

BROADENING THE GENOMIC LANDSCAPE OF SENSORY DISORDERS

Suzanne E. de Bruijn



Broadening the genomic landscape of sensory disorders

SUZANNE E. DE BRUIJN

The work presented in this thesis was carried out within the Donders Institute for Brain Cognition and Behaviour, at the Department of Human Genetics of the Radboud university medical center in Nijmegen, the Netherlands.

The printing of this thesis was generously supported by the Landelijke Stichting voor Blinden en Slechtienden, Rotterdamse Stichting Blindenbelangen, ChipSoft and the Radboud university medical center.

ISBN: 978-94-6416-789-4

Lay-out and cover design: James Jardine | www.jamesjardine.nl

Printing: Ridderprint | www.ridderprint.nl

© 2021, **Suzanne E. de Bruijn**. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission.

Broadening the genomic landscape of sensory disorders

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,

volgens besluit van het college voor promoties

in het openbaar te verdedigen op vrijdag 18 februari 2022

om 12.30 uur precies

door

Suzanne Eveline de Bruijn

geboren op 25 mei 1993

te Bemmelen

Promotoren:

Prof. dr. J.M.J. Kremer

Prof. dr. F.P.M. Cremers

Copromotoren:

Dr. S. Roosing

Dr. E. de Vrieze

Manuscriptcommissie:

Prof. dr. D.J. Lefeber (voorzitter)

Prof. dr. B. Wieringa

Prof. dr. E. De Baere (Universiteit Gent, België)

Broadening the genomic landscape of sensory disorders

Dissertation

to obtain the degree of doctor

from Radboud University Nijmegen

on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,

according to the decision of the Doctorate Board

to be defended in public on Friday, February 18, 2022

at 12.30 pm

by

Suzanne Eveline de Bruijn

born on May 25, 1993

in Bemmelen (the Netherlands)

Supervisors:

Prof. dr. J.M.J. Kremer
Prof. dr. F.P.M. Cremers

Co-supervisors:

Dr. S. Roosing
Dr. E. de Vrieze

Manuscript Committee:

Prof. dr. D.J. Lefeber (chair)
Prof. dr. B. Wieringa
Prof. dr. E. De Baere (Ghent University, Belgium)

TABLE OF CONTENTS

Chapter 1	Introduction	11
Chapter 1.1	General introduction	11
Chapter 1.2	The impact of modern technologies on molecular diagnostic success rates, with a focus on inherited retinal dystrophy and hearing loss	35
Chapter 2	Homozygous variants in <i>KIAA1549</i>, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa	77
Chapter 3	DFNA21	99
Chapter 3.1	A <i>RIPOR2</i> in-frame deletion is a frequent and highly penetrant cause of adult-onset hearing loss	99
Chapter 3.2	The development of a genetic therapy for DFNA21 using allele-specific antisense oligonucleotides	147
Chapter 4	Structural variants create new topological associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa	167
Chapter 5	Exploring the missing heritability in subjects with hearing loss, enlarged vestibular aqueducts, and a single or no pathogenic <i>SLC26A4</i> variant	231
Chapter 6	General discussion and perspectives	281
Chapter 7	Summary	319
Chapter 7.1	Summary	319
Chapter 7.2	Samenvatting	323
Chapter 8	Appendix	
	Data management plan	331
	List of publications	333
	PhD portfolio	335
	About the author	337
	Acknowledgements	339
	Abbreviations	347

CHAPTER 1

INTRODUCTION



Chapter 1.1

General introduction



INTRODUCTION

Hearing, smell, taste, touch and vision are considered the primary senses of the human body. Sensation is defined as the ability to transduce, encode and perceive information that is generated from our external and internal environment. A significant portion of the brain is devoted to this task; the organ is continuously distinguishing relevant input from background signals, a process called perception. This is a critical process, and the output serves as the most important determinant for behavior, cognition and thought. Vision and hearing are by many considered as the most important of the five senses. Already since the old days, they have been critical for the survival and evolution of mankind. Vision allows the early observation of potential threats, while hearing allows detection of enemies that are not even within sight yet. Today, these senses do not only warn us for danger; they have also become more and more important in social (e.g. communication) and cultural (e.g. music and art) aspects of life.

With this in mind, it is not surprising that vision and hearing disabilities have severe consequences for the quality of life.¹ Individuals with a visual disability are severely handicapped in today's society, which increasingly relies on rapid (visual) communication. Also the impact of hearing loss (HL) should not be underestimated; disabling HL has been linked with several mental conditions including depression.^{2,3} Especially when HL manifests before the development of speech, it can have serious negative consequences for the cognitive and socio-emotional development.⁴ Therefore, both the auditory system and the visual system are critical for a person's wellbeing. For both visual and hearing disabilities, genetic etiologies play a crucial role in the development of disease and are the topic of this thesis.

In this chapter, the basic principles of the visual and auditory systems are discussed. The first part addresses the visual system, starting with the anatomy of the eye followed by an overview of inherited retinal dystrophies (RD). Next, the anatomy of the ear is discussed and followed by the different aspects of (inherited) HL. In the final section of this chapter, the general aim and contents of the thesis will be described.

ANATOMY AND PHYSIOLOGY OF THE EYE

The human eye is a highly specialized sensory organ, responsible for the capture and detection of light signals. It allows for the visualization of objects and to distinguish them based on their color, shape, size, and texture. Visual perception is mediated by the retina, a light sensitive layer of highly-specialized cells that lines the back of the

eye. Light perception is a multi-step process which involves transmission and refraction of light rays, allowing these signals to reach the retina, the conversion of light energy into electrical signals and the refinement of these signals within the neural circuits of the retina and the brain.⁵ The eye and especially the retina are complex multi-layered structures, all designed to allow optimal image acquisition of our surroundings.

Anatomy of the eye

On the exterior side of the eyeball several anatomical structures can be distinguished: the pupil, the iris and the cornea.^{6,7} The pupil is a black-looking aperture that allows light rays to enter the eye. It is surrounded by the iris, a colored circular muscle that determines the color of the eyes. This circular muscle controls the size of the pupil, and the amount of light that enters the eye. The cornea, on the other hand, is a transparent external surface that covers both the pupil and the iris and functions as a powerful lens of the optical system.⁷

On the interior side, the eye is a fluid-filled sphere enclosed by three layers of tissues: an external, intermediate and internal layer (**Figure 1A**). The outermost layer consists of the sclera and the cornea. The sclera is a continuation of the cornea, and is a tough white fibrous tissue which forms a strong enclosure of the eye and provides protection for the inner layers. The intermediate tissue layer consists of a posterior (choroid) and an anterior part (iris and ciliary body). The choroid is the vascular wall of the eye, and provides nutrition for the sensory cells located in the inner tissue layer. The ciliary body is a ring of tissue that encircles the lens and has a muscular and vascular component that are required for adjustment of the refractive power of the lens and ciliary processes, respectively. The lens is a transparent body that is located behind the iris. Together with the cornea, the lens is responsible for refraction of the light that allows the creation of a sharp, focused image on the back of the eye. The muscles of the ciliary body act to flatten or thicken the lens, allowing focused visualization of objects observed at variable distances in a process called accommodation. The innermost tissue layer is the retina and is the sensory part of the eye. It contains the light-sensitive cells, called photoreceptors, and houses the biochemical processes that are initiated by light perception. This will ultimately lead to the transmission of signals via the optic nerve through the optic disc, directed to the visual cortex of the brain.⁵⁻⁷

Besides the different tissues, there are three different fluid compartments within the eyeball (**Figure 1A**). An anterior chamber, that is located between the cornea and the iris, a posterior chamber between the iris and the lens, and a vitreous chamber between the lens and the retina. The anterior and posterior chambers are filled with aqueous humor, whereas the vitreous chamber is filled with vitreous humor. Aqueous humor

is produced by the epithelial cells of the ciliary body and is a clear watery liquid that supplies nutrients to the cornea and the lens. This liquid has a high turnover and is replaced approximately twelve times a day. The vitreous humor is a thick gelatinous substance and makes up ~80% of the volume of the eye and is responsible for maintaining the shape of the eyeball. The vitreous humor contains phagocytic cells that remove unclarities that might interfere with light transmission.^{6,7} The different liquids, their contents, and the cell types located in the surrounding tissue layers all work together to allow optimal perception of light.

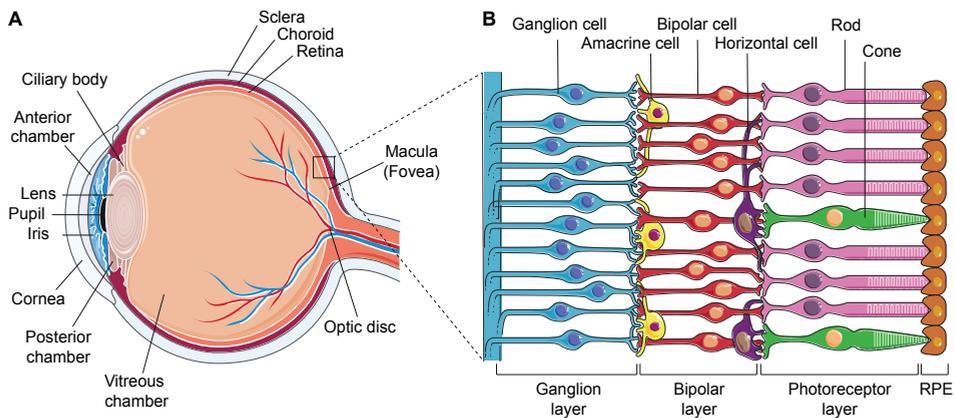


Figure 1. Anatomy of the human eye. (A) A schematic representation of the different cells and tissue types present within the eye. (B) The cellular organization of the human retina. The neuroretina contains five types of neurons that are divided over three nuclear layers. The photoreceptor layer contains the light sensitive rod and cone photoreceptor cells, the bipolar layer contains the bipolar cells and ganglion layer contains ganglion cells that will form the optic nerve. Communication between the different layers is supported by the horizontal and amacrine cells. RPE; retinal pigment epithelium. The figures are adapted and published with permission from Servier Medical Art.

The retina and photoreceptor cells

The sensory part of the eye, the retina, is part of the central nervous system and can be separated in the neural retina and the retinal pigment epithelium (RPE). The neural retina has a highly organized layered structure and contains five types of neurons that are organized in three nuclear layers and two synaptic layers (**Figure 1B**). The neurons are interconnected via synapses located in the inner and outer plexiform synaptic layers, the neuron cell bodies are divided over the three nuclear layers: the inner and the outer nuclear layer and the ganglion cell layer. The photoreceptor cells (1st layer) are the light-sensitive neuronal cells which communicate with the axons of bipolar cells (2nd layer), the terminals of the bipolar cells are connected to the synapses of the ganglion

cells (3rd layer). Finally, the larger axons of the ganglion cells form the optic nerve that will transmit the stimulus to the rest of the central nervous system. The horizontal neuron cells and the amacrine neuron cells enable the lateral interactions between the photoreceptor-bipolar cells and the bipolar-ganglion cells, respectively. The retinal cell types contribute in different ways to visual function, such as allowing a direction selective response.^{5,7}

There are two types of photoreceptors in the neural retina: the rod and cone photoreceptors (**Figure 1B**).⁸ Both types are composed of an outer segment, inner segment, cell body and a synaptic region. The outer segment consists of membranous discs that contain light-sensitive photopigments, and this is where the phototransduction processes take place. The inner segments house the metabolic machinery of the cell (the mitochondria, the endoplasmic reticulum, the Golgi complex, the lysosomes, and the ribosomes) that is responsible for protein synthesis and energy production. The outer and inner segment are connected by the connecting cilium. This structure plays a pivotal role in protein trafficking between both segments.⁹ Lastly, the cell body contains the nucleus of the photoreceptor cells, and the synaptic region contains the synaptic terminals through which the signals are propagated to the bipolar or horizontal cells.

The two photoreceptor cell types can be distinguished based on shape, type of bounded photopigment, and the distribution across the retina. Each cell type is specialized for a different aspect of vision. Rod photoreceptors allow light detection at extremely low luminance, but only with a low spatial resolution. Cone photoreceptors on the other hand, have a high spatial resolution but lack sensitivity for vision in low light conditions. Whereas one light photon is enough to initiate a stimulus in a rod photoreceptor, 100 light photons are required to activate a cone photoreceptor.⁵ Additionally, rods only mediate achromatic vision and cone photoreceptors are responsible for color vision. Cone photoreceptors can be divided in three types, that all contain a different type of photopigment that is able to absorb light with a different wavelength. They are referred to as short (S, "blue"), medium (M, "green") and long (L, "red") wavelength cones. Cones are highly concentrated at the macula, although present throughout the retina.⁷ The macula is located near the center of the retina and contains yellow pigment, and it supports high visual acuity. This acuity is highest at the center of the macula: the fovea, which can be observed as a small pit in the retina. Within the retina, there are 6 million cones and 120 million rod photoreceptors of which the latter are predominating in the mid-peripheral retina.⁸ The region of the optic disc does not contain any photoreceptors and is insensitive to light stimuli. It is referred to as the blind spot.⁵

The RPE layer surrounds the tips of the outer segments of the photoreceptors (**Figure 1B**). It has a key role in maintaining the structural integrity and normal function of the retina and is involved in a variety of functions: (1) it supports and nourishes the photoreceptor cells, and is firmly attached to the vascular layer (the choroid), (2) it contributes to the blood-retina barrier and controls the exchange of nutrients, growth factors and waste products, (3) it absorbs the excess of light that is projected on the retina, (4) it contributes to the regeneration of photopigment molecules after light absorption and (5) it plays a role in the phagocytosis of the outer membrane discs of the photoreceptors. The outer segments have a limited life span of ~12 days and are continuously renewed at their base. The oldest discs are shed from the tip and the RPE has a critical role in removing the expended discs.¹⁰

Phototransduction

Phototransduction is the biochemical process that is initiated after a photon is absorbed by the photopigment molecules in the photoreceptor cells. These photopigments contain a light-absorbing chromophore, retinal (a vitamin A derivative), which is coupled to an opsin protein. The different types of photoreceptor cells (rod, S-, M- and L-cones) differ in the type of opsin molecules. When retinal (i.e. 11-cis-retinal) absorbs a photon, it is converted to all-trans-retinal. The coupled opsin protein subsequently induces the activation of the intracellular messenger molecule transducin. The activation of transducin will trigger a signaling cascade, ultimately leading to the hydrolysis of cyclic guanosine monophosphate (cGMP). As a result, the cGMP concentration in the outer segments will be lowered, and the cGMP-gated cation channels in the outer segments will close. The subsequent reduction of inflow of cations will cause hyperpolarization of the cell membrane. This light-dependent change in photoreceptor membrane potential leads to lower neurotransmitter release at the cell's synaptic terminal.⁷ At dim light conditions, cGMP is not hydrolyzed and therefore the amount of open calcium channels is relatively high. Because of the high influx of cations (Na^+ and Ca^{2+}) in the outer segment discs, the photoreceptor cell is depolarized and neurotransmitter release is increased.

The signaling cascade is stopped when activated photopigment is rapidly phosphorylated by the responsible kinase, which leads to binding of the protein arrestin to the photopigment. This will block the ability of the photopigment to activate transducin. In this final stage of the retinoid acid cycle, restoration of the retinal molecule to its original configuration, 11-cis-retinal, takes place. All-trans-retinal dissociates from opsin and diffuses in the cytosol of the outer segment and is converted into all-

trans-retinol, transported to the RPE where it will be converted to 11-cis-retinal and transported back to the outer segment discs. Correct functioning of the retinoid cycle is crucial for maintaining light sensitivity.⁵

INHERITED RETINAL DYSTROPHIES

Inherited retinal dystrophies (RDs) are a group of clinically and genetically heterogeneous disorders that mainly involve the dysfunction or death of the photoreceptor and RPE cells.¹¹ Collectively, it is estimated that inherited RDs affect at least 2 million people worldwide.⁸ Clinically, RDs can be broadly subdivided in three categories based on the primarily affected cell type: (1) the rod photoreceptors (e.g. retinitis pigmentosa and choroideremia), (2) the cone photoreceptors (e.g. macular or central dystrophies) and (3) more generalized types of RDs that involve both photoreceptor types (e.g. cone-rod and rod-cone dystrophies and Leber congenital amaurosis). However in practice, this distinction is not always clear as RDs often display overlapping clinical features especially in later stages of disease. Additionally, RDs can be classified as stationary (e.g. congenital stationary night blindness and achromatopsia) or progressive (e.g. retinitis pigmentosa, cone-rod dystrophy, Stargardt disease). RD phenotypes can be involved in non-syndromic or syndromic forms of disease that affect multiple organs or tissues. Consequences of RDs are highly variable and range from mild retinal dysfunction to (congenital) legal blindness in the most severe cases.^{11,12}

Genetically, RDs are associated with pathogenic variants in more than 270 genes and can be inherited in an autosomal recessive, dominant, X-linked or digenic fashion.¹³ The RD-associated genes encode proteins that are involved in a multitude of biochemical processes. A gene ontology (GO) term-enrichment analysis was performed to visualize the biochemical processes that are most significantly enriched for RD-associated proteins (**Figure 2**). Most significantly enriched processes include sensory and visual perception of light stimuli and cilium assembly and organization.

Once a new potentially pathogenic variant is identified in a gene that has not been previously associated with RD, the interpretation and implication of this finding can be challenging. Variants within the same gene can be associated with different phenotypes, and the same phenotype can be caused by variants in different genes. Additionally, the phenotype can be affected by genetic modifiers as well as environmental factors, which can give rise to inter- and intrafamilial variability. The large genetic heterogeneity of RDs and the complexity of subsequent genetic analyses will be discussed in more detail in **chapter 1.2** of this thesis.

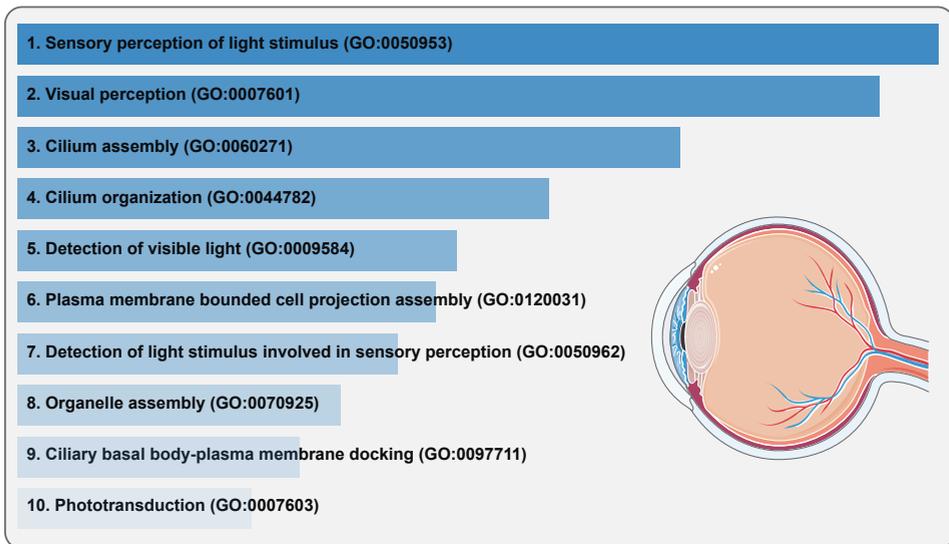


Figure 2. Biological processes linked to retinal dystrophy-associated genes. A gene ontology (GO) enrichment analysis was performed to visualize biochemical processes that are significantly associated with retinal dystrophies (RD). The top 10 of biochemical processes that show enrichment for RD-associated proteins are indicated. The length of the colored bars represents the association strength ($-\log_{10}(\text{p-value})$). GO-term enrichment analysis was performed using Enrichr¹⁴ and based on the GO biological processes term list (2018). RD-associated genes ($n=271$) were extracted from the RetNet website (May 2021).¹³

Non-syndromic retinitis pigmentosa

Retinitis pigmentosa (RP) is the most common RD, with an incidence of 1 in 4,000 individuals.¹⁵ It is caused by the progressive degeneration of the rod photoreceptor cells and the RPE. The classical RP phenotype involves night blindness in adolescence, followed by the loss of peripheral vision in young adulthood leading to tunnel vision or complete visual loss in some cases.¹⁶ In case of legal blindness, the peripheral vision is less than 20 degrees. However, as most inherited RDs, RP is a highly variable disorder and there are ample examples of individuals that remain asymptomatic until mid-adulthood.⁸

RP can be clinically diagnosed by performing imaging techniques and visual acuity measurements. Fundus photography, providing an image of the back of the eye, can be performed to check for the presence of characteristic RP-associated hallmarks: bone spicule pigmentation (predominantly midperipheral), attenuation of the retinal vessels and a waxy color of the optic nerve.¹⁵ Fundus autofluorescent (FAF) imaging can, among other characteristics, reveal a hyperautofluorescent ring that represents the transition zone between normal and abnormal retinal tissue. Optical coherence tomography

(OCT) is an imaging technique that allows the investigation of the (dis)organization of the retinal layers.¹⁵ In RP, the photoreceptor outer nuclear layer is severely attenuated, whereas the inner nuclear layer and the ganglion cell layer remain fairly well preserved until the late stages of disease.⁸ Additional examinations that can be performed to measure visual acuity include specialized charts to determine color vision, dark adaptation thresholds or contrast sensitivity.⁸ Electroretinography (ERG) on the other hand provides a more quantitative measure of visual acuity as it measures the electrical response of the retina to flashes of light recorded using an electrode contact lens or an electrode applied to the eye lid. Using dark-adapted dim light (scotopic) or light-adapted bright light (photopic) flashes that stimulate a cone- or rod-driven response, respectively, dysfunctionality of specific photoreceptor types can be distinguished.¹⁵

To date, there are 89 genes associated with non-syndromic RP.¹³ RP can be mostly explained by monogenic causes, but there are few cases described with a digenic or mitochondrial inheritance. Additionally, six loci have been described for which the genetic defects are still elusive. In 30-40% of the cases, RP is dominantly inherited, 50-60% recessively and 5-15% of all cases has an X-linked mode of inheritance. Frequently mutated genes are *RHO* (25% of dominant RP cases), *USH2A* (5-10% of recessive RP cases) and *RPGR* (80% of X-linked cases).^{11,15,17} Together these three genes genetically explain one-fifth of all non-syndromic RP cases.

Syndromic retinitis pigmentosa

In 20-30% of the cases, RP coexists with non-ocular symptoms.⁸ Usher syndrome is the most frequent syndromic form of RP. In this syndrome, RP is accompanied with HL and in part of the cases with vestibular dysfunction. Three clinical types of Usher syndrome can be distinguished based on the severity and age of onset of HL and the presence or absence of vestibular symptoms. HL can be profound and congenital, and combined with a vestibular dysfunction (Usher syndrome type I), moderate to severe and non-progressive, without clear vestibular dysfunction (type II) or progressive with a postlingual onset and a variable vestibular phenotype (type III).^{8,18} More recently, reports have appeared that describe atypical forms of Usher syndrome that do not fit within one of these categories. Examples of genes that are associated with atypical forms of Usher syndrome are *ARSG*¹⁹, *CEP250*²⁰ and *CEP78*^{21,22}. Pathogenic variants in these genes have only been reported in a small number of cases, and additional genotype-phenotype studies are required to describe the associated phenotypes in more detail. Other syndromic forms of RP include Bardet-Biedl syndrome, Joubert syndrome and Alström syndrome, which are ciliopathies that originate from defects in ciliary function, and involve multiple organ dysfunction.¹¹

ANATOMY AND PHYSIOLOGY OF THE EAR

The sense of hearing is an essential component of the human communication system. Well-functioning of the ear is critical for a person's well-being and defects can be socially debilitating and even lead to anxiety, cognitive decline and depression.^{2,23,24} The ear can be divided into three anatomical compartments: the outer, middle and the inner ear. Under normal circumstances, sound waves are captured by the outer ears, conducted through the middle ear in which they are amplified or attenuated (in case of excessive noise), and converted into neuronal activity by complex biochemical processes that take place in the inner ear. The inner ear also harbors the vestibular system that is essential for our sense of balance.

The outer and middle ear

The outer ear consists of the pinna (auricle) and the external ear canal (auditory meatus) (**Figure 3A**). Sound waves are collected by the pinna, and travel via the external ear canal to the middle ear. The external ear canal directs the waves to the tympanic membrane (eardrum) that separates the outer and the middle ear and starts vibrating upon capture of sound waves.²⁵

The middle ear is an air-filled cavity and forms a bridge between the outer ear and the fluid-filled inner ear. It contains the tympanic membrane that is connected with the oval window via three interconnected auditory ossicles (malleus, incus and stapes) (**Figure 3A**).²⁵ The main function of the middle ear is to efficiently transmit the sound wave energy across the air-fluid boundary of the outer and inner ear. Vibrations of the tympanic membrane are transmitted by the ossicles, in a pistol-like movement, to the oval window. The resulting inward movement of the oval window leads to displacement of the fluids that are present in the inner ear.^{25,26}

The inner ear: Cochlea

The inner ear contains the sensory organs that enable hearing (cochlea) and importantly contribute to the control of our balance and spatial orientation (vestibular system) (**Figure 3A**). The inner ear is embedded in the temporal bone, the strongest bone of the skull, which serves as protection for the delicate organ.

The cochlea is the auditory part of the inner ear and is a small, coiled structure. The cochlea consists of three fluid-filled chambers that run from base to apex: the scala media, the scala tympani and the scala vestibuli (**Figure 3B**). The fluid movements induced by pulses of the oval window and consequently, bulging of the round window

lead to deformation of the cochlear duct.²⁵ The traveling sound wave propagates from the base towards the apex of the cochlea, until maximum displacement is reached. The site of maximum displacement depends on the sound frequency; higher frequencies only reach the stiffer base of the cochlea while low frequency sounds are perceived in the apex. This is called the tonotopy of the cochlea and serves as a mechanical frequency analyzer.²⁵

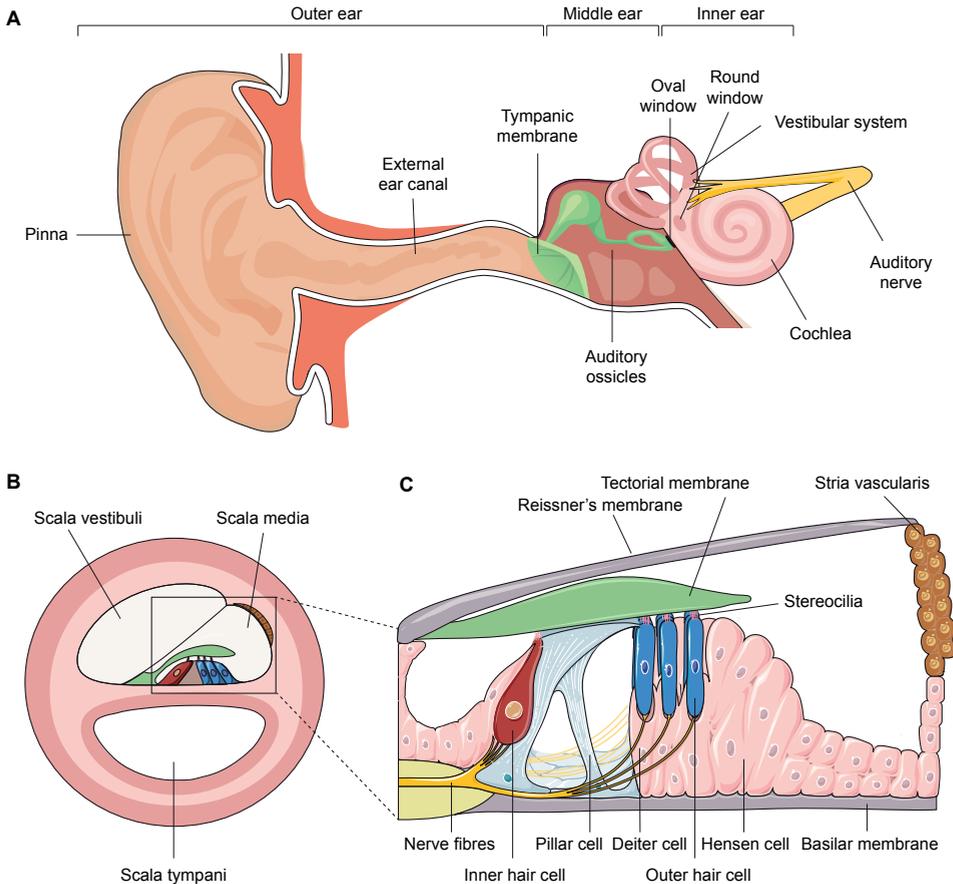


Figure 3. Schematic representation of the human ear. (A) An overview of the three anatomical compartments of the ear: the outer, middle and inner. **(B)** Intersection of the cochlea that contains three fluid-filled chambers: the scala media, the scala tympani and the scala vestibuli. The scala media is filled with potassium-rich endolymph fluid, the scala tympani and vestibuli are filled with sodium-rich perilymph fluid. The scala media houses the sensory epithelium of the inner ear: the organ of Corti. **(C)** The cellular structures of the organ of Corti. The epithelium contains one row of inner hair cells and three rows of outer hair cells. The hair cells are separated and flanked by different types of supporting cells (Deiter's, pillar, Hensen's, inner phalangeal and border cells). The figures are adapted and published with permission from Adobe Stock and Servier Medical Art.

The cochlear chambers are separated from each other by the Reissner's membrane and the basilar membrane which are essential for maintenance of the endocochlear potential. The scala media is filled with endolymph fluid (potassium rich), the scala tympani and vestibuli are filled with perilymph fluid (sodium rich). The stria vascularis, that is part of the lateral wall of the cochlear duct, is responsible for the secretion of a constant flow of potassium ions into the endolymph and maintaining its ion homeostasis. The scala vestibuli and scala tympani are joined via an opening (helicotrema) positioned at the apical end of these compartments, allowing their fluids to mix. The oval window is located at the basal part of the scala vestibuli (upper chamber), the round window is located at the basal part of the scala tympani.²⁵ The organ of Corti is the sensory epithelium of the inner ear and resides within the scala media (**Figure 3C**).

The epithelium consists of one row of inner hair cells and three rows of outer hair cells which are separated by supporting cells. The hair cells function as sensory receptors; the outer hair cells enhance sound sensitivity and selectivity whereas inner hair cells are responsible for signal transmission to the afferent nerve terminals of the cochlear nerve.²⁶ Actin-rich structures protrude from the apical surface of the hair cells: the stereocilia. These are anchored in the cuticular plate. Stereocilia are organized in a V-shape staircase-like pattern and are connected by horizontal links and tip-links.^{27,28} The kinocilium is the only true microtubule-based primary cilium and located lateral to the row of largest stereocilia. The kinocilium elongates and determines the orientation of the hair bundle during development, and disappears in mammals before or shortly after birth depending on the species.²⁸⁻³⁰ The different types of supporting cells that are present in the organ of Corti have a pivotal role in the development of the sensory epithelium and preserve its structural integrity and homeostasis.²⁹

Sensory hair cells are located between the basilar membrane and the tectorial membrane in the cochlea. Sound-evoked fluid movements trigger vertical (basilar membrane) or horizontal (tectorial membrane) displacement of these membranes. The combination of movements results in bending of the stereocilia. The displacement induces the opening of cation-selective mechanotransduction channels located at the lower end of tip links, allowing influx or efflux of potassium ions from the endolymph. Depending on the bending direction of the stereocilia, the hair cell will depolarize or hyperpolarize. As a result, voltage-gated calcium channels will open in the cell soma, allowing calcium-dependent neurotransmitter release at the synapses of the hair cells. An action potential is generated within the auditory nerve and a stimulus will be transmitted to the auditory cortex in the brain.²⁵ Upon repolarization of the hair cell, ion

pumps, ion channels and gap junctions located in the hair cells, the supporting cells, the lateral wall of the scala media and the stria vascularis are responsible for recycling of the potassium ions back to the endolymph.³¹

The inner ear: Vestibular system

The vestibular part of the inner ear, the vestibular labyrinth, includes three semicircular canals and two otolith organs (utricle and saccule) (**Figure 3A**). The semicircular canals allow detection of rotational movements whereas the otolith organs can sense linear accelerations. Similar to the cochlea, vestibular structures are filled with endolymph or perilymph fluids and contain supporting cells and vestibular hair cells with stereocilia at their apical surface. The hair cells are present in the otolith organs and in the ampullae, that are located at the base of the semicircular canals. The two otolith organs are responsible for detection of head displacement and linear accelerations. A gelatinous layer (the otolithic membrane) that contains calcium carbonate crystals (otoconia), the so-called ear stones, is overlying the hair cells. Gravity can cause the otolithic membrane to shift towards the sensory epithelium, this shearing motion leads to displacement of the hair bundles. The semicircular canals are filled with endolymph fluid. The stereocilia of the hair cells that reside in the ampullae of these canals are deflected due to fluid movements caused by head rotations.^{32,33} As in the cochlea, stereocilia deflection leads to hair cell depolarization that induces neurotransmitter release to the vestibular nerve.³²

HEARING LOSS

According to the World Health Organization (WHO), 460 million people suffer from disabling hearing loss (HL), among which 34 million children.³⁴ Based on these numbers, it is the most frequent sensory disorder worldwide, which has enormous negative consequences for society (economical challenges) in addition to those on the personal level (educational and social disadvantages). In most developed countries, newborn hearing screening programs have been implemented that allow the earliest possible identification of HL and intervention.^{35,36} Disabling HL is defined as a loss that is greater than 40 dB in the better hearing ear in adults, or greater than 30 dB in children. Several different causes can be underlying that include both acquired and congenital causes. Early detection and treatment or rehabilitation of HL is a key element in effective disease management. It is of benefit for the patients as it is essential to prevent and adverse the impact of HL.³ Nevertheless, numbers suggest that HL is still significantly undertreated.³⁷ Although no cure is currently available for sensorineural HL, hearing aids and cochlear implants can mitigate HL. Several studies have shown that the use of cochlear implants consistently leads to improvement of quality of life (reviewed in (37)).

Clinical characteristics of hearing loss

HL can be broadly categorized into three different types: conductive, sensorineural and mixed.³⁸ Conductive HL results from abnormalities in the outer or middle ear, such as obstructions or damaged structures due to trauma. In this type of HL, the efficiency of sound transmission to the inner ear is lowered. In case of sensorineural HL, the problem resides within inner ear structures or the auditory nerve, which impairs the conversion of sound to a signal that is detected in the auditory cortex of the brain.²⁵ In mixed HL, there is a combination of conductive and sensorineural HL.³⁸ Additionally, there are several other clinical characteristics that can be used to classify and describe the type of HL including age of onset, the sound frequencies for which thresholds are elevated, severity and (a)symmetry, which are summarized in **Table 1**.

Clinical evaluation

Different tests can be performed to determine the hearing status of an individual. These assessments can be divided into physical, electrophysiological and audiometric tests. Ideally, a combination is used to allow comparison and validation of the results.⁴ Often, first a general physical otoscopic examination will be performed to evaluate the condition of the ear canal, the tympanic membrane and the middle ear. In some instances, additionally a basic balance test (e.g. head-impulse test), and CT or MRI will be performed. Electrophysiological tests provide an objective measure of the functional status of an individual's auditory system. They include measurements of otoacoustic emissions (sounds generated by the outer hair cells) and auditory brain stem responses.^{4,38}

Audiometric tests, on the other hand, provide a more subjective measure of the hearing status of an individual. The most widely applied test is pure-tone audiometry (PTA), which determines the sensitivity of the auditory system for different sound frequencies. Pure-tone thresholds indicate the lowest sound intensity that can be perceived by an individual when seated in a sound attenuating booth.^{4,38} Both air conduction audiometry (sound is presented via earphones), and bone conduction audiometry (sound is presented through a vibrating device placed on the skull) which bypasses the outer and middle ear, can be performed. This allows the distinction between conductive and sensorineural HL.³⁸ After completion of a PTA, hearing sensitivity can be plotted on an audiogram which displays sound intensity (decibel hearing level (dB HL)) as a function of sound frequency (hertz (Hz)). Besides PTA, also word-recognition testing (speech audiometry test) can be performed.^{37,38}

Table 1. Clinical characteristics and classification of hearing loss

Characteristic	Class	Definition
Type <i>Thresholds averaged over 0.5, 1 and 2 kHz</i>	Conductive	HL resulting from abnormalities in outer or middle ear structures, normal bone-conduction thresholds (<20 dB HL), air-bone gap >15 dB HL
	Sensorineural	HL resulting from abnormalities in inner ear structures, air-bone gap <15 dB HL
	Mixed	HL with combinatory involvement of conductive and sensorineural HL, >20 dB HL bone-conduction threshold, >15 dB HL air-bone gap
Severity <i>Criteria should be applied to the better hearing ear, thresholds averaged over 0.5, 1, 2 and 4 kHz</i>	Mild	20-40 dB HL
	Moderate	41-70 dB HL
	Severe	71-95 dB HL
	Profound	>95 dB HL
Audiometric configuration	Low frequency, ascending	>15 dB HL difference between the poorer low frequency thresholds and those for the higher frequencies
	Mid frequency, U-shaped	>15 dB HL difference between the poorest thresholds in the mid-frequencies, and those at higher and lower frequencies
	High frequency, gently sloping	15-29 dB HL difference between the mean of 0.5 and 1 kHz and the mean of 4 and 8 kHz.
	High frequency, steeply sloping	>30 dB HL difference between the above frequencies
	Flat	<15 dB HL difference between the mean of 0.25 and 0.5 kHz thresholds, the mean of 1 and 2 kHz and the mean of 4 and 8 kHz
Frequency ranges	Low	≤ 0.5 kHz
	Mid	>0.5 kHz, ≤2 kHz
	High	>2 kHz, ≤8 kHz
	Extended high	>8 kHz
Symmetry of HI	Symmetrical	Bilateral HL, <10 dB HL difference between the ears
	Asymmetrical	Unilateral or bilateral HL, asymmetrical HL if >10 dB HL difference between the ears in at least two frequencies exists
Age of onset	Congenital/prelingual	Onset of HL before speech develops
	Postlingual	Onset of HL after speech developed
Progression <i>Averaged over 0.5, 1 and 2 kHz</i>	Progressive	Deterioration of >15 dB HL
	Non-progressive	Deterioration of ≤15 dB HL
Additional symptoms	Tinnitus	Presence of e.g. low or high tone pitch noise
	Vestibular symptoms	Presence of vestibular problems and/or abnormal vestibular function

Classifications and definitions of audiological phenotypes as recommended and described by the GENDEAF study group.⁴⁰ kHz, kilohertz; HL, hearing loss; dB, decibel.

Causes of hearing loss

HL is an etiologically heterogeneous trait as it involves both environmental and genetic causes. Environmental causes include complications at birth, infections, ototoxic drugs, head trauma and excessive noise exposure. In developed countries, genetic defects are estimated to underlie 50% and 60% of HL cases with a congenital onset or an onset at the age of 4 years, respectively.³⁶ For adult-onset HL these percentages are unknown, although it is generally accepted that the contribution of strong genetic factors decreases with age and that there is an increased exposure to damaging environmental factors during life. Consequently, adult-onset HL is generally considered a multifactorial condition, in contrast to congenital HL which is often either monogenic or caused by a non-genetic factor.^{36,39} Inherited HL can display an autosomal dominant (type DFNA), autosomal recessive (DFNB), mitochondrial or X-linked (DFNX) inheritance pattern. DFNB concerns 80% of the cases with prelingual (non-syndromic) HL, DFNA 19% and mitochondrial or X-linked inherited HL concern less than 1% of the cases.³⁸

Besides non-syndromic forms of HL, also many syndromic forms (>300 (OMIM⁴¹)) have been described, in which HL is associated with distinctive additional clinical features (such as RD in Usher syndrome).²⁶ The majority of cases however, are non-syndromic (~70% of cases with inherited prelingual HL).³⁸

Similar to RD, monogenic, non-syndromic and sensorineural HL displays a high genetic heterogeneity. The condition has been associated with pathogenic variants in >150 genes.⁴² Despite this large number of genes, up to 50% of all presumed genetic cases with early-onset HL can be attributed to pathogenic variants in the *GJB2* gene.⁴³ In the Netherlands, this is true for ~10% of such cases.⁴³ Variants in this gene are associated with profound non-syndromic, recessively-inherited HL, but the associated phenotype is highly variable.^{4,44} *GJB2* encodes the gap junction beta-2 protein, also known as connexin 26 (Cx26), that functions in gap junctions connecting the supporting cells in the organ of Corti, the mesenchymal cells in the stria vascularis and fibrocytes, located in the spiral ligament that underlies the stria vascularis.^{45,46} The gap junctions between these cell types are involved in the potassium-recycling pathway of the cochlea and *GJB2* is therefore essential for maintenance of the ion homeostasis.²⁶

Besides ion homeostasis, HL-associated proteins are involved in several other biological processes. Similarly as performed for RD, a GO-enrichment analysis was performed to visualize the most important biological processes that are enriched with HL-associated proteins (**Figure 4**). The most significantly enriched processes for HL-associated proteins are those that entail the sensory perception of mechanical stimuli and sound. Interesting to note is that the second-most enriched process concerns “the sensory perception of

light stimuli”, which indicates there are numerous proteins that play crucial roles in the sensory cells of both the retina as well as the inner ear. The majority of the proteins that are linked to this category consider proteins that have been linked to Usher syndrome (e.g. cadherin 23, myosin VIIa, and harmonin). Other biological processes enriched for HL-associated proteins include hair cell bundle morphogenesis, cochlear ion homeostasis, extracellular matrix formation, mechanostability, and maintenance and gene regulation.⁴⁷

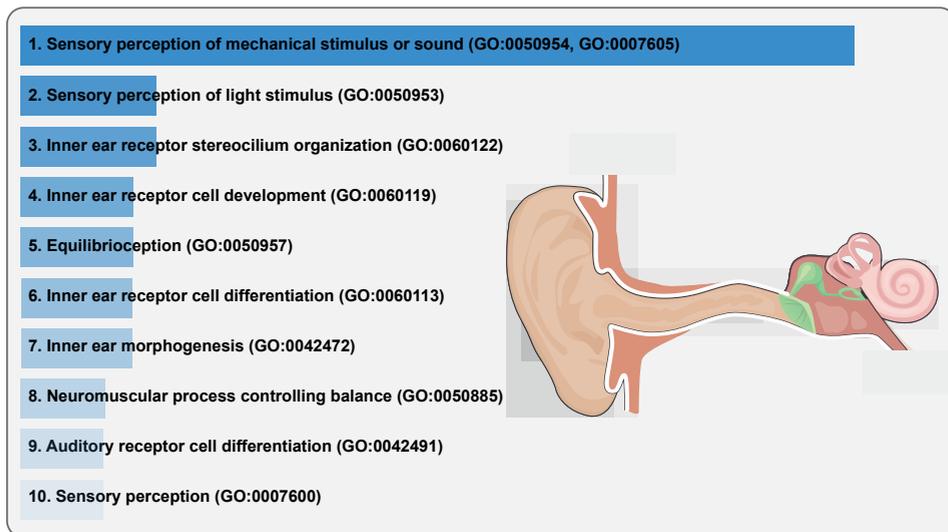


Figure 4. Biological processes linked to hearing loss-associated genes. A gene ontology (GO) enrichment analysis was performed to visualize biochemical processes that are significantly associated with hearing loss (HL). The top 10 of biochemical processes that show enrichment for HL-associated proteins has been indicated. The length of the colored bars represents the association strength ($-\log_{10}(p\text{-value})$). GO-term enrichment analysis was performed using Enrichr¹⁴ and based on the GO biological processes term list (2018). HL-associated genes ($n=154$) were extracted from the Hereditary Hearing Loss Homepage (May 2021).

In diagnostic practices, genetic prescreening approaches are currently still preferred before continuing with performing exome or genome sequencing. These prescreening approaches can involve screening for pathogenic variants in the *GJB2* gene only, or other candidate genes in case of a strong phenotype-genotype correlation. For instance, variants in the *SLC26A4* gene are associated with recessive HL and a unilateral or bilateral enlarged vestibular aqueduct. Also, some genetic forms of HL can be distinguished based on audiogram shape. Age Related Typical Audiograms (ARTA) have been developed, in which the latest pure tone thresholds of individuals are analyzed cross-sectionally

and plotted in a decade audiogram, illustrating hearing threshold development over time. This could serve as a phenotype fingerprint, especially for specific HL types with a high phenotype-genotype correlation.⁴⁸ Additionally, machine-learning efforts have been performed, in order to develop an algorithm that is able to predict the underlying cause based on an individual's audiogram. Using the machine-learning tool Audiogene, accuracies of up to 68% can be reached for dominantly inherited HL, which indicates there is still room for improvement.⁴⁹ Considering the extremely high genetic but also clinical heterogeneity of HL, the identification of a causative variant is still a complicated task. The different aspects of disease variant identification and interpretation will be discussed in more detail in **chapter 1.2**.

Despite all technological advancements and insights that are gained during thirty years of research directed towards the identification of disease-associated genes and variant interpretation, still a significant number of cases remain genetically unexplained for both inherited HL and RD. With this, it is clear that still many challenges are ahead of us that need to be overcome in the following years to finally be able to completely “*solve the unsolved*”.

OUTLINE AND AIMS OF THIS THESIS

Inherited RD and HL are both characterized by a large clinical and genetic heterogeneity. Despite the large number of genes that have been associated with these disorders (>270 and >150, respectively), literature reports still indicate a significant missing heritability. The research described in this thesis aimed to shed light on this missing heritability, and to solve the unsolved cases by combining established methods, implementing novel approaches and sometimes by thinking “outside the box”. This has led to several important findings and relevant contributions to the field that are described in the different chapters of this thesis.

Chapter 1.2 provides a comprehensive overview of available technologies that contributed over the years to the identification of RD and HL-associated genes and can be applied in modern diagnostics. Additionally, the step-wise approach of disease variant identification and interpretation is described, with a special focus on inherited RD and HL.

Chapter 2 describes the confirmation of a candidate gene for RD. Pathogenic variants in the *KIAA1549* gene have been detected using a combined approach involving both

whole exome sequencing (WES) and homozygosity mapping. RNA expression analyses and immunohistochemistry were performed to provide additional evidence for the pathogenicity of *KIAA1549* variants.

Chapter 3 focuses on the identification of the genetic defect and the development of a genetic therapy for adult-onset HL type DFNA21. The identification of the genetic defect using WES is described in **chapter 3.1**. The defect is an in-frame deletion in the *RIPOR2* gene and is positioned 0.9 Mb centromeric of the DFNA21 locus, a locus that was identified using linkage analysis over twenty years ago. The *RIPOR2* variant is associated with adult-onset dominantly HL, and presumably the most frequent cause of this type of HL in Northwest Europe. Because of this high prevalence, the first steps towards the development of an allele-specific therapy for *RIPOR2*-associated HL were initiated, which is described in **chapter 3.2**.

One of the limitations of WES as compared to whole genome sequencing (WGS) is the detection of structural variants (SV). In **chapter 4**, we describe the successful application of WGS leading to the identification of eight unique SVs in 22 families affected with autosomal dominant RP. These SVs show overlap with the previously described genetically unresolved RP17 locus. Epigenetic and chromatin analyses were performed to elucidate a novel mechanism of disease: the ectopic expression of the *GDPD1* gene.

In **chapter 5**, a phenotype-genotype correlation study was performed to explain the relatively high missing heritability for *SLC26A4*-associated HL. Sequencing technologies that were performed include both short and long-read WGS. Important clues for the missing heritability were identified, including a commonly shared disease haplotype and indications for possible digenic inheritance.

Finally, **chapter 6** provides the general discussion and, future perspectives whereas **chapter 7** provides the summary of this thesis.

REFERENCES

1. Keller, B.K., Morton, J.L., Thomas, V.S. & Potter, J.F. The effect of visual and hearing impairments on functional status. *Journal of the American Geriatrics Society* **47**, 1319-1325 (1999).
2. Mener, D.J., Betz, J., Genther, D.J., Chen, D. & Lin, F.R. Hearing loss and depression in older adults. *Journal of the American Geriatrics Society* **61**, 1627-1629 (2013).
3. Arlinger, S. Negative consequences of uncorrected hearing loss--a review. *International Journal of Audiology* **42**, 2S17-2S20 (2003).
4. Korver, A.M., Smith, R.J., Van Camp, G., Schleiss, M.R., Bitner-Glindzicz, M.A., Lustig, L.R. et al. Congenital hearing loss. *Nature reviews. Disease primers* **3**, 16094 (2017).
5. Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., Lamantia, A. & White, L. Vision: the eye. in *Neuroscience* (eds. Purves, D. & Augustine, G.) 229-256 (Sinauer Associates Inc, 2012).
6. Snell, R.S. & Lemp, M.A. The eyeball. in *Clinical Anatomy of the Eye* 132-213 (1997).
7. Kolb, H. Gross anatomy of the eye. in *Webvision: The Organization of the Retina and Visual System* (eds. Kolb, H., Fernandez, E. & Nelson, R.) (University of Utah Health Sciences Center, 1995).
8. Hartong, D.T., Berson, E.L. & Dryja, T.P. Retinitis pigmentosa. *Lancet* **368**, 1795-1809 (2006).
9. Wheway, G., Parry, D.A. & Johnson, C.A. The role of primary cilia in the development and disease of the retina. *Organogenesis* **10**, 69-85 (2014).
10. Strauss, O. The retinal pigment epithelium in visual function. *Physiological Reviews* **85**, 845-881 (2005).
11. Berger, W., KloECKener-Gruissem, B. & Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Progress in Retinal and Eye Research* **29**, 335-375 (2010).
12. den Hollander, A.I., Black, A., Bennett, J. & Cremers, F.P.M. Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. *The Journal of Clinical Investigation* **120**, 3042-3053 (2010).
13. RetNet. Available from: <https://sph.uth.edu/RetNet/>.
14. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
15. Verbakel, S.K., van Huet, R.A.C., Boon, C.J.F., den Hollander, A.I., Collin, R.W.J., Klaver, C.C.W. et al. Non-syndromic retinitis pigmentosa. *Progress in Retinal and Eye Research* **66**, 157-186 (2018).
16. Hamel, C. Retinitis pigmentosa. *Orphanet Journal of Rare Diseases* **1**, 40-40 (2006).
17. Stone, E.M., Andorf, J.L., Whitmore, S.S., DeLuca, A.P., Giacalone, J.C., Streb, L.M. et al. Clinically focused molecular investigation of 1000 consecutive families with inherited retinal disease. *Ophthalmology* **124**, 1314-1331 (2017).
18. Toulbi, L., Toms, M. & Moosajee, M. USH2A-retinopathy: from genetics to therapeutics. *Experimental Eye Research* **201**, 108330 (2020).
19. Khateb, S., Kowalewski, B., Bedoni, N., Damme, M., Pollack, N., Saada, A. et al. A homozygous founder missense variant in arylsulfatase G abolishes its enzymatic activity causing atypical Usher syndrome in humans. *Genetics in Medicine* **20**, 1004-1012 (2018).

20. Khateb, S., Zelinger, L., Mizrahi-Meissonnier, L., Ayuso, C., Koenekoop, R.K., Laxer, U. *et al.* A homozygous nonsense CEP250 mutation combined with a heterozygous nonsense C2orf71 mutation is associated with atypical Usher syndrome. *Journal of Medical Genetics* **51**, 460-469 (2014).
21. Nikopoulos, K., Farinelli, P., Giangreco, B., Tsika, C., Royer-Bertrand, B., Mbefo, M.K. *et al.* Mutations in CEP78 cause cone-rod dystrophy and hearing loss associated with primary-cilia defects. *American Journal of Human Genetics* **99**, 770-776 (2016).
22. Namburi, P., Ratnapriya, R., Khateb, S., Lazar, C.H., Kinarty, Y., Obolensky, A. *et al.* Bi-allelic truncating mutations in CEP78, encoding centrosomal protein 78, cause cone-rod degeneration with sensorineural hearing loss. *American Journal of Human Genetics* **99**, 777-784 (2016).
23. Gurgel, R.K., Ward, P.D., Schwartz, S., Norton, M.C., Foster, N.L. & Tschanz, J.T. Relationship of hearing loss and dementia: a prospective, population-based study. *Otology & Neurotology* **35**, 775-781 (2014).
24. Lin, F.R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T.B., Purchase-Helzner, E. *et al.* Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine* **173**, 293-299 (2013).
25. Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., Lamantia, A. & White, L. The auditory system. in *Neuroscience* (eds. Purves, D. & Augustine, G.) 278-302 (Sinauer Associates Inc, 2012).
26. Dror, A.A. & Avraham, K.B. Hearing loss: mechanisms revealed by genetics and cell biology. *Annual Review of Genetics* **43**, 411-437 (2009).
27. Zhao, B., Wu, Z. & Müller, U. Murine Fam65b forms ring-like structures at the base of stereocilia critical for mechanosensory hair cell function. *Elife* **5**, e14222 (2016).
28. Goodyear, R.J., Marcotti, W., Kros, C.J. & Richardson, G.P. Development and properties of stereociliary link types in hair cells of the mouse cochlea. *Journal of Comparative Neurology* **485**, 75-85 (2005).
29. Forge, A. & Wright, T. The molecular architecture of the inner ear. *British Medical Bulletin* **63**, 5-24 (2002).
30. Barr-Gillespie, P.-G. Assembly of hair bundles, an amazing problem for cell biology. *Molecular Biology of the Cell* **26**, 2727-2732 (2015).
31. Mittal, R., Nguyen, D., Patel, A.P., Debs, L.H., Mittal, J., Yan, D. *et al.* Recent advancements in the regeneration of auditory hair cells and hearing restoration. *Frontiers in Molecular Neuroscience* **10**, 1-16 (2017).
32. Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., Lamantia, A. & White, L. The vestibular system. in *Neuroscience* (eds. Purves, D. & Augustine, G.) 303-320 (Sinauer Associates Inc, 2012).
33. Kingma, H. & van de Berg, R. Anatomy, physiology, and physics of the peripheral vestibular system. *Handbook of Clinical Neurology* **137**, 1-16 (2016).
34. World Health Organisation. Deafness and hearing loss: Key facts 2021 update. (Geneva: World Health Organization, 2021).
35. McPherson, B. Newborn hearing screening in developing countries: needs & new directions. *Indian Journal of Medical Research* **135**, 152-153 (2012).

36. Morton, C.C. & Nance, W.E. Newborn hearing screening--a silent revolution. *New England Journal of Medicine* **354**, 2151-2164 (2006).
37. Carlson, M.L. Cochlear implantation in adults. *New England Journal of Medicine* **382**, 1531-1542 (2020).
38. Shearer, A.E., Hildebrand, M.S. & Smith, R.J.H. Hereditary hearing loss and deafness overview. in *GeneReviews* (eds. Adam, M.P. *et al.*) (University of Washington, 2017).
39. Tu, N.C. & Friedman, R.A. Age-related hearing loss: unraveling the pieces. *Laryngoscope Investigative Otolaryngology* **3**, 68-72 (2018).
40. M, M., Camp, G.U.Y., Newton, V., N, G., Declau, F. & Parving, A. Recommendations for the description of genetic and audiological data for families with nonsyndromic hereditary hearing impairment. *Audiological Medicine* **1**, 148-150 (2009).
41. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research* **43**, D789-D798 (2015).
42. Van Camp, G. & Smith, R. Hereditary Hearing Loss Homepage. Available from: <https://hereditaryhearingloss.org/>.
43. Chan, D.K. & Chang, K.W. GJB2-associated hearing loss: systematic review of worldwide prevalence, genotype, and auditory phenotype. *Laryngoscope* **124**, E34-E53 (2014).
44. Sloan-Heggen, C.M., Bierer, A.O., Shearer, A.E., Kolbe, D.L., Nishimura, C.J., Frees, K.L. *et al.* Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Human Genetics* **135**, 441-450 (2016).
45. Beach, R., Abitbol, J.M., Allman, B.L., Esseltine, J.L., Shao, Q. & Laird, D.W. GJB2 mutations linked to hearing loss exhibit differential trafficking and functional defects as revealed in cochlear-relevant cells. *Frontiers in Cell and Developmental Biology* **8**, 215 (2020).
46. Liu, W., Boström, M., Kinnefors, A. & Rask-Andersen, H. Unique expression of connexins in the human cochlea. *Hearing Research* **250**, 55-62 (2009).
47. Hilgert, N., Smith, R.J.H. & Van Camp, G. Function and expression pattern of nonsyndromic deafness genes. *Current Molecular Medicine* **9**, 546-564 (2009).
48. Huygen, P.L., Pennings, R.J. & Cremers, C.W. Characterizing and distinguishing progressive phenotypes in nonsyndromic autosomal dominant hearing impairment. *Audiological Medicine* **1**, 37-46 (2003).
49. Hildebrand, M.S., DeLuca, A.P., Taylor, K.R., Hoskinson, D.P., Hur, I.A., Tack, D. *et al.* A contemporary review of AudioGene audioprofiling: a machine-based candidate gene prediction tool for autosomal dominant nonsyndromic hearing loss. *Laryngoscope* **119**, 2211-2215 (2009).

CHAPTER 1

INTRODUCTION



Chapter 1.2

The impact of modern technologies on molecular diagnostic success rates with a focus on inherited retinal dystrophy and hearing loss

Suzanne E. de Bruijn*, Zeinab Fadaie*, Frans P.M. Cremers, Hannie Kremer, Susanne Roosing

*These authors contributed equally to this work.

International Journal of Molecular Sciences **22**, 2943 (2021)

ABSTRACT

The identification of pathogenic variants in monogenic diseases has been of interest to researchers and clinicians for several decades. However, for inherited diseases with extremely high genetic heterogeneity such as hearing loss and retinal dystrophies, establishing a molecular diagnosis requires an enormous effort. In this review, we use these two genetic conditions as examples to describe the initial molecular genetic identification approaches, as performed since the early 90s, and subsequent improvements and refinements introduced over the years. Next, the history of DNA sequencing from conventional Sanger sequencing to high-throughput massive parallel sequencing, a.k.a. next-generation sequencing, is outlined, including their advantages and limitations, and their impact on identifying the remaining genetic defects. Moreover, the development of recent technologies, also coined “third generation” sequencing, is reviewed, which holds the promise to overcome these limitations. Furthermore, we outline the importance and complexity of variant interpretation in clinical diagnostic settings concerning the massive number of different variants identified by these methods. Finally, we briefly mention the development of novel approaches such as optical genome mapping and multi-omics, which can help to further identify genetic defects in the near future.

INTRODUCTION

In previous decades, different methods for disease gene identification have been established and successfully employed. All these technologies have significantly contributed to identifying the large number of genes that are associated with inherited forms of hearing loss (HL) (>150 genes)¹ and retinal dystrophy (RD) (>270 genes).² Especially after the introduction next generation sequencing (NGS) techniques, it was anticipated that soon all HL- and RD-associated genes would be identified. Nevertheless, the diagnostic yield suggests a significant portion of missing heritability, which can potentially be explained by unrecognized disease genes or missed variants.^{3,4} To provide a genetic diagnosis for all inherited cases, it has become evident that there is no single technique that can serve as the gold standard. To be able to detect and interpret all genetic variation of the human genome, classical methods such as linkage analysis or homozygosity mapping should be combined with novel state-of-the-art techniques.^{5,6}

The observed high genetic heterogeneity is not unique for these inherited sensory disorders; they have also been described for other inherited disorders including intellectual disability, ciliopathies and inherited susceptibility for cancer.⁷⁻⁹ Although in general, disease gene identification strategies applied in these fields rely on the same principles and have undergone a similar development, an optimal diagnostic strategy depends heavily on key factors such as evolutionary pressure and involvement of multifactorial versus monogenic causes. For example, for intellectual disability, de novo causes are more frequent due to a strong reduction of reproductive fitness; this impacts the optimal diagnostic strategy. This review focuses on the identification of monogenic causes of inherited HL and RD.

In this review, we aim to provide an overview of the development of techniques that have enabled disease gene discovery throughout the years. Additionally, we evaluate and highlight the complexity and different aspects of candidate variant and candidate gene interpretation. Finally, we describe recent and upcoming improvements and innovations of existing technologies and the development of novel technologies in the field.

IDENTIFICATION OF GENES ASSOCIATED WITH HEARING LOSS AND RETINAL DYSTROPHY

Linkage analysis

The first HL- and RD-associated genes were identified using linkage analysis and candidate gene strategies in the early 90s.¹⁰⁻¹² Examples of candidate gene approaches include analysis of candidate disease-associated genes based on their function, gene expression or animal model studies (discussed in (13)). Linkage analysis was used to pinpoint a genomic region of interest likely to encompass the disease gene. The strategy is based on the key principle that a disease haplotype is shared between affected individuals within a family but is not present in unaffected individuals. The shared haplotype cosegregates with the disease according to the observed mode of inheritance. Initially, linkage regions were mapped using laborious genotyping of polymorphic microsatellite markers, but the process was optimized when microarray technologies became available. Microarrays, such as SNP-arrays, allow rapid genotyping of thousands of single nucleotide polymorphisms (SNPs) that are present across the genome and have a variant allele frequency higher than 1% in the healthy population. The higher the density of the SNPs on the array and the more SNPs that reside within the region showing linkage disequilibrium, the more precise the determination of a possible disease haplotype is. The distance between two SNPs can be expressed in centimorgans (cM). One cM is defined as the distance between chromosomal positions that have a 1% chance of being separated by chromosomal recombination during meiosis. A logarithm of the odds (LOD) score can be calculated to estimate the odds that two loci, or a locus and a disease-associated gene, are located at an assumed distance from each other (expressed as the recombination fraction θ). A LOD-score of 3.3 or higher is considered evidence for linkage in a genome-wide manner, with a probability of 95%.¹⁴ Nowadays, several tools (e.g. GENEHUNTER¹⁵ and PLINK¹⁶) are available to calculate the LOD-score and identify a linkage region. However, large family pedigrees and sufficient participating family members are required to reach a statistically significant linkage. When a disease-associated locus is defined, Sanger sequencing can be performed to evaluate the genomic region for causative variants. In this way, the linkage analysis strategy has been applied very effectively for disease gene identification for many years (reviewed in (13, 17)). Despite the introduction of higher throughput sequencing techniques, SNP genotyping can still be very useful to determine regions of genotype-sharing even in small families, especially to reduce the number of candidate variants.

Homozygosity mapping

Genome-wide homozygosity mapping has proven to be a powerful tool to identify disease-associated genes for autosomal recessive disorders. For both inherited HL and RD, a significant number of disease-associated genes were identified using this strategy.^{13,17} In consanguineous families, a pathogenic variant is often present in a homozygous state as it is inherited from a recent common ancestor (grandparent or great grandparent). Homozygosity mapping can be used to determine regions that contain consecutive homozygous variants^{16,18}, which is often performed using SNP-arrays. Although the average size of homozygous stretches is larger in consanguineous families (typically between 30 to 100 megabase (Mb)-sized regions)^{19,20}, several studies have indicated that this method is also an effective tool for non-consanguineous families (1-30 Mb-sized homozygous regions¹⁹⁻²¹). *EYS* is one of the most frequently mutated genes in RD and was identified using homozygosity mapping in a non-consanguineous family.^{22,23} Other examples of disease gene identification using homozygosity mapping in a non-consanguineous family include *PDE6C*²⁴ which is associated with RD, and *OTOG*²⁵ and *MYO15A*²⁶, which are implicated for HL. The size of a homozygous disease-associated haplotype decreases over subsequent generations due to meiotic recombination. Well-characterized families and detailed phenotypic information are prerequisites for the successful application of this technique.

Next generation sequencing

DNA studies have been revolutionized by the conventional Sanger sequencing technique, which was introduced in 1977.²⁷ It is known as an enzymatic sequencing or chain-termination method, which utilizes labeled di-deoxynucleotides acting as chain terminators.²⁷ The first human genome was sequenced based on the Sanger sequencing technology in 2001, which took almost 13 years to complete and cost \$2.7 billion, and was part of a large collaborative and international publicly funded project.²⁸ In parallel, efforts to sequence the first human genome were also performed in a commercial setting by the company Celera Corporation, whose results were revealed in joint publications with the public human genome project.^{29,30} The Celera project employed a whole genome shotgun sequencing approach and proceeded at a much faster pace and lower cost, although it benefited significantly from the data that were already generated by the public human genome project.^{29,31} As a result of both efforts to sequence the human genome, it became clear that the scale, efficiency, and cost needed to be vastly optimized for routine use in clinical diagnostics. Therefore, shortly after the release of the human genome sequence, the aim was re-established to achieve a \$1,000 human genome within 10 years.³²

Sanger sequencing is still routinely used for variant validation and has an extremely high accuracy of up to 99.999%.³³ However, it is considered a low-throughput technique as up to 1 kilobase (kb) of DNA can be sequenced in 96 or 384 parallel reactions.³⁴ The technique has been optimized by the application of nucleotide-specific fluorescent labels and automated detection^{35,36}, the invention of the polymerase chain reaction (PCR)³⁷, and the usage of polyacrylamide gels in capillary electrophoresis.³⁶ Therefore, DNA sequencing can be achieved within a shorter time frame and on a larger scale, in which the sequencing of millions of reads can be carried out in parallel called “massive parallel sequencing” or “NGS”.

The NGS technique has rapidly overcome the limitations of traditional sequencing. Since 2005, various sequencing platforms such as Illumina, Ion Torrent, Roche 454, and SOLiD sequencing have been developed, which has resulted in a rapidly changing landscape during this new era of sequencing. The read length of these different platforms is shorter than that of Sanger sequencing (approximately 50-500 bp) and with a higher error rate (0.1% in NGS compared to 0.001% in Sanger sequencing).³⁸ However, the fast development of NGS techniques and the generation of public reference datasets containing population allele frequency data allowed a widespread integration of NGS technology in the research community and later in the clinical diagnostics of genetic diseases. Nevertheless, as whole genome sequencing (WGS) is still relatively expensive and data interpretation is complex, a targeted sequencing approach (e.g. whole exome sequencing (WES)) is often preferred.

Targeted capture sequencing

Genomic regions of interest, such as the genes implicated in HL or RD, can be selectively enriched before sequencing is performed. There are various methods available to enrich for target regions such as hybridization-based, highly multiplexed PCR-based, and targeted circularization approaches. Extensive studies have been performed which have applied these techniques to unravel genetic defects involved in inherited HL and RD. In 2013, Chio et al. investigated 32 cases with familial non-syndromic HL, in which they reached a molecular diagnostic rate of 37% using a candidate gene sequencing approach of *GJB2*, *SLC26A4*, *POU3F4* or mitochondrial genes, based on observed clinical features and inheritance patterns. Later, by application of hybridization-based target capture sequencing for 80 HL-associated genes, they were able to increase the total diagnostic detection rate to 78% in this cohort.³⁹ In 2017, Dockery et al. utilized the hybridization-based enrichment method to sequence 254 RD-associated genes in over 750 affected individuals in Ireland, in which they could identify pathogenic variants in 68% of the cases.⁴⁰ A recent study by Khan et al. applied a highly multiplexed PCR-based approach with single-molecule molecular inversion probes (smMIPs) to sequence

the complete *ABCA4* gene (coding and non-coding regions) in 1,054 individuals with Stargardt disease, who were previously screened for variants in the coding regions and remained genetically unexplained. This study proved that a smMIP-based approach is a cost-effective approach in case of a strong genotype-phenotype correlation. The method allowed deep-sequencing of the region of interest, and causal structural and deep-intronic variants were identified in 25% of the investigated cases, who were genetically undiagnosed after pre-screening methods.⁴¹

Targeted NGS techniques have several advantages, such as less data storage, high sequencing accuracy due to high coverage, cost-, and time-effectivity.⁴² However, this approach is unable to detect variants in novel (candidate) disease-associated genes. Furthermore, pathogenic variants residing in non-coding regions, and structural variants (SVs) can be missed if only exons are analyzed. Due to decreasing prices of both WES and WGS, these approaches have become rapidly preferred to overcome the disadvantages of targeted NGS.

Whole exome sequencing versus whole genome sequencing

Protein coding regions comprise 1-2% of the human genome. However, it is estimated that they harbor approximately 85% of disease-causing variants.⁴³⁻⁴⁵ Therefore, the enrichment of coding regions utilized in WES quickly became an accurate and efficient method to investigate the coding regions of the genome for potential pathogenic variants, and this is now widely applied in genetic diagnostics.⁴⁶ One of the striking features of WES is in the success rates of genetic diagnostics of diseases with extensive locus heterogeneity, such as inherited HL and RD.^{4,47} Currently, the diagnostic yield for RD using WES is estimated to be between 50% and 80%, dependent on the phenotype studied.^{4,48-50} According to a study performed by Haer-Wigman et al., the highest yields were obtained for retinitis pigmentosa (63%) and lowest yields are obtained for macular dystrophy (28%) and rare unspecified types of RD (25%).⁴ For HL, the genetic diagnostic rate is also highly dependent on phenotype (e.g. syndromic or non-syndromic phenotype, mode of inheritance). The highest rates are observed in patients with a positive family history or with a congenital or symmetric type of HL.⁴⁷ The overall estimated detection rate for HL, when employing WES, varies between 30-40% based on different large cohort studies and largely depends on phenotypic diversity.^{3,51} The diagnostic yield for HL is importantly influenced by the involvement of environmental factors (e.g. noise, ototoxic drugs and trauma), which likely explains the difference in yield compared to RD. Genetic causes have been estimated to underlie approximately two-thirds of the cases of congenital and early childhood HL; the remaining cases can be explained by acquired causes.⁵² This genetic contribution decreases with the

patient's age due to increased exposure to damaging environmental factors during life. In line with this observation, there are several reports of a negative correlation between diagnostic yield and age of onset of HL.^{3,47}

Despite the successes of WES in clinical settings, this technology is inaccurate in detecting SVs, such as a deletion of a single exon, and does not allow variant detection in deep-intronic regions or regulatory elements. Therefore, WGS may be preferred as it provides a more evenly distributed and uniform read coverage, and it is capable of detecting different types of variants across the entire genome.⁵³⁻⁵⁶ In 2017, Carss et al. investigated a large cohort of RD patients, in which WGS was performed for 605 cases, WES for 72 cases, and for 45 cases both technologies were performed.⁵⁶ They identified disease-causing variants in 56% of all individuals (404/722), while by using WES alone the diagnostic yield was calculated to be 50%. Subsequently, 45/58 cases that remained unexplained by WES underwent WGS, and pathogenic variants were identified in 14 cases. The authors concluded that WGS has great power to detect pathogenic SVs, variants in non-coding and regulatory regions, and variants in GC-rich regions. The application of WGS revealed the pathogenic variants in 31% of the cases that remained unsolved after WES. These variants were missed mainly due to the poor quality of reads or the incapability of WES to identify SVs.⁵⁶ The prices for WGS keep decreasing⁵⁷ and the importance of the non-coding regions of the genome has become more evident. Therefore, a shift from exome to genome sequencing will be observed in clinical diagnostics in the near future to overcome the diagnostic gap observed in the application of WES. In 2020, Méjécase et al. provided a practical and cost-effective guideline for current and future genetic testing of RDs in which they proposed to utilize WES or targeted NGS for the initial screening of exons and flanking intronic regions of (candidate- or known RD) genes, reserving WGS solely for cases that remained unresolved.⁵⁸

Although NGS techniques have revolutionized the field of medical genetics, these short-read sequencing (SRS) approaches pose several limitations, such as (1) difficulties in the identification of complex and large SVs, (2) inability to sequence repetitive regions, (3) the lack of phasing of alleles, and (4) difficulties distinguishing highly homologous regions such as pseudogenes.⁵⁹ These limitations may play a significant role in the diagnostic gap in medical genetics.

Third generation sequencing

Due to the limitations of the aforementioned NGS techniques, there has been a need to develop new sequencing approaches to overcome these issues. The era of third generation sequencing arrived in 2011, when Pacific Biosciences (PacBio) released a

novel sequencing technique called “single molecule real-time” (SMRT) sequencing⁶⁰, and only three years later, Oxford Nanopore technologies introduced nanopore sequencing.⁶¹ Although these two techniques utilize different approaches to sequence genomic DNA, they share two major advantages compared to NGS. First, they are established on PCR-free and real-time sequencing processes, and second, they generate ultra-long sequencing reads, >10 kb.^{59,62} These long-read sequencing (LRS) technologies are revolutionizing the genetics field as they provide a further understanding of the normal and morbid anatomy of human genome and thereby can fill the gaps in the molecular diagnostics of genetic diseases.

Single molecule real-time (SMRT) sequencing

SMRT sequencing relies on ligating hairpin adapters to both ends of the double-stranded template DNA molecule (dsDNA), thereby circulating the dsDNA into the construct called the SMRT-bell. In the next step, primers and DNA polymerase are annealed to the adapter in the SMRT-bell, which will later be utilized for circular consensus sequencing (CCS) (**Box 1, Figure 1**). The CCS approach can obtain approximately 83% accuracy (10x coverage on average) with a 13-15% error rate dominated by small insertions and deletions.^{62,63} This can be improved to 99% accuracy by selectively sequencing a targeted region with an increased coverage of 15x.⁶⁴⁻⁶⁶ The SMRT technology is a PCR-free approach and requires minimal amounts of reagents and a simple library preparation procedure by which ultra-long dsDNA can be obtained. This technology can provide the result within hours compared to several days for previous approaches.⁶⁰ An average read length of 10-15 kb can be reached, which allows de novo assembly, phasing of variants and haplotyping and the detection of large SVs throughout the genome.^{59,67,68}

Nanopore sequencing

Nanopore sequencing is an advanced third generation sequencing technique that offers straightforward sample preparation, requiring minimal reagents or amplification processes.⁶⁹ This technology relies on transferring a DNA molecule through a pore and directly detecting each nucleotide by its effect on an electric current (**Box 2, Figure 2**).^{70,71}

BOX 1 - Single molecule real-time (SMRT) sequencing technique

To enable sequencing of single DNA-molecules in real-time, two obstacles had to be overcome. First, concentrating the DNA polymerase and its template, SMRT-bell (**Figure 1A**), to the very small observation chambers, which creates a higher signal-to-noise ratio. This problem has been solved by zero-mode waveguide (ZMW) technology, a small hole of approximately 45 nanometer (nm) in diameter.⁷² The DNA polymerase with its template is anchored by a strong biotin/streptavidin interaction to the bottom of the ZMW. Therefore, the laser illumination of incorporating nucleotides is limited to the bottom, which increases the signal-to-noise ratio⁶³ as ZMW can efficiently distinguish signals of nucleotide incorporation against the background of unincorporated nucleotides (**Figure 1B**). The second obstacle in real-time sequencing of single DNA-molecules was the large size of the fluorescent dye, which interfered with the normal activity of DNA polymerase and caused halting of the enzyme shortly after initiation of DNA synthesis. In the SMRT technology, the dye is attached to the phosphate chain instead of the nucleotide, which is naturally cleaved during DNA synthesis after nucleotide incorporation and results in a single long, natural DNA strand.⁶³ The real-time sequencing of the circular SMRT-bell is performed in each ZMW that generates continuous long reads (**Figure 1B**). During data processing, the adaptors are removed, and subreads are generated. Subsequently, the combined subreads enable the generation of one highly accurate consensus sequence called the circular consensus sequence (CCS).

There is no limit in DNA length that can be sequenced, since it does not require DNA amplification or synthesis. The challenge lies in library preparation, which needs to result in ultra-long dsDNA molecules.⁷⁵ The average size of reads is usually >10 kb and for some molecules, it can reach 1 Mb.⁵⁹ The main drawback of nanopore sequencing is its relatively high error rate of ~20%. Compared to SMRT technology, in which the error rate can be reduced by high coverage due to CCS, in Nanopore sequencing it is a systematic error and correction can only be achieved by comparison to short-read sequence data.⁷⁵ Nevertheless, this technology is rapidly improving to overcome current issues.⁷¹

Application of third generation sequencing in inherited HL and RD

Third generation sequencing has revolutionized the field of medical genetics by its superior performance in the analysis of repeated and highly homologous regions, SVs, haplotype phasing, and transcriptome analysis.⁷⁶ These technologies are currently mainly used in research applications and show great promise to overcome the disadvantages of SRS methods. In a systematic analysis, Ebbert et al. compared the performance of whole-genome SRS and LRS technologies at repetitive regions in the human genome. Amongst others, they showed that 8.6% of the protein-coding regions of *RPGR* (associated with X-linked RD) and 12.7% of the protein coding regions of *OTOA* (associated with HL) are within the unmapped reads of SRS-data, which were resolved by performing LRS. Specifically, they indicated that Nanopore sequencing outperforms PacBio sequencing by resolving 90.4% and 64.4% of the SRS-unmapped regions, respectively.⁷⁷

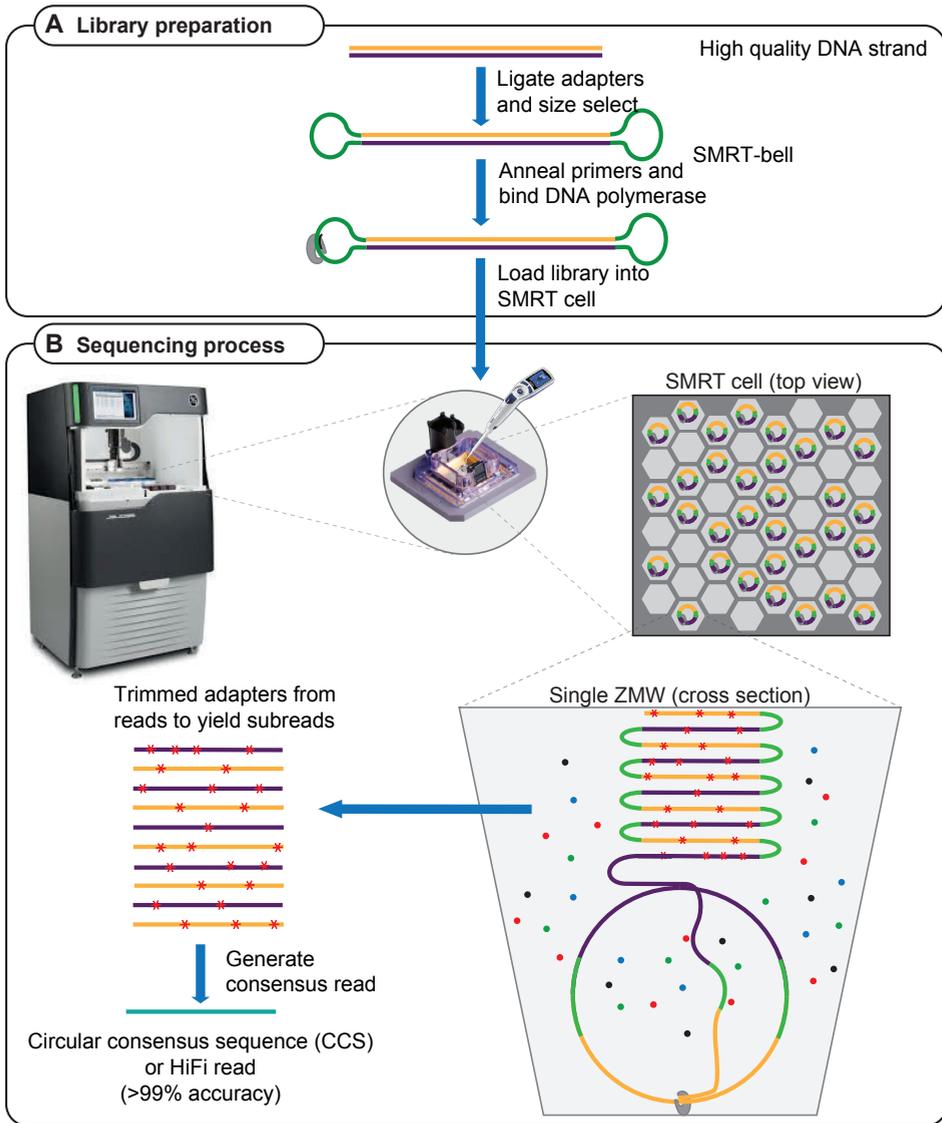


Figure 1. Overview of SMRT sequencing. (A) Sequencing starts with library preparation of ultra-long dsDNA. In the next step, adapters, DNA polymerase, and primers bind to the dsDNA creating the SMRT-bell, which later will be loaded to the SMRT-cell. (B) The library is randomly distributed in the SMRT-cell in the sequencer instrument, in the ideal condition 1/3 of the ZMWs will be loaded by an SMRT-bell. In each ZMW, DNA polymerase together with an SMRT-bell are bound to the bottom of the ZMW. The SMRT sequencing uses the circular DNA template to generate a continuous long read in each ZMW chamber. Afterwards, the adapters are trimmed from this long read and overlapping reads can be combined to one consensus sequence of high quality HiFi read.

BOX 2 - Nanopore sequencing technique

Nanopore sequencing occurs in a flow cell, in which two ionic solution compartments are separated by a membrane containing thousands of nanopores. The flow of electric current between these compartments depends on the molecule transferring through one of the pores. Since each nucleotide differs in shape, its effect on electric current is specific for each nucleotide.^{62,71,73} Library preparation in Nanopore sequencing includes end-repair of the ultra-long dsDNA, dA-tails addition, size selection and ligation of adapters (protein-DNA molecules). The first adapter is the leader-adapter, which contains a motor enzyme. It binds to the nanopore and ensures gradual movement of DNA through the pore. The dsDNA is then unwound at the pore and only one strand will pass through the pore. The second adapter is a hairpin-adapter containing a hairpin protein. It generates one long single-strand of DNA, which ensures the sequencing of the second strand of DNA in order to increase sequencing accuracy.^{61,62,74}

One important application has been to identify complex SVs associated with genetic diseases including HL and RD. Reiner et al. utilized SMRT LRS to detect a 72.8-kb deletion region in the *BBS9* gene and map the breakpoints at the nucleotide level in a patient diagnosed with Bardet-Biedl syndrome. This deletion was determined to be the causal variant and a founder mutation in the Guyanese population.⁷⁸

In another recent study, researchers utilized transcriptome sequencing followed by short- and long-read WGS to identify a 7.4-kb duplication in *NMNAT1*, which spans two out of five exons of this gene. The duplication caused a previously unrecognized autosomal recessive syndrome, symptoms of which are Leber congenital amaurosis and sensorineural HL, which occur together with other features such as spondylo-epiphyseal dysplasia, intellectual disability, and brain anomalies. The authors were able to determine the exact breakpoints of the duplicated region, missed by previous approaches, as well as two *Alu* elements flanking this segment which are potentially involved in the origin of the SV.⁷⁹

Recently, Nanopore sequencing enabled researchers to unravel the genetic defect in two unrelated patients diagnosed with mild-to-moderate HL. Nanopore sequencing revealed a gene conversion event between the *OTOA* gene and its pseudogene, in which exons 21 to 23 of *OTOA* were replaced by exons 1 to 3 of *OTOAP1*.⁸⁰ As pathogenic variants within the *OTOA* gene have been described to cause autosomal recessive HL (DFNB22), this gene conversion event was considered causative.⁸⁰

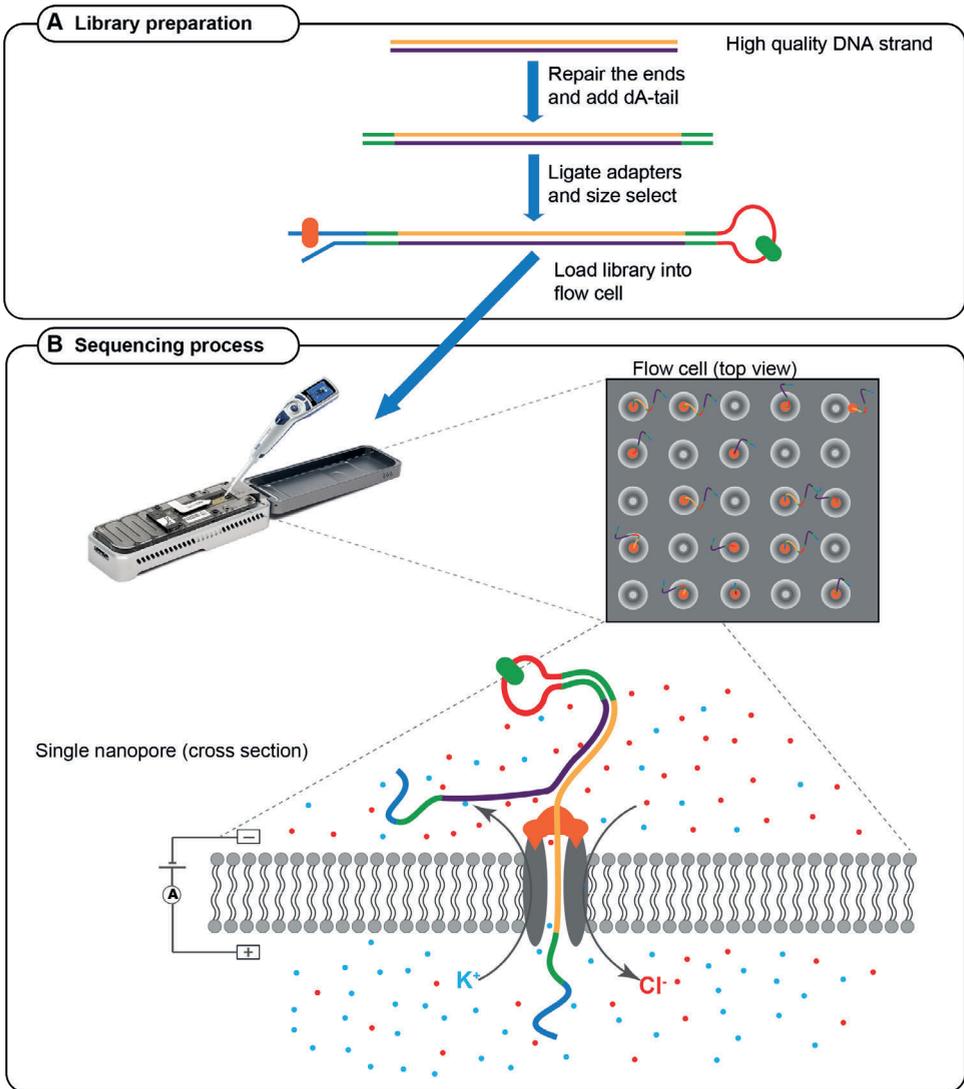


Figure 2. Overview of Nanopore sequencing. (A) Library preparation includes end-repairing, adding dA-tails and ligation of two types of adapters to both ends of the ultra-long dsDNA. The adapters carry the motor enzyme (orange) and hairpin-protein (green) to facilitate movement of DNA through the pore and ensure sequencing of the second strand of DNA, respectively. (B) The library is loaded into the flow cell in the instrument. The flow cell contains thousands of nanopores that allow Cl^- and K^+ ion-flow between the two compartments. The motor enzyme anchors to the nanopore and unwinds the DNA to pass it through the pore. Thereby, the electric current is influenced based on the unique shape of each nucleotide in ssDNA. These changes in the electric current are later translated to sequences.

Despite the advantages of LRS techniques, they possess multiple important drawbacks that prevent a wide range of uses outside research applications. One of these is the relatively high costs compared to SRS NGS technologies (\$800-2,000 per run depending on the different platforms and instruments), based on the lowest possible flow cell price and highest output.⁶² The other major disadvantage is the requirement for high quality ultra-long dsDNA, which can be challenging to obtain. In particular for Nanopore sequencing, the required fresh blood samples for DNA extraction can also be a hurdle. However, as LRS technologies are rapidly decreasing in price and are continuously improving in different aspects, such as optimized library preparation and error correction, it is expected that these technologies will eventually enter routine genetic diagnostics in Western countries. In addition, like SRS, targeted LRS can also be performed by targeted amplicon sequencing, CRISPR/Cas-based targeted enrichment, or using a “Read Until” approach, in order to enrich for genetic loci associated with a specific phenotype. Targeted LRS is a cost-effective and efficient strategy to investigate high-priority loci in unsolved cases.^{81,82} For both HL and RD, several associated genetic loci (44 and 36 loci, respectively) have been described for which the implicated genetic defect is still elusive^{1,2}, and therefore, a targeted LRS approach could be of interest.

Finally, as sequencing technologies develop and improve rapidly (**Figure 3**), the next challenge will lie in bioinformatics, data storage, data analysis, and variant interpretation of NGS or LRS data. A high number of different variants are revealed by these methods. However, not all these variants are disease causing. Therefore, special attention is being paid to prioritization processes that can aid in decreasing the number of putative candidate variants. In addition, developments in bioinformatic tools are needed to better interpret the effect of candidate variants. In the next section, we will discuss the importance and challenges of variant interpretation and the importance of this matter in clinical application.

VARIANT INTERPRETATION

The total length of human DNA is over 3 billion base pairs and holds on average 4-5 million variants compared to the healthy human reference genome, which highlights the obvious challenge of distinguishing potential disease-causing variants from benign variants or polymorphisms.⁸³ For protein-truncating variants, a potential pathogenic consequence is often evident, while missense, synonymous and non-coding variants are more challenging to interpret. Moreover, with the increased knowledge regarding the involvement of the non-coding DNA in human disease development, the complexity of data to be analyzed has gone through the roof.

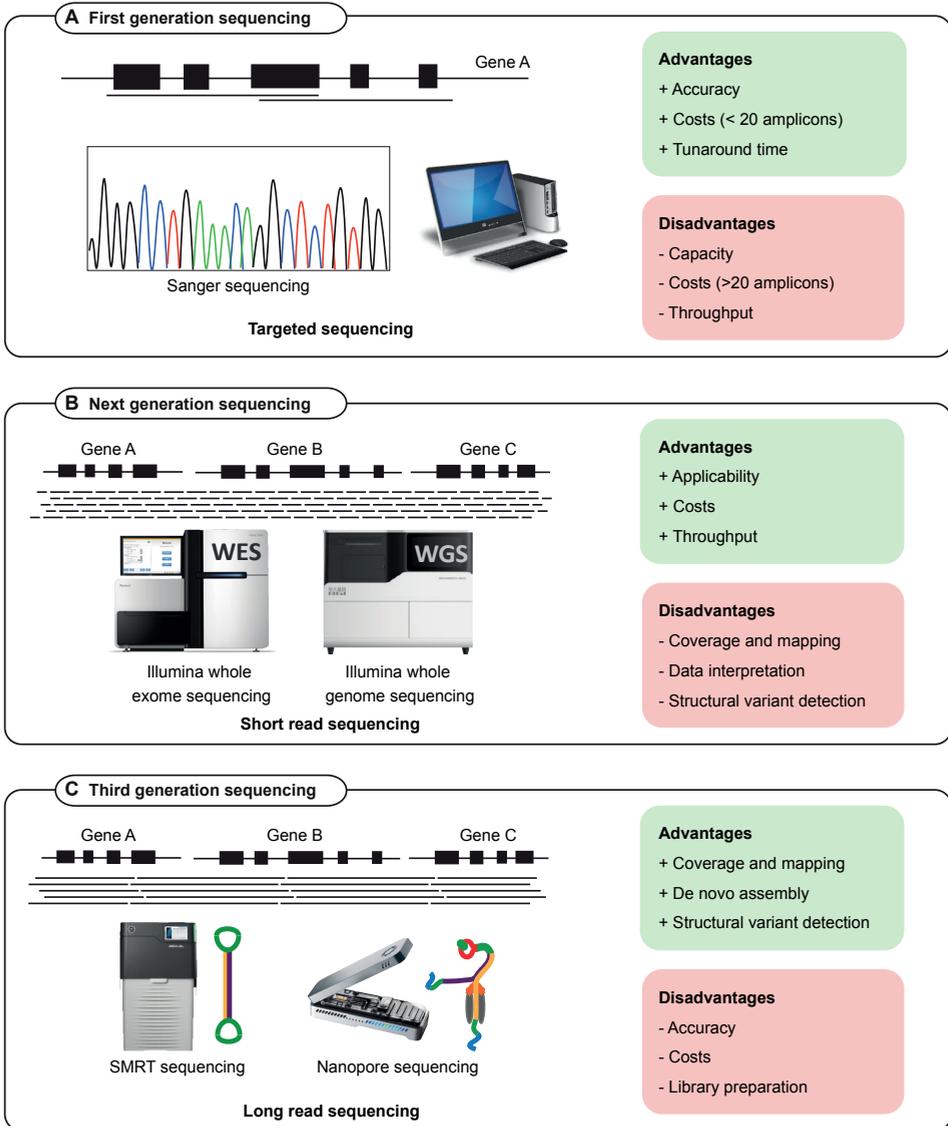


Figure 3. Comparison of conventional Sanger, next generation, and third generation sequencing. A schematic representation of **(A)** first generation sequencing (Sanger sequencing), **(B)** next generation sequencing (e.g. Illumina whole genome sequencing (WGS) and whole exome sequencing (WES)) and **(C)** third generation sequencing (e.g. SMRT sequencing as performed by Pacific Biosciences (PacBio) and nanopore sequencing by Oxford Nanopore Technologies (ONT)). For each technique, advantages (green) and disadvantages (red) are provided.

In 2015, the American College of Medical Genetics (ACMG) provided a framework that should utilize and standardize sequence variant interpretation for Mendelian disorders.⁸⁴ Each variant is categorized using a uniform scoring system: benign, likely benign, uncertain significance, likely pathogenic, or pathogenic. The classification system employs several hierarchical steps which include the use of literature and databases, computational and predictive data, functional data, and segregation analysis. Variant classification is the cornerstone of clinical molecular genetic testing. Therefore, the ACMG guidelines provide a consistent and well applicable system to guide this process. On the other hand, for research focused on the identification of novel gene-disease associations, the ACMG guidelines are more difficult to apply and less suitable.

Literature and database use

Variant frequency databases are useful resources for allele frequencies of variants in large populations. As a rule of thumb, the frequency of a disease-causing variant should not be higher than expected based on the incidence or prevalence of the genetic disorder.⁸⁵ The most comprehensive allele frequency database today is gnomAD (successor of ExAC), which contains frequency data for both SNVs and SVs based on 91,864 genomes and 125,748 exomes.⁸⁶ Additionally, this database provides variant frequencies for many subpopulations, which allows the usage of population-matched control data.

Nevertheless, some populations (e.g. African/African-American) remain underrepresented which limits efforts in precision and personalized medicine for these ethnicities. Several efforts are ongoing to obtain more (high-quality) genomes from these populations.^{87,88} Databases such as gnomAD⁸⁶, goNL⁸⁹, UK10K⁹⁰ and Welllderly⁹¹ contain sequencing data of (presumably) healthy cohorts. However, important caveats related to age-of-onset and reduced penetrance should not be ignored.

Unlike population databases, disease databases such as ClinGen⁹², ClinVar⁹³, Leiden Open (source) Variation Databases (LOVDs)⁹⁴, and the Human Gene Mutation Database (HGMD)⁴⁴ also provide genotype-phenotype information. All the variants collected in the HGMD database have been reported in patients and likely disease causing. They have been published in literature and manually curated. The Deafness Variation Database (DVD) provides a comprehensive catalog for genetic variation in genes associated with HL.⁹⁵ Efforts are ongoing to collect and annotate all published variants associated with inherited non-syndromic RDs, Bardet-Biedl syndrome and Usher syndrome into 195 gene-specific LOVDs.^{23,96-100}

Several studies have proven the value of incorporating population frequency data as a variant prioritization strategy and have successfully clarified variants of unknown

significance.^{56,101} However, an important caveat is that a reliable database should be frequently updated, and uploaded sequencing data should adhere to quality control criteria. An example of non-pathogenic variants mistakenly reported as pathogenic has been highlighted in a study performed by Hanany et al.¹⁰² The authors extracted up-to-date allele frequencies from gnomAD of variants in genes associated with dominantly inherited RD and concluded that the pathogenicity of variants in 19% of these genes should be debated. Inherited HL, on the other hand, is a more common condition, than RD and therefore the expected maximum allele frequency for a pathogenic variant should be adjusted accordingly.¹⁰³

Once a potentially disease-causing variant is identified, a rich source of available scientific and medical literature can be assessed. A first important step entails thorough comparisons between the observed phenotype in the investigated proband and phenotypic observations described in literature. Most well-described phenotype-genotype correlations can be found in data repositories: Online Mendelian Inheritance in Man (OMIM)¹⁰⁴, ClinGen⁹² and ClinVar.⁹³

Strong phenotype-genotype correlations are complicated by a phenomenon called allelism: different phenotypes can result from different alleles of the same gene.¹⁰⁵ For example, autosomal recessive Stargardt disease (STGD1), which is due to two variants or alleles in *ABCA4*, shows a wide clinical spectrum of maculopathies.¹⁰⁶ The most severe form is early-onset (onset <10 years) STGD1 or panretinal cone-rod dystrophy which is due to two deleterious *ABCA4* alleles. Classical or intermediate STGD1 (onset between 10 and 40 years) is due to a combination of a deleterious variant and a mild variant. Finally, late-onset STGD1 (onset >40 years) is caused by a deleterious variant and a mild variant (p.Asn1868Ile) showing reduced penetrance.¹⁰⁶⁻¹⁰⁸ Truncating variants in the *CDH23* gene are assumed to cause Usher syndrome type 1D, which consists of HL and retinitis pigmentosa, whereas missense variants cause non-syndromic HL.¹⁰⁹ However, several exceptions to this rule have been reported.^{110,111} For pathogenic variants in the *USH2A* gene that can cause both non-syndromic retinitis pigmentosa and Usher syndrome type IIa, the correlation with missense and truncating variants with the associated phenotypic expression is not always clear, although truncating *USH2A* variants are more frequently reported in patients diagnosed with a syndromic phenotype.^{112,113} Additionally, variants affecting genes that are implicated in ciliopathies (e.g. *BBS1*, *CEP290*, *IQCB1*) can cause a wide range of variable symptoms that are part of a (syndromic) phenotype. Symptoms described for ciliopathies often include retinal degeneration and less frequently HL (reviewed in (9)). To date, >80 forms of syndromic RD have been described which are linked to approximately 200 RD-associated genes¹¹⁴; for syndromic HL, these numbers are suggested to be even higher.¹¹⁵

Besides a phenotypic resemblance, the expected mode of inheritance and the involved pathogenic mechanism of the variant (e.g. haploinsufficiency, loss- or gain-of-function) should also be compared with literature reports. For genes that have not been previously associated with the disease of interest, OMIM¹⁰⁴ and GeneCards¹¹⁶ provide a summary of known clinical and functional information for the gene. For candidate disease genes, it may be valuable to investigate gene expression in the tissue of interest (e.g. SHIELD¹¹⁷, gEAR¹¹⁸, EyeGEx¹¹⁹) and explore associated protein interaction networks (e.g. STRING¹²⁰). Additionally, the initiative Genematcher¹²¹ and European Retinal Disease Consortium (ERDC)¹²² offer the opportunity for different research groups that share an interest in the same candidate disease gene to collaborate. It is hypothesized that the most prominent genetic causes of diseases have been identified and novel findings appear in few cases or families, which underlines the urgency for collaborations among research groups worldwide. It is of utmost importance to share candidate disease gene data to increase the likelihood of identifying multiple unrelated individuals affected by pathogenic variants in the same gene.^{123,124}

Computational and predictive data

The spectrum of human genetic variation is diverse, and a rich source of bioinformatics tools has been developed to evaluate the different potential consequences of a variant. Although the pathogenicity of SNVs has been most extensively studied, recent efforts into the characterization of SVs have revealed that pathogenic SVs are more abundant than initially thought.^{85,125} This has led to a gradual shift of attention from coding variation to structural variation and the non-coding regions of the genome.

