2019) 'A large data resource of genomic copy number variation across neurodevelopmental disorders', *NPJ Genom Med,* 4, pp. 26. doi: 10.1038/s41525-019-0098-3.

Zhan, X.*, et al.* (2016) 'RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data', *Bioinformatics,* 32(9), pp. 1423-6. doi: 10.1093/bioinformatics/btw079.

Zheng, Q. Y., Johnson, K. R. and Erway, L. C. (1999) 'Assessment of hearing in 80 inbred strains of mice by ABR threshold analyses', *Hear Res,* 130(1-2), pp. 94-107. doi: 10.1016/s0378-5955(99)00003-9.

Zimmerman, E. (2018) 'Do Infants Born Very Premature and Who Have Very Low Birth Weight Catch Up With Their Full Term Peers in Their Language Abilities by Early School Age?', *J Speech Lang Hear Res,* 61(1), pp. 53-65. doi: 10.1044/2017_jslhr-l-16-0150.

Zou, J.*, et al.* (2017) 'The roles of USH1 proteins and PDZ domain-containing USH proteins in USH2 complex integrity in cochlear hair cells', *Hum Mol Genet,* 26(3), pp. 624-636. doi: 10.1093/hmg/ddw421.

Zou, J.*, et al.* (2014) 'Deletion of PDZD7 disrupts the Usher syndrome type 2 protein complex in cochlear hair cells and causes hearing loss in mice', *Hum Mol Genet,* 23(9), pp. 2374-90. doi: 10.1093/hmg/ddt629.

Zou, J. H.*, et al.* (2015) 'Individual USH2 proteins make distinct contributions to the ankle link complex during development of the mouse cochlear stereociliary bundle', *Human Molecular Genetics,* 24(24), pp. 6944-6957. doi: 10.1093/hmg/ddv398.

Zwaenepoel, I.*, et al.* (2001) 'Identification of three novel mutations in the USH1C gene and detection of thirty-one polymorphisms used for haplotype analysis', *Hum Mutat,* 17(1), pp. 34-41. doi: 10.1002/1098-1004(2001)17:1<34::aid-humu4>3.0.co;2-o.