Null variants

Null variants are considered as very strong evidence of pathogenicity and often lead to open reading frame disruption and consequently complete loss of protein function. Null variants include nonsense, frameshift, canonical splice site, and initiation codon variants, as well as out-of-frame single and multi-exon deletions. Available *in silico* prediction tools are often not designed for the interpretation of null variants as pathogenicity already seems evident in most cases. However, some caveats should be considered, including the presence of alternative transcripts, the position of the variant with respect to the 3'UTR, and the inducement of alternative splicing such as in-frame exon skipping as a putative correction mechanism.¹²⁶⁻¹²⁸ For each gene, a loss-of-function intolerance (pLI) score is provided in gnomAD⁸⁶, which is based on observed (homozygous) loss-of-function variants in healthy cohorts compared to the expected number based on the gene size.

Missense, synonymous, indel and intronic variants

Substitution variants located in the coding (exonic) or non-coding (intronic) regions of a gene are more difficult to interpret. Missense variants and small in-frame insertions or deletions (indels) lead to changes in amino acid composition. Several computational tools have been developed to aid in the assessment of deleterious consequences of the identified variants. Output scores provided by these tools are usually based on the evolutionary conservation of the altered nucleotide or amino acid residues, biochemical consequences of the amino acid change, or the location and context of the residue within the protein sequence e.g. in a domain with a specific function. Most widely applied tools are Combined Annotation-Dependent Depletion (CADD)¹²⁹, Grantham¹³⁰, MutationTaster¹³¹, phyloP¹³², PolyPhen-2¹³³ and Sorting Intolerant From Tolerant (SIFT).¹³⁴

Alternatively, synonymous, missense and (deep)-intronic variants can disrupt the normal splicing machinery and alter pre-mRNA processing. Variants can introduce or strengthen cryptic splice sites, disrupt canonical donor or acceptor splice sites or disrupt the (binding) motifs that are essential for correct splicing processes, such as exonic splicing enhancers or silencers.¹⁰¹ This can lead to alternative splicing events, such as pseudo-exon inclusion, exon elongation or (partial) exon skipping. Potential splice-altering variants can be evaluated based on nucleotide conservation scores or by performing splicing assessments using predictive splicing algorithms, such as Human Splicing Finder¹³⁵, SpliceSiteFinder-like¹³⁶, MaxEntScan¹³⁷, GeneSplicer¹³⁸, NNSPLICE¹³⁹ and SpliceAI, a deep learning algorithm.¹⁴⁰ *In vitro* midi- or minigene splice assays can be performed to confirm the predicted alternative splicing events in HEK293T cells or, if transcript levels allow, aberrant splicing can be detected in RNA derived from (EBV-transformed) blood cells.^{141 142}

One pitfall of splice site prediction tools is tissue-specific splicing of exons, which prevents most current prediction tools from detecting cochlear or retina-specific splicing effects. Recently, Riepe et al. benchmarked different established and deep-learning tools on sets of variants in tissue-specific genes *ABCA4* and *MYBPC3* and observed that SpliceAI is the best performing splice prediction tool for both non-canonical splice site and deep-intronic variants in *ABCA4*.¹⁴³ Moreover, Rowlands et al. compared seven machine and deep learning-based splice prediction tools and demonstrated that SpliceAI is superior in both sensitivity and specificity.¹⁴⁴

Regulatory variants

Variants located in intergenic and intronic regions of the genome can exert their pathogenic effect through a variety of mechanisms. Variation can occur within characterized *cis* regulatory elements (CREs), such as promoters, enhancers, or

insulators.^{145,146} Regulatory elements are short DNA sequences (100-500 bp) that allow precise spatiotemporal control of gene expression levels.¹⁴⁵ Promoter and distant enhancer regions interact with each other via chromosomal looping, allowing the recruitment of the transcriptional machinery. Alternatively, insulators can block the interactions between promoters and enhancers. An enhancer element can be located up to 1 Mb away from its target gene and can serve as the transcriptional regulator of one or more genes.^{145,147-149} Usually, an enhancer displays a spatiotemporal pattern of activity. Transcription factors, that bind enhancer or promoter elements, are the key regulators of these processes and modulate gene expression. Pathogenic variants in *cis* regulatory elements could alter the transcription factor binding sites or chromatin landscape and therefore the activity of the enhancer or promoter.^{145,146} Databases such as JASPAR¹⁵⁰ that contain consensus sequences of transcription factor binding sites can be used to predict the effect of a potential regulatory variant on transcription factor binding.

Regulatory variants that impact transcription initiation usually lead to subtle changes in gene expression and are difficult to assess.¹⁴⁶ Therefore, a context-specific profiling of the tissue- and cell type-specific *cis* regulatory architecture is essential.¹⁵¹ Enhancer databases such as the ENCODE portal¹⁵², GeneHancer¹⁵³ and EnhancerAtlas¹⁵⁴ contain an overview of reported *cis* regulatory elements that are widespread through the genome. Potential interactions between promoter and enhancer elements can be assessed by evaluating available chromosome conformation capture data like Hi-C. Additionally, the presence of context-specific active enhancer hallmarks should be assessed. These include: (1) confirmed binding of transcription factors, (2) production of non-coding enhancer RNA, (3) an open chromatin conformation and (4) the presence of histone-modification marks that are associated with enhancer activity, such as histone 3 lysine 27 acetylation.^{145,151} **Figure 4** provides an overview of these hallmarks, suitable techniques to assess these, and a selection of available and relevant publicly available (epigenetic) datasets as used to interrogate the recently resolved autosomal dominant retinitis pigmentosa RP17 locus.¹⁵⁵ Once a candidate regulatory variant has been identified, experiments such as an *in vitro* luciferase reporter assay could be applied to confirm its effect on enhancer or promoter activity.¹⁴⁵

Structural variants

SVs are defined as genomic rearrangements that are larger than 50 bp.¹²⁵ SVs include deletions and duplications, also referred to as copy number variations, as well as inversions, translocations and insertions.¹⁴⁷ In 2020, an amendment of the ACMG guidelines was published to aid in the classification of SVs.¹⁵⁶ SVs can have direct consequences on gene dosage levels when (partially) overlapping with coding regions

of a gene, or can cause changes in gene expression levels or patterns when overlapping with regulatory elements, such as enhancers. Additionally, SVs that are limited to the non-coding regions of the DNA can interfere with the 3D genome structure and disrupt *cis* regulatory architecture.¹²⁵ Each chromosome is compartmentalized in regulatory units, so called topologically associating domains (TADs). Within each TAD, enhancers and gene promoters can interact. Neighboring TADs are shielded from each other by boundaries, which are typically occupied by the transcription factor CTCF.¹⁵⁷ Disruption of TAD-architecture by SVs can have severe pathogenic consequences. Deletions can lead to the fusion of neighboring TADs, inversions can result in the exchange of regulatory sequences, and duplications can generate novel TAD compartments leading to ectopic enhancer-promoter contacts (neo-TADs).^{147,158-160} These genomic rearrangements can result in pathogenic alterations of gene expression levels. Recently, it was shown that TAD rearrangements caused by SVs are an important cause of autosomal dominant retinitis pigmentosa (RP17).¹⁵⁵ Additional studies focused on the identification of copy number variants involved in HL or RD have also suggested a prominent role for pathogenic SVs.^{161,162} To predict the potential consequences of structural rearrangements, the epigenetic landscape of the region, including the presence of CTCF-sites, -interactions and -directionality, should be evaluated.

Segregation analysis

Once a candidate disease-causing variant is identified, segregation analysis should be performed, if possible, to confirm that the observed inheritance of the variant matches the family history. If a variant is segregating with the phenotype within the family, this could serve as supportive evidence for linkage of the identified variant to the disorder. However, the variant might still be in linkage equilibrium with the true pathogenic variant and the genetic locus should always be carefully screened for missed variants. Additionally, a careful clinical evaluation of all family members is essential to exclude mild symptoms of reportedly unaffected individuals as well as possible phenocopies whose phenotype can be explained by other (non-genetic) factors. The latter is especially relevant for cases diagnosed with inherited HL, as both genetic and environmental factors are significant contributors to the development of HL.¹⁶³

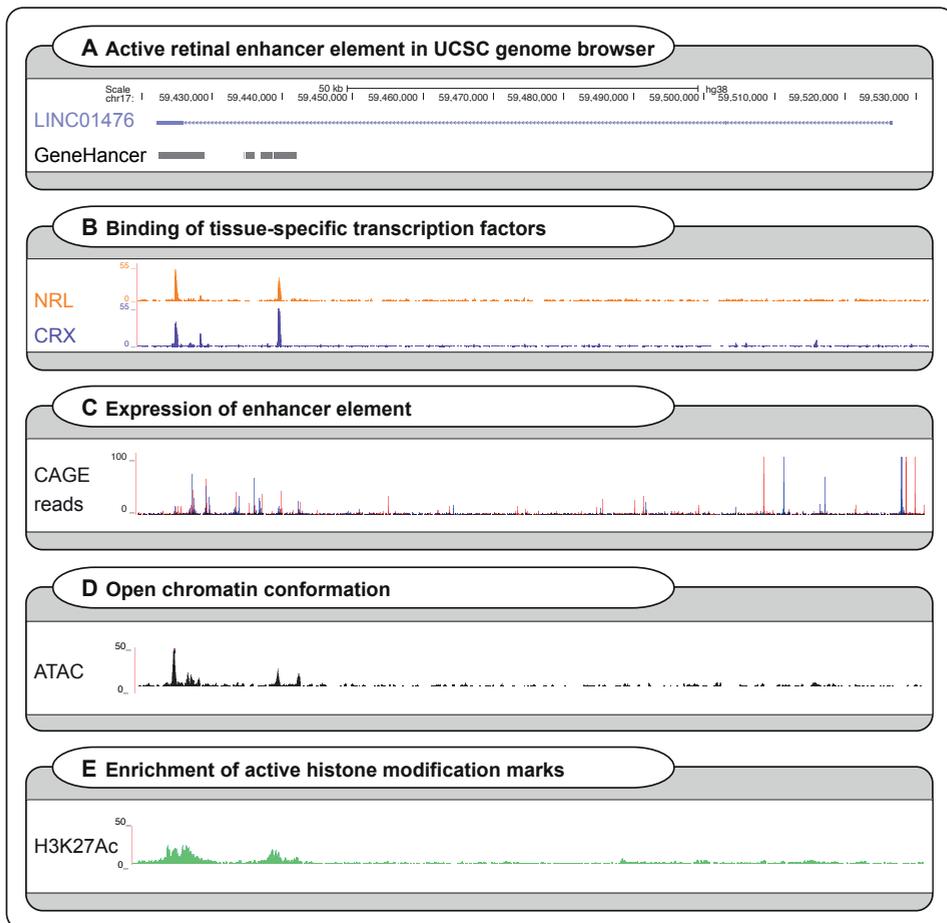


Figure 4. Interpretation of potential regulatory variants using the UCSC genome browser. To evaluate a potential regulatory variant, several publicly available datasets in the University of California Santa Cruz (UCSC) browser can be assessed to determine whether a variant is located within an active *cis* regulatory element based on the presence of active enhancer or promoter hallmarks.¹⁶⁴ **(A)** A known active retinal enhancer element described by de Bruijn et al. and included in the GeneHancer database is visualized in UCSC browser.^{153,155} The enhancer element overlaps with a long non-coding gene *LINC01476*, and is predicted to bind to the promoter region of the *YPEL2* gene (GeneHancer).¹⁵³ The enhancer element is enriched for several active enhancer hallmarks. **(B)** Firstly, the element is bound by retina-specific transcription factors (NRL and CRX) as confirmed by ChIP-seq performed in a human retina sample (GEO database: GSE137311).¹⁴⁵ **(C)** Secondly, cap analysis gene expression (CAGE) that allows 5' end sequencing of cDNAs confirms expression of the enhancer element, as shown by the FANTOM5 CAGE human dataset (Data available from: <https://fantom.gsc.riken.jp/data/>).¹⁶⁵ **(D)** Thirdly, an open chromatin conformation of the enhancer element is confirmed by ATAC-seq in a human retina sample (GSE137311).¹⁴⁵ **(E)** Lastly, the enhancer element is enriched for histone-modification marks that are associated with enhancer activity such as H3K27Ac as determined using ChIP-seq performed in human retina (GSE137311).¹⁴⁵

Other factors that might complicate the interpretation of segregation analysis results are age-related or reduced penetrance, modifiers, carrier females in X-linked disease and multigenic inheritance. Several studies have indicated that modifying variants can have higher allele frequencies than fully penetrant alleles and therefore are not recognized by diagnostic pipelines.^{166,167} Despite their high allele frequencies, it has been shown that these variants can still significantly modify Mendelian genotypes. For instance, the variants p.(Ser192Tyr) and p.(Arg402Gln) in *TYR* have an individual allele frequency of 36.4% and 27.3% in the gnomAD database (non-Finnish Europeans), respectively, while the p.[Ser192Tyr;Arg402Gln] allele has an allele frequency of 1.9%. Despite the relatively high population frequency, the pathogenicity of the p.[Ser192Tyr;Arg402Gln] allele has been suggested when present in homozygous state or in a tri-allelic genotype with a known pathogenic *TYR* variant in *trans*.^{168,169} Studies suggest that one in six genes implicated in RD is possibly associated with variable penetrance due to variability in expression levels.^{170,171} Examples of strong evidence for variants with reduced penetrance implicated in RD or HL have been reported for *ABCA4*^{107,172}, *COCH*¹⁷³, *PRPF31*¹⁷⁴ and *RIPOR2*.¹⁷⁵

Another complicating factor is uniparental disomy (UPD) where two homologous chromosomes are inherited from the same parent due to errors during meiosis. In 2020, Yauy et al. investigated the presence of UPD in exome sequencing data of 4,912 trios.¹⁷⁶ The authors detected UDPs in 0.05-0.2% of these trios, amongst which was a chromosome 1 UPD (*ABCA4*) in a Stargardt disease case, suggesting minimal contribution to the genetic diagnostic yield.¹⁷⁶ Thus far, there are four reported Stargardt disease cases showing UPD in chromosome 1.^{41,176-178} Moreover, in 2013, Roosing et al. described maternal UPD of chromosome 6, which included a pathogenic *TULP1* variant responsible for the cone dystrophy phenotype of the proband.¹⁷⁹ For HL, several cases of UPD have been described as well, affecting chromosome 1 (*USH2A*)¹⁸⁰, chromosome 13 (*GJB2*)¹⁷⁸ and chromosome 18 (*LOXHD1*).^{178,181}

Functional evaluation of variants

Functional assays can provide an extra line of evidence that can aid in the discrimination between (likely) pathogenic variants, (likely) benign variants or variants with unknown significance. For proteins with a well characterized subcellular localization or function, *in vitro* approaches can be considered to assess the impact of the variant on protein localization or function. Examples of the latter are assessments of transporter function, enzymatic activity or activity of metabolic pathways. *In vivo* experiments are ideal for studying the true biological context. However, as it is not always feasible to perform such studies, *in vitro* research can instead provide valuable information. Biochemical

data obtained from patient-derived biopsies might be more informative. However, for both HL and RD, samples derived from the tissues of interest are usually not available. For these purposes, animal models could provide a valuable alternative. Over the years, several studies have proven the suitability of studying ear- or eye-related disease in non-human primates and mouse models.^{33,182} The International Mouse Phenotyping Consortium (IMPC) aims to generate mouse knockout models for all known genes in the mouse genome.¹⁸³ Furthermore, the zebrafish has proven its suitability as an animal model. In this model, retinal and inner ear function can already be studied five days post fertilization.^{182,184} Limitations of the usage of animal models include ethical, time and financial considerations, in addition to the level of gene conservation.

Stem cell technology and the development of differentiation protocols over the past decades have enabled the *in vitro* generation of patient-derived cells resembling retinal photoreceptors or inner ear hair cells.^{185,186} These models can provide an alternative to method of studying the tissue of interest. Research has shown that differentiated cells can resemble the patient's retina or inner ear. Several 2D- and 3D-differentiation protocols have been successfully applied to study both HL and RD. Differentiation approaches are rapidly being optimized, as the involved processes are still very time consuming and expensive.^{185,186} More so, variability and cell heterogeneity are important hurdles, and these should be overcome in order to fully replace animal model studies.

FUTURE DEVELOPMENTS

Development of new technologies

Chromosomal abnormalities and SVs are among the main causes of genetic diseases, which are being addressed in clinical application using routine cytogenetics methods, such as karyotyping and fluorescent *in situ* hybridization (FISH), comparative genomic hybridization (CGH), and SNP-microarrays.^{187,188} However, these methods manifest significant limitations in the identification of SVs. For example, karyotyping allows the identification of different chromosomal abnormalities with a 5-10 Mb resolution. Although microarrays and CGH-arrays are able to identify gain and loss of chromosomal material as small as 10 kb, balanced rearrangements cannot be detected by these methods nor the exact location of the structural variation.^{189,190} It is estimated that only 15-20% of chromosomal abnormalities can be detected by the application of these techniques, which indicates the great need for new technologies in the field of cytogenetics.¹⁹¹

Although LRS techniques are rapidly developing and show a great ability to identify SVs, their routine application in clinical diagnostics still requires several improvements in terms of sequencing, and variant interpretation; it also requires a cost reduction. In addition, despite the fact that these technologies can provide substantial read length, the reads can only be assembled to the scaffold-level and not to the chromosome-level.¹⁹² Complementary approaches to identify SVs can be offered by cytogenetics.¹⁹⁰ One of these recent technologies is optical genome mapping (Bionano Genomics), which is a de novo assembly-based method that allows the visualization of the genomic structure in high resolution.¹⁹³ The approach is based on ultra-long dsDNA molecules that are fluorescently labeled at CTTAAG hexanucleotide motifs, which are found on average 15 times per 100 kb across the human genome. The distances and patterns of these labels can be compared to those in a reference genome. Therefore, copy number aberrations and other SVs, including insertions, inversions and translocations, can be detected (**Figure 5**).

Optical genome mapping has a much higher resolution compared to standard karyotyping and microarray technologies, and therefore enables much more precise data analysis. As it is an imaging method and not a sequencing method, SNVs cannot be detected. However, for the analysis of SVs, optical genome mapping can be used in a complementary manner to sequencing techniques.¹⁹⁰ With the ability to map ultra-long dsDNA molecules at a low cost, optical genome mapping has facilitated SV detection, haplotype phasing, and genome assembly.¹⁹² In a recent study, researchers utilized optical genome mapping to identify a 48-kb duplication at the *LAMA1* locus, that causes Poretti-Boltshauser syndrome. Affected individuals present with ataxia, cognitive impairment, and language delay as well as ocular phenotypes such as ocular motor apraxia, abnormal eye movement, and RD. The WES and chromosome microarray pre-screening methods failed to reveal the large SV in the studied family.¹⁹⁴ The authors reasoned that LRS technologies offer promising applications in comprehensive SV analysis, however the costs and accuracy may represent a burden. Therefore, they suggested that a combination of different technologies such as optical genome mapping and SRS provides a more comprehensive understanding of SVs when considering cost, time, and throughput.¹⁹⁴

Multi-omic approaches

Besides genome sequencing, other omic-technologies, such as transcriptomics, proteomics, metabolomics or epigenomics, hold the promise to further close the diagnostic gap for RD and HL. It is evident that for each identified disease-associated gene, the isoform landscape and levels of involved gene regulation are more

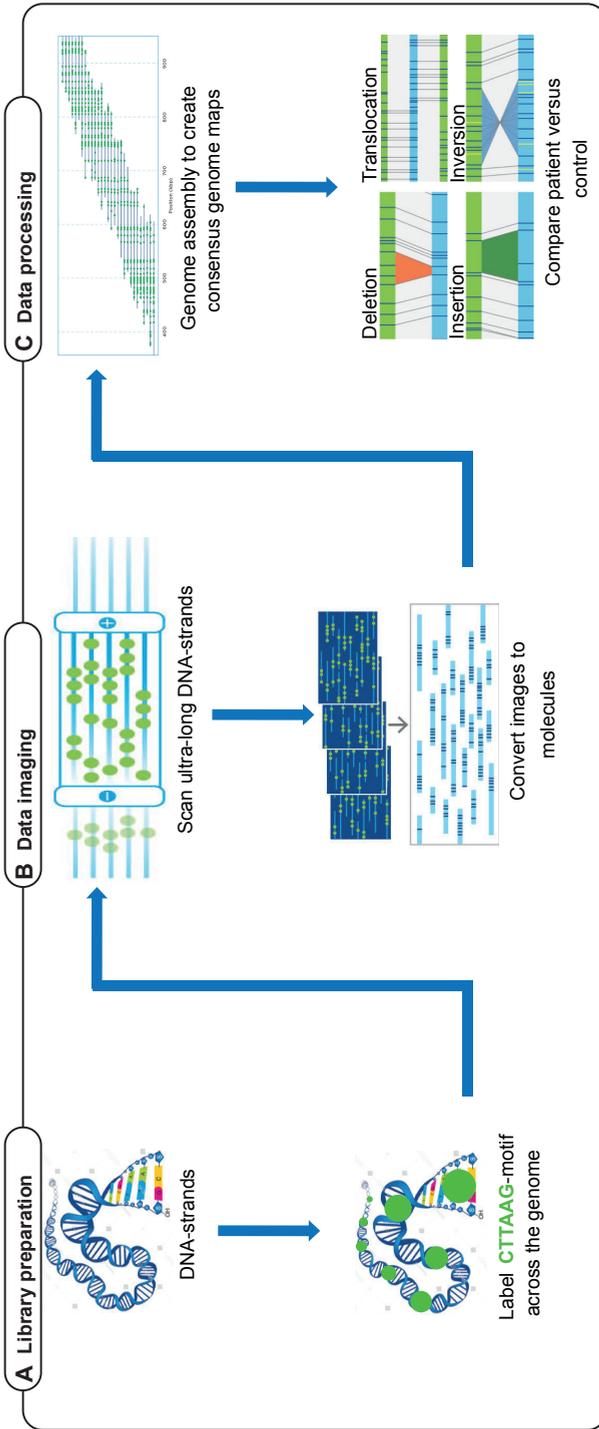


Figure 5. Overview of the optical genome mapping technology. (A) The high-quality DNA is isolated and later labeled at a 6-mer motif across the genome. (B) The labeled DNA is linearized in order to take images of the label patterns in the DNA molecules and subsequently, images converted to the molecules. (C) These molecules are then utilized for genome assembly to generate consensus genome maps. The pattern of labels can be compared between the reference genome and affected individuals to identify structural variants. The shorter or longer distance between two labels indicates deletion or insertion, respectively. Translocations can be identified by mapping of a single region in the patient genome to two genomic regions in the reference. The inverted pattern of labels in a patient compared to that in the reference genome indicates the presence of an inversion.

complex than initially thought. A quantitative (gene expression levels) or qualitative (isoform structures, novel exons) analysis of the transcriptomic landscape is valuable in enhancing diagnostic yield, as shown by several studies.^{195,196} In combination with genome sequencing, RNA sequencing can improve the interpretation of variants with unknown significance, although inaccessibility of cell types for RD and HL-associated genes is a major limitation.

LRS offers the potential for RNA analysis as well: for example, the Iso-Seq method of PacBio enables the sequencing of full transcripts and Nanopore offers the direct sequencing of RNA molecules.⁵⁹ LRS techniques have already shown to be successful in the identification of novel full-length transcripts. In a study performed by Ray et al., an abundant retina-specific *CRB1* transcript (*CRB1-B*) was detected which was not annotated in genome databases, such as the UCSC genome browser.^{164,197} The authors showed that the expression of the *CRB1-B* transcript is significantly higher in photoreceptors than the canonical *CRB1* transcript (*CRB1-A*). The newly identified transcript includes unique exons that are not present in *CRB1-A* and thereby represent important candidate regions for potentially missed pathogenic variants.¹⁹⁷ In addition, developments in the single-cell RNA sequencing field allow the identification and characterization of tissue-specific isoforms and regulatory events. The Single Cell Portal (Broad Institute) offers a valuable resource of tissue-specific single-cell RNA sequencing datasets.¹⁹⁸

Epigenomics is an emerging and promising development in the field of medical genetics. Analysis of epigenomic signatures can aid in the understanding of the 3D organization of the genome. Since base-modifications remain captured in native DNA molecules that are used for SMRT and Nanopore sequencing, investigation of the methylome and DNA base modification is possible.^{59,62} Ideally, multi-omics layers (e.g. genomics, transcriptomics and epigenomics) should be integrated (so called multi-omics), which aids in an ultimate understanding of the genomic landscape and provides valuable insights for (candidate) disease-associated genes.

CONCLUSIONS AND DISCUSSION

Fifty years after the arrival of the Sanger sequencing technique, the sequencing technology landscape is still rapidly evolving. However, genetic diagnostic yield still varies between 40-70% for inherited HL and RD, indicating that there are still opportunities for further improvement.^{3,4,47} Although novel disease-associated genes are being discovered, disease-gene identification curves are slowly reaching a plateau phase, suggesting more attention should be paid to currently missed or

misinterpreted variants within known HL- or RD-associated genes that reside within complex (non-coding) regions of the genome. Recent developments of LRS techniques and optical genome mapping, and improvements in WGS techniques offer valuable opportunities to investigate the non-coding landscape of the genome in more detail. Furthermore, the interpretation of SVs has been greatly advanced by developments in computational analysis and bioinformatics tools. Therefore, the emphasis will be on overcoming the limitations of the sequencing and bioinformatic techniques in the near future. Additionally, evidence suggests that complex factors, such as modifiers, digenic inheritance and variable penetrance, play an important role in disease-causing mechanisms in inherited HL and RD. The generation of larger high-quality datasets will allow a better understanding of these events as well.

We foresee that, in the near future, the new technologies and improved analytical tools will reinforce the clinical diagnostic setting in order to close the diagnostic gap, as it is of utmost importance for both the affected individuals and the involved clinicians and researchers. It will help to provide guidance to affected families with regard to family planning, providing them with an optimal prognosis and counseling. In addition, with recent developments in the field of genetic therapies, the importance of genetic diagnostics can no longer be underestimated. We have come a long way from linkage analysis starting in the early 90s, to the more recent LRS of single DNA molecules to unravel the genetic causes of HL and RD. Clinical diagnostics has significantly improved over these years, and the diagnostic yield is still increasing. We anticipate an extensive application of new technologies in the future, which will redirect traditional therapies towards precision or personalized medicine to improve treatments for HL and RD.

ACKNOWLEDGEMENTS

The authors would like to show their gratitude to the funding agencies for their support and to all the colleagues for their useful discussion. The work of Z.F. is funded by the Foundation Fighting Blindness USA Project Program Award, grant no. PPA-0517-0717-RAD (to F.P.M.C., S.R.). The work of S.E.d.B is funded by a DCMN Radboudumc grant (to F.P.M.C., H.K.). The research was supported by the Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, Oogfonds, Landelijke Stichting voor Blinden en Slechtzienden; Rotterdamse Stichting Blindenbelangen, Stichting Blindenhulp, Stichting tot Verbetering van het Lot der Blinden, and Stichting Blinden-Penning (to S.R. and F.P.M.C.). We thank Rebekkah Hitti-Malin for her expert input and proofreading.

REFERENCES

1. Van Camp, G. & Smith, R. Hereditary Hearing Loss Homepage. Available from: <https://hereditaryhearingloss.org/>.
2. RetNet. Available from: <https://sph.uth.edu/retnet/>.
3. Zazo Seco, C., Wesdorp, M., Feenstra, I., Pfundt, R., Hehir-Kwa, J.Y., Lelieveld, S.H. *et al.* The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. *European Journal of Human Genetics* **25**, 308-314 (2017).
4. Haer-Wigman, L., van Zelst-Stams, W.A., Pfundt, R., van den Born, L.I., Klaver, C.C., Verheij, J.B. *et al.* Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *European Journal of Human Genetics* **25**, 591-599 (2017).
5. Wesdorp, M., Murillo-Cuesta, S., Peters, T., Celaya, A.M., Oonk, A., Schraders, M. *et al.* MPZL2, encoding the epithelial junctional protein myelin protein zero-like 2, is essential for hearing in man and mouse. *American Journal of Human Genetics* **103**, 74-88 (2018).
6. Pierrache, L.H.M., Kimchi, A., Ratnapriya, R., Roberts, L., Astuti, G.D.N., Obolensky, A. *et al.* Whole-exome sequencing identifies biallelic IDH3A variants as a cause of retinitis pigmentosa accompanied by pseudocoloboma. *Ophthalmology* **124**, 992-1003 (2017).
7. McClellan, J. & King, M.C. Genetic heterogeneity in human disease. *Cell* **141**, 210-207 (2010).
8. Dawn Teare, M. & Barrett, J.H. Genetic linkage studies. *Lancet* **366**, 1036-1044 (2005).
9. Waters, A.M. & Beales, P.L. Ciliopathies: an expanding disease spectrum. *Pediatric Nephrology* **26**, 1039-1056 (2011).
10. Cremers, F.P.M., van de Pol, D.J., van Kerkhoff, L.P., Wieringa, B. & Ropers, H.H. Cloning of a gene that is rearranged in patients with choroïdæmia. *Nature* **347**, 674-677 (1990).
11. Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W. *et al.* A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**, 364-366 (1990).
12. de Kok, Y.J., van der Maarel, S.M., Bitner-Glindzicz, M., Huber, I., Monaco, A.P., Malcolm, S. *et al.* Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. *Science* **267**, 685-688 (1995).
13. Broadgate, S., Yu, J., Downes, S.M. & Halford, S. Unravelling the genetics of inherited retinal dystrophies: Past, present and future. *Progress in Retinal and Eye Research* **59**, 53-96 (2017).
14. Morton, N.E. Sequential tests for the detection of linkage. *American Journal of Human Genetics* **7**, 277-318 (1955).
15. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *American Journal of Human Genetics* **58**, 1347-1363 (1996).
16. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575 (2007).
17. Vona, B., Nanda, I., Hofrichter, M.A., Shehata-Dieler, W. & Haaf, T. Non-syndromic hearing loss gene identification: A brief history and glimpse into the future. *Molecular and Cellular Probes* **29**, 260-270 (2015).

18. Seelow, D., Schuelke, M., Hildebrandt, F. & Nürnberg, P. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic Acids Research* **37**, W593-W599 (2009).
19. Woods, C.G., Cox, J., Springell, K., Hampshire, D.J., Mohamed, M.D., McKibbin, M. *et al.* Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *American Journal of Human Genetics* **78**, 889-896 (2006).
20. Collin, R.W., van den Born, L.I., Klevering, B.J., de Castro-Miró, M., Littink, K.W., Arimadyo, K. *et al.* High-resolution homozygosity mapping is a powerful tool to detect novel mutations causative of autosomal recessive RP in the Dutch population. *Investigative Ophthalmology & Visual Science* **52**, 2227-2239 (2011).
21. Schraders, M., Lee, K., Oostrik, J., Huygen, P.L., Ali, G., Hoefsloot, L.H. *et al.* Homozygosity mapping reveals mutations of GRXCR1 as a cause of autosomal-recessive nonsyndromic hearing impairment. *American Journal of Human Genetics* **86**, 138-147 (2010).
22. Collin, R.W.J., Littink, K.W., Klevering, B.J., van den Born, L.I., Koenekoop, R.K., Zonneveld, M.N. *et al.* Identification of a 2 Mb human ortholog of *Drosophila* eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. *American Journal of Human Genetics* **83**, 594-603 (2008).
23. Messchaert, M., Haer-Wigman, L., Khan, M.I., Cremers, F.P.M. & Collin, R.W.J. EYS mutation update: in silico assessment of 271 reported and 26 novel variants in patients with retinitis pigmentosa. *Human Mutation* **39**, 177-186 (2018).
24. Thiadens, A.A., den Hollander, A.I., Roosing, S., Nabuurs, S.B., Zekveld-Vroon, R.C., Collin, R.W. *et al.* Homozygosity mapping reveals PDE6C mutations in patients with early-onset cone photoreceptor disorders. *American Journal of Human Genetics* **85**, 240-247 (2009).
25. Schraders, M., Ruiz-Palmero, L., Kalay, E., Oostrik, J., del Castillo, F.J., Sezgin, O. *et al.* Mutations of the gene encoding otogelin are a cause of autosomal-recessive nonsyndromic moderate hearing impairment. *American Journal of Human Genetics* **91**, 883-889 (2012).
26. Friedman, T.B., Liang, Y., Weber, J.L., Hinnant, J.T., Barber, T.D., Winata, S. *et al.* A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. *Nature Genetics* **9**, 86-91 (1995).
27. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**, 5463-5467 (1977).
28. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
29. Waterston, R.H., Lander, E.S. & Sulston, J.E. On the sequencing of the human genome. *Proceedings of the National Academy of Sciences USA* **99**, 3712-3716 (2002).
30. Roberts, P.J. Human genome project. *Annales Chirurgiae et Gynaecologiae* **90**, 3 (2001).
31. Venter, J.C., Smith, H.O. & Hood, L. A new strategy for genome sequencing. *Nature* **381**, 364-366 (1996).
32. Schloss, J.A. How to get genomes at one ten-thousandth the cost. *Nature Biotechnology* **26**, 1113-1115 (2008).
33. Vona, B., Müller, M., Dofek, S., Holderried, M., Löwenheim, H. & Tropitzsch, A. A big data perspective on the genomics of hearing loss. *Laryngo-rhino-otologie* **98**, S32-S81 (2019).

34. Levy, S.E. & Myers, R.M. Advancements in next-generation sequencing. *Annual Review of Genomics and Human Genetics* **17**, 95-115 (2016).
35. Smith, L.M., Fung, S., Hunkapiller, M.W., Hunkapiller, T.J. & Hood, L.E. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Research* **13**, 2399-2412 (1985).
36. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-9 (1986).
37. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51**, 263-273 (1986).
38. Buermans, H.P. & den Dunnen, J.T. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta* **1842**, 1932-1941 (2014).
39. Choi, B.Y., Park, G., Gim, J., Kim, A.R., Kim, B.J., Kim, H.S. *et al.* Diagnostic application of targeted resequencing for familial nonsyndromic hearing loss. *PLoS One* **8**, e68692 (2013).
40. Dockery, A., Stephenson, K., Keegan, D., Wynne, N., Silvestri, G., Humphries, P. *et al.* Target 5000: target capture sequencing for inherited retinal degenerations. *Genes (Basel)* **8**, 304 (2017).
41. Khan, M., Cornelis, S.S., Pozo-Valero, M.D., Whelan, L., Runhart, E.H., Mishra, K. *et al.* Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genetics in Medicine* **22**, 1235-1246 (2020).
42. Lin, X., Tang, W., Ahmad, S., Lu, J., Colby, C.C., Zhu, J. *et al.* Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hearing Research* **288**, 67-76 (2012).
43. Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences USA* **106**, 19096-19101 (2009).
44. Stenson, P.D., Mort, M., Ball, E.V., Evans, K., Hayden, M., Heywood, S. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics* **136**, 665-677 (2017).
45. Stranneheim, H. & Wedell, A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *Journal of Internal Medicine* **279**, 3-15 (2016).
46. Tucker, T., Marra, M. & Friedman, J.M. Massively parallel sequencing: the next big thing in genetic medicine. *American Journal of Human Genetics* **85**, 142-154 (2009).
47. Sloan-Heggen, C.M., Bierer, A.O., Shearer, A.E., Kolbe, D.L., Nishimura, C.J., Frees, K.L. *et al.* Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Human Genetics* **135**, 441-450 (2016).
48. Xu, Y., Guan, L., Shen, T., Zhang, J., Xiao, X., Jiang, H. *et al.* Mutations of 60 known causative genes in 157 families with retinitis pigmentosa based on exome sequencing. *Human Genetics* **133**, 1255-1271 (2014).

49. Tiwari, A., Bahr, A., Bähr, L., Fleischhauer, J., Zinkernagel, M.S., Winkler, N. *et al.* Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Scientific Reports* **6**, 28755 (2016).
50. Abu-Safieh, L., Alrashed, M., Anazi, S., Alkuraya, H., Khan, A.O., Al-Owain, M. *et al.* Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Research* **23**, 236-247 (2013).
51. Sang, S., Ling, J., Liu, X., Mei, L., Cai, X., Li, T. *et al.* Proband whole-exome sequencing identified genes responsible for autosomal recessive non-syndromic hearing loss in 33 Chinese nuclear families. *Frontiers in Genetics* **10**, 639 (2019).
52. Morton, C.C. & Nance, W.E. Newborn hearing screening--a silent revolution. *New England Journal of Medicine* **354**, 2151-2164 (2006).
53. Ellingford, J.M., Barton, S., Bhaskar, S., Williams, S.G., Sergouniotis, P.I., O'Sullivan, J. *et al.* Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology* **123**, 1143-1150 (2016).
54. Barbitoff, Y.A., Polev, D.E., Glotov, A.S., Serebryakova, E.A., Shcherbakova, I.V., Kiselev, A.M. *et al.* Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. *Scientific Reports* **10**, 2057 (2020).
55. Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q.B., Antipenko, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences USA* **112**, 5473-5478 (2015).
56. Carss, K.J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S. *et al.* Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *American Journal of Human Genetics* **100**, 75-90 (2017).
57. Wetterstrand, K. DNA sequencing costs: data from the NHGRI genome sequencing program (GSP).
58. Méjécasse, C., Malka, S., Guan, Z., Slater, A., Arno, G. & Moosajee, M. Practical guide to genetic screening for inherited eye diseases. *Therapeutic Advances in Ophthalmology* **12**, 1-28 (2020).
59. Mantere, T., Kersten, S. & Hoischen, A. Long-read sequencing emerging in medical genetics. *Frontiers in Genetics* **10**, 426 (2019).
60. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138 (2009).
61. Magi, A., Semeraro, R., Mingrino, A., Giusti, B. & D'Aurizio, R. Nanopore sequencing data analysis: state of the art, applications and challenges. *Briefings in Bioinformatics* **19**, 1256-1272 (2018).
62. van Dijk, E.L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The third revolution in sequencing technology. *Trends in Genetics* **34**, 666-681 (2018).
63. Schadt, E.E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Human Molecular Genetics* **19**, R227-240 (2010).
64. Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S. & Turner, S.W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Research* **38**, e159 (2010).

65. Ardui, S., Ameer, A., Vermeesch, J.R. & Hestand, M.S. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Research* **46**, 2159-2168 (2018).
66. Xu, M., Fujita, D. & Hanagata, N. Perspectives and challenges of emerging single-molecule DNA sequencing technologies. *Small* **5**, 2638-2649 (2009).
67. Chaisson, M.J.P., Huddleston, J., Dennis, M.Y., Sudmant, P.H., Malig, M., Hormozdiari, F. *et al.* Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608-611 (2015).
68. Seo, J.-S., Rhie, A., Kim, J., Lee, S., Sohn, M.-H., Kim, C.-U. *et al.* De novo assembly and phasing of a Korean human genome. *Nature* **538**, 243-247 (2016).
69. Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J.M., Couce, M.L. & Cocho, J.A. A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism* **110**, 3-24 (2013).
70. Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences USA* **106**, 7702-7707 (2009).
71. Kono, N. & Arakawa, K. Nanopore sequencing: review of potential applications in functional genomics. *Development, Growth & Differentiation* **61**, 316-326 (2019).
72. Levene, M.J., Korch, J., Turner, S.W., Foquet, M., Craighead, H.G. & Webb, W.W. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686 (2003).
73. Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T. *et al.* The potential and challenges of nanopore sequencing. *Nature Biotechnology* **26**, 1146-1153 (2008).
74. Ip, C.L.C., Loose, M., Tyson, J.R., de Cesare, M., Brown, B.L., Jain, M. *et al.* MinION Analysis and Reference Consortium: phase 1 data release and analysis. *F1000Research* **4**, 1075 (2015).
75. Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* **36**, 338-345 (2018).
76. Mardis, E.R. DNA sequencing technologies: 2006-2016. *Nature Protocols* **12**, 213-218 (2017).
77. Ebbert, M.T.W., Jensen, T.D., Jansen-West, K., Sens, J.P., Reddy, J.S., Ridge, P.G. *et al.* Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biology* **20**, 97 (2019).
78. Reiner, J., Pisani, L., Qiao, W., Singh, R., Yang, Y., Shi, L. *et al.* Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a Bardet-Biedl Syndrome 9 (BBS9) deletion. *NPJ Genomic Medicine* **3**, 3 (2018).
79. Bedoni, N., Quinodoz, M., Pinelli, M., Cappuccio, G., Torella, A., Nigro, V. *et al.* An Alu-mediated duplication in NMNAT1, involved in NAD biosynthesis, causes a novel syndrome, SHILCA, affecting multiple tissues and organs. *Human Molecular Genetics* **29**, 2250-2260 (2020).
80. Laurent, S., Gehrig, C., Nospikel, T., Amr, S.S., Oza, A., Murphy, E. *et al.* Molecular characterization of pathogenic OTOA gene conversions in hearing loss patients. *Human Mutation* **42**, 373-377 (2021).

81. Hafford-Tear, N.J., Tsai, Y.C., Sadan, A.N., Sanchez-Pintado, B., Zarouchlioti, C., Maher, G.J. *et al.* CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy-associated TCF4 triplet repeat. *Genetics in Medicine* **21**, 2092-2102 (2019).
82. Edwards, H.S., Krishnakumar, R., Sinha, A., Bird, S.W., Patel, K.D. & Bartsch, M.S. Real-time selective sequencing with RUBRIC: read until with basecall and reference-informed criteria. *Scientific Reports* **9**, 11475 (2019).
83. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
84. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J. *et al.* 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. *Genetics in Medicine* **17**, 405-424 (2015).
85. Eichler, E.E. Genetic variation, comparative genomics, and the diagnosis of disease. *New England Journal of Medicine* **381**, 64-74 (2019).
86. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
87. Tucci, S. & Akey, J.M. The long walk to African genomics. *Genome Biology* **20**, 130 (2019).
88. Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T. *et al.* High-depth African genomes inform human migration and health. *Nature* **586**, 741-748 (2020).
89. Boomsma, D.I., Wijmenga, C., Slagboom, E.P., Swertz, M.A., Karssen, L.C., Abdellaoui, A. *et al.* The Genome of the Netherlands: design, and project goals. *European Journal of Human Genetics* **22**, 221-227 (2014).
90. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).
91. Erikson, G.A., Bodian, D.L., Rueda, M., Molparia, B., Scott, E.R., Scott-Van Zeeland, A.A. *et al.* Whole-genome sequencing of a healthy aging cohort. *Cell* **165**, 1002-1011 (2016).
92. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J. *et al.* ClinGen—the clinical genome resource. *New England Journal of Medicine* **372**, 2235-2242 (2015).
93. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research* **42**, D980-D985 (2014).
94. Fokkema, I.F., Taschner, P.E., Schaafsma, G.C., Celli, J., Laros, J.F. & den Dunnen, J.T. LOVD v.2.0: the next generation in gene variant databases. *Human Mutation* **32**, 557-563 (2011).
95. Azaiez, H., Booth, K.T., Ephraim, S.S., Crone, B., Black-Ziegelbein, E.A., Marini, R.J. *et al.* Genomic landscape and mutational signatures of deafness-associated genes. *American Journal of Human Genetics* **103**, 484-497 (2018).
96. Cremers, F.P.M., den Dunnen, J.T., Ajmal, M., Hussain, A., Preising, M.N., Daiger, S.P. *et al.* Comprehensive registration of DNA sequence variants associated with inherited retinal diseases in Leiden Open Variation Databases. *Human Mutation* **35**, 147-148 (2014).

97. Cornelis, S.S., Bax, N.M., Zernant, J., Allikmets, R., Fritsche, L.G., den Dunnen, J.T. *et al.* In silico functional meta-analysis of 5,962 ABCA4 variants in 3,928 retinal dystrophy cases. *Human Mutation* **38**, 400-408 (2017).
98. Baux, D., Blanchet, C., Hamel, C., Meunier, I., Larrieu, L., Faugère, V. *et al.* Enrichment of LOVD-USHbases with 152 USH2A genotypes defines an extensive mutational spectrum and highlights missense hotspots. *Human Mutation* **35**, 1179-1186 (2014).
99. Astuti, G.D., Bertelsen, M., Preising, M.N., Ajmal, M., Lorenz, B., Faradz, S.M. *et al.* Comprehensive genotyping reveals RPE65 as the most frequently mutated gene in Leber congenital amaurosis in Denmark. *European Journal of Human Genetics* **24**, 1071-1079 (2016).
100. Mackay, D.S., Borman, A.D., Sui, R., van den Born, L.I., Berson, E.L., Ocaka, L.A. *et al.* Screening of a large cohort of leber congenital amaurosis and retinitis pigmentosa patients identifies novel LCA5 mutations and new genotype-phenotype correlations. *Human Mutation* **34**, 1537-1546 (2013).
101. Ellingford, J.M., Thomas, H.B., Rowlands, C., Arno, G., Beaman, G., Gomes-Silva, B. *et al.* Functional and in-silico interrogation of rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *bioRxiv* (2019).
102. Hanany, M. & Sharon, D. Allele frequency analysis of variants reported to cause autosomal dominant inherited retinal diseases question the involvement of 19% of genes and 10% of reported pathogenic variants. *Journal of Medical Genetics* **56**, 536-542 (2019).
103. Oza, A.M., DiStefano, M.T., Hemphill, S.E., Cushman, B.J., Grant, A.R., Siegert, R.K. *et al.* Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Human Mutation* **39**, 1593-1613 (2018).
104. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research* **43**, D789-D798 (2015).
105. McHugh, R.K. & Friedman, R.A. Genetics of hearing loss: allelism and modifier genes produce a phenotypic continuum. *Anatomical Record Part A Discoveries in Molecular Cellular and Evolutionary Biology* **288**, 370-81 (2006).
106. Cremers, F.P.M., Lee, W., Collin, R.W.J. & Allikmets, R. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. *Progress in Retinal and Eye Research* **79**, 100861 (2020).
107. Runhart, E.H., Khan, M., Cornelis, S.S., Roosing, S., Del Pozo-Valero, M., Lamey, T.M. *et al.* Association of sex with frequent and mild ABCA4 alleles in Stargardt disease. *JAMA ophthalmology* **138**, 1035-1042 (2020).
108. Runhart, E.H., Valkenburg, D., Cornelis, S.S., Khan, M., Sangermano, R., Albert, S. *et al.* Late-onset Stargardt disease due to mild, deep-intronic ABCA4 alleles. *Investigative Ophthalmology & Visual Science* **60**, 4249-4256 (2019).
109. Astuto, L.M., Bork, J.M., Weston, M.D., Askew, J.W., Fields, R.R., Orten, D.J. *et al.* CDH23 mutation and phenotype heterogeneity: A profile of 107 diverse families with Usher syndrome and nonsyndromic deafness. *American Journal of Human Genetics* **71**, 262-275 (2002).

110. Becirovic, E., Ebermann, I., Nagy, D., Zrenner, E., Seeliger, M.W. & Bolz, H.J. Usher syndrome type 1 due to missense mutations on both CDH23 alleles: investigation of mRNA splicing. *Human Mutation* **29**, 452 (2008).
111. Zhang, L., Cheng, J., Zhou, Q., Khan, M.A., Fu, J., Duan, C. *et al.* Targeted next-generation sequencing identified novel compound heterozygous variants in the CDH23 gene causing Usher syndrome type ID in a Chinese patient. *Frontiers in Genetics* **11**, 422 (2020).
112. Molina-Ramírez, L.P., Lenassi, E., Ellingford, J.M., Sergouniotis, P.I., Ramsden, S.C., Bruce, I.A. *et al.* Establishing genotype-phenotype correlation in USH2A-related disorders to personalize audiological surveillance and rehabilitation. *Otology & Neurotology* **41**, 431-437 (2020).
113. Rivolta, C., Sweklo, E.A., Berson, E.L. & Dryja, T.P. Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss. *American Journal of Human Genetics* **66**, 1975-1978 (2000).
114. Tatour, Y. & Ben-Yosef, T. Syndromic inherited retinal diseases: Genetic, clinical and diagnostic aspects. *Diagnostics (Basel)* **10**, 779 (2020).
115. Gettelfinger, J.D. & Dahl, J.P. Syndromic hearing loss: A brief review of common presentations and genetics. *Journal of Pediatric Genetics* **7**, 1-8 (2018).
116. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics* **54**, 1.30.1-1.30.33 (2016).
117. Shen, J., Scheffer, D.I., Kwan, K.Y. & Corey, D.P. SHIELD: an integrative gene expression database for inner ear research. *Database (Oxford)* **2015**, bav071 (2015).
118. gEAR portal. Available from: <https://umgear.org/>.
119. Ratnapriya, R., Sosina, O.A., Starostik, M.R., Kwicklis, M., Kapphahn, R.J., Fritsche, L.G. *et al.* Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nature Genetics* **51**, 606-610 (2019).
120. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**, D607-D613 (2019).
121. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Human Mutation* **36**, 928-930 (2015).
122. The European Retinal Disease Consortium (ERDC). Available from: <https://www.erd.c.info/>.
123. Di Stazio, M., Morgan, A., Brumat, M., Bassani, S., Dell'Orco, D., Marino, V. *et al.* New age-related hearing loss candidate genes in humans: an ongoing challenge. *Gene* **742**, 144561 (2020).
124. Astuti, G.D.N., van den Born, L.I., Khan, M.I., Hamel, C.P., Bocquet, B., Manes, G. *et al.* Identification of inherited retinal disease-associated genetic variants in 11 candidate genes. *Genes (Basel)* **9**, 21 (2018).
125. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C. *et al.* A structural variation reference for medical and population genetics. *Nature* **581**, 444-451 (2020).

126. Littink, K.W., Pott, J.W., Collin, R.W., Kroes, H.Y., Verheij, J.B., Blokland, E.A. *et al.* A novel nonsense mutation in CEP290 induces exon skipping and leads to a relatively mild retinal phenotype. *Investigative Ophthalmology & Visual Science* **51**, 3646-3652 (2010).
127. Roosing, S., Cremers, F.P.M., Riemsdag, F.C.C., Zonneveld-Vrieling, M.N., Talsma, H.E., Klessens-Godfroy, F.J.M. *et al.* A rare form of retinal dystrophy caused by hypomorphic nonsense mutations in CEP290. *Genes (Basel)* **8**, 208 (2017).
128. DiStefano, M.T., Hemphill, S.E., Cushman, B.J., Bowser, M.J., Hynes, E., Grant, A.R. *et al.* Curating clinically relevant transcripts for the interpretation of sequence variants. *Journal of Molecular Diagnostics* **20**, 789-801 (2018).
129. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. & Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310-315 (2014).
130. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862 (1974).
131. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**, 361 (2014).
132. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research* **20**, 110-121 (2010).
133. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249 (2010).
134. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. & Ng, P.C. SIFT missense predictions for genomes. *Nature Protocols* **11**, 1 (2015).
135. Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M. & Bérout, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research* **37**, E67-E67 (2009).
136. Shapiro, M.B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research* **15**, 7155-74 (1987).
137. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology* **11**, 377-394 (2004).
138. Pertea, M., Lin, X. & Salzberg, S.L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Research* **29**, 1185-1190 (2001).
139. Reese, M.G., Eeckman, F.H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *Journal of Computational Biology* **4**, 311-323 (1997).
140. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I. *et al.* Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535-548 (2019).
141. Sangermano, R., Khan, M., Cornelis, S.S., Richelle, V., Albert, S., Garanto, A. *et al.* ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Research* **28**, 100-110 (2018).

142. Collin, R.W., de Heer, A.M., Oostrik, J., Pauw, R.J., Plantinga, R.F., Huygen, P.L. *et al.* Mid-frequency DFNA8/12 hearing loss caused by a synonymous TECTA mutation that affects an exonic splice enhancer. *European Journal of Human Genetics* **16**, 1430-1436 (2008).
143. Riepe, T.V., Khan, M., Roosing, S., Cremers, F.P.M. & t Hoen, P.A.C. Benchmarking deep learning splice prediction tools using functional splice assays. *Human Mutation* **Online ahead of print** (2021).
144. Rowlands, C.F., Baralle, D. & Ellingford, J.M. Machine learning approaches for the prioritization of genomic variants impacting pre-mRNA splicing. *Cells* **8**, 1513 (2019).
145. Cherry, T.J., Yang, M.G., Harmin, D.A., Tao, P., Timms, A.E., Bauwens, M. *et al.* Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. *Proceedings of the National Academy of Sciences USA* **117**, 9001-9012 (2020).
146. van der Lee, R., Correard, S. & Wasserman, W.W. Deregulated regulators: disease-causing cis variants in transcription factor genes. *Trends in Genetics* **36**, 523-539 (2020).
147. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
148. de Kok, Y.J.M., Vossenaar, E.R., Cremers, C.W.R.J., Dahl, N., Laporte, J., Jia Hu, L. *et al.* Identification of a hot spot for microdeletions in patients with x-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Human Molecular Genetics* **5**, 1229-1235 (1996).
149. Naranjo, S., Voesenek, K., de la Calle-Mustienes, E., Robert-Moreno, A., Kokotas, H., Grigoriadou, M. *et al.* Multiple enhancers located in a 1-Mb region upstream of POU3F4 promote expression during inner ear development and may be required for hearing. *Human Genetics* **128**, 411-419 (2010).
150. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **48**, D87-D92 (2020).
151. Perez-Cervantes, C., Smith, L.A., Nadadur, R.D., Hughes, A.E.O., Wang, S., Corbo, J.C. *et al.* Enhancer transcription identifies cis-regulatory elements for photoreceptor cell types. *Development* **147**, 1-13 (2020).
152. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research* **46**, D794-D801 (2018).
153. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, bax028 (2017).
154. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Research* **48**, D58-D64 (2020).
155. de Bruijn, S.E., Fiorentino, A., Ottaviani, D., Fanucchi, S., Melo, U.S., Corral-Serrano, J.C. *et al.* Structural variants create new topological-associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa. *American Journal of Human Genetics* **107**, 802-814 (2020).

156. Brandt, T., Sack, L.M., Arjona, D., Tan, D., Mei, H., Cui, H. *et al.* Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number variants. *Genetics in Medicine* **22**, 336-344 (2020).
157. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
158. Spielmann, M., Lupiáñez, D.G. & Mundlos, S. Structural variation in the 3D genome. *Nature Reviews Genetics* **19**, 453-467 (2018).
159. Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269 (2016).
160. Ibrahim, D.M. & Mundlos, S. Three-dimensional chromatin in disease: What holds us together and what drives us apart? *Current Opinion in Cell Biology* **64**, 1-9 (2020).
161. Van Schil, K., Naessens, S., Van de Sompele, S., Carron, M., Aslanidis, A., Van Cauwenbergh, C. *et al.* Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genetics in Medicine* **20**, 202-213 (2018).
162. Shearer, A.E., Kolbe, D.L., Azaiez, H., Sloan, C.M., Frees, K.L., Weaver, A.E. *et al.* Copy number variants are a common cause of non-syndromic hearing loss. *Genome Medicine* **6**, 37 (2014).
163. World Health Organisation. Deafness and hearing loss: Key facts 2021 update. (Geneva: World Health Organization, 2021).
164. Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M. *et al.* ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research* **41**, D56-D63 (2013).
165. Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C.C. *et al.* Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Research* **47**, D752-D758 (2018).
166. Nikopoulos, K., Cisarova, K., Quinodoz, M., Koskiniemi-Kuendig, H., Miyake, N., Farinelli, P. *et al.* A frequent variant in the Japanese population determines quasi-Mendelian inheritance of rare retinal ciliopathy. *Nature Communications* **10**, 2884 (2019).
167. Yan, D. & Liu, X.-Z. Modifiers of hearing impairment in humans and mice. *Current Genomics* **11**, 269-278 (2010).
168. Norman, C.S., O'Gorman, L., Gibson, J., Pengelly, R.J., Baralle, D., Ratnayaka, J.A. *et al.* Identification of a functionally significant tri-allelic genotype in the Tyrosinase gene (TYR) causing hypomorphic oculocutaneous albinism (OCA1B). *Scientific Reports* **7**, 4415 (2017).
169. Grønskov, K., Jespersgaard, C., Bruun, G.H., Harris, P., Brøndum-Nielsen, K., Andresen, B.S. *et al.* A pathogenic haplotype, common in Europeans, causes autosomal recessive albinism and uncovers missing heritability in OCA1. *Scientific Reports* **9**, 645 (2019).
170. Green, D.J., Sallah, S.R., Ellingford, J.M., Lovell, S.C. & Sergouniotis, P.I. Variability in gene expression is associated with incomplete penetrance in inherited eye disorders. *Genes (Basel)* **11**, 179 (2020).
171. Llavona, P., Pinelli, M., Mutarelli, M., Marwah, V.S., Schimpf-Linzenbold, S., Thaler, S. *et al.* Allelic expression imbalance in the human retinal transcriptome and potential impact on inherited retinal diseases. *Genes (Basel)* **8**, 283 (2017).

172. Runhart, E.H., Sangermano, R., Cornelis, S.S., Verheij, J., Plomp, A.S., Boon, C.J.F. *et al.* The common ABCA4 variant p.Asn1868Ile shows nonpenetrance and variable expression of Stargardt disease when present in trans with severe variants. *Investigative Ophthalmology & Visual Science* **59**, 3220-3231 (2018).
173. Smits, J.J., van Beelen, E., Weegerink, N.J.D., Oostrik, J., Huygen, P.L.M., Beynon, A.J. *et al.* A novel COCH mutation affects the vWFA2 domain and leads to a relatively mild DFNA9 phenotype. *Otology & Neurotology* **42**, e399-e407 (2021).
174. Vithana, E.N., Abu-Safieh, L., Pelosini, L., Winchester, E., Hornan, D., Bird, A.C. *et al.* Expression of PRPF31 mRNA in patients with autosomal dominant retinitis pigmentosa: a molecular clue for incomplete penetrance? *Investigative Ophthalmology & Visual Science* **44**, 4204-4209 (2003).
175. de Bruijn, S.E., Smits, J.J., Liu, C., Lanting, C.P., Beynon, A.J., Blankevoort, J. *et al.* A RIPOR2 in-frame deletion is a frequent and highly penetrant cause of adult-onset hearing loss. *Journal of Medical Genetics* **58**, 96-104 (2021).
176. Yauy, K., de Leeuw, N., Yntema, H.G., Pfundt, R. & Gilissen, C. Accurate detection of clinically relevant uniparental disomy from exome sequencing data. *Genetics in Medicine* **22**, 803-808 (2020).
177. Fingert, J.H., Eliason, D.A., Phillips, N.C., Lotery, A.J., Sheffield, V.C. & Stone, E.M. Case of Stargardt disease caused by uniparental isodisomy. *Archives of Ophthalmology* **124**, 744-745 (2006).
178. Alvarez, A., del Castillo, I., Pera, A., Villamar, M., Moreno-Pelayo, M.A., Rivera, T. *et al.* Uniparental disomy of chromosome 13q causing homozygosity for the 35delG mutation in the gene encoding connexin26 (GJB2) results in prelingual hearing impairment in two unrelated Spanish patients. *Journal of Medical Genetics* **40**, 636-639 (2003).
179. Roosing, S., van den Born, L.I., Hoyng, C.B., Thiadens, A.A., de Baere, E., Collin, R.W. *et al.* Maternal uniparental isodisomy of chromosome 6 reveals a TULP1 mutation as a novel cause of cone dysfunction. *Ophthalmology* **120**, 1239-1246 (2013).
180. Fu, J., Shen, S., Cheng, J., Lv, H. & Fu, J. A case of Usher syndrome type IIA caused by a rare USH2A homozygous frameshift variant with maternal uniparental disomy (UPD) in a Chinese family. *Journal of Cellular and Molecular Medicine* **24**, 7743-7750 (2020).
181. Morgan, A., Lenarduzzi, S., Cappellani, S., Pecile, V., Morgutti, M., Orzan, E. *et al.* Genomic studies in a large cohort of hearing impaired Italian patients revealed several new alleles, a rare case of uniparental disomy (UPD) and the importance to search for copy number variations. *Frontiers in Genetics* **9**, 681 (2018).
182. Slijkerman, R.W., Song, F., Astuti, G.D., Huynen, M.A., van Wijk, E., Stieger, K. *et al.* The pros and cons of vertebrate animal models for functional and therapeutic research on inherited retinal dystrophies. *Progress in Retinal and Eye Research* **48**, 137-159 (2015).
183. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508-514 (2016).
184. Vona, B., Doll, J., Hofrichter, M.A.H., Haaf, T. & Varshney, G.K. Small fish, big prospects: using zebrafish to unravel the mechanisms of hereditary hearing loss. *Hearing Research* **397**, 107906 (2020).

185. Tang, P.C., Hashino, E. & Nelson, R.F. Progress in modeling and targeting inner ear disorders with pluripotent stem cells. *Stem Cell Reports* **14**, 996-1008 (2020).
186. Kruczek, K. & Swaroop, A. Pluripotent stem cell-derived retinal organoids for disease modeling and development of therapies. *Stem Cells* **38**, 1206-1215 (2020).
187. Vissers, L.E., Veltman, J.A., van Kessel, A.G. & Brunner, H.G. Identification of disease genes by whole genome CGH arrays. *Human Molecular Genetics* **14**, R215-223 (2005).
188. Cui, C., Shu, W. & Li, P. Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Frontiers in Cell and Developmental Biology* **4**, 89 (2016).
189. Hyon, C. Usefulness of CGH-array and SNP-array for the etiological diagnosis of premature ovarian insufficiency. *Biologie Aujourd'hui* **211**, 199-205 (2017).
190. Mantere, T., Neveling, K., Pebrel-Richard, C., Benoist, M., van der Zande, G., Kater-Baats, E. *et al.* Next generation cytogenetics: genome-imaging enables comprehensive structural variant detection for 100 constitutional chromosomal aberrations in 85 samples. *bioRxiv* (2020).
191. Lee, C.N., Lin, S.Y., Lin, C.H., Shih, J.C., Lin, T.H. & Su, Y.N. Clinical utility of array comparative genomic hybridisation for prenatal diagnosis: a cohort study of 3171 pregnancies. *BJOG* **119**, 614-625 (2012).
192. Yuan, Y., Chung, C.Y. & Chan, T.F. Advances in optical mapping for genomic research. *Computational and Structural Biotechnology Journal* **18**, 2051-2062 (2020).
193. Chan, S., Lam, E., Saghbini, M., Bocklandt, S., Hastie, A., Cao, H. *et al.* Structural variation detection and analysis using bionano optical mapping. *Methods in Molecular Biology* **1833**, 193-203 (2018).
194. Chen, M., Zhang, M., Qian, Y., Yang, Y., Sun, Y., Liu, B. *et al.* Identification of a likely pathogenic structural variation in the LAMA1 gene by Bionano optical mapping. *NPJ Genomic Medicine* **5**, 31 (2020).
195. Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science Translational Medicine* **9**, eaal5209 (2017).
196. Kremer, L.S., Bader, D.M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature Communications* **8**, 15824 (2017).
197. Ray, T.A., Cochran, K., Kozlowski, C., Wang, J., Alexander, G., Cady, M.A. *et al.* Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. *Nature Communications* **11**, 3328 (2020).
198. Single Cell Portal (Broad Institute). Available from: https://singlecell.broadinstitute.org/single_cell.



Chapter 2

Homozygous variants in *KIAA1549*, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa

Suzanne E. de Bruijn, Sanne K. Verbakel, Erik de Vrieze, Hannie Kremer, Frans P.M. Cremers,
Carel B. Hoyng, L. Ingeborgh van den Born, Susanne Roosing

Journal of Medical Genetics **55**, 705-712 (2018)

ABSTRACT

Retinitis pigmentosa (RP) shows substantial genetic heterogeneity. It has been estimated that in approximately 60-80% of RP cases the genetic diagnosis can be found using whole exome sequencing (WES). In this study, the purpose was to identify causative variants in individuals with genetically unexplained retinal disease, which included one consanguineous family with two affected siblings and one case with RP. To identify the genetic defect, WES was performed in both probands and clinical analysis was performed. To obtain insight into the function of KIAA1549 in photoreceptors, mRNA expression, knockdown and protein localization studies were performed. Through analysis of WES data, based on population allele frequencies, and *in silico* prediction tools we identified a homozygous missense variant and a homozygous frameshift variant in *KIAA1549* that segregate in two unrelated families. KIAA1549 was found to localize at the connecting cilium of the photoreceptor cells and the synapses of the mouse retina. Both variants affect the long transcript of *KIAA1549* which encodes a 1950-amino acid protein and shows prominent brain expression. The shorter transcript encodes a 734-amino acid protein with a high retinal expression and is affected by the identified missense variant. Strikingly, knockdown of the long transcript also leads to decreased expression of the short transcript likely explaining the nonsyndromic retinal phenotype caused by the two variants targeting different transcripts. In conclusion, our results underscore the causality of segregating variants in *KIAA1549* for autosomal recessive RP. Moreover, our data indicate that KIAA1549 plays a role in photoreceptor function.

INTRODUCTION

Retinitis pigmentosa (RP) encompasses a clinical and genetic heterogeneous group of progressive inherited retinal diseases (IRD). RP is characterized by the primary degeneration of rod photoreceptor cells, followed by the loss of cone photoreceptor cells and retinal pigment epithelium (RPE). With a prevalence of approximately 1 in 4,000 persons, it is considered the most common form of IRD.¹ RP typically displays night blindness in early adulthood or adolescence, followed by the progressive loss of the peripheral visual field. The visual acuity can be relatively preserved until the advanced disease stages, but RP leads to severe visual impairment or blindness in a large number of patients.²

Besides clinical heterogeneity, RP is also characterized by its broad range of genetic heterogeneity. A large number of genes has been implicated in the pathogenesis of RP, and pathogenic variants can be inherited in a recessive, dominant or X-linked manner (RetNet).³ Recently, it has been estimated that in only 60-80% of RP cases the genetic explanation can be found using whole exome sequencing (WES), which is currently the most widely applied method for disease gene identification.^{4,5} A better understanding of the underlying disease mechanisms, the role of variants in the pathogenesis of disease in currently known RP genes, and genotype-phenotype correlations are required to provide further insights towards developing therapeutic approaches.

Recently, *KIAA1549* (GenBank: NM_001164665) has been proposed as a candidate RP gene; however, supporting evidence is limited. In an autosomal recessive RP (arRP) family a homozygous frameshift variant in *KIAA1549* was described to be the only variant remaining after applying filtering criteria on WES data.⁵

In this study, we report on homozygous variants in *KIAA1549* in two families with arRP. In addition, protein localization studies have been performed to provide insight in the involvement of *KIAA1549* in photoreceptor function, supporting its role as an RP gene.

MATERIALS AND METHODS

Subjects and clinical examinations

Two families with individuals with genetically unexplained RP were included in this study; one consanguineous Iranian family with two affected siblings, and one case

from the Netherlands. This study was approved by the Institutional Review Boards of the participating centers, and adhered to the tenets of the declaration of Helsinki. All subjects provided informed consent prior to inclusion in the study.

Clinical data were collected from the medical records of two patients from Family A (A-II:1, A-II:2) and one patient from Family B (B-II:1), including information regarding best-corrected Snellen visual acuity, and results of slit-lamp biomicroscopy and ophthalmoscopy. In patient A-II:2 and B-II:1, fundus photography, spectral-domain optical tomography (SD-OCT; Spectralis, Heidelberg Engineering) and Goldmann kinetic perimetry were performed, and full-field electroretinography was recorded according to the International Society for Clinical Electrophysiology of Vision guidelines and assessed by applying local standard values.⁶ In addition, fundus autofluorescence (Spectralis, Heidelberg Engineering) images, were available for patient B-II:1.

Whole exome sequencing and variant interpretation

Genomic DNA was isolated from peripheral blood using standard isolation methods and WES was performed in both probands. For proband A-II:2, exome enrichment was performed using the Agilent SureSelect Human All Exome V6 kit. Read mapping along the hg19 reference genome (GrCH37/hg19) and variant calling were performed using BWA version 0.78⁷ and the haplotype caller module of GATK (Broad Institute).⁸ Copy number variant (CNV) detection was performed using CoNIFER version 0.2.2.⁹ Exome enrichment for proband B-II:1 was carried out with the Agilent SureSelect XT Human All Exon V5 enrichment kit. Mapping of sequencing reads along the hg19 reference genome and variant calling were performed using Lifescope version 1.3 (Life Technologies). CNV detection was performed using ExomeDepth version 1.1.1.¹⁰

For both datasets, the obtained variants were filtered based on population allele frequencies $\leq 0.5\%$ in gnomAD¹¹, ExAC¹², dbSNP¹³, and an in-house exome database (containing 15,576 alleles). Only nonsense, indels, splice site (-14/+14 nucleotides), missense and synonymous variants were assessed. Missense variants were only assessed when predicted to be possibly pathogenic by at least one *in silico* predictor; a Grantham score ≥ 80 , PhyloP ≥ 2.7 or CADD-Phred score ≥ 15 .¹⁴ Synonymous variants were only assessed when predicted to have an effect on splicing by one of the splice prediction tools that are embedded in the AlamutVisual software (version 2.10). Candidate genes in which remaining variants were found were compared to currently known IRD-associated genes listed on RetNet³ (accessed on 1st June 2018). Validation of found variants and segregation analysis were performed by Sanger sequencing. Primer sequences and PCR conditions are available upon request.

KIAA1549 expression in human tissues

KIAA1549 expression was determined in human adult tissues using commercially available cDNA panels. Total RNA derived from heart, lung, brain, kidney and bone marrow (Bio-Chain) and total RNA derived from skeletal muscle, liver, duodenum, stomach, spleen, thymus and testis (Stratagene) were utilized. Total RNA from retina was obtained from a healthy anonymous donor. Subsequently, cDNA was prepared using the iScript cDNA Synthesis kit (Bio-RaD) and purified with NucleoSpin Gel and PCR Clean-up Columns (Machery-Nagel). Quantitative PCR was performed using GoTaq qPCR Master Mix (Promega) according to manufacturer's protocol. Transcript-specific intron-spanning primers have been designed and validated for the long (NM_001164665) and short transcript (XM_935390) of *KIAA1549*, and for the reference gene *GUSB*. Primer locations and sequences can be found in **Table S1**. Amplifications were performed with the Applied Biosystem Fast 7900 System (Applied Biosystems). All PCR reactions were executed in duplicate and relative gene expression levels compared to the reference gene *GUSB* were determined with the delta-delta Ct method.

Immunofluorescence of KIAA1549 in mouse retinal sections

An eye obtained from a healthy 2-month-old mouse was dissected and cryoprotected for 30 minutes with 10% sucrose in PBS before embedding Tissue-Tek OCT (Sakura). Subsequently, sections were frozen in isopentane cooled by liquid nitrogen. For immunofluorescence, unfixed cryosections (7 μm) were permeabilized in 0.01% Tween20 in PBS for 20 minutes. After washing with PBS, blocking was performed for 1 hour using a blocking solution containing 0.1% ovalbumin and 0.5% fish gelatin in PBS. Primary antibodies against *KIAA1549* (1:500; cat.# HPA019560, Sigma-Aldrich) and Centrin (1:500; cat.# 04-1624, Millipore) were diluted in blocking solution and incubated on the sections overnight at 4°C. Subsequently, sections were rinsed with PBS and incubated with secondary antibodies goat-anti-rabbit Alexa 568 and goat-anti-mouse Alexa 488 (1:500; Molecular Probes) and DAPI (1:8000; Molecular Probes) in blocking solution for 45 minutes. Finally, sections were post-fixed with 4% paraformaldehyde (PFA) for 10 minutes before mounting with Prolong Gold (Molecular Probes). Sections were analyzed using a Zeiss Axio Imager Z2 fluorescence microscope equipped with an Apotome using several magnifications.

Knockdown of KIAA1549 in vitro using siRNAs

Silencer® Select siRNAs targeting *KIAA1549* (s33562 and s33563) and non-targeting Negative Control No. 1 were obtained from Thermo Fisher Scientific (**Table S2**). For transfection, hTERT-RPE1 or HEK293T cells (ATCC) were transfected with a single siRNA in duplicate (15 nM final concentration), using Lipofectamine RNAiMax transfection

reagent (Thermo Fisher Scientific) according to manufacturer's protocol. After 24 hours of transfection, cells were serum starved (0.2% FCS) for 48 hours to induce ciliogenesis. To assess the effect of the siRNAs on *KIAA1549* expression, RNA was isolated using the NucleoSpin RNA kit (Macherey-Nagel), and expression was quantified by qPCR. To evaluate the effect of knockdown of the long transcript on expression of the short transcript, HEK293T cells were used. HEK293T cells express both the long and short transcript abundantly, unlike hTERT-RPE1 cells which only express the long transcript.

For immunofluorescence, transfected hTERT-RPE1 cells were fixed with 2% PFA for 20 minutes, and permeabilized using 1% Triton X-100 in PBS for 5 minutes. Subsequently, cells were blocked with 2% bovine serum albumin (BSA) in PBS for 45 minutes. Primary antibodies against the primary cilium (anti-ARL13B; 1:500; cat.# 17711-1AP; ProteinTech) and the ciliary transition zone (anti-RPGRIP1L; 1:500; cat.# SNC039;⁶) diluted in blocking solution were incubated for 1 hour. After incubation with secondary antibodies in blocking solution for 45 minutes, samples were mounted by VECTASHIELD containing DAPI (Vector Laboratories). Cells were imaged using a Zeiss Axio Imager Z2 fluorescence microscope and a 63x magnification. Percentage of ciliated cells and cilium length were calculated using Fiji Is Just ImageJ (FIJI).¹⁵ Each experiment was performed three independent times.

RESULTS

Identification of *KIAA1549* variants

To identify the genetic defect underlying the arRP in two affected siblings of an Iranian consanguineous family (Family A; **Figure 1A**), exome sequencing was performed in individual A-II:2. After analysis of the WES data, the frameshift variant c.52del (Hg19:g.138,665,964del; p.(Arg18Alafs*64)) was detected in the candidate RP gene *KIAA1549* (Family A). This variant is located in the second largest homozygous region of 13.2 Mb. Presence of the homozygous variant was confirmed and segregation analysis was performed using Sanger sequencing. The variant is absent from population frequency databases gnomAD, ExAC, dbSNP and the in-house database. Moreover, the variant is absent from the Iranome database¹⁶, which contains whole exome sequencing data of 800 healthy individuals from eight major ethnic groups in Iran. The variant causes a frameshift in exon 1, and is predicted to result in degradation of *KIAA1549* mRNA due to nonsense-mediated decay. Previously, a heterozygous variant in *CRB1* was reported for Family A in both affected siblings.¹⁷ Analysis of the WES data of patient A-II:2 did not yield additional variants or copy number variants in this gene. Moreover, the c.2816A>G (p.(Asn894Ser)) variant was predicted to be benign by *in silico* predictions

suggesting that *CRB1*-variants are unlikely to cause disease in this family. No CNVs or other compound heterozygous or homozygous variants were detected in currently known IRD-associated genes. Also, no heterozygous candidate variants were found in causative genes related to the patient's phenotype. WES was performed in patient B-II:1 affected with RP (Family B, **Figure 1A**), and revealed a homozygous missense variant in *KIAA1549*; c.4686C>A (Hg19:g.138,554,373G>T; p.(His1562Gln)). After analysis, this was the only homozygous variant remaining in an IRD-associated gene and no compound heterozygous variants were observed. The homozygous variant was validated in the proband and segregation analysis was performed by Sanger sequencing in two unaffected siblings.

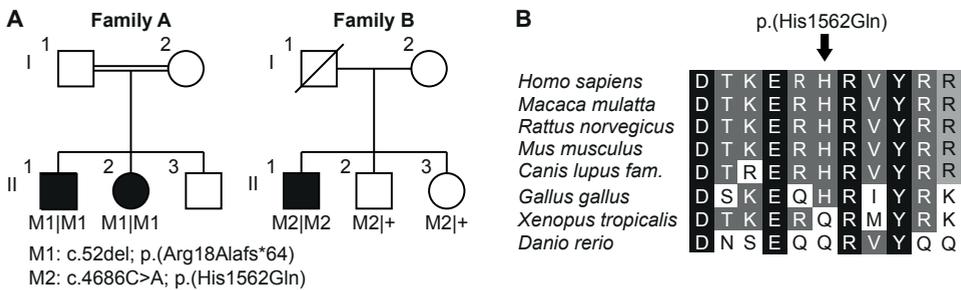


Figure 1. KIAA1549 variants detected in patients in two families with retinitis pigmentosa. (A) Pedigrees of two families with RP associated with homozygous variants in *KIAA1549* (NM_001164665). **(B)** Evolutionary conservation of the mutated amino acid (M2) identified in Family B. Black boxes indicate fully conserved amino acid residues, whereas dark grey boxes indicate highly conserved and light grey boxes moderately conserved amino acid residues. M, mutation.

The moderately conserved mutated histidine residue is located in a highly conserved region (**Figure 1B**), has a CADD-Phred score of 24.2, a PhyloP score of 0.53 and a Grantham score of 24. Additionally, the prediction tools MutationTaster¹⁸, PolyPhen-2¹⁹ and SIFT¹⁴ predicted the variant to be disease causing (p-value: 0.835), possibly damaging (HumDiv: 0.889; HumVar: 0.651) and tolerated (0.17), respectively. Moreover, putative changes are predicted by Human Splicing Finder.²⁰ The binding sites of the splicing factors SRp40 and SF2/ASF are no longer present, and a potential creation of an exonic splicing silencer site is predicted.²⁰ Heterozygous variants in currently known IRD-associated genes were found in *ABCA4* (c.2588G>C(;);5603A>T; p.[Gly863Ala, Gly863del] (;)(Asn1868Ile) and *CDHR1* (c.512C>G; p.(Thr171Ser)), but no second pathogenic alleles could be detected for these genes. No CNVs were detected in regions overlapping with known IRD-associated genes.

Clinical evaluation

Clinical data were collected from the medical records of two patients from Family A (A-II:1, A-II:2) and one patient from Family B (B-II:1). An overview of the clinical characteristics of the three affected individuals with damaging *KIAA1549* variants at the most recent examination is provided in **Table 1** and clinical images of patient A-II:2 and B-II:1 are shown in **Figure 2**. All affected individuals were diagnosed with RP. They all initially experienced night blindness, followed by a gradual decline of their visual fields and visual acuity. The age of onset varied from the first decade (patient A-II:1) to the fifth decade (patient B-II:1) and all patients were myopic. Cortical cataract was observed in patient A-II:1 (age 38), whereas patient A-II:2 underwent a cataract extraction at the age of 38 (right eye) and 52 years (left eye). Ophthalmoscopy revealed characteristic RP features in all three patients, including attenuated retinal vessels, waxy pallor of the optic disc, and bone spicule pigmentation (**Figure 2A, 2D**). In addition, nummular deep pigmentations were visible in the mid-periphery. SD-OCT imaging in patient A-II:2 showed profound atrophy of the outer retinal layers with preservation of the photoreceptors in the fovea. This patient was treated for Coats-like exudative vasculopathy related to her RP in the past. Fundus autofluorescence imaging in patient B-II:1 showed the characteristic hyperautofluorescent ring that represents the transition zone between intact and degenerated photoreceptor outer segments, corresponding with a preserved ellipsoid zone within the ring on SD-OCT, and loss of the ellipsoid zone external to the ring (**Figure 2E-F**).

Table 1. Clinical features at most recent examination in patients with segregating KIAA1549 variants

Patient	Sex/ Age (y)	Initial symptoms, age (y)	Visual acuity		Spherical equivalent		Lens status	Ophthalmoscopy results	ERG		Goldmann perimetry
			RE	LE	RE	LE			scot	phot	
A-II:1	M/38	Night blindness (8)	20/40	20/50	-4.38	-5.25	Cortical cataracts	Severely attenuated retinal vessels, RPE atrophy with bone spicule and nummular pigmentation in the periphery, preserved posterior pole, and absence of optic disc pallor	NP	NP	NP
A-II:2	F/54	Night blindness (28)	20/400	20/400	-6.50*	-7.25	Pseudophakia	Attenuated retinal vessels, severe RPE atrophy BE, with recognizable foveal island, moderate optic disc pallor, bone spicule and nummular pigmentations, white epiretinal changes, and old Coats-like exudative vasculopathy inferior quadrants BE	NR†	NR†	Constricted VF, central residue <5° (age 48)
B-II:1	M/54	Night blindness (~45)	20/60	20/20	-2.00	-2.13	Clear	Attenuated vessels, midperipheral bone spicule and nummular pigmentation, waxy pallor of the optic discs, and CME (RE>LE)	SR	MR	RE: midperipheral ring scotoma LE: partial midperipheral ring scotoma. VF affected temporal>nasal BE

BE, both eyes; CME, cystoid macular edema; ERG, electroretinography; F, female; LE, left eye; M, male; MR, moderately reduced; NP, not performed; NR, non-recordable; phot, photopic; RE, right eye; scot, scotopic; SR, severely reduced; VF, visual field; y, years. *Prior to cataract surgery. †ERG performed at the age of 32.

In addition, the SD-OCT image of patient B-II:1 showed evident cystoid macular edema, that was refractory to topical treatment with nonsteroidal anti-inflammatory drugs, steroids, as well as both topical and oral carbonic anhydrase inhibitor treatment (**Figure 2F**). Electrophysiology examination demonstrated a generalized retinal dystrophy with non-recordable rod and cone-driven responses in patient A-II:2 at the age of 32 years, and severely reduced rod and cone-driven responses in patient B-II:1 at 54 years of age. Finally, perimetric analysis revealed a severely constricted visual field up to 5 degrees in patient A-II:2, and a complete and partial ring scotoma in the right and left eye of patient B-II:1, respectively. No visual field testing was performed in patient A-II:1, yet he reported severe visual field constriction. All patients were in good general health, and no non-ocular conditions were reported.

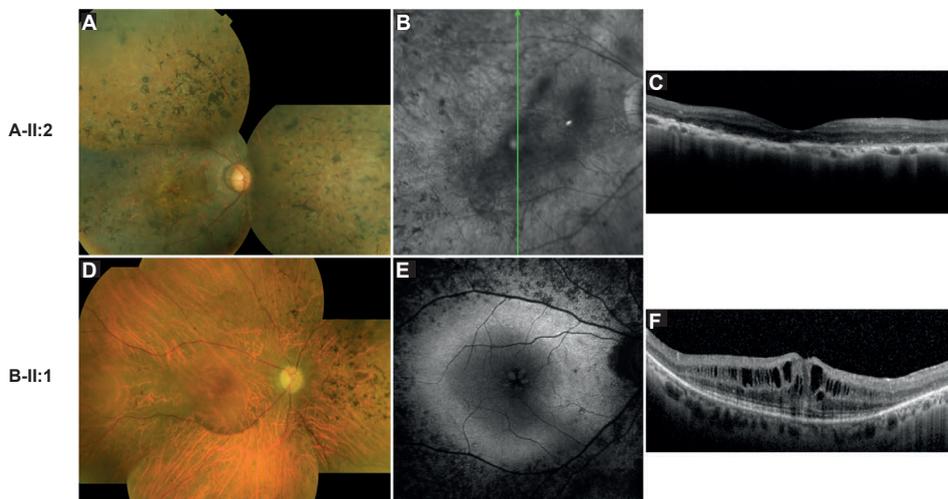


Figure 2. Multimodal retinal imaging of patients with segregating *KIAA1549* variants. (A–C), Clinical characteristics of patient II:2 of Family A. (A) Composite fundus photograph of the right eye of patient A-II:2 showing attenuated retinal vessels, bone spicule and nummular pigmentation, moderate optic disc pallor, and a remnant of RPE between optic disc and macula, and at the fovea. (B) Infrared image of the posterior pole of the right eye of patient A-II:2, indicating the position of the corresponding optical coherence tomography (OCT) examination. (C) Spectral-domain OCT showing atrophy of the outer retina with intact photoreceptors in the fovea. (D–F), Clinical characteristics of patient II:1 of Family B. (D) Composite fundus photograph of the right eye of patient B-II:1, showing attenuated retinal vessels, waxy pallor of the optic disc, and midperipheral bone spicule and nummular pigmentation. (E) Fundus autofluorescence of the right eye of patient B-II:1, showing a central hyperautofluorescent ring surrounding the normal appearing retina, hypoautofluorescence spots along and external to the vascular arcade, and a spoke wheel pattern in the fovea corresponding to the cystoid macular edema visible on OCT. (F) SD-OCT scan taken along the horizontal meridian of the central retina, revealing peripheral loss of the outer retina with central preservation of the ellipsoid band, and macular cysts.

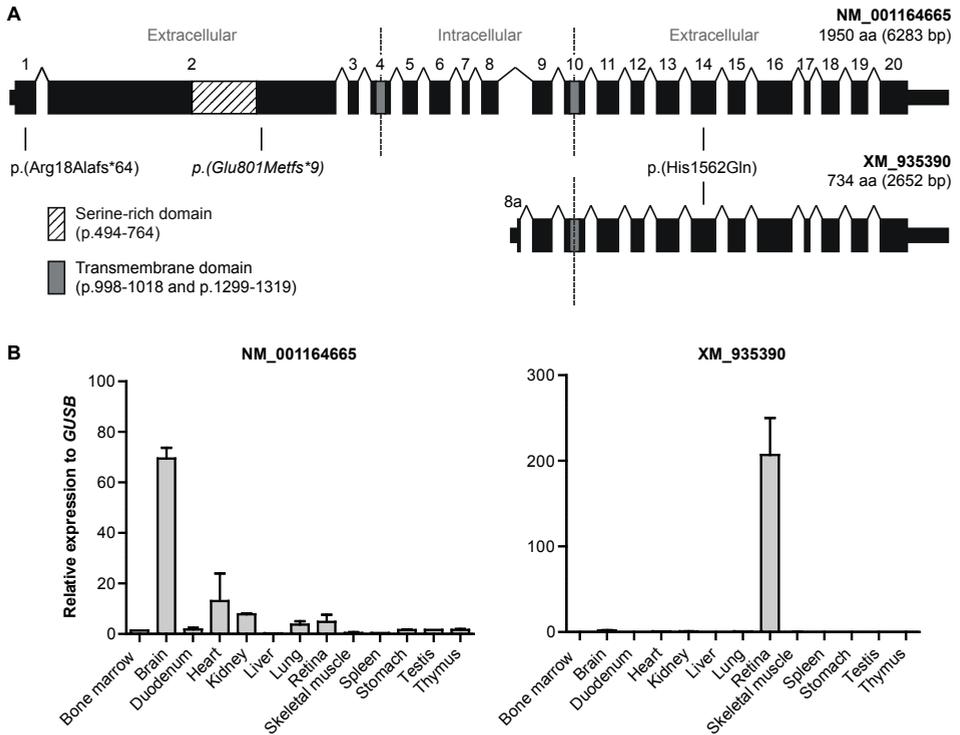


Figure 3. Expression of *KIAA1549* transcripts in human tissues. (A) Schematic representation of the major long and short *KIAA1549* transcripts and identified segregating variants. A previously reported *KIAA1549* variant by Abu-Safieh et al. is depicted in italics.⁴ The short transcript is transcribed from an alternative promoter sequence present in intron 8, and includes a transcript-specific exon referred to as exon 8a. (B) Relative *KIAA1549* expression in human tissues determined by qPCR. Expression of the transcript encoding the long isoform (NM_001164555) is depicted in the left panel, and expression of the transcript encoding the short isoform (XM_935390) is depicted in the right panel. The long transcript is predominantly expressed in the brain, whereas expression of the short transcript is significantly increased in the retina compared to other tissues.

Expression of *KIAA1549* transcripts in human tissues

To gain knowledge on the specific role of *KIAA1549*, its relative expression was determined in a set of human adult tissues. Two major *KIAA1549* isoforms have been identified (Uniprot: Q9HCM3), a long primary isoform (NM_001164665) of 1950 amino acids (aa) and a short isoform (XM_935390) of 734 aa (**Figure 3A**). The shorter isoform is produced from an alternative transcript transcribed from an alternative promoter sequence located in intron 8. The nomenclature of all genetic or protein elements is based on the long isoform. The expression of both *KIAA1549* transcripts and of the reference gene *GUSB* was evaluated in cDNA of human tissues by qPCR (**Figure 3B**). The

long transcript showed a low to moderate expression in retina and other tissues, such as heart and kidney, and is predominantly expressed in brain as has been previously described.²¹

Localization of KIAA1549 in mouse retina sections

To confirm the presence of KIAA1549 in the retina, as well as to define its specific localization in this tissue, immunofluorescence was performed in retina sections obtained from a healthy, 2-month-old mouse. Costaining was performed with anti-Centrin, a well-defined marker for the connecting cilium within the photoreceptor cell.²² Results showed that KIAA1549 colocalized with Centrin, and thus is located at the connecting cilium of the photoreceptor cells (**Figure 4**). Moreover, positive staining of KIAA1549 was also observed at the outer plexiform layer of the mouse retina. This layer contains neural synapses between the photoreceptors and the bipolar and horizontal cells in the retina.

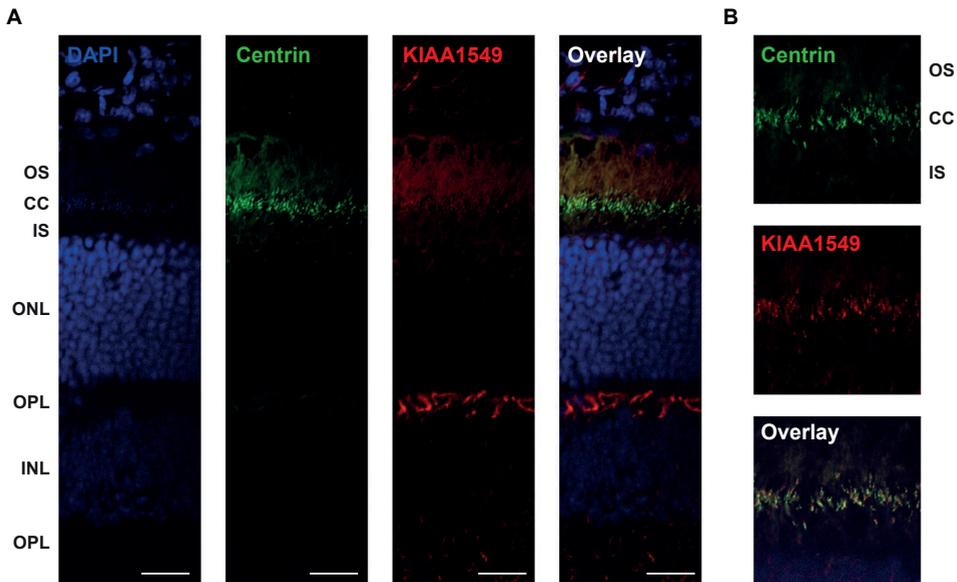


Figure 4. Localization of KIAA1549 in the retina of an adult mouse. Immunofluorescence analysis of KIAA1549 (red) stained on an unfixed retina section obtained from a healthy 2-month-old mouse. Costaining was performed with an antibody for Centrin (green), a connecting cilium marker of the photoreceptor. **(A)** Provides an overview of the retina section and was imaged using a 40x magnification. The scale bar represents 20 μ m. KIAA1549 is detected at the connecting cilium photoreceptor cells and the outer plexiform layer of the mouse retina. **(B)** Focused image of the photoreceptor region taken using a 63x magnification. The signal of KIAA1549 at the connecting cilium shows overlap with that of Centrin. CC, connecting cilium; INL, inner nuclear layer; IPL, inner plexiform layer; IS, inner photoreceptor segment; ONL, outer nuclear layer; OPL, outer plexiform layer; OS, outer photoreceptor segment.

Assessing the function of *KIAA1549*

To investigate whether *KIAA1549* plays a role in ciliogenesis, an *in vitro* study was performed in which *KIAA1549* was knocked down in hTERT-RPE1 cells with two different siRNAs targeting the long transcript. The efficiency of *KIAA1549* knockdown was validated by qPCR analysis, and induced ciliogenesis was studied using immunocytochemistry. The percentage of ciliated cells and average cilium length showed no significant difference when comparing cells treated with *KIAA1549*-targeting siRNAs or the non-targeting siRNA (**Figure S1**), suggesting that *KIAA1549* does not have a function in the formation of cilia however *KIAA1549* may be involved in other processes that are performed at the primary cilium.

DISCUSSION

In this study, we report on two families in which autosomal recessive RP is associated with homozygous variants in *KIAA1549*. The retinal phenotype in both families is typical for RP, and patients' complaints started with night blindness with subsequent constriction of the visual field. Fundus examination revealed the hallmark RP features. However, patients in Family A are more severely affected compared to the patient in Family B, which is displayed in a lower age at onset, severely constricted visual fields, and more severely reduced electroretinography (ERG) responses.

In Family A, a homozygous frameshift variant was found in exon 1 (c.52del; p.(Arg18Alafs*64)) by WES. Although putative alternative start codons are present in exon 2, exon 1 encodes the signal peptide of the protein (aa 1-60). Therefore, a shorter protein is potentially mislocalized, impairing protein function. In Family B, a homozygous missense variant was found in exon 14 (c.4686C>A; p.(His1562Gln)), that affects a highly conserved region of the protein. The damaging nature of these variants is supported by a probability of loss of function intolerance (pLI) score of 1.00 (Scale 0-1) in gnomAD (accessed on 1st June 2018) and that no homozygous variants have been reported in the entire *KIAA1549* gene. Combined, this suggests that the identified *KIAA1549* variants in both families can be associated to the RP-phenotype of the patients. Regardless, the presence of pathogenic variants present in non-coding regions uncovered by WES cannot be ruled out.

KIAA1549 encodes a transmembrane protein and is described to be predominantly expressed in the brain and is involved in oncogenesis when fused to *BRAF*.^{5,23} *BRAF-KIAA1549* in-frame fusion genes are caused by a 2 Mb tandem duplication at 7q34, and are found to induce *BRAF* kinase activity and consequently, activation of the

MAPK pathway which is involved in the development of cancer. For this reason, these fusion genes are the major cause (66%) for pilocytic astrocytomas, the most frequently occurring central nervous system tumor in children and young adolescents.

Besides this role in oncogenesis, knowledge about the function of *KIAA1549* is limited. Recently, a homozygous truncating variant in *KIAA1549* was found in an arRP family with two affected siblings in a study performed by Abu-Safieh et al.⁵ Involvement of *KIAA1549* in photoreceptor function was suggested, however no functional data was provided.⁵ Nevertheless, *KIAA1549* is reported to be among the top 4% of genes being enriched for binding sites for the photoreceptor specific transcription factor CRX.²⁴ *Kiaa1549* expression was evaluated in a *Nrl*^{-/-} knockout mouse that is characterized by degenerated rod photoreceptors. In this mouse, *Kiaa1549* expression was found to be reduced ~88% (wildtype: 106 reads, knockout: 13 reads) when compared to the wildtype mouse, based on number of sequencing reads.²⁴

In this study, expression levels of the major short and long transcripts of *KIAA1549* have been evaluated in a set of human tissues, which demonstrated that both isoforms are present in the retina, of which the expression of the transcript encoding the short isoform is significantly higher in retina compared to other tissues. We hypothesize that both isoforms are required for the correct function of the protein in the retina, as the homozygous frameshift variant affecting the long isoform has detrimental consequences as observed in Family A and the family previously described in the study of Abu-Safieh et al.⁵ By performing an *in vitro* experiment in which HEK293T cells were transfected with *KIAA1549*-targeting siRNAs that specifically recognize the long transcript of *KIAA1549* (**Table S1**), also a significant decrease in expression of the short transcript was observed (**Figure S2**) which suggests a functional dependency between the two transcripts. Hence, observed variants in the long transcript likely cause a decrease in the abundant retinal expression of the short transcript and thereby could lead to retinal degeneration. The fact that the identified variants have different consequences on the two *KIAA1549* transcripts could explain the phenotypic differences observed among the affected individuals. The phenotype of the family described by Abu-Safieh et al. (a non-recordable ERG at age 35) (personal communication Prof. F.S. Alkuraya and N. Patel, PhD) is more comparable with Family A (non-recordable ERG at age 32 in patient A-II:2) than Family B (severely reduced photopic and moderately reduced scotopic ERG at age 54), which may be in line with the genotype having a damaging variant in the long transcript. Identification of additional families with *KIAA1549*-associated RP are required to provide deeper insight into a possible phenotype-genotype correlation.²⁵

In addition, we showed localization of KIAA1549 at the connecting cilium of mouse photoreceptor cells, providing the first information on KIAA1549 function in photoreceptors. Moreover, KIAA1549 localization was also noted at the outer plexiform layer of the mouse retina. Proteins localized at the ribbon synapses of the outer plexiform layer are often structural or synaptic vesicle proteins or are involved synaptic vesicle trafficking.^{26,27} The KIAA1549 antibody will recognize both isoforms and thus does not provide additional knowledge on alternative localization of the isoforms. Hypothetically, the long and short isoforms may harbor a unique function at either one of the identified locations. Additional research is required to unravel the functional differences between the short isoform and the ubiquitously expressed long isoform of KIAA1549.

Besides localization in the photoreceptor, there is additional evidence for ciliary function is at the molecular level. A recent study based on proximity-dependent biotinylation revealed an interaction between KIAA1549 and TMEM17.²⁸ TMEM17 is a part of the Meckel syndrome (MKS) protein complex located in the ciliary transition zone, in which it facilitates cilium formation. Also, pathogenic variants in genes encoding proteins in this complex are known to cause (severe) ciliopathies.²⁹ *TMEM17* pathogenic variants have been reported to cause oral-facial-digital syndrome type 6.²⁹ The MKS complex contains both cytoplasmic and transmembrane proteins, and functions as a barrier preventing rapid diffusion of transmembrane proteins between cilia and plasma membranes.³⁰ The interaction between KIAA1549 and TMEM17 was only observed in cells in non-ciliated conditions, which suggests that the interaction is involved in a cilium-related process.²¹

We have studied the role of KIAA1549 in ciliogenesis, by knocking down the expression of the gene in hTERT-RPE1 cells using siRNAs. siRNA-transfected cells did not show a difference in percentage of ciliated cells or ciliumlength, suggesting that KIAA1549 does not have a direct role in the cilium formation explaining the nonsyndromic phenotype observed in the patients of Family A and B, as well as the family of Abu-Safieh et al., which is restricted to the retina. Pathogenic variants that do affect genes essential for ciliogenesis, such as *TMEM17*, would give rise to a phenotype likely affecting multiple organs as in ciliopathies. Transmembrane proteins present at the transition zone are often involved in the sensing and transducing of extracellular signals. Like TMEM17, KIAA1549 is a transmembrane protein, therefore it is plausible that KIAA1549 may be involved in these processes at the primary cilium of the photoreceptors specifically.³⁰

In conclusion, by employing WES we have identified that homozygous frameshift or missense variants in *KIAA1549* are associated with RP in two families. We demonstrated retina-specific expression of the short isoform of *KIAA1549* and provide evidence that damaging variants targeting the long transcript may cause RP by reducing the

expression of the short transcript. Moreover, we showed that KIAA1549 resided in the connecting cilium of the mouse retina, thereby providing supporting evidence that KIAA1549 might act as an essential photoreceptor protein.

ACKNOWLEDGEMENTS

We would like to thank Theo A. Peters, Sanne Broekman, Nisha Patel, Fozwan S. Alkuraya, Thanh-Minh T. Nguyen and Maartje van de Vorst for expert technical assistance. The study was financially supported by DCN Radboudumc grant (to F.P.M.C. and H.K.), as well as the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, the Stichting tot Verbetering van het Lot der Blinden, and the Stichting Blinden-Penning (to F.P.M.C. and S.R.).

REFERENCES

1. Hamel, C. Retinitis pigmentosa. *Orphanet Journal of Rare Diseases* **1**, 40-40 (2006).
2. Hartong, D.T., Berson, E.L. & Dryja, T.P. Retinitis pigmentosa. *The Lancet* **368**, 1795-1809 (2006).
3. RetNet. Available from: <https://sph.uth.edu/retnet/>.
4. Haer-Wigman, L., van Zelst-Stams, W.A.G., Pfundt, R., van den Born, L.I., Klaver, C.C.W., Verheij, J.B.G.M. *et al.* Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *European Journal of Human Genetics* **25**, 591-599 (2017).
5. Abu-Safieh, L., Alrashed, M., Anazi, S., Alkuraya, H., Khan, A.O., Al-Owain, M. *et al.* Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Research* **23**, 236-247 (2013).
6. McCulloch, D.L., Marmor, M.F., Brigell, M.G., Hamilton, R., Holder, G.E., Tzekov, R., and Bach, M. ISCEV Standard for full-field clinical electroretinography (2015 update). *Documenta Ophthalmologica* **130**, 1-12 (2015).
7. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
8. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
9. Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P. *et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Research* **22**, 1525-1532 (2012).
10. Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S. *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **28**, 2747-2754 (2012).
11. Whiffin, N., Karczewski, K.J., Zhang, X., Chothani, S., Smith, M.J., Evans, D.G. *et al.* Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nature Communications* **11**, 2523 (2020).
12. Karczewski, K.J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D.M., Kavanagh, D. *et al.* The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Research* **45**, D840-D845 (2017).
13. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* **29**, 308-311 (2001).
14. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. & Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310-315 (2014).
15. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T. *et al.* Fiji - an Open Source platform for biological image analysis. *Nature Methods* **9**, 678-682 (2012).

16. Fattahi, Z., Beheshtian, M., Mohseni, M., Poustchi, H., Sellars, E., Nezhadi, S.H. *et al.* Iranome: a catalog of genomic variations in the Iranian population. *Human Mutation* **40**, 1968-1984 (2019).
17. Hollander, A.I.d., Heckenlively, J.R., van den Born, L.I., de Kok, Y.J.M., van der Velde-Visser, S.D., Kellner, U. *et al.* Leber congenital amaurosis and retinitis pigmentosa with coats-like exudative vasculopathy are associated with mutations in the crumbs homologue 1 (CRB1) gene. *American Journal of Human Genetics* **69**, 198-203 (2001).
18. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**, 361-362 (2014).
19. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249 (2010).
20. Desmet, F.-O., Hamroun, D., Lalonde, M., Collod-Bérout, G., Claustres, M. & Bérout, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research* **37**, e67-e67 (2009).
21. Sadighi, Z. & Slopis, J. Pilocytic astrocytoma. *Journal of Child Neurology* **28**, 625-632 (2013).
22. Gießl, A., Trojan, P., Rausch, S., Pulvermüller, A. & Wolfrum, U. Centrin, gatekeepers for the light-dependent translocation of transducin through the photoreceptor cell connecting cilium. *Vision Research* **46**, 4502-4509 (2006).
23. Jones, D.T.W., Kocialkowski, S., Liu, L., Pearson, D.M., Bäcklund, L.M., Ichimura, K. *et al.* Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas. *Cancer Research* **68**, 8673-8677 (2008).
24. Özgül, Rıza K., Siemiatkowska, Anna M., Yücel, D., Myers, Connie A., Collin, Rob W., Zonneveld, Marijke N. *et al.* Exome sequencing and cis-regulatory mapping identify mutations in MAK, a gene encoding a regulator of ciliary length, as a cause of retinitis pigmentosa. *American Journal of Human Genetics* **89**, 253-264 (2011).
25. Kevany, B.M. & Palczewski, K. Phagocytosis of retinal rod and cone photoreceptors. *Physiology* **25**, 8-15 (2010).
26. Ullrich, B. & Südhof, T.C. Distribution of synaptic markers in the retina: implications for synaptic vesicle traffic in ribbon synapses. *Journal of Physiology-Paris* **88**, 249-257 (1994).
27. Mercer, A.J. & Thoreson, W.B. The dynamic architecture of photoreceptor ribbon synapses: cytoskeletal, extracellular matrix, and intramembrane proteins. *Visual Neuroscience* **28**, 453-471 (2011).
28. Gupta, Gagan D., Coyaud, É., Gonçalves, J., Mojarad, Bahareh A., Liu, Y., Wu, Q. *et al.* A dynamic protein interaction landscape of the human centrosome-cilium interface. *Cell* **163**, 1484-1499 (2015).
29. Li, C., Jensen, V.L., Park, K., Kennedy, J., Garcia-Gonzalo, F.R., Romani, M. *et al.* MKS5 and CEP290 dependent assembly pathway of the ciliary transition zone. *PLoS Biology* **14**, e1002416 (2016).
30. Chih, B., Liu, P., Chinn, Y., Chalouni, C., Komuves, L.G., Hass, P.E. *et al.* A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain. *Nature Cell Biology* **14**, 61-72 (2011).

SUPPLEMENTARY TABLES

Table S1. Sequences of primers used for qPCR to determine tissue-specific expression levels

Primer name	Location (NM_001164665)	Sequence (5'-3')
KIAA1549_long_Fw	Exon 2	ACACCAACACTGGCTACTGC
KIAA1549_long_Rev	Exon 3	TGATGTACTCCTGCACAGCTC
KIAA1549_short_Fw	Exon 8a	AGCTTCTGCAATGTGAATGG
KIAA1549_short_Rev	Exon 9	TGTACCGGATTGTCATCTCC

Table S2. siRNAs used for KIAA1549 knockdown *in vitro* experiments

siRNA ID	Location (NM_001164665)	Sequence (5'-3')
s33562 (siRNA 1)	Exon 3	GGAGUACAUCAUUACAGCAtt
s33563 (siRNA 2)	Exon 5	CAGGGAACGUUAUACCUCAtt

SUPPLEMENTARY FIGURES

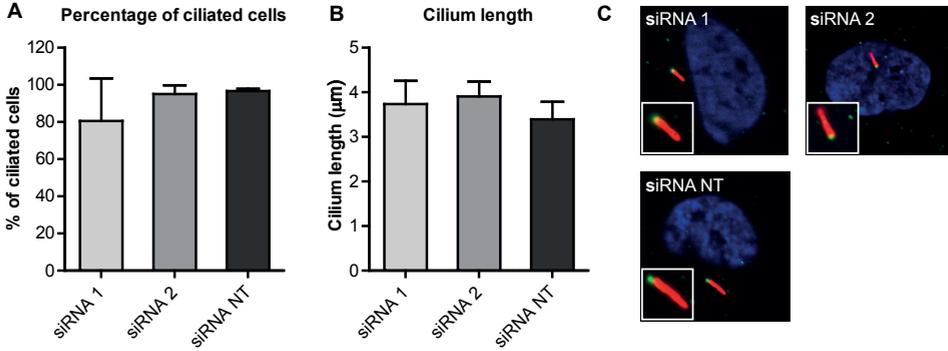


Figure S1. Effect of *KIAA1549* knockdown on ciliogenesis. hTERT-RPE1 cells were transfected with two different siRNAs targeting *KIAA1549* (siRNA 1 and 2) and one non-targeting siRNA (siRNA NT). After transfection of the cells, ciliogenesis was induced and immunofluorescence was used for analysis. Primary cilia were stained using an anti-ARL13B antibody (red), and the transition zone was stained using an anti-RPGRIP1L antibody (green). **(A)** Percentage of ciliated cells calculated and **(B)** measured cilium lengths for hTERT-RPE1 cells transfected with the different siRNAs. **(C)** Images of ciliated cells transfected with one of the siRNAs. A close-up picture of the primary cilium is shown to visualize the transition zone.

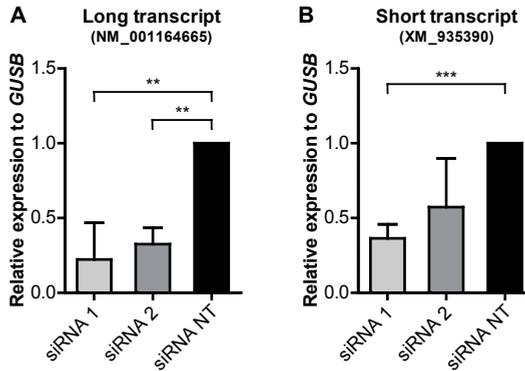
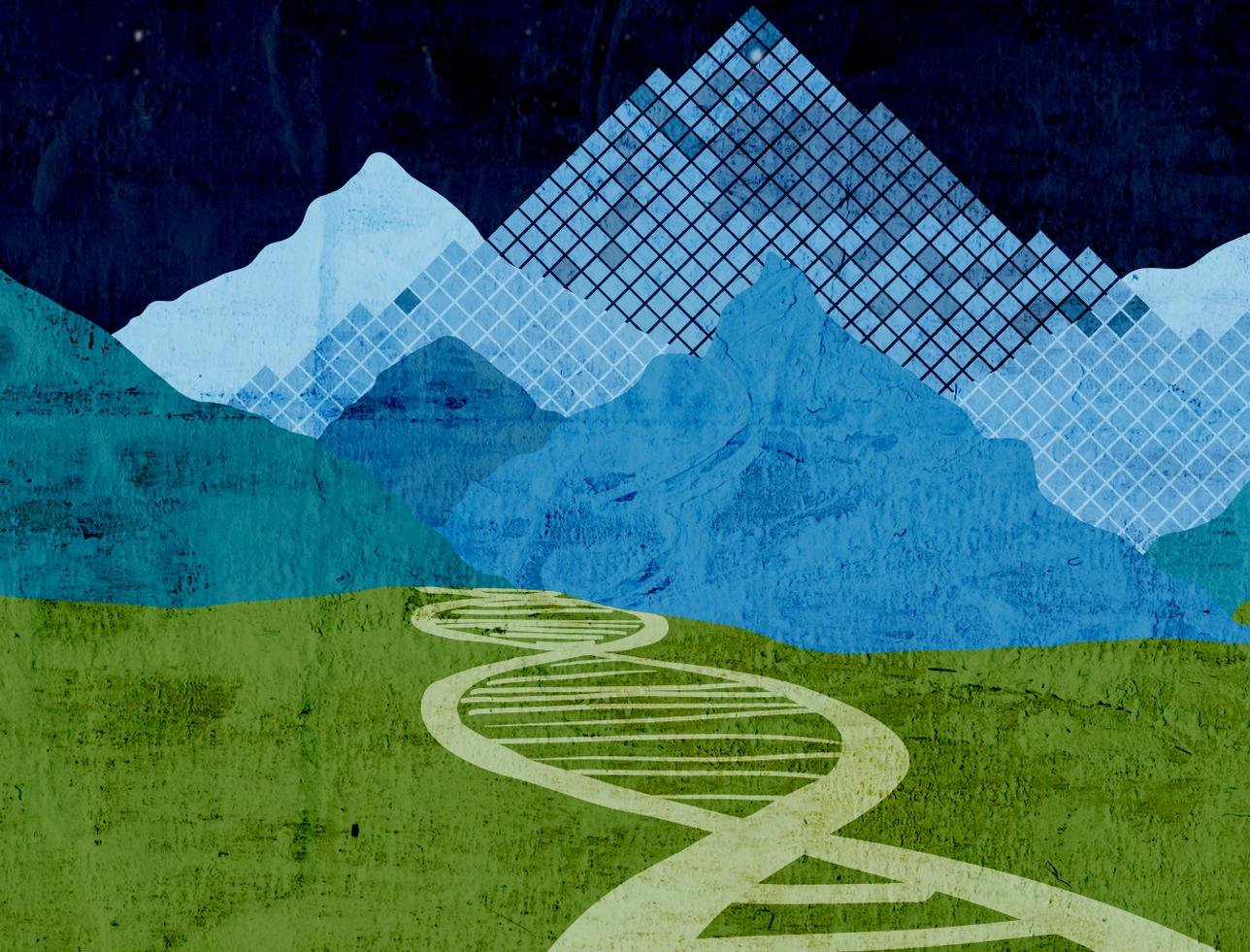


Figure S2. Expression levels of *KIAA1549* transcripts in HEK293T after knockdown using siRNAs. HEK293T cells were transfected with two different siRNAs targeting the long *KIAA1549* transcript specifically (siRNA 1 and 2) and one non-targeting siRNA (siRNA NT) in two independent experiments. After transfection of the cells, expression levels of both the short and the long transcript were quantified using qPCR. **(A)** Using siRNA 1 and 2, a significant knockdown of the long *KIAA1549* transcript was established. **(B)** Also, decreased expression levels of the short transcript were observed, which was found significant for siRNA 1. Significance was calculated using an unpaired t-test, ** p-value < 0.01, *** p-value < 0.001.

CHAPTER 3

DFNA21



Chapter 3.1

A *RIPOR2* in-frame deletion is a frequent and highly penetrant cause of adult-onset hearing loss

Suzanne E. de Bruijn*, Jeroen J. Smits[†], Chang Liu, Cornelis P. Lanting, Andy J. Beynon, Joëlle Blankevoort, Jaap Oostrik, Wouter Koole, Erik de Vrieze, DOOFNL Consortium, Cor W.R.J. Cremers, Frans P. M. Cremers, Susanne Roosing, Helger G. Yntema, Henricus P.M. Kunst, Bo Zhao, Ronald J.E. Pennings[‡], and Hannie Kremer[§]

*, [‡] These authors contributed equally to this work

Journal of Medical Genetics **58**, 96-104 (2021)

ABSTRACT

Hearing loss is one of the most prevalent disabilities worldwide, and has a significant impact on quality of life. The adult-onset type of the condition is highly heritable but the genetic causes are largely unknown, which is in contrast to childhood-onset hearing loss. An in-frame deletion of 12 nucleotides in *RIPOR2* was identified as a highly penetrant cause of adult-onset progressive hearing loss that segregated as an autosomal dominant trait in 12 families from the Netherlands. Hearing loss associated with the deletion in 63 subjects displayed variable audiometric characteristics and an average age of onset of 30.6 years (SD 14.9 years, range 0-70 years). A functional effect of the *RIPOR2* variant was demonstrated by aberrant localization of the mutant *RIPOR2* in the stereocilia of cochlear hair cells and failure to rescue morphological defects in *RIPOR2*-deficient hair cells, in contrast to the wildtype protein. Strikingly, the *RIPOR2* variant is present in 18 of 22,952 individuals not selected for hearing loss in the Southeast Netherlands. Collectively, the presented data demonstrate that an inherited form of adult-onset hearing loss is relatively common, with potentially thousands of individuals at risk in the Netherlands and beyond, which makes it an attractive target for developing a (genetic) therapy.

INTRODUCTION

Hearing loss (HL) is one of the most prevalent disabilities worldwide¹ and genetic factors importantly contribute to this condition. So far, 118 genes have been associated with nonsyndromic forms of sensorineural HL and variants in these genes explain a significant part of subjects with an early onset of HL, i.e., congenital or in childhood.²⁻⁴ Our knowledge of the genetic architecture of adult-onset HL is limited despite a high heritability which is estimated to be 30-70%.⁵⁻⁷ Differences in phenotypic parameters that are used and age ranges of study participants may well contribute to the variation in the reported heritability. As summarized by Lewis et al.⁸, genome-wide association studies (GWAS) of hearing status in adults and genetic analyses of families with dominantly inherited post-lingual onset HL indicate that both common variants and rare variants contribute to adult-onset HL with a small and large effect size, respectively. Such variants may or may not affect genes that are already known to function in the auditory pathway.

Previously, we identified a 12.4-Mb locus for adult-onset HL on chromosome 6 (p24.1-22.3): DFNA21.^{9,10} However, the underlying pathogenic variant in the studied family (W97-056) remained elusive. Here, we present the identification of an in-frame deletion (c.1696_1707del; NM_014722.3) in *RIPOR2* (RHO Family Interacting Cell Polarization Regulator 2) to underlie autosomal dominant nonsyndromic HL (adNSHL) in this family and in 11 additional (large) families of Dutch origin. The allele frequency (AF) of this variant suggests that it potentially explains adult-onset HL in thousands of individuals in the Netherlands and Northwest Europe. Our study expands the phenotypic spectrum associated with *RIPOR2* defects which had so far only been described to underlie early-onset recessively inherited HL.¹¹

MATERIALS AND METHODS

Study approval

The study of human subjects was approved by the medical ethics committee of the Radboudumc (registration number: NL33648.091.10) and performed in accordance with the principles of the World Medical Association Declaration of Helsinki. Written informed consent was obtained from all participants or their legal representatives. All animal experiments were approved by the Institutional Animal Care and Use Committee of Indiana University School of Medicine (registration number 19075).

DNA sequencing

Next generation sequencing was performed for identification of DNA variants. Details of employed sequencing techniques are provided in **Supplementary Methods**.

Variant interpretation

For exome sequencing and Molecular Inversion Probe (MIP) datasets, annotated variants were filtered based on a population AF of $\leq 0.5\%$ in the gnomAD database V.2.1, and our in-house exome database ($\sim 15,000$ alleles). Variants in coding and splice site regions ($-14/+14$ nucleotides) were analyzed. Interpretation of missense variants was performed using the *in silico* tools CADD-PHRED (≥ 15),¹² SIFT (≤ 0.05),¹³ PolyPhen-2 (≥ 0.450)¹⁴ and MutationTaster (deleterious)¹⁵ to predict potential functional effects. Variants were considered if a pathogenic effect was predicted by at least two different tools. Potential effects on splicing of missense and synonymous variants were evaluated using the algorithms embedded in the AlamutVisual software (V.2.10, Interactive Biosoftware). A change of $\geq 5\%$ in splice site scores predicted by at least two algorithms was considered significant. For candidate variants, segregation analysis was performed by Sanger sequencing. PCR conditions are available upon request.

Clinical evaluation

Medical history was taken from all participants with special attention paid to acquired and noise-induced HL. Both affected and unaffected participants underwent general Ear Nose and Throat examinations, or this medical information was taken from previous examinations. Age of onset of HL was reported by subjects themselves. Only reports of a specific age of onset were used in calculations. The audiometric data in this study are described according to GENDEAF guidelines.¹⁶ Pure tone- and speech- audiometry and click-evoked auditory brainstem response (ABR) was performed in a sound-attenuated booth, according to current standards (International Organisation for Standardization; ISO 8253-1:2010, ISO 389-1, ISO 389-5 and ISO 389-6).¹⁷ Individuals were considered affected when pure tone thresholds for at least three individual frequencies were below the frequency-specific 95th percentile of age- and sex-specific thresholds (ISO7029:2017) for the best hearing ear. HL was considered asymmetric if pure tone audiometry showed a difference of more than 10 dB between both ears at two individual frequencies.¹⁶ Longitudinal (individual) progression of HL was calculated if there was a follow-up duration of at least 10 years, after onset of HL. The progression rate is defined as the mean increase pure tone average at 0.5-4 kHz ($PTA_{0.5-4kHz}$) in dB/year between first and last audiometry. For symmetric HL, the average of both ears was used to calculate progression; for asymmetric HL, the best-hearing ear at first audiometry was used. In case of profound HL at 0.5-4 kHz at the latest audiometry, the most recent audiometry

at which all thresholds at 0.5-4 kHz could be measured, was selected. Cross-sectional linear regression analysis was applied on pure tone thresholds to calculate an Age Related Typical Audiogram (ARTA),¹⁸ using Prism 6.0 software (GraphPad). A k-means clustering analysis was performed as described in **Supplementary Methods**.

Injectoporation of *Ripor2*-constructs and immunostaining

The generation of *Ripor2*^{LacZ/LacZ} mice has been described previously.¹⁹ For *Ripor2* DNA construct generation, *Ripor2* cDNA (NM_029679.2, without exon 13) was amplified from a mouse-cochlear cDNA library and cloned into a pEGFP-N3-derived vector from which the EGFP coding sequence was deleted. Procedures for injectoporation and immunostaining have been described previously,¹⁹ and are detailed in Supplementary Methods.

Immunoprecipitations and western blots

Cell culture, immunoprecipitations and western blots were carried out as described.^{19,20} Experiments were carried out at least three times. Antibodies used are listed in **Supplementary Methods**.

Methods and materials for VNTR marker analysis, vestibular testing and allele-specific expression analysis are provided in **Supplementary Methods**.

RESULTS

Exome sequencing revealed an in-frame deletion in *RIPOR2*

To identify the genetic defect underlying the HL in family W97-056 (**Figure 1**), exome sequencing was performed in three affected family members (III:22, IV:20 and IV:25).

After applying the variant filtering and prioritization described above, two variants were shared between the three affected individuals (**Table S1**). A *SPATS1* variant (c.419G>A; p.(Gly140Glu); NM_145026.3), did not completely segregate with HL within the family as 7 out of 23 affected subjects did not harbor the variant (**Figure S1**). Also, *SPATS1* expression was not detected in the mammalian cochlea^{21,22} and *SPATS1* function has only been related to spermatogenesis.²³ Therefore, this variant was deemed non-causative. The in-frame deletion was present in exon 14 of *RIPOR2* (c.1696_1707del; p.(Gln566_Lys569del); NM_014722.3; Chr6:g.24,843,303_24,843,314del; rs760676508). It affects a highly conserved protein region of *RIPOR2* which is present in all *RIPOR2*

isoforms (**Figure S2**). *RIPOR2* has previously been associated with recessively inherited early-onset hearing loss and is positioned 0.9 Mb centromeric of the *DFNA21* locus.^{10,11} No copy number variants were detected that were shared by all three subjects.

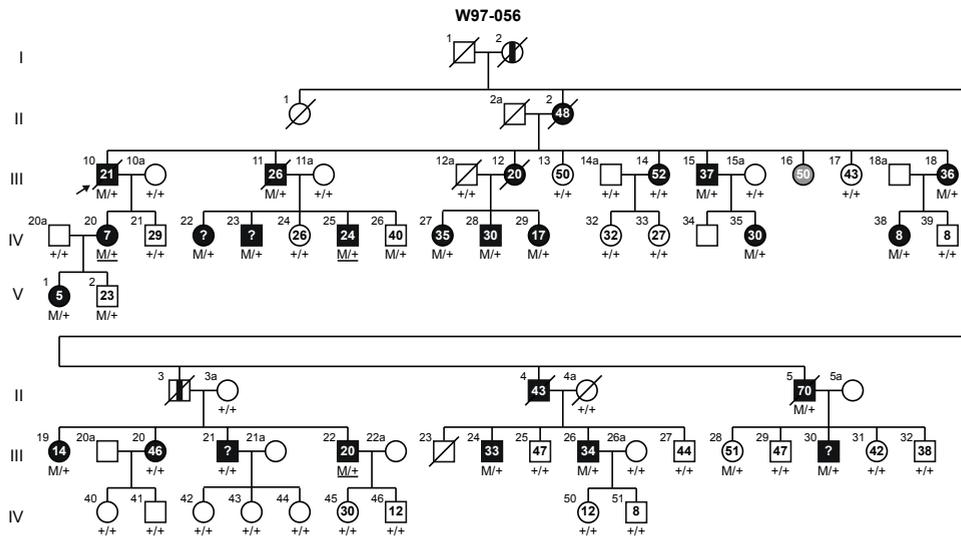


Figure 1. Pedigree of family W97-056 and segregation of *RIPOR2* variant c.1696_1707del. For affected and unaffected family members, the age of onset of hearing loss or the age at the most recent audiometric evaluation are indicated in the pedigree symbols, respectively. Subjects who did not report an age of onset are indicated with a question mark. The index case is marked by an arrow. Exome sequencing was performed in subjects with an underlined genotype. Subjects determined to be affected by heteroanamnesis are indicated with a vertical black bar. The subject marked in grey is diagnosed with intellectual disability and excluded from further participation in this study. Subject identifiers correspond to those in de Brouwer et al., 2005.¹⁰ M, c.1696_1707del; +, wildtype.

Segregation analysis identified the *RIPOR2* variant in 20 of 23 affected subjects of family W97-056 (**Figure 1**). The variant was not found in subjects III:14, III:20, and III:21; a recombination event in subject III:14 previously delimited the centromeric border of the *DFNA21* locus.¹⁰ The *RIPOR2* c.1696_1707del variant was also found in three unaffected family members (V:2, age 23 years; IV:26, age 40 years and III:28, age 51 years). The strong association of the *RIPOR2* variant with HL in this family urged us to further address this and other variants in *RIPOR2* in families with (adult-onset) HL.

The *RIPOR2* variant c.1696_1707del associates with adNSHL in eleven additional families

An exome sequencing dataset of 1,544 index cases with (presumed) hereditary HL was evaluated for rare *RIPOR2* variants. In these cases, (likely) pathogenic variants in known deafness genes were previously addressed in a clinical diagnostic setting. The c.1696_1707del variant was identified in 10 index cases, all diagnosed with adNSHL (**Figure 2**).

Analysis of a dataset obtained through MIP sequencing of 89 HL-associated genes in 64 index cases with (presumed) adNSHL revealed another subject (V:1, W08-1421; **Figure 2**) with this variant. No other rare *RIPOR2* variants ($AF \leq 0.5\%$) that met the variant filtering criteria were identified.

For 6 of the 11 index cases with the c.1696_1707del *RIPOR2* variant, family members were included in the study and segregation analysis was performed (**Figure 2**). The variant was detected in 39 of 40 affected subjects, but not in subject III:10 of family W04-262. As observed in family W97-056, the *RIPOR2* variant was also found in unaffected subjects namely III:14 of family W04-262 and III:4 of family W15-0495, aged 49 and 50 years respectively.

For all 11 index cases, targeted reanalysis of sequencing data for known adNSHL-associated genes⁴ was performed to reveal other (likely) pathogenic variants. No rare variants were identified that both segregated with HL in the family and were classified as (likely) pathogenic in ClinVar²⁴ (**Table S2**).

The presence of an identical *RIPOR2* variant in 12 families of Dutch origin is suggestive for a common ancestor. Indeed, a shared haplotype of ~0.71 Mb, flanking the variant (D6S2439-D6S1281), was observed in the seven families and potentially the five single cases (**Supplementary Results, Figure S3**).

Clinical evaluation of individuals with the c.1696_1707del variant and phenocopies

To characterize the HL associated with the c.1696_1707del *RIPOR2* variant, 200 affected and unaffected subjects from seven families and five single index cases were evaluated between 1997 and 2018. The *RIPOR2* variant was found to be present in 64 of the 200 subjects. Detailed clinical data per individual are provided in **Table S3**.

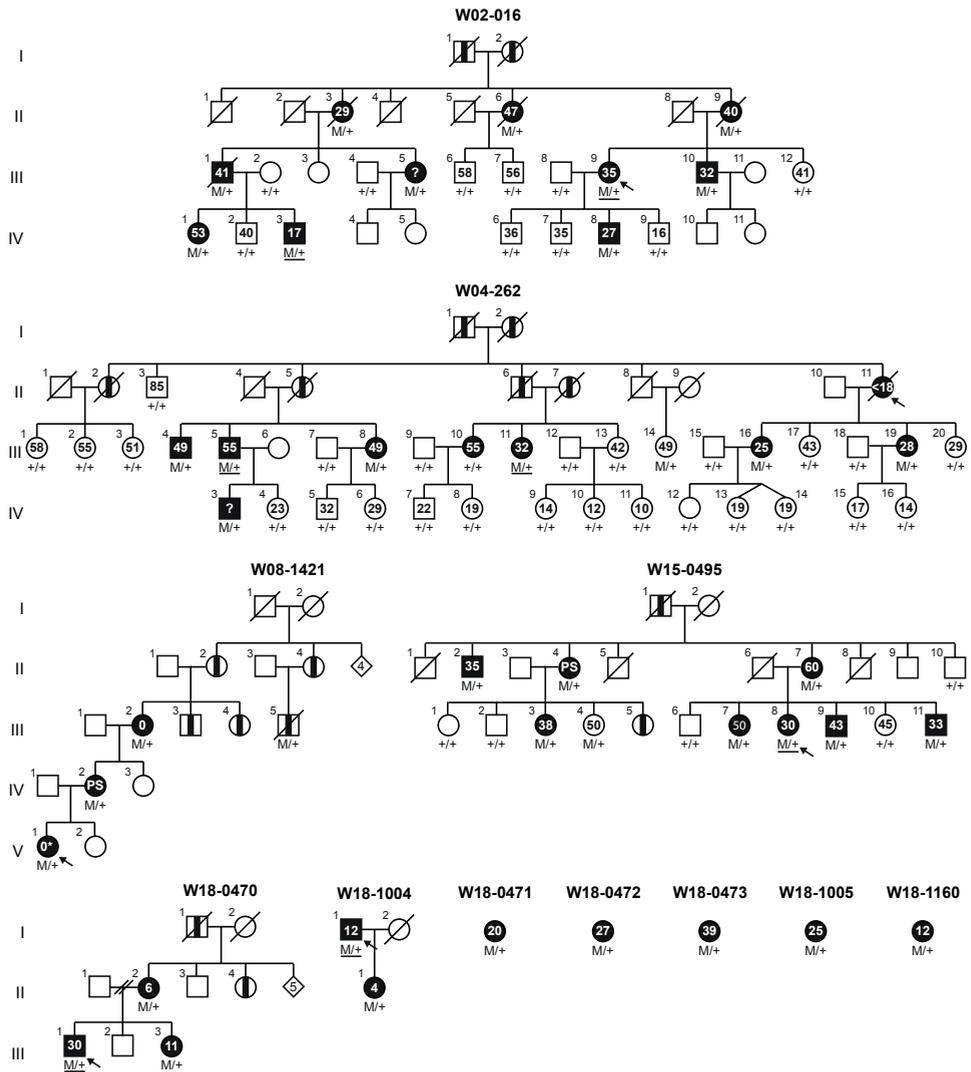


Figure 2. Family pedigrees and segregation of *RIPOR2* variant c.1696_1707del. For affected and unaffected family members, the age of onset of hearing loss or the age at the most recent audiometric evaluation are indicated in the pedigree symbols, respectively. Subjects who did not report an age of onset are indicated with a question mark. Index cases are marked by arrows. Exome sequencing was performed in subjects with an underlined genotype. Subjects determined to be affected by heteroanamnesis are indicated with a vertical black bar. Based on the information provided in the questionnaires, an autosomal dominant inheritance pattern of hearing loss is likely for each of the single cases. M, c.1696_1707del; +, wildtype; PS, primary school.

The mean reported age of onset is 30.6 years (standard deviation 14.9 years) with a wide range from congenital to 70 years (**Figure S4**). Evaluation of audiometric data showed that subjects with the *RIPOR2* variant have progressive sensorineural HL, ranging from mild to profound, with variable audiometric configurations (**Figure 3**, **Figure S5**).

In order to distinguish audiometric patterns, a k-means clustering algorithm, independent of subject age, was applied on the latest audiogram of each subject. This unbiased approach yielded four audiometric patterns, each with a distinct audiometric configuration (**Figure 4**). Asymmetry of HL was seen in 16 cases. Inter-aural differences in progression of HL were also seen (**Figure 3C**). For three subjects (III:30 of family W97-056, II:11 of family W04-262 and V:1 of W08-1421), an explanation for asymmetry was noted (**Table S3**).

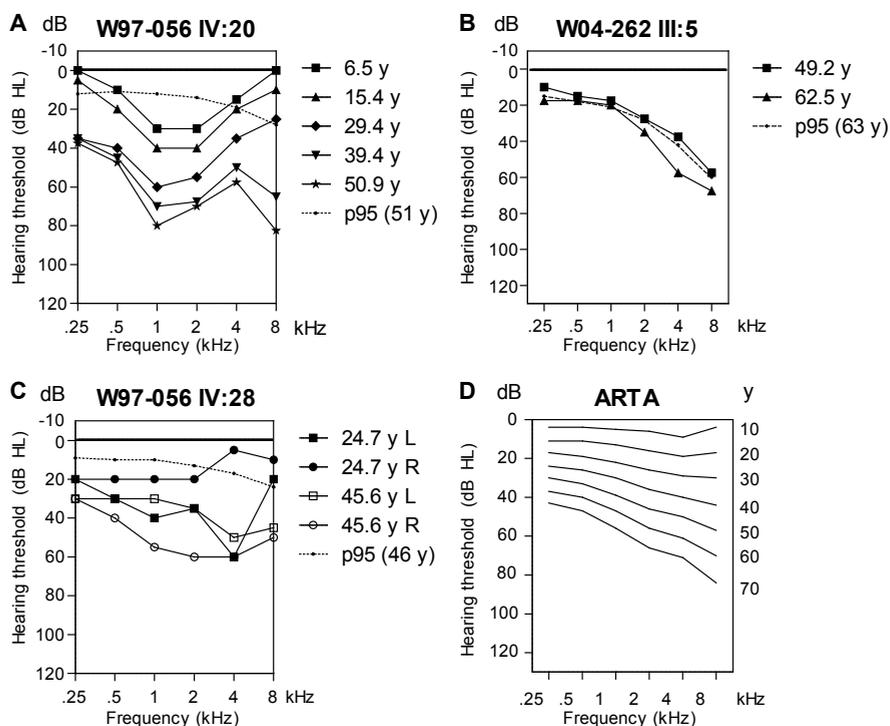


Figure 3. Selection of audiograms and the ARTA. (A-C) Air conduction thresholds of three selected individuals with the c.1696_1707del *RIPOR2* variant are depicted. For the individuals in panels A and B, hearing loss was symmetric and the average of left and right ear thresholds are depicted. For the individual in panel C, hearing loss was asymmetric and the thresholds for both right and left ears are depicted. The p95 values are matched to the individuals' sex and age at most recent audiometry, according to the ISO 7029:2017 standard. (D) Age Related Typical Audiogram (ARTA), cross-sectional linear regression analysis of last visit audiograms of affected subjects with the c.1696_1707del *RIPOR2* variant. y, age in years; R, right; L, left; dB HL, decibel hearing level; kHz; kiloHertz

Longitudinal analysis of HL in individual subjects revealed a large variation in progression of HL between subjects (**Table S3**). We could not identify a specific pattern, such as a certain progression (in dB/y) in certain decades. There was a median progression of 1.2 dB/y (range 0.5-2.7 dB/y), for the frequencies 0.5-4 kHz. Cross-sectional linear regression was applied to calculate ARTA (**Figure 3D**). Progression ranged from 0.7 dB/y (0.25 kHz) to 1.3 dB/y (8 kHz). Progression of HL was significant for all frequencies (*F*-test, *p*-value: <0.0001).

Speech reception thresholds were generally lower than, or comparable to, $PTA_{0.5-2kHz}$ (**Table S3**). This indicates absence of retrocochlear pathology and is in line with normal results of click-evoked ABR in four subjects (**Table S3**). CT and/or MRI of the bilateral temporal bones and cerebellopontine angle in six subjects revealed normal inner and middle ear anatomy (**Table S3**).

Vestibular testing, performed in 11 randomly selected subjects with the *RIPOR2* variant, aged 29 to 71 years, led to the conclusion that c.1696_1707del *RIPOR2* is not associated with vestibular dysfunction (**Table S4**). Further details are provided in the **Supplementary Results**.

Transcript levels of *RIPOR2* do not correlate with age of onset in affected subjects

We hypothesized that the variability in age of onset of the HL associated with the c.1696_1707del *RIPOR2* variant might be explained by differences in expression levels of the wildtype allele. Alternatively, variants in *cis* regulatory elements of the affected allele more distantly located from *RIPOR2*, could influence expression levels of the mutant allele and might thereby modulate the age of onset. To test these hypotheses, allele-specific transcript levels of *RIPOR2* were determined in peripheral blood cells of 33 subjects using quantitative RT-PCR. Subjects were divided into three groups based on self-reported age of onset: <20 years (n=7), 20-39 years (n=15) and ≥ 40 years (n=6). No significant differences were observed between the different subject groups, neither for the wildtype or c.1696_1707del variant *RIPOR2* alleles nor for total *RIPOR2* transcript levels (**Figure S7**). Also, no difference was observed between the ratios of *RIPOR2* mutant to wildtype relative transcript levels. A small difference was observed in total *RIPOR2* transcript levels between subjects with an early onset of HL and controls (*p*=0.0241). This could suggest a trend between low expression levels and an early onset of HL, however, considering the overall variability in transcript levels it is more likely that other factors play a role. A larger sample size would be required to confirm or negate the observed trend.

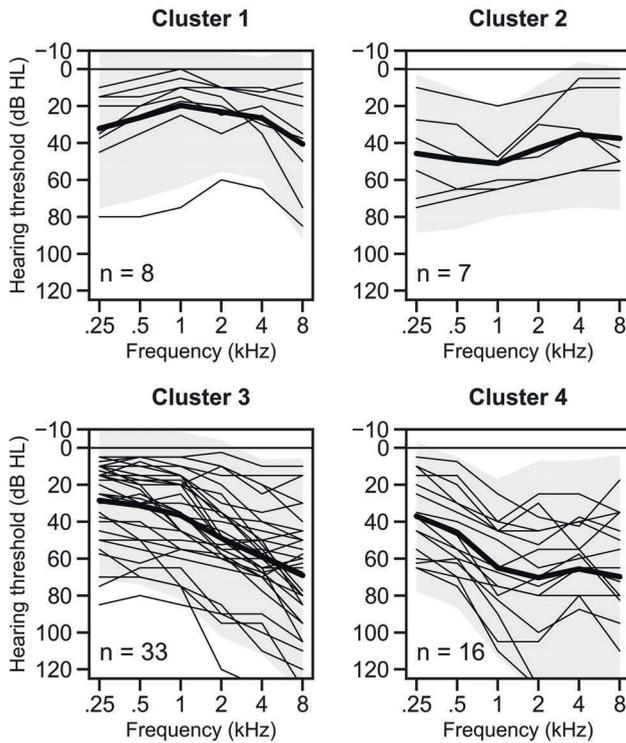


Figure 4. Four audiometric patterns of *RIPOR2*-associated hearing loss. Air conduction thresholds of all subjects were analyzed with a k-means clustering protocol. The thick black lines depict the average of each cluster, the transparent grey areas represent the ± 2 standard deviations. Cluster 1: mild hearing loss (average ($PTA_{0.5-4\text{ kHz}}$)) 23 dB hearing level (HL) with an inverse U-shape audiogram. Cluster 2: moderate hearing loss (average 48 dB HL), with relatively worse hearing in the lower frequencies. Cluster 3: moderate (average 39 dB HL) high-frequency hearing loss with a gently down sloping audiogram configuration (average of 28 dB HL difference between the mean of 0.5-1 and 4-8 kHz). Cluster 4: moderate (average 60 dB HL), mid-frequency hearing loss with a U-shape audiogram, individual audiometry (**Figure S5**) shows relatively faster deterioration of higher frequencies later in life, for example W97-056 IV:20. dB HL, decibel hearing level; kHz, kiloHertz.

The *Ripor2* variant prevents correct localization of the protein in mouse cochlear hair cells

Previous studies have shown that *RIPOR2* is specifically localized to the base of the stereocilia in mouse cochlear hair cells.¹⁹ *RIPOR2* is highly conserved between mouse and human (87% amino acid identity). To study whether the localization of mouse *RIPOR2* with a deletion of the orthologous four amino acid residues (p.584_587del) is altered, plasmids encoding wildtype- or mutant-*RIPOR2* were injectoporated into cochlear outer hair cells of wildtype mice (P2). Interestingly, mutant-*RIPOR2* was detected in the stereocilia but in none of the 12 evaluated cells it was retained at the stereocilia base where the wildtype protein was found to be localized in all 11 evaluated

cells (**Figure 5A**). Morphology of the stereocilia was not significantly affected two days after injectoporation of the mutant *Ripor2* construct, suggesting the mutant protein did not visibly affect the stereocilia structure in the short term.

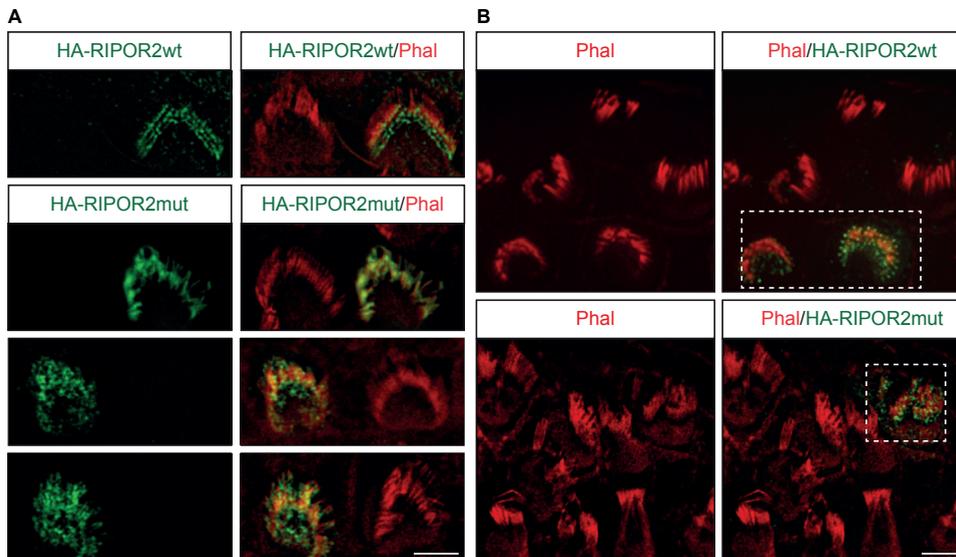


Figure 5. Functionality of mutant RIPOR2 is altered in mouse cochlear outer hair cells. A. Mutant RIPOR2 differed in localization from wildtype RIPOR2 in mouse cochlear outer hair cells. Outer hair cells of wildtype mice were injectoporated at P2 to express murine N-terminally HA-tagged wildtype RIPOR2 (RIPOR2wt) or mutant RIPOR2 (RIPOR2mut). Expression was evaluated after two days by immunohistochemistry and three representative images of cells expressing the mutant RIPOR2 are provided. Eleven cells expressing the wildtype construct and 12 cells expressing the mutant construct were evaluated. **B.** Mutant RIPOR2 did not rescue stereocilia defects in RIPOR2-deficient hair cells. Cochlear explants of RIPOR2-deficient mice were prepared at P2 and injectoporated with constructs RIPOR2wt or RIPOR2mut. After culturing for two days, five out of six cells expressing the wildtype RIPOR2 construct demonstrated rescued hair bundle morphology but none of the 13 cells expressing the mutant RIPOR2 construct. Cells expressing the constructs are boxed. HA-tagged protein was stained in green, stereocilia were stained using phalloidin (phal) conjugated with Alex Fluor 568 (red). Scale bar represents 5 μ m.

Potentially, the variant affects interactions of RIPOR2 that are essential for its localization. The four deleted amino acids are predicted to be part of a disorganized coiled coil structure (predicted using K MAD²⁵). Coiled coil regions are indicated to mediate protein-protein interactions, which supports the hypothesis that the variant affects RIPOR2 protein interactions.²⁶ Co-immunoprecipitation (Co-IP) assays demonstrated that both the dimerization ability (**Figure S8A**) and the interaction with RHOC (**Figure S8B**) of mutant RIPOR2 are intact.

Mutant *RIPOR2* cannot rescue morphological defects in outer hair cells from *Ripor2* knockout mice

In *Ripor2* knockout mice, morphological defects were previously observed in hair cells, which included hair bundle polarity and cohesion and length of stereocilia.¹⁹ After inyectoporation of the *Ripor2* mutant construct into outer hair cells of these mice, these defects could not be rescued in any of the 13 cells expressing mutant-RIPOR2. The typical V-shaped hair bundle was not formed, in contrast to the rescue effect observed in five out of six cells expressing wildtype-RIPOR2 (**Figure 5B**). This, together with the aberrant localization of mutant-RIPOR2, confirms an effect of the 4-amino acid deletion on RIPOR2 function in outer hair cells.

DISCUSSION

This study identified an in-frame 12 nucleotide deletion in *RIPOR2* as a prevalent and highly penetrant genetic factor for adult-onset HL in the Netherlands and beyond. HL associated with the deletion is highly variable in age of onset and audiometric characteristics. Our study exemplifies that an increasing contribution of environmental factors and of low-penetrance genetic factors to the hearing ability during life, complicates the identification of highly penetrant genetic factors in adult-onset HL. This is best illustrated by family W97-056 in which the linkage interval was falsely delimited by a phenocopy.

The *RIPOR2* variant was significantly enriched in an in-house dataset, previously coined "SE-NL" (Southeast Netherlands) with exomes of 22,952 unrelated individuals with unknown hearing abilities.²⁷ Eighteen individuals were heterozygotes for the variant (AF 0.0392%), as compared to 8 of 56,352 individuals (AF 0.0071%) and 5 of 32,287 individuals (AF 0.0077%) of non-Finnish European descent in the gnomAD exome database v2.1.1 and gnomAD genome database v3, respectively. As the variant was indicated to be inherited from a common ancestor, this individual might well be of Dutch origin or of neighboring regions.

Several lines of evidence indicate the association of the c.1696_1707del *RIPOR2* variant with HL. Firstly, the deletion affects four highly conserved amino acids of RIPOR2, which is known to have a crucial role in murine and zebrafish hair cell development, function, and maintenance.^{11,19,28} *Ripor2* knockout mice are already found to be deaf at four weeks of age due to impaired mechanotransduction.¹⁹ Also, knockdown of *ripor2* in zebrafish induced loss of hair cells, and consequently profound hearing loss.¹¹ Secondly, aberrant localization of the mutant RIPOR2 in early postnatal mouse hair cells, *ex vivo*, and failure

to rescue the stereocilia defects of *Ripor2* knockout mice indicate a functional effect of the variant. Thirdly, neither other rare potentially causative variants in protein coding regions and splice sites of the shared haplotype region, nor structural variants affecting this region were revealed in exome or genome sequencing.

RIPOR2 is localized at the taper region of the mechanically sensitive stereocilia of murine hair cells^{11,19,28} where it is organized in a ring-like fashion.¹⁹ The latter is thought to be achieved by homo-oligomerization in a head-to-head and tail-to-tail manner, regulated by RHOC.¹⁹ The oligomerization is essential for the structure of the taper region and for the morphology of the hair bundle as a whole, but the precise molecular mechanism is still elusive. The taper region is the specialized basal part of stereocilia that allows their deflection upon mechanical stimulation.²⁹ CLIC5, PTPRQ, MYO6, TPRN, RDX, GRXCR2, and RIPOR2 are described to concentrate and co-function in the taper region and to be crucial for its structure and/or for hair bundle development and maintenance in mice.^{19,30-35} Direct interactions of these proteins are indicated, e.g., of CLIC5, RDX and TPRN, but not RIPOR2.^{19,32} Also, interdependence for their concentration in the taper region was observed.^{19,31,32} In RIPOR2-deficient hair cells, for example, TPRN is no longer concentrated at the stereociliary base.^{19,30} Depletion of TPRN in *Tprn* knock-out mice leads to functional as well as (slowly) progressive morphological abnormalities of the stereocilia bundle.³⁰

Based on the above described molecular structure of the stereociliary taper, we hypothesize that p.(Gln566_Lys569)del RIPOR2 affects this taper region and thereby the durability of the hair bundle, potentially via an effect on TPRN. Additionally or alternatively, the *RIPOR2* variant might affect the amount of the RIPOR2-interaction partner MYH9 in stereocilia, as well as the abundance of phosphorylated MYH9 and acetylated α -tubulin in the kinocilia, as these proteins are reduced in RIPOR2-deficient mice.²⁸ Interestingly, *MYH9* defects in humans are also associated with progressive HL.³⁶

In light of developing therapeutic strategies, it is essential to determine whether the *RIPOR2* variant has a loss-of-function, a dominant negative or toxic gain-of-function effect. A haploinsufficiency effect of the variant seems to be the least plausible, as a loss-of-function *RIPOR2* variant in the heterozygous state was not indicated to be associated with HL.¹¹ Also, heterozygous *Ripor2* knockout mice displayed no significant hearing loss at four weeks¹⁹ and two months of age (Zhao, unpublished data). A dominant-negative effect of the p.(Gln566_Lys569del) variant cannot be excluded as an interaction between the mutant- and wildtype-RIPOR2 was detected in Co-IP assays.

However, a strong dominant negative effect would be expected to result in early-onset HL, comparable to that associated with the homozygous loss-of-function variant.¹¹ Therefore, we hypothesize that the variant has a toxic gain-of-function effect.

RIPOR2 is expressed in a wide-range of tissues and cell types.¹⁹ It is a known inhibitor of the small G-protein RHOA in neutrophils and T lymphocytes, where it regulates migration of these cells.³⁷ Additionally, *RIPOR2* is upregulated during muscle cell differentiation and induces the formation of filopodia.³⁸ We did not observe an effect of the four amino acid-deletion on filopodia formation (de Bruijn, unpublished data) which is in line with the fact that the deleted residues are not part of the RHOA-interaction domain.³⁸ This might, at least in part, explain that the *RIPOR2* variant leads to HL only. The variant could affect a cochlear-specific protein interaction that determines *RIPOR2* localization in the hair bundle. Furthermore, in tissues other than the inner ear loss of *RIPOR2* function might be compensated by *RIPOR1* and *RIPOR3* which are described to have redundant functions.^{39,40} Indeed, RNA levels of both *RIPOR1* and *RIPOR3* are low in hair cells (gEAR).⁴¹

The audiometric phenotype and age of onset of HL associated with c.1696_1707del *RIPOR2* displayed variation. Such intrafamilial phenotypic variation has also been reported for defects in several of the genes that can be associated with adult-onset adNSHL, e.g. *EYA4*, *MYO6* and *POU4F3*, and remains unexplained.⁴²⁻⁴⁴ Non-penetrance is an extreme of phenotypic variability. In our study, four subjects with the c.1696_1707del *RIPOR2* variant had normal hearing: V:2, IV:26, III:28, (W97-056), III:14 (W04-262), and III:4 (W15-0495). They are aged 23, 40, 51, 49, and 50 years, respectively, at the latest audiometric evaluation. The average reported age of onset in the studied families is 30 years (SD 14.9) and 70 years the highest reported onset age. Therefore, the unaffected subjects might develop HL in the future. However, incomplete penetrance of the variant cannot be excluded.

With a k-means cluster analysis, four distinct audiometric clusters could be distinguished. It is possible that subjects, due to increasing age, may go from one cluster to another cluster, which is not captured by the k-means clustering algorithm, since no longitudinal data are used. As no clear patterns of age of onset or audiometric configurations were observed within families or family branches with the *RIPOR2* variant, the phenotypic variability might well result from an interplay between environmental and genetic modifying factors. We have addressed differences in transcript levels of both wildtype and mutant *RIPOR2* alleles as potential modifiers of age of onset but no clear correlations were observed. As the analysis was performed on RNA extracted from peripheral blood, we cannot exclude that *RIPOR2* mRNA levels determined by cochlear-specific *cis* or *trans* regulatory elements modify the onset of HL. Other candidate genetic modifiers are

variants in the genes that encode proteins of the indicated complex of the stereocilia taper. As the taper region is thought to be essential for anchoring the mechanosensory stereocilia, noise exposure is an obvious candidate environmental modifying factor. Fourteen subjects with the *RIPOR2* variant reported noise exposure. However, we could not correlate onset or strong progression of HL with a preceding significant noise exposure.

Four subjects with HL who did not have the *RIPOR2* variant, are considered to be phenocopies. In the light of the heterogeneity in the etiology of HL, the occurrence of phenocopies is not unexpected. For individuals III:14 and III:20 (W97-056) a possible explanation for their HL is a Ménière-like disease and heavy smoking (COPD Gold III), respectively.^{45,46} Subject III:10 (W04-262) might have inherited a cause of HL associated with vestibular problems from her mother, who married into the family. HL in subject III:21 of family W97-056 remains unexplained.

The c.1696_1707del *RIPOR2* variant was only reported in non-Finnish Europeans, with the exception of a single individual of African origin (gnomAD v3 genomes). Assuming that the AF of 0.0392% determined in the SE-NL cohort is comparable throughout the Netherlands, the c.1696_1707del *RIPOR2* variant is estimated to be present in more than 13,000 individuals who are therefore at risk to develop HL or have developed HL already due to this variant. About 30,000 additional individuals can be calculated to be at risk, based on the AF of 0.0096% of the variant in Northwest Europe (gnomAD v2.1.1) with ~156 million inhabitants (United Nations Population Division estimates, 2019). This large number of individuals at risk to develop HL due to the c.1696_1707del *RIPOR2* variant illustrates the need to gain broader estimates of the penetrance of the variant which was ~90% at the age of 50 years in the studied families. However, this calculated penetrance cannot be excluded to be biased because these families were included based on index cases with HL. Further insight in the age-related penetrance of c.1696_1707del *RIPOR2* will pave the way for the identification of modifying factors which may convey handles for prevention.

In conclusion, we demonstrate that an adult-onset type of HL (DFNA21) is relatively common and associated with a “mild” variant in *RIPOR2*. Potentially, thousands of individuals in the Netherlands and beyond are at risk to develop HL. More such variants might well wait to be “unmasked” as (population-specific) frequent and highly penetrant causes of adult-onset HL. Because of the large number of subjects estimated to be at risk for HL due to the c.1696_1707del *RIPOR2* variant, it is an attractive target for the development of a genetic therapy. The great progress that is being made for this in hearing disorders is promising.⁴⁷

ACKNOWLEDGEMENTS

The authors thank José van den Broek, Herman Kok, Lianne Schenk, Daniëlle Manders, Karin Krommenhoek and Jacqueliën Jilissen for their contributions to clinical evaluations and Saskia van der Velde-Visser for her contributions in EBV transformations and cell culture. We would like to thank Sita Vermeulen, Galuh Astuti and Hanka Venselaar for their contribution to statistical and bioinformatic analyses. This study was financially supported by a DCMN Radboudumc grant, by a grant of the Heinsius-Houbolt foundation and an NIH/NIDCD (R01 DC017147) grant.

CONSORTIA

DOOFNL consortium: The DOOFNL consortium is a Dutch nationwide collaboration on hereditary hearing loss and consists of M.F. van Dooren, H.H.W. de Gier, E.H. Hoefsloot, M.P. van der Schroeff (ErasmusMC, Rotterdam, the Netherlands), S.G. Kant, L.J.C. Rotteveel, F.G. Ropers (LUMC, Leiden, the Netherlands), J.C.C. Widdershoven, J.R. Hof, E.K. Vanhoutte (MUMC+, Maastricht, the Netherlands), R.J.C. Admiraal, I. Feenstra, H. Kremer, C.P. Lanting, R.J.E. Pennings, H.G. Yntema (Radboudumc, Nijmegen, the Netherlands), R.H. Free and J.S. Klein Wassink-Ruiter (UMCG, Groningen, the Netherlands), R.J. Stokroos, A.L. Smit, M.J. van den Boogaard (UMC, Utrecht, the Netherlands) and F.A. Ebbens, S.M. Maas, A. Plomp, T.P.M. Goderie, P. Merkus and J. van de Kamp (Amsterdam UMC, Amsterdam, the Netherlands).

REFERENCES

1. World Health Organisation. The global burden of disease: 2004 update. (Geneva: World Health Organization, 2008).
2. Sloan-Heggen, C.M., Bierer, A.O., Shearer, A.E., Kolbe, D.L., Nishimura, C.J., Frees, K.L. *et al.* Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Human Genetics* **135**, 441-450 (2016).
3. Zazo Seco, C., Wesdorp, M., Feenstra, I., Pfundt, R., Hehir-Kwa, J.Y., Lelieveld, S.H. *et al.* The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. *European Journal of Human Genetics* **25**, 308-314 (2017).
4. Van Camp, G. & Smith, R. Hereditary Hearing Loss Homepage. Available from: <https://hereditaryhearingloss.org>.
5. Gates, G.A., Couropmitree, N.N. & Myers, R.H. Genetic associations in age-related hearing thresholds. *Archives of Otolaryngology-Head & Neck Surgery* **125**, 654-659 (1999).
6. Karlsson, K.K., Harris, J.R. & Svartengren, M. Description and primary results from an audiometric study of male twins. *Ear and Hearing* **18**, 114-120 (1997).
7. Wolber, L.E., Steves, C.J., Spector, T.D. & Williams, F.M.K. Hearing ability with age in northern European women: a new web-based approach to genetic studies. *PLoS One* **7**, e35500 (2012).
8. Lewis, M.A., Nolan, L.S., Cadge, B.A., Matthews, L.J., Schulte, B.A., Dubno, J.R. *et al.* 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 Medical Genomics* **11**, 77 (2018).
9. Kunst, H., Marres, H., Huygen, P., Van Duijnhoven, G., Krebsova, A., Van der Velde, S. *et al.* Non-syndromic autosomal dominant progressive non-specific mid-frequency sensorineural hearing impairment with childhood to late adolescence onset (DFNA21). *Clinical Otolaryngology & Allied Sciences* **25**, 45-54 (2000).
10. De Brouwer, A.P.M., Kunst, H.P.M., Krebsova, A., van Asseldonk, K., Reis, A., Snoeckx, R.L. *et al.* Fine mapping of autosomal dominant nonsyndromic hearing impairment DFNA21 to chromosome 6p24.1-22.3. *American Journal of Medical Genetics Part A* **137A**, 41-46 (2005).
11. Diaz-Horta, O., Subasioglu-Uzak, A., Grati, M.h., DeSmidt, A., Foster, J., Cao, L. *et al.* FAM65B is a membrane-associated protein of hair cell stereocilia required for hearing. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9864-9868 (2014).
12. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. & Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310-315 (2014).
13. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. & Ng, P.C. SIFT missense predictions for genomes. *Nature Protocols* **11**, 1 (2015).
14. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249 (2010).
15. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**, 361 (2014).

16. M. Mazzoli, G.V.C., V. Newton, N. Giarbini, F. Declau, A. Parving. Recommendations for the description of genetic and audiological data for families with nonsyndromic hereditary hearing impairment. *GeneReviews* (2003).
17. Smits, J.J., Oostrik, J., Beynon, A.J., Kant, S.G., de Koning Gans, P.A.M., Rotteveel, L.J.C. et al. De novo and inherited loss-of-function variants of ATP2B2 are associated with rapidly progressive hearing impairment. *Human Genetics* **138**, 61-72 (2019).
18. Huygen, P., Pennings, R. & Cremers, C. Characterizing and distinguishing progressive phenotypes in nonsyndromic autosomal dominant hearing impairment. *Audiological Medicine*, 37-46 (2003).
19. Zhao, B., Wu, Z. & Müller, U. Murine Fam65b forms ring-like structures at the base of stereocilia critical for mechanosensory hair cell function. *eLife* **5**, e14222 (2016).
20. de Bruijn, S.E., Verbakel, S.K., de Vrieze, E., Kremer, H., Cremers, F.P.M., Hoyng, C.B. et al. Homozygous variants in KIAA1549, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa. *Journal of Medical Genetics* **55**, 705-712 (2018).
21. Schrauwen, I., Hasin-Brumshtein, Y., Corneveaux, J.J., Ohmen, J., White, C., Allen, A.N. et al. A comprehensive catalogue of the coding and non-coding transcripts of the human inner ear. *Hearing Research* **333**, 266-274 (2016).
22. Scheffer, D.I., Shen, J., Corey, D.P. & Chen, Z.-Y. Gene expression by mouse inner ear hair cells during development. *Journal of Neuroscience* **35**, 6366-6380 (2015).
23. Capoano, C.A., Wettstein, R., Kun, A. & Geisinger, A. Spats 1 (*Srsp1*) is differentially expressed during testis development of the rat. *Gene Expression Patterns* **10**, 1-8 (2010).
24. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research* **42**, D980-D985 (2014).
25. Lange, J., Wyrwicz, L.S. & Vriend, G. KMAD: knowledge-based multiple sequence alignment for intrinsically disordered proteins. *Bioinformatics* **32**, 932-936 (2016).
26. Rose, A. & Meier, I. Scaffolds, levers, rods and springs: diverse cellular functions of long coiled-coil proteins. *Cellular and Molecular Life Sciences CMLS* **61**, 1996-2009 (2004).
27. Cremers, F.P.M., Cornelis, S.S., Runhart, E.H. & Astuti, G.D.N. Author response: penetrance of the ABCA4 p.Asn1868Ile allele in Stargardt disease. *Investigative Ophthalmology & Visual Science* **59**, 5566-5568 (2018).
28. Diaz-Horta, O., Abad, C., Cengiz, F.B., Bademci, G., Blackwelder, P., Walz, K. et al. *Ripor2* is involved in auditory hair cell stereociliary bundle structure and orientation. *Journal of Molecular Medicine* **96**, 1227-1238 (2018).
29. Barr-Gillespie, P.-G. Assembly of hair bundles, an amazing problem for cell biology. *Molecular Biology of the Cell* **26**, 2727-2732 (2015).
30. Men, Y., Li, X., Tu, H., Zhang, A., Fu, X., Wang, Z. et al. *Tprn* is essential for the integrity of stereociliary rootlet in cochlear hair cells in mice. *Frontiers of Medicine* **13**, 690-704 (2018).
31. Liu, C., Luo, N., Tung, C.-Y., Perrin, B.J. & Zhao, B. *GRXCR2* regulates taperin localization critical for stereocilia morphology and hearing. *Cell Reports* **25**, 1268-1280 (2018).

32. Salles, F.T., Andrade, L.R., Tanda, S., Grati, M., Plona, K.L., Gagnon, L.H. *et al.* CLIC5 stabilizes membrane-actin filament linkages at the base of hair cell stereocilia in a molecular complex with radixin, taperin, and myosin VI. *Cytoskeleton (Hoboken)* **71**, 61-78 (2014).
33. Sakaguchi, H., Tokita, J., Naoz, M., Bowen-Pope, D., Gov, N.S. & Kachar, B. Dynamic compartmentalization of protein tyrosine phosphatase receptor Q at the proximal end of stereocilia: implication of myosin VI-based transport. *Cell Motil Cytoskeleton* **65**, 528-38 (2008).
34. Pataky, F., Pironkova, R. & Hudspeth, A.J. Radixin is a constituent of stereocilia in hair cells. *Proceedings of the National Academy of Sciences* **101**, 2601-2606 (2004).
35. Gagnon, L.H., Longo-Guess, C.M., Berryman, M., Shin, J.B., Saylor, K.W., Yu, H. *et al.* The chloride intracellular channel protein CLIC5 is expressed at high levels in hair cell stereocilia and is essential for normal inner ear function. *Journal of Neuroscience* **26**, 10188-98 (2006).
36. Lalwani, A.K., Goldstein, J.A., Kelley, M.J., Luxford, W., Castelein, C.M. & Mhatre, A.N. Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in nonmuscle myosin MYH9. *American Journal of Human Genetics* **67**, 1121-1128 (2000).
37. Gao, K., Tang, W., Li, Y., Zhang, P., Wang, D., Yu, L. *et al.* Front-signal-dependent accumulation of the RHOA inhibitor FAM65B at leading edges polarizes neutrophils. *Journal of Cell Science* **128**, 992-1000 (2015).
38. Yoon, S., Molloy, M.J., Wu, M.P., Cowan, D.B. & Gussoni, E. C6ORF32 is upregulated during muscle cell differentiation and induces the formation of cellular filopodia. *Developmental Biology* **301**, 70-81 (2007).
39. Mardakheh, F.K., Self, A. & Marshall, C.J. RHO binding to FAM65A regulates Golgi reorientation during cell migration. *Journal of Cell Science* **129**, 4466-4479 (2016).
40. Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S. *et al.* IntAct: an open source molecular interaction database. *Nucleic Acids Research* **32**, D452-D455 (2004).
41. Cremers, F.P.M., den Dunnen, J.T., Ajmal, M., Hussain, A., Preising, M.N., Daiger, S.P. *et al.* Comprehensive registration of DNA sequence variants associated with inherited retinal diseases in Leiden Open Variation Databases. *Human Mutation* **35**, 147-148 (2014).
42. Oonk, A.M.M., Leijendeckers, J.M., Lammers, E.M., Weegerink, N.J.D., Oostrik, J., Beynon, A.J. *et al.* Progressive hereditary hearing impairment caused by a MYO6 mutation resembles presbycusis. *Hearing Research* **299**, 88-98 (2013).
43. Vahava, O., Morell, R., Lynch, E.D., Weiss, S., Kagan, M.E., Ahituv, N. *et al.* Mutation in transcription factor POU4F3 associated with inherited progressive hearing loss in humans. *Science* **279**, 1950-1954 (1998).
44. O'Neill, M.E., Marietta, J., Nishimura, D., Wayne, S., Van Camp, G., Van Laer, L. *et al.* A gene for autosomal dominant late-onset progressive non-syndromic hearing loss, DFNA10, maps to chromosome 6. *Human Molecular Genetics* **5**, 853-6 (1996).
45. Bayat, A., Saki, N., Nikakhlagh, S., Mirmomeni, G., Raji, H., Soleimani, H. *et al.* Is COPD associated with alterations in hearing? A systematic review and meta-analysis. *International Journal of Chronic Obstructive Pulmonary Disease* **14**, 149-162 (2018).

46. Hu, H., Sasaki, N., Ogasawara, T., Nagahama, S., Akter, S., Kuwahara, K. *et al.* Smoking, smoking cessation, and the risk of hearing loss: Japan epidemiology collaboration on occupational health study. *Nicotine & Tobacco Research* **21**, 481-488 (2018).
47. Omichi, R., Shibata, S.B., Morton, C.C. & Smith, R.J.H. Gene therapy for hearing loss. *Human Molecular Genetics* **28**, R65–R79 (2019).

SUPPLEMENTARY MATERIALS AND METHODS

DNA sequencing and variant identification

Genomic DNA was isolated from peripheral blood lymphocytes following standard procedures. Subsequently, exome enrichment was performed using the Agilent SureSelect Human All Exome V5 kit according to the manufacturer's protocols. Exome sequencing was performed on an Illumina HiSeq system by BGI Europe (Copenhagen, Denmark). Read mapping along the hg19 reference genome (GRCh37/hg19) and variant calling were performed using BWA V.0.78¹ and GATK HaplotypeCaller V.3.3². A coverage of >20 reads was reached for 85.1% to 97.8% of the enriched regions. For variant annotation an in-house developed annotation and variant evaluation pipeline was used. For sequencing data of family W97-056, copy number variant (CNV) detection was performed using CoNIFER V.0.2.2.³ Genome sequencing was performed by BGI (Hong Kong, China) on a BGISeq500 using a 2x 100 bp paired end module, with a minimal median coverage of 30-fold per genome. Structural variants were called using Manta V.1.1.0⁴ and CNVs using Control-FREEC.⁵ Variants were validated and visualized using the IGV Software (V.2.4).⁶

In the index case of family W08-1421, targeted DNA sequencing was performed using MIP sequencing.⁷ MIPs were designed covering exons and exon-intron boundaries of a panel of 89 HL genes (**Table S6**). Sequencing and data analysis were performed as previously described.⁸ For each targeted region, an average coverage of 420 reads was obtained. A coverage of >20 reads was reached for 85.4% of the MIPs. Only those called variants were considered that had a quality-by-depth >200 and that were present in less than 10% of the samples that were analyzed in the same sequence run (n=150).

VNTR marker analysis

Genotyping of Variable Number of Tandem Repeats (VNTR) markers was performed by genomic DNA amplification using touchdown PCR and analysis on an ABI Prism 3730 Genetic Analyzer (Applied Biosystems). Genomic positions of markers were determined using the UCSC genome browser (human genome assembly GRCh37/hg19). Alleles were assigned with the GeneMarker software (V.2.6.7, SoftGenetics) according to the manufacturer's protocol.

Audiometric cluster analysis

A k-means clustering algorithm was applied on the last audiogram of affected subjects with the *RIPOR2* variant.⁹ Each audiogram was first normalized by subtracting the average hearing threshold across all frequencies from the data, preventing the algorithm

to select clusters based on average hearing threshold, while retaining the overall shape of the audiogram. Next, the optimal number of clusters in the data was obtained by using the Elbow method;¹⁰ for a number of $k=1$ to $k=15$ clusters. The distortion, i.e. the sum of squared distances from each point to its assigned center, was determined for each value of k . The smallest number of k clusters with the lowest distortion (i.e. the elbow) was then taken as the optimal number of clusters. Finally, the audiograms of all the patients within each cluster were visualized and an average cluster prototype was obtained by averaging all audiograms within a cluster.

Vestibular testing

Vestibular function was assessed by electronystagmography, caloric irrigation testing, rotary chair stimulation and video head impulse tests, as described previously.¹¹ Cervical and ocular vestibular-evoked myogenic potentials (cVEMP/oVEMP) were measured to assess saccular and utricular function, respectively.^{12,13} When responses were seen at or below 100 dB Hearing Level during (air conducted) cVEMP testing, saccular function was considered to be present, otherwise absent.¹² For (bone conducted) oVEMP stimulation, this normal value is ≤ 140 dB Force Level.¹³

Allele-specific expression analysis

Peripheral blood (2.5 ml) was collected in PAXgene Blood RNA tubes (BD Biosciences). RNA was isolated using the PAXgene blood RNA kit (Qiagen) following the manufacturer's protocol. Subsequently, cDNA was prepared using the iScript cDNA synthesis kit (Bio-Rad) with 500 ng RNA. Allele-specific primer sets were designed and validated; the design was based on a forward primer that specifically hybridizes to either the mutant or wildtype *RIPOR2* alleles. Additionally, primers were designed for exons 3-4 of *RIPOR2*, as well as for exons 2-3 of the reference gene *GUSB* (NM_000181). Primer sequences are provided in **Table S7**. All qPCR reaction mixtures were prepared with the GoTaq qPCR Master Mix (Promega) according to the manufacturer's protocol. Amplifications were performed with the Applied Biosystem Fast 7900 System (Applied Biosystems). For all RNA samples, cDNA was synthesized twice, and all qPCR reactions were performed in duplicate. Relative gene expression levels, as compared to the internal reference gene *GUSB*, were determined with the ΔC_t method.¹⁴ Statistical analyses were performed using a one-way ANOVA followed by Tukey's multiple comparison test to test for significance between the groups.

Injectoporation of *Ripor2*-constructs, immunostaining and antibodies

For injectoporation, the organ of Corti was isolated and placed in DMEM/F12 medium with 1.5 µg/ml ampicillin. Glass electrodes (~2 µm diameter) were used to deliver the plasmid (500 ng/µl in Hank's Balanced Salt Solution (HBSS)) to the sensory epithelium. A series of 3 pulses were applied at 1 sec intervals with a magnitude of 60V and duration of 15 msec (ECM 830 square wave electroporator; BTX). Two days after injectoporation, samples were fixed in the fixative containing 4% paraformaldehyde in HBSS for 20 min. Tissues were then washed in HBSS and blocked for 20 min at room temperature in HBSS containing 5% BSA, 1% goat serum and 0.5% Triton X-100, and then incubated overnight at 4°C with primary antibodies in HBSS containing 1% BSA and 0.1% Triton X-100. Tissues were washed in HBSS and incubated 2 hours at room temperature with secondary antibodies. Tissues were mounted in ProLong® Antifade Reagents (ThermoFisher). Stacked images were then captured by fluorescence deconvolution microscope (Leica). Antibodies used were: anti-HA (mouse; 1:500; cat.#2367S; Cell Signaling), Alex Fluor 568-phalloidin (1:500; cat.#A12380; ThermoFisher) and Alexa Fluor 488 goat anti-mouse (1:1000; cat.#A11017; ThermoFisher).

Antibodies used for co-immunoprecipitations were: anti-HA (mouse; 1:500; cat.#2367S; Cell Signaling), anti-Myc (rabbit; 1:500; cat.#2278S; Cell Signaling), anti-Myc (mouse; 1:500; cat.#9E10; Santa Cruz); anti-GFP (mouse; 1:1000; cat.#SC-9996; Santa Cruz).

SUPPLEMENTARY RESULTS

The *RIPOR2* c.1696_1707del variant is derived from a common ancestor

VNTR marker analysis was performed to determine whether a haplotype of the chromosomal region flanking the variant was shared by the different families. Indeed, a shared haplotype of ~1.0 Mb, delimited by markers D6S2439 and D6S1281, was found in the seven families for which segregation analysis of the marker alleles could be performed (**Figure S3**). This haplotype was also potentially shared by the five single cases. For marker D6S1545, a different CA-repeat length was determined on the variant-carrying allele of family W18-0470 whereas the alleles of two more centromeric markers were still shared. Since a rare event that caused a repeat length change of the D6S1545 allele may have occurred, this marker locus was still considered to be part of the shared haplotype. To further refine the shared haplotype, we extracted homozygous SNPs present in the region between D6S2439 and D6S1281 from the exome sequencing datasets. Subsequently, homozygous SNP genotypes were compared between all

index cases and discordant alleles were seen for SNP rs6901322 (Chr6: 24,583,804) that is located between D6S2439 and D6S1554. This SNP was found in homozygous state in subject IV:20 (W97-056), but was absent in the index cases of families W02-016 (III:9) and W18-0473. Based on these results, the shared haplotype is delimited by SNP rs6901322 at the telomeric side and comprises a region of 0.713 Mb. Genome sequencing in two members of family W97-056 (IV:25 and III:22) excluded potentially causative CNVs or other structural variants that are present within the shared chromosomal region.

The c.1696_1707del *RIPOR2* variant is not associated with vestibular dysfunction

Four of 64 subjects with the *RIPOR2* variant had vestibular complaints. Subjects III:1, III:9, IV:1 (W02-016) and the index cases of family W18-0472 reported infrequent vertigo attacks, complaints after cochlear implant surgery, a diagnosis of benign paroxysmal positional vertigo and migrainous vertigo, respectively (**Table S3**). Vestibular testing was randomly performed in 9 subjects with the *RIPOR2* variant, aged 29 to 71 years, and included the abovementioned subjects III:1 and III:9 of family W02-016. No abnormalities were found, except for a mild hyporeflexia in subject III:1 of family W02-016 (**Table S4**), which is appropriate for the subject's age of 71 years. Based on these results, we conclude that c.1696_1707del *RIPOR2* is not associated with vestibular dysfunction. This is in line with the lack of vestibular dysfunction in *Ripor2*^{-/-} mice despite expression of the gene in the vestibular organ of wildtype mice.¹⁵ Also, humans with recessively inherited HL caused by a homozygous loss-of-function defect in *RIPOR2*, did not report balance problems, vertigo or dizziness but absence of a vestibular phenotype was not confirmed by objective vestibular testing.¹⁶

SUPPLEMENTARY TABLES

Table S1. Shared rare WES variants in family W97-056

Genome	Gene	Transcript	cDNA	Protein	In-house AF (%)	gnomAD_E AF (%)	gnomAD_G AF (%)	CADD_PHRD	SIFT	PPH2	MutationTaster (prob)
Chr1: 248,059,798T>A	OR2W3	NM_001001957.2	c.910T>A	p.(Leu304Met)	0.07	0.02	0.03	6.504	0.1	0.032	Polymorphism (1.0)
Chr6: 15,501,310C>T	JARID2	NM_004973.3	c.2118C>T	p.(Leu706=)	0.18	0.03	0.04	NA	NA	NA	NA
Chr6: 16,146,884C>T	MYLIP	NM_013262.3	c.1249-9C>T	-	0.01	-	-	NA	NA	NA	NA
Chr6: 24,843,303_24,843,314del	RIPOR2	NM_014722.3	c.1696_1707del	p.(Gln566_Lys569del)	0.08	0.00	-	NA	NA	NA	NA
Chr6: 41,196,733C>T	TREML4	NM_198153.2	c.345C>T	p.(Ser115=)	0.47	0.31	0.30	NA	NA	NA	NA
Chr6: 44,329,574G>A	SPATS1	NM_145026.3	c.419G>A	p.(Gly140Glu)	0.18	0.11	0.12	19.16	0.02	1.0	Disease causing (0.98)

Variants identified by whole exome sequencing (WES) that are shared by all three index cases of family W97-057 and have an allele frequency of $\leq 0.5\%$ in gnomAD and the in-house database (~7,500 exomes). For none of the variants is an effect on transcript splicing predicted nor are any reported in the ClinVar database. Scores that meet the thresholds for pathogenicity as described in the methods section are indicated in red. Thresholds for pathogenicity: CADD-PHRD (≥ 15), SIFT (≤ 0.05), PolyPhen-2 (≥ 0.450) and MutationTaster (disease causing). Genome, Genomic positions according to GRCh37/hg19; In-house AF, allele frequency (%) in the in-house database (~15,000 alleles); GnomAD_E AF and GnomAD_G AF, allele frequencies (%) in respectively gnomAD exome or genome databases; CADD_PHRD, Combined Annotation Dependent Depletion PHRED score; SIFT, Scale-Invariant Feature Transform; PPH2, Poly-Phen-2 score; MutationTaster (prob), MutationTaster score with probability (0-1); -, frequency not available; NA, not applicable.

Table S2. Rare variants in the index cases in genes known to be associated with adHL

Family	Genome	Gene	Transcript	cDNA	Protein	In-house AF (%)	gnomAD_E AF (%)	gnomAD_G AF (%)	CADD_PHRED	SIFT	PPH2	MutationTaster (prob)	ClinVar
W18-0470	Chr22: 36681327T>C	MYH9	NM_002473.4	c.5323A>G	p.(Lys1775Glu)	0.18	0.15	0.19	22.2	0.03	0.120	Disease causing (1.0)	UV2
W18-0473	Chr4: 6303119C>T	WFS1	NM_006005.3	c.1597C>T	p.(Pro533Ser)	0.22	0.07	0.08	19.64	0.00	1.000	Disease causing (1.0)	UV1-UV3
	Chr22: 36700183G>A	MYH9	NM_002473.4	c.2248G>A	p.(Asp750Asn)	0.01	0.00	-	20.80	0.00	0.997	Disease causing (1.0)	NA
W18-1160	Chr11: 76873225A>G	MYO7A	NM_000260.3	c.1403A>G	p.(His468Arg)	0.10	0.01	0.02	19.61	0.01	0.993	Disease causing (1.0)	UV3

For none of the variants is an effect on transcript splicing predicted. Scores that meet the thresholds for pathogenicity as described in the methods section are indicated in red. Thresholds for pathogenicity: CADD-PHRED (≥ 15), SIFT (≤ 0.05), PolyPhen-2 (≥ 0.450) and MutationTaster (deleterious). Genome, Genomic positions according to GRCh37/hg19; In-house AF, allele frequency (%) in in-house database ($\sim 7,500$ exomes); GnomAD_E AF and GnomAD_G AF, allele frequencies (%) in respectively gnomAD exome or genome databases; CADD_PHRED, Combined Annotation Dependent Depletion PHRED score; SIFT, Scale-Invariant Feature Transform; PPH2, Poly-Phen-2 score; MutationTaster (prob), MutationTaster score with probability (0-1); ClinVar, American College of Medical Genetics and Genomics (ACMG) classification of variants as in ClinVar; UV1, benign; UV2, likely benign; UV3, variant with unknown significance; NA, not available.

Table S3. Individual results of otoscopic examination, audiometry, imaging and progression of HL

Family	Subject	Age of onset (y)	Otosopic examination	Clinical remarks	CT	MRI	Audiometry				Progression of HL					
							PTA		SRT		Maximum SRS (%)		Progression rate (dB/y)	YOY (y)	General remarks	
							R	L	R	L	R	L				R
W97-056	III:2	48	NT				81	72	63	70	72	87	65	0.6	53-82	
	III:4	43	NT	Ab, NE, T			72	93	90	NT	NT	NT	NT	NA	63-72	
	III:5	70	NT	NE			79	63	52	NT	NT	NT	NT	NA	0	Professional noise exposure
	III:10	21	NT	NE			41	47	48	61	70	95	95	0.5	21-56	
	III:11	26	NT	O, T			54	43	38	NT	NT	NT	NT	NA	0	
	III:12	20	NT	T			41	40	43	43	42	90	95	NA	32-40	
	III:15	37	N	NE, T			67	50	48	57	51	92	75	1.0	46-67	
	III:18	36	NT	T			62	50	51	50	NT	95	90	0.9	41-62	
	III:19	14	N	A			42	NA	73	NA	NA	NA	41	1.4	21-42	R ear profoundly deaf
	III:22	20	NT	T			62	75	80	70	70	75	88	0.8	42-62	
	III:24	33	N				43	62	60	63	67	93	93	1.8	36-48	
	III:26	34	N	O, T			45	50	52	NT	NT	NT	NT	NA	0	
	III:28	NOHL	NT				52	7	10	NT	NT	NT	NT	NA	47-51	
	III:30	NR	N	A			64	53	35	42	25	100	100	1.2	48-64	Otosclerosis, infrequent balance complaints
	IV:20	7	N	T			50	70	63	62	52	88	90	0.9	7-51	
	IV:22	NR	N	T			27	20	25	NT	NT	NT	NT	NA	0	
	IV:23	NR	NT				36	7	8	NT	NT	NT	NT	0.5	25-35	
	IV:25	24	NT				42	25	18	18	22	NT	NT	0.4	22-42	

Table S3. Continued

Family	Subject	Age of onset (y)	Otoscopic examination	Clinical remarks	CT	MRI	Subject age (y)	Audiometry						Progression of HL		
								PTA		SRT		Maximum SRS (%)		Progression rate (dB/y)	YOF (y)	General remarks
								R	L	R	L	R	L			
	IV:26	NOHL	NT				41	7	7	NT	NT	NT	NT	NA	NA	
	IV:27	35	N				47	17	18	12	18	100	100	0.3	26-47	
	IV:28	30	N	A, NE, T	N	N	45	32	52	37	55	100	92	1.8	25-46	Professional noise exposure
	IV:29	17	N	NE			26	5	10	NT	NT	NT	NT	NA	22-26	
	IV:35	30	N	NE, T			37	15	15	10	10	100	100	NA	0	Recreational noise exposure
	IV:38	8	N	T			31	8	7	8	7	100	100	NA	NA	
	V:1	5	N	NE, T			25	40	37	28	27	100	97	1.3	8-25	
	V:2	NOHL	N				23	3	5	<10	<10	100	100	NA	NA	
W02-016	II:3	29	NT	T			82	68	68	65	70	100	93	0.9	82-92	
	II:6	47	N				89	80	83	80	80	52	67	2	76-86	
	II:9	40	N				73	NA	NA	67	72	75	55	NA	69-73	R ear profoundly deaf above 2 kHz
	III:1	41	N	A, V			73	40	67	35	77	88	62	NA	66-73	Infrequent vertigo attacks since the age of 65 years
	III:5	NR	N				65	20	27	18	17	95	96	NA	0	Balance complaints after CI surgery
	III:9	35	N	O, A, T, V	N		39	48	42	47	38	80	100	3	24-47	
	III:10	32	N	T			32	15	20	10	15	100	95	1.4	32-60	
	IV:1	53	NT	V			42	12	10	NT	NT	NT	NT	NA	0	Benign paroxymal positional vertigo

Table S3. Continued

Family	Subject	Age of onset (y)	Otoscopic examination	Clinical remarks	CT	MRI	Imaging				Audiometry				Progression of HL			
							Subject age (y)		PTA		SRT		Maximum SRS (%)		Progression rate (dB/y)	YOF (y)	General remarks	
							R	L	R	L	R	L	R	L				
	IV:3	17	N					27	25	25	25	17	100	100	1	34-47		
	IV:8	27	N					38	37	32	30	30	100	95	2.7	17-30		
W04-262	II:11	<18	R atelectasis L sclerotic	A, T				85	87	>95	>95	37	44	NA	0	0	Multiple ear surgeries, a.o. ear drum surgery	
	III:4	49	N	T				40	47	28	35	100	90	NA	0			
	III:5	55	N	NE, T				32	23	NT	NT	NT	NT	NT	0.6	49-63		
	III:8	49	N	A				43	32	37	22	100	100	1.4	46-59			
	III:11	32	N	A, T		N		62	55	60	70	95	70	2	47-60			
	III:14	NOHL	N	Ab				7	7	NT	NT	NT	NT	NA	NA			
	III:16	25	N	A, T				70	65	80	77	55	60	0.9	29-47			
	III:19	28	N	A		N		83	92	NA	NA	42	42	2.1	33-49			
	IV:3	NR	N					12	7	NT	NT	NT	NT	NA	0			
W15-0495	II:2	35	NT	T				73	75	65	70	60	70	0.9	74-84			
	II:4	PS	N					77	85	105	105	60	50	2.4	58-81			
	II:7	60	NT	A				60	73	55	72	73	55	1.5	70-80			
	III:3	38	N	T, A				22	15	19	14	100	97	NA	51-55			
	III:4	NOHL	NT	NE				8	7	NT	NT	NT	NT	NA	0		Noise trauma	
	III:7	50	NT					22	20	NT	NT	NT	NT	NA	0			
	III:8	30	N	A, T				62	58	52	48	92	92	NA	44-51			
	III:9	43	N	NE				38	37	42	44	96	90	NA	0		Professional noise exposure	

Table S3. Continued

Family	Subject	Age of onset (y)	Otoscopy examination	Clinical remarks	CT	MRI	Imaging				Audiometry				Progression of HL			
							PTA		SRT		Maximum SRS (%)		Progression rate					
							R	L	R	L	R	L	R	L	(dB/y)	YOF (y)	General remarks	
W08-1421	III:11	33	N	NE, T				38	37	40	40	100	95	NA	36-38			
	III:2	cong.	N	T			67	42	43	30	40	87	65	1.8	50-67			
	IV:2	PS	N	A, Ns, O, T			34	27	28	27	32	95	93	NA	6			
W18-0470	V:1	0	R eardrum perforation	A, Ab, Ns, O, T			9	73	67	58	60	NT	NT	NA	NA	Neonatal intensive care, surgeries, ototoxic antibiotics		
	II:2	6	N				66	42	38	27	27	100	95	NA	0			
	III:1	30	N	O			31	19	18	16	14	100	100	NA	0			
W18-1004	III:3	11	NT				23	23	22	15	13	100	100	NA	5			
	I:1	12	N				54	17	23	NT	NT	NT	NT	NA	0			
	II:1	4	N				20	38	35	30	27	100	95	NA	5			
W18-0471	I:1	20	N	A, T			47	72	117	64	NA	83	0	NA	0			
	I:1	27	N	T, V			31	35	37	18	22	100	100	NA	0	Migrainous vertigo		
	I:1	39	NT		N	N	48	35	42	NT	NT	NT	NT	NA	0			
W18-1005	I:1	25	N		N	N	26	42	45	32	37	100	97	NA	0			
	I:1	12	N	T			50	47	48	42	43	100	100	NA	7			

Age of onset (AoO), age of onset in years as reported by the subjects. Subject age, the age at which the audiometric data of column 9 to 14 were obtained, in general the last audiogram. If no speech audiometry was performed during the latest pure tone audiometry, the latest audiogram in which both were measured, was selected. Progression rate of HL, calculated as described in the methods section, if there was at least a follow-up duration of 10 years. Y, years; PTA, pure tone average, mean of 0.5, 1 and 2 kHz air conduction thresholds; R, right; L, left; SRT, speech reception threshold; SRS, speech recognition score in %; YOF, years of follow up; MOHL, unaffected subject; NR, age of onset of HL not reported; PS, subject reported onset of HL during primary school; NT, not tested; N, no abnormalities; T, Tinnitus; NE, extensive exposure to noise; O, recurrent otitis; A, asymmetric HL; Ab, subject reported long-term antibiotics usage, but no details about duration and which antibiotics; V, vestibular complaints; NA, not applicable.

Table S4. Results of vestibular testing

Family	Subject (age)	Click-evoked ABR	Remarks and history of vestibular complaints	Oculo-motor testing	vHIT (gain)	Rotating chair											
						SPV Caloric irrigation					Rotating chair						
						Warm (°/s) (10-52) ^a		Cold (°/s) (7-31) ^a		Conclusion	MC	MC	MC	MC	MC	MC	MC
W97-056	III:24 (48)	NT	no	normal	NT	NT	12	13	normal	NA	NA	49	50	15	15	normal	
	III:11 (53)	NT	no	normal	NT	31	18	16	9	normal	NA	NA	28	30	14	14	Underestimated ^b
W02-016	III:1 (71)	N	Infrequent vertigo attacks since the age of 65 years	normal	NT	NT	3	11	hyporeactive	NA	NA	51	33	10	11	Hyporeactive	
	III:9 (63) ^c	N	Balance complaints after CI surgery	normal	normal	NT	19	24	normal	70	61	63	56	15	19	normal	
W04-262	III:10 (60)	N	no	normal	normal	32	39	NT	NT	normal	46	57	42	52	12	12	normal
	III:16 (60)	NT	no	normal	normal	31	30	37	46	normal	NA	NA	42	65	22	15	normal
	III:19 (48)	NT	no	normal	NT	9	7	9	8	normal	65	78	59	71	21	15	normal
W18-1421	IV:2 (47)	NT	no	normal	normal	23	19	12	19	normal	75	80	68	72	11	11	normal
W18-0470	III:1 (32)	N	no	normal	normal	12	26	12	12	normal	82	60	75	55	10	12	normal

Auditory brainstem response; vHIT, video head impulse test; °/s, degrees per seconds; SPV, slow phase velocity; Tau, time constant; R, right ear; L, left ear; CW, clockwise; CCW, counterclockwise; N, no abnormalities; NT, not tested; NA, not applicable. ^a, normative values at our institute ^b, nystagmus was suppressed due to stress ^c, Subject was tested after CI surgery and also had c- and oVEMP testing, no abnormalities were objectified (data not shown).

Table S5. Individual results of age of onset, otoscopy, audiometry and progression of HL of the phenocopies

Family	Subject	Age of onset (y)	Otoscopy examination	Clinical remarks	CT	MRI	Audiometry				Progression of HL				
							PTA		SRT		Maximum SRS (%)		Progression rate (dB/y)	YOF (y)	General remarks
							R	L	R	L	R	L			
W97-056	III:14	52	N	V			42	40	48	42	85	92	0.8	39-70	Ménière-like phenotype
	III:20	46	N	Ab, T			45	40	37	37	100	100	0.7	50-71	Smoking, COPD Gold III, often antibiotics
	III:21	NR	NT				38	37	20	25	NT	NT	0.8	45-68	
W04-262	III:10	55	N	T, V			33	38	33	41	100	95	NA	0	

Family members with HL that do not carry the RIPOR2 variant are considered a phenocopy. Age of onset (AoO) is the age of onset in years as reported by the subjects. Subject age is the age at which the audiometric data of column 9 to 14 were obtained. If no speech audiometry was performed during the latest audiometric testing, the penultimate audiogram was selected. Progression rate of HL was calculated if there was at least a follow-up duration of 10 years. Y, years; PTA, pure tone average, mean of 0.5, 1 and 2 kHz air conduction thresholds; R, right; L, left; SRT, speech reception threshold; SRS, speech recognition score in %; YOF, years of follow up; NR, age of onset of HL not reported; NT, not tested; N, no abnormalities; V, vestibular complaints; Ab, subject reported long-term antibiotics usage, but no details about duration and which antibiotics; T, Tinnitus; NA, not applicable.

Table S6. Genes analyzed by MIP sequencing

<i>ACTG1</i>	<i>GPSM2</i>	<i>PNPT1</i>
<i>ADCY1</i>	<i>GRHL2</i>	<i>POU3F4</i>
<i>BDP1</i>	<i>GRM7</i>	<i>POU4F3</i>
<i>BSND</i>	<i>GRM8</i>	<i>PRPS1</i>
<i>CABP2</i>	<i>GRXCR1</i>	<i>PTPRQ</i>
<i>CCDC50</i>	<i>GRXCR2</i>	<i>RDX</i>
<i>CDH23</i>	<i>HGF</i>	<i>RIPOR2</i>
<i>CEACAM16</i>	<i>ILDR1</i>	<i>SERPINB6</i>
<i>CIB2</i>	<i>KARS</i>	<i>SIX1</i>
<i>CLDN14</i>	<i>KCNQ4</i>	<i>SLC17A8</i>
<i>CLIC5</i>	<i>LHFPL5</i>	<i>SLC26A4</i>
<i>COCH</i>	<i>LOXHD1</i>	<i>SLC26A5</i>
<i>COL11A2</i>	<i>LRTOMT</i>	<i>SMPX</i>
<i>COL4A6</i>	<i>MARVELD2</i>	<i>STRC</i>
<i>CRYM</i>	<i>MIR96</i>	<i>SYNE4</i>
<i>DCDC2</i>	<i>MSRB3</i>	<i>TBC1D24</i>
<i>GSDME</i>	<i>MYH14</i>	<i>TECTA</i>
<i>DFNB31</i>	<i>MYH9</i>	<i>TJP2</i>
<i>DFNB59</i>	<i>MYO15A</i>	<i>TMC1</i>
<i>DIABLO</i>	<i>MYO3A</i>	<i>TMEM132E</i>
<i>DIAPH1</i>	<i>MYO6</i>	<i>TMIE</i>
<i>ELMOD3</i>	<i>MYO7A</i>	<i>TMPRSS3</i>
<i>EPS8</i>	<i>NAT2</i>	<i>TNC</i>
<i>ESPN</i>	<i>OSBPL2</i>	<i>TPRN</i>
<i>ESRRB</i>	<i>OTOA</i>	<i>TRIOBP</i>
<i>EYA4</i>	<i>OTOF</i>	<i>TSPEAR</i>
<i>GIPC3</i>	<i>OTOG</i>	<i>USH1C</i>
<i>GJB2</i>	<i>OTOGL</i>	<i>USH1G</i>
<i>GJB3</i>	<i>P2RX2</i>	<i>WFS1</i>
<i>GJB6</i>	<i>PCDH15</i>	

Table S7. Primer sequences

Target	Primer	Oligonucleotides (5'-3')
<i>RIPOR2</i> exon 14, wt allele	Forward	aagcagctggtaagagg
	Reverse	gcagccttcagattctcc
<i>RIPOR2</i> exon 14, mut allele	Forward	ggaaggaaacatcacaagag
	Reverse	gcagccttcagattctcc
<i>RIPOR2</i> exons 3-4, mRNA	Forward	ggccttgaaaaatggacttg
	Reverse	ccaggcgagagtttctttc
<i>GUSB</i> exons 2-3, mRNA	Forward	agagtggctgctgaggattgg
	Reverse	ccctcatgctctagcgtgc

Primer sequences for *RIPOR2* are based on reference sequence NM_00147722.3 and for *GUSB* on NM_00181.3. wt, wildtype; mut, mutant.

SUPPLEMENTARY FIGURES

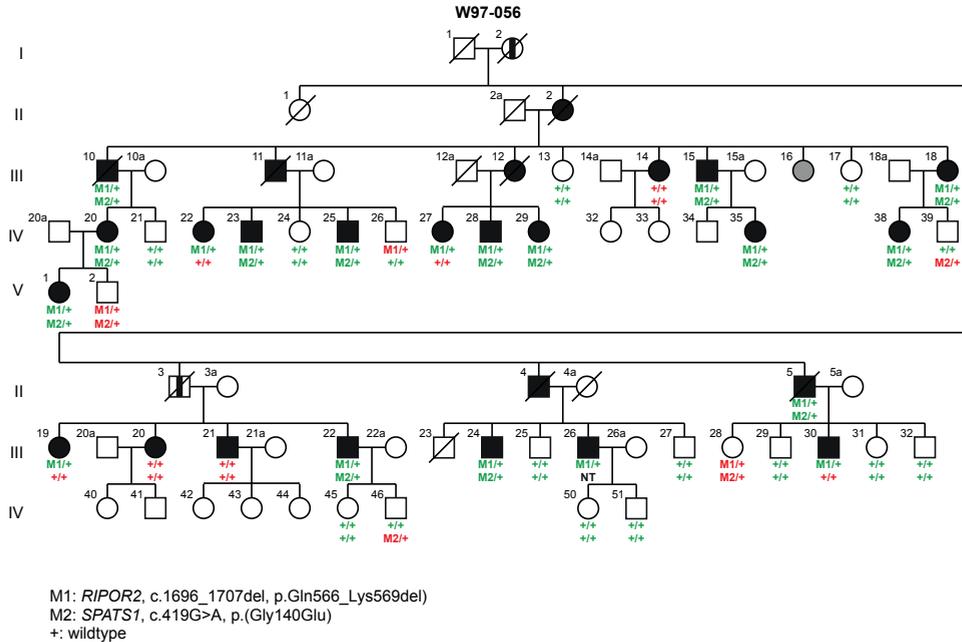


Figure S1. Segregation analysis of the *SPATS1* and *RIPOR2* variants in W97-056. Subjects determined to be affected by heteroanamnesis are indicated with a vertical black line. The subject marked in grey is diagnosed with intellectual disability and excluded from further participation in this study. Subject identifiers correspond to those in de Brouwer et al., 2005.¹⁷ Genotypes in green correspond to a co-occurrence of the variant and hearing impairment and those in red to lack of co-occurrence. NT, not tested.

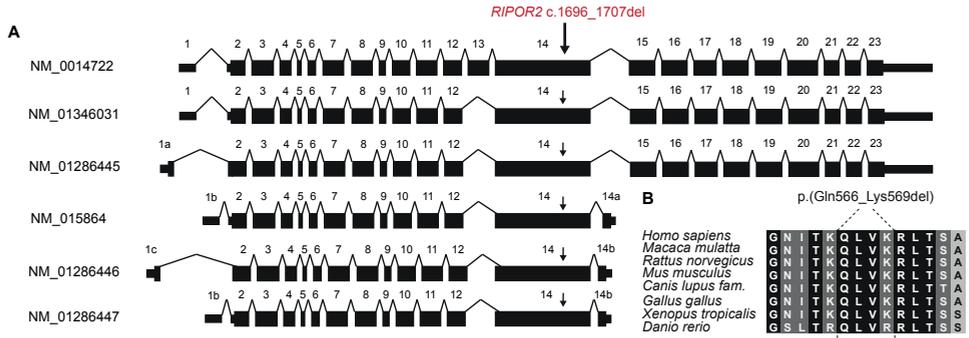


Figure S2. Schematic overview of major *RIPOR2* transcripts. (A) Transcripts are extracted from the Ensembl genome browser. The position of the c.1696_1707del *RIPOR2* variant is indicated with an arrow. **(B)** Evolutionary conservation of the amino acids that are affected by the variant. Fully conserved amino acid residues are shown on a black background. A dark grey background marks chemical similarity of residues and a light grey background indicates chemical dissimilarity.

A

Position (bp)	Marker	W97-056: IV:25	W02-016: IV:3	W04-262: III:9	W08-1421: V:1	W15-0495: III:9	W18-0470: III:3	W18-1004: I:1	W18-0471	W18-0472	W18-0473	W18-1005
24,185,805	D6S276	5	9	10	8	5	8	5	10	5	8	5
24,306,692	D6S2439	1	6	6	3	1	7	1	6	2	3	3
24,843,752	D6S1554	1	1	1	1	1	1	1	1	1	1	1
24,964,249	D6S1571	1	1	1	1	1	1	1	1	1	1	1
24,983,204	D6S1545	4	4	4	4	4	3	4	4	4	4	4
25,296,948	D6S1281	2	2	2	1	2	2	2	2	2	1	2
25,495,470	D6S1621	1	1	1	1	1	1	1	1	1	3	2

B

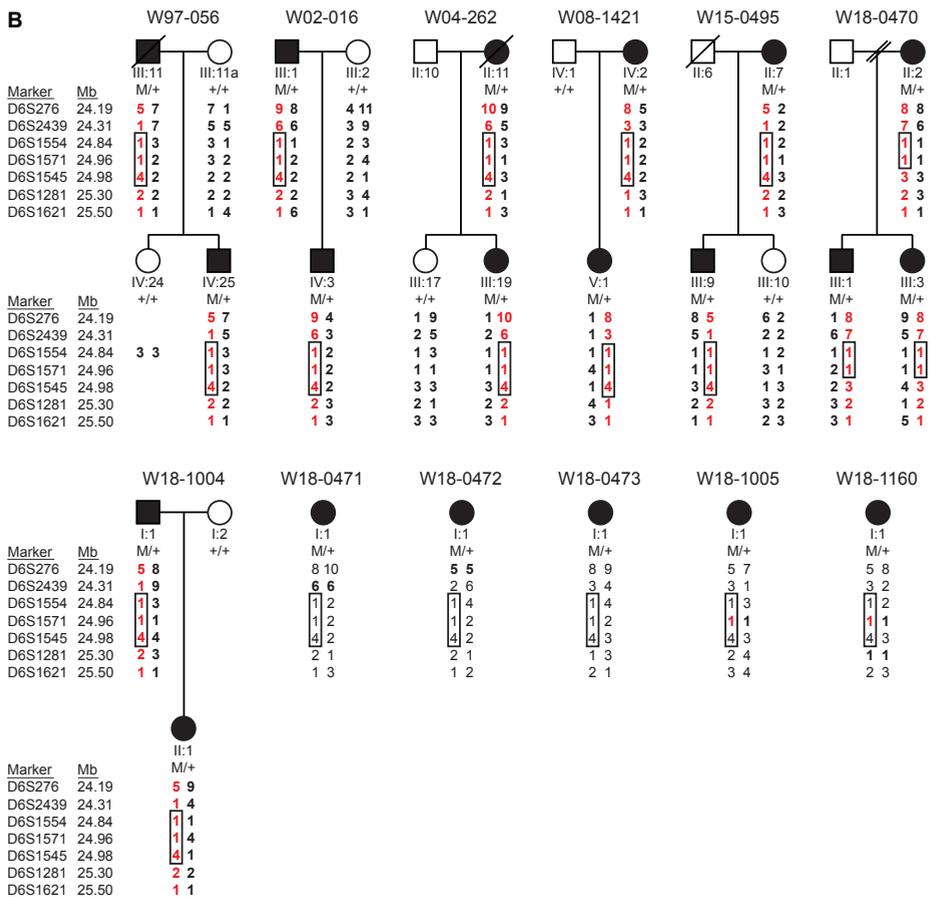


Figure S3. Family pedigrees with genotypes and haplotypes of VNTR markers. (A) The shared haplotype is marked in grey. For marker D6S1545 (light-grey), a different CA-repeat length was determined in one family, but the marker is considered to be potentially part of the shared haplotype as a change of repeat length cannot be excluded. Markers for which the phase of the alleles could conclusively be determined via segregation in the family are marked in bold. The *RIPOR2* c.1696_1707del variant is located between the markers D6S2439 and D6S1554. Genomic positions (bp) are according to the UCSC Genome Browser (GRCh37/hg19). (B) The haplotypes carrying the *RIPOR2* c.1696_1707del variant are shown in red. A haplotype of 1.1 Mb was found to be shared (boxed) and is delimited by the markers D6S1554 and D6S1545. Alleles for which the parent of origin could conclusively be determined are marked in bold.

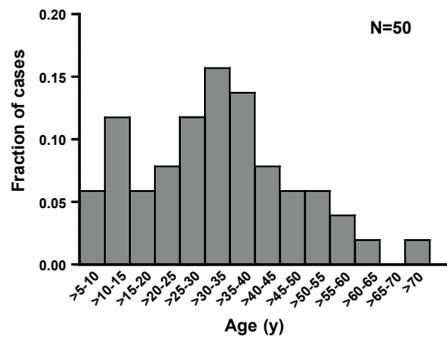
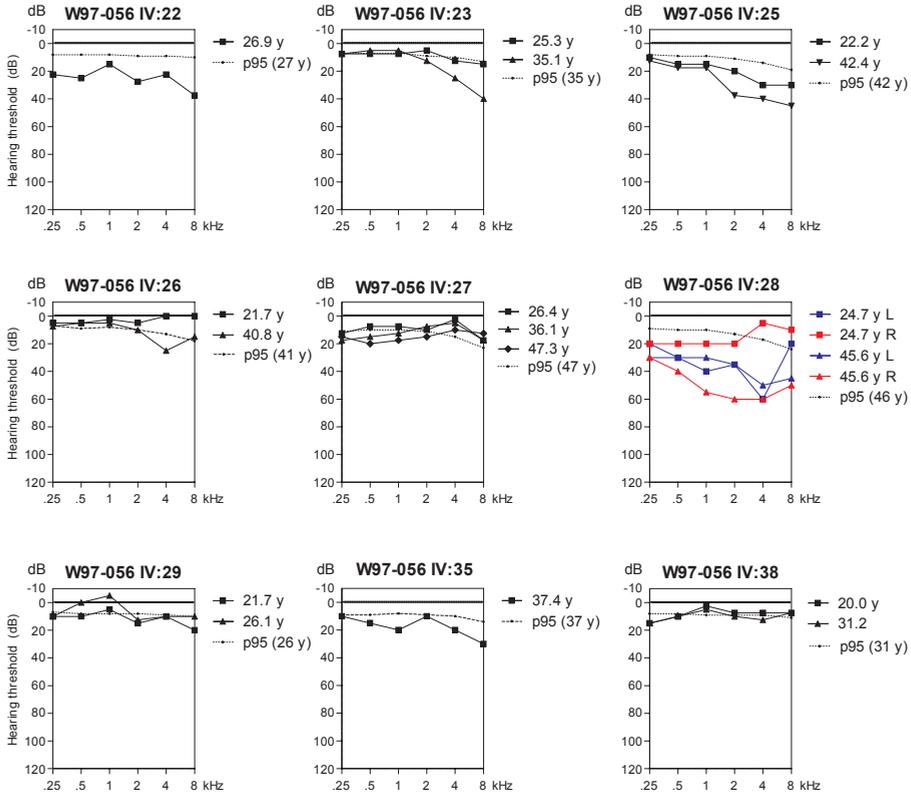
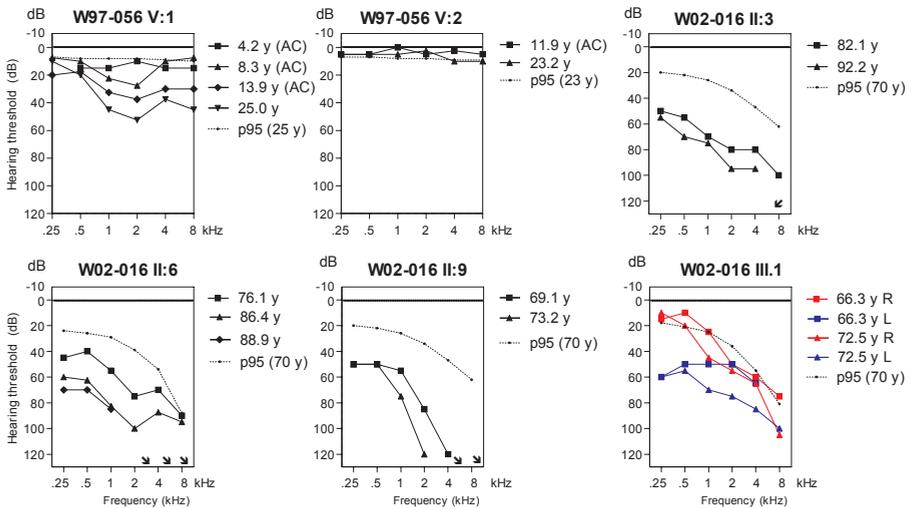


Figure S4. Reported age of onset of hearing loss of subjects with the *RIPOR2* c.1696_1707del variant. Distribution of the reported ages of onset of *RIPOR2*-associated hearing loss per 5 years for 52 subjects who reported a specific age of onset, y, years.

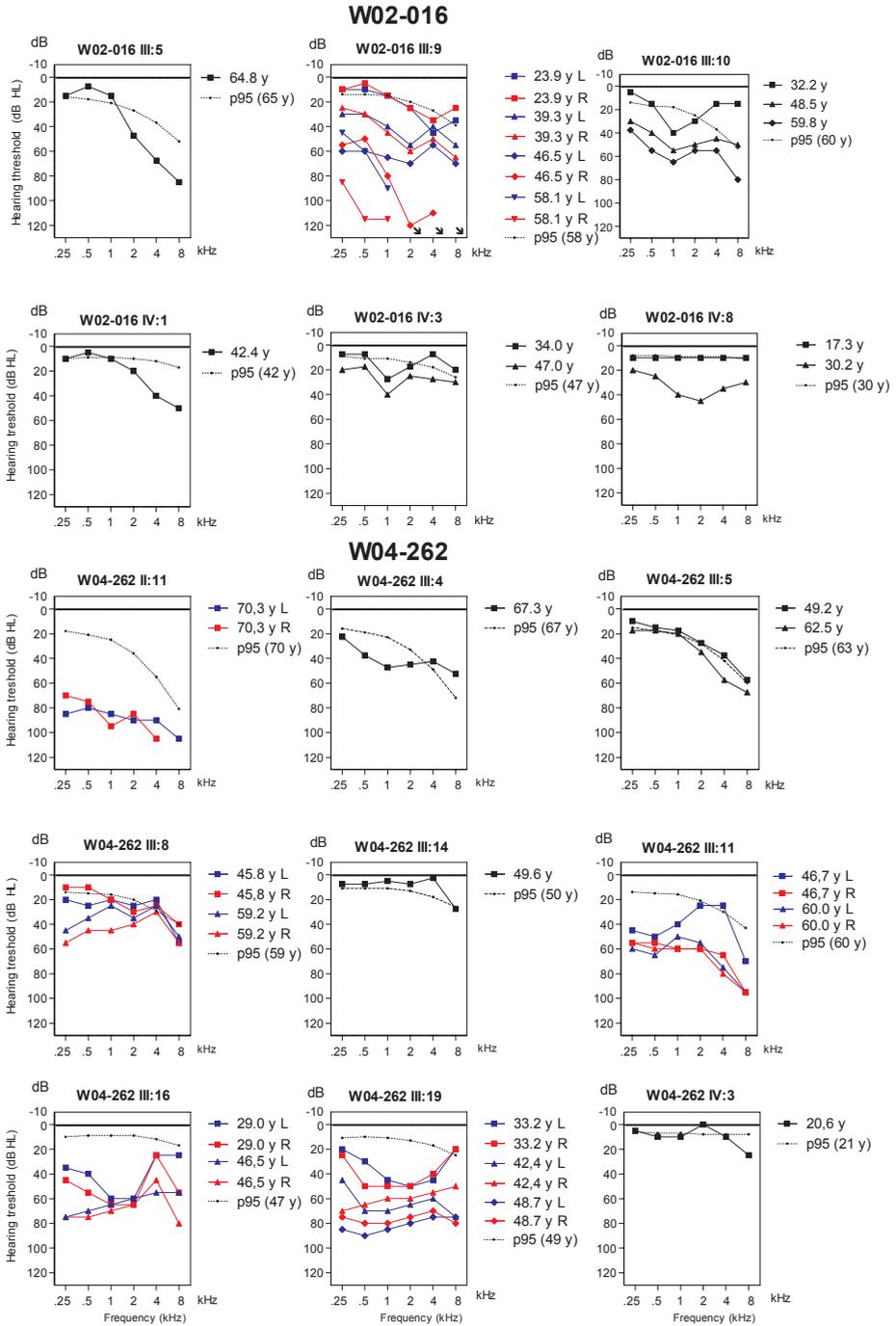
W97-056

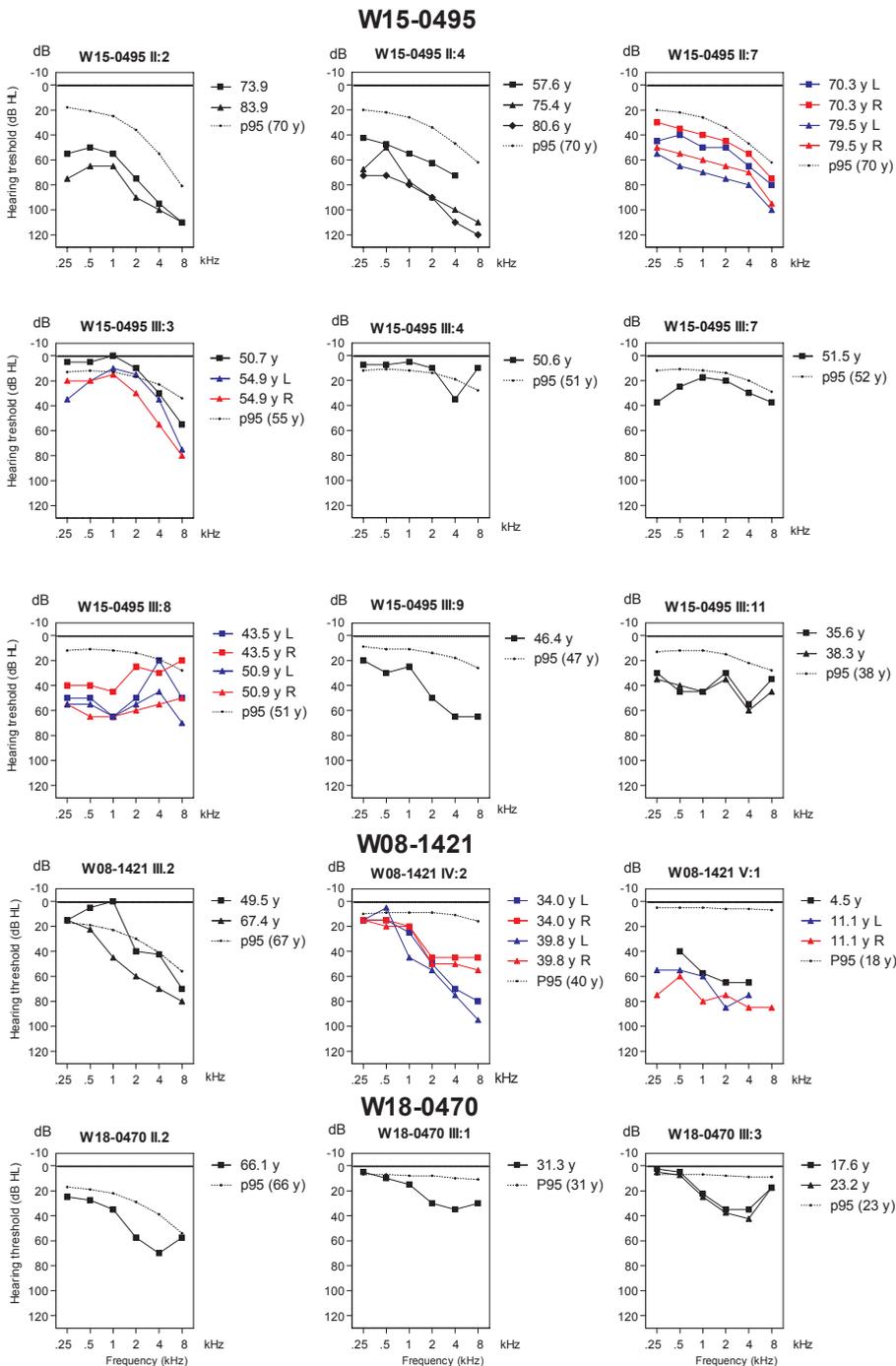


W97-056, W02-016



A RIPOR2 deletion is a frequent cause of adult-onset hearing loss





W18-1004, single cases

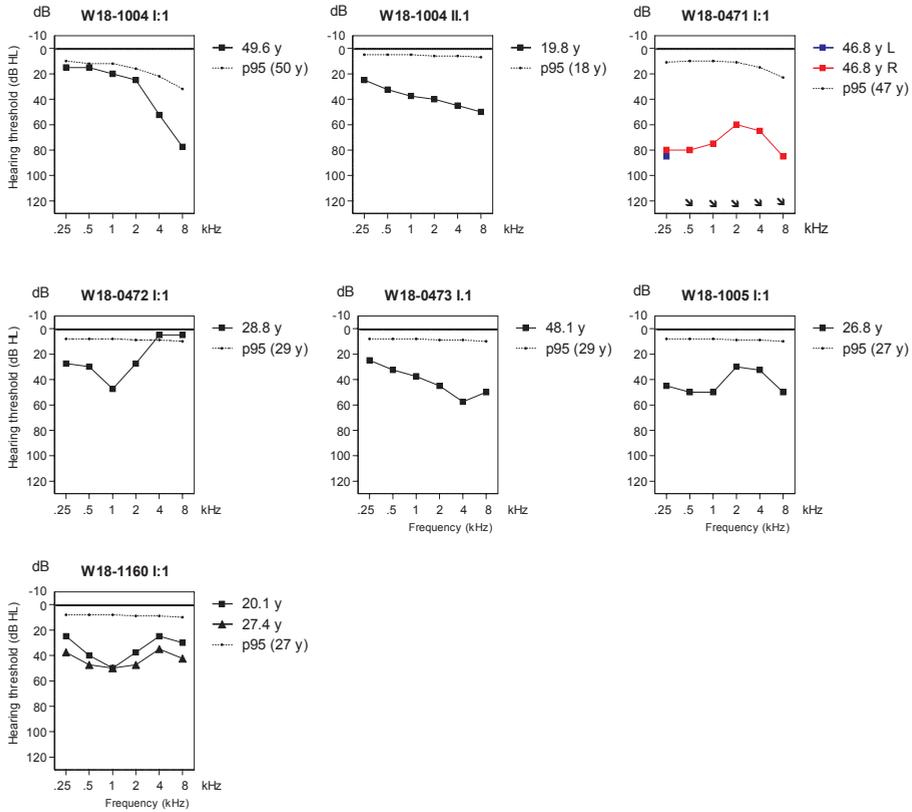


Figure S5. Pure tone audiometry. Air conduction thresholds of all subjects (except III.5 of W08-1421) with the c.1696_1707del *RIPOR2* variant are depicted. In case of symmetry, the averages of left and right ear thresholds are shown. Otherwise, colorized (right red, left blue) audiograms of both ears are depicted. The p95 values are matched to the individuals' sex and age at the most recent audiometry, according to the ISO 7029:2017 standard. The age range for which the ISO 7029:2017 can be applied is 18 to 70 years. Part of the pure tone audiometry for family W97-056 has been published previously by Kunst et al. 2000.¹⁸ y, age in years; R, right; L, left; dB HL, decibel hearing level; kHz, kiloHertz; AC; only air conduction levels available, no additional bone conduction thresholds have been measured.

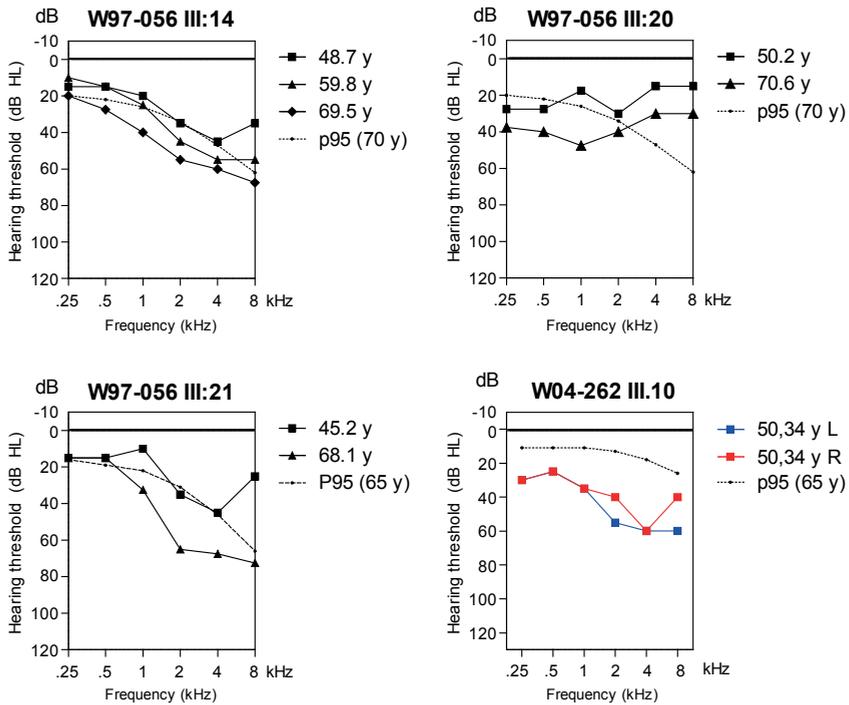


Figure S6. Audiograms of hearing impaired individuals without the *RIPOR2* c.1696_1707del variant. Air conduction thresholds are depicted of all hearing impaired subjects who did not have the c1696_1707del *RIPOR2* variant. In **Table S5** information about possible explanations for their hearing impairment is provided. In case of symmetry, the averages of left and right ear thresholds are shown. Otherwise, colorized (right red, left blue) audiograms of both ears are depicted. The p95 values are matched to the individuals' sex and age at the most recent audiometry, according to the ISO 7029:2017 standard. The age range for which the ISO 7029:2017 can be applied is 18 to 70 years. y, age in years; R, right; L, left; dB HL, decibel hearing level; kHz, kiloHertz.

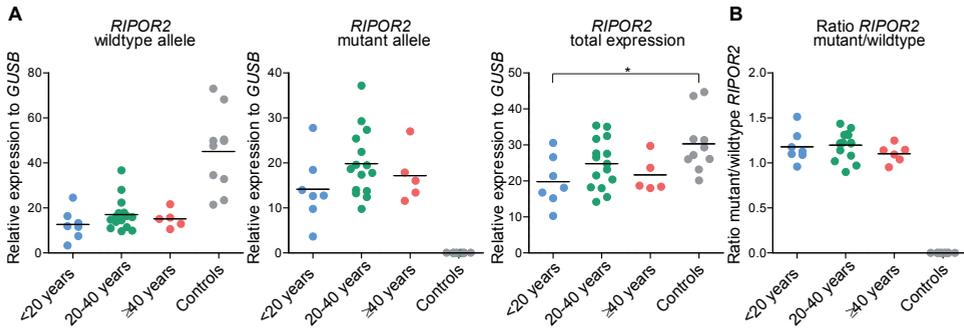


Figure S7. Transcript levels of *RIPOR2* alleles determined by RT-qPCR. (A) Subjects were divided in three groups based on the reported ages of onset: early onset (<20 years, n=7), middle onset, (20-39 years, n=15) and late onset (≥40, n=6) hearing impairment. RNA samples isolated from peripheral blood of individuals without the *RIPOR2* variant were used as controls (n=10). **(B)** Calculated ratio of *RIPOR2* mutant to wildtype relative expression analysis. A one-way ANOVA followed by Tukey's multiple comparison test was employed to identify potentially significant differences between the transcript levels of the groups. * p-value = 0.0214.

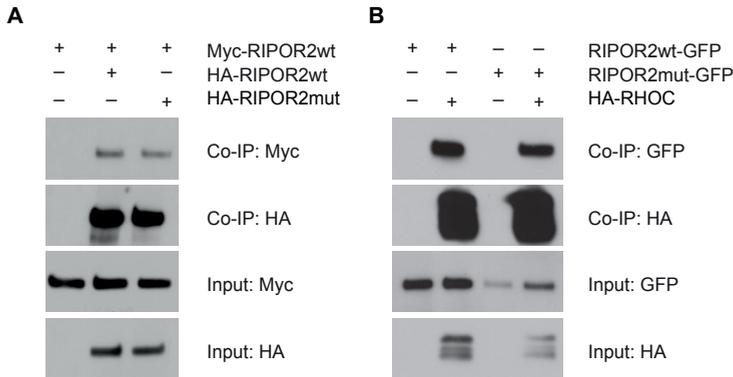


Figure S8. *RIPOR2* dimerization and interaction with RHOC. (A) Interaction of murine *RIPOR2*-wildtype (*RIPOR2*wt) and -mutant (*RIPOR2*mut) was studied using CoIP assays. HEK293T cells were transfected with constructs encoding N-terminally tagged proteins as indicated above each panel. Immunoprecipitations were performed using anti-HA antibodies, followed by western blotting. **(B)** Interaction with RHOC was studied using C-terminally GFP-tagged *RIPOR2*-wildtype or -mutant and N-terminally HA-tagged RHOC. Immunoprecipitations were performed using anti-HA antibodies, followed by western blotting.

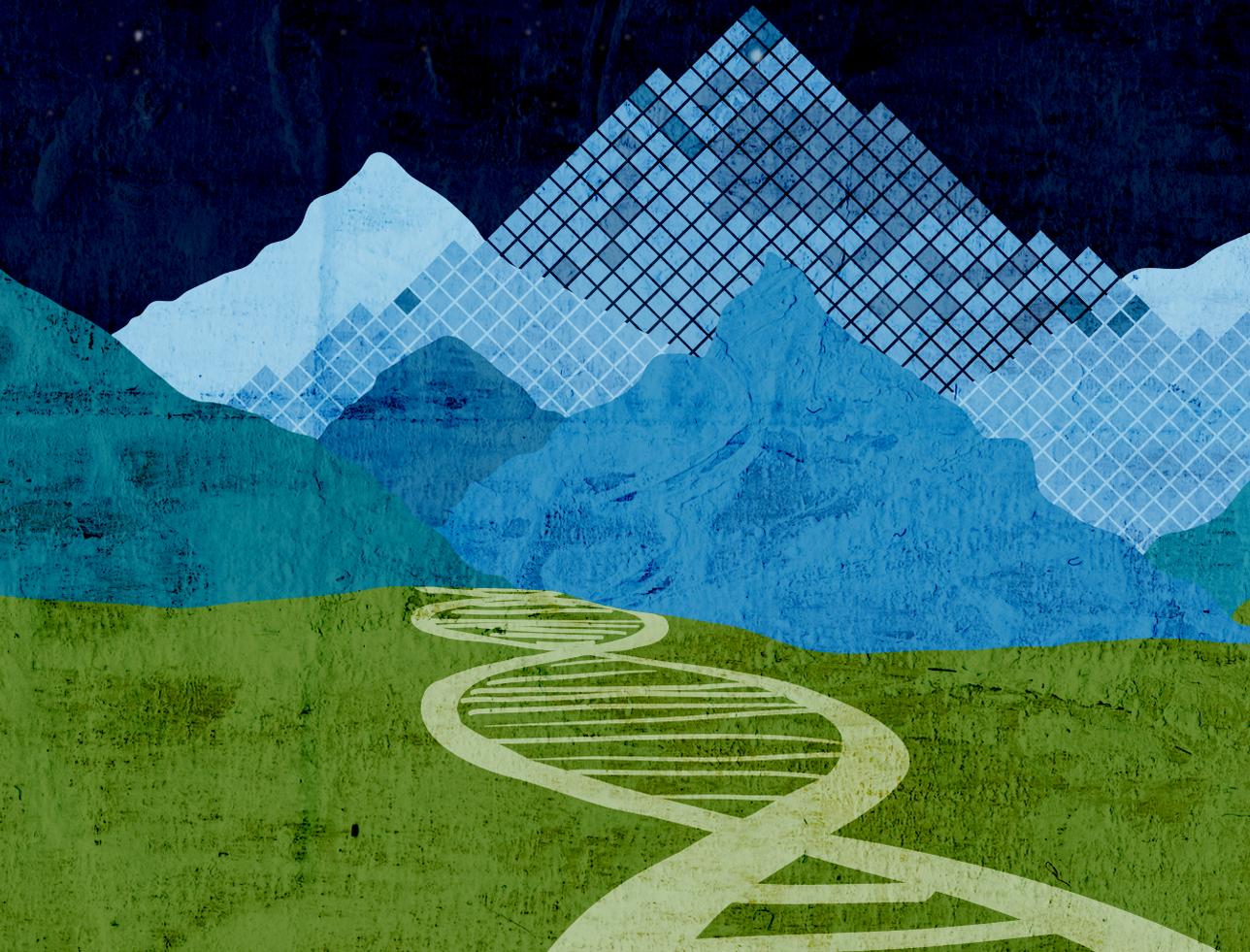
SUPPLEMENTARY REFERENCES

1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
2. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
3. Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P. *et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Research* **22**, 1525-1532 (2012).
4. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).
5. Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423-425 (2012).
6. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24 (2011).
7. Hardenbol, P., Banér, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnology* **21**, 673 (2003).
8. Khandelwal, K.D., Ishorst, N., Zhou, H., Ludwig, K.U., Venselaar, H., Gilissen, C. *et al.* Novel IRF6 mutations detected in orofacial cleft patients by targeted massively parallel sequencing. *Journal of Dental Research* **96**, 179-185 (2016).
9. Tarpey, T. Linear transformations and the k-means clustering algorithm: Applications to clustering curves. *Journal of the American Statistical Association* **61**, 34-40 (2007).
10. Ketchen, D.J. & Shook, C.L. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal* **17**, 441-458 (1996).
11. Oonk, A.M., Beynon, A.J., Peters, T.A., Kunst, H.P., Admiraal, R.J., Kremer, H. *et al.* Vestibular function and temporal bone imaging in DFNB1. *Hearing Research* **327**, 227-34 (2015).
12. Vanspauwen, R., Wuyts, F.L., Krijger, S. & Maes, L.K. Comparison of different electrode configurations for the oVEMP with bone-conducted vibration. *Ear and Hearing* **38**, 205-211 (2017).
13. Papatthanasidou, E.S., Murofushi, T., Akin, F.W. & Colebatch, J.G. International guidelines for the clinical application of cervical vestibular evoked myogenic potentials: an expert consensus report. *Clin Neurophysiol* **125**, 658-666 (2014).
14. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**, e45-e45 (2001).
15. Diaz-Horta, O., Abad, C., Cengiz, F.B., Bademci, G., Blackwelder, P., Walz, K. *et al.* Ripor2 is involved in auditory hair cell stereociliary bundle structure and orientation. *Journal of Molecular Medicine* **96**, 1227-1238 (2018).

16. Diaz-Horta, O., Subasioglu-Uzak, A., Grati, M.h., DeSmidt, A., Foster, J., Cao, L. *et al.* FAM65B is a membrane-associated protein of hair cell stereocilia required for hearing. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9864-9868 (2014).
17. De Brouwer, A.P.M., Kunst, H.P.M., Krebsova, A., van Asseldonk, K., Reis, A., Snoeckx, R.L. *et al.* Fine mapping of autosomal dominant nonsyndromic hearing impairment DFNA21 to chromosome 6p24.1-22.3. *American Journal of Medical Genetics Part A* **137A**, 41-46 (2005).
18. Kunst, H., Marres, H., Huygen, P., Van Duijnhoven, G., Krebsova, A., Van der Velde, S. *et al.* Non-syndromic autosomal dominant progressive non-specific mid-frequency sensorineural hearing impairment with childhood to late adolescence onset (DFNA21). *Clinical Otolaryngology & Allied Sciences* **25**, 45-54 (2000).

CHAPTER 3

DFNA21



Chapter 3.2

The development of a genetic therapy for DFNA21 using allele- specific antisense oligonucleotides

Suzanne E. de Bruijn, Jill R.P.M. van Wolferen, Frans P.M. Cremers, Susanne Roosing, Ronald J.E. Pennings, Erwin van Wijk, Hannie Kremer, Erik de Vrieze

Manuscript in preparation

ABSTRACT

DFNA21 is a type of dominantly inherited adult-onset hearing loss and caused by a 12-nucleotide deletion in the *RIPOR2* gene. Previously, the c.1696_1707del *RIPOR2* variant was reported to be the most frequent cause of inherited adult-onset HL in Northwest Europe. There are strong indications that the *RIPOR2* deletion acts via a dominant, non-haploinsufficiency disease mechanism. Mutant *RIPOR2* is aberrantly localized in the stereocilia of murine auditory hair cells, suggesting a toxic gain-of-function effect. Both humans and mice carrying heterozygous loss-of-function alleles do not display hearing loss. This implies that inhibiting the synthesis of mutant *RIPOR2* protein, by selectively degrading the (pre)-mRNA transcribed from the mutant allele, can alleviate the negative consequences of mutant *RIPOR2* on auditory function. In this study, gapmer antisense oligonucleotides (AONs) were designed to specifically target mutant *RIPOR2* transcripts for degradation by the endogenous RNase H1 enzyme. The molecular efficacy of the AONs was validated in DFNA21 patient-derived fibroblasts and HEK293T cells. This revealed a lead AON molecule that was able to significantly reduce mutant *RIPOR2* transcript levels (up to ~90% in HEK293T cells), whilst leaving the level of wildtype *RIPOR2* mRNA intact. Additionally, western blot analyses showed that the decrease in mutant *RIPOR2* transcripts leads to a marked decrease in mutant protein synthesis. These studies provide proof-of-concept for the induction of rapid and specific degradation of mutant-*RIPOR2* transcripts by gapmer AONs. The identified lead molecule is a strong candidate for further preclinical development to ultimately establish a treatment for DFNA21.

INTRODUCTION

DFNA21, caused by an in-frame deletion in the *RIPOR2* gene, is a progressive form of dominantly inherited adult-onset hearing loss (HL). Recently, this deletion (c.1696_1707del, p.(Gln566_Lys569del), NM_014722.3) was identified as the most frequent cause of this type of HL within Europe. It is estimated that >30,000 individuals within Northwest Europe carry this variant, and therefore are at risk to develop DFNA21.¹ DFNA21 is highly variable; the age of onset ranges from congenital to the 7th decade of life, and HL can progress from mild in early stages of the disease, to severe or profound later in life.¹⁻³ Currently, there is no cure available to treat DFNA21.

RIPOR2 encodes the RHO Family Interacting Cell Polarisation Regulator 2 protein (RIPOR2). The protein is expressed in a wide variety of tissues and cell types, including the cochlea.⁴ In mouse cochlear hair cells, RIPOR2 is described to be localized at the basal taper region of stereocilia, where it is organized in a ring-like fashion. In the absence of RIPOR2, development, function and maintenance of murine hair cells is severely affected and morphological defects in both hair cells and stereociliary structures can be observed. *Ripor2* knockout mice are already deaf at 4 weeks of age, and *ripor2* knockdown in zebrafish leads to profound HL and loss of saccular hair cells.⁴⁻⁶

Pathogenic variants in *RIPOR2* have not only been associated with dominantly inherited, but with recessively inherited HL (DFNB104) as well.⁶ A loss-of-function variant in exon 3 of the gene is associated with profound, prelingual HL that corresponds with the phenotypes observed in the mouse knockout and zebrafish knockdown models.^{4,6} The c.1696_1707del variant that affects exon 14, on the other hand, can be considered a milder variant and is associated with a less severe, later onset, phenotype.¹ The mutated RIPOR2 was demonstrated to lack complete functionality due to aberrant localization in the stereocilia of wildtype mouse cochlear hair cells and the inability to rescue morphological defects of the hair bundle of RIPOR2-deficient hair cells. Although the pathogenic mechanism of the in-frame deletion is not yet completely understood, there are strong indications that the variant acts via a non-haploinsufficiency mechanism. The mislocalization of mutant RIPOR2 in the stereocilia suggests a toxic gain-of-function effect. In line with this, both humans and mice carrying heterozygous loss-of-function alleles do not display HL.^{4,6}

The non-haploinsufficiency disease mechanism implies that blocking the synthesis of mutant RIPOR2 proteins can alleviate their negative effects on auditory function. The lack of HL in heterozygous carriers of loss-of-function mutations, both in human and mice⁴⁻⁶, indicates that RIPOR2 synthesis from a single wildtype allele is sufficient for

normal inner ear function. Antisense oligonucleotides (AON)-mediated therapy can be considered a suitable strategy to modulate mutant *RIPOR2* expression. AONs with DNA-like properties can be specifically designed to bind transcripts harboring pathogenic variants, and will subsequently recruit RNase H1 endonuclease. This endonuclease degrades RNA molecules that are part of RNA:DNA duplexes, which will lead to a sequence-specific decrease in protein synthesis.^{7,8} The 5' and 3' wings of the RNase H1-dependent AONs can be chemically modified to increase thermodynamic stability and nuclease resistance, while maintaining a central gap region of DNA nucleotides that ensures RNase H1 activity.⁹ These modified AONs are referred to as gapmers, and have shown great therapeutic potential in treatment strategies for other inherited disorders including amyotrophic lateral sclerosis (ALS) and Huntington disease.¹⁰ More recently, gapmers were designed to specifically decrease mutant *COCH* transcript levels as a potential future treatment for HL type DFNA9.¹¹ To date, there are four AON-gapmers on the market that have been FDA- or EMA-approved and many more gapmers are under investigation in clinical trials.¹²

An AON-mediated therapy holds important advantages over other genetic therapy strategies such as CRISPR/Cas9-mediated DNA editing. The effect of the AONs is transient, and there is no risk for the introduction of non-reversible DNA changes (reviewed in (13,14)). The transient nature of AONs also significantly decreases the risk of severe adverse effects. Treatments can be halted when adverse effects occur, and off-target transcript degradation is only temporary, unlikely to induce biologically significant changes in protein levels.¹³ Additionally, previous studies achieved successful delivery of AONs to and uptake of AONs in relevant cochlear cells after intratympanic delivery.^{15,16} This method of AON delivery could successfully restore *Ush1c* pre-mRNA splicing in adult *Ush1c* knockin mice¹⁵, which confirms the feasibility of an AON approach as a potential treatment for (inherited) hearing disorders.

In the current study, we designed several (gapmer) AONs aimed to specifically target the transcripts of the c.1696_1707del *RIPOR2* allele for degradation. The designed molecules were evaluated in both patient-derived fibroblast cells and in transiently-transfected HEK293T cells. We successfully identified an AON with a strong potency to downregulate mutant *RIPOR2* transcript levels and thereby reduce the synthesis of mutant *RIPOR2* protein, whilst leaving the level of wildtype *RIPOR2* transcript intact. The identified lead molecule is a strong candidate for further preclinical development to ultimately establish a treatment for DFNA21.

MATERIALS AND METHODS

Study approval

The study of human subjects was approved by the medical ethics committee of the Radboudumc and performed in accordance with the principles of the World Medical Association Declaration of Helsinki. Written informed consent was obtained from all participants or their legal representatives.

Design of antisense oligonucleotides

AONs were designed following previously described criteria.^{11,17-19} Exon 14, harboring the 12-nucleotide (nt) target deletion, was examined for open configuration (mfold Web Server²⁰). Thermodynamic properties of potential 20-mer AON molecules (40-60% GC content) were assessed using the RNAstructure software as previously described.¹⁹ Uniqueness of AON target sequences was validated using BLAST (NCBI), allowing a maximum of two mismatches. AONs were purchased from Eurogentec and dissolved in PBS before use. Sequences and AON chemistry are provided in **Table S1**.

Cell culture conditions and AON delivery

Patient-derived primary fibroblast cells were cultured in standard fibroblast medium consisting of DMEM (Gibco) supplemented with 20% fetal calf serum, 1% sodium pyruvate and 1% penicillin-streptomycin. Prior to AON treatment, cells were seeded in 12 wells plates and cultured to a confluency of ~80%. Cells were transfected with AON molecules (final concentrations 50-250 nM in the culture medium) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions, following a 1:2 ratio (1 mg AON:2 ml Lipofectamine reagent). After 24 hours, cells were harvested for RNA isolation and subsequent transcript analysis.

For the assessment of AON efficacy in HEK293T cells, plasmids containing cDNA sequences encoding an N-terminally FLAG-tagged wildtype or mutant RIPOR2 (NM_015864.3) were generated with Gateway Technology (Life Technologies). RNA isolated from patient-derived EBV-transformed lymphoblastoid cells was used as input for Gateway-adapted RT-PCR. The sequence of both mutant and wildtype entry clones was verified with Sanger sequencing. HEK293T cells were co-transfected with AONs and the generated DNA constructs (500 ng per well) using 45 µl polyethyleneimine (PEI).²¹ Treated cells were collected 24 hours after AON delivery for transcript and protein analyses.

RNA isolation and RT-qPCR

Total RNA was isolated from treated cells using the Nucleospin RNA kit (Machery-Nagel) according to the manufacturer's instructions. Subsequently, cDNA was synthesized using the iScript cDNA synthesis kit (Bio-Rad) and 250 ng RNA as input material. Obtained cDNA was diluted twice (fibroblast cells) or five times (HEK293T cells) for quantitative PCR analysis.

QPCR analysis was performed using GoTaq qPCR Master Mix (Promega) according to standard procedures. Allele-specific primer sets were employed that allowed to distinguish the mutant and wildtype *RIPOR2* transcripts and included allele-specific forward primers (5'-GGAAGGAAACATCACAAAGAG-3', mutant allele; 5'-AAGCAGCTGGTCAAGAGG-3', wildtype allele), and a universal reverse primer (5'-GCAGCCTTCAGATTCTCC-3'). Relative expression levels as compared to the housekeeping gene *GUSB* (primers 5'-AGAGTGGTGCTGAGGATTGG-3' and 5'-CCCTCATGCTCTAGCGTGTC-3') were determined with the $2^{-\Delta\Delta Ct}$ method.

Western blot

For protein isolation, cells were treated with a lysis buffer containing 50 mM Tris pH 7.5, 150 mM NaCl and 0.5% (v/v) Triton X-100. Protein lysates were supplemented with protein sample loading buffer (LI-COR) and DTT (final concentration 2 mM) and denatured at 70°C for 10 minutes. Proteins were separated on a 4-12% NuPAGE Bis-Tris gel (ThermoFisher) in NuPAGE MOPS SDS running buffer (ThermoFisher) and transferred overnight at 4°C in NuPAGE transfer buffer supplemented with 20% methanol to a nitrocellulose membrane (0.45 mm, Sigma-Aldrich). Membranes were blocked with 5% Blotto, non-fat dry milk, blocking buffer (Santa Cruz), and incubated with primary antibodies (1 hour, RT) and secondary antibodies (45 minutes, RT), in 0.5% blocking buffer. In between incubation steps, the membranes were washed three times for 10 minutes with PBS supplemented with 0.2% Tween-20. After a final wash with PBS, proteins were visualized using the Odyssey Infrared Imaging System (LI-COR). RIPOR2 protein was quantified using the Fiji software (version 1.47), and normalized against tubulin. Antibodies: anti-RIPOR2 (1:1000; cat.#17015-1-AP, Proteintech), anti-tubulin (1:2000; cat.# ab7291, Abcam), goat anti-rabbit Alexa Fluor 680 (1:20000; cat.#A21076, Molecular Probes) and goat anti-mouse IRDye800 (1:20000; cat.# 926-32210, LI-COR).

RESULTS

Design of allele-specific RNase H1-dependent AONs

We first employed several *in silico* analyses in order to design RNase H1-dependent AONs complementary to transcripts of the c.1696_1707del mutant, but not wildtype, *RIPOR2* allele. First, internal hybridization of the (pre-)mRNA of mutant *RIPOR2* exon 14 was predicted *in silico* to determine accessibility of the target region (**Figure 1A**). In the putative AON target region, 17 of the 30 nucleotides are unpaired in the most probable 3D structure. Several candidate AONs were selected that span the 12nt-deletion region, which is a prerequisite for allele-discrimination. Next, thermodynamic properties were calculated, among which the free energy of on- and off-target AON binding, and the potential of the AONs for hairpin formation and dimerization. Based on these *in silico* predictions (**Table S1**), three AON molecules were ordered as phosphorothioate (PS)-linked DNA-bases for an initial screening. **Figure 1B** provides a schematic overview of the target regions of all AONs that are used in this study.

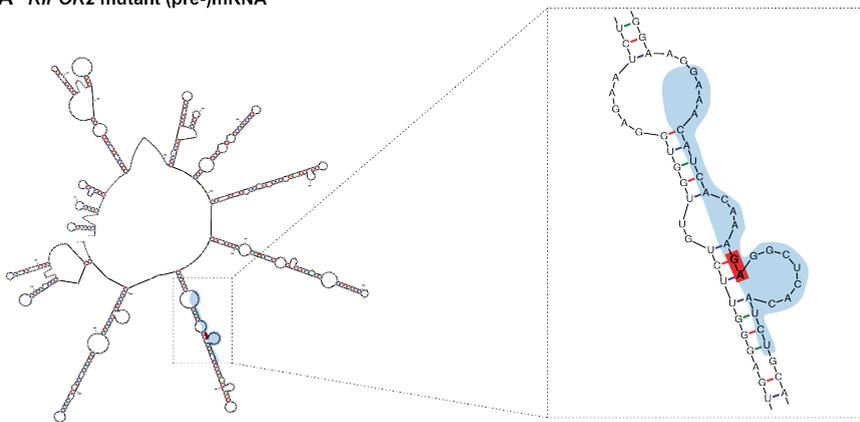
AONs induce efficient knockdown of mutant *RIPOR2* transcripts in patient-derived fibroblast cells

To assess the potency of the AON strategy and the accessibility of the AON-target region of the mutant *RIPOR2* transcript, DFNA21 patient-derived primary fibroblast cells were treated with PS-DNA oligos at a final concentration of 100 nM in the culture medium. This dose was selected based on earlier findings described by de Vrieze et al.¹¹ where significant reduction in mutant *COCH* transcripts was achieved with similarly designed AONs. The non-gapmer composition was selected for its strong ability to recruit RNase H1. All three c.1696_1707del-targeting AON molecules were able to induce a significant reduction in transcript levels of the mutant allele (ranging from 60-78%) compared to cells that were treated with transfection reagent only, indicating that the target region in mutant *RIPOR2* is indeed accessible to AONs (**Figure 2A**). However, a reduction of wildtype *RIPOR2* transcript levels was observed for all three AONs. The strongest concomitant reduction in wildtype *RIPOR2* transcripts was observed for AON 3, for which a 78% reduction in mutant transcripts coincides with a 58% reduction of wildtype transcripts.

As a next step, AONs 1 and 2 were ordered with a gapmer chemistry that included a central gap region of PS-DNA bases flanked by 2'-O-methyl RNA wings (AON 6 and AON 7, respectively). Although the gapmer composition is known to decrease RNase H1 cleavage efficiency, their improved stability and reduced toxicity, make them particularly attractive for clinical applications. In addition, the inability of RNase H1 to

cleave 2'-O-methyl RNA bases was exploited to improve allele-specificity. In view of the preferred cleavage preference of RNase H1²², an asymmetric wing design was preferred for AON 6 (7-10-3), whereas symmetric wings were chosen for AON 7 (5-10-5) (**Figure 1**). Additionally, the AON 7 gapmer sequence was shifted by one nucleotide compared to AON 2. Treatment with AON 6 and 7 both led to a significant reduction of mutant *RIPOR2* transcript levels (33% and 51%, respectively) as compared to treatment with transfection reagent only. Only AON 6 showed allele-specific knockdown as no significant reduction of wildtype transcripts was observed (**Figure 2B**).

A *RIPOR2* mutant (pre-)mRNA



B *RIPOR2* AON design



Figure 1. Design of *RIPOR2*-targeting AONs. (A) *In silico* prediction of the most-probable structure of the mutant *RIPOR2* (pre-)mRNA. The 12-nucleotide target deletion breakpoint-nucleotides are marked in red. The mRNA conformation was analyzed using the mfold Web Server, which revealed a mixture of open (non-base paired) and closed nucleotides. (B) Design of 20-mer antisense oligonucleotides (AONs) spanning the 12-nucleotide target deletion (dotted lines). AONs 1-3 consist of a complete phosphorothioate (PS)-linked DNA backbone, whereas AONs 6 and 7 are PS-linked gapmer molecules that contain a DNA gap flanked by 2'-O-methyl RNA wings (2'-OMe, depicted in blue).

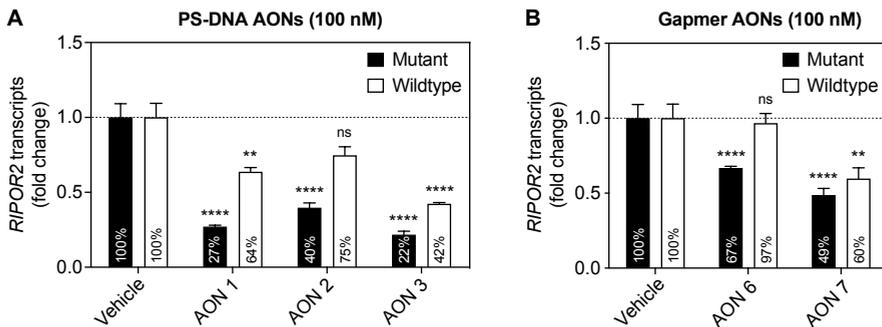


Figure 2. Screening of candidate antisense oligonucleotides. Degradation of mutant and wildtype *RIPOR2* transcripts in DFNA21 patient-derived fibroblast cells by antisense oligonucleotides (AONs) (100 nM in the culture medium), directed against the c.1696_1707del *RIPOR2* variant. **(A)** All three phosphorothioate (PS)-modified DNA AONs (1-3) were able to significantly reduce mutant transcript levels, 24 hours after transfection. **(B)** Gapmer AONs (AON 6 and AON 7) were designed based on sequences of AON 1 and AON 2, respectively. Both AONs were able to significantly decrease mutant *RIPOR2* transcript levels, but only AON 6 showed allele-discriminative potential as it did not significantly reduce wildtype transcript levels, 24 hours after transfection. Data are expressed as mean \pm SEM of three replicate transfections, normalized to the expression of *GUSB* and displayed as the fold change compared to cells treated with transfection reagent only (vehicle). ** $p < 0.01$, **** $p < 0.0001$, one-way ANOVA with Tukey's post-test.

AON 6 induces a dose-dependent and specific decrease in mutant *RIPOR2* transcript levels

To further investigate the efficacy and allele-specificity of AON 6, a dose-response analysis was performed. The AONs were transfected in patient-derived fibroblast cells, with a final concentration in the culture medium ranging from 50 nM to 250 nM. A significant reduction in transcripts of the mutant allele could be observed when treated with concentrations ≥ 150 nM (**Figure 3**). A maximum knockdown of 68% was achieved and a clear dose-dependent response could be observed. For none of the concentrations, a significant reduction in transcripts of the wildtype allele was observed. Although a non-significant decrease in wildtype transcripts at 50 nM compared to transfection reagent only was found, this change was not dose-dependent. This suggests that AON 6 has the ability to specifically target the mutant *RIPOR2* allele for RNase H1-mediated degradation.

Validation of the lead AON molecule in HEK293T cells

We questioned whether the lower knockdown efficiency of AON 6 as compared to the full PS-DNA AON is, in part, due to a lower transfection efficiency. To overcome this potential limitation, we co-transfected AON 6 with vectors encoding FLAG-tagged mutant or wildtype *RIPOR2* in HEK293T cells.

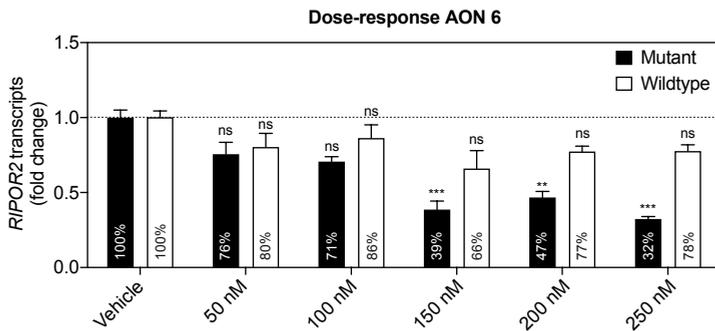


Figure 3. Dose-response analysis of AON 6. To validate the potency of the previously identified lead molecule, AON 6, the efficacy and specificity of the molecule was evaluated at different concentrations (50 nM – 250 nM in the medium) in DFNA21 patient-derived fibroblast cells. A significant decrease in mutant *RIPOR2* transcripts 24 hours after transfection was observed at concentrations ≥ 150 nM. No significant reduction of wildtype transcripts was observed for any of the doses. Data are expressed as mean \pm SEM of three replicate transfections, normalized to the expression of *GUSB* and displayed as the fold change compared to cells treated with transfection reagent only (vehicle). ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, one-way ANOVA with Tukey's post-test.

Although this experimental paradigm does not optimally reflect the patient situation, the ability to assess the effect of the AONs on protein translation is an important advantage over patient-derived fibroblasts.

We co-supplemented HEK293T cells with DNA constructs encoding FLAG-tagged mutant or wildtype *RIPOR2*, together with AON 6. The maximum AON concentration (250 nM) that was tested in fibroblast cells was selected for this experiment. Again, an allele-specific knockdown was observed on the transcript level (88% mutant transcript reduction (p -value < 0.0001 , ****) versus 8% wildtype transcript reduction (p -value 0.5591, not significant) (**Figure 4A**). Additionally, we performed a western blot analysis to confirm the effect of the AON molecule on *RIPOR2* translation. In three replicate AON deliveries, western blot analysis of *RIPOR2* revealed a strong decrease in mutant *RIPOR2* translation (95% reduction). However, also a decrease in wildtype *RIPOR2* protein levels was observed (80% reduction), larger than anticipated based on the knockdown of wildtype transcript levels.

DISCUSSION

DFNA21, resulting from an in-frame deletion in the *RIPOR2* gene (c.1696_1707del), is a frequent form of dominantly inherited HL, and is estimated to affect thousands of individuals within Europe.¹ The DFNA21 HL phenotype is highly variable, but in

general displays an adult-onset of symptoms which provides a window of opportunity for therapeutic intervention. Inspired by the non-haploinsufficiency pathogenic mechanism of the variant, we explored the specific silencing of the mutant *RIPOR2* allele as a potential therapeutic strategy for DFNA21. Several AON molecules were designed and evaluated for their ability to target mutant *RIPOR2* transcripts for degradation by the RNase H1 enzyme, thereby blocking the synthesis of (toxic) mutant *RIPOR2* proteins whilst leaving wildtype *RIPOR2* transcripts intact. We identified a lead molecule that showed the potency to efficiently and specifically reduce mutant *RIPOR2* transcript levels and, to a lesser extent, protein synthesis.

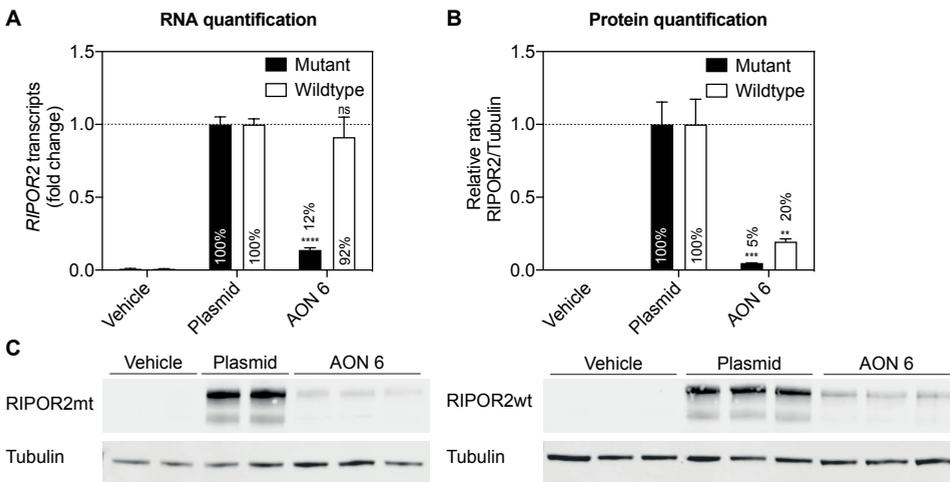


Figure 4. Evaluation of efficacy and specificity of AON 6 in HEK293T cells. HEK293T cells were co-transfected with mutant- or wildtype *RIPOR2* cDNA constructs and AON 6 (250 nM in the medium). RNA and protein were isolated 24 hours afterwards. **(A)** RT-qPCR analysis revealed a significant reduction of mutant *RIPOR2* transcript levels when HEK293T cells were treated with AON 6. Data are expressed as mean \pm SEM of six replicate transfections, normalized to the expression of *GUSB* and displayed as the fold change compared to cells treated with *RIPOR2* plasmid and transfection reagent only (plasmid). Vehicle transfected cells were transfected with transfection reagent only and show no endogenous *RIPOR2* expression. **(B-C)** Western blot analyses were performed using anti-RIPOR2 and anti-tubulin antibodies. Analyses confirmed a reduction in mutant *RIPOR2* protein synthesis in cells treated with AON 6. Also a reduction in wildtype *RIPOR2* protein levels was observed. Quantification of western blot results was performed using the Fiji software (v1.47). Data are expressed as mean \pm SEM of three replicate transfections, the relative ratio of *RIPOR2* protein to tubulin protein was calculated and compared to cells treated with transfection reagent and *RIPOR2* only (plasmid). ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, one-way ANOVA with Tukey's post-test.

The ability of AONs to specifically target mutant transcripts for degradation is of key importance for the development of an AON-based therapy for the c.1696_1707del *RIPOR2* gain-of-function allele. The therapeutic strategy must induce a large enough

decrease in mutant *RIPOR2* synthesis, but allow sufficient synthesis of wildtype protein to sustain normal inner ear function. We previously developed and *in vitro* validated an allele-specific AON for the single nucleotide change c.151C>T in *COCH* that is associated with hearing loss and vestibular dysfunction type DFNA9.¹¹ The 12-nt difference between mutant and wildtype *RIPOR2* would offer, in theory, a better opportunity to design mutant allele-specific AONs.

All of the AON molecules that were included in this study, both full PS-DNA AONs and gapmers, were able to significantly reduce *RIPOR2* transcript levels. Surprisingly, a clear mutant allele-specific effect was only observed for gapmer AON 6. In patient-derived fibroblasts, 250 nM of AON 6 resulted in a 68% reduction of mutant *RIPOR2* transcripts accompanied by a mild, statistically not significant, decrease in wildtype *RIPOR2* transcripts. Upon *in silico* predictions of the 3D structure of the AON-target duplexes, a potential explanation of the lack of mutant allele-specificity of the PS-DNA AONs emerged. Unlike for AONs directed to a specific missense variant, all nucleotides of the AON can form base pairs with their complementary nucleotides of both the mutant and the wildtype transcript. Indeed, base pairing of the PS-DNA AON with wildtype *RIPOR2* is predicted by RNAfold²³ to lead to the formation of 12-nucleotide loop in the target RNA, with 11- to 15-nucleotide stretches of PS-DNA/RNA duplexes that can still recruit RNase H1. The slightly lower predicted binding affinity for the wildtype transcript is perfectly reflected in the slightly stronger decrease in mutant transcript levels than in wildtype transcript levels that is observed upon delivery of the PS-DNA AONs to patient-derived primary fibroblasts. 2'-O-methyl (2'-OMe) sugar modified RNA nucleotides are resistant to RNase H1 activity, and were used in the gapmer AONs to reduce the ability of the AONs to recruit RNase H1 to the stretches flanking the 12-nucleotide loop. The resulting gapmer AON 6 displayed stronger allele-specificity, indicating that the ability of the AON wings to recruit RNase H1 was indeed sufficiently decreased. The specificity of AON 6 for the mutant allele was confirmed in HEK293T cells. Co-transfection of AON 6 with a mutant *RIPOR2* cDNA construct led to a decrease of 88% in mutant *RIPOR2* transcripts, whereas co-transfection of AON 6 with a wildtype *RIPOR2* cDNA construct did not significantly affect wildtype *RIPOR2* transcript levels (8% reduction). Unfortunately, based on the western blot results, this effect was less specific at the protein level as a significant reduction of wildtype protein was observed (80%). A potential explanation could be that the delivery of an AON, independent of the sequence, has an effect on protein translation. While discrepancies between transcript and protein levels are common, the observed mutant *RIPOR2* reduction (95%) is also higher than observed at the transcript level (88%). This indicates that the difference might be (partially) attributed to a general effect of AON delivery, and should be further investigated using a control AON treatment in future experiments.

Interestingly, when tested at the same concentration of 100 nM, the target knockdown induced by AON 1 (non-gapmer sequence of AON 6) was much higher than that of AON 6 (73% mutant reduction AON 1, versus 29-33% AON 6). Because of their complete DNA chemistry, PS-DNA AONs are known to have a higher RNase H1-recruiting activity than gapmer sequences. In addition, the differences in effect between these two AONs could also be attributed to differences in transfection efficiency, chemical toxicity of the PS-DNA AON, subcellular distribution, or any combination of these factors. Although to the best of our knowledge not reported before, the strong and highly mutant allele-specific knockdown of *RIPOR2* induced by AON 6 in cDNA construct-transfected HEK293T cells suggests that transfection efficiency may have well been a limitation in the patient-derived fibroblasts. The use of a fluorescently-conjugated AON, and gymnotic delivery of AON 6, will shed more light on this. Additionally, alternative AON chemistries could be explored in parallel to enhance the on-target efficiency of the gapmer molecule. In this study, 2'-O-methyl (2'-OMe) sugar modified RNA wings were employed in the gapmer design. In several studies, alternative sugar modifications have been described that can potentially increase the potency of the molecule. A study performed in HeLa cells indicated that the 2'-O-methoxyethyl (2'-MOE) modified AONs are consistently more effective in suppressing *CTNFB1* RNA levels compared to the corresponding 2'OMe AONs.²⁴ Alternatively, Rukov et al. reported the use of locked-nucleic acid (LNA)-modified gapmer wings as a promising approach to increase binding affinity and stability of the molecule.²⁵ By extending the free binding energy of the AON and the intended RNA target, the observed knockdown was increased and even higher than observed using full PS-DNA AONs. However, increasing the binding affinity of the AON may also increase the affinity for the wildtype *RIPOR2* transcript. Therefore, adjusting the AON wing design or gap size, may offer a better chance at optimizing efficiency and specificity for the mutant *RIPOR2* transcript (reviewed in (26)).

Importantly, it should be acknowledged that the fibroblast cell model is probably not ideal to study inner ear disease. While the patient-specific genotype is advantageous over engineered cell models such as plasmid-transfected HEK293T cells, or stable transgenic cell models, they poorly reflect the transcriptional and cellular nature of the *RIPOR2*-expressing hair cells in the inner ear. Therefore, alternative more relevant cell models, such as otic progenitor cells or inner ear organoids²⁷, should be considered in future studies to reliably determine the necessity to optimize the chemistry and gapmer design of the lead AON. The RNA profile of these cell types is more representative for the auditory hair cells, and therefore more potent to screen for any potential adverse effects.

Essential questions that remain are how much reduction of mutant allele expression is needed to achieve a therapeutic effect and how much expression of the wildtype allele should be maintained for normal hair cell function. Here, even organoid models likely fall short, and the generation of a knockin animal model should be considered. An animal model could provide valuable insights in the feasibility and efficacy of the lead AON to relieve the burden of mutant RIPOR2 proteins in the auditory hair cells of the cochlea. These studies could also shed light on the balance between the knockdown of mutant RIPOR2 and maintained wildtype RIPOR2 synthesis that is needed for a clinically meaningful improvement. While the therapeutic efficacy of near complete and highly mutant allele-specific knockdown of *RIPOR2* seems apparent, this does not necessarily reflect the minimal decrease in mutant *RIPOR2* transcript levels that is required to lower the amount of mutant RIPOR2 protein to a level that can halt, or significantly delay, disease progression. With the current knowledge, it can only be speculated to which extent mutant allele knockdown would be required to achieve therapeutic potential. DFNA21 is an adult-onset, progressive disease, suggesting that mutant protein toxicity and burden is only slowly accumulating. Removing a small percentage of this burden could potentially already delay the onset of HL for several years, especially when treatment is initiated at an early age. Therefore, we speculate that a >50% reduction of mutant transcript levels, achieved upon delivery of 250 nM of the lead AON to patient-derived fibroblasts, might already be sufficient to provide a clinically meaningful outcome to DFNA21 patients especially when treatments are started at an early stage. Additionally, exon-skipping studies have indicated that already ~20% of wildtype allele expression can be enough for phenotypic rescue in recessively inherited hearing loss and retinal degeneration.^{28,29} Hypothetically, an approximate 50% reduction in wildtype *RIPOR2* transcripts could be tolerated in order to achieve a strong reduction in mutant *RIPOR2* transcripts.

The transient nature of gapmer-mediated gene knockdown is both an advantage and a potential drawback for future clinical applications. It lowers the risk of sustained adverse or off-target effects that could accompany genome editing techniques, but repeated delivery of gapmers is likely required to achieve maximum efficacy. However, a repeated surgical delivery directly into the cochlea is not feasible. The recent study by Lentz et al. indicated that, in a mouse model of USH1C, AONs can diffuse over the round window membrane to reach auditory hair cells of the adult cochlea.¹⁵ Also, intratympanic injections of steroids such as dexamethasone, are routinely and repeatedly used to treat e.g. Meniere's disease or idiopathic sudden sensorineural hearing loss which suggests this method could be very promising for repeated AON delivery as well.^{30,31} Recent studies on splice-modulation AONs in the mouse revealed a relatively stable effect on transcript level up to 200 days post-delivery. While the delivery method and

AON chemistry are not completely comparable, these data suggest a repetition rate of gapmer treatments for inner ear disorders of once or twice a year. In comparison, weekly or daily intratympanic injections, albeit for a short period of time, are an established treatment regimen for several other inner ear therapeutics.³²⁻³⁴ Despite the technical feasibility of repeated intratympanic injections, there are significant knowledge gaps in terms of AON uptake, biodistribution and half-life in the cochlea and auditory hair cells that need to be addressed before further clinical development of gapmer AON treatments for DNFA21 and other forms of dominantly inherited HL can be initiated.

In conclusion, based on the results obtained in this study and the developments in inner ear therapeutics, we consider AON 6 as a potent molecule to treat DFNA21 in the future. Although there are still some avenues to be explored to further optimize the molecular efficacy of the AON, the results obtained in the current study warrant further preclinical development of the lead molecule in advanced cellular models and animal models. With successful application of AON therapies for many other inherited disorders, and the rapid developments in the fields of AON therapeutics and cochlear drug delivery, we are confident that the specific challenges for inner ear AON therapeutics will be resolved in the future.

ACKNOWLEDGEMENTS

The authors thank Hedwig Velde for her help with tissue collection and Saskia van der Velde-Visser, Marlie Jacobs-Camps and Anita Roelofs for their contributions to cell culture. This study was funded by a DCMN Radboudumc grant (to H.K. and F.P.M.C.).

REFERENCES

1. de Bruijn, S.E., Smits, J.J., Liu, C., Lanting, C.P., Beynon, A.J., Blankevoort, J. *et al.* A RIPOR2 in-frame deletion is a frequent and highly penetrant cause of adult-onset hearing loss. *Journal of Medical Genetics* **58**, 96-104 (2021).
2. Kunst, H., Marres, H., Huygen, P., Van Duijnhoven, G., Krebsova, A., Van der Velde, S. *et al.* Non-syndromic autosomal dominant progressive non-specific mid-frequency sensorineural hearing impairment with childhood to late adolescence onset (DFNA21). *Clinical Otolaryngology & Allied Sciences* **25**, 45-54 (2000).
3. De Brouwer, A.P.M., Kunst, H.P.M., Krebsova, A., van Asseldonk, K., Reis, A., Snoeckx, R.L. *et al.* Fine mapping of autosomal dominant nonsyndromic hearing impairment DFNA21 to chromosome 6p24.1-22.3. *American Journal of Medical Genetics* **137A**, 41-46 (2005).
4. Zhao, B., Wu, Z. & Müller, U. Murine Fam65b forms ring-like structures at the base of stereocilia critical for mechanosensory hair cell function. *eLife* **5**, e14222 (2016).
5. Diaz-Horta, O., Abad, C., Cengiz, F.B., Bademci, G., Blackwelder, P., Walz, K. *et al.* Ripor2 is involved in auditory hair cell stereociliary bundle structure and orientation. *Journal of Molecular Medicine* **96**, 1227-1238 (2018).
6. Diaz-Horta, O., Subasioglu-Uzak, A., Grati, M.h., DeSmidt, A., Foster, J., Cao, L. *et al.* FAM65B is a membrane-associated protein of hair cell stereocilia required for hearing. *Proceedings of the National Academy of Sciences* **111**, 9864-9868 (2014).
7. Vickers, T.A. & Crooke, S.T. The rates of the major steps in the molecular mechanism of RNase H1-dependent antisense oligonucleotide induced degradation of RNA. *Nucleic Acids Research* **43**, 8955-8963 (2015).
8. Crooke, S.T. Molecular mechanisms of action of antisense drugs. *Biochimica et Biophysica Acta* **1489**, 31-44 (1999).
9. Vickers, T.A. & Crooke, S.T. Antisense oligonucleotides capable of promoting specific target mRNA reduction via competing RNase H1-dependent and independent mechanisms. *PLOS ONE* **9**, e108625 (2014).
10. Crooke, S.T., Baker, B.F., Crooke, R.M. & Liang, X.-h. Antisense technology: an overview and prospectus. *Nature Reviews Drug Discovery* **Online ahead of print** (2021).
11. de Vrieze, E., Cañas Martín, J., Peijnenborg, J., Martens, A., Oostrik, J., van den Heuvel, S. *et al.* AON-based degradation of c.151C>T mutant COCH transcripts associated with dominantly inherited hearing impairment DFNA9. *Molecular Therapy - Nucleic Acids* **24**, 274-283 (2021).
12. Hammond, S.M., Aartsma-Rus, A., Alves, S., Borgos, S.E., Buijsen, R.A.M., Collin, R.W.J. *et al.* Delivery of oligonucleotide-based therapeutics: challenges and opportunities. *EMBO Molecular Medicine* **13**, e13243 (2021).
13. Vázquez-Domínguez, I., Garanto, A. & Collin, R.W.J. Molecular therapies for inherited retinal diseases-current standing, opportunities and challenges. *Genes (Basel)* **10**, 654 (2019).
14. Delmaghani, S. & El-Amraoui, A. Inner ear gene therapies take off: Current promises and future challenges. *Journal of Clinical Medicine* **9**, 2309 (2020).

15. Lentz, J.J., Pan, B., Ponnath, A., Tran, C.M., Nist-Lund, C., Galvin, A. *et al.* Direct delivery of antisense oligonucleotides to the middle and inner ear improves hearing and balance in usher mice. *Molecular Therapy* **28**, 2662-2676 (2020).
16. Delprat, B., Boulanger, A., Wang, J., Beaudoin, V., Guitton, M.J., Ventéo, S. *et al.* Downregulation of otospiralin, a novel inner ear protein, causes hair cell degeneration and deafness. *Journal of Neuroscience* **22**, 1718-25 (2002).
17. Pallan, P.S. & Egli, M. Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a non-specific enzyme-dsDNA complex. *Cell Cycle* **7**, 2562-2569 (2008).
18. Aartsma-Rus, A., van Vliet, L., Hirschi, M., Janson, A.A.M., Heemskerk, H., de Winter, C.L. *et al.* Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms. *Molecular Therapy* **17**, 548-553 (2009).
19. Slijkerman, R., Kremer, H. & van Wijk, E. Antisense oligonucleotide design and evaluation of splice-modulating properties using cell-based assays. *Methods in Molecular Biology* **1828**, 519-530 (2018).
20. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415 (2003).
21. Roosing, S., Lamers, I.J., de Vrieze, E., van den Born, L.I., Lambertus, S., Arts, H.H. *et al.* Disruption of the basal body protein POC1B results in autosomal-recessive cone-rod dystrophy. *American Journal of Medical Genetics* **95**, 131-142 (2014).
22. Kiełpiński, Ł.J., Hagedorn, P.H., Lindow, M. & Vinther, J. RNase H sequence preferences influence antisense oligonucleotide efficiency. *Nucleic Acids Research* **45**, 12932-12944 (2017).
23. Reuter, J.S. & Mathews, D.H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
24. Wang, B. An ASO modification that enhances nuclease resistance, lowers toxicity, and increases binding affinity - the 2'MOE analog. (Integrated DNA Technologies, 2018).
25. Rukov, J.L., Hagedorn, P.H., Høy, I.B., Feng, Y., Lindow, M. & Vinther, J. Dissecting the target specificity of RNase H recruiting oligonucleotides using massively parallel reporter analysis of short RNA motifs. *Nucleic Acids Research* **43**, 8476-8487 (2015).
26. Hagedorn, P.H., Hansen, B.R., Koch, T. & Lindow, M. Managing the sequence-specificity of antisense oligonucleotides in drug discovery. *Nucleic Acids Research* **45**, 2262-2282 (2017).
27. Tang, P.C., Hashino, E. & Nelson, R.F. Progress in modeling and targeting inner ear disorders with pluripotent stem cells. *Stem Cell Reports* **14**, 996-1008 (2020).
28. Dulla, K., Slijkerman, R., van Diepen, H.C., Albert, S., Dona, M., Beumer, W. *et al.* Antisense oligonucleotide-based treatment of retinitis pigmentosa caused by USH2A exon 13 mutations. *Molecular Therapy* **In press** (2021).
29. Lentz, J.J., Jodelka, F.M., Hinrich, A.J., McCaffrey, K.E., Farris, H.E., Spalitta, M.J. *et al.* Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. *Nature Medicine* **19**, 345-350 (2013).
30. de Cates, C. & Winters, R. Intratympanic steroid injection. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2021).

31. Cao, Z., Yue, F., Huang, W., Rajenderkumar, D. & Zhao, F. Different medications for the treatment of Ménière's disease by intratympanic injection: A systematic review and network meta-analysis. *Clinical Otolaryngology* **44**, 619-627 (2019).
32. Staecker, H., Morelock, M., Kramer, T., Chrbolka, P., Ahn, J.H. & Meyer, T. Safety of repeated-dose intratympanic injections with AM-101 in acute inner ear tinnitus. *Otolaryngology-Head and Neck Surgery* **157**, 478-487 (2017).
33. Leng, Y., Liu, B., Zhou, R., Liu, J., Liu, D., Zhang, S.-L. *et al.* Repeated courses of intratympanic dexamethasone injection are effective for intractable Meniere's disease. *Acta Otolaryngologica* **137**, 154-160 (2017).
34. Liu, B., Leng, Y., Zhou, R., Liu, J., Liu, D., Zhang, S.L. *et al.* Intratympanic steroids injection is effective for the treatment of drop attacks with Ménière's disease and delayed endolymphatic hydrops: A retrospective study. *Medicine* **95**, e5767 (2016).

SUPPLEMENTARY TABLE

Table S1. Chemical and thermodynamic properties of AON molecules

AON	AON sequence	2 nd structure free energy (kcal/mol)		Self-dimerization (kcal/mol)		Free energy of bimolecular structure (kcal/mol)		Delta
		As DNA	As RNA	As DNA	As RNA	Mutant allele	Wildtype allele	
AON 1	G*A*G*C* T *G* T *T*G* T *G*A* T *G* T *T* T *C	1.5	-1.6	-4.4	-6.8	-33.9	-26.9	7.0
AON 2	T*G* T *G*A*G*C* C * T * C *T* T *T*G* T *G*A* T *G* T	1.6	0.1	-4.4	-6.8	-35.2	-28.2	7.0
AON 3	A*G*A* T *G* T *G*A*G*C* C * T * C *T* T *T*G* T *G*A	1.6	1.6	-4.4	-6.8	-35.3	-28.3	7.0
AON 6	(G*A*G*)C* T * C *T* T *T*G* T *G*(A*U*G*U*U*U*C)	1.5	-1.6	-4.4	-6.8	-33.9	-26.9	7.0
AON 7	(A*U*G*U*G*)A*G*C* C * T * C *T* T *T*G*(U*G*A*U*G)	1.6	1.6	-4.4	-6.8	-34.5	-27.5	7.0

Phosphorothioate links in the antisense oligonucleotide (AON) sequences are indicated by the asterisks between bases. The 2'-O-methyl RNA bases are placed between brackets. Allele-discriminating bases for the c.1696_1707delR1692 allele are indicated by bold underlined fonts. The RNAstructure webserver (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) was used to predict the secondary structure free and self-dimerization energies of each molecule. Gapmer molecules were analyzed as both DNA and RNA molecules, since the RNA-structure webserver cannot take the chemical modification or gapmer composition into account. Candidate AON molecules were selected according to the criteria presented by Slijkerman et al.¹⁹ A Nucleotide BLAST analysis (NCBI) was performed to confirm specificity of the designed molecules.



Chapter 4

Structural variants create new topological associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa

Suzanne E. de Bruijn*, Alessia Fiorentino*, Daniele Ottaviani, Stephanie Fanucchi, Uirá S. Melo, Julio C. Corral-Serrano, Timo Mulders, Michalis Georgiou, Carlo Rivolta, Nikolas Pontikos, Gavin Arno, Lisa Roberts, Jacquie Greenberg, Silvia Albert, Christian Gilissen, Marco Aben, George Rebello, Simon Mead, F. Lucy Raymond, Jordi Corominas, Claire E.L. Smith, Hannie Kremer, Susan Downes, Graeme C. Black, Andrew R. Webster, Chris F. Inglehearn, L. Ingeborgh van den Born, Robert K. Koenekoop, Michel Michaelides, Raj S. Ramesar, Carel B. Hoyng, Stefan Mundlos, Musa M. Mhlanga, Frans P.M. Cremers, Michael E. Cheetham, Susanne Roosing[#], Alison J Hardcastle[#]

*, [#] These authors contributed equally to this work

American Journal of Human Genetics **107**, 802-814 (2020)

ABSTRACT

The cause of autosomal dominant Retinitis Pigmentosa (adRP), which leads to loss of vision and blindness, was investigated in families lacking a molecular diagnosis. A refined locus for adRP on Chr17q22 (RP17) was delineated through genotyping and genome sequencing, leading to the identification of structural variants (SVs) that segregate with disease. Eight different complex SVs were characterized in 22 adRP families with >300 affected individuals. All RP17-SVs had breakpoints within a genomic region spanning *YPEL2* to *LINC01476*. To investigate the mechanism of disease, fibroblasts from affected individuals and controls were reprogrammed into induced pluripotent stem cells (iPSCs) and differentiated into photoreceptor precursor cells (PPCs) or retinal organoids (ROs). Hi-C was performed on ROs, and differential expression of regional genes and a retinal enhancer RNA at this locus was assessed by qPCR. The epigenetic landscape of the region, and Hi-C RO data, showed that *YPEL2* sits within its own topologically associating domain (TAD), rich in enhancers with binding sites for retinal transcription factors. The Hi-C map of RP17 ROs revealed creation of a neo-TAD with ectopic contacts between *GDPD1* and retinal enhancers, and modelling of all RP17-SVs was consistent with neo-TADs leading to ectopic retinal specific enhancer-*GDPD1* accessibility. qPCR confirmed increased expression of *GDPD1*, and increased expression of the retinal enhancer that enters the neo-TAD. Altered TAD structure resulting in increased retinal expression of *GDPD1* is the likely convergent mechanism of disease, consistent with a dominant gain-of-function. Our study highlights the importance of SVs as a genomic mechanism in unsolved Mendelian diseases.

INTRODUCTION

Despite recent advances in next generation sequencing, approximately 30-40% individuals with inherited retinal diseases (IRDs) lack a molecular diagnosis. This is probably due to a combination of rare novel disease genes, which require large cohorts for validation, and previously intractable mutation classes such as intronic variants, structural variants (SVs), and variants in regulatory regions.^{1,2}

The most common form of IRD is Retinitis Pigmentosa (RP), which is genetically heterogeneous, with a prevalence of 1 in 4,000.³ RP is defined as a retinal degeneration that primarily affects rod photoreceptors, resulting in night blindness and progressive loss of peripheral vision, often progressing into the central retina and affecting cone photoreceptors, leading to severe visual impairment or blindness.⁴ Autosomal dominant RP (adRP) accounts for 25-40% of cases, depending on the population studied, and has been associated with mutations in 30 genes, including *CA4* on Chr17q23.1 (RP17).⁵⁻⁷ Following initial publications defining this locus^{8,9} a variant in *CA4* was implicated as the cause of adRP in families of South African origin, however pathogenicity of the reported variant has been questioned as it has a population frequency of 4% in healthy controls in Northern Sweden.¹⁰⁻¹² Subsequently reported *CA4* variants in individuals with RP were identified by targeted Sanger sequencing, and do not fully exclude variants in other genes as a cause of disease (**Table S1**).

We investigated the cause of adRP in unsolved families, including the first pedigree (GC1, referred to as UK1) drawn up at Moorfields Eye Hospital over 35 years ago, and the original Dutch family (W97-079, referred to as NL1), which showed linkage to the RP17-locus, but lacked a mutation in *CA4*.⁹

Here, we report identification and characterization of complex SVs on Chr17q22, through whole genome sequencing (WGS), as the genomic cause of adRP at the RP17-locus in a large number of families including the families of South African origin. To explore a convergent mechanism of disease, we investigated the effect of RP17-SVs on three-dimensional (3D) chromatin organization that results in the compartmentalization of the genome into topologically associating domains (TADs), and the epigenetic landscape of the region. TADs are chromatin domains within the genome that facilitate enhancer promoter contacts within the nuclear 3D space.¹³ Disruption of TAD structures can lead to loss of chromosomal contact between regulatory regions and their target genes, or the formation of novel active domains with ectopic contacts occurring between regulatory regions and a new target gene, resulting in pathogenic alterations in gene expression.¹⁴⁻¹⁷ We demonstrate that altered TAD structure at the RP17-locus

leads to ectopic retinal enhancer-gene interactions, consistent with a dominant gain-of-function. Our study highlights the pathogenicity of SVs that alter 3D chromatin organization and gene expression by rearranging TAD structures, and the need to revisit rare Mendelian diseases where genes and variants have not been substantiated in other cohorts.

MATERIALS AND METHODS

Study cohort

The study was approved by the medical ethics committee of the ErasmusMC Rotterdam, Radboudumc Nijmegen and Moorfields Eye Hospital, and was performed in accordance with the principles of the World Medical Association Declaration of Helsinki. Informed consent was obtained from all participants or their legal representatives.

Genetic analyses

SNP genotyping was performed for index families NL1 and UK1 to define and refine the RP17-locus. Genomic DNA from affected individuals and their family members was analyzed by whole exome and genome sequencing. Sequence data was aligned to the Human Reference Genome build hg19. Variants were prioritized based on a minor allele frequency (MAF) ≤ 0.0001 in gnomAD. SVs were called using ExomeDepth, Manta Structural Variant Caller, Canvas Copy Number Variant Caller and Control-FREEC. Details of genotyping, sequencing and analysis pipelines are provided in **Supplementary Materials and Methods**.

Characterization and validation of structural variants

SV breakpoint junctions were PCR amplified and validated with Sanger sequencing. Primer sequences and coordinates are listed in **Table S2**. SV breakpoint regions were assessed for the presence of microhomology and repetitive elements. To validate a triplicated region for UK-SV6, quantitative real-time PCR (qPCR) was performed on genomic DNA from affected individuals from family UK13 and unaffected controls (**Supplementary Materials and Methods**).

Clinical analysis

Available clinical notes of cases for the pedigrees identified at Radboudumc, Moorfields Eye Hospital, University of Cape Town and McGill University Health Centre were reviewed, as well as detailed retinal imaging, fundus autofluorescence and optical coherence tomography. Age of onset is defined as the age at which symptoms were first experienced.

Interrogation of the genomic region

We interrogated chromatin and genome regulation datasets to explore the epigenomic landscape of the region. Available datasets were obtained and analyzed using UCSC genome browser (details of datasets used are provided in **Supplementary Materials and Methods**).

Reprogramming fibroblasts into iPSCs and differentiation into photoreceptor progenitor cells and 3D retinal organoids

Fibroblasts were cultured from skin biopsies of two individuals with NL-SV1, one individual with UK-SV2, and five anonymous control individuals. Cell lines were reprogrammed into induced pluripotent stem cells (iPSCs) and differentiated into photoreceptor progenitor cells (PPCs) following the previously described 60-day protocol (**Supplementary Materials and Methods**).^{18,19} 3D retinal organoids (ROs) were differentiated for UK-SV2 and controls, as previously described (**Supplementary Materials and Methods**).²⁰

Preparation of low input Hi-C libraries (Low-C)

Hi-C was performed on UK-SV2 and control 3D ROs using a low input protocol (Low-C) with few modifications (**Supplementary Materials and Methods**).²¹ Two libraries per sample were sequenced for 200 million fragments in a 100 bp paired-end run on a NovaSeq 6000 (Illumina). Paired-end sequencing data was processed using Juicer²² and the Hi-C maps were created using a bin size with 10 kb resolution. Further information about the bioinformatics pipeline is detailed in Melo et al., 2020.²³

Expression analysis of genes and enhancer RNA within the RPI7-locus

To assess expression of genes, qPCR was performed for different human tissues, including retina (**Table S3, Supplementary Materials and Methods**). Single cell RNA sequencing data of human²⁴ and primate²⁵ retinal cell types was obtained and visualized using the Broad Institute Single Cell Portal (**Supplementary Materials and Methods**).

cDNA was synthesized from total RNA extracted from PPCs, ROs and fibroblasts. Differential expression of genes implicated in the SVs, and control housekeeping and retinal progenitor genes, were assessed by qPCR (**Table S3, Supplementary Materials and Methods**). Primers were designed to the enhancer region containing multiple retinal transcription factor binding sites implicated in all SVs, to analyze targeted enhancer RNA expression by qPCR (**Table S3, Supplementary Materials and Methods**).

RESULTS

Refinement of the RP17-locus in two unrelated adRP families

The affected haplotype for a Dutch adRP family (NL1) (**Figure 1A**) was previously mapped to a 7.18 Mb region spanning the RP17-locus on chromosome 17.⁹ The RP17-locus was refined to a 5.16 Mb interval by SNP haplotyping in an extended pedigree (**Figure 1D, Supplementary Results**). No rare coding or splice site heterozygous variants (MAF ≤ 0.0001) shared between affected individuals were found through whole exome sequencing (WES). Subsequently WGS was performed, and similarly no rare candidate coding, splice site, intronic or intergenic heterozygous single nucleotide variants were identified (**Table S4, Supplementary Results**).

In parallel, WES and WGS was performed for affected individuals from a genetically unexplained UK adRP family (UK1) (**Figure 1B**). This also failed to identify a rare causative variant, however, a disease-associated haplotype on chromosome 17 was identified (**Figure 1E, Table S5, Supplementary Results**). Interrogation of unsolved IRD sequence data generated through the UK IRDC, UCL-Ex, NIHR-Bioresource, and Genomics England identified other adRP probands that shared the same haplotype of Chr17 SNVs, and established this as a founder haplotype in eleven additional UK adRP families (**Figure 1C**). The adRP locus was refined to a 4.4 Mb interval on Chr17q22 (**Figure 1E**). This genomic interval overlaps the previously described RP17-locus in families of Dutch and South African origin (**Figure 1F**).

A missense variant in *CA4* (c.40C>T; p.Arg14Trp; NM_000717.4) was previously described as the cause of adRP at the RP17-locus in families of South African origin.¹⁰ No rare coding, intronic or upstream variants in *CA4* were identified in the Dutch and UK families.

Identification of structural variants within the RP17-locus

Next, we analyzed genome and exome data for copy number variants and SVs (**Supplementary Results**). In family NL1, WGS revealed a 226-kb duplication within the RP17-locus: chr17:57,291,905_57,518,137dup (NL-SV1). This SV involves two duplicated genes (*GDPD1* and *YPEL2*), an intragenic microRNA (*MIR4729*), and partial duplication of *SMG8* and the long non-coding RNA *LINC01476*. The duplication creates a breakpoint junction (chr17:g.57,518,137-57,291,905) specific for the mutated allele in NL1 (**Figure 2A-B, Figure S1**), which was used to confirm segregation of the SV with the adRP phenotype in this family. No overlapping SVs in the RP17-locus were observed in WES of ~7,500 individuals without retinal disease generated in-house at the Department of Human Genetics, Radboudumc.

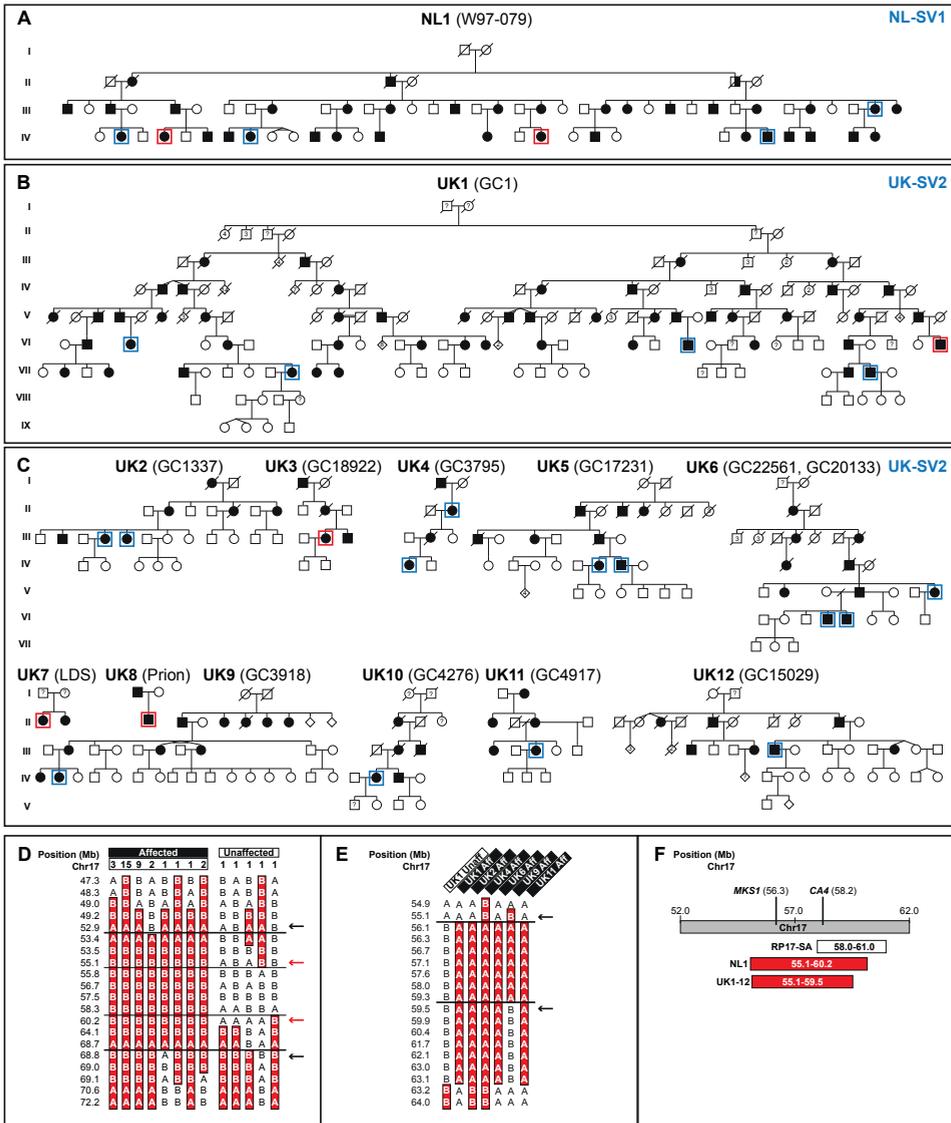


Figure 1. Mapping of the RP17-locus in two unrelated families. (A) Pedigree of Dutch NL1 family. **(B)** Pedigree of UK1 family. **(C)** Pedigrees of additional UK families with the founder haplotype on Chr17q. WGS or WES was performed in individuals highlighted in blue or red, respectively. **(D)** SNP haplotyping results for NL1. The refined RP17-locus (rs8078110-rs9910672) is shared by all affected individuals (n=35) and not present in unaffected individuals (n=28, only individuals with recombination close to or refining the critical region are depicted), with a maximum LOD-score of 15.0. The horizontal numbers represent the number of individuals with this haplotype. **(E)** UK founder haplotype refining the RP17-locus in UK families. Representative haplotypes from several unrelated families are shown with affected (aff) individuals compared to an unaffected (unaff) individual. Black lines and arrows indicate recombination events. Shared haplotype in individuals is shaded red. **(F)** overlap of refined RP17 loci in UK, NL and previously described SA families.¹⁰

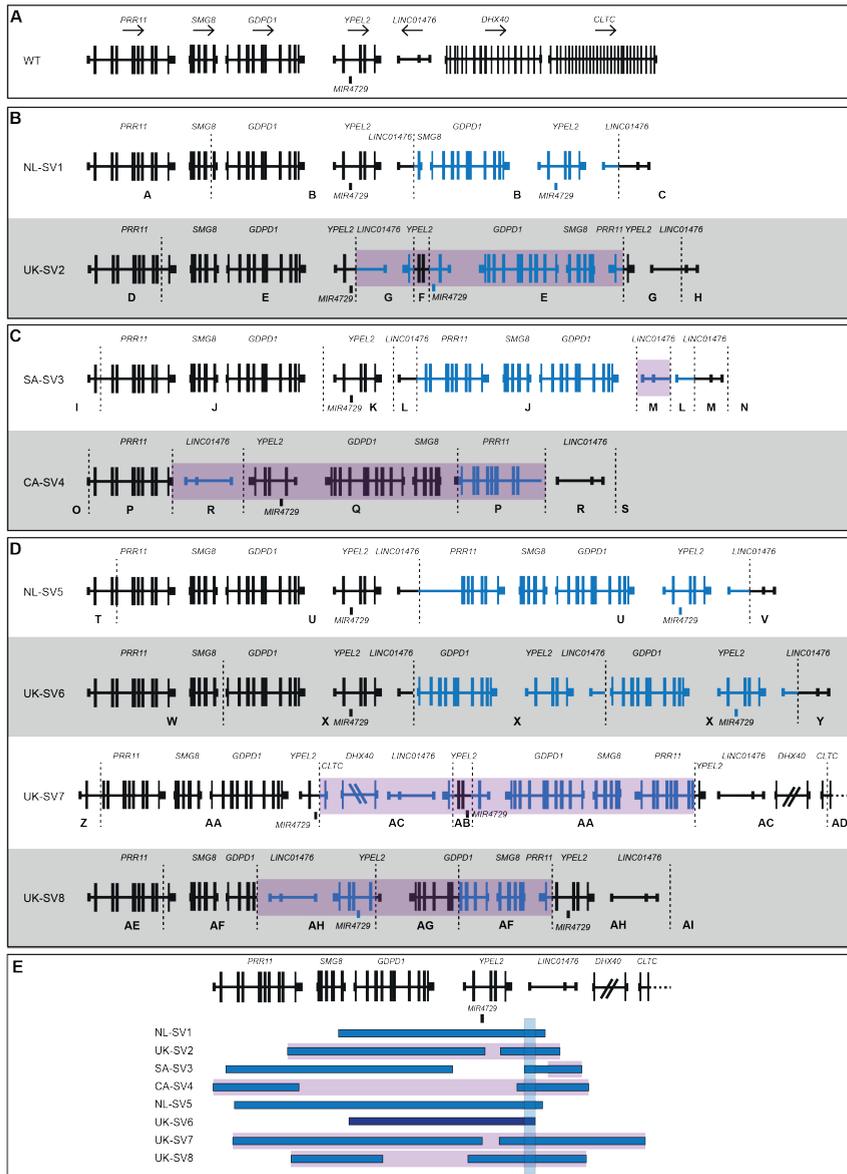


Figure 2. Overview of structural variants within the RP17-locus in adRP families. Breakpoints are indicated with dashed lines. Blue segments represent duplicated or triplicated regions, whereas inversions are highlighted in purple. **(A)** Wildtype (WT) chromosomal organization. **(B)** Structural variants identified in NL1 (NL-SV1) and UK founder haplotype families (UK-SV2). **(C)** Structural variants identified in adRP families that were previously linked to the RP17-locus; SA-SV3¹⁰, and CA-SV4 (unpublished data). **(D)** Structural variants found in a cohort of unsolved adRP families; NL-SV5, UK-SV6, UK-SV7, UK-SV8. Letters A-AI depict the genomic intervals for each SV used to analyze and annotate SV breakpoints. **(E)** Overview of all SV breakpoints identified in the RP17-locus. An overlapping genomic region that is duplicated or triplicated in all SVs was identified (chr17:57,499,214-57,510,765) and is highlighted by a light-blue vertical bar. The size of *DHX40* is reduced and *CLTC* is partially shown for the purpose of this figure.

For the 12 UK RP17 founder haplotype families, WGS revealed a duplicated inversion: chr17:57,456,098-57,468,960delins57,275,839_57,559,114inv (UK-SV2) (**Figure 2B**). The SV was characterized and breakpoint junctions were validated (**Figure S1, Supplementary Results**). This SV involved four coding genes (*PRR11*, *SMG8*, *GDPD1* and *YPEL2*) and two non-coding RNA genes (*MIR4729* and *LINC01476*) (**Figure 2B**). UK-SV2 segregated with adRP in all families where DNA was available for analysis. UK-SV2 was absent in WGS control genome data generated for 58,000 UK individuals (Genomics England).

Different structural variants within the RP17-locus in multiple adRP families

These data prompted us to investigate whether SVs were present in the two original South African families (SA1 and SA2) that were linked to the RP17-locus (**Figure S2A**).^{8,10} In addition, a Canadian adRP family (CA1) was also mapped to the RP17-locus (unpublished data, **Figure S2B**). WGS was performed for affected individuals from these families, and inversion duplication events were identified in all samples analyzed (**Figure 2C**). In SA1 and SA2, an identical SV, SA-SV3, was revealed suggesting this is a founder variant in this population. SA-SV3 was also found by breakpoint PCR in two additional families of South African origin (SA3 and SA4), confirming the founder effect (**Figure S2A**). In the Canadian family, a different inversion duplication event was identified, CA-SV4. SA-SV3 and CA-SV4 breakpoints were characterized and validated (**Figure S1**), and segregation of the SVs with the adRP phenotype was confirmed.

Our data suggested that SVs at the RP17-locus are an important cause of adRP. Therefore, WGS and WES data for genetically unexplained adRP-affected families were analyzed for SVs within this locus. In four unrelated families of Dutch or UK origin, four additional unique complex SVs were discovered (**Figure 2D, S2C**). For individuals that had only undergone WES, WGS was performed to determine the breakpoint junctions and identify potential inversions or other SVs. In all families, breakpoints were validated and segregation analysis was performed where possible. Triplication for UK-SV6 was confirmed by qPCR in family UK13 (**Figure S3, Supplementary Results**).

Details of all SVs identified in this study are shown in **Table S6, Figure 2** and **Figure S4**, and an overview of SV-specific breakpoint junctions is shown in **Figure S1**. All RP17-SVs share a common duplicated (or triplicated) region of 11.5 kb and harbor unique breakpoints disrupting the genomic region spanning *YPEL2* to *LINC01476* (chr17:57,499,214-57,510,765) (**Figure S4**). We analyzed all breakpoint junction sequences to investigate the potential mechanism(s) that created RP17-SVs. No single

mechanism could account for the RP17-SVs, as a combination of (micro)homology-mediated repair and non-homologous end joining events were identified (**Table S7-S8, Figure S5, Supplementary Results**).

Consistent autosomal dominant retinitis pigmentosa phenotype for RP17-affected families

The SVs identified were fully penetrant in all families. Available clinical data are presented in **Table S9**. Twenty-four cases from seventeen pedigrees were evaluated retrospectively. There is significant correlation of phenotype across all genotypes, with relatively mild disease, decreased visual acuity, visual field constriction, nyctalopia, and slow progression consistent with adRP. Many cases have preserved central visual function and acuity until the 6th-7th decade. Foveal sparing and cystoid macular edema were a common finding in individuals with UK-SV2. Based on a small number of cases (n=2) UK-SV6 (with a triplicated SV) may be associated with an earlier age of onset and more severe phenotype (**Figure S6**).

Topologically associating domain structure and epigenetic landscape of the RP17 genomic region

All of the RP17-SVs lead to disruption of the genomic region spanning *YPEL2* to *LINC01476* (**Figure 2E**). SVs that interfere with genome structure can have distinct effects on gene regulation, depending on the type and extent of the SV and landscape of the genomic region.¹⁷ TADs are separated by boundaries, regions of low chromatin interaction that insulate the regulatory activities of neighboring TADs. The transcription factor CTCF (CCTC-binding factor) typically binds in these regions where it plays a pivotal role in the maintenance of boundaries. SVs can cause loss-of-function by disconnecting enhancers from their target genes; however, disruption of TAD structures and boundaries can also exert a gain-of-function effect. Deletions, for example, can lead to the fusing of two previously separated TADs (TAD-fusion), inversions can result in the exchange of regulatory material between TADs (TAD-shuffling), whereas duplications can give rise to the generation of novel domains, so called neo-TADs.^{14,15} In each case, SVs result in the generation of ectopic contacts of enhancers with the promoters of novel target genes resulting in aberrant gene activation. The human limb malformations caused by SVs that alter the CTCF-associated boundary of the *WNT6/IHH/EPHA4/PAX3* locus are a prominent example. The SVs result in ectopic interactions between *EPHA4* limb enhancers and the neighboring developmental genes that are normally insulated, driving ectopic expression in the limb.¹⁶ Similarly, the deletion of a CTCF site located between the *Xist* and *Tsix* TADs on the X-chromosome resulted in a novel domain by

fusion of the adjacent TADs (fused-TAD).²⁶ As a consequence, previously insulated enhancers activated genes in the adjacent TAD, leading to the dysregulation of these genes.

Hi-C data were not available for human retina, therefore we generated Hi-C maps of control human 3D ROs to obtain maps of the chromatin organization of our region of interest. Hi-C revealed a structured domain containing *YPEL2* (*YPEL2* TAD) flanked by less structured neighboring domains (**Figure 3A**). CTCF binding is present on both boundaries (**Figure 3B**) supporting the TAD structure at this locus. CTCF ChIA-PET data highlighted interactions between the CTCF binding sites at the 5' and the 3' boundary of the *YPEL2* TAD (**Figure S7B**).

Assay for transposase accessible chromatin using sequencing (ATAC)-seq data from human retina shows that the chromatin in the *YPEL2* TAD is accessible and H3K27Ac ChIP-seq data revealed that there are several active enhancers located within the *YPEL2* TAD that are expected to drive *YPEL2* expression in the retina (**Figure 3B**).²⁷ Importantly, the *YPEL2* TAD harbors two regions of active enhancers with binding sites for transcription factor (TFs) known to be required for photoreceptor function, including NRL, CRX, and OTX2 (**Figure 3B**). NRL is a TF that is preferentially expressed in rod photoreceptors. These TF binding sites correlated with H3K27Ac and ATAC-seq peaks in retina. The published GeneHancer dataset shows that these regulatory elements have interactions with the *YPEL2* promoter (**Figure S7C**).²⁸ Collectively, these analyses revealed that *YPEL2* is located within an active compartment that contains retinal-specific enhancers (**Figure 3C**).

Expression of *YPEL2* and *GDPD1*

Expression of *YPEL2* and *GDPD1* was assessed by qPCR in multiple healthy human tissues, including retina (**Figure S8**). *YPEL2* is ubiquitously expressed in the tissues studied, including retina, with highest relative expression in brain. Single cell retina RNA-seq datasets revealed *YPEL2* is expressed at higher levels in rod photoreceptor cells, which is the primary cell type affected in retinitis pigmentosa, compared to cone photoreceptors (**Figure S8**).²⁵ *GDPD1* is detected at low levels in all tissue types, with higher levels of expression in testis and the brain. These data support the hypothesis that *YPEL2* expression is regulated by retinal enhancers within the *YPEL2* TAD.

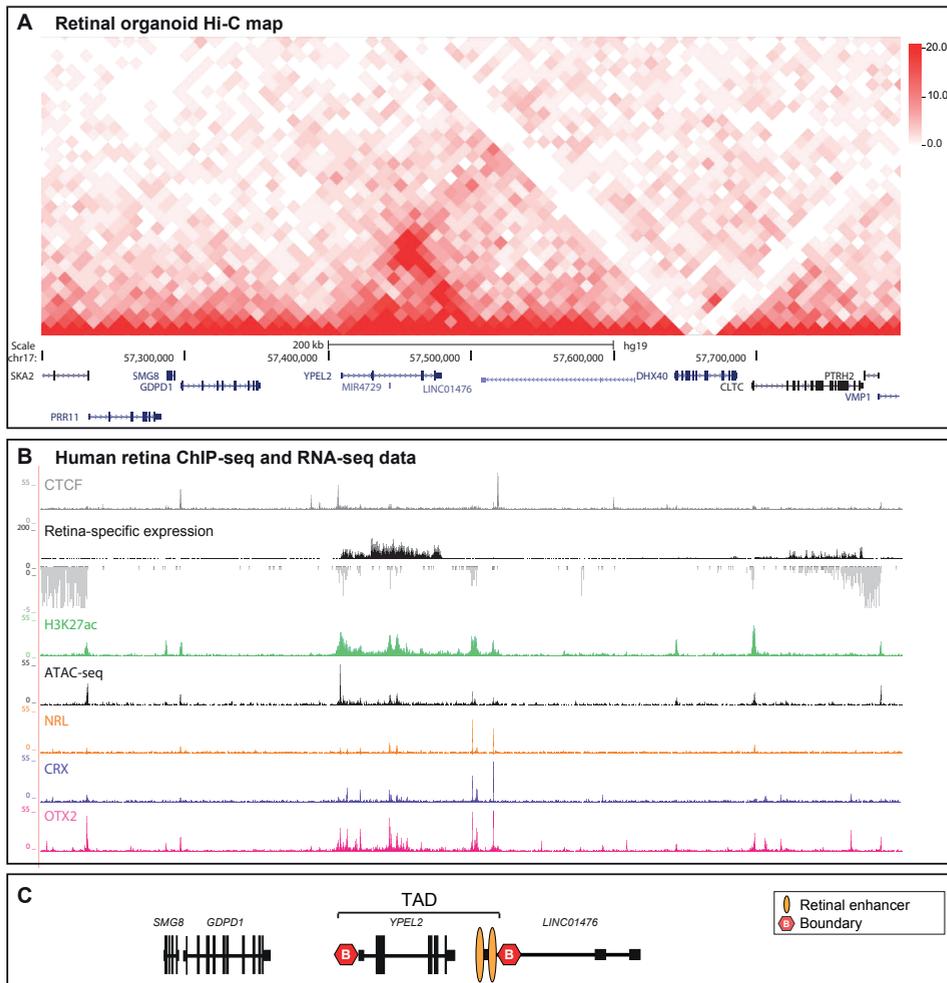


Figure 3. *YPEL2* is located within a structured active compartment that contains retinal-specific enhancers. (A) The TAD landscape of the genomic region disrupted by the RP17-SVs. Hi-C map of control retinal organoids revealed a structured domain containing *YPEL2* **(B)** *YPEL2* TAD boundaries correspond with CTCF sites identified in human retina. Analysis of RNA-seq and assay for transposase accessible chromatin using sequencing (ATAC)-seq data across the *YPEL2* region shows *YPEL2* retinal expression and an accessible chromatin configuration. Analysis of H3K27Ac ChIP-seq data in the same region revealed several active enhancers located within the *YPEL2* TAD, which are enriched for retinal transcription factor binding sites, including NRL, CRX, and OTX2 (Cherry et al. 2020).²⁷ These enhancers were located 5' of the CTCF boundary site within *LINC01476*. **(C)** Schematic representation of the *YPEL2* TAD structure.

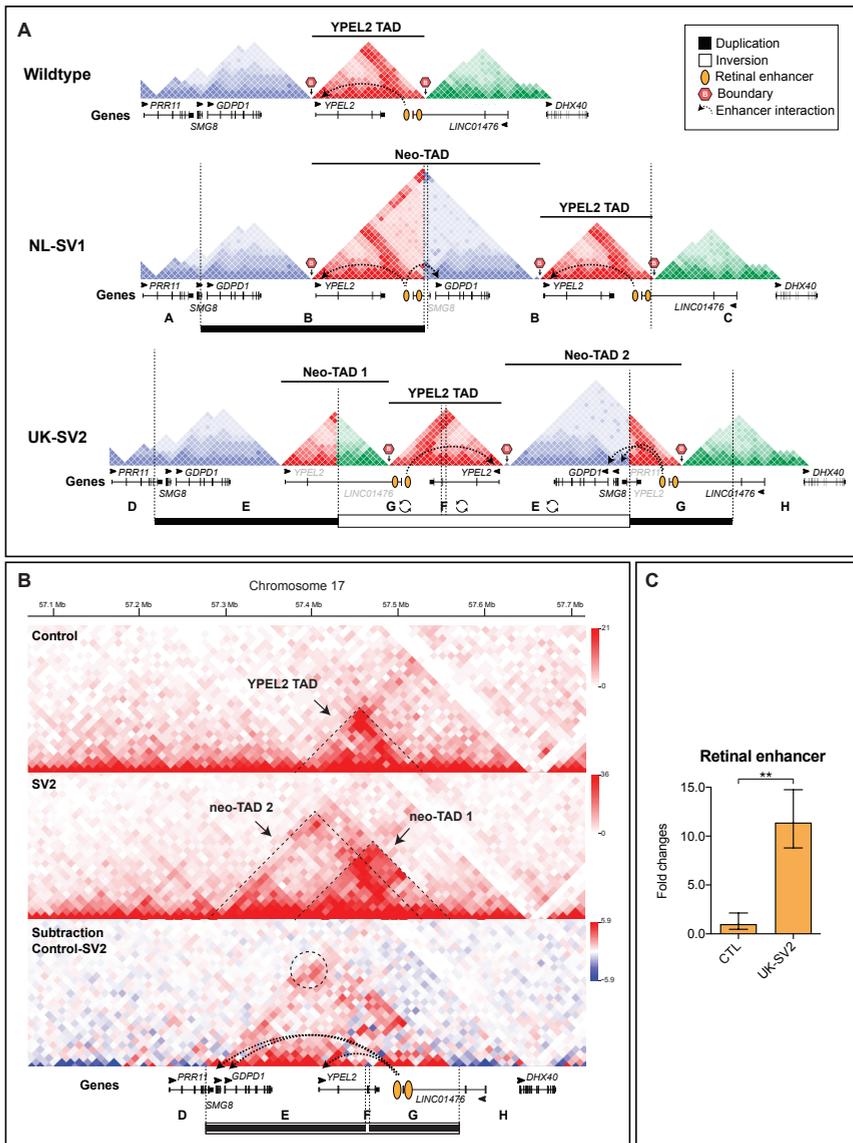


Figure 4. RP17-SVs create novel domains (neo-TADs) and hyper-activation of retinal enhancers. (A) Schematic modeling of the genome architecture spanning the RP17 region using Hi-C maps. The wildtype Hi-C map derived from neuronal tissue shows a TAD with CTCF boundaries containing YPEL2 and retinal enhancers, flanked by unstructured domains. TAD models of NL-SV1 and UK-SV2 (dotted vertical lines represent SV breakpoints) predict the formation of neo-TADs and ectopic interactions of the retinal enhancer with *GDPD1*. **(B)** Hi-C performed on retinal organoids (ROs) derived from control (top) and RP17 UK-SV2 individuals (bottom) (10 kb resolution; raw count map). The chromatin organization in control ROs shows the YPEL2 TAD (indicated by dashed lines). Two novel domains (neo-TAD 1 and 2) are visible in the UK-SV2 ROs, and neo-TAD 2 allows ectopic retinal enhancer contacts to *GDPD1* and *SMG8*. Dashed circle indicates the strong chromatin contact between retinal enhancers and the *GDPD1* promoter. **(C)** qPCR revealed significantly upregulated retinal enhancer RNA expression in UK-SV2 ROs compared to controls. (n=3; mean \pm standard error of the mean, **p<0.01).

RP17-SVs create new topologically associating domains and ectopic enhancer-gene interactions

Using the wildtype retinal organoid Hi-C map, we modelled the TAD boundaries, CTCF site orientation and retinal TF binding site positions for each unique RP17-SV (**Figure 4A, S9**). In NL-SV1 the duplication contains part of the *YPEL2* TAD, the boundary to the neighboring region and *GDPD1*. This results in the creation of a neo-TAD which now contains the previously separated *YPEL2* enhancers and the *GDPD1* gene in one domain. To directly investigate the effect of the SVs in retinal cells, dermal fibroblasts from UK-SV2 were reprogrammed to iPSC and differentiated to 3D ROs thus creating an *in vitro* model (**Supplementary Results**).

In this case, the duplicated regions are also inverted. Hi-C of RP17 ROs (UK-SV2) revealed the creation of two neo-TADs, compared to control ROs (**Figure 4B**). The rearrangement of CTCF sites caused by the SV creates boundaries for two novel domains (neo-TAD 1 and 2), where neo-TAD 2 contains a duplicated copy of *GDPD1* and *SMG8*, and the retinal enhancers, confirming the modelling for this SV (**Figure 4A**). Furthermore, based on our predictions, neo-TADs are created in each of the RP17 cases and *GDPD1* is predicted to gain ectopic access to the retinal-specific enhancers (**Figure 4A, S9**). Therefore, the potential convergent mechanism for retinal degeneration is transcriptional activation and expression of *GDPD1* through juxtaposition of retinal TF binding sites within active compartments, bounded by CTCF sites. This model would also fit with a dominant gain-of-function mechanism of disease.

Next, we assessed retinal enhancer expression in control and UK-SV2 ROs, by enhancer RNA qPCR (**Supplementary Results**). A significant increase of the retinal enhancer was detected in RP17 ROs (**Figure 4C**), demonstrating that this transcriptionally active retinal enhancer in the neo-TAD could drive retinal expression of *GDPD1*.

Differential expression of *GDPD1* in RP17 iPSC derived photoreceptor precursors and 3D retinal organoids

Our experimental data and modelling predict *GDPD1* enters a neo-TAD with retinal enhancers in all RP17-SVs. An extra copy of *YPEL2* enters the neo-TAD of NL-SV1, and *SMG8* enters this domain in UK-SV2 (**Figure 5, S9**).

To experimentally validate the consequence of RP17-SVs in genomic and cellular context, qPCR was performed to assess differential expression in PPCs (NL-SV1) and ROs (UK-SV2). The expression of *GDPD1*, *YPEL2* and *SMG8*, was compared to controls (**Supplementary Results**).

In both experimental models, the expression levels of *GDPD1* were significantly increased compared to controls. *YPEL2* was increased in NL-SV1 only (**Figure 5C**), whereas *SMG8* was increased in UK-SV2 (**Figure 5E**), which correlates with our TAD modelling and Hi-C experimental data for UK-SV2 ROs (**Figure 5B, 5D, 59**). To further explore the tissue specific effect of this transcriptional upregulation, we performed the same qPCR assays on fibroblasts of the same individuals. None of these genes had increased expression levels in affected individuals compared to controls (data not shown).

DISCUSSION

Previous genetic studies of adRP families mapping to the RP17-locus have implicated missense variants in the *CA4* gene as the cause of disease, or have been unable to confirm pathogenicity.^{9,10} Here, we describe the discovery of SVs as the cause of adRP at the RP17-locus in a large number of families, suggesting this is a previously unrecognized major locus for adRP. Our results show how complex rearrangements can result in the disruption of 3D genome architecture, the re-wiring of enhancer-promoter interactions and consequent gene misexpression.

Following identification of SVs in NL1 and UK1 using short-read WGS, our search for similar complex SVs in the RP17 genomic interval of genetically unexplained adRP families identified six other complex SVs that segregated with disease. SVs are a major source of normal variation in the human genome and are often benign,^{29,30} however, none of the RP17-SVs are found in the population database gnomAD³¹ or the Database of Genomic Variants (DGV).³² Although overlapping canonical SVs (deletions and duplications) have been identified, they do not have breakpoints within the *YPEL2-LINC01476* region, as observed for all RP17-SVs reported in this study. This is in line with observations that different SVs can have different consequences, depending on the characteristics of specific SVs in local 3D chromatin and epigenetic context.^{33,34}

Base level resolution of breakpoint junctions and interrogation of the DNA sequence signatures revealed the mechanisms of the chromosomal rearrangements. Repetitive elements are key factors in facilitating unequal crossover of genomic segments or to provide microhomology that induces fork stalling and subsequent template switching.^{35,36} Consistent with this model, repetitive elements were present in the flanking sequences of breakpoint junctions. In addition, microhomology, larger stretches of homology and small insertions-deletions were found at all breakpoints.³⁶⁻³⁸ Therefore, repetitive elements may explain why the RP17-locus is prone to such structural variation, which is

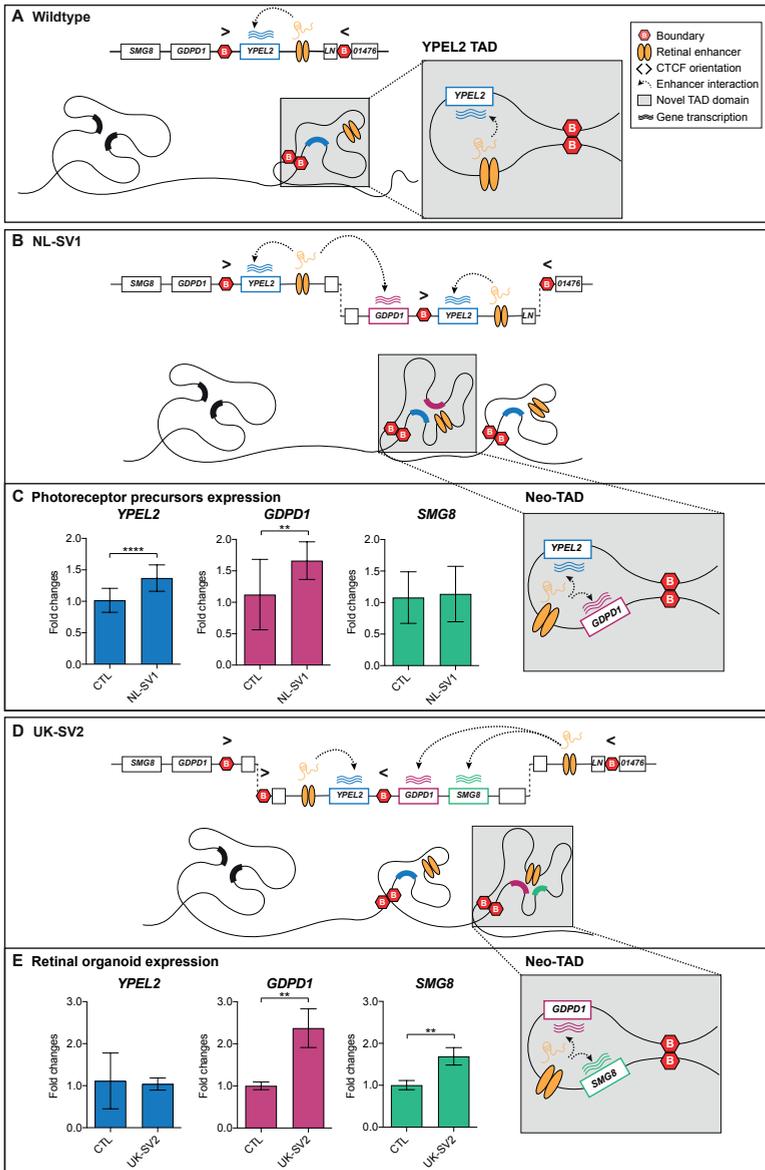


Figure 5. Convergent mechanism of ectopic retinal enhancer-*GDPD1* interaction caused by RP17 SVs. (A) In wildtype genomic context, *YPEL2* expression in retina is driven by retinal enhancers in a TAD with CTCF boundaries. Neighboring genes are insulated from retinal enhancer activation. **(B)** The NL-SV1 duplication creates a neo-TAD with a full-length copy of *YPEL2*, *GDPD1* and the retinal enhancers. This enables retinal-specific enhancers to ectopically interact with *GDPD1*, that drives its misexpression **(C)** qPCR analysis of photoreceptor precursor cells (PPCs) revealed a significant upregulation of *GDPD1* in NL-SV1 PPCs compared to controls. **(D)** The UK-SV2 duplication and inversion creates a neo-TAD with a full-length copy of *GDPD1* and *SMG8* and the retinal enhancers bounded by CTCF sites **(E)** qPCR analysis ROs revealed a significant upregulation of *GDPD1* expression in UK-SV2 ROs compared to controls. (n=3 independent ROs, mean ± standard error of the mean, **p≤0.01, ****p≤0.0001).

supported by the presence of breakpoint “hotspots”, as seen in *LINC01476* intron 2 and *YPEL2* intron 4, with some breakpoints only differing by a small number of base pairs (e.g. for UK-SV2 and UK-SV7).

None of the genes implicated in the RP17-SVs have been previously associated with retinal disease. *YPEL2* is expressed in retina, and single cell RNA sequencing of human and primate retina revealed expression in photoreceptors, with highest expression in rod photoreceptors.^{24,25} Although the function of *YPEL2* in the retina is unknown, we show that retinal expression is controlled by a number of retinal TF binding sites, including *NRL* which is predominantly expressed in rod photoreceptors. Furthermore, Hi-C data show that *YPEL2* and the retinal enhancer binding sites are insulated from the surrounding region in a structured *YPEL2* TAD in control ROs and other tissues.

Hi-C analyses of UK-SV2 ROs revealed the generation of new chromatin domains (neo-TADs), with altered structure and repositioning of the boundaries enabling *GDPD1* promoter-retinal enhancer contacts and consequent *GDPD1* misexpression in the retina. The molecular disease mechanism in these cases is similar to the reported duplications at the *SOX9/KCNJ2* locus.¹⁵ As described for the rearrangements reported here, the duplications at the *SOX9* locus also encompass a regulatory domain (of *SOX9*), a boundary (between the *SOX9* and the *KCNJ2* TADs) and the neighboring gene (*KCNJ2*). This results in the formation of a novel chromatin domain (neo-TAD) containing the *SOX9* regulatory elements and the new target gene (*KCNJ2*) that are now free to interact. In the *SOX9* case, this leads to misexpression of *KCNJ2* in a *SOX9* pattern and consecutive limb malformation, whereas in the RP cases the interaction of *GDPD1* with *YPEL2* enhancers leads to misexpression in the retina. However, in some of the RP cases, such as UK-SV2, the situation is more complex because the duplications are inverted. Inversions can lead to the exchange of regulatory material from one end of the breakpoint to the other (also called TAD-shuffling).¹⁴ In UK-SV2 the duplication creates two neo-TADs but the content is reorganized by the inversion. Again, the *GDPD1* gene and retinal enhancers are brought together in one new TAD. Thus, the pathogenetic principle remains the same, as all the RP17 SVs are predicted to create new TADs allowing access of the retinal enhancers to *GDPD1*. This suggests that increased expression of *GDPD1* in photoreceptors is the convergent mechanism of disease. Consistent with this hypothesis, PPCs from NL-SV1 and ROs from UK-SV2 showed significant increased expression of *GDPD1* in RP17 families with different SVs compared to controls. In UK-SV2 ROs, an increased expression of *SMG8* was observed, which is also introduced into the active neo-TAD of UK-SV2. Conversely, *YPEL2* shows upregulation in NL-SV1, which is in line with the complete duplication of

YPEL2 in NL-SV1. Importantly, qPCR provided evidence for the increased expression of the retinal enhancer in UK-SV2 ROs, with TF binding sites for NRL which is preferentially expressed in rod photoreceptors, the primary cell type affected in RP.

Although increased expression of *SMG8*, *YPEL2*, or the retinal enhancer cannot be excluded from contributing to the phenotype in individual families, these experimental data support the hypothesis of a convergent mechanism of *GDPD1* entry into the active neo-TAD with retinal enhancers for all eight complex RP17-SVs. This is further supported by the observation that the two affected individuals in family UK13 with an earlier age of onset and more severe phenotype, compared to all other families, have a triplication (UK-SV6) where two copies of *GDPD1* are predicted to enter the active neo-TAD.

Our data implicate increased retinal expression of *GDPD1* as a dominant gain-of-function mechanism leading to adRP. *GDPD1* encodes a glycerophosphodiesterase, that can hydrolyze lysophosphatidylcholine (lyso-PC) to lysophosphatidic acid (LPA)³⁹ with lysophospholipase D (lysoPLD) activity on various lysophospholipids.⁴⁰ *GDPD1* is detected at low expression in the healthy retina, therefore, increased expression of *GDPD1* could lead to dysregulation of lipid metabolism, which is known to be critical for photoreceptor function although the exact mechanisms of photoreceptor cell death are not known.^{41,42,43} Disruption of lipid metabolism leading to adRP, combined with the adult age of onset, opens avenues for therapeutic intervention to preserve vision by restoring lipid homeostasis.

ACKNOWLEDGEMENTS

We thank the affected individuals and their families for participating in this study and our funding bodies. We thank Saskia D. van de Velde-Visser for cell culture work, Marijke N. Zonneveld-Vrieling, Mubeen Khan, Rolph Pfundt, M. Imran Khan, for experimental contributions, Alex Hoischen for expert contribution, Galuh Dyah Nur Astuti for bioinformatics support, Edward Bloch for taking skin biopsies and Karin W. Littink, Suzanne IJzer, Irma Lopez, Anneke I. den Hollander and Samantha Malka for their effort in sample collection. This research was supported by grants from Retina UK, Fight for Sight (UK), Moorfields Eye Charity (UK), the National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust, UCL Institute of Ophthalmology, GOSH and Cambridge (UK), and The Wellcome Trust. The study was also financially supported by DCN Radboudumc grant (to FPMC and HK), the Foundation Fighting Blindness (PPA-0517-0717-RAD, to SR and FPMC), and the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, the

Stichting tot Verbetering van het Lot der Blinden and the Stichting Blinden-Penning (to SR and FPMC), Foundation Fighting Blindness, Réseau de Vision and MCH Foundation (to RKK), Swiss National Science Foundation (176097, to CR), the MRC (UK, to SM), Retina South Africa and the South African Medical Research Council (MRC, to LR). The Steunfonds RP17 financially supported experiments for this project (to SR and FPMC). This research accessed data generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by affected individuals and collected by the National Health Service as part of their care and support.

CONSORTIA

UK Inherited Retinal Dystrophy Consortium: A.J. Hardcastle, M.E. Cheetham, M. Michaelides, A.R. Webster, N. Pontikos, A. Fiorentino, G. Arno, C.F. Inglehearn, C. Toomes, M. Ali, M. McKibbin, C.E.L. Smith, S. Downes, J. Yu, S. Halford, S. Broadgate, G.C. Black, R.L. Taylor, P.I. Sergouniotis.

Genomics England Research Consortium: J.C. Ambrose, P. Arumugan, E.L. Baple, M. Bleda, F. Boardman-Pretty, J.M. Boissiere, C.R. Boustred, H. Brittain, M.J. Caulfield, G.C. Chan, C.E.H. Craig, L.C. Daugherty, A. de Burca, A. Devereau, G. Elgar, R.E. Foulger, T. Fowler, P. Furió-Tarí, J.M. Hackett, D. Halai, A. Hamblin, S. Henderson, J.E. Holman, T.J.P. Hubbard, K. Ibanez Garikano, R. Jackson, L.J. Jones, D. Kasperaviciute, M. Kayikci, L. Lahnstein, K. Lawson, S.E.A. Leigh, I.U.S. Leong, F.J. Lopez, F. Maleady-Crowe, J. Mason, E.M. McDonagh, L. Moutsianas, M. Mueller, N. Murugaesu, A.C. Need, C.A. Odhams, C. Patch, D. Perez-Gil, D. Polychronopoulos, J. Pullinger, T. Rahim, A. Rendon, P. Riesgo-Ferreiro, T. Rogers, M. Ryten, K. Savage, K. Sawant, R.H. Scott, A. Siddiq, A. Sieghart, D. Smedley, K.R. Smith, A. Sosinsky, W. Spooner, H.E. Stevens, A. Stuckey, R. Sultana, E.R.A. Thomas, S.R. Thompson, C. Tregidgo, A. Tucci, E. Walsh, S.A. Watters, M.J. Welland, E. Williams, K. Witkowska, S.M. Wood, M. Zarowiecki.

REFERENCES

1. Haer-Wigman, L., van Zelst-Stams, W.A., Pfundt, R., van den Born, L.I., Klaver, C.C., Verheij, J.B. *et al.* Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *European Journal of Human Genetics* **25**, 591-599 (2017).
2. Verbakel, S.K., van Huet, R.A.C., Boon, C.J.F., den Hollander, A.I., Collin, R.W.J., Klaver, C.C.W. *et al.* Non-syndromic retinitis pigmentosa. *Progress in Retinal and Eye Research* **66**, 157-186 (2018).
3. Hamel, C. Retinitis pigmentosa. *Orphanet Journal of Rare Diseases* **1**, 40-40 (2006).
4. Fahim, A., Daiger, S. & Weleber, R. Nonsyndromic retinitis pigmentosa overview. *GeneReviews* (2017).
5. Snoeckx, R.L., Kremer, H., Ensink, R.J.H., Flothmann, K., de Brouwer, A., Smith, R.J.H. *et al.* A novel locus for autosomal dominant non-syndromic hearing loss, DFNA31, maps to chromosome 6p21.3. *Journal of Medical Genetics* **41**, 11-13 (2004).
6. Sullivan, L.S., Bowne, S.J., Reeves, M.J., Blain, D., Goetz, K., Ndifor, V. *et al.* Prevalence of mutations in eyeGENE probands with a diagnosis of autosomal dominant retinitis pigmentosa. *Investigative Ophthalmology & Visual Science* **54**, 6255-6261 (2013).
7. Daiger, S.P., Bowne, S.J. & Sullivan, L.S. Genes and mutations causing autosomal dominant retinitis pigmentosa. *Cold Spring Harbor Perspectives in Medicine* **5**, a017129 (2014).
8. Bardenb, S., Ebenezer, N., Greenberg, J., Inglehearn, C.F., Bartmann, L., Goliath, R. *et al.* An eighth locus for autosomal dominant retinitis pigmentosa is linked to chromosome 17q. *Human Molecular Genetics* **4**, 1459-1462 (1995).
9. den Hollander, A.I., van der Velde-Visser, S.D., Pinckers, A.J.L.G., Hoyng, C.B., Brunner, H.G. & Cremers, F.P.M. Refined mapping of the gene for autosomal dominant retinitis pigmentosa (RP17) on chromosome 17q22. *Human Genetics* **104**, 73-76 (1999).
10. Rebello, G., Ramesar, R., Vorster, A., Roberts, L., Ehrenreich, L., Oppon, E. *et al.* Apoptosis-inducing signal sequence mutation in carbonic anhydrase IV identified in patients with the RP17 form of retinitis pigmentosa. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6617-6622 (2004).
11. Köhn, L., Burstedt, M.S.I., Jonsson, F., Kadzhaev, K., Haamer, E., Sandgren, O. *et al.* Carrier of R14W in carbonic anhydrase IV presents bothnia dystrophy phenotype caused by two allelic mutations in RLBP1. *Investigative Ophthalmology & Visual Science* **49**, 3172-3177 (2008).
12. Golovleva, I., Köhn, L., Burstedt, M., Daiger, S. & Sandgren, O. Mutation spectra in autosomal dominant and recessive retinitis pigmentosa in northern Sweden. *Advances in experimental medicine and biology* **664**, 255-262 (2010).
13. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
14. Spielmann, M., Lupiáñez, D.G. & Mundlos, S. Structural variation in the 3D genome. *Nature Reviews Genetics* **19**, 453-467 (2018).

15. Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269 (2016).
16. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
17. Ibrahim, D.M. & Mundlos, S. Three-dimensional chromatin in disease: what holds us together and what drives us apart? *Current Opinion in Cell Biology* **64**, 1-9 (2020).
18. Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N.M., Hoynig, C.B. *et al.* Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. *American Journal of Human Genetics* **102**, 517-527 (2018).
19. Sangermano, R., Bax, N.M., Bauwens, M., van den Born, L.I., De Baere, E., Garanto, A. *et al.* Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461-10T C mutation in Stargardt disease. *Ophthalmology* **123**, 1375-1385 (2016).
20. Schwarz, N., Lane, A., Jovanovic, K., Parfitt, D.A., Aguila, M., Thompson, C.L. *et al.* Arl3 and RP2 regulate the trafficking of ciliary tip kinesins. *Human Molecular Genetics* **26**, 3451-3451 (2017).
21. Díaz, N., Kruse, K., Erdmann, T., Staiger, A.M., Ott, G., Lenz, G. *et al.* Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. *Nature Communications* **9**, 4938-4938 (2018).
22. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* **3**, 95-98 (2016).
23. Melo, U., Schöpflin, R., Acuna-Hidalgo, R., Mensah, M.A., Fischer-Zirnsak, B., Holtgrewe, M. *et al.* Hi-C identifies complex genomic rearrangements and TAD-shuffling in developmental diseases. *American Journal of Human Genetics* **106**, 872-884 (2020).
24. Lukowski, S.W., Lo, C.Y., Sharov, A.A., Nguyen, Q., Fang, L., Hung, S.S. *et al.* A single-cell transcriptome atlas of the adult human retina. *The EMBO Journal* **38**, e100811 (2019).
25. Peng, Y.-R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G.S. *et al.* Molecular classification and comparative taxonomics of foveal and peripheral cells in primate retina. *Cell* **176**, 1222-1237 (2019).
26. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
27. Cherry, T.J., Yang, M.G., Harmin, D.A., Tao, P., Timms, A.E., Bauwens, M. *et al.* Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. *Proceedings of the National Academy of Sciences* **117**, 9001-9012 (2020).
28. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database: The Journal of Biological Databases and Curation* **2017**, bax028 (2017).
29. Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C. *et al.* Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nature Genetics* **42**, 385-391 (2010).

30. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81 (2015).
31. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
32. MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L. & Scherer, S.W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research* **42**, D986-D992 (2014).
33. Lohan, S., Spielmann, M., Doelken, S.C., Flöttmann, R., Muhammad, F., Baig, S.M. *et al.* Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clinical Genetics* **86**, 318-325 (2014).
34. Ngcungcu, T., Oti, M., Sitek, J.C., Haukanes, B.I., Linghu, B., Bruccoleri, R. *et al.* Duplicated enhancer region increases expression of CTSB and segregates with keratolytic winter erythema in South African and Norwegian families. *American Journal of Human Genetics* **100**, 737-750 (2017).
35. Gu, S., Yuan, B., Campbell, I.M., Beck, C.R., Carvalho, C.M.B., Nagamani, S.C.S. *et al.* Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. *Human Molecular Genetics* **24**, 4061-4077 (2015).
36. Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D. & Lupski, J.R. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nature Genetics* **41**, 849 (2009).
37. Lupski, J.R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in Genetics* **14**, 417-422 (1998).
38. Lieber, M.R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry* **79**, 181-211 (2010).
39. Ohshima, N., Kudo, T., Yamashita, Y., Mariggìò, S., Araki, M., Honda, A. *et al.* New members of the mammalian glycerophosphodiester phosphodiesterase family: GDE4 and GDE7 produce lysophosphatidic acid by lysophospholipase D activity. *Journal of Biological Chemistry* **290**, 4260-4271 (2015).
40. Tsuboi, K., Okamoto, Y., Rahman, I.A.S., Uyama, T., Inoue, T., Tokumura, A. *et al.* Glycerophosphodiesterase GDE4 as a novel lysophospholipase D: a possible involvement in bioactive N-acylethanolamine biosynthesis. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1851**, 537-548 (2015).
41. Friedman, J.S., Chang, B., Krauth, D.S., Lopez, I., Waseem, N.H., Hurd, R.E. *et al.* Loss of lysophosphatidylcholine acyltransferase 1 leads to photoreceptor degeneration in rd11 mice. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15523-15528 (2010).
42. Kmoch, S., Majewski, J., Ramamurthy, V., Cao, S., Fahiminiya, S., Ren, H. *et al.* Mutations in PNPLA6 are linked to photoreceptor degeneration and various forms of childhood blindness. *Nature Communications* **6**, 5614-5614 (2015).

43. Lidgerwood, G.E., Morris, A.J., Conquest, A., Daniszewski, M., Rooney, L.A., Lim, S.Y. *et al.* Role of lysophosphatidic acid in the retinal pigment epithelium and photoreceptors. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1863**, 750-761 (2018).

SUPPLEMENTARY MATERIALS AND METHODS

SNP genotyping

The RP17-locus was previously established using polymorphic markers selected from the Généthon genetic map, that were genotyped in 23 individuals from index family NL1.¹ Subsequently, we collected DNA from 27 individuals (18 affected and 9 unaffected subjects) from the fourth generation of the family. SNP-genotyping was performed on these 27 DNA samples from generation four, and for 36 individuals (17 affected and 10 unaffected subjects and 9 spouses) from the second and third generation using the HumanCore-24V.1.0 array (Illumina). The RP17-locus was further refined by determining phase in a two-parent-sib dataset.

Exome and Genome sequencing

Index family NL1; Whole exome sequencing (WES) was performed for three affected individuals from different branches of the family. Exome enrichment was performed using the Aligent SureSelect Human All Exome V5 kit following manufacturer's instructions. Subsequently, WES was executed on an Illumina HiSeq2000TM system by BGI Europe (Copenhagen, Denmark). BWA V.0.78² and GATK HaplotypeCaller V.3.3³ were used for read mapping along the hg19 reference genome (GRCh37/hg19) and variant calling, respectively. Variants were annotated using an in-house developed pipeline.

WGS was performed by BGI (Hongkong, China) on a BGISeq500 using a 2x 100 bp paired end module, with a minimal median coverage per genome of 30-fold. SVs were called using Manta Structural Variant Caller V.1.1.0 (Illumina; paired end and split read evidence for SVs) and copy number variants (CNVs) using Control-FREEC (detection of copy number changes and allelic imbalances based on read depth).⁴ Variants were validated and visualized using the IGV software (V.2.4).⁵ Shared single nucleotide variants (SNVs) or SVs located in or spanning the refined RP17-locus were assessed for putative pathogenicity. Variants were prioritized based on a minor allele frequency (MAF) ≤ 0.0001 in gnomAD.⁶

Index family UK1; WES was performed for one affected individual as previously described.⁷ WGS was subsequently performed for four affected individuals from distant branches of the family by Edinburgh Genomics using TruSeq Nano with a minimal median coverage of 30-fold per genome. Variants were assessed and filtered using the Variant Annotation and Filter Tool (VarAFT).⁸ Variants were prioritized based on a MAF

≤ 0.0001 in gnomAD. CNVs and SVs were analyzed from WES data using ExomeDepth⁹ and WGS data using Canvas Copy Number Variant Caller¹⁰ (Illumina; copy number gain or loss based on read depth) and Manta Structural Variant Caller.¹¹

For additional unsolved adRP families, or families suspected to harbor RP17-SVs, WES or WGS was performed. Families of Canadian (CA) or South African (SA) origin were analyzed in the Netherlands with additional families of Dutch origin. WGS was performed as described for NL1. For UK families, WGS was executed as described for UK1 or through the NIHR-Bioresource and Genomics England pipelines as previously described.^{7,12}

Characterization and validation of structural variants

Primer sequences and coordinates are listed in **Table S2** and PCR conditions for all breakpoint junctions are available upon request.

SV breakpoint regions were assessed for presence of microhomology and repetitive elements. Breakpoint regions and junctions were defined as 150 bp flanking sequence surrounding the breakpoint, which were used as input sequences for subsequent analyses. The presence of microhomology at the breakpoints was assessed using multiple sequence alignment between the junction fragment and the 5' and 3' breakpoint regions using Clustal Omega.¹³ The presence of repetitive elements at the breakpoint regions was assessed using RepeatMasker.¹⁴

To validate the presence of a triplicated region for UK-SV6, a quantitative real-time PCR (qPCR) experiment was performed on genomic DNA from affected individuals from family UK13 (n=2), and unaffected controls (n=2). qPCR was performed using SYBR Green labTAQ Green mix (labTAQ) on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). Primer pairs were designed to amplify genes in the suspected triplicated regions and distal and proximal regions on Chr17 outside the triplicated areas as a reference for standard quantity. Primer sequences and chromosomal positions are listed in **Table S3**.

Each reaction was run in triplicate and was comprised of 2x labTAQ Green mix (labTAQ), 0.8 μ l of each primer (10 mM) and 25 ng DNA in a final reaction volume of 20 μ l. Cycling conditions were as follows: 95°C for 2 min, followed by 40 cycles at 95°C for 15 s and 60°C for 20s. Dissociation curves were generated by heat denaturation over a temperature gradient from 60-95°C to ensure no primer-dimers had formed and to check for a single amplicon. To verify the presence of a single PCR product, samples were also electrophoresed on a 2% agarose gel. Data were obtained using the QuantStudio™ Real-Time PCR Software (Applied Biosystems) to generate an amplification plot and a

melting curve for each reaction. The fold difference of the target region was normalized to the wildtype reference genomic region with respect to the calibrator sample, and was calculated using the $\Delta\Delta C_t$ method.¹⁵

Interrogation of the genomic region

Available Hi-C, ChIP-seq and RNA-seq datasets were downloaded, analyzed and visualized using UCSC genome browser.¹⁶ Human retina ChIP-seq and RNA-seq datasets were obtained from Cherry et al. 2020.¹⁷ CTCF ChIP-seq datasets for GM12878 and K562 were retrieved from the ENCODE project/Broad Institute¹⁸ and for MCF-7 from the ENCODE project/University of Washington.¹⁹ CTCF ChiA-PET libraries for K562 and MCF-7 (GSM970215) were obtained from the ENCODE/GIS-Ruan dataset.²⁰

Reprogramming fibroblasts into iPSCs and differentiation into photoreceptor progenitor cells and 3D retinal organoids

Fibroblasts were cultured from skin biopsies of individuals with NL-SV1, UK-SV2, and anonymous control individuals. Cell lines were reprogrammed into iPSCs and differentiated in PPCs (NL-SV1) or ROs (UK-SV2).

For NL-SV1, fibroblasts of two affected and four anonymous control individuals were reprogrammed into iPSCs. Reprogramming into iPSCs was performed by lentiviral transduction as previously described²¹, for one control cell line, reprogramming was performed using episomal vectors (Addgene).²² iPSC lines for each affected and control individual were then differentiated into PPCs following the previously described 60-day protocol.^{21,23} For each iPSC line, differentiation was performed for two iPSC clonal lines in triplicate. Differentiation of PPCs was confirmed by RT-qPCR for neural (*PAX6*) and photoreceptor progenitor (*CRX* and *NRL*) markers (data not shown).

For one affected individual with UK-SV2 and one control individual, fibroblasts were reprogrammed into iPSCs using episomal vectors (Addgene), as described previously.²² Retinal organoids were differentiated from iPSC, following a previously described protocol with slight modifications.²⁴ iPSCs were seeded on plates coated with Geltrex (ThermoFisher Scientific) until neuronal retinal vesicles (NRVs) appeared. NRVs were excised by a sterile scalpel and distributed in single wells in 25 wells low-attachment plates. NRVs were then cultured in Retinal differentiation media; 3:1 v/v of DMEM:F12, 2% B27 supplement, 1% Non-Essential Amino Acid, 1% Penicillin-Streptomycin (Gibco) for one week. Optic vesicles were then cultured in Retinal Maturation Medium 1 (3:1 v/v of DMEM:F12, 2% B27 supplement, 1% Non-Essential Amino Acid, 1% Penicillin-Streptomycin, 10% Fetal Bovine Serum (Labtech), 100 μ M Taurine, 2 mM GlutaMAX) until day 70, then changed to Retinal Maturation Medium 2 (3:1 v/v of DMEM:F12,

1% N2 supplement, 1% Non-Essential Amino Acid, 1% Penicillin-Streptomycin, 10% Fetal Bovine Serum (Labtech), 100 μ M Taurine, 2 mM GlutaMAX) until maturation and collection of the ROs for experimental procedures. Media was supplemented with 1 μ M retinoic acid from day 50 to day 70, then changed to 0.5 μ M from day 70 to day 100. After day 100 no further supplement was added to the media.

Preparation of low input Hi-C libraries (Low-C)

Four UK-SV2 and four control 200-day old ROs were harvested and dissociated to single cells by gentle trituration in 150 μ L PBS. Total volume was brought up to 500 μ L with PBS before fixation with 2% PFA/PBS for 10 min while tumbling. Next, 100 μ L of 1.425 M glycine were added and incubated in rotation for 5 min. To quench the cross-linking reaction, cells were placed on ice for 10 min. Then, cells were centrifuged for 8 min at 500 g and 4°C, and supernatant was removed. The pellet was resuspended in cold lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1.15 Triton X-100, 5% Protease inhibitor cocktail) and incubated for 15 min on ice. Cells were centrifuged for 5 min at 500 g and 4°C, and the supernatant was discarded. Finally, lysed cells were washed in 500 μ L PBS and centrifuged for 2 min at 500 g and 4°C. Cells were snap frozen in liquid N2 before restriction enzyme digestion. Next, RO fixed chromatin (2×10^5 cells) from UK-SV2 and controls was digested for 2h at 37 °C with a 4bp cutter (*DpnII*; New England Biolabs - NEB). The DNA overhangs generated by the restriction enzyme were marked with biotin-14-dATP (Thermo Fischer Scientific) and the proximity ligation step was performed for 4h at 18°C using T4 DNA ligase (NEB). Crosslink reversal was performed overnight at 65°C with vigorous shaking (1,000 rpm). The DNA was precipitated by adding Phenol-Chloroform-Isoamyl alcohol mix (25:24:1) (Merck) and then sheared to fragments of 300-600 bp using Covaris S220 (2 cycles, each 50sec long; 10% duty; 4 intensity; 200 cycles/burst). The biotin-filled DNA fragments were pulled down using Dynabeads MyOne Streptavidin T1 beads (Thermo Fischer Scientific) and the products were prepared for Illumina short-reads sequencing using the NEBNext Ultra DNA Library Prep kit (NEB).

Quantitative real time PCR of genes and enhancer RNA within the RP17-locus

Expression of genes located in the RP17-locus was assessed using RT-qPCR in human tissues, affected individual and control PPCs and ROs. Commercially available RNA panels were used to determine the expression of *GDPD1* and *YPEL2* in healthy human adult tissues. RNA isolation and cDNA preparation were performed as previously described.²⁵ Single cell RNA sequencing data of human²⁶ and primate retinal cell types²⁷ was obtained and visualized using the Broad Institute Single Cell Portal.

For the PPCs, total RNA was extracted using a Nucleospin RNA kit (Machery-Nagel) and cDNA was synthesized using an iScript cDNA synthesis kit (Bio-Rad). qPCR analysis was performed using GoTaq qPCR Master Mix (Promega) following manufacturer's instructions. 100 day old ROs were harvested and RNA was extracted using RNeasy Mini Kits (Qiagen). cDNA was synthesized using Tetro cDNA Synthesis kits (Bioline) and qPCR analysis was performed using the SYBR Green labTAQ Green mix (labTAQ) following manufacturer's instructions on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems).

Primers were designed to assess differential expression of genes implicated in the SVs, and control reference genes and retinal progenitor genes (**Table S3**). Primers to detect retinal enhancer expression were designed based on observed transcriptional activity of the enhancer RNA in the FANTOM5 Cap Analysis of Gene Expression (CAGE) human dataset (**Table S3**).²⁸ Relative gene expression levels, compared to the reference genes *GUSB* and *ACTB*, were determined with the $\Delta\Delta C_t$ method.¹⁵ Statistical analyses were performed using an unpaired Student t-test to test for significance between groups.

SUPPLEMENTARY RESULTS

Refinement of the RP17-locus in two unrelated adRP families

Index family NL1; In total, 35 affected and 28 unaffected individuals were included. Assuming complete penetrance of the phenotype, a refined locus of 5.16 Mb was identified; chr17:g.55,112,092-60,271,924 (rs8078110-rs9910672) (**Figure 1D**), with a maximum LOD-score of 15.0. Next, WES was performed in three affected family members from different branches of the family. No rare coding or splice site heterozygous variants ($MAF \leq 0.0001$) located within the defined locus were identified that were shared by all three individuals. In addition, no rare shared heterozygous variants were found in IRD-associated genes (RetNet). Subsequently, WGS was performed in three additional affected individuals. Shared variants within the locus between the three affected individuals were prioritized based on population frequency ($MAF \leq 0.0001$), and coding, splice site, intronic and intergenic heterozygous variants were assessed (**Table S4**).

Index family UK1; WES was performed for an affected individual from a genetically unexplained UK adRP family (UK1). No rare coding or splice site heterozygous variants ($MAF \leq 0.0001$) were identified in IRD-associated genes, so WGS was performed for four affected individuals (**Figure 1B**). Prioritization of rare heterozygous variants in genome data shared by affected individuals in this family failed to identify a candidate rare shared heterozygous variant in IRD-associated genes ($MAF \leq 0.0001$); however, a disease

associated haplotype on chromosome 17 spanning 8 Mb (17q22-17q24.1) was identified (**Figure 1E**). No shared rare (MAF ≤ 0.0001) coding or splice-site variants were identified within the haplotype (**Table S5**). A deep intronic shared rare (absent from gnomAD) variant (g.56293716G>A; c.262-112C>T; NM_001321269.1), in the ciliopathy gene *MKS1*, was initially considered a candidate. This variant was assessed for its potential to alter splicing using lymphoblast RNA extracted from affected individuals and controls; however, no difference in pre-mRNA splicing was observed (data not shown). This rare variant was used as a flag SNV to detect this haplotype in other families. Twelve UK adRP families were found to carry the same founder haplotype (**Figure 1B** and **Figure 1C**). We then refined the adRP locus, by genotyping SNPs in the extended pedigrees, to a 4.4 Mb interval on Chr17q22 (chr17:55,139,138-59,536,883) (**Figure 1E**).

Identification of structural variants within the RP17-locus

We analyzed the genome and exome data for CNVs and SVs using Manta, Control-FREEC, Canvas and ExomeDepth. In all families, SVs within the RP17-locus were identified. Triplication was suspected from read depth of SNVs observed in IGV for UK-SV6. To validate the presence of a triplicated region, qPCR was performed on affected and control genomic DNA for genes and genomic regions implicated in this SV, and proximal and distal genomic regions (as additional controls for copy number).

For all families that harbor SVs in the RP17-locus, reanalysis of sequencing data was performed to exclude other potentially pathogenic variants in IRD-associated genes. No pathogenic heterozygous coding or splice site variants were observed in genes that have been associated with IRDs (MAF ≤ 0.0001). NL2 consists of distantly related affected individuals, who were identified as having a common ancestor following the identification of the NL-SV5. In the middle branch of this pedigree, a plausible candidate variant in *ZNF513* was described previously.²⁹ This variant was absent in WGS data of the other two affected individuals of this family, and therefore does not segregate with disease and is no longer a candidate variant.

A combination of mutational mechanisms created the RP17-SVs

Different mutational mechanisms have been described for the formation of complex SVs in the genome; including replication-based mechanisms, such as microhomology-mediated break-induced replication.^{30,31} Therefore, we analyzed all breakpoint junction sequences to investigate the potential mechanism(s) that created RP17-SVs. Analysis of breakpoint sequences in the reference genome using the algorithm RepeatMasker identified an enrichment for long repetitive elements (e.g. *Alu*-elements) in all SVs

(**Table S6**). In addition, breakpoint sequences revealed several DNA signatures that are indicative of distinct underlying mechanisms. For some SVs (e.g. NL-SV1), microhomology (2-5bp) was identified at the breakpoints, whereas longer stretches of homology (>100bp) were identified for breakpoints of UK-SV6 and UK-SV7. In these cases, (micro)homology-mediated repair is the likely mechanism giving rise to the SV. For other breakpoints (e.g. UK-SV1), small insertions and deletions were observed at breakpoint junctions, suggesting non-homologous end joining events (**Table S8**, **Figure S4**). In all SVs, there is a high content of repetitive elements, suggesting these play a role in both repair mechanisms.

SUPPLEMENTARY TABLES

Table S1. Previously reported CA4 variants

Genome	cDNA	Protein	Ethnicity	gnomAD AF all	gnomAD AF subpopulation	CADD_PHRD	Detection method	References
g_58227429G>A	c.4C>T	p.Ala12Thr	Chinese	0.000004471	- (other EAS)	6.280	Targeted sequencing of CA4	Tian et al., 2010
g_58227435C>T	c.40C>T	p.Arg14Trp	South-African	0.0002410	0.0001368 (AFR)	15.94	Locus gene sequencing (RP17)	This study (SA1-4), Rebello et al., 2004; Yang et al., 2005
g_58234014G>A	c.206G>A	p.Arg69His	Chinese	0.00004374	0.0001087 (other EAS)	0.005	Targeted sequencing of CA4	Alvarez et al., 2007
g_58235718C>A	c.655C>A	p.Arg219Ser	Northern European	0.00003186	0.0001163 (NWE)	26.6	Targeted sequencing of CA4	Yang et al., 2005
g_58235763G>A	c.700G>A	p.Val234Ile	Spanish	0.01015	0.01757 (NWE)	9.468	Targeted sequencing of 12 adRP genes	De Sousa Dias et al., 2013

Overview of CA4 (NM_000717.4) variants reported in literature. A CADD_PHRD score of ≥ 15 and allele frequency ≤ 0.0001 are considered as pathogenicity criteria. Values that meet these criteria are indicated in red. The p.Arg14Trp variant was found in families SA1-4 that are included in this manuscript and carriers of SA-SV3. Genome, genomic position based on hg19; gnomAD AF all, allele frequency in gnomAD v2.1.1 database; gnomAD AF subpopulation, allele frequency in gnomAD v2.1.1 based on subpopulation corresponding to the ethnicity of the affected individual in which the variant was reported in literature; CADD_PHRD, Combined Annotation Dependent Depletion PHRED score; other EAS, other East Asian population; AFR, African population; NWE, Northwest European population.

Table S2. Primer sequences utilized to validate and characterize breakpoints

SV	Breakpoint	Coordinates	F primer (5'-3')	R primer (5'-3')	Amplicon size (bp)
NL-SV1	A-B	57,291,905	GCCTGGGTGACTAAGAAGACTCCATTCCC	CCACGGAGCACCTTAGCTCATTAAGTGC	720
	B-B	57,518,137-57,291,905	GGCCTAATGAACACAGAAAGACACTTGGC	CCACGGAGCACCTTAGCTCATTAAGTGC	839
	B-C	57,518,137	GGCCTAATGAACACAGAAAGACACTTGGC	TAGTCATAGTCCCTGATTCCTTAAAGCGG	831
UK-SV2	D-E	57,275,839	CATGACAAAACCTGTCTCC	CCTATCCAGTAAATGCCCTCTCC	881
	E-[G]	57,456,098-57,559,114	ATCAGGCAACGACACCCAT	AGAGTGTTAACAAAGTAGACTCGAT	1262
SA-SV3	[G]-[F]	57,468,960	GGAGCCTGAAGGAGTTGTCAAA	AATCCAACAATCTTCAGGGCA	999
	[F]-[E]	57,456,098	ATCAGGCAACGACACCCAT	TCTTCCACATGGGGACATAGG	894
	[E]-G	57,275,839-57,468,960	CCTATCCAGTAAATGCCCTCTCC	AATCCAACAATCTTCAGGGCA	1427
	G-H	57,559,114	AGAGTGTTAACAAAGTAGACTCGAT	ACTGGCCAAAGAAGACCCT	989
	I-J	57,247,615	GGGTGCAGTCAITTCATTCCT	TCTCTTGAGCCCAAGGAAATC	513
CA-SV4	J-K	57,391,678	TCATGTGAAATGCCACCTTC	GAGTGTAACGGCATGGTCTC	1530
	K-L	57,499,214	TTCTTTAAGGGGACCTTG	AAGCCAAGATCATCCAACC	694
	L-J	57,516,678-57,247,615	TGCCACTTCCATATGTGTG	TCTCTTGAGCCCAAGGAAATC	657
	J-[M]	57,391,678-57,612,711	TTATGAATCTGCCCAAGATCAC	AATGATTTGCCCTTGGCTTTC	1022
	[M]-L	[57,516,678]-57,499,214	GAATTTGCTGAAGGGCTTG	AAGCCAAGATCATCCAACC	486
	L-M	57,516,678	TGCCACTTCCATATGTGTG	GAATTTGCTGAAGGGCTTG	702
CA-SV4	M-N	57,612,711	AATGATTTGCCCTTGGCTTTC	CAATGCCAATCTGGACACC	823
	O-P	57,233,035	GAAAGCCAACCAATCACAC	AACAGGCCAGCTACTCAAG	368
	P-[R]	57,280,008-[57,634,900]	ATACAGGGAGACCCCGTTTC	CTGATCGAAGTGCAAAATGG	1801
	[R]-[Q]	57,483,883	CTACACAGGGACTGACACC	CAGACGAGCATTATCAACC	677
	[Q]-[P]	57,280,008	ATACAGGGAGACCCCGTTTC	AGATGATTTCTTGCTCTGTTGC	653
	[P]-R	[57,233,035]-57,483,883	CAGACGAGCATTATCAACC	AACAGGCCAGCTACTCAAG	493
	R-S	57,634,900	CTGATCGAAGTGCAAAATGG	TGGAGGGAAGTTATCTTGG	2299

Table S2. Continued

SV	Breakpoint	Coordinates	F primer (5'-3')	R primer (5'-3')	Amplicon size (bp)
NL-SV5	T-U	57,260,511	TTCAATCCACCCTCT	TCCATGGACTCCCTGAAACT	893
	U-U	57,515,862-57,260,511	TTCACCGCTGTTAAGAAAG	GAAGAGGAGACCCCAAAATG	648
UK-SV6	U-V	57,515,862	CCGCTGTTAAGAAAGGCTCT	CCCACTCAAGGAGCTTGA	971
	W-X	57,295,969	TAAGGGATCCAGGAACCTAAATG	AAAAATTCAGGGGTGG	767
	X-X	57,510,765-57,295,969	TGAGAGCTGGAGGCTAGT	AAAATTTGCCAGGGGTGG	1295
UK-SV7	X-Y	57,510,765	TGAGAGCTGGAGGCTAGT	AGCAACTGCAACTGAACTCCT	1013
	Z-AA	57,259,525	TC TTCGTATCTGTCTCAG	TGGGAGCTCAAGTGGACAAC	1097
	AA-[AC]	57,453,630-57,710,821	GCTGGACTCAGAGGGTGT	AAGCATCTAGGGCACATCCT	1554
	[AC]-[AB]	57,468,931	GGAGCCTGAAGGAGTTGTCAAA	AATCCAACACATCTTCAGGGCA	999
	[AB]-[AA]	57,453,630	GCTGGACTCAGAGGGTGT	AACAGTCATGGCTCACACTCA	1200
UK-SV8	[AA]-AC	57,259,525-57,468,931	TGGGACTCAAGTGGACAAC	AATCCAACACATCTTCAGGGCA	1504
	AC-AD	57,710,821	AAGCATCTAGGGCACATCCT	ACCCTATACTGAGGGACCTGC	990
	AE-AF	57,277,347	GCTGACACTTCCACCCC	CCAATGCAAAACCTGATACAGT	551
	AF-[AH]	57,326,234-57,631,659	TGGGATGTTCTGCTAAGGG	TGCCTAGTCCAATTTCTCAG	458
	[AH]-[AG]	57,413,153	CAGTGGTGATCTGCTCA	CACCAAGCATTTTCAGCAGC	478
	[AG]-[AF]	57,326,234	TGGGATGTTCTGCTAAGGG	TGTGCCAGCCCTTTTCATT	525
	[AF]-AH	57,277,347-57,413,153	CCAATGCAAAACCTGATACAGT	CACCAAGCATTTTCAGCAGC	569
AH-AI	57,631,659	TGCTGTAGTCCAATTTCTCAG	GTGGGAAAGGGTGTCTTAT	433	

SV Structural variant; Breakpoint, Breakpoints between genomic regions as illustrated in Figure 2; Coordinates, genomic positions of breakpoints according to hg19; F primer and R primer, primer sequences used for PCR amplification and Sanger sequencing. Amplicon size, size of amplified PCR product in base pairs (bp). [] indicates inverted segments. Allele-specific mutant breakpoint junctions are indicated in bold.

Table S3. qPCR primers

Target	Primer	Oligonucleotides (5'-3')
<i>SMG8</i> exons 3-4, mRNA	Forward	ACTAATGCCTCAGGTTACAGC
	Reverse	ATCTCAAACCCAAAGGCCA
<i>GDPD1</i> exons 3-5, mRNA	Forward	ATACTGTGAGCTCCACCTTAC
	Reverse	GGAGTGTTAGGAAAGGCCTCAA
<i>YPEL2</i> exons 2-4, mRNA	Forward	TCACTGCAGAGCTCACTTGG
	Reverse	CCACAGCCCACATTAACACTGA
<i>TRIM37</i> exons 11-12, mRNA	Forward	GCGTCAGAGAGCAGATCC
	Reverse	GCACAACCTCCATTTCCATCTG
<i>NRL</i> exons 3-3, mRNA	Forward	GGCTCCACACCTTACAGCTC
	Reverse	AGCCAGTACAGTCTCTCCAG
<i>CRX</i> exons 2-3, mRNA	Forward	GCCCCACTATTCTGTCAACG
	Reverse	CTTCAGAGCCACCTCTCAC
<i>ACTB</i> exons 3-4, mRNA	Forward	CCAACCCGCGAGAAGATGA
	Reverse	CCAGAGGCGTACAGGGATAG
<i>GUSB</i> exons 2-3, mRNA	Forward	AGAGTGGTGCTGAGGATTGG
	Reverse	CCCTCATGCTCTAGCGTGTC
Retinal enhancer, eRNA	Forward	ACCTGCCTGTACGAATCCAA
	Reverse	CTGGGAGGAGGCAAATTGTA
<i>SMG8</i> exon 4	Forward	CCTGGAAGAGAAGTGCGGT
<i>Triplication qPCR 1</i>	Reverse	AGGCCAGAGCACATGAATC
<i>GDPD1</i> intron 1	Forward	TGTGAATTGAGGCTCTTCCG
<i>Triplication qPCR 2</i>	Reverse	ACCGTGTCTTTCCCGTTCA
Downstream of <i>YPEL2</i>	Forward	AAGTTCAGCGTTCTCTCAGAAG
<i>Triplication qPCR 3</i>	Reverse	TGTTGAGTTCTGTCTGCCTCG
<i>LINC01476</i> intron 2	Forward	CCTGCAACCTAACCCTAAGC
<i>Triplication qPCR 4</i>	Reverse	GCATGCCAGATCGCTGTTG

Table S4. Shared heterozygous variants (MAF ≤ 0.0001 , 3 affected individuals) located within the Dutch RP17-locus (NL1)

Chr	Start	End	Ref	Var	gnomAD_G AF	Component	Gene name
chr17	55225642	55225642	G	C	.	intergenic	
chr17	55518925	55518925	C	T	.	intronic	<i>MSI2</i>
chr17	55625150	55625150	C	T	.	intronic	<i>MSI2</i>
chr17	55645018	55645018	G	A	.	intronic	<i>MSI2</i>
chr17	55774511	55774511	A	T	.	intergenic	
chr17	55774518	55774518	C	T	.	intergenic	
chr17	55815842	55815842	C	A	.	intergenic	
chr17	55875136	55875136	C	A	.	ncRNA_intronic	
chr17	56372707	56372707	T	A	.	intergenic	
chr17	56726598	56726598	G	C	.	intronic	<i>TEX14</i>
chr17	56769489	56769489	G	C	0.00008155	upstream	<i>TEX14</i>
chr17	56811950	56811950	T		.	UTR3	<i>RAD51C</i>
chr17	56878006	56878006	C	T	.	intronic	<i>PPM1E</i>
chr17	56970337	56970337	G	T	.	intronic	<i>PPM1E</i>
chr17	56970342	56970342	A	T	0.00006367	intronic	<i>PPM1E</i>
chr17	56970349	56970349	A	C	.	intronic	<i>PPM1E</i>
chr17	56970362	56970362	G	A	.	intronic	<i>PPM1E</i>
chr17	57232150	57232150	C	G	.	intronic	<i>SKA2</i>
chr17	57315768	57315774	TTATTTT		.	intronic	<i>GDPD1</i>
chr17	57333198	57333198		TG	.	intronic	<i>GDPD1</i>
chr17	57403137	57403137		T	.	intergenic	
chr17	57482817	57482817	A	T	.	upstream	<i>AC091059.1</i>
chr17	57510654	57510654	A	T	.	ncRNA_intronic	
chr17	57717126	57717126	C	T	.	intronic	<i>CLTC</i>
chr17	57788781	57788783	ACT		.	intronic	<i>VMP1</i>
chr17	57788784	57788784	C	T	.	intronic	<i>VMP1</i>
chr17	57812263	57812263		TT	.	intronic	<i>VMP1</i>
chr17	57827828	57827828		C	0.00003192	intronic	<i>VMP1</i>
chr17	58092315	58092315	T	C	.	ncRNA_intronic	
chr17	58093706	58093706	A	C	.	ncRNA_intronic	
chr17	58203846	58203846	G	A	.	upstream	<i>AC025048.5</i>
chr17	58691382	58691382	T	C	.	intronic	<i>PPM1D</i>
chr17	59551786	59551796	CTACCAGCATT		.	intronic	<i>TBX4</i>
chr17	59646261	59646261	G	T	.	intergenic	
chr17	59652939	59652939	A	T	.	intergenic	

Table S4. Continued

Chr	Start	End	Ref	Var	gnom		Component	Gene name
					AD	G AF		
chr17	59654667	59654667	A	T	.	.	intergenic	
chr17	59935737	59935737	T	G	.	.	intronic	BRIP1
chr17	59987185	59987192	TGTGTGTG		.	.	intronic	INTS2
chr17	60065110	60065110	A	T	.	.	intronic	MED13
chr17	60223074	60223074	G	C	.	.	intergenic	
chr17	60223078	60223078	G	T	.	.	intronic	

Chr, chromosome; Start, End, genomic positions based on hg19; Ref, reference allele; Var, variant; GnomAD_G AF, minor allele frequency according to gnomAD v.2.1.1; Component, genomic position.

Table S5. Shared heterozygous variants (MAF \leq 0.0001, 3 affected individuals) located in the founder haplotype in family UK1

Chr	Start	End	Ref	Var	gnomAD_G AF	Component	Gene name
chr17	56059537	56059537	T	C	0.00003228	intronic	VEZF1
chr17	56122144	56122144	T	C	0.00003228	intergenic	
chr17	56293716	56293716	G	A	.	intronic	MKS1
chr17	56478605	56478605	T	C	.	intronic	RNF43
chr17	56731111	56731111	G	A	.	intronic	TEX14
chr17	56775478	56775478	T	A	.	intronic	RAD51C
chr17	56783547	56783547	T	C	.	intronic	RAD51C
chr17	56834462	56834462	C	A	.	intronic	PPM1E
chr17	57107553	57107553	G	A	.	intronic	TRIM37
chr17	57260755	57260755	A	G	0.0001	intronic	PRR11
chr17	57548764	57548764	T	C	.	ncRNA_intronic	LINC01476
chr17	57616479	57616479	A	G	0.00009681	intergenic	
chr17	57641653	57641653	G	A	0.000097	intergenic	
chr17	57688592	57688592	T	C	.	intergenic	
chr17	57918969	57918969	G	C	.	UTR3	VMP1
chr17	58024808	58024808	A	G	.	UTR3	RPS6KB1
chr17	58108605	58108605	G	A	.	intergenic	
chr17	58932373	58932373	C	T	.	intronic	BCAS3
chr17	59279276	59279276	C	T	0.00006532	intronic	BCAS3
chr17	59328755	59328755	C	A	.	intronic	BCAS3
chr17	59913924	59913924	C	T	.	intronic	BRIP1
chr17	60391209	60391209	G	A	0.0001	intergenic	
chr17	60404484	60404484	C	G	0.0001	intergenic	
chr17	60428327	60428327	G	T	.	intergenic	
chr17	61687237	61687237	C	T	0.00003228	intergenic	
chr17	61696765	61696765	T	A	0.00006906	intergenic	
chr17	62009718	62009718	C	T	0.00009688	upstream	CD79B
chr17	62075612	62075612	A	G	.	ncRNA_intronic	PRR29-AS1
chr17	62113494	62113494	C	G	.	intergenic	
chr17	62826064	62826064	A	C	.	ncRNA_intronic	PLEKHM1P1
chr17	62834157	62834157	C	G	.	upstream	PLEKHM1P1
chr17	62855508	62855508	G	A	0.0001	intronic	LRRC37A3

Chr, chromosome; Start, End, genomic positions based on hg19; Ref, reference allele; Var, variant; GnomAD_G AF, minor allele frequency according to gnomAD v.2.1.1; Component, genomic position.

Table S6. Genomic details of RP17-SVs

SV	Type	Chr	Start	End	Event	Size (Mb)	Genomic regions	Genes involved	Detection method
NL-SV1	Dup	17	57,291,905	57,518,137	Dup	0.23	B	SMG8, GDPD1, YPEL2, MIR4729, LINC01476	FREEC, Manta
UK-SV2	DupINVDup	17	57,275,839	57,456,098	Dup	0.18	E	PRR11, SMG8, GDPD1, YPEL2, MIR4729	Canvas
			57,275,839	57,559,114	Inv	0.28	E, F, G	PRR11, SMG8, GDPD1, YPEL2, MIR4729, LINC01476	Manta
SA-SV3	DupINVDup	17	57,468,960	57,559,114	Dup	0.09	G	YPEL2, LINC01476	Canvas
			57,247,615	57,391,678	Dup	0.14	J	PRR11, SMG8, GDPD1	FREEC
			57,516,678	57,612,711	Inv	0.10	M	LINC01476	Manta
CA-SV4	DupINVDup	17	57,499,214	57,612,711	Dup	0.11	L, M	LINC01476	FREEC
			57,233,035	57,280,008	Dup	0.05	P	PRR11	FREEC
			57,233,035	57,634,900	Inv	0.40	P, Q, R	PRR11, SMG8, GDPD1, YPEL2, MIR4729, LINC01476	Manta
NL-SV5	Dup	17	57,483,883	57,634,900	Dup	0.15	R	LINC01476	FREEC
			57,260,511	57,515,862	Dup	0.13	U	PRR11, SMG8, GDPD1, YPEL2, MIR4729, LINC01476	FREEC, Manta
UK-SV6	Trip	17	57,295,969	57,510,765	Trip	0.21	X	GDPD1, YPEL2, MIR4729, LINC01476	Canvas, Manta
UK-SV7	DupINVDup	17	57,259,525	57,453,630	Dup	0.16	AA	PRR11, GDPD1, YPEL2, MIR4729	Canvas
			57,259,525	57,710,821	Inv	0.42	AA, AB, AC	PRR11, GDPD1, YPEL2, MIR4729, LINC01476, DHX40, CLTC	Manta
			57,468,931	57,710,821	Dup	0.24	AC	YPEL2, LINC01476, DHX40, CLTC	Canvas
UK-SV8	DupINVDup	17	57,277,347	57,326,234	Dup	0.05	AF	PRR11, SMG8, GDPD1	Canvas
			57,277,347	57,631,659	Inv	0.35	AF, AG, AH	PRR11, GDPD1, YPEL2, MIR4729, LINC01476	Manta
			57,413,153	57,631,659	Dup	0.22	AH	YPEL2, MIR4729, LINC01476	Canvas

SV, Structural variant; Type, complex structural rearrangements specified as duplications (dup), inversions (inv), triplications (trip) rearrangements; Start, End, genomic positions of structural rearrangements according to GRCh37/hg19; Size, of structural rearrangement in Mb; Genomic regions annotated as illustrated in **Figure 2**; Detection method, specific tools employed for identification of structural rearrangements are described in materials and methods section.

Table S7. Repetitive elements identified in sequence flanking the breakpoints

SV	Breakpoint	Coordinates	Repetitive elements
NL-SV1	A-B	57,291,905	97.01% SINE/ALU
	B-C	57,518,137	24.92% SINE/MIR
UK-SV2	D-E	57,275,839	99.34% SINE/ALU
	E-F	57,456,098	16.28% SINE/ALU
	F-G	57,468,960	44.19% SINE/ALU
	G-H	57,559,114	46.18% SINE/ALU
SA-SV3	I-J	57,247,615	54.49% SINE/ALU, 39.53% DNA
	J-K	57,391,678	85.86% SINE/ALU
	K-L	57,499,214	22.92% SINE/ALU
	L-M	57,516,678	55.15% SINE/ALU, 14.29% DNA/hAT-Charlie
	M-N	57,612,711	50.17% SINE/ALU, 20.27% small RNA
CA-SV4	O-P	57,233,035	NP
	P-Q	57,280,008	25.58% SINE/ALU
	Q-R	57,483,883	NP
	R-S	57,634,900	96.01% LTR/ERV-class I
NL-SV5	T-U	57,260,511	NP
	U-V	57,515,862	50.17% SINE/ALU, 38.21% LINE/L1
UK-SV6	W-X	57,295,969	72.76% SINE/ALU
	X-Y	57,510,765	81.06% SINE/ALU
UK-SV7	Z-AA	57,259,525	95.68% SINE/ALU
	AA-AB	57,453,630	67.11% SINE/ALU
	AB-AC	57,468,931	33.22% SINE/ALU
	AC-AD	57,710,821	66.45% SINE/ALU
UK-SV8	AE-AF	57,277,347	NP
	AF-AG	57,326,234	NP
	AG-AH	57,413,153	68.44% SINE/ALU
	AH-AI	57,631,659	41.86% SINE/ALU, 56.48% LINE/L1

Presence of repetitive elements was assessed using RepeatMasker from the reference sequence, 150 bp reference sequences flanking each side of the breakpoint were used as input. SV, Structural variant; Breakpoint, Breakpoint annotation of genomic regions as illustrated in **Figure 2**; Coordinates, Genomic position of breakpoint according to GRCh37/hg19 coordinates; Repetitive elements. Percentage of repetitive elements present in input sequence per specified element (class/family); NP, Not present.

Table S8. Assessment of microhomology, insertions and deletions at allele-specific breakpoints

SV	Breakpoint junction	3' Coordinates	5' Coordinates	(Micro)homology	Insertion	Deletion
NL-SV1	B-B	57,518,137	57,291,905	5 bp (AGGCA)	-	-
UK-SV2	E-[G]	57,456,098	57,559,114	NP	9 bp (TTTTATGAC)	-
	[E]-G	57,275,839	57,468,960	NP	9 bp (AGGCTGGTC)	-
SA-SV3	L-J	57,516,678	57,247,615	NP	23 bp (AAAAAAAACTTGAAGAAAGATT)	-
	J-[M]	57,391,678	57,612,711	4 bp (TTCAG)	-	-
	[M]-L	57,516,678	57,499,214	1 bp (C)	13 bp (GGTCCAGATTGTG)	4 bp (AGAG)
CA-SV4	P-[R]	57,280,008	57,634,900	1 bp (T)	-	-
	[P]-R	57,233,035	57,483,883	2 bp (GC)	5 bp (TAAAGC)	-
NL-SV5	U-U	57,515,862	57,260,511	5 bp (ATCCT)	-	-
UK-SV6	X-X	57,510,765	57,295,969	>100 bp	-	-
UK-SV7	AA-[AC]	57,453,630	57,710,821	>100 bp	-	-
	[AA]-AC	57,259,525	57,468,931	NP	10 bp (GTAATTTTC)	-
UK-SV8	AF-[AH]	57,326,234	57,631,659	NP	2 bp (CT)	-
	[AF]-AH	57,277,347	57,413,153	2 bp (CT)	-	-

SV, Structural variant; Breakpoint junction, Allele-specific breakpoint junction between genomic regions as illustrated in Figure 2; Coordinates, genomic position of breakpoints according to hg19; Microhomology, presence of microhomology was assessed using ClustalOmega; Insertion and deletion, presence of insertions or deletions as determined by Sanger sequencing. [] Indicate inverted segments, bp, base pairs; NP, not present.

Table S9. Clinical findings

Supplemental Table S9 can be found online via the following link: <http://bit.ly/ClinicalDataRP17>.

SUPPLEMENTARY FIGURES

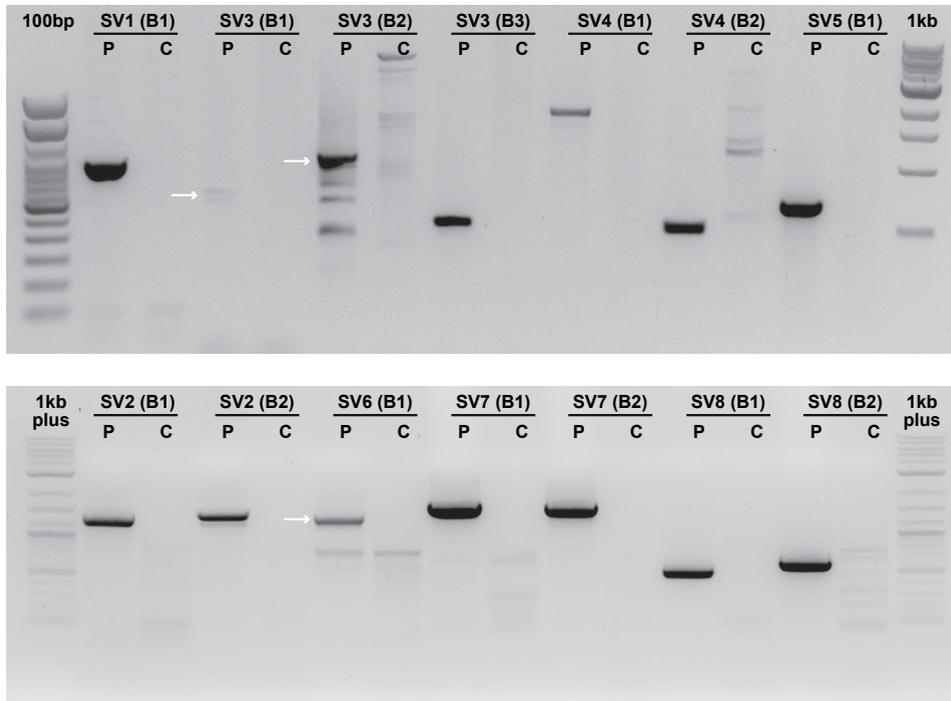


Figure S1. Agarose gels of allele-specific breakpoint junction PCR amplifications. PCR amplification of allele-specific breakpoints was performed, and gel electrophoresis was used for visualization. For each structural variant, allele-specific breakpoints (B) could be amplified in affected individual DNA (P) but not in DNA obtained from anonymous controls (C). Breakpoint junction nomenclature corresponds to those illustrated in **Figure 2**. Primer sequences used for PCR amplification are listed in **Table S2**.

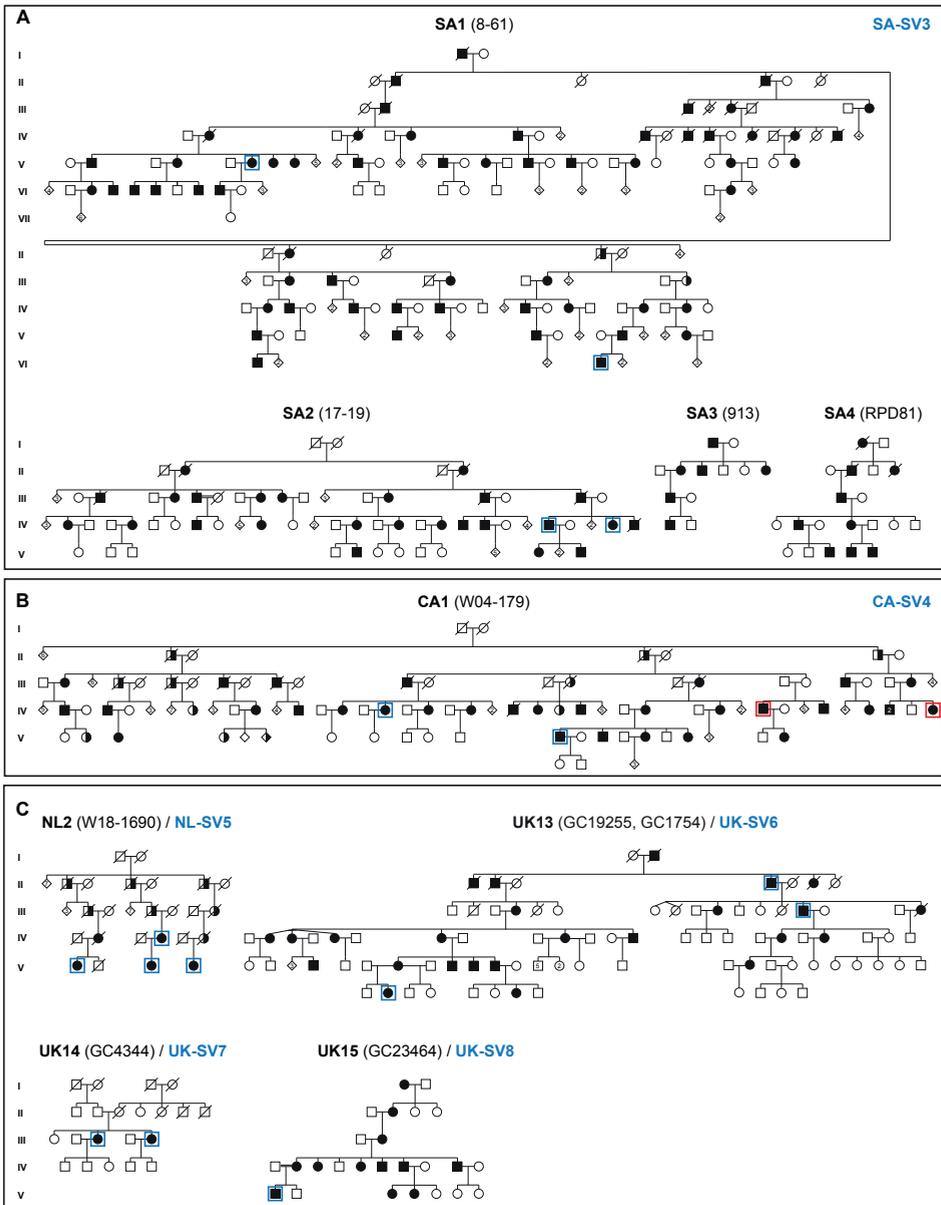


Figure S2. Additional adRP families with structural variants within the RP17-locus. (A) Pedigrees of South African origin with structural variant 3 (SA-SV3) **(B)** Canadian family with structural variant 4 (CA-SV4). **(C)** Dutch family (NL2) with structural variant 5 (NL-SV5), UK family (GC19255/GC1754) with structural variant 6 (UK-SV6), UK family (GC4344) with structural variant 7 (UK-SV7) and UK family (GC23464) with structural variant 8 (UK-SV8). WGS or WES was performed in individuals in blue or red, respectively.

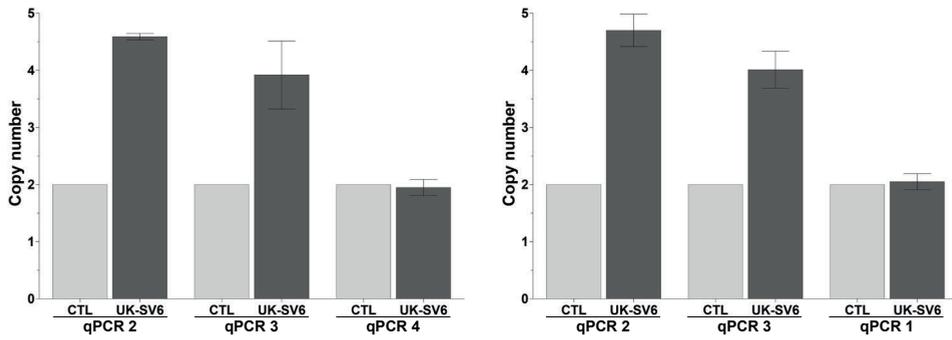


Figure S3. Triplication of region in UK-SV6. Quantitative real-time PCR for individuals from family UK13, UK-SV6. Primer pairs qPCR 2 (first intron of *GDPD1*) and qPCR 3 (downstream of *YPEL2*) confirmed triplication (four copies in the genome) of UK-SV6, compared to control unaffected DNA samples and additional control qPCR assays for genomic regions distal and proximal to this structural variant qPCR 4 (last intron of *LINC01476*) and qPCR1 (exon 3 of *SMG8*). CTL, unaffected control DNA sample; UK-SV6, affected individual DNA sample. Primer sequences are listed in **Table S3**.

Structural variants cause ectopic enhancer-gene contact in retinitis pigmentosa

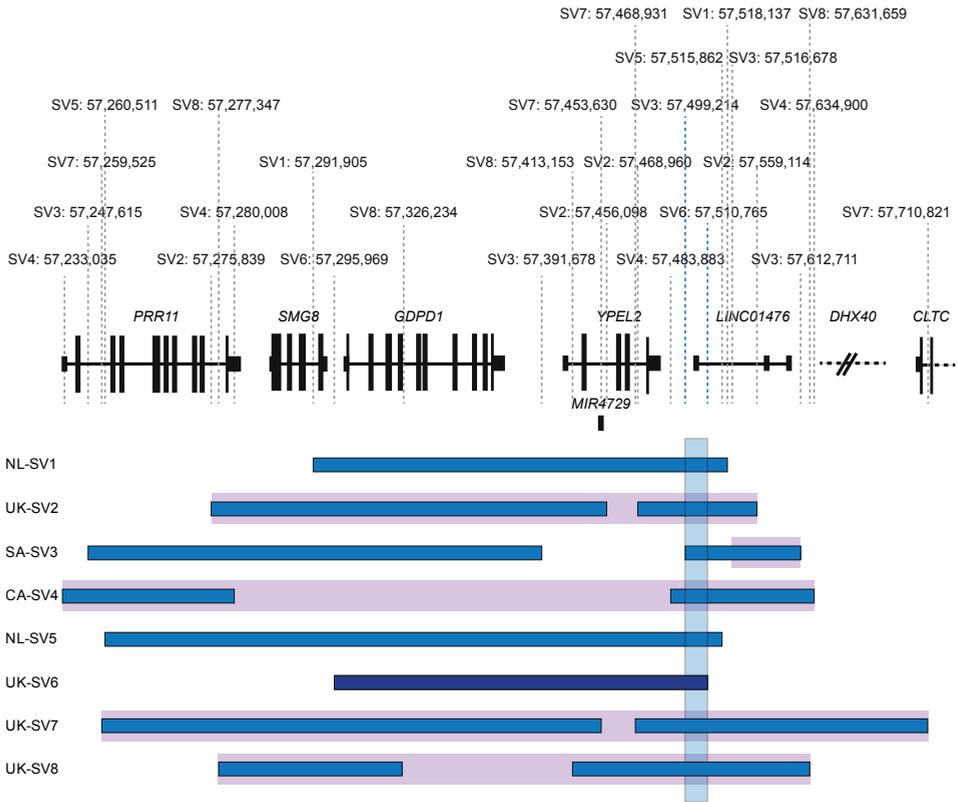
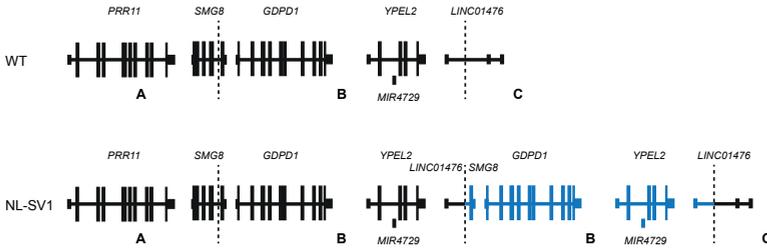


Figure S4. Overview of all SV breakpoints identified in the RP17-locus. Breakpoints are indicated with dashed lines. Duplicated, triplicated or inverted genomic regions for each SV are illustrated with blue, dark blue or purple bars, respectively. An overlapping genomic region that is duplicated or triplicated in all SVs was identified (chr17:57,499,214-57,510,765) and highlighted by a light-blue vertical bar. The size of *DHX40* is reduced and *CLTC* is only partially shown in this figure.

A



```

A-B ---TATATATTCTTTTATTTATTTATTTGTTTGTGTTTTTTTGGAGACAGTCTCGCTCTGTTG
B-B -----AACCCTGTCTCAAAAAAAAAAAAAATAAATAAATAAAGGAGAGTCA-----AGTCT
B-C -----AACCCTGTCTCAAAAAAAAAAAAAATAAATAAATAAAGGAGAGTCA-----AGTCT
      * * * * *

```

```

A-B CACAGGCTGGAGGGCAGTGGTGC-AATCTCAGCTCACTGCACCTCCGCCCTCTGGGTTC
B-B C-----TCAGGGCCTTGGTTACTTTATCTGTAAAAATGAAGGT-----ATTGGGCTAGA
B-C C-----TCAGGGCCTTGGTTACTTTATCTGTAAAAATGAAGGT-----ATTGGGCTAGA
      * * * * *

```

```

A-B AGCTGTTCTCCTGCCCCAGCC-----TCCCCAGTAACAGATTACAGGCAACAGCCA
B-B AGATTCCCTAGCAGCCCTAACAGGTTAACTAACAGCCAACCTGGGAGAAAGGCAACAGCCA
B-C AGATTCCCTAGCAGCCCTAACAGGTTAACTAACAGCCAACCTGGGAGAAAGGCAACAGGAC
      * * * * *

```

```

A-B CCACACCTGACTAATTTTTTGTATTTTAGTAGACATGGGGTT---TCTCCATGTTGGC
B-B CCACACCTGACTAATTTTTTGTATTTTAGTAGACATGGGGTT---TCTCCATGTTGGC
B-C AAGAATTCCAACGCTGGCTACTTGAC---AGCACAATGTGACTGGAGTCTCTTGGCTGAC
      * * * * *

```

```

A-B CAGGCTGGTCTCGAACTCCT--GACCTCGTGATTCACCTGCCTCAGCCTC-CCAAAGTGC
B-B CAGGCTGGTCTCGAACTCCT--GACCTCGTGATTCACCTGCCTCAGCCTC-CCAAAGTGC
B-C TTGTATTCTCTCACTGGTCTCAGCCAAATGCTTGCCTGATATGGTACTAAGTTATCCC
      * * * * *

```

```

A-B TGGAAATTACAGGCGTGAGCCA-----
B-B TGGAAATTACAGGCGTGAGCCA-----
B-C TTCTAGAAAAGTTGGTGCCTCACAT
      * * * * *

```

Figure continues on next page

Structural variants cause ectopic enhancer-gene contact in retinitis pigmentosa

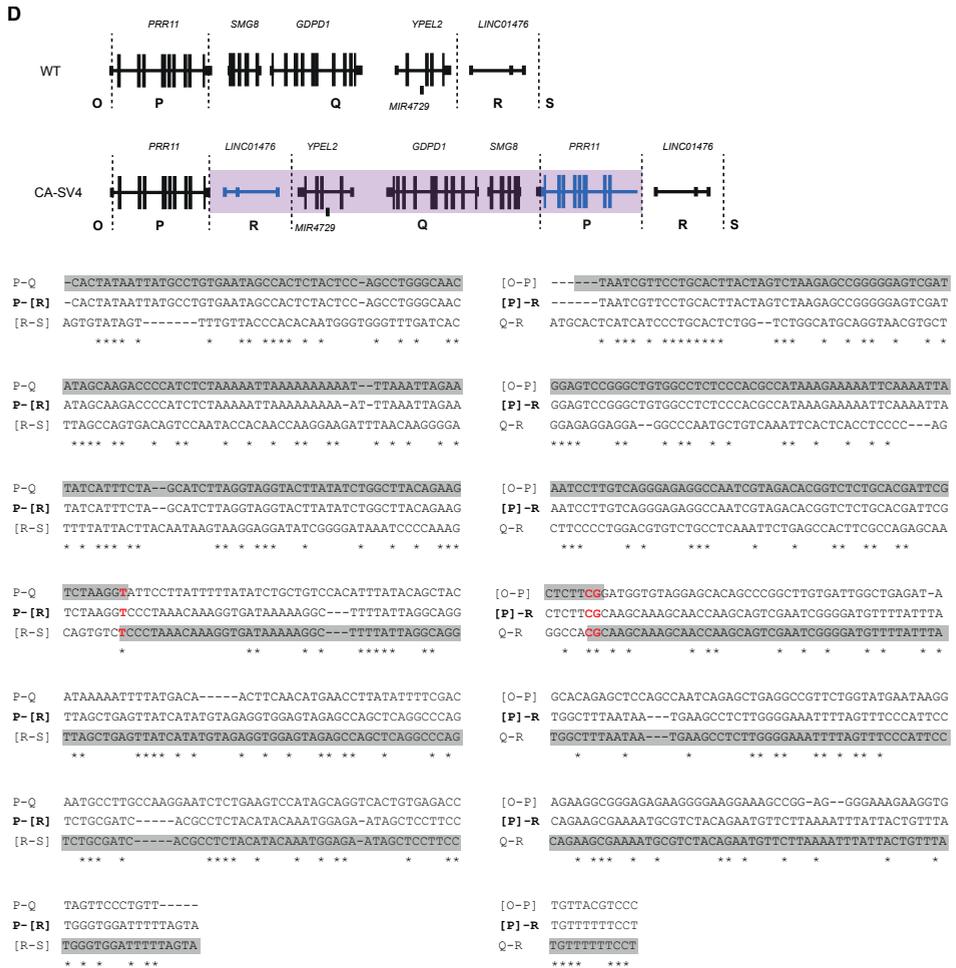
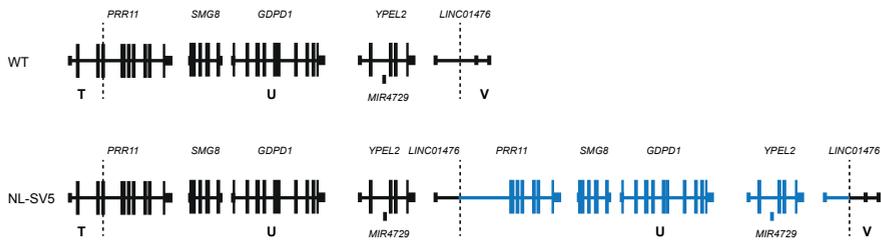


Figure continues on next page

E



```

T-U  GCAGTTGTCCACTTGAGCTCCCAGATGCCCATGGAGTCCAGTCTTCCAATCAGGAAGT
U-U  -CAGATTTGCTTGATCCGTCACCTTGCAATGCCCTTCTCTGTCTTTCCTTCTCCATTAA
U-V  CAGATTTGCTTGATCCGTCACCTTGCAATGCCCTTCTCTGTCTTTCCTTCTCCATTAA
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

T-U  TGGAAATCTCTGATGTCATTGGTCATTCACCTGGCAACCAGTTTGAAGAAAAACACATG
U-U  T--ACATGCTTATCACCTTATTC AAGAGCCAAGGAGAACTCACCACCTGCAGGAAGACT
U-V  T--ACATGCTTATCACCTTATTC AAGAGCCAAGGAGAACTCACCACCTGCAGGAAGACT
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

T-U  TAACTGCCAGGCTGGTCTCTTGTCTGGAGATCCTGGGTGAATGGTATCTCCTGCCACTG
U-U  GCCCAGACTGACCCCTCTTAACTGTAATCAATCCTGGGTGAATGGTATCTCCTGCCACTG
U-V  GCCCAGACTGACCCCTCTTAACTGTAATC-ATCCTTTTTTTTTTTTTTCTTTTTTAAGA
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

T-U  TCCCAACCTCAGACCACCATCCAAAAGCATCTT-----C-----AGGGTCTCCGCAT
U-U  TCCCAACCTCAGACCACCATCCAAAAGCATCTT-----C-----AGGGTCTCCGCAT
U-V  -CAGAACTCACTCTGTTCATCTAGGCTGGAGTGCACCTGGCCGATCTGGGCTCACTGCAA
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

T-U  CCATCTGTTCCTGTCCAGCAGAGGCTGTGTCTTCTCCACTCAAAGCCTGAAGCATTTT
U-U  CCATCTGTTCCTGTCCAGCAGAGGCTGTGTCTTCTCCACTCAAAGCCTGAAGCATTTT
U-V  CCTTCGCCTCCTGGGTCAA-----GCAATGTCCTGCCTCAGCCTCCAGAGTAGC
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

T-U  TGGGGTCTCC-----
U-U  TGGGGTCTCC-----
U-V  TGAGATTACAGGTGC
      * * * *
    
```

Figure continues on next page

F

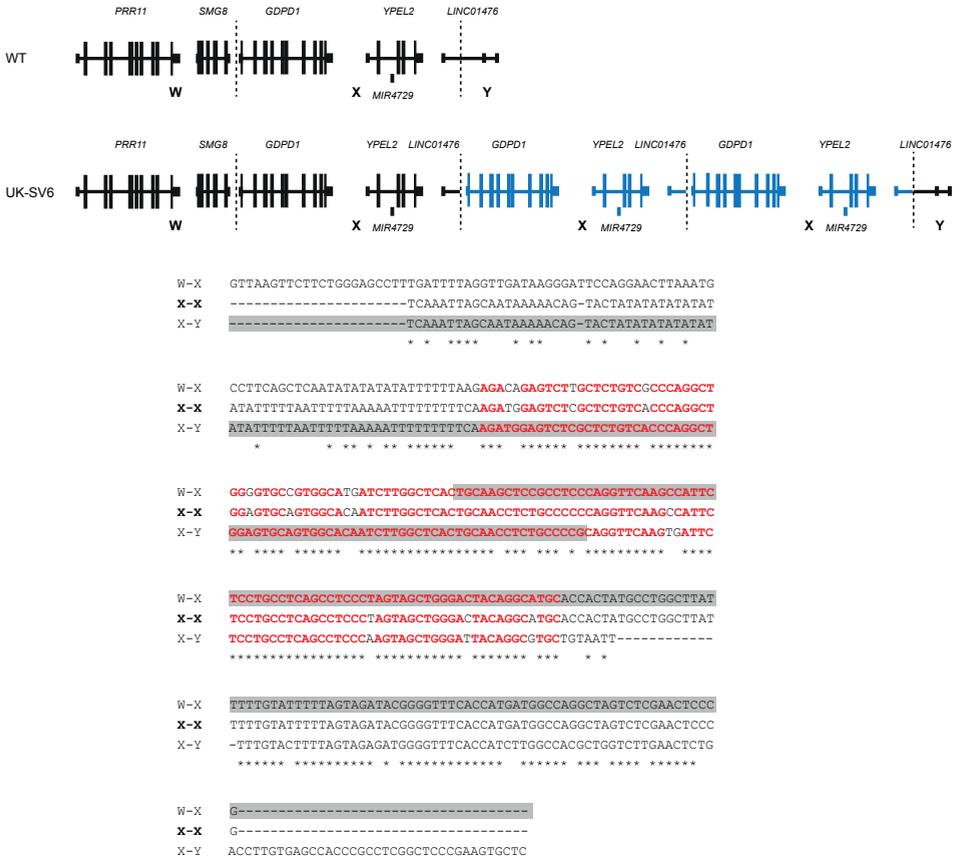
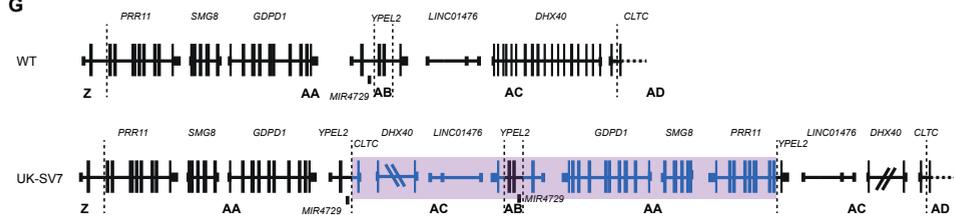


Figure continues on next page

G



AA-AB	<u>-GAAGTCATTGGGATGACACCTTCATGACCCCATCTTCATTCTCCCTAG</u>	[Z-AA]	<u>GGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGCGGGCAGATCA</u>
AA-[AC]	-GAAGTCATTGGGATGACACCTTCATGACCCCATCTTCATTCTCCCTAG	[AA]-AC	GGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGCGGGCAGATCA
[AC-AD]	TTATTCTAGTTAGATCAATCAT---TTAAGGCATACTCAGTCTACCAGGT	AB-AC	---AAAAAATATATTTCACACATAT---AAATATGTAGCGATCA
	* * * * *		* * * * *
AA-AB	<u>ACTTTAGAGCCAGACTGCCTGGAC--CTCTGTATCATTIATTTGATTGTG</u>	[Z-AA]	<u>CTTGAGCCAGGAGTT-TGAGACCA---GCCTGGTAAACATGGTGAAC</u>
AA-[AC]	ACTTTAGAGCCAGACTGCCTGGAC--CTCTGTATCATTIATTTGATTGTG	[AA]-AC	CTTGAGCCAGGAGTT-TGAGACCA---GCCTGGTAAACATGGTGAAC
[AC-AD]	GCTA-CAAGTTATAAGACTAGAACACATTTTTTTTACAATTTAGTCACTA	AB-AC	TGCCAGGCACTGAATATGATTCATTGAAATAGTTTGAATATGTGGAAA
	* * * * *		* * * * *
AA-AB	<u>GGTTTGTGTGGTTGTGTTGTTTTGAGATGGAGTCTCTCTGTCGCCCCAGG</u>	[Z-AA]	<u>CCCAT--CTCTACTAAAAGTACAAAATTGAGCTGGCATTATGGCAGATA</u>
AA-[AC]	GGTTTGTGTGGTTGTGTTGTTTTGAGATGGAGTCTCTCTGTCGCCCCAGG	[AA]-AC	CCCAT--CTCTACTAAAAGTACAAAATTGAGCTGGCATTATGGCAGATA
[AC-AD]	GAGTATTTTTTTTCTTTTTTTGAGATGGAGTCTCTGTCGCCCCAGG	AB-AC	TCATTAATAAAAAAAAAATGTCAGGTGGAGCTGTGTAGCACTGCCT-TT
	* * * * *		* * * * *
AA-AB	<u>CTGGCATGCAATAGCGTATCTGGGCTCACTGCAATCTCTGCCTCCCAGG</u>	[Z-AA]	<u>CCTGTAATCCAGCTACCTGAGAGGCTGAGGCATGAGAATAACTGGAATC</u>
AA-[AC]	CTGGAGTGCATGGCATGATCTCAGCTACCCACAACTCTGCTGCCCCGGG	[AA]-AC	CCTGTAATTTTCCCATCTCCCTGGAAACATTCAGGATTACAAAACAAA
[AC-AD]	CTGGAGTGCATGGCATGATCTCAGCTACCCACAACTCTGCTGCCCCGGG	AB-AC	AGATTCATTTTCCCATCTCCCTGGAAACATTCAGGATTACAAAACAAA
	* * * * *		* * * * *
AA-AB	<u>FTCAAGTGATTCCTGCCTCAACCTCCCTAGTAGCTGGATTA-CAGGCA</u>	[Z-AA]	<u>CGGGAGACAGATTTGCAGTGAACCTGAGATTGCACCCTACACTCCAGCC</u>
AA-[AC]	FTCAAGTGCATGGCATGATCTCAGCTACCCACAACTCTGCTGCCCCGGG	[AA]-AC	ATGGGGCCAGGCGTAAATAGCTCAGGCCTATAACCCAGCAGTTTGGGA--
[AC-AD]	FTCAAGCCATTCCTGCCTCAACCTCCCTAGTAGCTGGAGTACAGGCA	AB-AC	ATGGGGCCAGGCGTAAATAGCTCAGGCCTATAACCCAGCAGTTTGGGA--
	* * * * *		* * * * *
AA-AB	<u>CCCGCCACCACCTCCGGTAATTTTTGTATTTTTGTATAG--AGACGGGGT</u>	[Z-AA]	<u>CGGGCGACAAAGCAAGATTCTGTCTCGAAACAAAAAACAATCAAT</u>
AA-[AC]	TGCACCACCACGCCAGCTAATTTTTGTATTTTTTTCAGTAGACGGGGT	[AA]-AC	-GGCCGAGTGGGTGGATCACCTGAGGTCA--AGAGTTGAGACCCCGCT
[AC-AD]	TGCACCACCACGCCAGCTAATTTTTGTATTTTTTTCAGTAGAGACGGGGT	AB-AC	-GGCCGAGTGGGTGGATCACCTGAGGTCA--AGAGTTGAGACCCCGCT
	* * * * *		* * * * *
AA-AB	TTCACCA	[Z-AA]	AACAAAAA-----
AA-[AC]	TTCAC--	[AA]-AC	GGCCACATATAGT
[AC-AD]	TTCAC--	AB-AC	GGCCACATATAGT
	*****		* * * *

Figure continues on next page

Structural variants cause ectopic enhancer-gene contact in retinitis pigmentosa

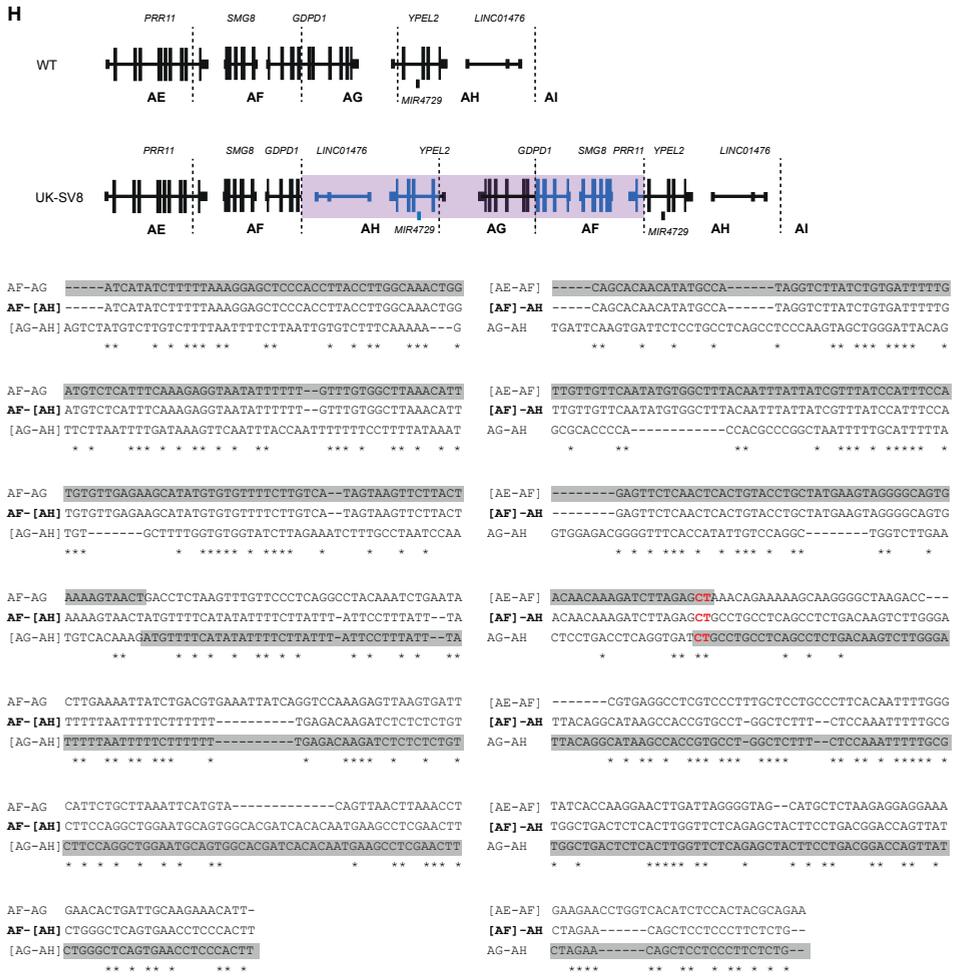


Figure S5. Assessment of microhomology at breakpoint sequences. (A-H) Breakpoint regions of structural variants (SVs) were assessed for presence of microhomology by using multiple sequence alignment between the junction fragment and the 5' and 3' breakpoint regions using Clustal Omega. 150 bp reference sequences flanking each side of the breakpoint were used as input. Regions of microhomology are indicated in red. Breakpoint annotation of genomic regions as illustrated in **Figure 2**.

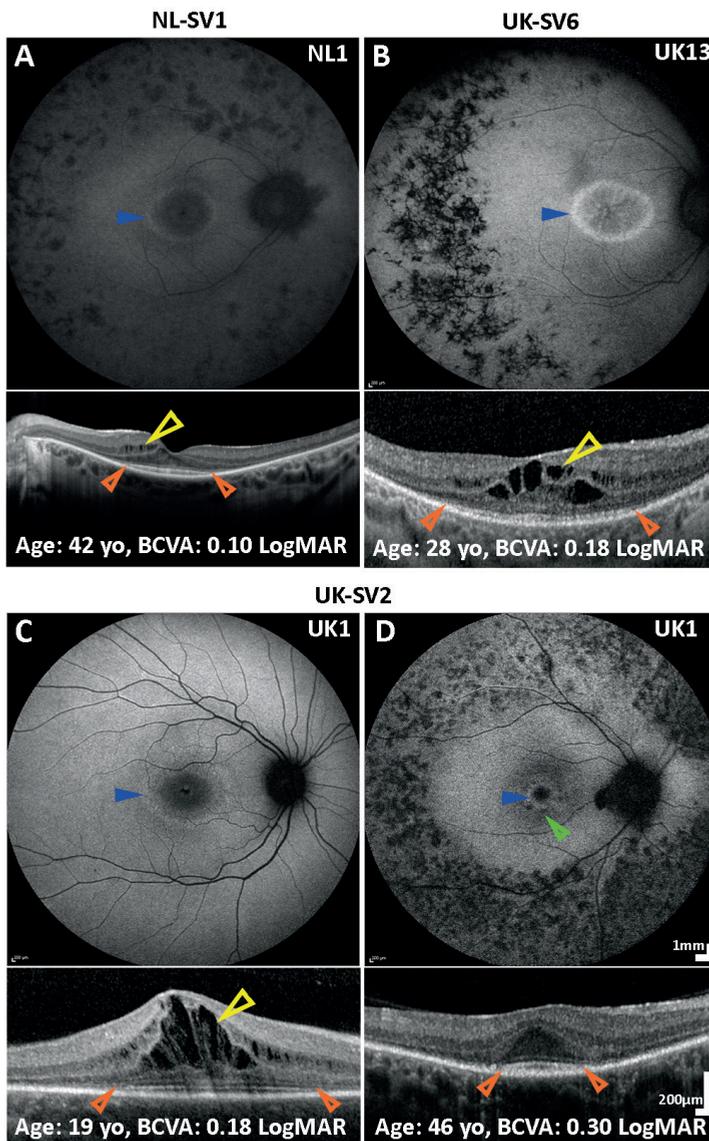


Figure S6. Detailed Retinal Imaging with Fundus Autofluorescence (FAF) and Optical Coherence Tomography (OCT). The blue arrow heads mark the temporal border of the ring of increased signal where present. The yellow arrow heads mark areas of cystoid macular edema (**A, B, C**). The orange arrow heads mark on the OCT scans the border of the residual ellipsoid zone (EZ). Bone spicules are visible in the mid periphery on FAF in cases (**A, B, D**). (**A**) Affected individual from family NL1 (NL-SV1). (**B**) Affected individual from family UK13 (UK-SV6). Affected individuals shown in (**C**) and (**D**) are from UK1 (UK-SV2). Two consecutive generations are shown, mother (**C**) and daughter (**D**). Note the slow structural disease progression indicated by these cases, with minimal change in the BCVA, due to sparing of the foveal EZ (orange arrow heads). The ring of increased signal decreases in size over time (blue arrow heads), with small areas of decreased signal (atrophy) developing and increasing in these same regions over time (green arrow heads). yr, years old; BCVA, best corrected visual acuity; LogMAR, Logarithm of the minimum angle of resolution.

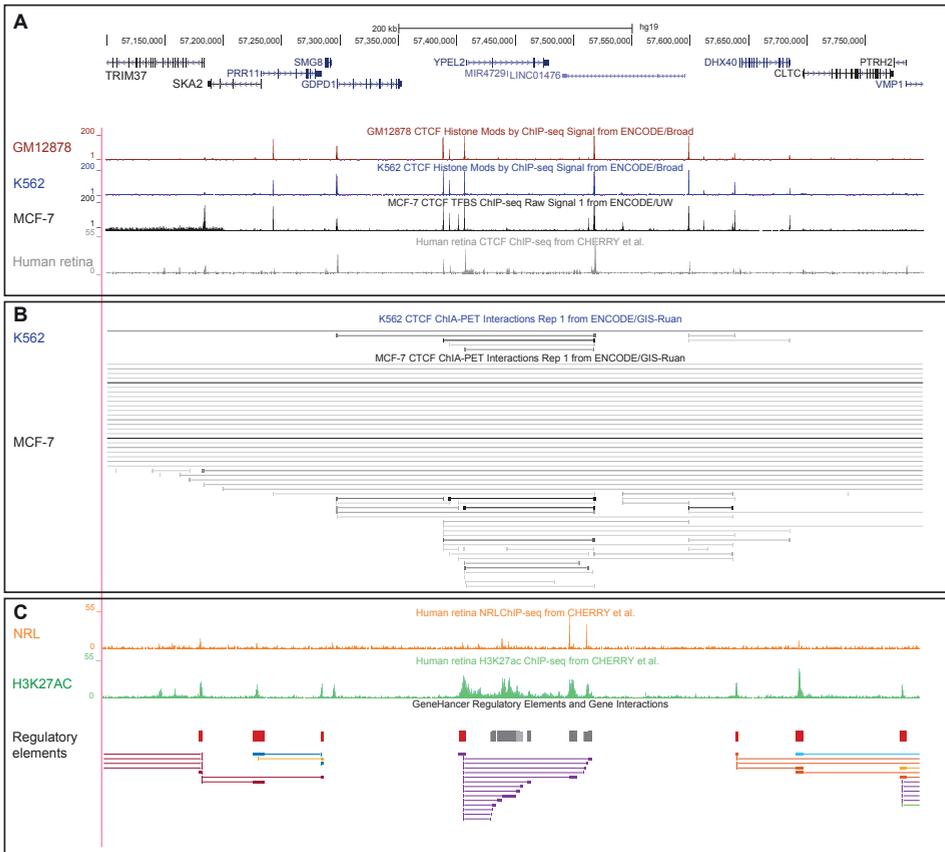


Figure S7. CTCF sites insulate contacts within the *YPEL2* TAD. (A) CTCF ChIP-seq data showed that *YPEL2* is located within an insulated TAD that is present in multiple cell and tissue types, including human retina. **(B)** ChIA-PET CTCF interaction data established in K562 and MCF-7 cells revealed strong interactions between the CTCF binding sites on the 5' side of the *YPEL2* TAD, and the single CTCF binding site on the 3' side of the *YPEL2* TAD. **(C)** The structured *YPEL2* TAD contains retina-specific enhancer elements as shown in **Figure 3**. These regulatory elements are also described in the GeneHancer database³² and interactions with the *YPEL2* promoter region were experimentally validated.

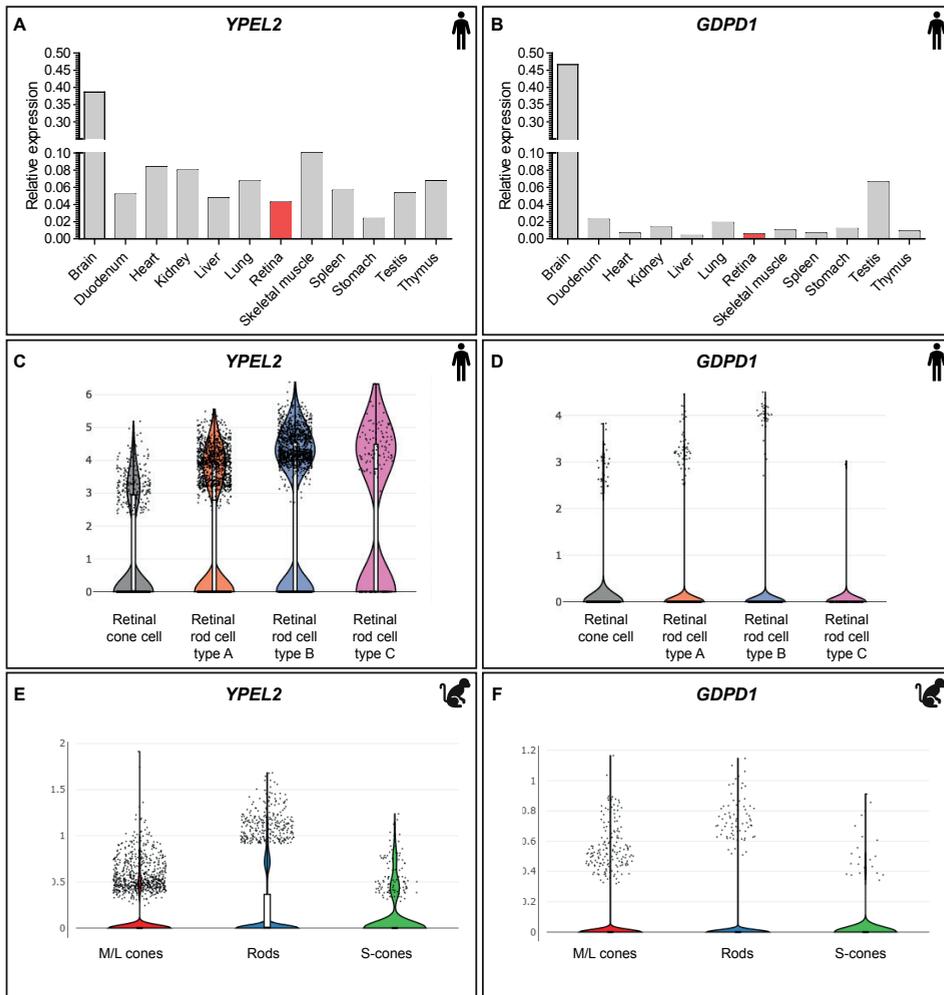


Figure S8. *YPEL2* and *GDPD1* expression across tissues and retinal cell types. (A) qPCR expression levels of *YPEL2* across healthy human tissues. *YPEL2* is ubiquitously expressed in the tissues studied, including retina, with highest expression in brain. (B) *GDPD1* is detected at low levels in all tissue types, with higher levels of expression in brain and testis. (C-F) *YPEL2* has higher expression in rod photoreceptor cells compared to cones from single cell RNA sequence data. *GDPD1* has low expression in all photoreceptor cells. Single cell expression levels and plots were obtained from the Broad Institute Single Cell Portal, and is based on single cell RNA sequencing results of human²⁶ (C-D) and macaque²⁷ (E-F) retinal cell types.

Structural variants cause ectopic enhancer-gene contact in retinitis pigmentosa

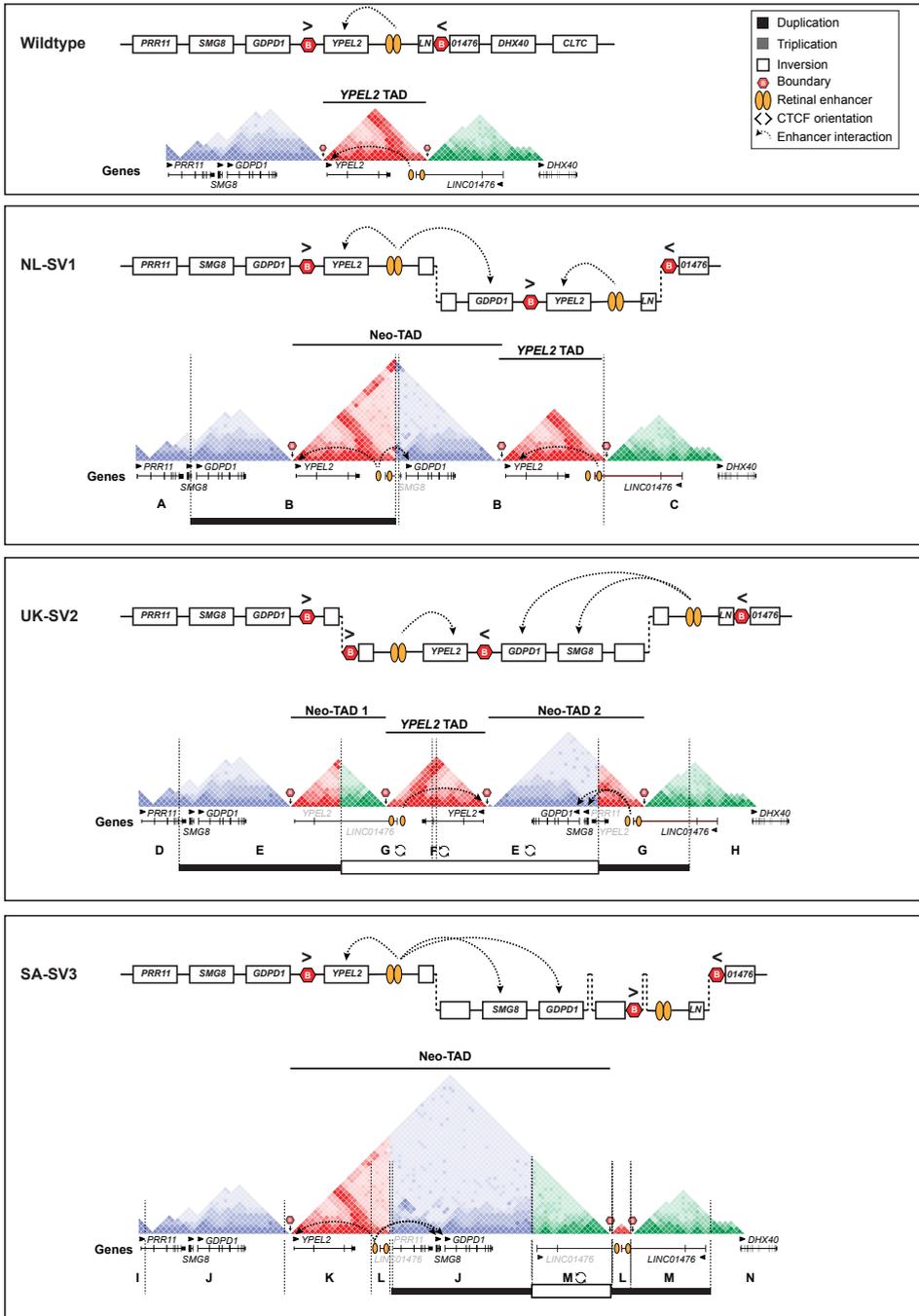


Figure continues on next page

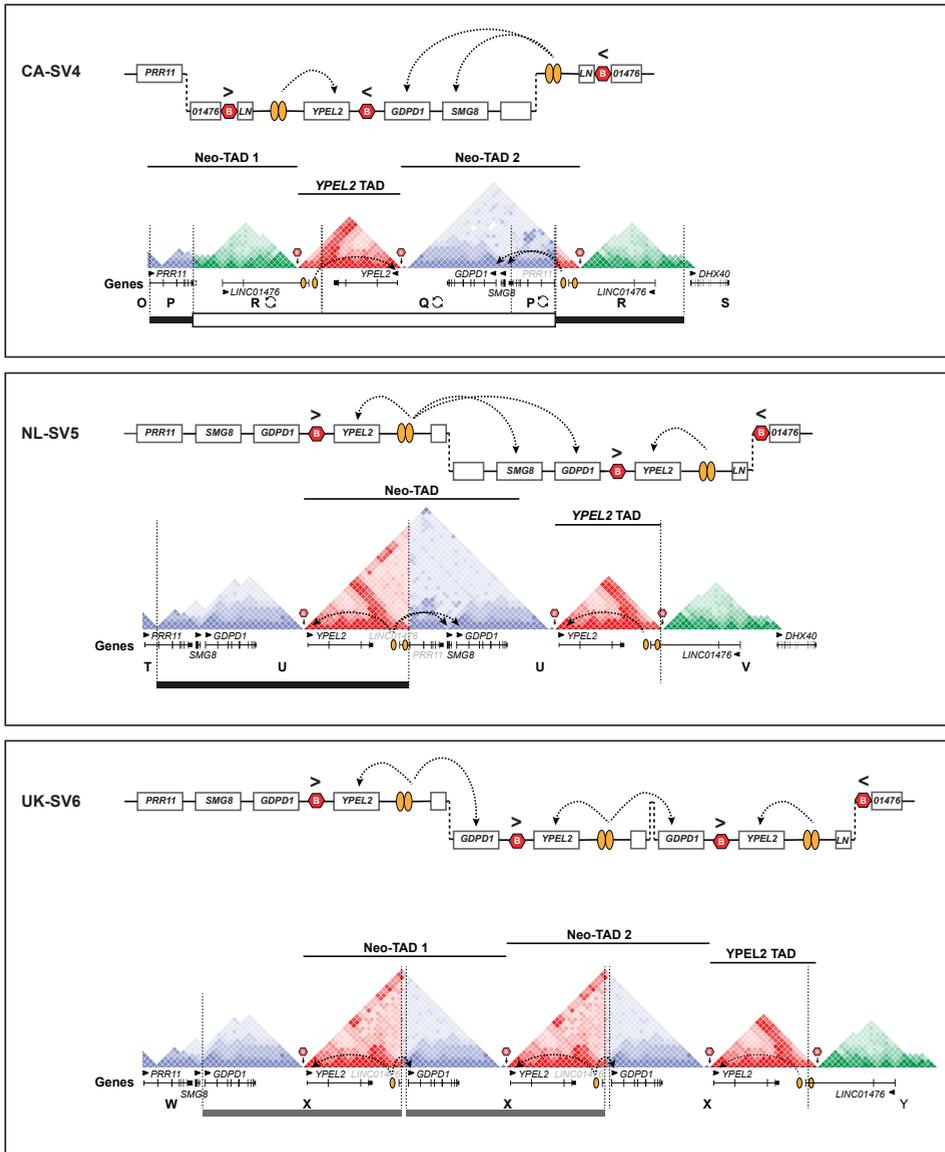


Figure continues on next page

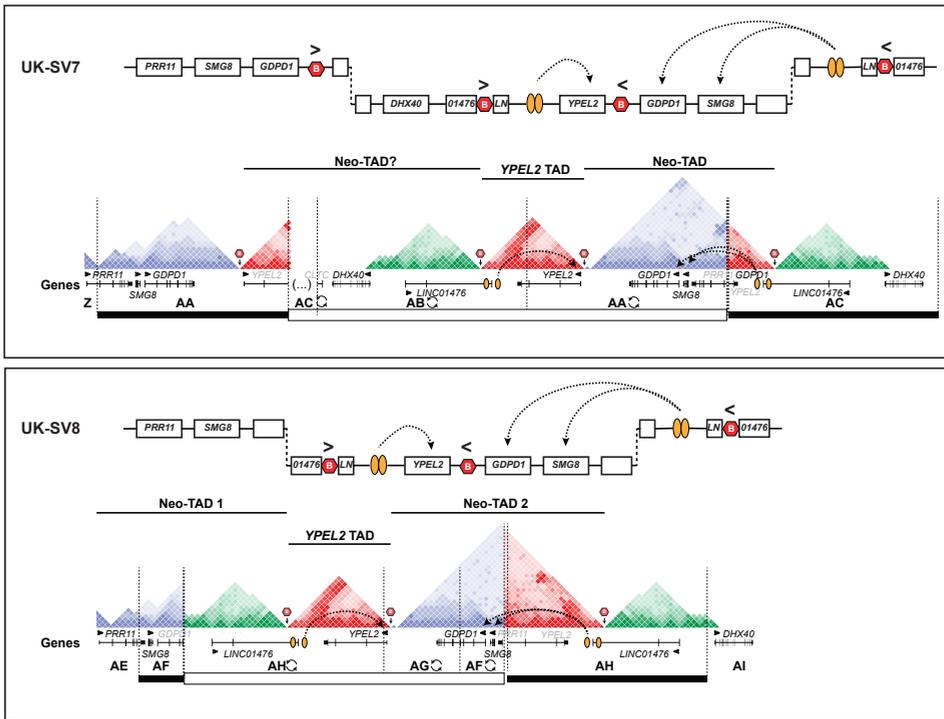


Figure S9. RP17-SVs are predicted to disrupt 3D chromatin organisation and create neo-TADs with ectopic retinal enhancer-gene contacts. Modelling of TAD boundaries, CTCF site orientation, gene position and orientation and retinal-specific enhancers for each unique RP17-SV is shown. Wildtype chromatin organisation is depicted schematically, based on Hi-C maps. Schematic models of the genome architecture for each RP17-SV are shown above Hi-C map models (dotted vertical lines represent SV breakpoints). Shaded bars represent duplicated (black) or triplicated (grey) regions, whereas inversions are indicated by open bar below the TAD maps, with nomenclature corresponding to those described in **Figure 2**. In all RP17-SVs, new domains (neo-TADs) are created with ectopic contacts between retinal-specific enhancers and *GDPD1*. For NL-SV1, NL-SV5 and UK-SV6, an extra copy of *YPEL2* is also introduced into the neo-TAD. For UK-SV2, SA-SV3, CA-SV4, NL-SV5, UK-SV7 and UK-SV8, one copy of *SMG8* is introduced into the neo-TAD.

SUPPLEMENTARY REFERENCES

1. den Hollander, A.I., van der Velde-Visser, S.D., Pinckers, A.J.L.G., Hoyng, C.B., Brunner, H.G. & Cremers, F.P.M. Refined mapping of the gene for autosomal dominant retinitis pigmentosa (RP17) on chromosome 17q22. *Human Genetics* **104**, 73-76 (1999).
2. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
3. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
4. Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423-425 (2012).
5. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26 (2011).
6. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
7. Fiorentino, A., Fujinami, K., Arno, G., Robson, A.G., Pontikos, N., Arasanz Armengol, M. *et al.* Missense variants in the X-linked gene PRPS1 cause retinal degeneration in females. *Human Mutation* **39**, 80-91 (2018).
8. Desvignes, J.-P., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Bérout, C. *et al.* VarAFT: a variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Research* **46**, W545-W553 (2018).
9. Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S. *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **28**, 2747-2754 (2012).
10. Roller, E., Ivakhno, S., Lee, S., Royce, T. & Tanner, S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* **32**, 2375-2377 (2016).
11. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).
12. Fiorentino, A., Yu, J., Arno, G., Pontikos, N., Halford, S., Broadgate, S. *et al.* Novel homozygous splicing mutations in ARL2BP cause autosomal recessive retinitis pigmentosa. *Molecular Vision* **24**, 603-612 (2018).
13. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539 (2011).
14. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>.

15. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**, e45-e45 (2001).
16. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. *et al.* The Human Genome Browser at UCSC. *Genome Research* **12**, 996-1006 (2002).
17. Cherry, T.J., Yang, M.G., Harmin, D.A., Tao, P., Timms, A.E., Bauwens, M. *et al.* Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. *Proceedings of the National Academy of Sciences* **117**, 9001-9012 (2020).
18. Ram, O., Goren, A., Amit, I., Shores, N., Yosef, N., Ernst, J. *et al.* Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* **147**, 1628-1639 (2011).
19. Sabo, P.J., Hawrylycz, M., Wallace, J.C., Humbert, R., Yu, M., Shafer, A. *et al.* Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16837-16842 (2004).
20. Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98 (2012).
21. Sangermano, R., Bax, N.M., Bauwens, M., van den Born, L.I., De Baere, E., Garanto, A. *et al.* Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461-10T C mutation in Stargardt disease. *Ophthalmology* **123**, 1375-1385 (2016).
22. Schwarz, N., Carr, A.-J., Lane, A., Moeller, F., Chen, L.L., Aguilà, M. *et al.* Translational read-through of the RP2 Arg120stop mutation in patient iPSC-derived retinal pigment epithelium cells. *Human Molecular Genetics* **24**, 972-986 (2015).
23. Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N.M., Hoyng, C.B. *et al.* Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. *American Journal of Human Genetics* **102**, 517-527 (2018).
24. Gonzalez-Cordero, A., Kruczek, K., Naeem, A., Fernando, M., Kloc, M., Ribeiro, J. *et al.* Recapitulation of human retinal development from human pluripotent stem cells generates transplantable populations of cone photoreceptors. *Stem Cell Reports* **9**, 820-837 (2017).
25. de Bruijn, S.E., Verbakel, S.K., de Vrieze, E., Kremer, H., Cremers, F.P.M., Hoyng, C.B. *et al.* Homozygous variants in KIAA1549, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa. *Journal of Medical Genetics* **55**, 705 (2018).
26. Lukowski, S.W., Lo, C.Y., Sharov, A.A., Nguyen, Q., Fang, L., Hung, S.S. *et al.* A single-cell transcriptome atlas of the adult human retina. *The EMBO Journal* **38**, e100811 (2019).
27. Peng, Y.-R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G.S. *et al.* Molecular classification and comparative taxonomics of foveal and peripheral cells in primate retina. *Cell* **176**, 1222-1237 (2019).
28. Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C.C. *et al.* Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Research* **47**, D752-D758 (2018).
29. Astuti, G.D.N., van den Born, L.I., Khan, M.I., Hamel, C.P., Bocquet, B., Manes, G. *et al.* Identification of inherited retinal disease-associated genetic variants in 11 candidate genes. *Genes* **9**, 21 (2018).

30. Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D. & Lupski, J.R. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nature Genetics* **41**, 849 (2009).
31. Sen, S.K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P.A. *et al.* Human genomic deletions mediated by recombination between alu elements. *American Journal of Human Genetics* **79**, 41-53 (2006).
32. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database: The Journal of Biological Databases and Curation* **2017**, bax028 (2017).



Chapter 5

Exploring the missing heritability in subjects with hearing loss, enlarged vestibular aqueducts, and a single or no pathogenic *SLC26A4* variant

Jeroen J. Smits*, **Suzanne E. de Bruijn***, Cornelis P. Lanting, Jaap Oostrik, Luke O’Gorman, Tuomo Mantere, DOOFNL Consortium, Frans P. M. Cremers, Susanne Roosing, Helger G. Yntema, Erik de Vrieze, Ronny Derks, Alexander Hoischen, Sjoert A.H. Pegge, Kornelia Neveling, Ronald J.E. Pennings, and Hannie Kremer

* These authors contributed equally to this work

Human Genetics, in press

ABSTRACT

Pathogenic variants in *SLC26A4* have been associated with autosomal recessive hearing loss (arHL) and a unilateral or bilateral enlarged vestibular aqueduct (EVA). *SLC26A4* is the second most frequently mutated gene in arHL. Despite the strong genotype-phenotype correlation, a significant part of cases remains genetically unresolved. In this study, we investigated a cohort of 28 Dutch index cases diagnosed with HL in combination with an EVA but without (M0) or with a single (M1) pathogenic variant in *SLC26A4*. To explore the missing heritability, we first determined the presence of the previously described EVA-associated haplotype (Caucasian EVA (CEVA)), characterized by 12 single nucleotide variants located upstream of *SLC26A4*. We found this haplotype and a delimited V1-CEVA haplotype to be significantly enriched in our M1 patient cohort (10/16 cases). The CEVA haplotype was also present in two M0 cases (2/12). Short- and long-read whole genome sequencing and optical genome mapping could not prioritize any of the variants present within the CEVA haplotype as the likely pathogenic defect. Short-read whole genome sequencing of the six M1 cases without this haplotype and the two M0/CEVA cases only revealed previously overlooked or misinterpreted splice-altering *SLC26A4* variants in two cases, who are now genetically explained. No deep-intronic or structural variants were identified in any of the M1 subjects. With this study, we have provided important insights that will pave the way for elucidating the missing heritability in M0 and M1 *SLC26A4* cases. For pinpointing the pathogenic effect of the CEVA haplotype, additional analyses are required addressing defect(s) at the RNA, protein, or epigenetic level.

INTRODUCTION

SLC26A4 encodes the transmembrane anion transporter pendrin and is most abundantly expressed in the inner ear, thyroid gland, kidney, and airways epithelia.¹⁻⁵ The 780 amino acid protein is part of the solute carrier family 26 and plays a pivotal role in chloride, bicarbonate and iodine transport. In the inner ear, pendrin functions as a $\text{Cl}^-/\text{HCO}_3^-$ exchanger. The protein is expressed in the epithelial cells of the cochlea (outer sulcus and spindle cells), the vestibular labyrinth (transitional cells), and the endolymphatic duct and sac (mitochondrial-rich cells).^{6,7} Expression of pendrin is essential for the development of the (murine) auditory and vestibular system and for maintaining ion homeostasis in the endolymphatic fluid and the endocochlear potential.^{2,7-9}

Defects in *SLC26A4* are among the most frequent causes (up to 10%) of early-onset autosomal recessive hearing loss (arHL); non-syndromic DFNB4 (MIM: 600791) and Pendred syndrome (MIM: 274600).¹⁰ Individuals carrying biallelic pathogenic *SLC26A4* variants are affected by variable, often progressive and predominantly sensorineural HL with a congenital or childhood-onset.^{11,12} In Pendred syndrome, the HL phenotype is accompanied by an iodine organification defect that can lead to thyroid goiter.¹³ In individuals affected by either syndromic or non-syndromic *SLC26A4*-associated HL, a unilateral or bilateral enlarged vestibular aqueduct (EVA) is observed, which is the most common imaging abnormality in individuals with HL.^{14,15} In some cases, EVA can be part of Mondini dysplasia: an inner ear malformation that includes both EVA and cochlear incomplete partition type II. Although Mondini dysplasia can be observed in both Pendred syndrome and DFNB4 cases, cases with the syndromic type of HL are more likely to present Mondini dysplasia than those with non-syndromic HL.^{16,17}

Pathogenic variants in *SLC26A4* have a loss-of-function effect, leading to malfunctioning of the pendrin ion transporter. Besides the antenatal formation of an EVA, this ultimately leads to acidification of the endolymphatic fluids in the inner ear during embryonic development.^{7,18} Although the exact molecular pathogenic mechanism remains to be elucidated, the lack of pendrin function ultimately leads to degeneration of the sensory cells in the inner ear.⁷

Despite the strong association between defects of *SLC26A4* and HL combined with an EVA, genetic screening of subjects with this combination of defects often does not reveal biallelic pathogenic variants in *SLC26A4* (coined M2). Cohort studies report that 14-31% of the subjects with an EVA and HL carry a monoallelic pathogenic variant in *SLC26A4* (M1), whereas in 10-65% of the subjects, no potentially pathogenic variant in the coding or splice site regions of the gene can be identified (M0).^{16,19,20} Segregation

analyses performed in family members of M1 subjects, however, do suggest that in 98% of M1 subjects an unidentified or unrecognized variant is present on the *trans SLC26A4* allele.^{20,21} In line with this hypothesis, Chattaraj and coworkers reported a haplotype, referred to as the Caucasian EVA (CEVA) haplotype, that was present in 13 of 16 (81%) of the studied M1 families and that was also enriched in M0 subjects.²² The haplotype is defined by the combination of 12 single nucleotide polymorphisms (SNPs; allele frequency (AF) 1.9-4.0%) spanning a 613 kb region. The 12 SNPs are located within a region of linkage disequilibrium that extends from upstream of *PRKAR2B* to intron 3 of *SLC26A4* and are either intergenic or intronic of the genes *SLC26A4*, *BCAP29*, *DUS4L*, *COG5*, *GPR22*, *HBP1*, *PRKAR2B* and *PIK3CG*.²² The true genetic defect of the CEVA allele has not been identified yet, but it cannot be excluded that a potential defect was missed due to the technical limitations of short-read sequencing and other standard-of-care tests. The CEVA haplotype was reported to be associated with a less severe HL phenotype as compared to variants in the protein-coding or splice site regions of *SLC26A4*.²³

We investigated a Dutch cohort of M1 and M0 subjects with HL and a unilateral or bilateral EVA. All subjects were tested for the presence of the CEVA haplotype, and whole genome sequencing (WGS) was performed to detect potentially missed single nucleotide variants (SNVs), structural variants (SVs), and regulatory or deep-intronic variants. Long-read sequencing and optical genome mapping were performed to reveal a potentially missed SV located on the CEVA haplotype. Variants located within the haplotype were subjected to *in silico* analyses to investigate potential effects on the regulation of *SLC26A4* expression or on splicing. With this study, we provided further insights into *SLC26A4*-associated disease.

MATERIAL AND METHODS

Inclusion criteria and clinical evaluation

This study was approved by the medical ethics committee of the Radboud University Medical Center (registration number: NL33648.091.10) and was carried out according to the Declaration of Helsinki. Subjects diagnosed with unilateral or bilateral HL and a unilateral or bilateral EVA on CT or MRI and for whom medical genetic testing only revealed a heterozygous (M1, n=16) or no pathogenic variant (M0, n=12) in *SLC26A4* were eligible to participate in this study. A retrospective cohort of nine subjects with confirmed pathogenic (biallelic) variants in *SLC26A4* was added as a reference cohort (**Table S1**).

Medical history was taken from all participants and special attention was paid to non-genetic causes of HL. Results of pure tone, speech, and brainstem evoked response audiometry, performed in a sound-attenuated booth, were collected. Air and bone conduction pure tone thresholds were determined for frequencies ranging from 0.25 to 8 kHz. Threshold estimates based on brainstem evoked response audiometry were used when pure tone audiometry was not available. Individuals were considered affected when pure tone thresholds for at least three frequencies were above the frequency-specific 95th percentile of age- and sex-specific thresholds (ISO 7029:2017) for the best hearing ear. In the Netherlands, routine newborn hearing screening is carried out by the detection of transient evoked otoacoustic emissions.²⁴ When available, these data were used to determine whether the HL was congenital.

Previously performed CT and MRI scans were retrieved and reassessed by an experienced neuroradiologist (S.A.H.P.). An EVA was defined as a vestibular aqueduct that measured ≥ 2 mm at the operculum and/or ≥ 1 mm at the midpoint²⁵, in accordance with previously published reports on this topic.^{22,23} Analyses of pair-wise differences between patient groups were performed with R (R Foundation) using multivariate linear regression analysis (using *lsmeans* 2.3.0) with a correction for multiple comparisons using the Holm method.²⁶

Next generation sequencing and variant interpretation

Genomic DNA was isolated from peripheral blood lymphocytes and analyzed by molecular inversion probe (MIP) sequencing, whole exome sequencing (WES) or whole genome sequencing (WGS) (**Table S2**). For WES, exome enrichment was performed using the Agilent SureSelect Human All Exome V4 or V5 kits according to the manufacturer's instructions. Subsequently, sequencing was executed on an Illumina HiSeq system by BGI Europe (Copenhagen, Denmark), with a minimal coverage of 20x for 93.77% of the targets and an average coverage of >100 reads. Read mapping along the hg19 reference genome (GRCh37/hg19) and variant calling was performed using BWA V.0.78²⁷ and GATK HaplotypeCaller V.3.3²⁸, respectively. An in-house developed pipeline was used for variant annotation and copy number variant (CNV) detection was performed using CoNIFER V.0.2.2.3²⁹. WGS was performed by BGI (Hongkong, China) on a BGISeq500 using a 2x 100 bp paired end module, with a minimal median coverage of 30-fold per genome. Read mapping (GRCh37/hg19) and variant calling was performed as described for WES. Structural variants (SVs) were called using the Manta Structural Variant Caller V.1.1.0 (SV detection based on paired end and split read evidence)³⁰ and CNVs using Control-FREEC (CNV detection based on alterations in read depth.³¹ MIP design, sequencing and data analysis were performed as previously described.^{32,33}

MIPs were designed to cover exons and exon-intron boundaries of a panel of 120 HL genes (**Table S3**). For each targeted region an average coverage of >500 reads was obtained. A minimal coverage of 20x was reached for 91.78% of the MIPs. CNV detection for *SLC26A4* was performed using a read coverage analysis as previously described.³⁴ Additionally, coding and splice site regions of *FOXI1* and the regions harboring reported pathogenic variants in *EPHA2* were sequenced using Sanger sequencing as previously described³⁵, since these genes are not included in the MIP panel. Primer sequences and PCR conditions are available upon request.

Variant prioritization was based on an AF of $\leq 0.5\%$ (gnomAD V2.1.1³⁶ and our in-house exome database (~15,000 alleles)), unless specified otherwise. Variant visualization was performed using the IGV software V.2.4 (Broad Institute).³⁷ Interpretation of missense variants was performed using the *in silico* tools CADD-PHRED (≥ 15)³⁸, SIFT (≤ 0.05)³⁹, PolyPhen-2 (≥ 0.450)⁴⁰ and MutationTaster (deleterious)⁴¹ to predict potentially deleterious effects. Variants were prioritized if a deleterious effect was predicted by at least two of these tools. Candidate variants were validated by Sanger sequencing and segregation analysis was performed when DNA of family members was available. Primer sequences and PCR conditions are available upon request. Potential effects on splicing of missense, synonymous and intronic variants were assessed using the deep-learning splice prediction algorithm SpliceAI (≥ 0.1).⁴² The maximum distance between the variant and potential gained or lost splice sites was set to 1000 bp. Predicted splice altering defects were evaluated using an *in vitro* splice assay in HEK293T cells as previously described.⁴³

Detection of the CEVA haplotype

Initial identification of the CEVA haplotype²² was performed with SNP-genotyping by Sanger sequencing in index cases for whom parental DNA was available for segregation analysis. Subsequently, the corresponding VNTR marker haplotype was determined in CEVA-positive families. For additional cases, VNTR marker analysis was performed to enable a fast and cost-effective detection of the CEVA haplotype. For the VNTR marker analysis, DNA segments were amplified by employing touchdown PCR, and subsequent analysis was carried out on an ABI Prism 3730 Genetic Analyzer (Applied Biosystems). Genomic positions of the markers were determined using the UCSC genome browser (GRCh37/hg19).⁴⁴ Alleles were assigned with the GeneMarker software (V.2.6.7, SoftGenetics) according to the manufacturer's instructions. When an individual was suspected of carrying the CEVA haplotype based on VNTR-marker alleles,

SNP genotyping by Sanger sequencing was performed to confirm the presence of the twelve SNPs that are located within the haplotype.²² SNP-phasing was performed if DNA samples of family members was available.

Optical genome mapping

Optical genome mapping (Bionano Genomics) was performed as previously described.^{45,46} Ultra-high molecular weight DNA was isolated from whole peripheral blood (collected in EDTA tubes) using the SP Blood & Cell Culture DNA Isolation Kit (Bionano Genomics). CTTAAG labeling was performed using the DLS (Direct Label and Stain) DNA Labeling Kit (Bionano Genomics) and the labeled sample was analyzed using a 3x 1,300 Gb Saphyr chip (G2.3) on a Saphyr instrument (Bionano Genomics). An effective coverage of 124x was reached, with a label density of 14.63/100 kb and an average N50 of 279 kb. *De novo* assembly (using GRCh37 and GRCh38) and variant annotation were performed using Bionano Solve version 3.4, which includes two separate algorithms for SV and CNV detection. Annotated variants were filtered for rare events as described previously.⁴⁵ In addition, the genomic region spanning the CEVA haplotype was analyzed visually in Bionano Access version 1.4.3.

PacBio long-read sequencing

Genomic DNA was isolated from peripheral blood according to standard procedures and subjected to long-read genome Hi-Fi sequencing using the SMRT sequencing technology (Pacific Biosciences). Library preparation was performed using the SMRTbell™ Template Prep Kit 2.0 (Pacific Biosciences) following manufacturer's instructions. Size selection was performed using a BluePippin DNA size selection system (target fragments ~15-18 kb). Sequence primer V2 and polymerase 2.0 were used for binding. Subsequently, the SMRTbell library was loaded on an 8M SMRTcell and sequencing was performed on a Sequel II system (Pacific Biosciences). Circular consensus sequencing (CCS), Hi-Fi reads, were generated using the CCS (v4.2.0) tool and were aligned to the GRCh37/hg19 reference genome with pbmm2 (v.1.3.0). The unique molecular yield was 93.46 Gb and the post-alignment Hi-Fi- coverage was 12x (Mosdepth v0.3.1⁴⁷). SV calling was performed using PBSV (v2.4.0) and annotation was applied using an in-house SV annotation pipeline.

RESULTS

Patient inclusion and genetic prescreening

In this study, we included 28 Dutch index cases diagnosed with a unilateral or bilateral EVA and unilateral or bilateral HL. All individuals were prescreened for pathogenic variants in *SLC26A4* (NM_000441.1) in a diagnostic setting and complete coverage of the coding and splice site (+/- 14 nucleotides) regions of *SLC26A4* was confirmed. In 16 individuals, a heterozygous (likely) pathogenic *SLC26A4* variant was reported and these cases were deemed M1. In the remaining 12 individuals, no potentially pathogenic variants were found in the coding or splice site regions of this gene, and these subjects were therefore considered M0. Causative variants in other genes associated with arHL⁴⁸ were addressed and excluded by analyzing available sequencing data (WES or MIPs-based) or in WGS data obtained in this study (**Table S2**). This revealed no homozygous or compound heterozygous variants that were known or predicted to be pathogenic, except two compound heterozygous variants in *OTOGL* (NM_173591.3) in individual SLC012 (**Table S4**). The c.890C>T (p.(Pro297Leu)) variant in *OTOGL* has, however, been reported as (likely) benign in ClinVar⁴⁹ and the Deafness Variation Database⁵⁰ and is classified as likely benign according to the ACMG guidelines.⁵¹ The c.1369G>T (p.(Val457Leu)) is considered as a variant of unknown significance (ACMG classification). Furthermore, subject SLC012 has progressive high-frequency HL, which differs from the symmetric, moderate, and stable HL associated with *OTOGL* (DFNB84B).^{52,53} Therefore, we considered the identified *OTOGL* variants as non-causative. For none of the cases, (likely) pathogenic variants (UV4/UV5, ClinVar) were identified in genes associated with autosomal dominant HL or syndromic HL.⁴⁸

The CEVA haplotype is enriched in Dutch monoallelic *SLC26A4* cases

In 2017, Chattaraj et al. described the ≥ 613 -kb CEVA haplotype located centromeric of the *SLC26A4* gene to be enriched in M1 *SLC26A4* cases and M0 cases with HL and EVA.²² To investigate whether this haplotype is also enriched in the selected Dutch cohort of M0 and M1 *SLC26A4* cases, we screened for the presence of this haplotype using VNTR marker analysis followed by Sanger sequencing of the 12 CEVA-associated SNPs. The CEVA haplotype was detected in 8 out of 16 (50%) M1 individuals and 2 out of 12 (16.7%) M0 subjects (**Figure 1, Table 1**). In two additional M1 individuals (SLC040 and SLC071), only a partial CEVA haplotype was found, harboring 9/12 SNPs. We will refer to this smaller haplotype as the variant 1-CEVA (V1-CEVA) haplotype.

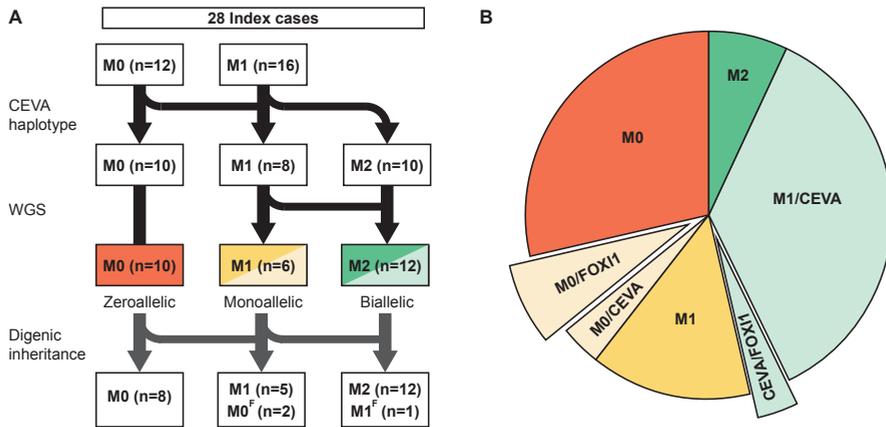


Figure 1. Overview of genetic analyses performed in zeroallelic and monoallelic *SLC26A4* cases. (A-B) To explain the missing heritability in zeroallelic (M0, n=12) and monoallelic (M1, n=16) *SLC26A4* cases, different genetic analyses were performed. Firstly, individuals were screened for the presence of the CEVA haplotype (M0/CEVA, n=2; M1/CEVA, n=10). Secondly, whole genome sequencing (WGS) was performed in all monoallelic cases (M0/CEVA, M1) to identify potential structural, splice (M2, n=2) or regulatory variants. Lastly, sequencing data were screened for potentially pathogenic variants in the *EPHA2*, *FOX11* and *KCNJ10* genes. Digenic inheritance has been previously suggested for variants in these genes and the *SLC26A4* gene. In three cases (M0/FOX11 (M0^f), n=2, CEVA/FOX11 (M1^f), n=1), a potentially pathogenic variant in *FOX11* (NM_012188.4, c.677C>T) was identified.

The CEVA haplotype has an AF of 2.8% in the 1000G database (28 in 1006 alleles)^{22,54}, and an AF of 3.3% in an in-house control cohort consisting of 322 healthy unrelated individuals (21 in 644 unphased alleles). This implies a significant enrichment of the CEVA haplotype in our M1 cohort (8 in 32 alleles) compared to the 1000G database (p-value 5.419×10^{-6}) and the control cohort (p-value 2.187×10^{-5}) as determined by a two-sided Fisher's exact test). The two M1 cases with the V1-CEVA haplotype were not included in this statistical analysis. Also this V1-CEVA allele is significantly enriched in our M1 cohort as only a single V1-CEVA allele is reported in the 1000G database (1 in 1006 alleles)²² (p-value 0.0027). The CEVA haplotype was not found to be significantly enriched in the M0 cohort (2 in 24 alleles). Although the pathogenicity of the CEVA haplotype is unclear, the significant enrichment of the haplotype within this M1 patient cohort and the patient cohorts (M1 and M0) previously described by Chattaraj and co-workers strongly suggests that a pathogenic defect resides within this haplotype.²² Because of this strong association of the CEVA haplotype with HL and EVA, we considered the M1 individuals carrying the CEVA or the V1-CEVA haplotype as genetically explained (M1/CEVA), and M0 individuals with the CEVA haplotype (M0/CEVA) as monoallelic in further steps of this study. For six M1 individuals, it could not be conclusively determined whether the CEVA haplotype was present *in trans* with the pathogenic *SLC26A4* variant, as the genetic material of family members was not available (**Table 1**).

Table 1. Detection of the CEVA haplotype in M1 and M0 individuals

Case	Allele 1	Allele 2	
	Variant	ACMG	CEVA
Zeroallelic SLC26A4 cases			
SLC014	c.2059G>T; p.(Asp687Tyr)	UV3	
SLC015	-	-	
SLC017	-	-	
SLC039	-	-	<u>ACACATG-GC-C</u> (CEVA)
SLC043	-	-	
SLC052	-	-	
SLC069	-	-	
SLC070	-	-	
SLC073	-	-	
SLC080	-	-	<u>ACACATG-GC-C</u> (CEVA)
SLC084	-	-	
SLC086	-	-	
Monoallelic SLC26A4 cases			
SLC002	c.412G>T; p.(Val138Phe)	UV5	
SLC003	c.131 dup; p.(Thr45Aspfs*42)	UV5	<u>ACACATG-GC-C</u> (CEVA)
SLC012*	c.707T>C; p.(Leu236Pro)	UV5	<u>ACACATG-GC-C</u> (CEVA)
SLC013	c.1001+1G>A; p.(?)	UV5	<u>ACACATG-GC-C</u> (CEVA)
SLC018	c.349C>T; p.(Leu117Phe)	UV5	
SLC031	c.1001+1G>A; p.(?)	UV5	<u>ACACATG-GC-C</u> (CEVA)
SLC032	c.1334T>G; p.(Leu445Trp)	UV5	
SLC036*	c.1246A>C; p.(Thr416Pro)	UV5	<u>ACACATG-GC-C</u> (CEVA)
SLC040*	c.655_656dup; p.(Phe223Alafs*15)	UV5	<u>GTTTCATG-GC-C</u> (V1-CEVA)
SLC045	c.1334T>G; p.(Leu445Trp)	UV5	
SLC048	c.706C>G; p.(Leu236Val)	UV4	
SLC056	c.707T>C; p.(Leu236Pro)	UV5	<u>ACACATG-GC-C</u> (CEVA)
SLC071*	c.1334T>G; p.(Leu445Trp)	UV5	<u>GTTTCATG-GC-C</u> (V1-CEVA)
SLC078	c.304G>C; p.(Gly102Arg)	UV4	<u>ACACATG-GC-C</u> (CEVA)
SLC079	c.1001+1G>A; p.(?)	UV5	<u>ACACATG-GC-C</u> (CEVA)
SLC085	c.706C>G; p.(Leu236Val)	UV4	

Presence of the CEVA haplotype was tested in zeroallelic (M0) and monoallelic (M1) SLC26A4 cases with a unilateral or bilateral enlarged vestibular aqueduct. SLC26A4 (NM_000441.1) variants reported in ClinVar as (likely) pathogenic (UV4, UV5) were considered causative, whereas variants reported as (likely) benign or of unknown significance were considered non-causative. In ten individuals, the complete CEVA haplotype was detected (ACACATG-GC-C), whereas in two individuals a shorter version of the haplotype was found, consisting of 9/12 CEVA SNPs (GTTTCATG-GC-C; V1). For individuals marked with an *, it could be conclusively determined that the (V1)-CEVA haplotype is present on the trans SLC26A4 allele. ACMG, variant classification according to the American College of Medical Genetics and Genomics (ACMG) classification guidelines⁵¹; UV3, uncertain significance; UV4, likely pathogenic; UV5, pathogenic.

Whole genome sequencing reveals potential *SLC26A4* splice and regulatory variants in M1 subjects without the CEVA haplotype

To detect any potentially missed coding or unidentified intronic *SLC26A4* variants or variants located *in cis* regulatory elements of the gene, WGS analysis was performed for all six M1 individuals who could not be genetically explained by the presence of the CEVA haplotype. Additionally, WGS analysis was performed for the two M0/CEVA individuals. In none of these eight cases, SVs overlapping with the *SLC26A4* gene were identified by WGS.

To identify any variants with a potential effect on splicing, the deep-learning algorithm SpliceAI was employed.⁴² In two M1 individuals (SLC048 and SLC085), a rare heterozygous potentially splice altering *SLC26A4* variant was identified (**Table 2**). For both variants, the predicted splice defect was investigated using an *in vitro* splice assay performed in HEK293T cells. For SLC048, a canonical splice site variant (c.1342-2A>C), that was overlooked during prescreening efforts, was predicted to remove the splice acceptor site. This variant was previously reported in a study performed by Van Beeck Calkoen and coworkers and in ClinVar.⁵⁵ Indeed, the splice assay revealed loss of the acceptor site and usage of an alternative splice acceptor site located thirteen nucleotides downstream (**Figure S1A**). This leads to the formation of an out-of-frame exon 12 and premature protein truncation (p.(Ser448Leufs*3)). Based on these results, the variant was classified as pathogenic according to the ACMG guidelines.⁵¹ In SLC085, a synonymous variant (c.471C>T, classified as likely benign in ClinVar) was identified in exon 5. SpliceAI predicts that this variant strengthens an alternative splice acceptor site (27 nucleotides downstream of the variant). Indeed, an *in vitro* splice assay confirmed that the alternative splice acceptor site is used, which leads to the partial deletion of exon 5 and a truncated protein (p.(Gly139Aafs*6)) (**Figure S1B**). Therefore, this variant is now classified as pathogenic according to the ACMG classification guidelines.⁵¹ The observed splice defect resulting from this synonymous variant underlines the importance of evaluating potential splice effects of all rare variants in coding sequences, using *in silico* prediction splice tools. We considered the two identified splice variants as pathogenic and the HL of the two individuals as genetically explained, thus M2.

To explore variants that are potentially located within a *cis* regulatory element of *SLC26A4*, we extracted all (predicted) human enhancer and promoter elements that are associated with the *SLC26A4* gene from the GeneHancer⁵⁶ and EnhancerAtlas⁵⁷ databases (**Table S5**). GeneHancer V5 is a collection of both predicted and experimentally validated enhancer-to-gene and promoter-to-gene interactions, based on information integrated from multiple resources: ENCODE⁵⁸, Ensembl⁵⁹, FANTOM⁶⁰, VISTA⁶¹, dbSuper⁶², EPDnew⁶³,

UCNEbase⁶⁴ and CraniofacialAtlas⁶⁵. For each regulatory element, a gene interaction score (>7) and element confidence score (>0.7) are provided. The EnhancerAtlas V2 is a database providing enhancer annotations in different species based on experimental datasets determined in several tissues and cell types.

WGS data were analyzed for variants located within these elements and two rare potentially regulatory variants (Chr7:107220628C>A, Chr7:107384987C>G) were identified in two M1 individuals (SLC002 and SLC045) (**Table S6**). Both variants are located in a predicted enhancer element of *SLC26A4* according to GeneHancer. We did not find any strong indication of a functional effect for the two variants based on (nucleotide) conservation scores (PhyloP, UCSC genome browser⁴⁴) or loss of transcription factor binding sites (JASPAR database⁶⁶). Therefore, the variants were considered non-pathogenic, although only a reporter assay can completely exclude a potential regulatory effect of the variants on *SLC26A4* expression.

A *FOXI1* missense variant is revealed in three unrelated index cases

Several studies have suggested a potential digenic inheritance for *SLC26A4* variants and variants in *KCNJ10* and *FOXI1*.⁶⁷⁻⁶⁹ Additionally, a more recent study suggested digenic inheritance with pathogenic variants in *EPHA2*.⁷⁰ We screened all remaining genetically unexplained individuals (M1, M0/CEVA and M0) for variants in these genes with an AF ≤5% (gnomAD V2.1.1). In cases for which only MIP sequencing data was available, coding regions and exon-intron boundaries of *FOXI1* and the regions harboring the reported pathogenic variants in *EPHA2* (c.1063G>A; p.(G355R), c.1532C>T; (p.T511M), NM004431.4) were analyzed using Sanger sequencing. In three individuals (SLC039; M0/CEVA, SLC052; M0 and SLC069; M0) a c.677C>T (p.(Thr226Ile)) *FOXI1* (NM_012188.4) missense variant was identified (**Table 3**).

Table 2. WGS revealed two heterozygous splice variants in SLC26A4

Case	Class	Genome	cDNA	Protein	In-house AF (%)	gnomAD AF (%)	CADD_PHRD	SIFT	PPH2	Mutation Taster	SpliceAI	ACMG
SLC048	M1	Chr7:107335064A>C	c.1342-2A>C	p.Ser448Leufs*3	0.00	-	21.7	NA	NA	NA	0.99 (AS loss)	UV5
SLC085	M1	Chr7:107314664C>T	c.471C>T	p.Gly139Alafs*6,=	-	0.00	0.725	NA	NA	NA	0.59 (AS gain)	UV5

Whole genome sequencing (WGS) revealed two potentially splice altering variants in SLC26A4. Variants are selected based on an allele frequency of $\leq 0.5\%$ in gnomAD and the in-house database. Scores that meet the thresholds for pathogenicity as described in the methods section are indicated in red. The predicted effect on splicing was confirmed in an in vitro splice assay that was performed in HEK293T cells (Figure S1). Genome; Genomic position according to GRCh37/hg19; In-house AF, allele frequency (%) in an in-house database ($\sim 7,500$ exomes); gnomAD AF, allele frequency (%) in gnomAD database V2.1.1; CADD_PHRD, Combined Annotation Dependent Depletion PHRED score; SIFT, Scale-Invariant Feature Transform; PPH2, PolyPhen-2 score; MutationTaster (prob), MutationTaster score with probability (0-1); spliceAI, splicing prediction score; AS, acceptor site; ACMG, variant classification according to the American College of Medical Genetics and Genomics (ACMG) classification guidelines⁵¹; UV5, pathogenic; NA, not applicable.

Table 3. Rare variants identified in EPHA2, FOXI1 and KCNJ10

Case	Class	Gene	Transcript	cDNA	Protein	In-house AF (%)	gnomAD AF (%)	CADD_PHRD	SIFT	PPH2	Mutation Taster	SpliceAI	ACMG
SLC017	M0	EPHA2	NM_004431.4	c.2627G>A	p.(Arg876His)	2.36	1.70	32	0	0.769	NA	0.03	UV2
SLC039	M0/CEVA	FOXI1	NM_012188.4	c.677C>T	p.(Thr226Ile)	0.56	0.37	11	0.14	0.109	P	0.03	UV2
SLC052	M0	EPHA2	NM_004431.4	c.1941G>T	p.(Thr647=)	1.09	0.55	7.309	NA	NA	NA	0.05	UV2
SLC052	M0	EPHA2	NM_004431.4	c.1896G>A	p.(Leu632=)	0.76	0.05	3.197	NA	NA	NA	0.05	UV2
SLC052	M0	FOXI1	NM_012188.4	c.677C>T	p.(Thr226Ile)	0.56	0.37	11	0.14	0.109	P	0.03	UV2
SLC069	M0	FOXI1	NM_012188.4	c.677C>T	p.(Thr226Ile)	0.56	0.37	11	0.14	0.109	P	0.03	UV2

Available sequencing datasets of monoallelic (M1, M0/CEVA) and zeroallelic (M0) individuals were screened for variants in EPHA2, FOXI1 and KCNJ10 with an allele frequency of $\leq 5\%$ in gnomAD (V2.1.1). Scores that meet the thresholds for pathogenicity as described in the methods section are indicated in red. In-house AF, allele frequency (%) in in-house database ($\sim 7,500$ exomes); gnomAD AF, allele frequency (%) in gnomAD database V2.1.1; CADD_PHRD, Combined Annotation Dependent Depletion PHRED score; SIFT, Scale-Invariant Feature Transform; PPH2, PolyPhen-2 score; MutationTaster (prob), MutationTaster score with probability (0-1); spliceAI, splicing prediction score; ACMG, variant classification according to the American College of Medical Genetics and Genomics (ACMG) classification guidelines⁵¹; UV2, likely benign; NA, not available; P, polymorphism.

The variant was not identified in any of the M1/CEVA or the two M2 cases. *FOXI1* encodes the Forkhead transcription factor FOXI1, a key transcriptional regulator of *SLC26A4*.⁶⁹ Segregation analysis has confirmed that the *FOXI1* variant is not co-inherited with the CEVA allele in individual SLC039, which is in line with digenic inheritance. The Thr226 residue is located outside of the conserved forkhead DNA-binding domain of FOXI1 (amino acids 94–211)⁶⁹ and none of the *in silico* tools used for analysis predicted a deleterious effect of the c.677C>T variant. Nevertheless, the variant is enriched in individuals diagnosed with HL and EVA (3 in 56 alleles in the study cohort versus 165 in 26,590 alleles of the in-house WES cohort, p-value 0.0004), and we consider the c.677C>T *FOXI1* variant as an interesting candidate for functional validation.

In case SLC017, a heterozygous missense variant in *EPHA2* was detected (c.2627G>A (p.(Arg876His))). Although the variant is predicted to be pathogenic by *in silico* prediction tools, it has a relatively high AF of 1.70% (gnomAD) and 2.36% (in-house database) and is classified as likely benign according to the ACMG classification guidelines. Because the variant was only found in an M0 *SLC26A4* case, a potential digenic inheritance of pathogenic *SLC26A4* variants and the newly identified *EPHA2* variant could not be addressed.

To summarize, the CEVA haplotype or a short CEVA haplotype (V1-CEVA) was detected in 12 of the 28 index cases (16 M1, 12 M0) that were included in our study (**Figure 1**). In two individuals (M1), an *SLC26A4* splice variant was identified using WGS. After performing these genetic analyses by which the enrichment of the (V1-)CEVA haplotype in M1 cases was demonstrated, we consider the HL in 12 individuals to be associated with *SLC26A4* defects and these subjects to be genetically explained (2 M2, 10 M1/CEVA), six individuals are considered M1 (4 M1, 2 M0/CEVA), and ten individuals are still considered M0. Additionally, in three individuals (1 M0/CEVA, 2 M0) a potentially pathogenic variant in *FOXI1* was found.

Determination of boundaries of CEVA haplotype

To identify the true pathogenic defect located on the CEVA haplotype, an in-depth analysis of this genomic region was performed. Firstly, the exact boundaries of the genomic region shared by CEVA haplotype carriers were determined using VNTR marker analysis. For two individuals with the complete CEVA haplotype and the two subjects with the V1-CEVA haplotype, DNA samples of family members were available, allowing reliable determination of the marker alleles located within the haplotype. A shared haplotype of 0.89 Mb delimited by markers D7S501 and D7S2459 was identified (**Figure 2** and **Figure S2**). Although the V1-CEVA haplotype shares the marker alleles with the complete CEVA haplotype, the absence of SNPs 1-3 potentially delimits the

shared haplotype even more (0.57 Mb, CEVA SNP 3-D7S2459). The remaining eight individuals with the complete CEVA haplotype share identical marker alleles in the 0.89 Mb-sized region, although they could not be conclusively assigned to the haplotype as no segregation analysis could be performed. For individual SLC003, a deviating repeat length was identified for marker D7S2420. As we cannot exclude a rare event to be responsible for the change in allele length, this marker was still considered part of the shared CEVA haplotype.

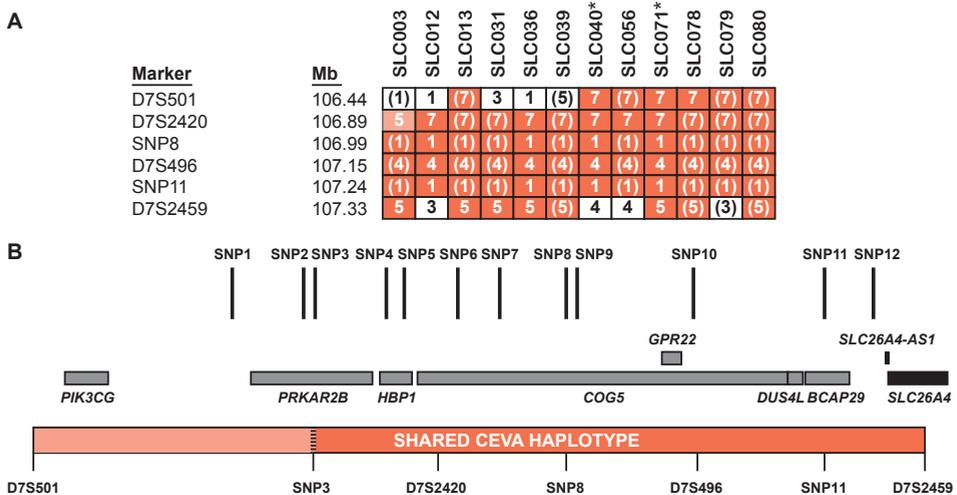


Figure 2. Determination of the boundaries of the shared CEVA haplotype. (A) The CEVA haplotype was detected in 10 individuals, in an additional 2 individuals (SLC040 and SLC071, indicated with *), a smaller haplotype was found, termed V1-CEVA. To determine the boundaries of the CEVA haplotype, VNTR marker analysis was performed. The shared haplotype (0.89 Mb, CEVA; 0.57 Mb V1-CEVA is marked in orange). For marker D7S2420 (light-orange) a deviating CA-repeat length was determined in SLC003. Nevertheless, the marker is still considered to be potentially part of the shared haplotype as a change or repeat length cannot be excluded. Genomic positions (Mb) are according to the UCSC Genome Browser (GRCh37/hg19). (B) A schematic overview of the identified shared CEVA haplotype (D7S501-D7S2459). Positions of the CEVA-associated SNPs and the genes located within the haplotype region (CEVA, D7S501-D7S2459; V1-CEVA, SNP3-D7S2459) have been indicated. All SNPs are located within intronic or intergenic regions. Genomic positions of the CEVA-associated SNPs are provided in **Table S7**. *SLC26A4* (NM_000441.1) is only partially included (exons 10/21) in the shared haplotype.

Short-read WGS did not reveal a pathogenic defect on the CEVA haplotype

Because of the significant enrichment of the CEVA haplotype in M1 cases, we hypothesized that the subjects with the CEVA haplotype share a yet elusive pathogenic defect. To identify this defect on the CEVA haplotype, short-read WGS was performed in two individuals (SLC012 & SLC036) carrying the CEVA haplotype *in trans* with a

pathogenic variant in *SLC26A4* (M1/CEVA). All heterozygous variants with an AF $\leq 5\%$ in gnomAD that were shared between the two individuals and located within the determined boundaries of the CEVA haplotype were analyzed (**Table S7**).

In total, 20 shared variants remained and included the 12 original SNPs that previously defined the CEVA haplotype.²² Sixteen of the shared variants are located in intronic regions, but for none of them, a significant effect (score ≥ 0.1) on splicing is predicted by SpliceAI. Two variants are located within a *cis* regulatory element of *SLC26A4* according to the GeneHancer database, however, these also show overlap with a long interspersed nuclear element (LINE) repeat element. One variant (CEVA SNP9) has a high nucleotide evolutionary conservation score (PhyloP, 2.769 [range -14, 3]). No SVs or CNVs were detected within or overlapping with the CEVA haplotype and shared by the two individuals.

Regions harboring heterozygous variants with an AF $\leq 5\%$ in gnomAD that were not shared between SLC012 and SLC036 had sufficient coverage to exclude that these variants were only called in one of the subjects but present in both of them. None of the variants identified in either SLC012 or SLC036 were within the *SLC26A4* gene or were obviously deleterious. SVs and CNVs within the CEVA boundaries were analyzed separately for the two subjects which did not reveal any of such variants that were not shared by the two studied subjects. To fully exclude that the CEVA haplotype harbors different pathogenic variants in the studied individuals, a study design including short- and long-read WGS in several nuclear families has to be applied.

Optical genome mapping & long-read sequencing

To investigate the possibility that SVs were missed using short-read sequencing, optical genome mapping (Bionano Genomics) was performed using ultra-high molecular weight DNA isolated from peripheral blood cells of individual SLC012 (M1/CEVA). Optical genome mapping identified a total of 6,565 SVs, of which none were within the CEVA region (D7S501-D7S2459; chr7:106,440,266-107,360,254). Two SV calls (both calling the same 2,196 bp insertion between chr7:107,367,549 and 107,373,585) were located just outside this region (**Figure S3A**). This insertion was also called in 100% of our current optical genome mapping control cohort⁷¹, strongly suggesting that this reflects a reference problem rather than a real SV. Additionally, there were 22 CNV calls, of which none were within the CEVA region.

Subsequently, PacBio long-read sequencing was performed on genomic DNA isolated from individual SLC079 (M1/CEVA; *in trans* status unknown). SV analysis of the long-read sequencing data revealed a total of 55,205 SVs, of which 12 within the CEVA

region. After careful interrogation of the SVs, all of them were considered false positives based on SV length, and presence of the SVs in an in-house control dataset. The CEVA haplotype region was also manually inspected in the IGV software, which did not reveal any indications for potential SVs (**Figure S3B**). Interestingly, the insertion event that was detected with optical genome mapping and located just outside the CEVA region was also present in the long-read sequencing data (chr7:107,370,573, 1,612 bp insertion). This insertion was also present in available in-house control sequencing data, supporting the hypothesis the variant concerns a reference problem and is not a true SV.

A comparable severity of hearing loss in the M1/CEVA and M2 cohorts

As the CEVA haplotype was reported to be associated with a less severe HL phenotype than variants in the protein-coding or splice site regions of *SLC26A4*²³, we addressed genotype-phenotype correlations in our cohort. We were able to retrieve pure tone audiometry of all subjects except for SLC071; for this subject, complete audiometry from only one ear was available (**Figure S2**). The original CT or MRI scans of subjects SLC018 and SLC032 could not be retrieved. However, written reports of the imaging were available. Data on thyroid gland function were not consistently available and were therefore not included in this study. We applied the methods of Chao et al. to compare the severity of HL between four subject groups (M0, M1, M1/CEVA, and M2 **Figure 3, Table 4**).²³ The M1/CEVA group includes the M1/V1-CEVA subjects. Bilateral EVA was present in 7 of 10 (70%) M1/CEVA subjects, in all 4 M1 subjects without the CEVA haplotype, and 7 of 10 (70%) M0 subjects without the CEVA haplotype. All 11 M2 subjects (reference cohort, SLC048 and SLC085) had bilateral EVA. The median pure tone audiometry in the M2 group (85 dB HL, n = 20) was not significantly different from that of the M1/CEVA group (84 dB HL, n = 16) and the M1 group (79.5 dB HL, n = 8) (p-values 0.8300 and 0.7142, respectively, all adjusted for age). Also, no difference was observed between the M1/CEVA group and the M1 group (p = 0.8782). In contrast, when we compare the M2 and M1/CEVA groups with the M0 group, we observed significant differences in the severity of HL (p = 0.0015 and p = 0.0135, respectively). When compared to Chao et al, subjects in our study displayed a similar degree of median HL in the M2 group (86.3 and 85 dB in (23) and the present study, respectively), more severe HL in the M1/CEVA group (47.5 and 84 dB, respectively) and less severe HL in the M0 group (54.4 and 42 dB HL, respectively). Slight age differences were seen between the groups presented in Chao et al. and those in the current study (**Table S8**). Chao and co-workers did not report audiometric data for the M1 group without the CEVA haplotype *in trans*, presumably due to the small sample size. Overall, in contrast to the study by

Chao et al., the present study showed that subjects with biallelic pathogenic variants in the coding regions and splice sites of *SLC26A4* have a degree of HL that is similar to that of subjects with a monoallelic *SLC26A4* variant and the CEVA haplotype. Due to the small sample size, we could not test the hypothesis that the CEVA haplotype acts as a modifier in M0 subjects as reported previously.²³

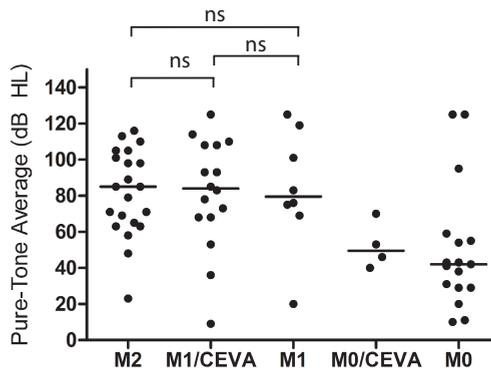


Figure 3. Results of audiometric evaluation in affected individuals. PTA_{0.5-4 kHz} for ears with an EVA. Each dot represents the hearing level of an ear with an enlarged vestibular aqueduct, allocated to genotype class (M2, M1/CEVA, M1, M0/CEVA and M0). The M1/CEVA group also includes subjects with an M1/V1-CEVA genotype. For an objective comparison, the same methods as used by Chao et al. (2019) were applied.

DISCUSSION

In this study we investigated 28 genetically unexplained Dutch index cases with HL and a unilateral or bilateral EVA. To elucidate the missing heritability in monoallelic *SLC26A4* cases, who represent 14-31% of subjects with HL and EVA^{16,20}, extensive genomic analyses as well as phenotyping were performed. Important findings in this study were (1) the enrichment of a shared (V1-)CEVA haplotype in M1 *SLC26A4* cases, (2) two *SLC26A4* splice variants and (3) the identification of a *FOXI1* variant in three subjects suggesting a contribution of this variant to the etiology of HL and EVA. Furthermore, the genotype-phenotype analyses revealed that the severity of the HL associated with biallelic variants (M2) in *SLC26A4* is comparable to the HL associated with a monoallelic variant in *SLC26A4* with or without the CEVA haplotype (M1 and M1/CEVA).

For six M1 individuals, it could not be conclusively determined whether the CEVA haplotype was present *in trans* with the (likely) pathogenic *SLC26A4* variant, as no genetic material of family members could be obtained. However, we anticipated that

most if not all of the six M1 cases carry the CEVA haplotype *in trans* with the *SLC26A4* variants because it seems highly unlikely that the *SLC26A4* variants all have occurred on an allele with a frequency of <3% in the population.²² Furthermore, the co-occurrence of two partial CEVA haplotypes that together exactly mimic a heterozygous CEVA haplotype in 6 of 16 individuals is highly unlikely as the frequencies of partial CEVA haplotypes in the European population are all (far) below 1%.²² The same holds true for the two M0/CEVA cases for whom we could not determine the phase of the 12 SNPs in the CEVA haplotype.

In two cases, the V1-CEVA haplotype was identified. This smaller CEVA haplotype was also reported previously in a single M1 case by Chattaraj and coworkers and likely refines the CEVA haplotype. Alternatively, the V1-CEVA haplotype harbors a different genetic defect. The shared VNTR marker alleles of the V1-CEVA and the CEVA haplotype suggest that V1-CEVA refines the boundaries of the shared genomic region to 0.57 Mb.

Table 4. Clinical evaluation of affected individuals

Case	Class	Gender	Age of onset (yr)	Otosopic examination	Newborn hearing screening	Motor development	Imaging		Audiometric evaluation			
							CT	MRI	Subject age (yr)	PTA (0.5-4kHz)	R	L
SLC014	M0	M	PC	N	3rd time pass	Delayed	B EVA			4	>120	55
SLC015		F	PS	myringosclerosis	N	NR		U EVA R		25	29	66
SLC017		M	PC	N	NA	NR	B EVA			5	10	59
SLC043		F	PC	N	N	NR	U EVA L			17	10	95
SLC052		F	33	N	NA	NR	B EVA			33	20	>120
SLC069		F	15	N	NA	NR	B EVA			19	43	42
SLC070		M	PC	N	NA	N		B EVA		19	43	31
SLC073		M	U	L: atelectatic middle ear, retracted eardrum	NA	NR		U EVA R		12	11	18
SLC084		F	PC	N	3rd time pass	Delayed	B EVA			6	29	38
SLC086		M	PC	N	N	Delayed	B EVA			6	54	41
SLC039	M0/CEVA	F	2-4	N	NA	NR		B EVA		24	46	70
SLC080		F	5	N	N	N	B EVA			6	40	53
SLC002	M1	F	U	N	NA	NR	B EVA			18	83	76
SLC018		M	PC	N	NA	N	B EVA*			16	75	101
SLC032		F	PC	N	NA	NR	B EVA*			59	119	>120
SLC045		F	PC	N	N	N	B EVA			7	20	69
SLC003	M1/CEVA	M	15	N	NA	NR	U EVA R			17	83	-1
SLC012		M	PC	N	NA	NR	B EVA			17	>120	53
SLC013		F	U	N	NA	NR	B EVA			22	36	9
SLC031		M	PS	N	2nd time pass	N	U EVA L			16	0	68

Table 4. Continued

Case	Class	Gender	Age of onset (yr)	Otosopic examination	Newborn hearing screening	Motor development	Imaging		Audiometric evaluation		
							CT	MRI	Subject age (yr)	R	L
SLC036		F	PC	N	NA	NR	B EVA		20	73	68
SLC040		M	U	N	U	N	U EVAL		7	5	78
SLC056		M	PC	N	NA	NR	B EVA		14	108	93
SLC071		M	PC	N	N	N	B EVA		3	85	NT
SLC078		F	PC	N	NA	U	B EVA		10	114	93
SLC079		F	C	N	R	N	B EVA		2	110	108
SLC048	M2	M	PC	N	NA	NR	B EVA		8	105	71
SLC085		M	C	N	R	N	B EVA		2	23	85
SLC087		F	C	N	R	Delayed		B EVA	5	65	63
SLC088		F	4	N	NA	N		B EVA	17	85	105
SLC089		M	U	N	U	U		B EVA	10	58	98
SLC090		F	3	N	NA	U		B EVA	41	116	110
SLC091		F	2-4	N	N	Delayed		B EVA	12	63	69
SLC092		F	C	N	R	N		B EVA	4	79	48
SLC093		M	PC	N	U	Delayed		B EVA	8	71	98
SLC094		M	PC	sclerotic eardrum	L NA	U		B EVA	37	101	113
SLC095		F	C	N	R	N		B EVA	1	NT	89

Age of onset (AoO), age of onset in years as reported by the subjects. Subject age, the age at which the audiometric data of the last two columns were obtained, in general the last audiogram. Newborn hearing screening was introduced in the Netherlands in 2006. *only written report available. Y, years; PTA, pure tone average, mean of 0.5, 1, and 4 kHz air conduction thresholds; M, male; F, female; R, right; L, left; PC, age of onset of HL is presumably congenital, based on anamnesis; C, age of onset of HL is congenital, based on newborn hearing screening; PS, subject reported onset of HL during primary school, exact age unknown; NR, not reported; NT, not tested; N, no abnormalities; R, refer in newborn hearing screening, failed in test; U, unknown; CT, computed tomography; MRI, magnetic resonance imaging; U EVA L/R, unilateral enlarged vestibular aqueduct in left or right ear; B EVA, bilateral enlarged vestibular aqueduct.

We anticipated that a pathogenic variant co-segregates with the CEVA haplotype. Therefore, we subjected the shared genomic region to extensive genomic analyses that included WES, short- and long-read WGS, and optical genome mapping, to reveal any potential variants missed or misinterpreted in earlier studies. None of the applied sequencing or imaging techniques revealed rare SVs that overlap or are present within the CEVA haplotype. In the light of the proven accuracy and efficacy of especially optical genome mapping and long-read sequencing in SV detection⁷², we deem it unlikely that any SVs within the CEVA region escaped detection. Additionally, we evaluated all SNVs with an AF $\leq 5\%$ (gnomAD) present within the region for predicted regulatory or splice altering effects but for none of the 20 SNVs a potential effect was predicted by SpliceAI. Two SNVs overlap with a potential regulatory element of *SLC26A4* (GeneHancer, EnhancerAtlas), and one variant is present within the intronic regions of this gene. However, all three variants are located within a highly repetitive element (LINE). Although little is known about the effects of genetic variation within LINE elements, a potential effect on the methylation landscape and consequently gene expression levels has been suggested⁷³ and such an effect can therefore not be excluded for the three indicated variants. For the remaining SNVs, no potential effects on transcript splicing or gene regulation were predicted. Nevertheless, we cannot rule out combinatory effects of the SNVs, since they are all located in *cis*. A thorough experimental (multi-omic) analysis is required to optimally assess the effects of the identified variants. RNA studies can be performed to detect quantitative or qualitative changes affecting the *SLC26A4* transcripts. A defect observed on the RNA level could provide valuable insights that may point towards the true pathogenic defect, and prioritize one, or a combination, of the variants on the CEVA allele. However, *SLC26A4* is not or at extremely low levels expressed in readily accessible patient cell types (e.g., fibroblasts and blood cells). The same holds true for induced pluripotent stem cells or otic progenitor cells.⁷⁴ However, Hosoya and co-workers have successfully developed a protocol that allows the differentiation of otic progenitor cells into outer sulcus-like cells that express *SLC26A4* at high levels. This protocol could potentially be a powerful tool to evaluate the consequences of CEVA haplotype at the RNA level.

SLC26A4 is not the only gene present within the CEVA haplotype, which also spans *BCAP29*, *COG5*, *DUS4L*, *HBP1*, *PIK3CG*, and *PRKAR2B*. For none of these genes, pathogenic variants associated with (syndromic) HL have been reported, nor has a function in the inner ear been described. The majority of the CEVA-associated SNVs (16/20) are located within an intronic region of these genes, however, for none of these variants a splice altering effect is predicted by SpliceAI.

Since the genetic defect on the CEVA haplotype could not be pinpointed by the genetic analyses, we could not determine whether the AF of the defect is lower than that of the CEVA haplotype and more in line with the expected frequency based on the prevalence of HL (1: 1,000 newborns⁷⁵) and the genetic heterogeneity of the condition. Alternatively, the CEVA haplotype could be considered a hypomorphic allele, of which the penetrance depends on the contribution of other co-existing (common) variants. Not all M0 or M1 *SLC26A4* cases could be genetically explained by the presence of the CEVA haplotype. Therefore, digenic inheritance with variants in *EPHA2*, *FOXI1*, and *KCN10* was also explored as a potential explanation for the missing heritability. Digenic inheritance of *SLC26A4* and *EPHA2* has recently been reported in two Japanese Pendred syndrome cases.⁷⁰ A c.1063G>A (p.(Gly355Arg)) and a c.1532C>T (p.(Thr511Met)) variant in *EPHA2* were each found *in trans* with a reported pathogenic variant in *SLC26A4* (Deafness Variation Database⁵⁰). *EPHA2* was identified as a binding partner of pendrin, with a crucial role in regulating pendrin localization.⁷⁰ The identified variants in *EPHA2* were predicted to be pathogenic by several *in silico* predictions tools. However, the c.1532C>T variant has a relatively high allele frequency of 3.03% in the East Asian population, including 11 homozygotes (gnomAD). Yet, in the present study, we did not obtain indications for digenic inheritance of variants in *SLC26A4* and *EPHA2* in subjects with HL and EVA. Besides for *EPHA2*, a digenic mechanism has also been reported and debated for variants in *SLC26A4* and *KCNJ10* or *FOXI1*, with currently no consensus.^{67-69,76,77} *FOXI1* is a transcriptional regulator of *SLC26A4*.⁶⁹ We identified a c.677C>T (p.(Thr226Ile)) *FOXI1* variant in three subjects (2 M0/*FOXI1* and 1 M0/CEVA/*FOXI1*). This variant was previously detected in an individual diagnosed with Pendred syndrome and a monoallelic pathogenic *SLC26A4* variant.⁶⁷ The variant has an allele frequency of 0.71% in non-Finnish Europeans (gnomAD) and affects an amino acid residue located outside the DNA-binding domain but close to the nuclear localization signal (NucPred⁷⁸). Previously reported pathogenic *FOXI1* variants have been shown to affect the DNA-binding properties of the protein.⁷⁹ We speculate that a variant affecting the localization motif of the protein could potentially have a loss of function effect as well. Although the variant is classified as likely benign according to the ACMG classification guidelines, we identified the variant three times in our cohort of genetically unexplained *SLC26A4* cases and combined with the fact that it has been reported in a previous study⁶⁷, this suggests that the variant might actually contribute to the etiology of HL and EVA although not in a monogenic pattern. Interestingly, in *Foxi1*^{-/-} mice, the expansion of the endolymphatic compartment and an audio-vestibular phenotype was observed.⁸⁰ *In situ* hybridization of the endolymphatic duct and sac of these mice revealed complete absence of *Slc26a4* mRNA expression. Functional studies, among

which cellular localizations assays, are warranted to evaluate the effect of the c.677C>T *FOXI1* variant. We did not identify likely pathogenic variants in *KCNJ10* (AF \leq 5%) in our cohort.

WGS did not reveal strong candidate regulatory variants based on data derived from enhancer databases and transcription factor binding site predictions. Nevertheless, interpretation of regulatory variants is still considered complex and is limited by the lack of available epigenetic datasets for the inner ear. Also, no SVs overlapping with *SLC26A4* were detected using WGS, suggesting a limited contribution of SVs to the mutational landscape of *SLC26A4*. This is in line with earlier observations described in literature.^{67,81} For the monoallelic cases (M1, M0/CEVA), no long-read sequencing or optical genome mapping was performed. As it is generally accepted that most SVs could not be accurately detected using short-read sequencing approaches only⁷², it cannot be excluded that causative SVs are present but missed due to technical limitations.

The present study did not confirm that the CEVA allele is associated with a milder HL compared to *SLC26A4* variants affecting the protein-coding sequences, as indicated by Chao et al.²³ They discerned a significantly milder HL in their cohort of M1/CEVA subjects (n=20 ears, median 47.5 dB HL) than we have seen in our cohort of M1/CEVA subjects (n=16 ears, median 84 dB HL). A possible explanation for this discrepancy could be the progression of HL combined with a ~5-year difference in average age between the cohorts (7.5 and 12.8 years, respectively). Progression of HL is seen in up to 39.6% of EVA-ears⁸², with progression rates of ~3.5 – ~5.5 dB/y.^{83,84} On the other hand, the older subjects in our M1/CEVA cohort show less severe HL than the younger subjects, which is questioning the relationship with age. Furthermore, there is also an average age difference of 5 years between the M2 groups in both studies (13.2 years and 18.4 years, respectively), while the severity of HL is comparable (85 and 86.3 dB HL, respectively).

The reported variability of the auditory phenotype associated with EVAs^{18,85,86} may be another explanation for the observed differences in severity of HL in both studies. In literature, many prognostic factors such as genotype, EVA size and morphology, age, head trauma, and gender are reported as underlying explanations for this variability, although some of these studies draw contradicting conclusions.^{82,85,87-91} In the same line, Song et al. reported intrafamilial differences in the severity of hearing loss in siblings with the same biallelic variants in *SLC26A4*.⁹² Larger sample sizes are needed to confirm or reject the hypothesis that the CEVA haplotype is associated with a milder HL phenotype. The significant difference in HL severity between the M2 and M1/CEVA

groups versus the M0 group suggests that *SLC26A4* defects have a prognostic value which can be strengthened in the future by the identification of the underlying genetic defects in subjects of the M0 group.

In conclusion, the HL and EVA in 12 of the 28 studied subjects could be associated with *SLC26A4*. In addition, we have identified genetic factors that might (partially) explain the phenotype in four additional subjects. However, we could not pinpoint the genetic defect that is present on the CEVA haplotype. The arrival of third-generation sequencing techniques, the expansion of epigenetic and transcriptomic datasets and the increasing understanding of non-coding, structural, and regulatory variants will aid in solving the missing heritability in *SLC26A4* in the coming years. This is of great importance for counseling patients about the underlying cause and expected prognosis of their HL. Furthermore, as variants in *SLC26A4* are a frequent cause of HL¹⁰, it is an interesting target for the development of a genetic therapy.⁹³ Although the involved molecular defect of the CEVA haplotype is still not resolved, the high prevalence of the CEVA haplotype suggests that a significant portion of monoallelic *SLC26A4* cases can be associated with *SLC26A4* defects by testing for the presence of this haplotype.

ACKNOWLEDGEMENTS

The authors thank Lieke Lamers, Jolinda Put and Evelien Verwoerd for their assistance in genetic analyses and Ronald van Beek, Ellen Kater-Baats, Marcel Nelen and Michiel Oorsprong for technical support. We would like to thank Mieke Wesdorp for including subjects in the study and Galuh D.N. Astuti and Christian Gilissen for their contribution to statistical and bioinformatic analyses. We thank the Radboudumc Genomics Technology Center, Radboud University Medical Center Nijmegen, for their technical assistance. This study was financially supported by a DCMN Radboudumc grant (to H.K. and F.P.M.C.) and by a grant of the Heinsius-Houbolt foundation (to R.J.E.P. and H.K). TM was supported by the Sigrid Jusélius Foundation. This research was part of the Netherlands X-omics Initiative and partially funded by NWO (The Netherlands Organization for Scientific Research; project 184.034.019).

CONSORTIA

DOOFNL consortium: The DOOFNL consortium is a Dutch nationwide collaboration on hereditary hearing loss and consists of M.F. van Dooren, S.G. Kant, H.H.W. de Gier, E.H. Hoefsloot, M.P. van der Schroeff (ErasmusMC, Rotterdam, the Netherlands), L.J.C. Rotteveel, F.G. Ropers (LUMC, Leiden, the Netherlands), J.C.C. Widdershoven, J.R. Hof, E.K. Vanhoutte (MUMC+, Maastricht, the Netherlands), I. Feenstra, H. Kremer, C.P. Lanting, R.J.E. Pennings, H.G. Yntema (Radboudumc, Nijmegen, the Netherlands), R.H. Free and J.S. Klein Wassink-Ruiter (UMCG, Groningen, the Netherlands), R.J. Stokroos, A.L. Smit, M.J. van den Boogaard (UMC, Utrecht, the Netherlands) and F.A. Ebbens, S.M. Maas, A. Plomp, T.P.M. Goderie, P. Merkus and J. van de Kamp (Amsterdam UMC, Amsterdam, the Netherlands).

REFERENCES

1. Everett, L.A., Glaser, B., Beck, J.C., Idol, J.R., Buchs, A., Heyman, M. *et al.* Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). *Nature Genetics* **17**, 411-422 (1997).
2. Everett, L.A., Morsli, H., Wu, D.K. & Green, E.D. Expression pattern of the mouse ortholog of the Pendred's syndrome gene (Pds) suggests a key role for pendrin in the inner ear. *Proceedings of the National Academy of Sciences* **96**, 9727-9732 (1999).
3. Royaux, I.E., Suzuki, K., Mori, A., Katoh, R., Everett, L.A., Kohn, L.D. *et al.* Pendrin, the protein encoded by the Pendred syndrome gene (PDS), is an apical porter of iodide in the thyroid and is regulated by thyroglobulin in FRTL-5 cells. *Endocrinology* **141**, 839-845 (2000).
4. Royaux, I.E., Wall, S.M., Karniski, L.P., Everett, L.A., Suzuki, K., Knepper, M.A. *et al.* Pendrin, encoded by the Pendred syndrome gene, resides in the apical region of renal intercalated cells and mediates bicarbonate secretion. *Proceedings of the National Academy of Sciences* **98**, 4221-4226 (2001).
5. Pedemonte, N., Caci, E., Sondo, E., Caputo, A., Rhoden, K., Pfeffer, U. *et al.* Thiocyanate transport in resting and IL-4-stimulated human bronchial epithelial cells: role of pendrin and anion channels. *Journal of Immunology* **178**, 5144-5153 (2007).
6. Wangemann, P., Nakaya, K., Wu, T., Maganti, R.J., Itza, E.M., Sanneman, J.D. *et al.* Loss of cochlear HCO₃⁻ secretion causes deafness via endolymphatic acidification and inhibition of Ca²⁺ reabsorption in a Pendred syndrome mouse model. *American Journal of Physiology - Renal Physiology* **292**, f1345-f1353 (2007).
7. Wangemann, P. The role of pendrin in the development of the murine inner ear. *Cellular Physiology and Biochemistry* **28**, 527-534 (2011).
8. Dou, H., Xu, J., Wang, Z., Smith, A.N., Soleimani, M., Karet, F.E. *et al.* Co-expression of pendrin, vacuolar H⁺-ATPase alpha4-subunit and carbonic anhydrase II in epithelial cells of the murine endolymphatic sac. *Journal of Histochemistry & Cytochemistry* **52**, 1377-1384 (2004).
9. Royaux, I.E., Belyantseva, I.A., Wu, T., Kachar, B., Everett, L.A., Marcus, D.C. *et al.* Localization and functional studies of pendrin in the mouse inner ear provide insight about the etiology of deafness in pendred syndrome. *Journal of the Association for Research in Otolaryngology* **4**, 394-404 (2003).
10. Sloan-Heggen, C.M., Bierer, A.O., Shearer, A.E., Kolbe, D.L., Nishimura, C.J., Frees, K.L. *et al.* Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Human Genetics* **135**, 441-450 (2016).
11. Lee, H.J., Jung, J., Shin, J.W., Song, M.H., Kim, S.H., Lee, J.H. *et al.* Correlation between genotype and phenotype in patients with bi-allelic SLC26A4 mutations. *Clinical Chemistry* **86**, 270-275 (2014).
12. Suzuki, H., Oshima, A., Tsukamoto, K., Abe, S., Kumakawa, K., Nagai, K. *et al.* Clinical characteristics and genotype-phenotype correlation of hearing loss patients with SLC26A4 mutations. *Acta Oto-Laryngologica* **127**, 1292-1297 (2007).

13. Fraser, G.R. Association of congenital deafness with goitre (pendred's syndrome): A study of 207 families. *Annals of Human Genetics* **28**, 201-249 (1965).
14. van Beeck Calkoen, E.A., Sanchez Aliaga, E., Merkus, P., Smit, C.F., van de Kamp, J.M., Mulder, M.F. *et al.* High prevalence of abnormalities on CT and MR imaging in children with unilateral sensorineural hearing loss irrespective of age or degree of hearing loss. *International Journal of Pediatric Otorhinolaryngology* **97**, 185-191 (2017).
15. van Beeck Calkoen, E.A., Merkus, P., Goverts, S.T., van de Kamp, J.M., Mulder, M.F., Sanchez Aliaga, E. *et al.* Evaluation of the outcome of CT and MR imaging in pediatric patients with bilateral sensorineural hearing loss. *International Journal of Pediatric Otorhinolaryngology* **108**, 180-185 (2018).
16. Mey, K., Muhamad, A.A., Tranebjaerg, L., Rendtorff, N.D., Rasmussen, S.H., Bille, M. *et al.* Association of SLC26A4 mutations, morphology, and hearing in pendred syndrome and NSEVA. *The Laryngoscope* **129**, 2574-2579 (2019).
17. Forli, F., Lazzerini, F., Auletta, G., Bruschini, L. & Berrettini, S. Enlarged vestibular aqueduct and Mondini Malformation: audiological, clinical, radiologic and genetic features. *European Archives of Oto-Rhino-Laryngology* **278**, 2305-2312 (2021).
18. Griffith, A.J. & Wangemann, P. Hearing loss associated with enlargement of the vestibular aqueduct: mechanistic insights from clinical phenotypes, genotypes, and mouse models. *Hearing Research* **281**, 11-17 (2011).
19. Choi, B.Y., Madeo, A.C., King, K.A., Zalewski, C.K., Pryor, S.P., Muskett, J.A. *et al.* Segregation of enlarged vestibular aqueducts in families with non-diagnostic SLC26A4 genotypes. *Journal of Medical Genetics* **46**, 856-861 (2009).
20. Azaiez, H., Yang, T., Prasad, S., Sorensen, J.L., Nishimura, C.J., Kimberling, W.J. *et al.* Genotype-phenotype correlations for SLC26A4-related deafness. *Human Genetics* **122**, 451-457 (2007).
21. Pryor, S.P., Madeo, A.C., Reynolds, J.C., Sarlis, N.J., Arnos, K.S., Nance, W.E. *et al.* SLC26A4/PDS genotype-phenotype correlation in hearing loss with enlargement of the vestibular aqueduct (EVA): evidence that Pendred syndrome and non-syndromic EVA are distinct clinical and genetic entities. *Journal of Medical Genetics* **42**, 159-165 (2005).
22. Chattaraj, P., Munjal, T., Honda, K., Rendtorff, N.D., Ratay, J.S., Muskett, J.A. *et al.* A common SLC26A4-linked haplotype underlying non-syndromic hearing loss with enlargement of the vestibular aqueduct. *Journal of Medical Genetics* **54**, 665-673 (2017).
23. Chao, J.R., Chattaraj, P., Munjal, T., Honda, K., King, K.A., Zalewski, C.K. *et al.* SLC26A4-linked CEVA haplotype correlates with phenotype in patients with enlargement of the vestibular aqueduct. *BMC Medical Genetics* **20**, 118 (2019).
24. van der Ploeg, C.P., Uilenburg, N.N., Kauffman-de Boer, M.A., Oudesluys-Murphy, A.M. & Verkerk, P.H. Newborn hearing screening in youth health care in the Netherlands: national results of implementation and follow-up. *International Journal of Audiology* **51**, 584-590 (2012).
25. Boston, M., Halsted, M., Meinzen-Derr, J., Bean, J., Vijayasekaran, S., Arjmand, E. *et al.* The large vestibular aqueduct: a new definition based on audiology and computed tomography correlation. *Otolaryngology - Head and Neck Surgery* **136**, 972-977 (2007).

26. Lenth, R.V. Least-Squares Means: The R package lsmmeans. *Journal of Statistical Software* **69**, 1-33 (2016).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
28. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
29. Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P. *et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Research* **22**, 1525-1532 (2012).
30. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).
31. Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappelletti, J., Schleiermacher, G. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423-425 (2012).
32. de Bruijn, S.E., Smits, J.J., Liu, C., Lanting, C.P., Beynon, A.J., Blankevoort, J. *et al.* A RIPOR2 in-frame deletion is a frequent and highly penetrant cause of adult-onset hearing loss. *Journal of Medical Genetics* **58**, 96 (2021).
33. Neveling, K., Mensenkamp, A.R., Derks, R., Kwint, M., Ouchene, H., Steehouwer, M. *et al.* BRCA testing by single-molecule molecular inversion probes. *Clinical Chemistry* **63**, 503-512 (2017).
34. Khan, M., Cornelis, S.S., Pozo-Valero, M.D., Whelan, L., Runhart, E.H., Mishra, K. *et al.* Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genetics in Medicine* **22**, 1235-1246 (2020).
35. Wesdorp, M., Murillo-Cuesta, S., Peters, T., Celaya, A.M., Oonk, A., Schraders, M. *et al.* MPZL2, encoding the epithelial junctional protein myelin protein zero-like 2, is essential for hearing in man and mouse. *American Journal of Human Genetics* **103**, 74-88 (2018).
36. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210 (2019).
37. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26 (2011).
38. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. & Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310-315 (2014).
39. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. & Ng, P.C. SIFT missense predictions for genomes. *Nature Protocols* **11**, 1 (2015).
40. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249 (2010).
41. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**, 361 (2014).

42. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I. *et al.* Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535-548 (2019).
43. Sangermano, R., Khan, M., Cornelis, S.S., Richelle, V., Albert, S., Garanto, A. *et al.* ABCA4 midgenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Research* **28**, 100-110 (2018).
44. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. *et al.* The human genome browser at UCSC. *Genome Research* **12**, 996-1006 (2002).
45. Mantere, T., Neveling, K., Pebrel-Richard, C., Benoist, M., van der Zande, G., Kater-Baats, E. *et al.* Optical genome mapping enables constitutional chromosomal aberration detection. *American Journal of Human Genetics* **108**, 1409-1422 (2021).
46. Neveling, K., Mantere, T., Vermeulen, S., Oorsprong, M., van Beek, R., Kater-Baats, E. *et al.* Next-generation cytogenetics: comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping. *American Journal of Human Genetics* **108**, 1423-1435 (2021).
47. Pedersen, B.S. & Quinlan, A.R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867-868 (2018).
48. Van Camp, G. & Smith, R. Hereditary Hearing Loss Homepage. Available from: <https://hereditaryhearingloss.org>.
49. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* **46**, d1062-d1067 (2018).
50. Azaiez, H., Booth, K.T., Ephraim, S.S., Crone, B., Black-Ziegelbein, E.A., Marini, R.J. *et al.* Genomic landscape and mutational signatures of deafness-associated genes. *American Journal of Human Genetics* **103**, 484-497 (2018).
51. Oza, A.M., DiStefano, M.T., Hemphill, S.E., Cushman, B.J., Grant, A.R., Siegert, R.K. *et al.* Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Human Mutation* **39**, 1593-1613 (2018).
52. Yariz, K.O., Duman, D., Zazo Seco, C., Dallman, J., Huang, M., Peters, T.A. *et al.* Mutations in OTOGL, encoding the inner ear protein otogelin-like, cause moderate sensorineural hearing loss. *American Journal of Human Genetics* **91**, 872-882 (2012).
53. Oonk, A.M., Leijendeckers, J.M., Huygen, P.L., Schraders, M., del Campo, M., del Castillo, I. *et al.* Similar phenotypes caused by mutations in OTOG and OTOGL. *Ear and Hearing* **35**, e84-e91 (2014).
54. Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
55. van Beeck Calkoen, E.A., Engel, M.S.D., van de Kamp, J.M., Yntema, H.G., Goverts, S.T., Mulder, M.F. *et al.* The etiological evaluation of sensorineural hearing loss in children. *European Journal of Pediatrics* **178**, 1195-1205 (2019).
56. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, bax028 (2017).

57. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Research* **48**, d58-d64 (2020).
58. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
59. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T. & Flicek, P.R. The Ensembl regulatory build. *Genome Biology* **16**, 56 (2015).
60. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).
61. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Research* **35**, d88-d92 (2007).
62. Khan, A. & Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Research* **44**, d164-d171 (2016).
63. Dreos, R., Ambrosini, G., Cavin Périer, R. & Bucher, P. EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic Acids Research* **41**, d157-d164 (2013).
64. Dimitrieva, S. & Bucher, P. UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Research* **41**, d101-d109 (2013).
65. Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J.P. & Cotney, J. High-resolution epigenomic atlas of human embryonic craniofacial development. *Cell Reports* **23**, 1581-1597 (2018).
66. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **48**, d87-d92 (2020).
67. Pique, L.M., Brennan, M.L., Davidson, C.J., Schaefer, F., Greinwald, J., Jr. & Schrijver, I. Mutation analysis of the SLC26A4, FOXI1 and KCNJ10 genes in individuals with congenital hearing loss. *PeerJ* **2**, e384 (2014).
68. Yang, T., Gurrola, J.G., Wu, H., Chiu, S.M., Wangemann, P., Snyder, P.M. *et al.* Mutations of KCNJ10 together with mutations of SLC26A4 cause digenic nonsyndromic hearing loss associated with enlarged vestibular aqueduct syndrome. *American Journal of Human Genetics* **84**, 651-657 (2009).
69. Yang, T., Vidarsson, H., Rodrigo-Blomqvist, S., Rosengren, S.S., Enerback, S. & Smith, R.J. Transcriptional control of SLC26A4 is involved in Pendred syndrome and nonsyndromic enlargement of vestibular aqueduct (DFNB4). *American Journal of Human Genetics* **80**, 1055-1063 (2007).
70. Li, M., Nishio, S.-y., Naruse, C., Riddell, M., Sapski, S., Katsuno, T. *et al.* Digenic inheritance of mutations in EPHA2 and SLC26A4 in Pendred syndrome. *Nature Communications* **11**, 1343 (2020).
71. Levy-Sakin, M., Pastor, S., Mostovoy, Y., Li, L., Leung, A.K.Y., McCaffrey, J. *et al.* Genome maps across 26 human populations reveal population-specific patterns of structural variation. *Nature Communications* **10**, 1025 (2019).

72. Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T. *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nature Communications* **10**, 1784 (2019).
73. Xie, H., Wang, M., Bischof, J., Bonaldo, M.d.F. & Soares, M.B. SNP-based prediction of the human germ cell methylation landscape. *Genomics* **93**, 434-440 (2009).
74. Hosoya, M., Fujioka, M., Sone, T., Okamoto, S., Akamatsu, W., Ukai, H. *et al.* Cochlear cell modeling using disease-specific iPSCs unveils a degenerative phenotype and suggests treatments for congenital progressive hearing loss. *Cell Reports* **18**, 68-81 (2017).
75. Morton, C.C. & Nance, W.E. Newborn hearing screening—a silent revolution. *New England Journal of Medicine* **354**, 2151-2164 (2006).
76. Landa, P., Differ, A.-M., Rajput, K., Jenkins, L. & Bitner-Glindzicz, M. Lack of significant association between mutations of KCNJ10 or FOXI1 and SLC26A4 mutations in pendred syndrome/enlarged vestibular aqueducts. *BMC Medical Genetics* **14**, 85 (2013).
77. Jonard, L., Niasme-Grare, M., Bonnet, C., Feldmann, D., Rouillon, I., Loundon, N. *et al.* Screening of SLC26A4, FOXI1 and KCNJ10 genes in unilateral hearing impairment with ipsilateral enlarged vestibular aqueduct. *International Journal of Pediatric Otorhinolaryngology* **74**, 1049-1053 (2010).
78. Brameier, M., Krings, A. & MacCallum, R.M. NucPred—predicting nuclear localization of proteins. *Bioinformatics* **23**, 1159-1160 (2007).
79. Enerbäck, S., Nilsson, D., Edwards, N., Heglind, M., Alkanderi, S., Ashton, E. *et al.* Acidosis and deafness in patients with recessive mutations in FOXI1. *Journal of the American Society of Nephrology: JASN* **29**, 1041-1048 (2018).
80. Hulander, M., Kiernan, A.E., Blomqvist, S.R., Carlsson, P., Samuelsson, E.J., Johansson, B.R. *et al.* Lack of pendrin expression leads to deafness and expansion of the endolymphatic compartment in inner ears of Foxi1 null mutant mice. *Development* **130**, 2013-2025 (2003).
81. Liu, Y.L., Wang, L.L., Wen, J., Mei, L.Y., He, C.F., Jiang, L. *et al.* Application value of high-throughput gene copy number variation detection in the diagnosis of enlarged vestibular aqueduct. *Zhonghua Yi Xue Za Zhi* **101**, 103-107 (2021).
82. Alemi, A.S. & Chan, D.K. Progressive hearing loss and head trauma in enlarged vestibular aqueduct: A systematic review and meta-analysis. *Otolaryngology - Head and Neck Surgery* **153**, 512-517 (2015).
83. Jackler, R.K. & de la Cruz, A. The large vestibular aqueduct syndrome. *The Laryngoscope* **99**, 1238-1243 (1989).
84. Govaerts, P.J., Casselman, J., Daemers, K., De Ceulaer, G., Somers, T. & Offeciers, F.E. Audiological findings in large vestibular aqueduct syndrome. *International Journal of Pediatric Otorhinolaryngology* **51**, 157-164 (1999).
85. Gopen, Q., Zhou, G., Whittemore, K. & Kenna, M. Enlarged vestibular aqueduct: review of controversial aspects. *The Laryngoscope* **121**, 1971-1978 (2011).
86. Arjmand, E.M. & Webber, A. Audiometric findings in children with a large vestibular aqueduct. *Archives of Otolaryngology-Head & Neck Surgery* **130**, 1169-1174 (2004).

87. Archibald, H.D., Ascha, M., Gupta, A., Megerian, C. & Otteson, T. Hearing loss in unilateral and bilateral enlarged vestibular aqueduct syndrome. *International Journal of Pediatric Otorhinolaryngology* **118**, 147-151 (2019).
88. Ascha, M.S., Manzoor, N., Gupta, A., Semaan, M., Megerian, C. & Otteson, T.D. Vestibular aqueduct midpoint width and hearing loss in patients with an enlarged vestibular aqueduct. *JAMA Otolaryngology – Head & Neck Surgery* **143**, 601-608 (2017).
89. Saeed, H.S., Kenth, J., Black, G., Saeed, S.R., Stivaros, S. & Bruce, I.A. Hearing loss in enlarged vestibular aqueduct: a prognostic factor systematic review of the literature. *Otology & Neurotology* **42**, 99-107 (2021).
90. Rah, Y.C., Kim, A.R., Koo, J.-W., Lee, J.H., Oh, S.-h. & Choi, B.Y. Audiologic presentation of enlargement of the vestibular aqueduct according to the SLC26A4 genotypes. *The Laryngoscope* **125**, e216-e222 (2015).
91. Miyagawa, M., Nishio, S.-y., Usami, S.-i. & The Deafness Gene Study, C. Mutation spectrum and genotype–phenotype correlation of hearing loss patients caused by SLC26A4 mutations in the Japanese: a large cohort study. *Journal of Human Genetics* **59**, 262-268 (2014).
92. Song, M.H., Shin, J.-W., Park, H.-J., Lee, K.-A., Kim, Y., Kim, U.-K. *et al.* Intrafamilial phenotypic variability in families with biallelic SLC26A4 mutations. *The Laryngoscope* **124**, e194-e202 (2014).
93. Kim, M.-A., Kim, S.H., Ryu, N., Ma, J.-H., Kim, Y.-R., Jung, J. *et al.* Gene therapy for hereditary hearing loss by SLC26A4 mutations in mice reveals distinct functional roles of pendrin in normal hearing. *Theranostics* **9**, 7184-7199 (2019).

SUPPLEMENTARY TABLES

Table S1. Genotype of reference cohort with biallelic pathogenic variants in SLC26A4

Case	Allele 1		Allele 2	
	Variant	ACMG	Variant	ClinVar
SLC087	c.1147del; p.(Gln383Argfs*49)	UV5	c.1147del; p.(Gln383Argfs*49)	UV5
SLC088	c.2T>C; p.(Met1?)	UV5	c.707T>C; p.(Leu236Pro)	UV5
SLC089	c.890del; p.(Pro297Glnfs*6)	UV5	c.1246A>C; p.(Thr416Pro)	UV5
SLC090	c.1001+1G>A; p.(?)	UV5	c.1001+1G>A; p.(?)	UV5
SLC091	c.1225C>T; p.(Arg409Cys)	UV5	c.707T>C; p.(Leu236Pro)	UV5
SLC092	c.1694G>A; p.(Cys565Tyr)	UV5	c.707T>C; p.(Leu236Pro)	UV5
SLC093	c.754T>C; p.(Ser252Pro)	UV4	c.1174A>T; p.(Asn392Tyr)	UV5
SLC094	c.2048T>C; p.(Phe683Ser)	UV5	c.707T>C; p.(Leu236Pro)	UV5
SLC095	c.1246A>C; p.(Thr416Pro)	UV5	c.707T>C; p.(Leu236Pro)	UV5

Genotype of a control cohort of nine subjects with two (likely) pathogenic variants in the coding or splice site regions of SLC26A4 and a Pendred syndrome phenotype. Segregation analysis to confirm biallelic occurrence of the variants could be carried out in all subjects, except for subjects SLC091 and SLC092. ACMG, variant classification according to the American College of Medical Genetics and Genomics (ACMG) classification guidelines¹; UV4, likely pathogenic; UV5, pathogenic.

Table S2. Details of applied next generation sequencing methods

Case	Sequencing method	Platform	% Reads coverage $\geq 20x$	Mean coverage (x reads)
SLC002	WGS	BGISeq500	88.14	36
SLC003	MIPS	NextSeq500	94.78	920
SLC012	WGS	BGISeq500	88.45	37
SLC013	MIPS	NextSeq500	91.78	900
SLC014	MIPS	NextSeq500	92.28	815
SLC015	WES	Illumina HiSeq2000	96.84	115
SLC017	WES	Illumina HiSeq2000	96.70	125
SLC018	WGS	BGISeq500	88.77	39
SLC031	MIPS	NextSeq500	93.31	517
SLC032	WGS	BGISeq500	89.62	43
SLC036	WGS	BGISeq500	89.21	41
SLC039	MIPS	NextSeq500	93.33	676
	WGS	BGISeq500	89.21	41
SLC040	WES	Illumina HiSeq4000	93.50	136
SLC043	WES	Illumina HiSeq2000	94.85	111
SLC045	MIPS	NextSeq500	92.28	590
	WGS	BGISeq500	83.83	30
SLC048	WGS	BGISeq500	88.82	38
SLC052	WES	Illumina HiSeq2000	93.77	103
SLC056	MIPS	NextSeq500	95.21	901
SLC069	WES	Illumina HiSeq2000	96.62	130
SLC070	WES	Illumina HiSeq4000	97.18	118
SLC071	WES	Illumina HiSeq4000	97.33	121
SLC073	MIPS	NextSeq500	94.99	880
SLC078	MIPS	NextSeq500	95.63	1017
SLC079	WES	Illumina HiSeq4000	97.17	101
	LRS	Sequel II PacBio	NA	12
SLC080	WES	Illumina HiSeq4000	97.40	115
	WGS	BGISeq500	85.27	30
SLC084	WES	Illumina HiSeq4000	98.01	123
SLC085	WGS	BGISeq500	80.33	30
SLC086	WES	Illumina HiSeq4000	97.34	123

WES, whole exome sequencing; WGS, short-read whole genome sequencing; MIPS, molecular inversion probe sequencing; LRS, long-read whole genome sequencing; NA, not applicable.

Table S3. Genes analyzed by MIP sequencing

<i>ACTG1</i>	<i>EPS8</i>	<i>LRTOMT</i>	<i>RIPOR2</i>
<i>ADCY1</i>	<i>ESPN</i>	<i>MARVELD2</i>	<i>S1PR2</i>
<i>ADGRV1</i>	<i>ESRRB</i>	<i>MCM2</i>	<i>SERPINB6</i>
<i>AIFM1</i>	<i>EYA1</i>	<i>MIR96</i>	<i>SIX1</i>
<i>ATP1A2</i>	<i>EYA4</i>	<i>MITF</i>	<i>SIX5</i>
<i>BDP1</i>	<i>GIPC3</i>	<i>MPZL2</i>	<i>SLC9A1</i>
<i>BSND</i>	<i>GJB2</i>	<i>MSRB3</i>	<i>SLC17A8</i>
<i>CABP2</i>	<i>GJB3</i>	<i>MYH14</i>	<i>SLC22A4</i>
<i>CCDC50</i>	<i>GJB6</i>	<i>MYH9</i>	<i>SLC26A4</i>
<i>CD164</i>	<i>GPSM2</i>	<i>MYO15A</i>	<i>SLC26A5</i>
<i>CDH23</i>	<i>GRHL2</i>	<i>MYO3A</i>	<i>SMPX</i>
<i>CEACAM16</i>	<i>GRM7</i>	<i>MYO6</i>	<i>SNAI2</i>
<i>CIB2</i>	<i>GRM8</i>	<i>MYO7A</i>	<i>SOX10</i>
<i>CLDN14</i>	<i>GRXCR1</i>	<i>NARS2</i>	<i>STRC</i>
<i>CLIC5</i>	<i>GRXCR2</i>	<i>NAT2</i>	<i>SYNE4</i>
<i>CLPP</i>	<i>GSDME</i>	<i>OSBPL2</i>	<i>TBC1D24</i>
<i>CLRN1</i>	<i>HARS2</i>	<i>OTOA</i>	<i>TECTA</i>
<i>COCH</i>	<i>HGF</i>	<i>OTOF</i>	<i>TJP2</i>
<i>COL11A2</i>	<i>HOMER2</i>	<i>OTOG</i>	<i>TMC1</i>
<i>COL4A6</i>	<i>HSD17B4</i>	<i>OTOGL</i>	<i>TMEM132E</i>
<i>CRYM</i>	<i>ILDR1</i>	<i>P2RX2</i>	<i>TMIE</i>
<i>DCDC2</i>	<i>KARS</i>	<i>PAX3</i>	<i>TMPRSS3</i>
<i>DFNB31</i>	<i>KCNE1</i>	<i>PCDH15</i>	<i>TNC</i>
<i>DFNB59</i>	<i>KCNJ10</i>	<i>PDZD7</i>	<i>TPRN</i>
<i>DIABLO</i>	<i>KCNQ1</i>	<i>PNPT1</i>	<i>TRIOBP</i>
<i>DIAPH1</i>	<i>KCNQ4</i>	<i>POU3F4</i>	<i>TSPEAR</i>
<i>DSPP</i>	<i>KITLG</i>	<i>POU4F3</i>	<i>USH1C</i>
<i>EDN3</i>	<i>LARS2</i>	<i>PRPS1</i>	<i>USH1G</i>
<i>EDNRB</i>	<i>LHFPL5</i>	<i>PTPRQ</i>	<i>USH2A</i>
<i>ELMOD3</i>	<i>LOXHD1</i>	<i>RDX</i>	<i>WFS1</i>

Table S4. Compound heterozygous or homozygous variants in arHL-associated genes

Case	Class	Gene	Transcript	cDNA	Protein	In-house AF (%)	gnomAD AF (%)	CADD_PHRED	SIFT	PPH2	Mutation Taster	SpliceAI	ACMG
SLC012	M1	OTOGL	NM_173591.3	c.890C>T	p.(Pro297Leu)	0.09	0.12	22.5	0	1.0	Disease causing	-	UV2
		OTOGL	NM_173591.3	c.1369G>T	p.(Val457Leu)	0.02	0.00	15.4	0	0.683	Disease causing	-	UV3

Homozygous or compound heterozygous variants detected in coding or splice site regions (± 14 nucleotides) of genes associated with autosomal recessive hearing loss (arHL). Variants are selected based on an allele frequency of $\leq 0.5\%$ in gnomAD and the in-house database. Scores that meet the thresholds for pathogenicity as described in the methods section are indicated in red. Thresholds for pathogenicity: CADD-PHRED (≥ 15), SIFT (≤ 0.05), PolyPhen-2 (≥ 0.450), MutationTaster (deleterious) and spliceAI (≤ 0.1). In-house AF, allele frequency (%) in in-house database ($\sim 7,500$ exomes); GnomAD AF, allele frequency (%) in gnomAD database V2.1.1; CADD_PHRED, Combined Annotation Dependent Depletion PHRED score; SIFT, Scale-Invariant Feature Transform; PPH2, Poly-Phen-2 score; MutationTaster (prob), MutationTaster score with probability (0-1); spliceAI, splicing prediction score; ClinVar, ACMG, variant classification according to the American College of Medical Genetics and Genomics (ACMG) classification guidelines; UV2, likely benign; UV3, uncertain significance.

Table S5. List of *cis* regulatory elements of SLC26A4

Gene	Start	End	Source
SLC26A4, SLC26A4-AS1	106740447	106742845	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	106743446	106747050	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	106762501	106763480	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107103661	107105444	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107120646	107123445	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107199656	107223646	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107219645	107223646	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107232401	107238444	GeneHancer V5.0
SLC26A4-AS1	107234760	107236310	EnhancerAtlas 2.0
SLC26A4, SLC26A4-AS1	107254046	107255844	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107262447	107263690	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107276447	107280445	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107301300	107302040	EnhancerAtlas 2.0
SLC26A4, SLC26A4-AS1	107301445	107302845	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107330247	107335644	GeneHancer V5.0
SLC26A4	107334930	107335060	EnhancerAtlas 2.0
SLC26A4	107336480	107338480	EnhancerAtlas 2.0
SLC26A4	107350640	107352980	EnhancerAtlas 2.0
SLC26A4, SLC26A4-AS1	107382558	107387330	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107495047	107499844	GeneHancer V5.0
SLC26A4, SLC26A4-AS1	107531740	107533640	EnhancerAtlas 2.0
SLC26A4	107564530	107564670	EnhancerAtlas 2.0
SLC26A4, SLC26A4-AS1	107643420	107643550	EnhancerAtlas 2.0
SLC26A4	107743680	107744940	EnhancerAtlas 2.0

List of human *cis* regulatory elements associated with SLC26A4 or SLC26A4-AS1 that are collected in the GeneHancer database V5.0² or the EnhancerAtlas 2.0³. Only *cis* regulatory elements with an enhancer score >0.7 and an enhancer-gene interaction score >7 were extracted from Genehancer. For EnhancerAtlas 2.0, all enhancer elements that were experimentally determined in human tissues or cell types were selected. Start and End; Genomic positions on chromosome 7 according to GRCh37/hg19.

Table S6. Heterozygous variants in (predicted) cis regulatory elements of SLC26A4

Case	Class	Variant	gnomAD AF (%)	Regulatory element	Source	Identifier	Enhancer score	Enhancer-gene score	PhyloP
SLC002	M1	Chr7:107220628C>A	-1	Chr7:107219645-107223646	GeneHancerV5	GH07J107579	2.05	10.54	-1.143
SLC045	M1	Chr7:107384987C>G	0.19	Chr7:107382558-107387330	GeneHancerV5	GH07J107742	2.25	10.63	0.183

A list of potential cis regulatory elements of SLC26A4 (GeneHancer V5² and EnhancerAtlas V2³) was screened for the presence of rare heterozygous variants (allele frequency $\leq 0.5\%$) in available whole genome sequencing datasets. For none of the variants, the loss of a transcription factor binding site (TFBS) is predicted (JASPAR database⁴, >80% TFBS confidence score and a delta score of > 10%). gnomAD AF, allele frequency (%) in gnomAD database V2.1.1; Regulatory element, genomic position of regulatory element according to GRCh37/hg19; Identifier, unique identifier of regulatory element as accessible in GeneCards⁵; Enhancer score and Enhancer-gene score of regulatory element as provided by the GeneHancer database; PhyloP⁶, nucleotide evolutionary conservation score.

Table S7. Rare genetic variants located within the CEVA haplotype

Genome	RefSNP	Location	gnomAD AF (%)	SpliceAI	PhyloP	Repeatmasker	Regulatory element
Chr7:106622156T>A	rs6961007	Intergenic	-	NA	-5.094	SINE	-
Chr7:106669858G>A (SNP1)	rs17424561	Intergenic	3.04	NA	-1.806	-	-
Chr7:106690778CTTTT>T	NA	Intronic (PPKAR2B)	-	0.01	0.556	-	-
Chr7:106736863C>T	rs149440050	Intronic (PPKAR2B)	3.07	0.01	0.135	LINE	-
Chr7:106741374T>C (SNP2)	rs79579403	Intronic (PPKAR2B)	3.03	0.00	0.8	LINE	GeneHancer
Chr7:106741580ATT>A	NA	Intronic (PPKAR2B)	-	0.01	0	LINE	GeneHancer
Chr7:106764419T>A (SNP3)	rs17425867	Intronic (PPKAR2B)	3.05	0.00	0.852	-	-
Chr7:106807591TAAAA>T	NA	Intergenic	-	NA	0.621	-	-
Chr7:106812322A>AA	NA	Intronic (HBP1)	-	0.00	-2.377	SINE	-
Chr7:106815154T>C (SNP4)	rs117113959	Intronic (HBP1)	2.93	0.05	-0.481	-	-
Chr7:106837681G>A (SNP5)	rs17349280	Intronic (HBP1)	2.90	0.01	0.275	-	-
Chr7:106930234C>T (SNP6)	rs117386523	Intronic (COG5)	2.92	0.00	0.641	-	-
Chr7:106967931A>G (SNP7)	rs80149210	Intronic (COG5)	2.91	0.00	0.838	LINE	-
Chr7:106993159AT>A (SNP8)	rs199667576	Intronic (COG5)	2.96	0.00	-100	-	-
Chr7:107014419A>G (SNP9)	rs9649298	Intronic (COG5)	2.90	0.00	2.769	-	-
Chr7:107081658G>A	rs188905420	Intronic (COG5)	2.31	0.00	-2.019	SINE	-
Chr7:10714762T>C (SNP10)	rs117714350	Intronic (COG5)	2.32	0.00	0.238	LINE	-
Chr7:107242636CT>C (SNP11)	rs199915614	Intronic (BCAP29)	1.91	NA	-100	-	-
Chr7:107282469A>C (SNP12)	rs150942317	Intergenic	2.31	NA	0.241	LTR	-
Chr7:107316164G>A	rs185507318	Intronic (SLC26A4)	2.01	0.00	0.089	SINE	-

Rare genetic variants (allele frequency $\leq 5\%$ gnomAD) that are shared between two individuals that harbor the CEVA allele. SNPs that are part of the previously described CEVA (SNP 1-12) or the V1-CEVA haplotype (SNP 4-12) are underlined. Genome, Genomic position according to GRCh37/hg19; RefSNP, dbSNP reference SNP number; GnomAD AF, allele frequency (%) in gnomAD database V.2.1.1; spliceAI, highest splicing prediction score; RepeatMasker, the interspersed repeats or low complexity DNA sequence at the genomic position according to RepeatMasker; Regulatory element, overlap of the genomic position with a predicted cis regulatory element according to GeneHancer V5 or EnhancerAtlas 2.0; NA, not available.

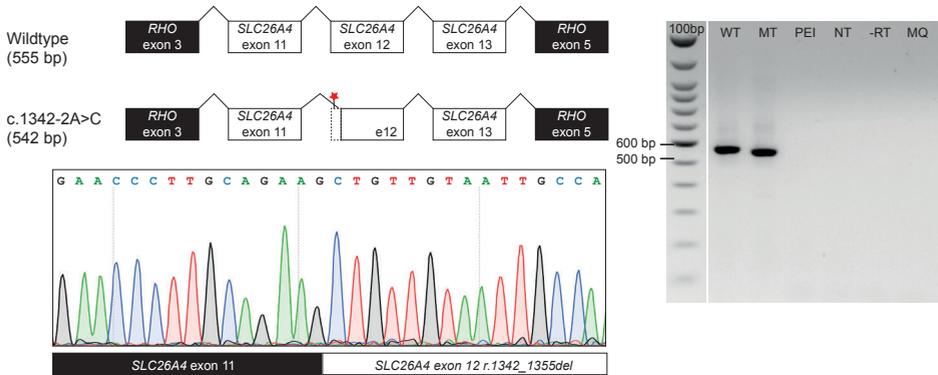
Table S8. Analyzed ears of affected individuals

Class	Number of subjects (male/female)	Number of EVA ears (male/female)	Number of analyzed ears (male/female)	Average age of subjects and analyzed ears (years)	Average age of analyzed ears (Chao et al. 2019) (years), amount of analyzed ears between brackets
M2	11 (5/6)	22 (10/12)	21 (10/11)	13.2	18.4 (n = 48)
M1/CEVA	10 (6/4)	17 (10/7)	16 (9/7)	12.8	7.5 (n = 20)
M1	4 (1/3)	8 (2/6)	8 (2/6)	25 (7, 16, 18 and 59)	15.8 (n = 5)
M0/CEVA	2 (0/2)	3 (0/4)	3 (0/4)	15 (6 and 24)	10.1 (n = 6)
M0	10 (5/5)	17 (9/8)	17 (9/8)	14.6	12.9 (n = 94)

Table adapted from Chao et al. 2019.⁷ Only ears with sufficient audiometric data were used in the analysis.

SUPPLEMENTARY FIGURES

A SLC048



B SLC085

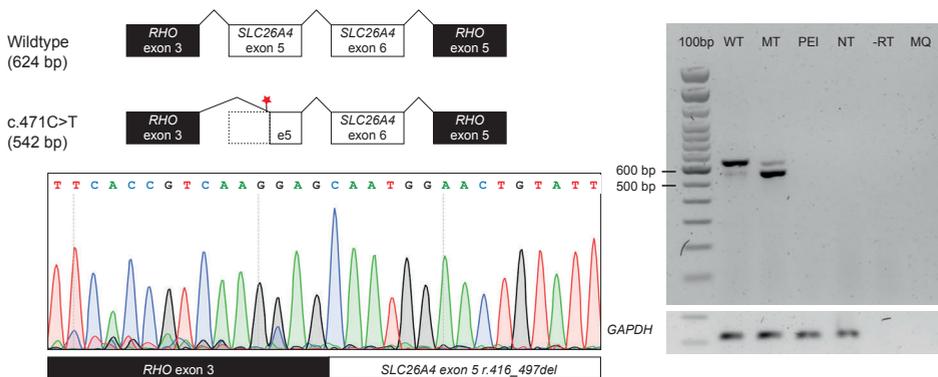


Figure S1. Results of *in vitro* splice assays for variants in SLC048 and SLC085. *In vitro* splice assays were performed in HEK293T cells to validate predicted splice defects. **(A)** In SLC048, a canonical splice site SLC26A4 variant (c.1342-2A>C) was detected. According to SpliceAI predictions, the splice variant (MT) weakens the canonical splice acceptor site. Splice assay results revealed usage of an alternative splice acceptor site located 13 nucleotides downstream. This leads to the formation of a truncated out-of-frame exon 12 (NM_000441.1:r.1342_1355del; p.Ser448Leufs*3). **(B)** In SLC085, a synonymous variant was detected (c.471C>T, p.(Pro157=)). According to SpliceAI, the variant (MT) potentially strengthens an alternative splice acceptor site. The *in vitro* splice assay confirmed that the alternative splice acceptor site (located 27 nucleotides downstream of the variant) is used, which leads to partial exon 5 skipping (NM_000441.1:r.416_497del; p.Gly139A1afs*6,=). Bp, base pair; wt, wildtype; mt, mutant; PEI, transfection reagent-only; RT, reverse transcriptase control; MQ, water control.

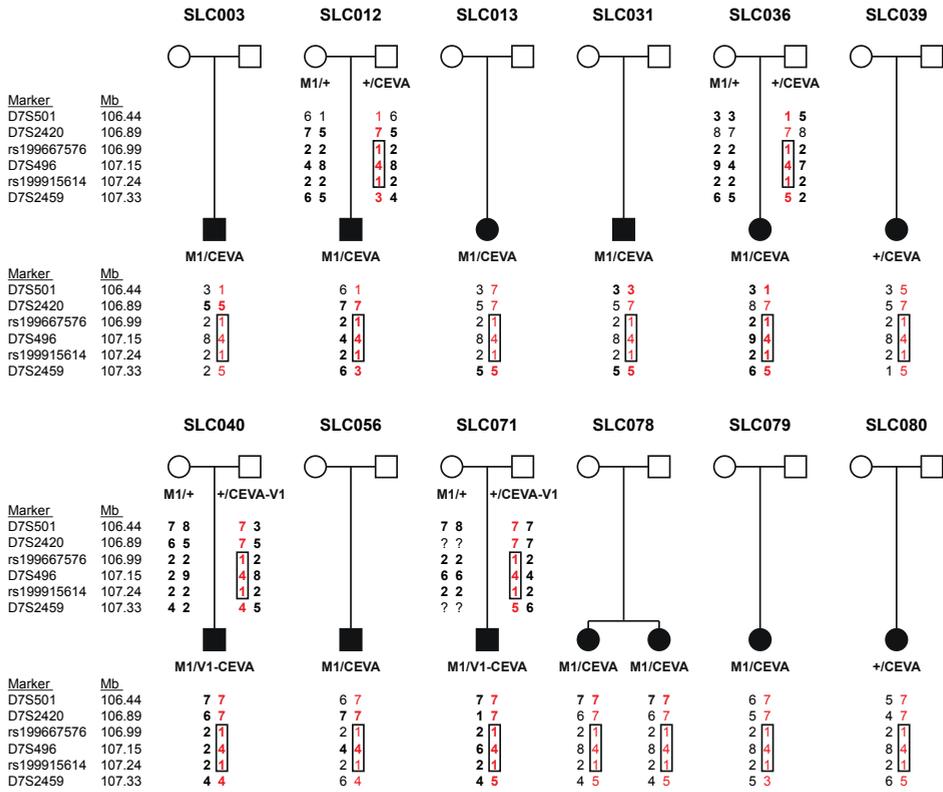


Figure S2. Family pedigrees with haplotypes of VNTR markers. The allele carrying the CEVA haplotype is marked in red. VNTR markers for which the phase of the alleles could be conclusively determined via segregation in the family are marked in bold. Genomic positions (Mb) on chromosome 7 are according to the UCSC genome browser (GRCh37/hg19). VNTR markers of the CEVA haplotype are marked in red. A shared haplotype of 0.89 Mb delimited by the markers D7S501 and D7S2459 was identified. A different repeat length was determined for marker D7S2420 in individual SLC003, the marker is still considered to be potentially part of the shared haplotype as a change of repeat length cannot be excluded. +, wildtype allele; M1, (likely) pathogenic *SLC26A4* variant.

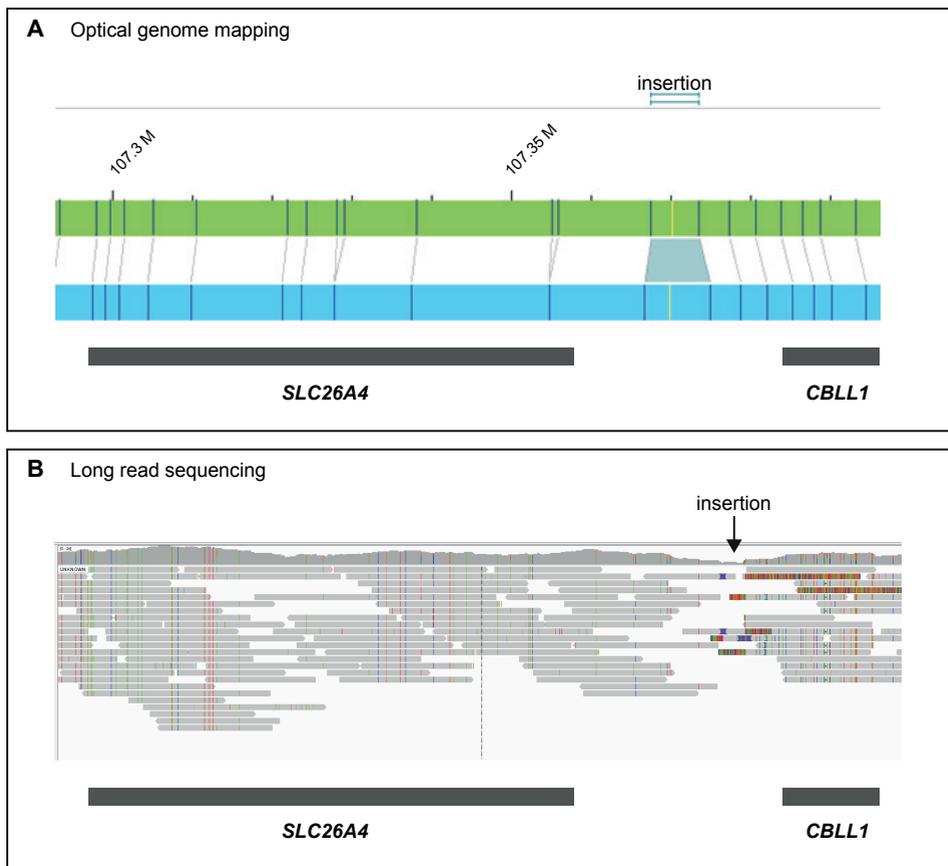


Figure S3. Optical genome mapping and long-read sequencing. Optical genome mapping (Bionano Genomics) and long-read sequencing (PacBio) were performed to detect potential structural variants (SVs) that could be present on the CEVA allele. **(A)** Optical genome mapping was performed using ultra-high molecular weight DNA isolated from peripheral blood cells from individual SLC012 (M1/CEVA). No SVs within the CEVA region or *SLC26A4* were called. One insertion event was called just telomeric from the CEVA-haplotype, but was also called in 100% of the control samples. **(B)** Hi-Fi sequencing reads were visualized in IGV software. PacBio long-read sequencing was performed on genomic DNA isolated from peripheral blood from individual SLC079 (M1/CEVA). After sequencing analyses, no SVs remained that were present within the CEVA region or *SLC26A4*. The same insertion event as depicted in A was detected using PacBio sequencing, but was also present in control data. The insertion event is therefore considered a reference problem and not a true SV.

M0

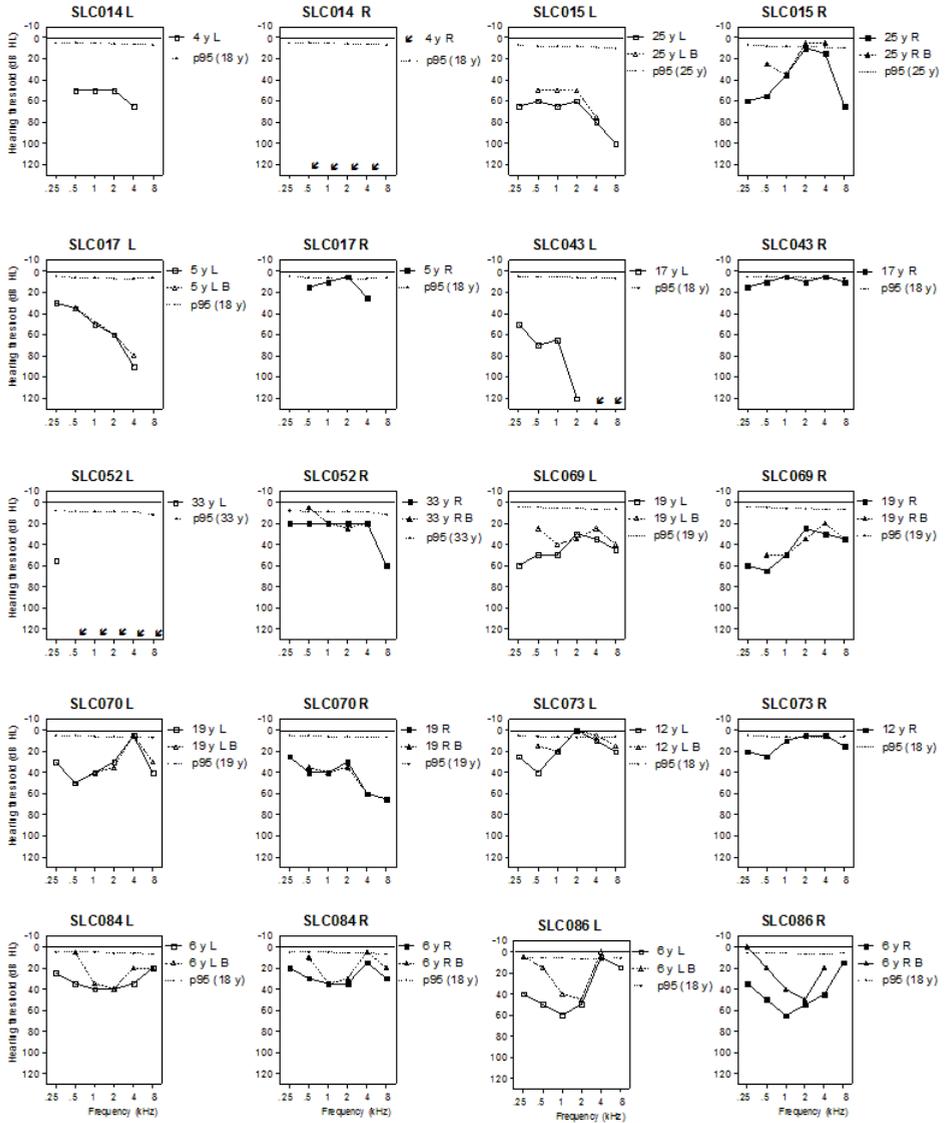
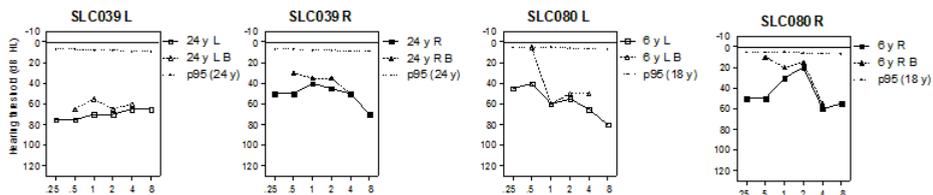
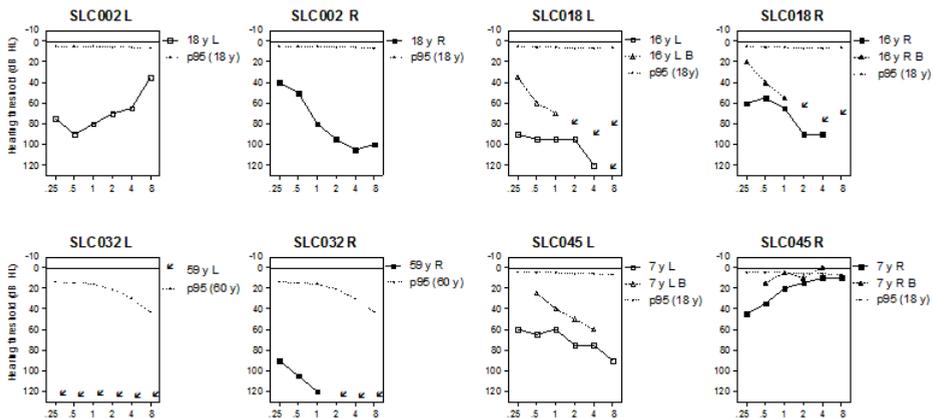


Figure continues on next page

M0/CEVA



M1



M1/(V1)-CEVA

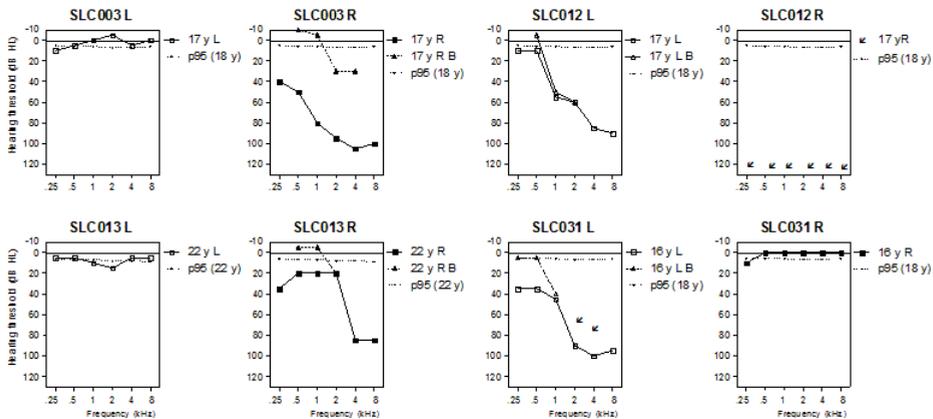


Figure continues on next page

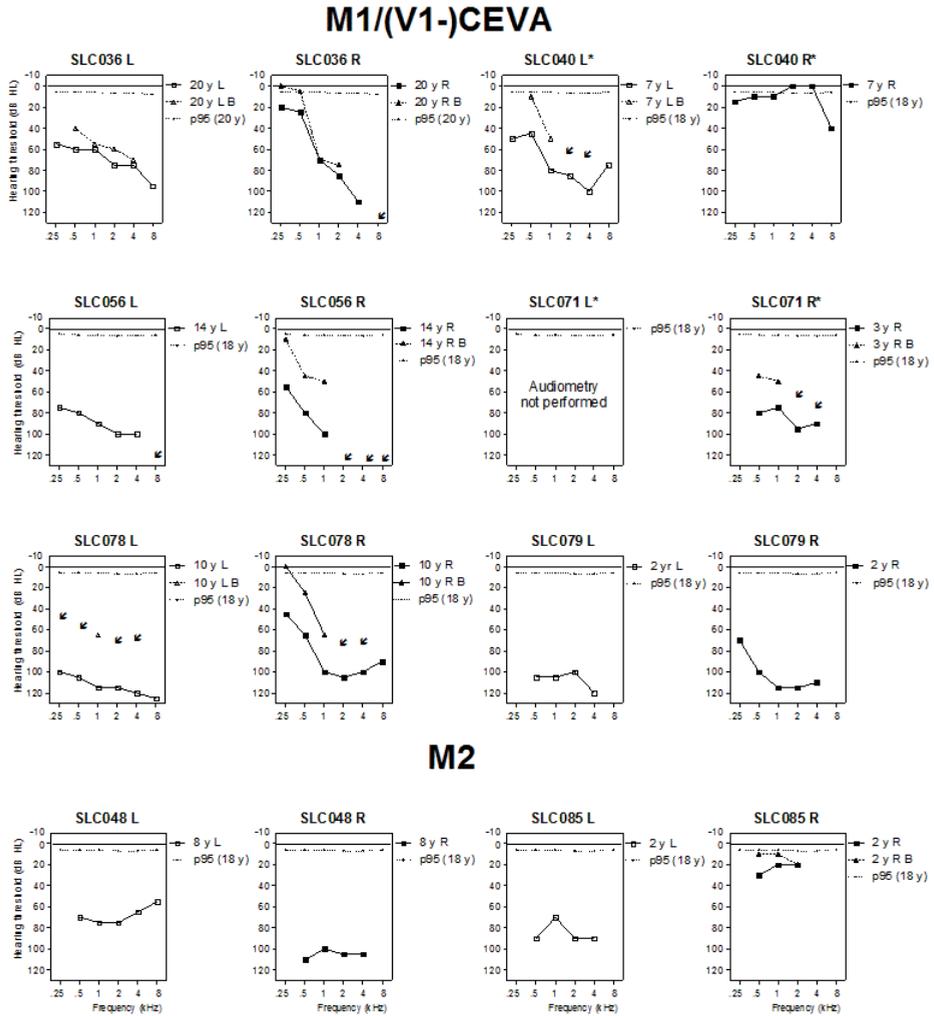


Figure S4. Pure tone audiometry of affected individuals. Air, and if available, bone conduction thresholds of all subjects are depicted, except for subjects of the M2 reference cohort. The p95 values are matched to the individuals' sex and age at the most recent audiometry, according to the ISO 7029:2017 standard. The age range for which the ISO 7029:2017 can be applied is 18 to 70 years. Black arrows: threshold could not be measured. The CEVA haplotype was detected in 8 individuals, in an additional 2 individuals (SLC040 and SLC071, indicated with *), a smaller haplotype was found, termed V1-CEVA. y, age in years; R, right; L, left; B, bone conduction; dB HL, decibel hearing level; kHz, kiloHertz.

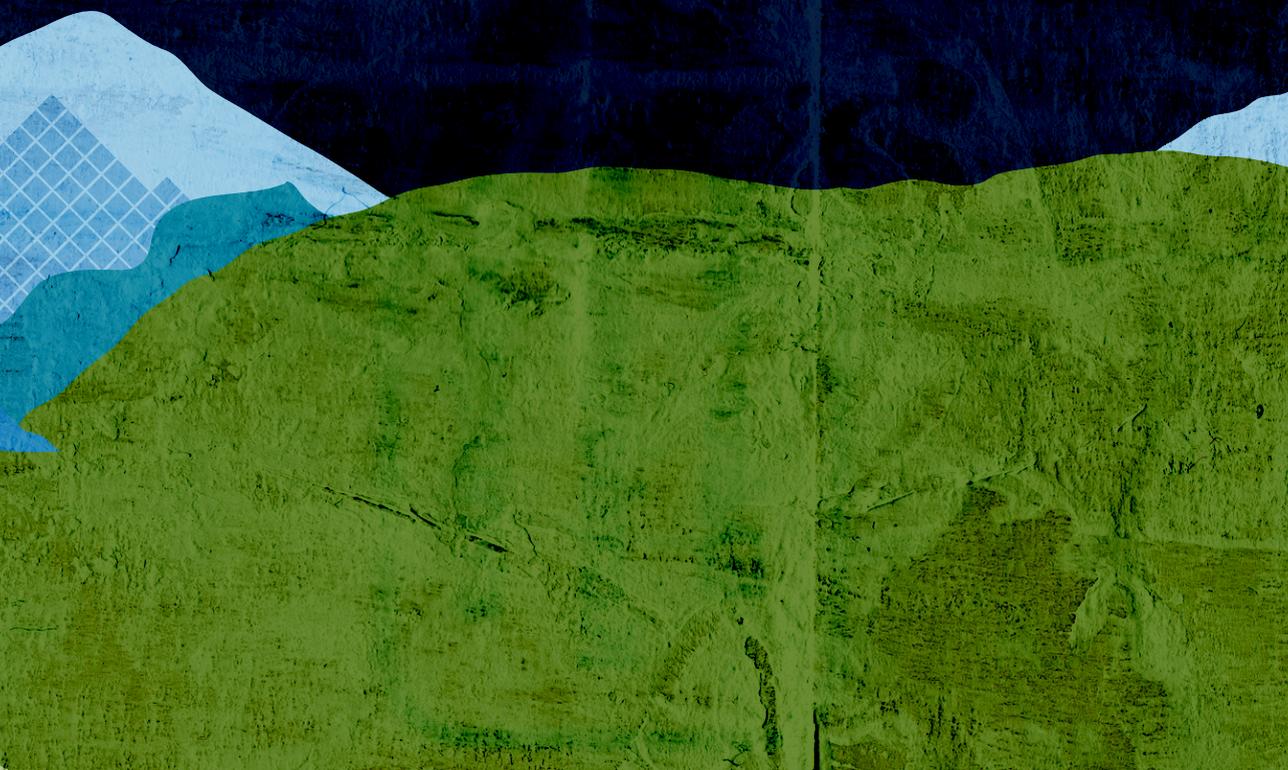
SUPPLEMENTARY REFERENCES

1. Oza, A.M., DiStefano, M.T., Hemphill, S.E., Cushman, B.J., Grant, A.R., Siegert, R.K. *et al.* Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Human Mutation* **39**, 1593-1613 (2018).
2. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, bax028 (2017).
3. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Research* **48**, d58-d64 (2020).
4. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **48**, d87-d92 (2020).
5. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S. *et al.* The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics* **54**, 1.30.1-1.30.33 (2016).
6. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research* **20**, 110-121 (2010).
7. Chao, J.R., Chattaraj, P., Munjal, T., Honda, K., King, K.A., Zalewski, C.K. *et al.* SLC26A4-linked CEVA haplotype correlates with phenotype in patients with enlargement of the vestibular aqueduct. *BMC Medical Genetics* **20**, 118 (2019).



Chapter 6

General discussion and perspectives



This year, 2021, we celebrate the 20th anniversary of the human genome. It has been twenty years since the public Human Genome Project and Celera Corporation jointly released the first sequences of our genome.^{1,2} As expected, this greatly impacted modern genetic diagnostics. Nowadays, a putative variant can be readily interrogated by extracting all relevant information such as genomic position and population allele frequencies. The costs to perform whole genome sequencing (WGS) have dropped from \$2.7 billion for the first complete human genome sequence, to less than \$1,000.^{2,3} In combination with other technological advancements and functional analyses, the field of human genetics revolutionized. Finally, we are slowly reaching a complete understanding of the human genome and the impact of genetic variation.

Despite all these advancements, interpretation of genetic variation can still be challenging (reviewed in **chapter 1.2**). Although we are able to detect the full spectrum of genetic variation, the knowledge that is required to interpret all these variants is still lagging behind, which prevents optimal translation of findings to clinical care. This knowledge gap is considered the most important contributor to the missing heritability that is described for both inherited retinal dystrophies (RD) and hearing loss (HL). About 20-50%^{4,7} of RD cases and 60-70%^{8,9} of HL cases still lack a genetic diagnosis when whole exome sequencing (WES) is performed.

The focus of this thesis was to shed light on the missing heritability for the inherited sensory disorders HL and RD. To do so, a variety of genetic and functional strategies was employed. In **chapter 2**, functional evidence was collected to validate the association of *KIAA1549* with retinitis pigmentosa (RP). Novel pathogenic *KIAA1549* variants were identified and the number of known patients affected by these variants was increased. To prove causality of *KIAA1549* pathogenic variants in RP, *in vitro* expression studies and immunohistochemistry were performed. The studies described in the following chapters of this thesis were focused on previously identified loci for RD and HL, for which the genetic defects were still elusive. After decades of research, the defects underlying two of these loci were finally identified: DFNA21, that is associated with dominantly-inherited HL (**chapter 3**) and RP17, associated with dominantly-inherited RP (**chapter 4**). As a continuation of this research, the first steps towards a genetic therapy for DFNA21 have been described in **chapter 3.2** with the development of a gene silencing strategy using RNase H1-dependent antisense oligonucleotides (AONs). The findings described in **chapters 3** and **4** do not only provide a genetic explanation for DFNA21 and RP17; the important insights that are gained can also be applied to other unsolved Mendelian diseases.

Not all cases included in this thesis could be (completely) resolved. **Chapter 5** described the investigation of a Dutch cohort of individuals affected by HL and monoallelic pathogenic *SLC26A4* variants. In the majority of the monoallelic cases, a common shared haplotype (previously coined as the “CEVA” haplotype) was detected. Despite the characterization and detailed interrogation of the haplotype using both short-read and long-read sequencing as well as optical genome mapping, no putative causative defect was found that could be associated with HL. This indicates that despite all the advancements, the understanding of the human genome is still incomplete.

In the current chapter, the importance of genetic diagnostics, its limitations and the most important challenges are discussed. Special attention is paid to upcoming developments that are necessary to address these challenges in the years to come. Some might believe that closing the diagnostic gap will soon become within reach. Nevertheless, several hurdles have to be overcome first.

THE IMPORTANCE OF GENETIC DIAGNOSTICS

Why do we strive to provide a genetic explanation for all individuals affected with a disorder that is thought to have a genetic cause? In most cases, receiving a genetic diagnosis has a significant impact on a patient’s life. Clinical care and counseling can be optimally fitted to the patient’s needs, possibilities regarding family planning can be discussed, and in some cases, therapeutic options become available.

Patient counseling

A conclusive genetic diagnosis is crucial for a clinician to be able to provide a patient with optimal care and counseling. Based on the genetic diagnosis, different questions may arise related to disease progression, genetic risks of family members and the risk to develop a syndromic phenotype. Special attention can be paid to fitted rehabilitation options, such as hearing aids for individuals affected by HL or light-filtering glasses for RD. In case of an (anticipated) syndromic phenotype, additionally, pre-symptomatic care can be initiated (e.g. thyroid evaluation in individuals affected with HL and pathogenic *SLC26A4* variants (**chapter 5**)).

A genetic diagnosis can help individuals to set their personal expectations and can be experienced as an important relief. Studies have revealed that a genetic diagnosis can have significant emotional consequences and that special attention should be paid to psychological well-being.¹⁰ Unfortunately, not every genetic diagnosis will provide conclusive answers to all questions. Variants in numerous HL- or RD-associated genes

(e.g. *CDH23*¹¹, *USH2A*¹²) are associated with highly variable phenotypes. Part of this variability can be attributed to the variant type or genomic position of the variant, but it cannot be explained in all cases. Even a single genetic variant can lead to a wide spectrum of phenotypic expression, as observed for the in-frame *RIPOR2* deletion described in **chapter 3.1**. Therefore, a clinician must always be cautious when providing counseling on prognosis.

Family planning

Depending on the genetic diagnosis, the severity of the disorder and the expected inheritance pattern, reproductive options can be discussed. For severe disorders that meet strict criteria that differ per country, a preimplantation genetic diagnosis (PGD) procedure can be initiated for a couple with a known affected status (dominant disorders) or when both parents are known carriers of disease-causing variants affecting the same gene (recessive disorders). PGD is a form of *in vitro* fertilization (IVF) performed in the laboratory to yield several embryos. These embryos are genetically tested to determine the genotype, and only embryos without the specific genetic defect(s) will be used for implantation.¹³ Alternatively, if a couple wishes to conceive a child naturally, prenatal testing can be performed. There are several options available that allow screening of the genotype of the fetus. When it appears that the embryo carries a genetic defect, it can be decided to terminate the pregnancy. Invasive prenatal genetic tests (e.g. chorionic villus sampling) have been available for some decades now. More recently, also a non-invasive test has been implemented in clinical settings as well.¹⁴ When performing non-invasive prenatal testing (NIPT), a blood sample is drawn from the pregnant mother, which contains cell-free fetal DNA from the placenta that is subjected to genetic testing.^{15,16}

To be eligible for PGD or prenatal testing, the genetic diagnosis has to leave no room for doubt. One cannot risk to erroneously select an embryo with a misinterpreted genetic status. For some years, *KIAA1549* was only considered a candidate disease gene for RP. With the identification of additional RP cases affected by pathogenic *KIAA1549* variants (**chapter 2**), causality of these variants in RP is now considered conclusive and *KIAA1549* is included in diagnostic gene panels for RD. Likewise, only after functional evidence was obtained that showed causality of RP17-associated structural variants (SVs) (**chapter 4**), the first PGD request has been recently approved for a member of the Dutch index RP17 family. This is one of the striking examples within this thesis of how molecular genetic findings can be swiftly translated to the clinic and thereby can have a significant impact on an individual's life.

Genetic therapies

Recently, the first retinal genetic therapy called Luxturna™ was approved for Leber congenital amaurosis or early-onset RP caused by pathogenic *RPE65* variants and can now be prescribed to eligible individuals. A viral vector (adeno-associated virus (AAV)) that contains a wildtype copy of *RPE65* cDNA is delivered to the retina by subretinal injection, which halts disease progression and, in some cases, even leads to improved visual function in treated individuals.¹⁷⁻¹⁹ Additionally, hundreds of clinical trials are currently ongoing (<https://www.clinicaltrials.gov/>) to assess the safety and efficacy of genetic therapies to treat other forms of inherited RD and HL. Many scientists and clinicians anticipate that gene- or variant-specific therapeutic strategies hold the future for inherited sensory disorders.

Genetic therapies are often personalized therapeutic strategies that combine knowledge of the underlying pathogenic variant and the molecular mechanisms involved. In case of loss-of-function, such as for the abovementioned pathogenic *RPE65* variants, timely supplementation of a wildtype copy of the affected gene (gene augmentation) suffices. In principle, this approach may also be used in haploinsufficiency (dominant) cases, but it may then be a challenge to correct the defect with the right dose as too little protein may be ineffective and too much may also be detrimental. In case of toxic gain-of-function or dominant-negative mechanisms, a different approach is required. In **chapter 3.2**, the design of an allele-specific approach to treat DFNA21 is described. AONs were designed that specifically bind the mRNA derived from the mutant allele implicated in DFNA21 and target the transcript for RNase H1-dependent degradation. A lead AON molecule was designed and validated *in vitro*. In the near future, the ability of the molecule to specifically knockdown mutant allele expression will be assessed in a mouse model of DFNA21.

Over the years, a wide range of strategies to optimally design a genetic therapy have been explored, each holding its own pros and cons (extensively reviewed in references (20) and (21)). DNA therapies include gene augmentation (e.g. Luxturna™) and genome editing strategies. A gene augmentation approach based on AAVs, holds the advantage that there is no risk for gene integration in the genome. However, important challenges include delivery, immunogenicity and longevity. Despite these challenges, AAV vectors are considered the most safe and efficacious therapeutic vectors for retinal and inner ear delivery.

Genome editing strategies have also been widely explored, especially the CRISPR/Cas9 system has received great attention in the past decade. This system that allows, for example, the correction of disease-causing variants or disruption of genes harboring

gain-of-function variants using *in vivo* or *ex vivo* approaches. Nowadays, CRISPR/Cas9 is commonly used for mammalian genomic editing and for the efficient generation of knockin or knockout animal models. Evidently, several ethical concerns are attached to human gene editing, including the risk to introduce permanent off-target effects. CRISPR/Cas9 has clearly revolutionized the field of genetic research but the direct application of the technique in human subjects will most likely still remain a topic of ongoing debate for quite some years.

In light of limitations of DNA therapies, RNA-targeting therapies, such as the AON-strategy for DFNA21 (**chapter 3.2**), can be considered a more attractive alternative. AONs can be designed to redirect splicing (so called splice-AONs) or to silence gene expression (e.g. RNase H1-dependent AONs). Especially the transient and reversible nature of treatments with these molecules is considered an important advantage over DNA editing strategies. Delivery, longevity and potential off-target effects on the other hand, remain important challenges. Nevertheless, the recent report of a single AON injection (seporfarsen) that leads to visual improvement with an effect lasting for over 15 months as a treatment for *CEP290*-associated RD is very promising.²² Drug delivery, and specifically repeated delivery, to the inner ear is still less advanced as compared to the retina. However, with many ongoing efforts this limitation can be expected to be overcome soon (reviewed in (21)).

The currently available genetic therapies are extremely expensive. The treatment of a single injection with Luxturna costs \$425,000 per eye.²³ Additionally, for each mutated gene, pathogenic variant, or subsets of pathogenic variants, a personalized therapy needs to be designed depending on the pathogenic variants involved. This entails a time- and money-consuming process of clinical testing, that on average takes more than 10 years. An exception to this might be specific “N=1” situations. Last year, an AON-based treatment (milasen) received FDA approval within one year after the first contact with the patient.²⁴ The molecule was designed to treat an inherited lethal form of Batten disease, an autosomal recessive disorder that progressively affects the retina and the central nervous system. In this case, WGS revealed biallelic pathogenic variants in *MFSD8*, among which a retrotransposon insertion in intron 6. The chemistry of the AON was based on an existing FDA-approved molecule, and combined with the urgency of the patient’s clinical situation and the sequence-specific design of the drug, rapid development of the treatment was possible.

Despite all these issues, it is evident that genetic therapies entail great promises for the future. Nevertheless, there is first a crucial initial step to be taken: the establishment of a genetic diagnosis for all inherited disorders. Considering the important benefits of

genetic diagnostics related to genetic therapy, but also patient counseling and family planning, it is evident that genetic research is of utmost importance and significantly contributes to the patient's quality of life. Additionally, with the rising number of commercially available DNA tests (e.g. 23andme) and ongoing debate of the value and ethical concerns of genetic risk factor screening²⁵, complete understanding of the genomic variation will become more and more important. A study performed by Hanany et al. estimated that about 36% of all individuals carry a pathogenic variant that can cause autosomal recessive RD.²⁶ Preconception screening (e.g. "clinical exome" analysis) in consanguineous²⁷⁻²⁹ and non-consanguineous³⁰ couples has shown its value. Unfortunately, after 30 years of research, it is not possible yet to provide a genetic diagnosis to all individuals affected by a genetic disorder. Which crucial challenges have been overcome in the past and which challenges are remaining that should be conquered to solve all the unsolved? These questions will be addressed in the following sections.

GENETIC DIAGNOSTICS: CHALLENGES AND LESSONS LEARNED FROM THE PAST

Thirty years after the discovery of the first HL- and RD-associated genes, novel disease-associated genes are still reported. Based on today's count (May 2021), 271 RD-associated genes³¹ and 154 HL-associated genes³² have been described (**Figure 1**). The number of genes gradually increased over the years, with developing technologies, increasing knowledge and decreasing sequencing costs being considered as important contributing factors. Noteworthy high-impact events over the years are the completion of the Human Genome Project and the introduction of WES, WGS and long-read sequencing technology. It was anticipated that WES and WGS would rapidly revolutionize the field of genetic diagnostics, lead to an exponential increase in disease-associated genes, and resolve the missing heritability. Unfortunately, this was not the case. In fact, in recent years, gene identification curves seem to have reached a plateau phase, suggesting that the majority of disease-associated genes have been discovered and attention should be shifted to (re)investigation of known disease genes in more depth, with specific emphasis on the involvement of non-coding elements.

First generation sequencing

As reviewed in **chapter 1.2** of this thesis, the first HL- and RD-associated genes were identified by linkage studies and positional cloning (*CHM* (1990)³³, *POU3F4* (1995)³⁴). Large and small deletions on the X-chromosome were linked to the respective diseases and enabled pinpointing the candidate disease genes and subsequent cloning.

Pathogenic variants in other genes, including *RHO* (1990)³⁵, were identified after candidate gene cloning, given the known organ-specific key function of for instance rhodopsin. Also studies in natural mouse mutants were employed successfully, in which a mutated gene was identified and the human counterpart turned out to be mutated in affected individuals. Such studies contributed significantly to an impressive list of candidate disease genes in which often one or two variants were found in a single case or family.³⁶⁻³⁸ Efforts undertaken by the International Mouse Phenotyping Consortium (IMPC) are still ongoing to generate and phenotype a knockout mouse model for all genes.³⁹

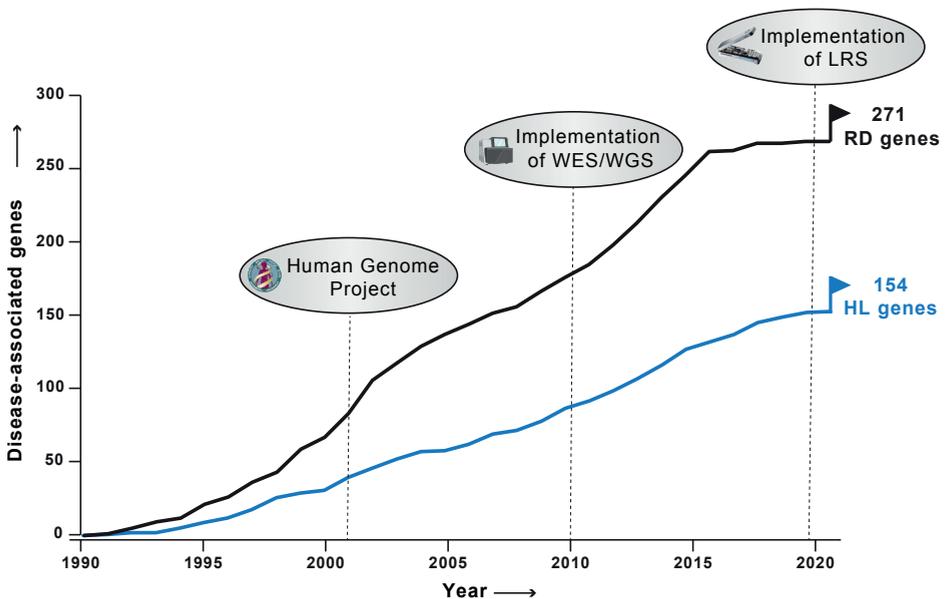


Figure 1. Discovery of inherited RD- and HL-associated genes over the years. The first inherited RD- and HL-associated genes were discovered in the early 90s using linkage analysis and positional cloning strategies. Supported by the arrival of novel technologies (next generation sequencing (e.g. whole exome sequencing (WES), whole genome sequencing (WGS) and long-read sequencing (LRS)), the number of disease-associated genes gradually increased over the years. Today (May 2021), 271 RD-associated genes (RetNet³¹) and 154 HL-associated (HHL homepage³²) genes have been reported.

Although the methods employed in this first era of DNA sequencing are generally considered to be time-consuming and labor-intensive as compared to contemporary parallelized sequencing efforts, they provided crucial insights from which we still benefit today. Most techniques applied in families with inherited RD or HL were focused on limiting the critical region of interest; pinpointing only a small genomic region that

needed to be interrogated for putative pathogenic variants. Also today, where more and more sequencing data is generated, these approaches could still be valuable and might even be necessary in order to keep data analyses feasible. Nevertheless, this advantage also immediately represents the most important pitfall of these old techniques. When focusing on a specific genomic region, a genetic defect located outside the region of interest is easily overlooked. In this thesis, an important example is described in **chapter 3**, where the genetic defect underlying DFNA21 was found to be located 0.9 Mb centromeric of the determined DFNA21 locus. In this case, the true locus within the family was masked by the presence of both phenocopies (individuals mimicking the HL phenotype, with other (non-genetic) causes underlying) and non-penetrance (asymptomatic individuals carrying the putative genetic defect). Because of this, the DFNA21 locus was falsely delimited in the past and the underlying pathogenic variant could not be identified by targeted sequencing of genes located within the locus.⁴⁰ Now, twenty years later, the introduction of WES and decreased sequencing costs allowed us to sequence the complete coding regions of the genome of multiple affected family members. Only by combining the genetic approaches (WES and linkage analysis) with the detailed phenotypic data that were collected over the years, we were able to finally resolve the genetic defect of DFNA21: an in-frame deletion in the *RIPOR2* gene.

Second generation sequencing

Second generation sequencing, or next generation sequencing, allows the analysis of all genetic variants in the coding elements (WES) or of the human genome sequence (WGS) in a parallelized fashion. The first reports of RD- and HL-associated genes discovered by WES appeared in 2010 (e.g. *TSPAN12*⁴¹ associated with familial exudative vitreoretinopathy or *GPSM2*⁴² associated with non-syndromic HL). The genetic solve rate of inherited HL and RD by WES ranges from 30% to 80%, depending on the specific phenotype studied.⁴⁻⁹ In general, the added diagnostic value of WGS for inherited disorders is estimated to be ~21% (i.e. 28% of cases solved by WES versus 49% by WGS).⁴³ In a first study in which the application of WES and WGS was compared for diagnostic purposes of inherited RD specifically, the added value of WGS was calculated to be 31%.⁴⁴ Considering the higher diagnostic yield, improved read coverage of both coding and non-coding regions and decreasing sequencing costs of WGS, it can be anticipated that soon WGS will replace WES in genetic diagnostics.

Some important limitations that were hampering the field for a long time can be resolved by the implementation of WGS. Most evidently, WGS allows the identification of non-coding variants that can have splice altering or regulatory effects. Additionally, the technique can identify SVs at a base-pair resolution. Although micro-array technologies

(first generation sequencing) or WES do allow the detection of copy number changes, these methods are relatively insensitive as breakpoints of SVs cannot be readily determined and inversions or translocation events cannot be detected. As described in **chapter 4**, where RP17-associated SVs were characterized, breakpoint resolution has appeared to be crucial for variant interpretation. SVs can lead to disruption of the 3D chromosome landscape; new regulatory domains can be formed (neo-topologically associating domains (TADs)), regulatory domains can fuse (TAD-fusion) or be deleted (TAD-deletion). Only when the breakpoints of an SV are precisely determined, the effects on the 3D chromosome landscape, and consequently enhancer-promoter interactions, can be correctly interpreted. Therefore, the arrival of WGS has been essential to resolve the genetic mysteries underlying RP17.

The findings in **chapter 4** indicate that SVs can be an important and unrecognized cause of inherited RD and HL. An important contribution of SVs to the genetic spectrum of both RD and HL has been recently suggested in literature as well.^{45,46} Based on this, it can be speculated that by expanding the application of WGS to clinical practices, a significant number of pathogenic SVs will be reported in the next years.

Third generation sequencing

Although SV-detection has been significantly improved by the implementation of WGS, there is still room for further improvement. As extensively described in **chapter 1.2**, short-read WGS approaches are inefficient in the detection of complex SVs. Long-read sequencing techniques (single-molecule real-time (SMRT) (Pacific Biosciences) and nanopore sequencing (Oxford Nanopore)) hold the promise to allow detection of all SVs. Additionally, these techniques have a superior performance in the analysis of repeat-rich and homologous regions, which is an important drawback of short-read WGS. Most likely, the implementation of these techniques is necessary to determine the complete genetic variation landscape in an individual. On the other hand, SMRT and nanopore sequencing still include a relatively high false-positive rate for the detection of single nucleotide variants and small insertions and deletions (≤ 15 bp) and therefore are still unfit to fully replace short-read WGS approaches. Nevertheless, the implementation of HiFi reads, also called circular consensus reads, in SMRT sequencing promises to improve the accuracy up to 99% (reviewed in **chapter 1.2**).

The future of variant detection

With the implementation of innovative sequencing techniques, and consequently, the generation of larger sequencing datasets, an increasing need and a rapid development of additional computational tools can be anticipated. Artificial intelligence software,

including deep-learning and machine-learning tools, have been developed to aid in the interpretation of candidate variants. An example is the SpliceAI deep-learning tool designed to assess putative splice variants.⁴⁷ This deep-learning tool shows significant advances in *in silico* splice effect predictions, as it does not depend on preselected features.⁴⁸ There is a growing interest to implement the deep-learning strategy to assess regulatory variants and SVs as well.^{49,50} Additionally, machine learning tools are also employed for predictive and diagnostic purposes. Audio profiling software has been reported that can accurately predict, based on audiometric data, whether a person is a carrier of pathogenic variants associated with specific types of HL.^{51,52}

Artificial intelligence also offers possibilities to investigate oligogenic or multifactorial diseases. Machine learning tools to predict potential pathogenicity of a combination of oligogenic variants in an individual have been described (e.g. ORVAL⁵³ and Variant Combinations Pathogenicity Predictor⁵⁴). Using a machine-learning approach such as Variant Combinations Pathogenicity Predictor, it is possible to accurately predict potential diallelic inheritance. With the increasing sizes of available sequence datasets and computational power, statistical testing has become possible. Statistical tools can be used to e.g. investigate gene-specific enrichment of *de novo* pathogenic variants⁵⁵, or genome-wide association studies (GWAS) studies can be performed to identify genetic risk loci for multifactorial disease. Several GWAS studies have been reported to investigate potential genetic risk factors underlying age-related HL. Whereas initially no loci could be identified with genome-wide significance⁵⁶, by investigating a larger patient cohort, 44 independently associated genomic loci were found.⁵⁷ While still in an early stage, it can be speculated that computational and artificial intelligence tools will increasingly contribute to variant identification and interpretation in the following years.

WHICH CHALLENGES ARE STILL AHEAD?

Last year, the first complete gapless human chromosome sequence was obtained. The X chromosome was successfully sequenced from telomere-to-telomere, which was achieved by combining high-coverage ultra-long nanopore sequencing (Oxford Nanopore).⁵⁸ Scientists were finally able to resolve the remaining 29 gaps of the X chromosome that were still present in the reference genome, which mostly consisted of repeat-rich sequences, and herewith completion of the human genome has become within reach. This achievement is considered an important step forward towards the

complete understanding of the human genome, chromosome function, and potentially of the impact of genomic variation on disease. Will this finally resolve all missing heritability?

While soon we will be able to detect all genetic variation, most likely still a significant part will be categorized as “variants of unknown significance”. Sequencing technology is no longer holding us back, but knowledge is. A complete understanding of all (non-coding) elements of the human genome and their functional interdependencies is lacking. What are the steps that are required to collect this missing fundamental knowledge, and which next developments can we expect?

1. The introduction of integrative multi-omics approaches

The involvement of non-coding elements in disease such as enhancers, microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) is generally accepted these days.^{59,60} miRNAs are a class of small noncoding RNAs of about 22 nucleotides long with a tightly regulated expression pattern. They are considered the “fine tuners” of gene regulation, and mediate posttranscriptional gene silencing.⁶¹ Various studies have indicated the essential role of miRNAs in normal eye and ear development.^{62,63} Pathogenic variants that affect the seed region of *MIR96* have been described to underlie HL type DFNA50.⁶⁴ Also for RD, pathogenic variants affecting miRNAs have been implicated in disease: the disruption of *MIR204* results in a unique eye phenotype that includes retinal degeneration and coloboma.⁶⁵

The biological function of lncRNAs is more diverse as compared to miRNAs. LncRNAs are multi-exon transcripts of usually 1,000-10,000 nucleotides in length, and are subjected to alternative splicing and post-translational modification such as polyadenylation.⁶⁶ LncRNAs are coined “the regulators of the regulators” and are involved in regulation of gene transcription (by functioning as enhancers, enhancer regulators, or pseudogenes), in post-transcriptional regulation of gene expression (by modulating splicing or acting as miRNA decoys) and in epigenetic regulation (by altering chromatin histone modifications).⁶⁷ The involvement of lncRNAs in disease is less explored, although this topic is gradually receiving more attention. LncRNAs have a high degree of tissue specificity and thousands are shown to be expressed in human inner ear and retinal tissues.⁶⁸⁻⁷⁰ Over 20 of these lncRNAs show potential to influence expression of genes already known to be important for inner ear function in humans and mice among which a lncRNA that controls expression of *miR96* (DFNA50).⁷¹ Also, the lncRNA *MALAT1* has been suggested to have a protective function in the development of glaucoma⁷², whereas elevated *MALAT1* expression levels have been reported in patients diagnosed

with proliferative vitreoretinopathy.⁷³ These examples already illustrate the diverse roles and functionalities of lncRNAs, and that several lncRNA genes could be potential candidate genes for inherited RD or HL.

Although the involvement of non-coding RNAs in disease is evident, it is extremely challenging to establish causality between variants affecting these elements and disease. To be able to interpret the effect of a non-coding variant, all connections should be mapped between regulatory genes and their downstream target genes. Unfortunately, our current knowledge of these regulatory connections is far from complete. When a variant is found in close proximity of a disease-associated gene, a potential effect on gene function may seem evident, but for variants located in distantly located enhancer elements or non-coding genes, this association is less straightforward and is likely to remain unrecognized. A *cis* regulatory element can be located up to 1 Mb away from its target gene, and in some cases even can be located in intronic regions of other coding genes.^{74,75}

To fully comprehend the genomic and regulatory landscape of disease-associated genes, the generation of a tissue-specific omics-framework is required. Omics analyses, such as transcriptomics, proteomics and epigenomics, can provide a deeper understanding of genetic variation by studying the consequences of putative variants at multiple levels (**Figure 2**). An omics-framework would furthermore provide a complete atlas of all active (tissue-specific) coding and non-coding genomic elements and allow the elucidation of potential *cis*-acting events of variants of unknown significance. The next sections will provide several examples that illustrate how and why the different omics technologies should be adopted in genetic diagnostics.

Transcriptomics

The integration of RNA studies with genome sequencing data has been most widely explored. Co-expression analyses, functional reporter assays or CRISPR-mediated perturbation studies can be performed to reveal interdependencies between expression of non-coding genes and coding (disease-associated) genes. Additionally, whole transcriptome analysis with total RNA-seq can detect quantitative (up- or downregulation) or qualitative (alternative splicing) abnormalities. Several studies have proven the added value of incorporation of RNA-seq in WES or WGS analyses and showed successful implementation in a clinical setting. RNA-seq was performed on patient-derived blood samples and muscle or skin biopsies. The success rates of RNA-seq analyses reported in these studies range from 7.5%-36% when performed in cases that could not be genetically resolved using WES or WGS.⁷⁶⁻⁸⁰ By combining genomic and transcriptomic datasets, candidate splice variants could be readily validated or

discarded, differential expression levels could be detected, and in RNA-seq data novel variants could be identified that are located in or close to retained intronic sequences (pseudoexons) not covered by WES.

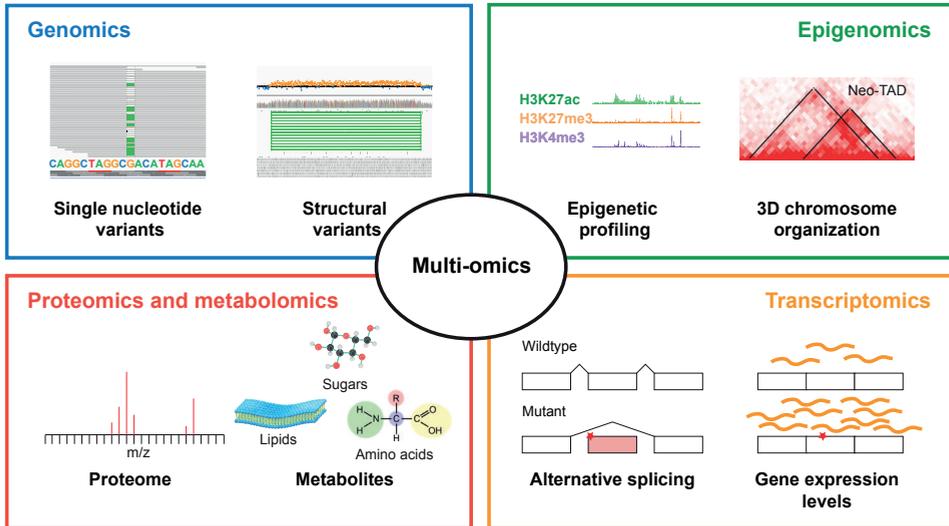


Figure 2. A schematic overview of multi-omics approaches. A multi-omics approach combines the strengths of the different omics technologies: (1) genomics allows the detection of single nucleotide variants and structural variants, (2) epigenomics allows visualization of the effect of variants on the regulatory chromosome landscape (e.g. epigenetic profiling or 3D chromosome organization) whereas (3) transcriptomics detects changes on the RNA level, this entails both qualitative changes like alternative splicing and quantitative changes that lead to up- or downregulation of gene expression. (4) Proteomics and metabolomics can be employed to investigate consequences on the proteome, aberrant protein structure or overexpression, or changes in the metabolite composition (e.g. lipids, sugars or amino acids).

A whole transcriptome analysis might not even be necessary to address all genetically unexplained cases or research questions. Mini- and midi-gene *in vitro* splice assays in HEK293T cells, or patient-derived induced pluripotent stem cells and photoreceptor precursor cells, have proven their value in the elucidation of putative splice-modulating variants in known disease genes (e.g. *ABCA4*⁸¹⁻⁸³), and the validation of candidate disease genes such as *CLRN2*, implicated in autosomal recessive HL.⁸⁴ To accelerate genetic diagnostics, the implementation of (whole) transcriptome analyses should be considered in future studies, and an informed decision should be made that includes weighing all the pros and cons of the diverse technologies to select the most suitable approach.

Proteomics and metabolomics

The integration of proteomics or metabolomics studies with genomics has been less well studied, which can mostly be attributed to the complexity of the corresponding analyses. Whereas proteomics focuses on studying changes at the proteome level (e.g. aberrant protein structure or overexpression), metabolomics can be employed to detect changes in metabolite composition (e.g. lipids, sugars or amino acids). Recently, a first study was reported, employing a combined genomics, transcriptomics and proteomics approach using fibroblast cells in a diagnostic setting for cases suspected to suffer from a mitochondrial disease.⁸⁵ Independent cumulative evidence was derived from the three techniques and a success rate of 21% was reached using the combined approach. In general, mitochondrial dysfunction is known to significantly affect cell metabolism for which the consequences can be relatively easily detected in a proteomics approach. Considering the immune-privileged nature of the eye and the ear, it is less plausible that proteomic or metabolic abnormalities caused by RD or HL-associated genetic defects, can be detected in readily accessible cells or tissues such as fibroblast and blood (**Box 1**). Therefore, the implementation of proteomics or metabolomics in HL- and RD-diagnostics is likely to be less effective. Nevertheless, in case of syndromic phenotypes that include both mitochondrial dysfunction and HL or RD, success stories have been described (e.g., *CLN3* implicated in lethal, syndromic RD (Batten disease) and *RMND1* in syndromic HL with chronic kidney disorder).^{86,87} Therefore, a combined genomics and proteomics or metabolomics approach should not be completely discarded and could still be considered dependent on the patient's phenotype.

Epigenomics

Epigenomic studies can be performed to make an inventory of chromatin modifications (epigenetic hallmarks), to reveal active regulatory elements, or the 3D chromosome organization (chromosome conformation capture techniques). The presence or absence of chromatin modifications that are associated with active *cis* regulatory elements (e.g. H3K27Ac, chromatin accessibility and transcription factor binding) can be explored with a variety of techniques such as chromatin immunoprecipitation (ChIP)-sequencing, an assay for transposase accessible chromatin (ATAC)-sequencing, and DNase I hypersensitive sites (DNase)-sequencing. However, these methods require millions of cells as input material. A recently developed Cleavage Under Targets and Tagmentation (CUT&Tag) protocol allows high-resolution profiling of important epigenetic hallmarks in a single-tube experiment.⁸⁸ This makes it possible to perform epigenomic profiling in small samples and even single cells, which is of particular interest for tissues and cell-types with limited availability as is the case for both the inner ear and the retina (**Box 1**). Pathogenic regulatory variants could have an effect on transcription factor

binding (e.g. reported for Stargardt disease⁸⁹) or alter chromatin modifications such as DNA methylation (e.g. reported for age-related macular degeneration⁹⁰). Epigenomic profiling should be performed to recognize and detect the *cis* regulatory elements that are affected by these variants.

Chromosome conformation capture techniques, such as 3C or Hi-C allow the investigation of interactions between genomic elements. With a 3C approach, putative interactions between two specific regions of interest can be investigated (one-to-one), whereas in a Hi-C approach genome-wide interactions between all elements can be mapped (all-to-all). The organization of the genome in TADs, enriched for intra-domain interactions, is crucial for maintaining promoter-enhancer interaction specificity. Using standard “C” techniques, (long-distance) interactions between two independent elements can be captured. Latest developments reporting ligation-free methods even allow the investigation of multi-region interactions (>2 elements) at an increased resolution (reviewed in (91)).

An important proof-of-concept, illustrating the value of chromosome conformation capture techniques is described in **chapter 4** of this thesis. Using a WGS approach, eight unique SVs affecting the RP17 locus were revealed in 22 families diagnosed with dominant RP. None of the genes implicated in the SVs were previously associated with retinal function or retinal disease, and without further investigation, the variants would be deemed variants of unknown significance. Instead, a low-C analysis on wildtype and patient-derived retinal organoids was performed to elucidate the 3D organization of the locus. Low-C, an adapted form of Hi-C, is optimized for low cell input which is necessary to allow its application in retinal organoids. Only after the incorporation of this technique, the pathogenic mechanism behind the RP17-associated SVs could be elucidated. Low-C profiles of retinal organoids carrying an RP17 SV indicated rewiring of enhancer-promoter contacts, leading to ectopic gene expression which was confirmed by RNA analyses. This describes a novel disease mechanism, that includes the ectopic and toxic expression of a gene. This mechanism of disease could also be involved in other unsolved Mendelian diseases. However, since this mechanism cannot be detected using a genomics-only approach, the implementation of multi-omics technologies is an important prerequisite.

2. The replacement of bulk-sequencing with single-cell approaches

Both the inner ear and the retina are considered highly complex tissues consisting of a variety of highly specialized cells. Spatiotemporal differences in gene expression (levels) between and within the distinct cell types allow the coordination of various functions that enable hearing or vision. Each cell type expresses many context-specific transcripts and splice variants and thus protein isoforms that all contribute to the overall gene expression profile. In traditional expression studies, the transcriptome or proteome of the retina or inner ear is analyzed in bulk; mRNA or protein content of the complete tissue is pooled and analyzed together.⁹² Gene expression changes that can be attributed to only small subsets of cells are easily missed. This severely complicates the analysis of expression profiles between two conditions such as affected and non-affected tissues.

Single-cell approaches such as single-cell RNA sequencing (scRNA-seq) addresses this limitation by investigating the transcriptome of individual cells in a parallelized, high-throughput fashion. Cell sorting-based approaches allow capturing of the mRNA profile of hundreds to thousands of selected cells simultaneously. Expression profiles of single cells can be directly compared, and clustered to discriminate cell populations. scRNA-seq has been successfully performed to study both ocular⁹³ and inner ear⁶² tissues. The approach allowed identification of cellular subclasses within cell types that were previously considered identical. For instance, 21 distinct clusters of amacrine cells could be distinguished in an scRNA-seq study in the mouse neural retina.⁹⁴ Important and significant differences could be observed by performing spatial and temporal expression analysis between foveal and peripheral retinal gene expression. Also in the inner ear, scRNA-seq data identified novel cellular subtypes: three subtypes of murine spiral ganglion neurons type I could be found.^{95,96}

scRNA-seq approaches can be applied to study retinal and inner ear disease, and can be expected to replace bulk-sequencing approaches in the future. Unannotated and alternatively splice transcripts in specialized cell types can be revealed, that may contain novel exons or have a regulatory function. Pathogenic variants in one of these previously unannotated exons can be easily overlooked or may have been wrongly deemed intronic and non-causative in the past. Both the inner ear and the retina express tissue-specific splicing factors that produce a large number of alternatively-spliced transcripts,^{97,98} which are still poorly characterized. scRNA-sequencing approaches are necessary to close this knowledge gap, and are required to interpret and complete our understanding of (potentially) disease-associated variants. Ray et al. stated that thousands of mRNAs are present in the retina, including the previously overlooked *CRB1*

transcript that turned out to be the predominant transcript in the retina.⁹⁹ Considering the high percentage of alternative spliced transcripts in both the retina and inner ear^{98,99}, it can be speculated that a considerable percentage of missing heritability resides in these “hidden” exons or in sequences that determine the splicing pattern.

Single cell approaches can also be employed to map the single-molecule 3D or regulatory landscape. Previously, TAD domains were considered to be conserved among tissues and cells, but there is accumulating evidence that this view is incomplete. Recent reports indicate that 20-80% of the CTCF-enriched TAD boundaries are not shared between cell types.¹⁰⁰ Additionally, there is great intra-TAD variability of interactions within TADs that is mainly caused by differences in enhancer accessibility.¹⁰¹ Single-cell platforms are rapidly being developed that allow mapping of the epigenome landscape of cell subpopulations via e.g. single cell CUT&Tag¹⁰², ATAC sequencing¹⁰³ and Hi-C.¹⁰⁴

Chapter 2 describes the presence of a long, ubiquitous *KIAA1549* transcript, and a short, retina-specific, *KIAA1549* transcript. Depending on the variant type and the affected *KIAA1549* transcript, pathogenic variants in *KIAA1549* lead to mild or more severe forms of RP.

This study left some questions unresolved, including the function of the individual isoforms, the signaling pathways involved and the cellular localization of the encoded proteins within the retina. A scRNA-seq approach could help to elucidate the function of the different transcripts and map the transcripts to retinal cellular subtypes. This could increase our understanding related to the different transcripts, their interplay and possibly the regulatory elements involved. With this, potentially, essential insights can be gained that will help in variant interpretation and patient counseling.

A complete understanding of tissue-specific alternative splicing patterns is not only crucial for diagnostic purposes, modulation of alternative splicing has also been shown to be instrumental in the design of novel genetic therapeutic strategies such as exon skipping.¹¹⁶ Together with the arrival of long-read sequencing technologies, that allow full-length transcript annotation, the application of scRNA-seq will allow the generation of a complete atlas optimally representing the diverse retina and inner ear transcriptome landscape. Nevertheless, as discussed in **Box 1**, the limited availability of tissue samples to study vision and hearing disorders severely hampers an efficient (diagnostic) application of this technique.

BOX 1 - A limitation: Tissue availability to study vision and hearing diseases

For both multi-omics approaches and single cell approaches, it is crucial that the optimal tissue of interest is selected for investigation. However, tissue availability imposes a major limitation to study both retinal and inner ear diseases. Over the years, different animal models have been employed to study sensory diseases, which led to the successful validation of several candidate disease genes (for example *EYS*-associated RD¹⁰⁵ and *TRRAP*-associated HL¹⁰⁶ in zebrafish). The use of animal models is time and money consuming, and raises several ethical concerns. Due to the evolutionary distance, and potential differences in physiology, findings in animal models are not always easily translatable to the patient situation. For example, disease genes are not always conserved among species (e.g. *EYS*¹⁰⁵), or exert the same function (e.g. *DFNA5*¹⁰⁷).

Patient-derived induced pluripotent stem cells (iPSC) provide a valuable alternative. Several protocols are available that allow the differentiation of iPSCs into cells that optimally mimic the sensory (precursor) cells of the eye and ear.¹⁰⁸⁻¹¹⁰ Many of the omics and single cell technologies mentioned in sections 1 and 2 have been successfully applied in stem cell-derived models and crucial insights were gained regarding disease mechanisms. Novel splice defects have been identified using 2D photoreceptor precursor cells⁸³, and enhanced splice and ciliation defects were previously observed in 3D retinal organoids compared to fibroblast cells.¹¹¹ Additionally, these cell models have proven value for *in vitro* validation of personalized therapies: an AON treatment in 3D retinal organoid and RPE cells derived from a patient with pathogenic *CEP290* variants rescued the ciliary defects.¹¹²

Although efficient protocols are available to produce cochlear hair cells or retinal photoreceptors, not all cellular specializations or cell types are equally represented or generated in the differentiated cell models. In most retinal organoid models, only limited development of photoreceptor outer segment discs is described, and neuronal cell types often display a fast degeneration.¹⁰⁸ Also, the development of retinal organoids is significantly slower than that of the human native retina.¹¹³ Direct interactions of photoreceptor cells with the retinal pigment epithelial (RPE) layer are absent, which are essential when studying multifactorial diseases that affect the multiple cell layers. The latter may be resolved by generating multi-layer tissue models in a retina-on-a-chip platform.¹¹⁴

The established differentiation protocols for inner ear organoids still have some important limitations as well. Current iPSC-derived inner ear models only represent a small subset of the diverse cell populations in the inner ear and the main focus has been on the generation of sensory cell types.¹¹⁰ However, many of the HL-associated genes are expressed in non-sensory or mesenchymal cell populations and therefore more attention should be paid to a broader set of cell types to study pathogenesis for defects of these genes as well. To illustrate, an important example is the investigation of *SLC26A4*-associated pathogenic mechanisms). As described in **chapter 5**, the genetic defect of the CEVA haplotype found in the majority of monoallelic patients is still unknown. A logical next step would be to interrogate the RNA for aberrant expression levels of *SLC26A4* or splice alterations, which would require a suitable cell model with detectable expression levels. *SLC26A4* is not expressed in readily available cell types (e.g. blood and fibroblasts) and also not induced in iPSCs or otic progenitor cells.¹¹⁵ Fortunately, a differentiation protocol aimed to specifically generate sulcus-like cells that do express *SLC26A4* has been described.^{11,115} The sulcus-like cells generated by this protocol are less complex than the multi-cellular organoid 3D systems. However, they can still potentially help to overcome the current deadlock in transcript analysis to unravel *SLC26A4* pathogenicity.

3. Looking beyond Mendelian genetics

Inherited HL and RD are generally considered monogenic disorders that follow the laws of Mendel. However, the increasing reports of high inter- and intrafamilial phenotypic variability that can be associated with a single pathogenic variant, and the occurrence of incomplete penetrance, show that this picture is far from complete. In view of missing heritability, shifting our attention from monogenic (Mendelian) inheritance towards oligogenic, polygenic or multifactorial inheritance could be a crucial next step.

Instead of a black-white division between monogenic and multifactorial diseases, one should consider a continuum from monogenic diseases, via near-Mendelian traits to complex multifactorial diseases with a large environmental component (**Figure 3**). There is an inverse relationship between variant allele frequencies and variant effect sizes. For some disorders, such as age-related HL and macular degeneration, multifactorial inheritance is clear. For other disorders assumed to be monogenic for decades, such as DFNA21 and a subset of Stargardt disease cases, a near-Mendelian inheritance is likely to be a better fit. The heritability of age-related HL is estimated to be 35% to 70%.¹¹⁷ GWAS studies have indicated that age-related HL is highly polygenic in nature, although several genetic risk loci could be identified.^{56,57} These data confirm that the genetic risk to develop age-related HL is determined by a combination of common variants (population allele frequency >5%) with small effect sizes. Stargardt disease, on the other hand, has been considered a monogenic disorder for years, where biallelic pathogenic variants in *ABCA4* cause macular disease. Today, Stargardt disease cases carrying the mild p.(Asn1868Ile) variant in *trans* with a severe *ABCA4* variant, are considered being non-Mendelian. This is based on the high population allele frequency of this mild variant (~6%), the identification of unaffected individuals carrying the p.Asn1868Ile variant in *trans* with a severe *ABCA4* variant, and an imbalance in sex ratio in patients with this genotype.¹¹⁸ As proposed by Runhart et al., approximately 25% of the Stargardt disease cases show oligogenic, polygenic or multifactorial inheritance, and together with other (unidentified) modifiers, sex should be considered as a potential disease-modifying variable.^{118,119}

Similarly as for Stargardt disease, it can be argued that DFNA21 also is a near-Mendelian disorder. Recently, we learned that the c.1696_1707del *RIPOR2* variant that underlies DFNA21 is presumed to be the most frequent cause of inherited adult-onset HL in Northwest Europe. In the Netherlands, 13,000 individuals are estimated to carry this variant and they are therefore at risk to develop HL. One of the most essential remaining questions is whether all carriers will develop HL during life, or that other (non-)genetic protective or damaging modifiers are involved as well. Based on the DFNA21-affected

families described in **chapter 3.1**, the *RIPOR2* related phenotype can be considered highly variable with an age of onset that ranges from congenital to 70 years. Also within families, significant phenotypic variability in age of onset and severity of HL can be observed. These data suggest that genetic and/or environmental modifiers are likely to contribute to the phenotypic outcome. In **chapter 3.1**, some effort was taken to identify potential modifiers such as noise exposure or *RIPOR2* transcript levels in blood, but no potential modifier was found yet. More in-depth studies, that are focused on including a larger patient cohort and investigating more relevant tissue samples, should be performed. Additionally, in view of genotype-phenotype correlations and the understanding the potential near-Mendelian inheritance of *DFNA21*, a large natural history study should be initiated to gain a full understanding of all aspects of *RIPOR2*-associated HL.

Besides the possibilities of digenic, oligogenic, polygenic, and multifactorial inheritance patterns, also other non-Mendelian inheritance patterns should receive more attention in diagnostic pipelines.

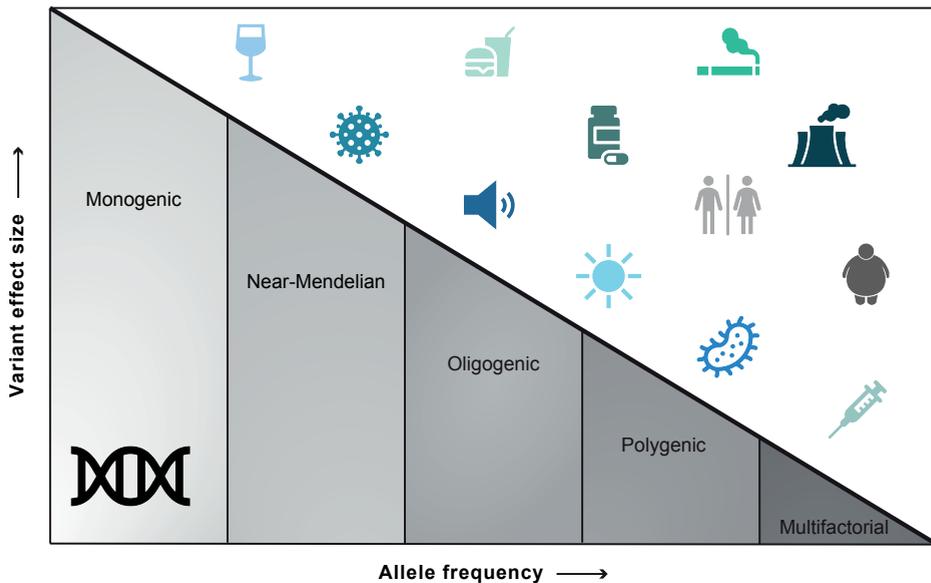


Figure 3. Continuum of monogenic to polygenic disorders. The spectrum of inherited disorders can be considered a continuum, with monogenic disorders as one of the extremes and associated with rare variants with a large effect size, via near-Mendelian disorders, towards oligogenic, polygenic or multifactorial disorders that are caused by a combination of common variants with small effects sizes and an increased contribution of environmental factors.

Examples are mitochondrial inheritance and somatic mosaicism that could lead to mosaic presence of the pathogenic variant in different cellular subtypes. A case study performed by de Kok et al. reported somatic mosaicism for a *POU3F4* variant (associated with HL type DFNX2); semiquantitative analysis revealed that the variant was present in approximately 50% of peripheral blood lymphocytes of the affected individual.¹²⁰ The authors hypothesized that most likely, the variant occurred in early embryogenesis before occurrence of the otic vesicle. Mitochondrial DNA is not routinely investigated in clinical diagnostics, however, pathogenic variants in mitochondrial genes have been described for both HL and RD.^{31,121}

Based on current observations, for the many genetically unexplained isolated cases diagnosed with HL and RD, a non-Mendelian inheritance pattern should be considered. To readily implement this in genetic diagnostics, data analysis pipelines and workflows for sequencing data should be adapted. Variants considered too common should no longer be readily disregarded as they could be part of a complex oligogenic or polygenic inheritance pattern. In case of monoallelic pathogenic variants, one should not only focus on the mutated gene, but should also consider pathogenic variants in other disease-associated (e.g. pathogenic *FOXI1* variants for monoallelic *SLC26A4* patients, **chapter 5**). Digenic inheritance has also been suggested for other HL- or RD-associated genes (e.g. *CDH23/PCDH15*¹²² and *CEP250/PCARE*¹²³) which, however, could not be confirmed due to the lack of additional cases.¹²⁴ Last but not least, to make this process feasible, a complete work-up of family history is required, and careful clinical evaluation of all family members (if possible), including unaffected siblings, should be performed to allow to establish phenotype-genotype correlations. As is clear from the above, there is an important need to start looking beyond “standard” Mendelian genetics. If we do this, hopefully, a significant part of genetically unsolved cases will receive a genetic, a non-genetic or combinatory explanation for their phenotype.

FINAL REMARKS

With the developments discussed in the previous sections (multi-omics approaches, single-cell sequencing technologies and non-Mendelian inheritance patterns), a significant portion of missing heritability could be resolved. However, how realistic is a fast implementation of these technologies in clinical practice? Implementation often requires collection of additional tissue biopsies, (fresh) blood samples, the establishment of new lab workflows and analytic pipelines.¹²⁵ Incorporation of the new technologies and approaches in standard procedures will require extreme efforts that are time- and money-consuming. Also, once implemented, the increased amount of data that need

to be collected for an individual case also requires a significant amount of time and expertise to be analyzed. How cost- and time-effective will the application of these technologies be? Currently, in a diagnostic setting, the sequencing and analysis costs of a single exome are around €1,000 and even higher when counseling is included.¹²⁶ The estimated turnaround time for a diagnostic report in the Netherlands is already three months. How much will these costs and turnaround time increase when multi-omics approaches will be adopted? Most realistically, the described technologies will remain in the research domains for many years to come, and a genomics-only approach (based on WES, or potentially WGS) will remain the first-line of screening in routine diagnostics.

Besides time and money concerns, another important hurdle is the lack of interpretation guidelines. Currently, ACMG guidelines are not fitted to validate non-coding variation and there is an urgent need for an established framework.^{127,128} For example, when aberrant gene expression levels are detected in RNA-seq experiment, which cut-off should be used to distinguish pathogenic from non-pathogenic differential expression changes? What is the minimum level of wildtype gene expression that is required to prevent disease? Exon-skipping studies have indicated that already ~20% of wildtype transcript can be enough for phenotypic rescue.^{116,129} Additionally, when can an elevated expression level be considered toxic? Furthermore, there are tissue-related concerns. Most likely, routine RNA-seq would be performed in readily available cell types such as blood cells and fibroblasts. Although a significant number of HL- and RD-associated genes are expressed in these tissue types, the majority of these genes are not or at a too low extent and differentiation of patient-derived iPSCs will increase costs and turnaround times even more. Also, how translatable will the findings be?

Long lists of candidate disease-associated genes are available for both HL and RD, for which putative pathogenic variants have only been reported in single cases or families. Evidently, the “low hanging fruit” has already been picked, and only the rare causes of disease remain. While collecting additional cases, which requires worldwide data sharing efforts (e.g. ClinGen¹³⁰ or GeneMatcher¹³¹) is still an option that should be explored, we should also consider the possibility that these mutational events are unique for these specific families. In these cases, functional and multi-omics analyses and potentially testing in animal models are required to provide conclusive evidence for causality of variants in these candidate disease genes. Is it feasible to perform all these analyses for all unique cases?

Despite the fact that a wide implementation is unlikely, this thesis exemplifies that the use of the indicated technologies will provide important insights that will aid in the identification of the underlying genetic defects. Crucial lessons have been learned from

the past, and pathogenic variants that used to be overlooked are now finally recognized as potentially pathogenic. Important examples are described in this thesis. Frequent causes of RD and HL that were previously missed or misinterpreted can now be readily detected with a targeted analysis strategy (RP17-SVs for dominant RP cases, an in-frame *RIPOR2* deletion for dominant HL). A novel disease mechanism, involving the ectopic expression of a gene, was identified. With these findings, hundreds of individuals finally received a genetic diagnosis. Additionally, we have learned to think outside of the box, and to question the established paradigms on the heritability of RD and HL.

The ultimate key to future success, is to acknowledge that we do not understand all the mysteries of the human genome yet. There is still enough to learn, and there are phenomena that cannot be explained yet. Revisiting the “junk DNA” to recognize that essential, functional elements are hidden in the non-coding world was a crucial first step. To be able to find a genetic explanation for all unsolved cases of inherited HL or RD in the years to come, we should continue this road trip and be ready for new adventures. We have to dive into the depths of the human genome and be open for alternative genetic explanations that can involve complex regulatory and non-Mendelian mechanisms.

We have to *broaden the genomic landscape*.

REFERENCES

1. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
2. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
3. Schloss, J.A. How to get genomes at one ten-thousandth the cost. *Nature Biotechnology* **26**, 1113-1115 (2008).
4. Haer-Wigman, L., van Zelst-Stams, W.A., Pfundt, R., van den Born, L.I., Klaver, C.C., Verheij, J.B. *et al.* Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *European Journal of Human Genetics* **25**, 591-599 (2017).
5. Tiwari, A., Bahr, A., Bähr, L., Fleischhauer, J., Zinkernagel, M.S., Winkler, N. *et al.* Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Scientific Reports* **6**, 28755 (2016).
6. Xu, Y., Guan, L., Shen, T., Zhang, J., Xiao, X., Jiang, H. *et al.* Mutations of 60 known causative genes in 157 families with retinitis pigmentosa based on exome sequencing. *Human Genetics* **133**, 1255-1271 (2014).
7. Abu-Safieh, L., Alrashed, M., Anazi, S., Alkuraya, H., Khan, A.O., Al-Owain, M. *et al.* Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Research* **23**, 236-247 (2013).
8. Sloan-Heggen, C.M., Bierer, A.O., Shearer, A.E., Kolbe, D.L., Nishimura, C.J., Frees, K.L. *et al.* Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Human Genetics* **135**, 441-450 (2016).
9. Zazo Seco, C., Wesdorp, M., Feenstra, I., Pfundt, R., Hehir-Kwa, J.Y., Lelieveld, S.H. *et al.* The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. *European Journal of Human Genetics* **25**, 308-314 (2017).
10. Oonk, A.M.M., Ariens, S., Kunst, H.P.M., Admiraal, R.J.C., Kremer, H. & Pennings, R.J.E. Psychological impact of a genetic diagnosis on hearing impairment-An exploratory study. *Clinical Otolaryngology* **43**, 47-54 (2018).
11. Astuto, L.M., Bork, J.M., Weston, M.D., Askew, J.W., Fields, R.R., Orten, D.J. *et al.* CDH23 mutation and phenotype heterogeneity: A profile of 107 diverse families with usher syndrome and nonsyndromic deafness. *American Journal of Human Genetics* **71**, 262-275 (2002).
12. Toualbi, L., Toms, M. & Moosajee, M. USH2A-retinopathy: from genetics to therapeutics. *Experimental Eye Research* **201**, 108330 (2020).
13. Bayefsky, M.J. Comparative preimplantation genetic diagnosis policy in Europe and the USA and its implications for reproductive tourism. *Reproductive Biomedicine & Society Online* **3**, 41-47 (2016).
14. van der Meij, K.R.M., Siermans, E.A., Macville, M.V.E., Stevens, S.J.C., Bax, C.J., Bekker, M.N. *et al.* TRIDENT-2: national implementation of genome-wide non-invasive prenatal testing as a first-tier screening test in the Netherlands. *American Journal of Human Genetics* **105**, 1091-1101 (2019).

15. Chiu, R.W.K., Chan, K.C.A., Gao, Y., Lau, V.Y.M., Zheng, W., Leung, T.Y. *et al.* Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proceedings of the National Academy of Sciences* **105**, 20458-20463 (2008).
16. Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L. & Quake, S.R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proceedings of the National Academy of Sciences* **105**, 16266-16271 (2008).
17. Le Meur, G., Lebranchu, P., Billaud, F., Adjali, O., Schmitt, S., Bézieau, S. *et al.* Safety and long-term efficacy of AAV4 gene therapy in patients with RPE65 leber congenital amaurosis. *Molecular Therapy* **26**, 256-268 (2018).
18. Testa, F., Maguire, A.M., Rossi, S., Pierce, E.A., Melillo, P., Marshall, K. *et al.* Three-year follow-up after unilateral subretinal delivery of adeno-associated virus in patients with Leber congenital Amaurosis type 2. *Ophthalmology* **120**, 1283-1291 (2013).
19. Weleber, R.G., Pennesi, M.E., Wilson, D.J., Kaushal, S., Erker, L.R., Jensen, L. *et al.* Results at 2 years after gene therapy for RPE65-deficient leber congenital amaurosis and severe early-childhood-onset retinal dystrophy. *Ophthalmology* **123**, 1606-1620 (2016).
20. Vázquez-Domínguez, I., Garanto, A. & Collin, R.W.J. Molecular therapies for inherited retinal diseases-current standing, opportunities and challenges. *Genes (Basel)* **10**(2019).
21. Delmaghani, S. & El-Amraoui, A. Inner ear gene therapies take off: Current promises and future challenges. *Journal of Clinical Medicine* **9**, 2309 (2020).
22. Cideciyan, A.V., Jacobson, S.G., Ho, A.C., Garafalo, A.V., Roman, A.J., Sumaroka, A. *et al.* Durable vision improvement after a single treatment with antisense oligonucleotide sepfarsen: a case report. *Nature Medicine*, 785-789 (2021).
23. Darrow, J.J. Luxturna: FDA documents reveal the value of a costly gene therapy. *Drug Discovery Today* **24**, 949-954 (2019).
24. Kim, J., Hu, C., Moufawad El Achkar, C., Black, L.E., Douville, J., Larson, A. *et al.* Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *New England Journal of Medicine* **381**, 1644-1652 (2019).
25. Haer-Wigman, L., van der Schoot, V., Feenstra, I., Vulto-van Silfhout, A.T., Gilissen, C., Brunner, H.G. *et al.* 1 in 38 individuals at risk of a dominant medically actionable disease. *European Journal of Human Genetics* **27**, 325-330 (2019).
26. Hanany, M., Rivolta, C. & Sharon, D. Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases. *Proceedings of the National Academy of Sciences* **117**, 2710-2716 (2020).
27. Sallevelt, S.C.E.H., Stegmann, A.P.A., de Koning, B., Velter, C., Steyls, A., van Esch, M. *et al.* Diagnostic exome-based preconception carrier testing in consanguineous couples: results from the first 100 couples in clinical practice. *Genetics in Medicine* **Online ahead of print** (2021).
28. Kirk, E.P., Barlow-Stewart, K., Selvanathan, A., Josephi-Taylor, S., Worgan, L., Rajagopalan, S. *et al.* Beyond the panel: preconception screening in consanguineous couples using the TruSight One "clinical exome". *Genetics in Medicine* **21**, 608-612 (2019).

29. Teeuw, M., Waisfisz, Q., Zwijnenburg, P.J., Sistermans, E.A., Weiss, M.M., Henneman, L. *et al.* First steps in exploring prospective exome sequencing of consanguineous couples. *European Journal of Human Genetics* **57**, 613-616 (2014).
30. Mor-Shaked, H., Rips, J., Gershon Naamat, S., Reich, A., Elpeleg, O., Meiner, V. *et al.* Parental exome analysis identifies shared carrier status for a second recessive disorder in couples with an affected child. *European Journal of Human Genetics* **29**, 455-462 (2021).
31. RetNet. Available from: <https://sph.uth.edu/RetNet/>.
32. Van Camp, G. & Smith, R. Hereditary Hearing Loss Homepage. Available from: <https://hereditaryhearingloss.org/>.
33. Cremers, F.P.M., van de Pol, D.J., van Kerkhoff, L.P., Wieringa, B. & Ropers, H.H. Cloning of a gene that is rearranged in patients with choroïdæmia. *Nature* **347**, 674-677 (1990).
34. de Kok, Y.J., van der Maarel, S.M., Bitner-Grindzicz, M., Huber, I., Monaco, A.P., Malcolm, S. *et al.* Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. *Science* **267**, 685-688 (1995).
35. Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W. *et al.* A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**, 364-366 (1990).
36. Bowl, M.R., Simon, M.M., Ingham, N.J., Greenaway, S., Santos, L., Cater, H. *et al.* A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nature Communications* **8**, 886 (2017).
37. Moore, B.A., Leonard, B.C., Sebbag, L., Edwards, S.G., Cooper, A., Imai, D.M. *et al.* Identification of genes required for eye development by high-throughput screening of mouse knockouts. *Communications Biology* **1**, 236-236 (2018).
38. Ingham, N.J., Pearson, S.A., Vancollie, V.E., Rook, V., Lewis, M.A., Chen, J. *et al.* Mouse screen reveals multiple new genes underlying mouse and human hearing loss. *PLoS Biology* **17**, e3000194 (2019).
39. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508-514 (2016).
40. Kunst, H., Marres, H., Huygen, P., van Duijnhoven, G., Krebsova, A., van der Velde, S. *et al.* Non-syndromic autosomal dominant progressive non-specific mid-frequency sensorineural hearing impairment with childhood to late adolescence onset (DFNA21). *Clinical Otolaryngology and Allied Sciences* **25**, 45-54 (2000).
41. Nikopoulos, K., Gilissen, C., Hoischen, A., Erik van Nouhuys, C., Boonstra, F.N., Blokland, E.A.W. *et al.* Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. *American Journal of Human Genetics* **86**, 240-247 (2010).
42. Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M.K., Thornton, A.M., Roeb, W. *et al.* Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *American Journal of Human Genetics* **87**, 90-94 (2010).
43. Mattick, J.S., Dinger, M., Schonrock, N. & Cowley, M. Whole genome sequencing provides better diagnostic yield and future value than whole exome sequencing. *Medical Journal of Australia* **209**, 197-199 (2018).

44. Carss, K.J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S. *et al.* Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *American Journal of Human Genetics* **100**, 75-90 (2017).
45. Van Schil, K., Naessens, S., Van de Sompele, S., Carron, M., Aslanidis, A., Van Cauwenbergh, C. *et al.* Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genetics in Medicine* **20**, 202-213 (2018).
46. Shearer, A.E., Kolbe, D.L., Azaiez, H., Sloan, C.M., Frees, K.L., Weaver, A.E. *et al.* Copy number variants are a common cause of non-syndromic hearing loss. *Genome Medicine* **6**, 37 (2014).
47. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I. *et al.* Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535-548 (2019).
48. Riepe, T.V., Khan, M., Roosing, S., Cremers, F.P.M. & 't Hoen, P.A.C. Benchmarking deep learning splice prediction tools using functional splice assays. *Human Mutation* **In press** (2021).
49. Telenti, A., Lippert, C., Chang, P.-C. & DePristo, M. Deep learning of genomic variation and regulatory network data. *Human Molecular Genetics* **27**, r63-r71 (2018).
50. Cai, L., Wu, Y. & Gao, J. DeepSV: accurate calling of genomic deletions from high-throughput sequencing data using deep convolutional neural network. *BMC Bioinformatics* **20**, 665 (2019).
51. Weininger, O., Warnecke, A., Lesinski-Schiedat, A., Lenarz, T. & Stolle, S. Computational analysis based on audioprofiles: A new possibility for patient stratification in office-based otology. *Audiology Research* **9**, 27-32 (2019).
52. Salah, M., de Varebeke, S.J., Franssen, E., Topsakal, V., Van Camp, G. & Van Rompaey, V. Predictive sensitivity and concordance of machine-learning tools for diagnosing DFNA9 in allLarge series of p.Pro51Ser variant carriers in the COCH-gene. *Otology & Neurotology*, 671-677 (2021).
53. Renaux, A., Papadimitriou, S., Versbraegen, N., Nachtegael, C., Boutry, S., Nowé, A. *et al.* ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. *Nucleic Acids Research* **47**, w93-w98 (2019).
54. Papadimitriou, S., Gazzo, A., Versbraegen, N., Nachtegael, C., Aerts, J., Moreau, Y. *et al.* Predicting disease-causing variant combinations. *Proceedings of the National Academy of Sciences* **116**, 11878 (2019).
55. Kaplanis, J., Samocha, K.E., Wiel, L., Zhang, Z., Arvai, K.J., Eberhardt, R.Y. *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757-762 (2020).
56. Franssen, E., Bonneux, S., Corneveaux, J.J., Schrauwen, I., Di Berardino, F., White, C.H. *et al.* Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. *European Journal of Human Genetics* **23**, 110-115 (2015).
57. Wells, H.R.R., Freidin, M.B., Zainul Abidin, F.N., Payton, A., Dawes, P., Munro, K.J. *et al.* GWAS identifies 44 independent associated genomic loci for self-reported adult hearing difficulty in UK biobank. *American Journal of Human Genetics* **105**, 788-802 (2019).

58. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A. *et al.* Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **585**, 79-84 (2020).
59. Booth, K.T., Azaiez, H., Jahan, I., Smith, R.J.H. & Fritzsche, B. Intracellular regulome variability along the organ of corti: Evidence, approaches, challenges, and perspective. *Frontiers in Genetics* **9**, 156-156 (2018).
60. Zhang, L., Dong, Y., Wang, Y., Gao, J., Lv, J., Sun, J. *et al.* Long non-coding RNAs in ocular diseases: new and potential therapeutic targets. *FEBS Journal* **286**, 2261-2272 (2019).
61. Guo, H., Ingolia, N.T., Weissman, J.S. & Bartel, D.P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840 (2010).
62. Mittal, R., Liu, G., Polineni, S.P., Bencie, N., Yan, D. & Liu, X.Z. Role of microRNAs in inner ear development and hearing loss. *Gene* **686**, 49-55 (2019).
63. Karali, M., Persico, M., Mutarelli, M., Carissimo, A., Pizzo, M., Singh Marwah, V. *et al.* High-resolution analysis of the human retina miRNome reveals isomiR variations and novel microRNAs. *Nucleic Acids Research* **44**, 1525-1540 (2016).
64. Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L. *et al.* Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nature Genetics* **41**, 609-613 (2009).
65. Conte, I., Hadfield, K.D., Barbato, S., Carrella, S., Pizzo, M., Bhat, R.S. *et al.* MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proceedings of the National Academy of Sciences* **112**, e3236-e3245 (2015).
66. Novikova, I.V., Hennelly, S.P. & Sanbonmatsu, K.Y. Sizing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? *Bioarchitecture* **2**, 189-199 (2012).
67. Quinn, J.J. & Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics* **17**, 47-62 (2016).
68. Schrauwen, I., Hasin-Brumshstein, Y., Corneveaux, J.J., Ohmen, J., White, C., Allen, A.N. *et al.* A comprehensive catalogue of the coding and non-coding transcripts of the human inner ear. *Hearing Research* **333**, 266-274 (2016).
69. Farkas, M.H., Grant, G.R., White, J.A., Sousa, M.E., Consugar, M.B. & Pierce, E.A. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics* **14**, 486 (2013).
70. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research* **22**, 1775-1789 (2012).
71. Ushakov, K., Koffler-Brill, T., Rom, A., Perl, K., Ulitsky, I. & Avraham, K.B. Genome-wide identification and expression profiling of long non-coding RNAs in auditory and vestibular systems. *Scientific Reports* **7**, 8637 (2017).
72. Wang, L., Gong, J., Wang, J., Dan, J. & Wang, P. Long non-coding RNA MALAT1 alleviates the elevated intraocular pressure (eiop)-induced glaucoma progression via sponging miR149-5P. *Current Eye Research*, 1-9 (2020).

73. Zhou, R.M., Wang, X.Q., Yao, J., Shen, Y., Chen, S.N., Yang, H. *et al.* Identification and characterization of proliferative retinopathy-related long noncoding RNAs. *Biochemical and Biophysical Research Communications* **465**, 324-330 (2015).
74. de Kok, Y.J., Vossenaar, E.R., Cremers, C.W., Dahl, N., Laporte, J., Hu, L.J. *et al.* Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Human Molecular Genetics* **5**, 1229-1235 (1996).
75. Naranjo, S., Voeseinek, K., de la Calle-Mustienes, E., Robert-Moreno, A., Kokotas, H., Grigoriadou, M. *et al.* Multiple enhancers located in a 1-Mb region upstream of POU3F4 promote expression during inner ear development and may be required for hearing. *Human Genetics* **128**, 411-419 (2010).
76. Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science Translational Medicine* **9**, eaal5209 (2017).
77. Kremer, L.S., Bader, D.M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature Communications* **8**, 15824 (2017).
78. Frésard, L., Smail, C., Ferraro, N.M., Teran, N.A., Li, X., Smith, K.S. *et al.* Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nature Medicine* **25**, 911-919 (2019).
79. Gonorazky, H.D., Naumenko, S., Ramani, A.K., Nelakuditi, V., Mashouri, P., Wang, P. *et al.* Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *American Journal of Human Genetics* **104**, 466-483 (2019).
80. Yépez, V.A., Gusic, M., Kopajtich, R., Mertes, C., Smith, N.H., Alston, C.L. *et al.* Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *medRxiv*, 21254633 (2021).
81. Sangermano, R., Garanto, A., Khan, M., Runhart, E.H., Bauwens, M., Bax, N.M. *et al.* Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genetics in Medicine* **21**, 1751-1760 (2019).
82. Bauwens, M., Garanto, A., Sangermano, R., Naessens, S., Weisschuh, N., De Zaeytijd, J. *et al.* ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genetics in Medicine* **21**, 1761-1771 (2019).
83. Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N.M., Hoyng, C.B. *et al.* Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. *American Journal of Human Genetics* **102**, 517-527 (2018).
84. Vona, B., Mazaheri, N., Lin, S.J., Dunbar, L.A., Maroofian, R., Azaiez, H. *et al.* A biallelic variant in CLRN2 causes non-syndromic hearing loss in humans. *Human Genetics*, 915-931 (2021).
85. Kopajtich, R., Smirnov, D., Stenton, S.L., Loipfinger, S., Meng, C., Scheller, I.F. *et al.* Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. *medRxiv*, 21253187 (2021).

86. Gaboon, N.E.A., Banaganapalli, B., Nasser, K., Razeeth, M., Alsaedi, M.S., Rashidi, O.M. *et al.* Exome sequencing and metabolomic analysis of a chronic kidney disease and hearing loss patient family revealed RMND1 mutation induced sphingolipid metabolism defects. *Saudi Journal of Biological Sciences* **27**, 324-334 (2020).
87. Schmidtke, C., Tiede, S., Thelen, M., Käkälä, R., Jabs, S., Makrypidi, G. *et al.* Lysosomal proteome analysis reveals that CLN3-defective cells have multiple enzyme deficiencies associated with changes in intracellular trafficking. *Journal of Biological Chemistry* **294**, 9592-9604 (2019).
88. Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications* **10**, 1930 (2019).
89. Cherry, T.J., Yang, M.G., Harmin, D.A., Tao, P., Timms, A.E., Bauwens, M. *et al.* Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. *Proceedings of the National Academy of Sciences* **117**, 9001-9012 (2020).
90. Oliver, V.F., Jaffe, A.E., Song, J., Wang, G., Zhang, P., Branham, K.E. *et al.* Differential DNA methylation identified in the blood and retina of AMD patients. *Epigenetics* **10**, 698-707 (2015).
91. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nature Reviews Genetics* **21**, 207-226 (2020).
92. Voigt, A.P., Mulfaul, K., Mullin, N.K., Flamme-Wiese, M.J., Giacalone, J.C., Stone, E.M. *et al.* Single-cell transcriptomics of the human retinal pigment epithelium and choroid in health and macular degeneration. *Proceedings of the National Academy of Sciences* **116**, 24100-24107 (2019).
93. Lukowski, S.W., Lo, C.Y., Sharov, A.A., Nguyen, Q., Fang, L., Hung, S.S. *et al.* A single-cell transcriptome atlas of the adult human retina. *EMBO Journal* **38**, e100811 (2019).
94. Macosko, Evan Z., Basu, A., Satija, R., Nemesht, J., Shekhar, K., Goldman, M. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202-1214 (2015).
95. Sun, S., Babola, T., Pregernig, G., So, K.S., Nguyen, M., Su, S.-S.M. *et al.* Hair cell mechanotransduction regulates spontaneous activity and spiral ganglion subtype specification in the auditory system. *Cell* **174**, 1247-1263 (2018).
96. Shrestha, B.R., Chia, C., Wu, L., Kujawa, S.G., Liberman, M.C. & Goodrich, L.V. Sensory neuron diversity in the inner ear is shaped by activity. *Cell* **174**, 1229-1246 (2018).
97. Aísa-Marín, I., García-Arroyo, R., Mirra, S. & Marfany, G. The alter retina: alternative splicing of retinal genes in health and disease. *International Journal of Molecular Sciences* **22**, 1855 (2021).
98. Ranum, P.T., Goodwin, A.T., Yoshimura, H., Kolbe, D.L., Walls, W.D., Koh, J.-Y. *et al.* Insights into the biology of hearing and deafness revealed by single-cell RNA sequencing. *Cell Reports* **26**, 3160-3171 (2019).
99. Ray, T.A., Cochran, K., Kozłowski, C., Wang, J., Alexander, G., Cady, M.A. *et al.* Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. *Nature Communications* **11**, 3328 (2020).

100. McArthur, E. & Capra, J.A. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *American Journal of Human Genetics* **108**, 269-283 (2021).
101. Barrington, C., Georgopoulou, D., Pezic, D., Varsally, W., Herrero, J. & Hadjur, S. Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nature Communications* **10**, 2908-2908 (2019).
102. Wu, S.J., Furlan, S.N., Mihalas, A.B., Kaya-Okur, H.S., Feroze, A.H., Emerson, S.N. *et al.* Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nature Biotechnology* **Online ahead of print** (2021).
103. Lareau, C.A., Duarte, F.M., Chew, J.G., Kartha, V.K., Burkett, Z.D., Kohlway, A.S. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nature Biotechnology* **37**, 916-924 (2019).
104. Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).
105. Messchaert, M., Dona, M., Broekman, S., Peters, T.A., Corral-Serrano, J.C., Slijkerman, R.W.N. *et al.* Eyes shut homolog is important for the maintenance of photoreceptor morphology and visual function in zebrafish. *PLOS ONE* **13**, e0200789 (2018).
106. Xia, W., Hu, J., Ma, J., Huang, J., Wang, X., Jiang, N. *et al.* Novel TRRAP mutation causes autosomal dominant non-syndromic hearing loss. *Clinical Genetics* **96**, 300-308 (2019).
107. Hosoya, M., Fujioka, M., Ogawa, K. & Okano, H. Distinct expression patterns of causative genes responsible for hereditary progressive hearing loss in non-human primate cochlea. *Scientific Reports* **6**, 22250 (2016).
108. Kruczek, K. & Swaroop, A. Pluripotent stem cell-derived retinal organoids for disease modeling and development of therapies. *Stem Cells* **38**, 1206-1215 (2020).
109. Tang, P.C., Hashino, E. & Nelson, R.F. Progress in modeling and targeting inner ear disorders with pluripotent stem cells. *Stem Cell Reports* **14**, 996-1008 (2020).
110. van der Valk, W.H., Steinhart, M.R., Zhang, J. & Koehler, K.R. Building inner ears: recent advances and future challenges for in vitro organoid systems. *Cell Death & Differentiation* **28**, 24-34 (2021).
111. Parfitt, D.A., Lane, A., Ramsden, C.M., Carr, A.-J.F., Munro, P.M., Jovanovic, K. *et al.* Identification and correction of mechanisms underlying inherited blindness in human iPSC-derived optic cups. *Cell Stem Cell* **18**, 769-781 (2016).
112. Dulla, K., Aguila, M., Lane, A., Jovanovic, K., Parfitt, D.A., Schulkens, I. *et al.* Splice-modulating oligonucleotide QR-110 restores CEP290 mRNA and Function In Human c.2991+1655A>G LCA10 models. *Molecular Therapy - Nucleic Acids* **12**, 730-740 (2018).
113. Cui, Z., Guo, Y., Zhou, Y., Mao, S., Yan, X., Zeng, Y. *et al.* Transcriptomic analysis of the developmental similarities and differences between the native retina and retinal organoids. *Investigative Ophthalmology & Visual Science* **61**, 6-6 (2020).
114. Achberger, K., Probst, C., Haderspeck, J., Bolz, S., Rogal, J., Chuchuy, J. *et al.* Merging organoid and organ-on-a-chip technology to generate complex multi-layer tissue models in a human retina-on-a-chip platform. *Elife* **8**, e46188 (2019).

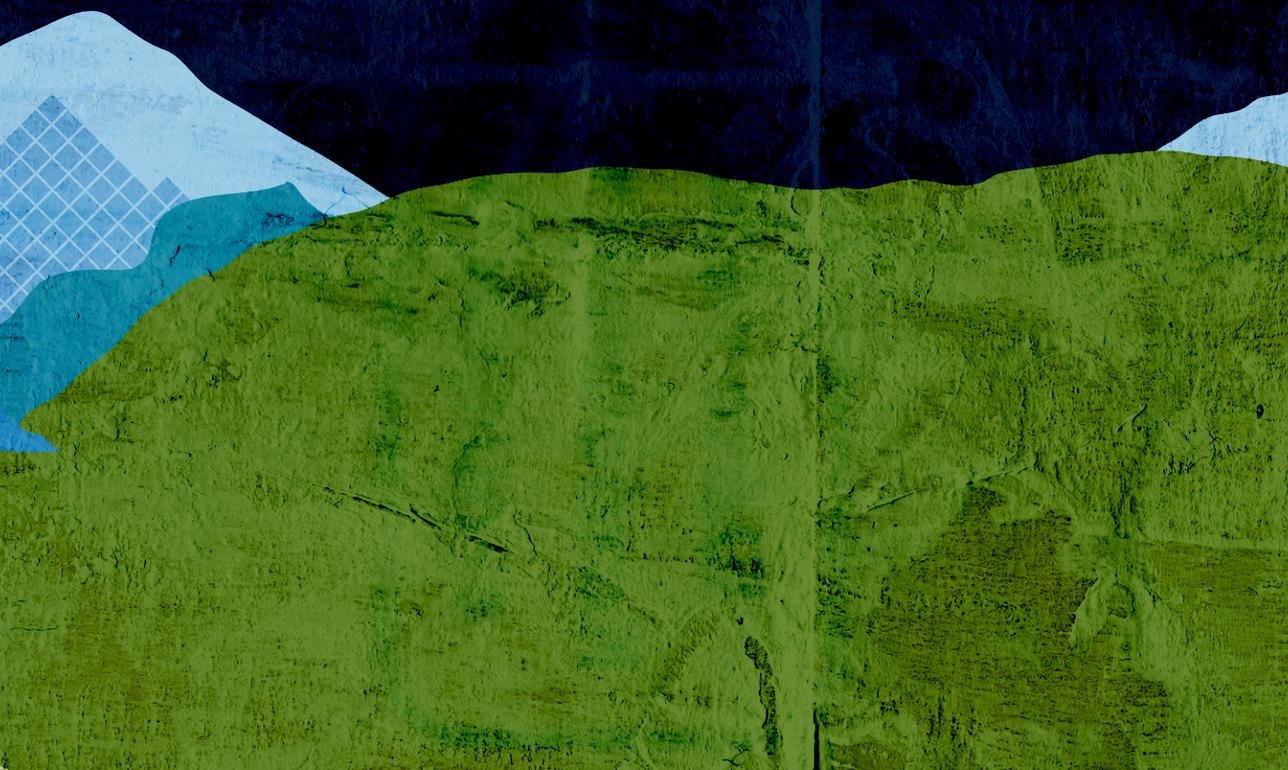
115. Hosoya, M., Fujioka, M., Sone, T., Okamoto, S., Akamatsu, W., Ukai, H. *et al.* Cochlear cell modeling using disease-specific iPSCs unveils a degenerative phenotype and suggests treatments for congenital progressive hearing loss. *Cell Reports* **18**, 68-81 (2017).
116. Dulla, K., Slijkerman, R., van Diepen, H.C., Albert, S., Dona, M., Beumer, W. *et al.* Antisense oligonucleotide-based treatment of retinitis pigmentosa caused by USH2A exon 13 mutations. *Molecular Therapy In press* (2021).
117. Boucher, S., Tai, F.W.J., Delmaghani, S., Lelli, A., Singh-Estivalet, A., Dupont, T. *et al.* Ultrarare heterozygous pathogenic variants of genes causing dominant forms of early-onset deafness underlie severe presbycusis. *Proceedings of the National Academy of Sciences* **117**, 31278-31289 (2020).
118. Runhart, E.H., Khan, M., Cornelis, S.S., Roosing, S., Del Pozo-Valero, M., Lamey, T.M. *et al.* Association of sex with frequent and mild ABCA4 alleles in Stargardt disease. *JAMA Ophthalmology* **138**, 1035-1042 (2020).
119. Runhart, E.H., Sangermano, R., Cornelis, S.S., Verheij, J., Plomp, A.S., Boon, C.J.F. *et al.* The common ABCA4 variant p.Asn1868Ile shows nonpenetrance and variable expression of Stargardt disease when present in trans with severe variants. *Investigative Ophthalmology & Visual Science* **59**, 3220-3231 (2018).
120. de Kok, Y.J., Cremers, C.W., Ropers, H.H. & Cremers, F.P. The molecular basis of X-linked deafness type 3 (DFN3) in two sporadic cases: identification of a somatic mosaicism for a POU3F4 missense mutation. *Human Mutation* **10**, 207-211 (1997).
121. Kokotas, H., Petersen, M.B. & Willems, P.J. Mitochondrial deafness. *Clinical Genetics* **71**, 379-391 (2007).
122. Zheng, Q.Y., Yan, D., Ouyang, X.M., Du, L.L., Yu, H., Chang, B. *et al.* Digenic inheritance of deafness caused by mutations in genes encoding cadherin 23 and protocadherin 15 in mice and humans. *Human Molecular Genetics* **14**, 103-111 (2005).
123. Khateb, S., Zelinger, L., Mizrahi-Meissonnier, L., Ayuso, C., Koenekoop, R.K., Laxer, U. *et al.* A homozygous nonsense CEP250 mutation combined with a heterozygous nonsense C2orf71 mutation is associated with atypical Usher syndrome. *Journal of Medical Genetics* **51**, 460-469 (2014).
124. Le Quesne Stabej, P., Saihan, Z., Rangesh, N., Steele-Stallard, H.B., Ambrose, J., Coffey, A. *et al.* Comprehensive sequence analysis of nine Usher syndrome genes in the UK National Collaborative Usher Study. *Journal of Medical Genetics* **49**, 27-36 (2012).
125. Boycott, K.M. & Ardigó, D. Addressing challenges in the diagnosis and treatment of rare genetic diseases. *Nature Reviews Drug Discovery* **17**, 151-152 (2018).
126. Radboudumc Genome Diagnostics. Turnaround times, materials and prices (www.radboudumc.nl).
127. Brandt, T., Sack, L.M., Arjona, D., Tan, D., Mei, H., Cui, H. *et al.* Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number variants. *Genetics in Medicine* **22**, 336-344 (2020).

128. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J. *et al.* 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. *Genetics in Medicine* **17**, 405-424 (2015).
129. Lentz, J.J., Jodelka, F.M., Hinrich, A.J., McCaffrey, K.E., Farris, H.E., Spalitta, M.J. *et al.* Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. *Nature Medicine* **19**, 345-350 (2013).
130. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J. *et al.* ClinGen--the Clinical Genome Resource. *New England Journal of Medicine* **372**, 2235-2242 (2015).
131. de Macena Sobreira, N.L. & Hamosh, A. Next-generation sequencing and the evolution of data sharing. *American Journal of Human Genetics* **Online ahead of print** (2021).



Chapter 7

Summary



ENGLISH SUMMARY

The inherited sensory disorders retinal dystrophies (RD) and hearing loss (HL) display an enormous degree of clinical and genetic heterogeneity. The clinical heterogeneity is characterized by the variability in age of onset, disease symptoms and disease progression, whereas the genetic heterogeneity is demonstrated by the large number of disease-associated genes (i.e. >270 RD-associated genes, >150 HL-associated genes). Despite this large number of genes, still many affected individuals lack a genetic diagnosis as no pathogenic variant (for autosomal dominant diseases), or not both pathogenic variants (for autosomal recessive diseases), could be identified using existing methodologies. A genetic diagnosis is essential for prognosis, counseling and eligibility for novel (genetic) therapies, and therefore it is of utmost importance to accommodate a genetic explanation for all affected individuals. With this in mind, the aim of this thesis was to shed light on the missing heritability involved in both inherited RD and HL. To increase the genetic diagnostic yield for these disorders, a combination of various techniques and approaches was applied.

Chapter 1 provided a general background of inherited sensory disorders and genetic diagnostics. **Chapter 1.1** addressed the anatomy of the eye and the inner ear, and the physiological processes involved that play a role in both health and disease. Special attention was paid to the different (genetic) causes that can underlie visual or hearing disorders. **Chapter 1.2** reviewed the available approaches for disease gene identification, and the subsequent improvements and refinements that have been introduced in the past decades. The advantages and limitations of different (sequencing) techniques and approaches were discussed, including their impact on genetic diagnostics. In the final part of this chapter, the importance and complexity of variant interpretation processes were discussed.

Chapter 2 described the identification of pathogenic *KIAA1549* variants using whole exome sequencing (WES). *KIAA1549* was previously suggested to be a candidate gene associated with autosomal recessive retinitis pigmentosa (RP), however additional evidence was lacking. In this chapter, novel pathogenic *KIAA1549* variants were described which strengthened the evidence for association of *KIAA1549* with RP. RNA analyses confirmed expression of *KIAA1549* in the human retina, and a retina-specific transcript was identified. Additionally, immunohistochemistry performed on murine retina sections revealed the presence of *KIAA1549* in the photoreceptor connecting cilium and the outer plexiform layer. Altogether, there now is very strong evidence that *KIAA1549* variants can underlie RP.

In **chapter 3**, the genetic defect underlying HL type DFNA21 was investigated. DFNA21 is a dominantly inherited type of adult-onset HL. Genetic diagnostics of adult-onset HL is severely complicated by the large involvement of environmental factors affecting hearing. These factors mask inheritance patterns in families with inherited adult-onset HL, thereby impeding the identification of potential monogenic defects. DFNA21 clearly illustrates this complexity. The implicated genetic defect of DFNA21 was previously localized to the so-called DFNA21 locus on chromosome 6. In the past, no putatively pathogenic variants could be identified within this locus. In **chapter 3.1**, WES was performed to screen all coding regions for putatively pathogenic variants and an in-frame deletion in the *RIPOR2* gene, located 0.9 Mb centromeric of the DFNA21 locus, was identified. Analyses revealed that the locus was wrongly delimited in the past, due to the presence of phenocopies and genocopies within the DFNA21-affected family. The newly identified *RIPOR2* variant was detected in 12 Dutch families, all diagnosed with dominantly-inherited and predominantly adult-onset HL. *Ex vivo* protein expression studies in murine cochlear hair cells confirmed a functional effect of the *RIPOR2* variant. An aberrant localization of mutant RIPOR2 could be observed in the stereocilia of these hair cells, and unlike wildtype RIPOR2, the mutant protein was unable to rescue the morphological defects in RIPOR2-deficient hair cells. Based on allele frequency datasets, the *RIPOR2* in-frame deletion is presumably the most frequent cause of adult-onset HL in Northwest Europe. More than 30,000 individuals worldwide are estimated to be at risk to develop HL due to this variant. For this reason, the opportunity to design a potential genetic therapy was explored in **chapter 3.2**. Based on the expected non-haploinsufficiency pathogenic mechanism of the variant, allele-specific RNaseH1-dependent antisense oligonucleotides (AONs) were designed and tested for their potency to suppress mutant RIPOR2 expression. A lead AON molecule was identified that is able to efficiently and specifically reduce mutant transcript levels as observed in patient-derived fibroblasts and HEK293T cells. Based on these results, future steps including the validation of the lead AON in animal model studies can be initiated.

Despite the success stories of WES as exemplified by the results described in **chapters 2** and **chapter 3**, an important limitation of WES remains the inability to detect structural variants (SVs). In **chapter 4**, whole genome sequencing (WGS) was performed after WES failed to identify the genetic defect associated with RP type 17 (RP17). In 22 unrelated families with autosomal dominant RP, eight distinct SVs (i.e. tandem duplications, duplication-inversion and triplication events) were identified. All SVs overlap with the RP17 locus on chromosome 17. None of the genes implicated in the SVs were previously associated with RP and therefore the genetic mechanism involved was not readily understood. To unravel this genetic disease mechanism, 3D chromosome mapping was performed in retinal organoids. By comparing the 3D maps of wildtype and RP17-

patient-derived cells, disruption of the chromosome organization could be observed. Rewiring of enhancer-promoter contacts were shown to cause the ectopic expression of *GDPD1*, a gene that is normally not expressed in the retina. Ectopic expression of this gene was confirmed by RNA analyses in both patient-derived photoreceptor progenitor cells and retinal organoids. From these results it was concluded that the reorganization of the chromosome landscape and consequently ectopic expression of *GDPD1* most likely was the true cause of disease. This work did not only explain the genetic defect involved in RP17, but it also highlighted a novel mechanism of disease that involves ectopic gene expression.

Not all affected individuals can be genetically explained by the implementation of WES or WGS. A significant number of patients with monoallelic *SLC26A4* variants, and affected with HL and a uni- or bilateral enlarged vestibular aqueduct, remain unsolved after screening of all coding regions of the gene. In **chapter 5**, 28 individuals with one or no pathogenic *SLC26A4* variants were subjected to extensive genetic analyses that included short- and long read WGS and optical genome mapping. Important insights that could (partially) explain the missing heritability were obtained: a significant enrichment of a 0.89-Mb haplotype (previously coined as the CEVA-haplotype) in monoallelic cases, potential digenic inheritance for *SLC26A4* and *FOXI1* variants, and two novel splice variants. Unfortunately, the true genetic defect present on the CEVA-haplotype could not be pinpointed despite the fact that it is highly unlikely that a genomic variant was missed by the combination of the applied tools. This underlines that our understanding of the human genome and genomic variation is still incomplete as we currently fail to recognize the causative variant.

In line with the conclusions drawn from **chapters 2 to 5**, **chapter 6** elaborated on the developments that can be expected in the next years and that will help to gain a complete understanding of all genomic variation. Strategies and methods that should be explored include multi-omics approaches, single-cell RNA and epigenetic technologies. Also the potential non-Mendelian inheritance patterns should be addressed. With the advent of genetic therapies, complete knowledge of the genomic landscape of sensory disorders is even more pressing. The work described in this thesis indicated that by implementing and combining novel and existing technologies, and integrating information from the (epi)genome, RNA and protein level, a significant portion of the missing heritability can potentially be resolved. Hopefully, by applying all the knowledge that was collected during decades of research, every individual affected with inherited RD or HL can obtain a genetic diagnosis in the near future.

NEDERLANDSE SAMENVATTING

De erfelijke sensorische aandoeningen retinale dystrofie en slechthorendheid worden gekenmerkt door een grote mate van klinische en genetische heterogeniteit. De klinische heterogeniteit wordt weerspiegeld in onder meer de variatie in progressie van symptomen en de leeftijd waarop de eerste symptomen worden gerapporteerd. De genetische heterogeniteit wordt weerspiegeld in het grote aantal genen dat betrokken kan zijn bij het ontwikkelen van deze aandoeningen. Er zijn meer dan 270 genen bekend die geassocieerd zijn met de ontwikkeling van retinale dystrofie en meer dan 150 genen met de ontwikkeling van slechthorendheid. Ondanks dit grote aantal genen is bij veel patiënten het onderliggend genetisch defect nog onbekend. In deze gevallen kan er geen pathogene variant (dominante overerving) of kunnen niet beide pathogene varianten (recessieve overerving) geïdentificeerd worden met behulp van de bestaande methodieken. Een genetische diagnose is van groot belang voor het welzijn van deze mensen en is essentieel voor het bepalen van prognose, het afstemmen van counseling en in sommige gevallen biedt de diagnose ook mogelijkheden voor genetische therapie of andere behandelingen. Er wordt daarom veel waarde gehecht aan genetische diagnostiek, en het is een belangrijk speerpunt van (klinisch) moleculair genetici om voor alle individuen met een genetische aandoening een verklaring te kunnen vinden. Daarom was het doel van het onderzoek beschreven in dit proefschrift om op zoek te gaan naar de ontbrekende genetische verklaringen voor erfelijke slechthorendheid en retinale dystrofie. Om dit doel te verwezenlijken werd er gebruik gemaakt van een combinatie van verschillende technieken en strategieën.

Hoofdstuk 1 geeft algemene achtergrondinformatie met betrekking tot erfelijke sensorische aandoeningen en genetische diagnostiek. In **hoofdstuk 1.1** wordt allereerst de anatomie van het oog en het binnenoer besproken, alvorens ook de verschillende betrokken (patho)fysiologische processen worden toegelicht. Er wordt speciale aandacht besteed aan de verschillende (genetische) oorzaken van netvliesdegeneratie of slechthorendheid. **Hoofdstuk 1.2** gaat in op de beschikbare methoden voor identificatie van ziektegenen en de verbeteringen en verfijningen van die methoden in de afgelopen decennia. De voordelen en beperkingen van de verschillende benaderingen worden uiteengezet, inclusief de impact die ze hebben gehad op de genetische diagnostiek. In het laatste deel van dit hoofdstuk worden de verschillende aspecten en de complexiteit van het beoordelen van varianten in het genoom besproken.

Hoofdstuk 2 beschrijft de identificatie van pathogene *KIAA1549*-varianten met behulp van exoomsequencing (WES). Al eerder werd het *KIAA1549*-gen gesuggereerd als

een kandidaatgen voor recessief overervende retinitis pigmentosa (RP), maar tot dusverre ontbrak het aan voldoende bewijs. De nieuwe pathogene *KIAA1549*-varianten beschreven in dit hoofdstuk leverden aanvullend bewijs voor de associatie van *KIAA1549* met RP. Tevens werd met behulp van RNA-analyses de expressie van *KIAA1549* in de humane retina bevestigd en werd er een retina-specifiek transcript geïdentificeerd. Daarnaast werd ook immunohistochemie uitgevoerd op retina coupes van de muis, waarmee aanwezigheid van het *KIAA1549*-eiwit werd aangetoond in het "connecting cilium" van de fotoreceptoren en in de buitenste plexiforme laag. Samenvattend kan gesteld worden dat bevestigd is dat *KIAA1549* varianten ten grondslag kunnen liggen aan RP.

In **hoofdstuk 3** werd het genetische defect onderzocht dat de slechthorendheid type DFNA21 veroorzaakt. DFNA21 is een vorm van dominant overervende slechthorendheid die met name op volwassen leeftijd ontstaat. Genetische diagnostiek van slechthorendheid bij volwassenen is zeer complex vanwege de grotere invloed van omgevingsfactoren. Deze factoren kunnen het overervingspatroon maskeren in families met deze vorm van slechthorendheid, waardoor de identificatie van mogelijke genetische oorzaken wordt belemmerd. DFNA21 is hiervan een duidelijk voorbeeld. De locatie van het onderliggend genetisch defect van DFNA21, de zogenaamde DFNA21-locus, was al eerder bepaald, en is gelegen op chromosoom 6. In het verleden konden geen mogelijk pathogene varianten worden geïdentificeerd binnen deze locus. In **hoofdstuk 3.1** werd WES uitgevoerd om alle coderende regio's van het genoom te screenen op mogelijk pathogene varianten. Een in-frame deletie in het *RIPOR2*-gen, gelegen op 0,9 Mb afstand van de DFNA21-locus, werd geïdentificeerd. Analyses lieten zien dat de locus in het verleden niet correct was bepaald vanwege de aanwezigheid van verschillende fenokopieën en genokopieën binnen de DFNA21-familie. De geïdentificeerde *RIPOR2*-variant werd vervolgens nog gevonden in 11 andere Nederlandse families waarvan verschillende familieleden op volwassen leeftijd gediagnosticeerd zijn met een dominant overervende vorm van slechthorendheid. Analyse van de expressie van het *RIPOR2*-eiwit in haarcellen uit de cochlea van muizen bevestigden een functioneel effect van de *RIPOR2*-variant. Een afwijkende lokalisatie van mutant *RIPOR2*-eiwit werd waargenomen in de stereocilia van deze haarcellen. Daarnaast konden morfologische defecten in *RIPOR2*-deficiente haarcellen niet worden hersteld met behulp van het mutant-eiwit, maar wel met wildtype-eiwit. Op basis van allelfrequenties in verschillende populaties wordt geschat dat de in-frame deletie in *RIPOR2* de meest voorkomende oorzaak van erfelijke slechthorendheid bij volwassenen is binnen Noordwest-Europa. Wereldwijd lopen mogelijk meer dan 30.000 personen het risico om slechthorend te worden als gevolg van deze variant. Dit was de aanleiding om de mogelijkheden te onderzoeken voor het ontwikkelen van een genetische

therapie voor DFNA21, zoals beschreven is in **hoofdstuk 3.2**. Gebaseerd op het zeer waarschijnlijke niet-haploïnsufficiëntie ziektemechanisme van de variant, werden allel-specifieke RNase H1-afhankelijke antisense oligonucleotiden (AONs) ontworpen en getest op hun vermogen om de expressie van mutant-RIPOR2 te onderdrukken. Er werd een kandidaat-AON geïdentificeerd die in staat is om efficiënt en specifiek de niveaus van het mutant-transcript te verlagen. Dit werd bepaald in zowel cellen gekweekt uit huidbiopten van DFNA21-patiënten als in het HEK293T celmodel. Op basis van deze resultaten kunnen de volgende stappen in het ontwikkelen van een therapie voor DFNA21 worden geïnitieerd, waaronder de validatie van de kandidaat-AON in een diermodel.

Ondanks de successen die behaald zijn met WES, zoals blijkt uit de resultaten beschreven in **hoofdstuk 2 en 3**, blijft een van de belangrijke nadelen van WES het beperkte vermogen om structurele varianten te detecteren. In **hoofdstuk 4** werd genomsequencing (WGS) uitgevoerd nadat het niet gelukt was om met behulp van WES het onderliggend genetisch defect voor RP type 17 (RP17) te identificeren. In 22 niet-verwante families met een dominante vorm van RP werden 8 verschillende structurele veranderingen (o.a. tandemduplicaties, duplicatie-inversies en een triplicatie) gevonden. Al deze structurele varianten zijn (ten dele) gelegen in de RP17-locus op chromosoom 17. Geen van de genen die betrokken zijn bij de structurele varianten waren al eerder geassocieerd met RP. Om deze reden was het onderliggende pathogene mechanisme niet meteen duidelijk. Om het ziektemechanisme verder te ontrafelen werd de 3D-organisatie van de chromosomen onderzocht in netvliesorganoids. Door het 3D-landschap in wildtype cellen en cellen afkomstig van RP17-patiënten te vergelijken, kon een verstoring in de chromosoomorganisatie worden waargenomen. Deze verstoring heeft tot gevolg dat nieuwe contacten gevormd kunnen worden tussen enhancer- en promoter-elementen, wat uiteindelijk leidt tot een ectopische en mogelijk toxische expressie van het *GDPD1*-gen. Dit gen komt normaalgesproken niet tot expressie in de cellen van het netvlies. De verhoogde expressie van dit gen werd bevestigd met behulp van RNA-analyses van gedifferentieerde stamcellen (fotoreceptor voorlopercellen en netvliesorganoids) die gekweekt werden uit huidbiopten van patiënten. Op basis van deze resultaten werd geconcludeerd dat de reorganisatie van de betreffende chromosomale regio en de daaropvolgende verhoogde expressie van *GDPD1* hoogstwaarschijnlijk de oorzaak is van RP17. Het werk in dit hoofdstuk beschrijft zowel het genetisch defect van deze netvliesandoening als een nieuw ziektemechanisme: de verhoogde, toxische expressie van een gen.

Niet alle patiënten kunnen genetisch worden verklaard na toepassing van WES of WGS. Een groot aantal patiënten met gehoorverlies, een uni- of bilateraal vergroot

vestibulair aquaduct en een heterozygote *SLC26A4*-variant blijft genetisch onopgelost na volledige analyse van alle coderende regio's van het *SLC26A4*-gen. In **hoofdstuk 5** werden 28 individuen met één of zonder een pathogene *SLC26A4*-variant onderworpen aan uitgebreide genetische analyses bestaande uit o.a. short-read en long-read WGS en "optische genoom mapping". Er werden belangrijke inzichten verkregen die de ontbrekende erfelijke factoren binnen deze patiëntengroep (gedeeltelijk) kunnen verklaren: een significante verrijking van een haplotype van 0,89 Mb (eerder beschreven als het CEVA-haplotype) in patiënten met één pathogene *SLC26A4* variant, aanwijzingen voor digene overerving van *SLC26A4*- en *FOXI1*-varianten, en twee nieuwe varianten met een effect op splicing van het transcript. Helaas kon het genetische defect dat gelegen is op het CEVA-haplotype niet worden vastgesteld, hoewel het niet waarschijnlijk is dat een genetische variant is gemist met het gebruik van de diverse methoden voor sequentie-analyse en DNA-imaging. Het feit dat we de oorzakelijke variant nog niet konden herkennen, bevestigt dat onze kennis van het menselijke genoom en genetische variatie nog steeds onvolledig is.

Voortbordurend op de conclusies die zijn getrokken in de **hoofdstukken 2 tot en met 5**, gaat **hoofdstuk 6** dieper in op de te verwachten ontwikkelingen in de komende jaren en hoe die zullen bijdragen aan het interpreteren van genetische variatie. Veelbelovend zijn onder meer "multi-omics" benaderingen en "single-cell" technologieën. Daarnaast moet meer aandacht worden besteed aan het herkennen van niet-Mendeliaanse overervingspatronen. Met de snelle ontwikkelingen op het gebied van genetische therapie is volledig inzicht in het genomische landschap van erfelijke sensorische aandoeningen nog urgenter. Het onderzoek beschreven in dit proefschrift, laat zien dat door het implementeren en combineren van nieuwe en bestaande technologieën, en door het verzamelen van informatie afkomstig uit het (epi)genoom, transcriptoom en proteoom, een aanzienlijk deel van de ontbrekende erfelijke oorzaken kan worden geïdentificeerd. Door alle kennis toe te passen die decennialang met behulp van onderzoek verzameld is, kunnen hopelijk alle mensen met een erfelijke vorm van retinale dystrofie of slechthorendheid in de toekomst een genetische diagnose krijgen.



Chapter 8

Appendix

Data management plan

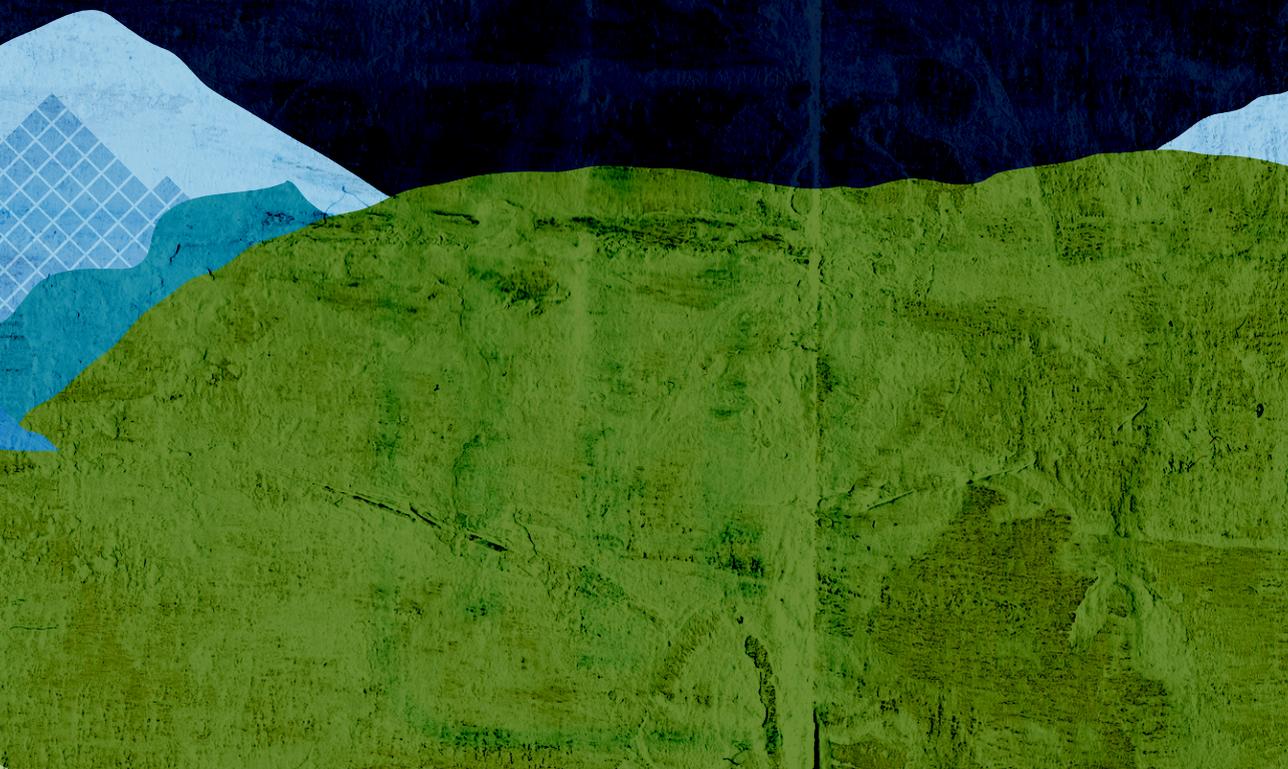
List of publications

PhD portfolio

About the author

Acknowledgments

Abbreviations



DATA MANAGEMENT PLAN

Type of data	Subject to privacy (Yes/No)	Way of anonymization	Storage
DNA samples	Yes	A DNA number was assigned to each individual by the cell culture facility of the Department of Human Genetics.	All DNA samples were stored at the cell culture facility of the Department of Human Genetics. Contact person for the DNA samples is Saskia van der Velde-Visser, Saskia.vanderVelde-Visser@radboudumc.nl
Cell lines	Yes	A cell line number was assigned to each cell line by the cell culture facility of the Department of Human Genetics.	Cell lines were frozen in liquid nitrogen and stored in an assigned -80°C freezer at the Department of Human Genetics. After registration at the cell culture facility each cell line received a unique cell line number. Contact person for the cell lines is Saskia van der Velde-Visser, Saskia.vanderVelde-Visser@radboudumc.nl
Sequencing data	Yes	All patients received an untraceable number, the identity of the patient is only known by the research PI and treating physician	All raw and analyzed sequencing data were stored at the private network of the Blindness Genetics and Deafness Genetics workgroups <i>Blindness Genetics</i> : T:\PIgroup-Frans-Cremers-and-Susanne-Roosing\NAS\03 NGSdata <i>Deafness Genetics</i> : T:\PIgroup-Hannie-Kremer\Deafness genetics
Clinical data of the patients	Yes	A study-ID was assigned to all patients	Clinical data of patients described in the chapters of these thesis can be found at the private network of the Blindness Genetics or Deafness Genetics workgroups <i>Blindness Genetics</i> : H:\GR Theme groups\05 PI Group Frans Cremers\06 Manuscripts <i>Deafness Genetics</i> : H:\GR Theme groups\07 PI Group Hannie Kremer\06 Manuscripts
Experimental procedures	No	Not applicable	All experimental procedures and data are described and can be accessed in an electronic lab journal (https://radboudumc.labguru.com)
Files for the publications presented in this thesis	No	Not applicable	All files that are part of these thesis can be found at the private network of the Blindness Genetics or Deafness Genetics workgroups <i>Blindness Genetics</i> : H:\GR Theme groups\05 PI Group Frans Cremers\06 Manuscripts <i>Deafness Genetics</i> : H:\GR Theme groups\07 PI Group Hannie Kremer\06 Manuscripts

LIST OF PUBLICATIONS

Wesdorp M*, Murillo-Cuesta S*, Peters T, Celaya AM, Oonk A, Schraders M, Oostrik J, Gomez-Rosas E, Beynon AJ, Hartel BP, Okkersen K, Koenen HJPM, Weeda J, Lelieveld S, Voermans NC, Joosten I, Hoyng CB, Lichtner P, Kunst HPM, Feenstra I, **de Bruijn SE**, DOOFNL Consortium, Admiraal RJC, Yntema HG, van Wijk E, Del Castillo I, Serra P, Varela-Nieto I#, Pennings RJE#, Kremer H#. MPZL2, encoding the epithelial junctional protein myelin protein zero-like 2, is essential for hearing in man and mouse. *American Journal of Human Genetics* **108**, 74-88 (2018).

de Bruijn SE, Verbakel SK, de Vrieze E, Kremer H, Cremers FPM, Hoyng CB, van den Born LI, Roosing S. Homozygous variants in *KIAA1549*, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa. *Journal of Medical Genetics* **55**, 705-712 (2018).

de Bruijn SE*, Fiorentino A*, Ottaviani D, Fanucchi S, Mulders T, Georgiou M, Rivolta C, Pontikos N, Arno G, Roberts L, Greenberg J, Albert S, Gilissen C, Aben M, Rebello G, Mead S, Raymond FJ, Corominas J, Smith CEL, Kremer H, Downes S, Black GC, Webster AR, Inglehearn CF, van den Born LI, Koenekoop RK, Michaelides M, Ramesar RS, Hoyng CB, Mhlanga MM, Cremers FPM, Cheetham ME, Roosing S#, Hardcastle AJ#. Structural variants create new topological associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa. *American Journal of Human Genetics* **107**, 802-814 (2020).

de Bruijn SE*, Smits JJ*, Liu C, Lanting CP, Beynon AJ, Blankevoort J, Oostrik J, Koole W, de Vrieze E, Cremers CWRJ, Cremers FPM, Roosing S, Yntema HG, Kunst HPM, Zhao B, Pennings RJE#, Kremer H#. A RIPOR2 in-frame deletion is a frequent and highly penetrant cause of adult-onset hearing loss. *Journal of Medical Genetics* **58**, 96-104 (2021).

de Bruijn SE*, Fadaie Z*, Cremers FPM, Kremer H, Roosing S. The impact of modern technologies on molecular diagnostic success rates, with a focus on inherited retinal dystrophy and hearing loss. *International Journal of Molecular Sciences* **22**, 2943-2969 (2021).

Smits JJ*, **de Bruijn SE***, Lanting CP, Oostrik J, O'Gorman L, Mantere T, DOOFNL Consortium, Cremers FPM, Roosing S, Yntema HG, de Vrieze E, Derks R, Hoischen A, Pegge SAH, Neveling K, Pennings RJE, Kremer H. Exploring the missing heritability in subjects with hearing loss, enlarged vestibular aqueducts, and a single or no pathogenic *SLC26A4* variant. *Human Genetics*, in press.

de Vrieze*, **de Bruijn SE***, Reurink J, Broekman S, van de Riet V, Aben M, Kremer H, van Wijk E. Efficient generation of knock-in zebrafish models for inherited disorders using CRISPR-Cas9 ribonucleoprotein complexes. *International Journal of Molecular Sciences* **22**, 9429-9445 (2021).

PHD PORTFOLIO

Name PhD student:	<i>S.E. de Bruijn</i>	PhD period:	1/3/2017-31/5/2021
Department:	<i>Human Genetics</i>	Promoters:	<i>Prof. dr. J.M.J. Kremer, Prof. dr. F.P.M. Cremers</i>
Graduate school:	<i>Donders Institute for Brain, Cognition and Behavior</i>	Co-promoters:	<i>Dr. S. Roosing Dr. E. de Vrieze</i>

COURSES AND WORKSHOPS	Year(s)
Graduate school introduction day	2017
How to write a medical scientific paper	2017
Eye genetics course (Bertinoro, IT)	2017
Management voor promovendi	2017
Introduction to R	2017
Scientific integrity course	2018
Graduate school day	2018, 2019
Education in a nutshell	2019
Scientific writing	2019
Genetics R course	2019
Genetics Linux course	2019
Introduction to laboratory animal science	2019
Workshop UCSC genome browser and the single cell browser (by Robert Kuhn)	2020
Grant writing and presenting for funding committees	2021

SEMINARS AND LECTURES	Year(s)
Sensory disease meeting, Human genetics department	2017-2021
Theme discussion, Human genetics department	2017-2021
Radboud Research Round Sensory Disorders, Radboudumc	2017-2021
"Meet the expert" meeting, RIMLS	2017-2021
International guest lectures, Radboudumc	2017-2021

(INTER)NATIONAL SYMPOSIA & CONGRESSES	Year(s)
Belgian society for stem cell research, 4 th annual meeting (Liège, BE) - <i>participant</i>	2017
ERDC meeting (Lausanne, CH) - <i>participant</i>	2018
Molecular Biology of Hearing and Deafness (Göttingen, DE) - <i>participant</i>	2018
ARVO annual meeting (Vancouver, CA) - <i>poster presentation</i>	2019
ARO midwinter meeting (San Jose, USA) - <i>oral presentation</i>	2020
ESHG conference (virtual) - <i>participant</i>	2020
ASHG conference (virtual) - <i>poster presentation</i>	2020
Retina 2020 meeting, Fighting Blindness Ireland (virtual) - <i>oral presentation</i>	2020
AVRO annual meeting (virtual) - <i>invited speaker</i>	2021

TEACHING	Year(s)
Instructor NGS workshop (Eye Genetics course, Bertinoro, Italy)	2017
Instructor Translational Genomics computer practical (master course Medical Biology)	2017, 2018, 2020
Supervisor Research Project Proposal (bachelor course Biomedical Sciences)	2018
Supervisor Genetic Lab Practice (bachelor course Biomedical Sciences)	2018, 2019
Supervisor Meet the PhD (bachelor course Biomedical Sciences)	2019, 2020
Supervisor Meet the Expert (bachelor course Medical Sciences)	2020

SUPERVISION	Year(s)
Supervision master thesis Molecular and Cellular Life Sciences (Joëlle Blankevoort)	2017-2018
Supervision master student Medical Biology (Vince van de Riet)	2018-2019
Supervision master student Medical Biology (Jill van Wolferen)	2019-2020

PRIZES AND GRANTS	Year(s)
Travel grant Simonsfonds	2017, 2019, 2020
Travel grant Association for Research in Otolaryngology (ARO)	2020
Don Henderson award for best travel grant application, ARO annual meeting	2020
Finalist early investigator award, Retina meeting Fighting Blindness Ireland	2020
Sensory disease talent award (scientific quality and social impact), Radboudumc	2021

ABOUT THE AUTHOR

Suzanne de Bruijn was born on the 25th of May in 1993 in Angeren, the Netherlands. After completing her secondary education at the Over Betuwe College in Bommel, she started her studies in Molecular Life Sciences at the Radboud University Nijmegen in 2011. Suzanne completed her Bachelor studies with an internship at the Department of Molecular Biology at the Radboud Institute of Molecular Life Sciences (RIMLS) in 2014. Afterwards, she decided to continue her studies and enrolled in the master program of Molecular Life Sciences at the Radboud



University, with a specialization in functional genomics. She performed a first master internship at the Department of Experimental Urology (RIMLS) where she investigated the role of the androgen receptor in the development of bladder cancer. During her studies, Suzanne developed a strong interest in Human Genetics, and therefore decided to visit the Vandenberghe lab for Gene Therapy (dr. Luk Vandenberghe) at Harvard Medical School (Boston, USA) for her second research internship. During this time, she investigated the development of a genetic therapy strategy for Usher syndrome type IIa. Suzanne returned to Nijmegen in September 2016, and graduated with the honor "*cum laude*" early 2017. In March 2017, Suzanne started as a PhD candidate at the Department of Human Genetics at the Radboudumc in Nijmegen. Under the supervision of prof. dr. Hannie Kremer, prof. dr. Frans Cremers, dr. Susanne Roosing and dr. Erik de Vrieze, she investigated the missing heritability in the sensory disorders hearing loss and retinal dystrophies. The work performed as part of this PhD project is described in this thesis and has led to several publications in international scientific journals. Suzanne received several prizes for her work, including the Don Henderson award at the annual Association for Research in Otolaryngology (ARO) meeting in 2020, and she was invited to speak at the Association for Research in Vision and Ophthalmology (ARVO) meeting in 2021. In June 2021, Suzanne started as a postdoctoral fellow at the Department of Human Genetics and will continue her research into the involvement of non-coding variants in the development of retinal disease.

ACKNOWLEDGEMENTS

Aan elk avontuur komt een einde, en zo ook aan deze PhD. Tijdens dit avontuur hebben vele waardevolle en onmisbare mensen mij bijgestaan, en zonder hun hulp had ik de eindstreep zeker niet gehaald. Daarom wil ik dit laatste hoofdstuk gebruiken om al deze mensen te bedanken, en een aantal daarvan in het bijzonder.

Allereerst wil ik mijn promotoren en copromotoren bedanken: Prof. Hannie Kremer, Prof. Frans Cremers, Dr. Susanne Roosing en Dr. Erik de Vrieze. Dank dat jullie 4,5 jaar geleden mij de kans hebben geboden om met dit project aan de slag te gaan en dat jullie altijd voor mij hebben klaargestaan.

Hannie, wat heb ik veel van jou geleerd de afgelopen jaren. Dankjewel voor alle uren die je hebt gespendeerd aan het reviseren van mijn hoofdstukken en manuscripten, de vele (urenlange) brainstormsessies en de interessante discussies die we hebben gehad over een van mijn projecten. Ik ben erg trots op de resultaten die we hebben behaald. Wie had dat gedacht toen we 4 jaar geleden de verschillende dossiers van DFNA21 gingen opgraven? Ik bewonder je professionaliteit en had het idee dat we vaak op een lijn bleken te zitten met onze ideeën. Onder andere om deze redenen heb ik onze samenwerking als zeer prettig ervaren. Ook buiten de wetenschap was je altijd in voor een goed gesprek of een gezellig praatje, en was je geïnteresseerd in hoe het nou met me gaat. Ook van ons tripje, toch nog op de valreep, naar San Jose heb ik erg genoten. Ook al ben ik qua werkzaamheden nu wel een andere richting in geslagen, ik hoop en ben ervan overtuigd dat we de komende jaren nog veel met elkaar te maken zullen hebben!

Susanne, wat een geslaagd team als Suus1 en Suus2 hebben wij weten te vormen. Je hebt me wegwijs gemaakt in de wereld van WES, WGS en verder. Wat een geweldige momenten waren het dat we samen letterlijk de puzzelstukjes voor het RP17 project in elkaar aan het leggen waren, gewapend met onze vilstiften en natuurlijk het koptelefoonsnoer! Ik bewonder jouw inzet om ons werk te vertalen naar de maatschappij, en hoe je ons zo dwingt om uit onze researchbubbel te treden. Tijdens de verschillende reizen voor congressen of bijeenkomsten die wij samen hebben gemaakt is gebleken dat we ook buiten de werkvloer om een goede match zijn. Wat een mooie roadtrip rond Vancouver Island was dat! Zelfs achter het stuur heb ik er nog van genoten ☺. Ik verheug me erop om samen weer de volgende genetische puzzel op te lossen en nog lang op zo'n prettige manier met je samen te werken. Hopelijk met vele mooie champagnemomentjes tot gevolg.

Erik, jij hebt je officieel pas later aangesloten bij mijn promotieteam, maar toch was je vanaf het begin al sterk betrokken bij mijn projecten. Je hebt me geholpen om het genetische werk naar een hoger niveau te tillen. Hier heb ik erg veel van geleerd. Zonder jouw hulp of input hadden de meeste functionele studies beschreven in dit proefschrift niet bestaan. Tijdens de laatste loodjes van mijn PhD hebben we samen als een goed geoliede machine nog grote experimenten over de finishlijn weten te trekken. Ondanks de lange uren heb ik erg genoten van deze intensieve, maar ook gezellige dagen. Ik verheug me alweer op het volgende ambitieuze project dat we ongetwijfeld weer zullen bedenken. Dankjewel voor al je onmisbare hulp!

Frans, jij hebt je meer op de achtergrond met mijn project beziggehouden, maar als ik je nodig had dan stond de deur altijd open. Wanneer een van de projecten vastliep had je altijd wel een goed advies, een nieuwe invalshoek of een interessant artikel dat we konden raadplegen. Ook van de onovertroffen manier waarop jij je verschillende projecten op de rails houdt heb ik ontzettend veel geleerd. Ik bewonder hoe je ondanks alle drukte toch altijd alles onder controle en stressvrij lijkt te kunnen houden. Ook heb je de groepsdynamiek hoog in het vaandel staan, iets wat ik erg in je waardeer. Je hebt me de kans geboden om mijn onderzoek-avontuur voort te zetten binnen de afdeling. Hier ben ik nog steeds ontzettend blij mee en dankbaar voor, want er is nog genoeg te leren en te ontdekken. Dank voor al je hulp gedurende de afgelopen jaren.

Daarnaast wil ik de leden van de manuscriptcommissie, **Prof. Dirk Lefeber**, **Prof. Bé Wieringa** en **Prof. Elfride De Baere**, en de overige leden van de corona bedanken voor hun tijd en aandacht die ze hebben besteed aan het lezen en beoordelen van mijn proefschrift.

Dan mijn paranimfen, **Jeroen** en **Janine**, bedankt dat jullie mij deze dag bij willen staan en dat eigenlijk al jarenlang hebben gedaan. **Jeroen**, letterlijk vanaf dag één zijn we samen dit avontuur aangegaan, en wat heb ik van onze samenwerking genoten. We vulden elkaar altijd perfect aan en zaten eigenlijk altijd op één lijn. Je hebt me de andere kant van het onderzoek laten zien: een inkijkje in de kliniek, een stoomcursus audiogrammen, maar ook het opgraven van oude dossiers in de krommen van het archief. Dankjewel dat je me hierin hebt betrokken, ik zie dit als een zeer waardevolle aanvulling op mijn PhD. Ook buiten het werk om hebben we veel mooie en gezellige momenten meegemaakt. Je nam altijd het initiatief voor de volgende sociale activiteit met vele borrels, spelletjesavonden en ook culinaire etentjes tot gevolg. Alleen al hierom zullen we je erg gaan missen. Ik wens je veel succes met je nieuwe baan in Utrecht en ik hoop dat we nog lang contact blijven houden.

Janine, eigenlijk vanaf het moment dat jij bij ons op de afdeling bent gekomen zijn we samen opgetrokken. We zaten dan ook in hetzelfde schuitje, bungelend tussen twee werkgroepen in, en ook zijn we (op een kort intermezzo na) altijd kantoorgenootjes geweest. Het zegt genoeg wanneer je na een lange werkdag nog niet uitgepraat bent, en besluit de gezelligheid voort te zetten tijdens een vrijmibo, etentje, of (sporadisch) een sportievere activiteit. Ik heb in jou een ware *partner in crime* gevonden wat betreft het organiseren van verschillende groepsuitjes, pubquizen of zelfs interactieve workshops. Wat een plezier hebben we hieraan beleefd! Het is fijn om iemand om je heen te hebben bij wie je altijd je ei kwijt kunt, maar ook je overwinningen kunt vieren.

I would also like to thank all my other (former) colleagues from the Deafness and Blindness genetics workgroups. **Claire-Marie, Daan, Duaa, Ellen, Erica, Erwin, Femke, Ferry, Hedwig, Imran, Jaap, Jessie, Ketan, Laura, Lisette, Marco, Margit, Margo, Marijke, Marta, Merel, Mieke, Mubeen, Ralph, Rebekkah, Renske, Riccardo, Sanne, Silvia, Stéphanie, Tabea, Theo, Tomasz, Victor, Zelia** and **Zeinab**: THANK YOU ALL. I feel blessed to be able to work with so many great and talented people every day.

Jaap en **Sanne**, jullie hebben me wegwijs op het lab gemaakt. **Jaap**, de koning van de PCR, dankjewel voor het meedenken over al mijn onmogelijke PCR-reacties maar ook voor het altijd bereid zijn om bij te springen of een paar extra samples (of kloneringen ☺) mee te nemen wanneer de druk voor mij te hoog werd. Ik heb veel van je geleerd en kon altijd bij je aankloppen. Ik hoop dat we de spelletjesavond-traditie nog lang in ere kunnen houden, want er staan nog genoeg bordspellen in de kast die nog uitgeprobeerd moeten worden! **Sanne**, dankjewel dat ik jouw schaduw mocht zijn tijdens de eerste weken op de afdeling, jij hebt me alle basisprincipes bijgebracht. Nog steeds kan ik altijd bij je terecht voor al mijn vragen, zelfs als het niks met zebravissen te maken heeft. Bovendien vind ik het veel te gezellig bij jullie op het lab, dus ik ben blij dat ik nog steeds welkom ben! **Theo**, ook van jou heb ik veel geleerd en je hebt me kennis laten maken met de wondere wereld van de immunohistochemie. Wat is het mooi om jou over je passie en met name de structuren van het binnenoor te horen praten! Dankjewel voor alle onmisbare hulp met alle immuno's door de jaren heen.

Ook mijn mede PhD-studenten (**Renske, Merel, Hedwig** en **Jessie**), dank jullie wel voor alle gezelligheid, een luisterend oor wanneer het nodig was, of de mogelijkheid om juist mooie resultaten te kunnen delen. Ik kon altijd bij jullie terecht! Ik heb erg genoten van de promovendi-etentjes en -borrels, laten we ook deze traditie erin houden.

Daarnaast wil ik ook alle andere collega's van Hearing&Genes bedanken. Het was leuk jullie te leren kennen tijdens de jaarlijkse uitjes en om een inkijkje te krijgen van de

andere kant van het onderzoek. Met name **Ronald**, dankjewel dat ik met jou heb mogen meekijken tijdens de DFNA21-familie dagen. Ik zie dit nog steeds als een van de meest leerzame en bijzondere dagen van mijn PhD en ben blij dat ik hierbij mocht zijn. Ook de studenten die ik de afgelopen jaren heb mogen begeleiden wil ik bedanken voor hun onmisbare hulp. **Jill, Joëlle en Vince**: jullie hebben allemaal een significante bijdrage geleverd aan een van de hoofdstukken van dit proefschrift, zonder jullie had dit eindproduct er niet gestaan.

Ook van "de andere kant" wil ik een aantal essentiële mensen bedanken. **Marco**, bedankt voor al het werk dat je hebt geleverd voor het RP17-project. Ik herinner me nog de blik op jouw gezicht toen ik voor het eerst aan kwam zetten met een van de vele grote familie stambomen. De impact was duidelijk, en we waren vastbesloten dit samen op te lossen, en dat is gelukt! Dankjewel voor je belangrijke bijdrage hieraan. **Ellen**, jij zorgt ervoor dat alles altijd op rolletjes loopt. Heel veel dank voor het beantwoorden van de tientallen vragen die ik de afgelopen jaren gesteld moet hebben. **Zeinab**, our PhD-trajects were almost in sync and I am really happy we could always discuss our struggles, problems, doubts, or victories together. It was nice having someone around that was in the same situation. Sometimes it may have felt as a never-ending story, but we made it! Thank you for being such a warm and friendly colleague and I wish you all the best with your new position in Utrecht. Also the other PhD-students in the group (**Mubeen, Tabea, Tom, Stéphanie** and **Zelia**) and of course my fellow Postdocs (**Daan** and **Rebekkah**), thank you for being such nice colleagues and creating a great group atmosphere. I am looking forward to keep working together in the future! **Rebekkah**, thank you for starting up the EJPRD-project when I was still too busy writing this thesis. Because of you, and your amazing organizational skills, we could really kick-start this project. I would also like to thank all other members from the **Sensory Diseases research groups**: thank you for all your help, useful discussions, and nice chats at the coffee machine that we had over the past years! Also, all colleagues from the genetic diagnostics and ophthalmology departments: thank you for your involvement, input, and significant contributions to all my different projects.

Also from outside the department, and outside Nijmegen, there are a few people who I would like to thank. I am very grateful to have met and worked together with the **RP17 dream-team: Alison Hardcastle, Mike Cheetham, Alessia Fiorentino, Daniele Ottaviani** and **Vasileios Toulis**. Thank you for all your enthusiasm, encouragements, and invaluable input during our numerous zoom-calls. Without your contribution, we could not have solved this puzzle. What a rollercoaster was that! I am glad we will continue the RP17 journey together and am convinced this story is not over yet. Also all other co-authors and collaborators that contributed to one of my projects (specifically

members from the DOOFNL consortium, European Retinal Disease Consortium (ERDC), **Mhlanga lab** (Nijmegen/Cape Town), **Mundlos lab** (Berlin) and **Zhao lab** (Indianapolis), thank you for all your valuable contribution, feedback, ideas, and suggestions.

Veel projecten zouden niet hebben bestaan zonder de bereidheid van alle patiënten en familieleden om deel te nemen aan ons onderzoek. Ik wil alle patiënten bedanken voor hun tijd, inzet en medewerking. Met name de oprichters en leden van de DFNA21- en RP17-stichting, dank voor jullie ongelofelijke doorzettingsvermogen, betrokkenheid en inzet. Dit geeft mij een belangrijke stimulans en geeft extra betekenis aan mijn werk en motivatie om door te gaan. **Ester** en **Jenny**, ik ben blij dat ik jullie heb mogen leren kennen de afgelopen jaren. Jullie maken mijn werk extra waardevol en ik voel me vereerd dat ik een steentje heb mogen bijdragen aan dit onderzoek.

Ik wil ook alle andere collega's van het lab bedanken (met name iedereen op **Pre-lab 2** en het **ZAT-lab**) voor het beantwoorden van al mijn vragen, meedenken met mijn experimenten, maar vooral ook voor alle gezelligheid, koffie- en lunchpauzes. Ook alle andere collega's met wie ik leuke en spontane momenten heb beleefd tijdens de verschillende borrels, dagjes uit of in de Aesculaaf, dank voor alle gezelligheid! In het bijzonder de leden van de dagje-uit-commissie 2018, **Jettie**, **Laurens**, **Chantal**, **Marc** en **Stephan**, dankjewel dat ik dit samen met jullie heb mogen organiseren en dat we er zo'n geslaagde dag van gemaakt hebben. **Jettie**, ik vind het ontzettend leuk dat hier een vriendschap uit is voortgekomen en hoop nog vaak samen te borrelen nu je ook in Nijmegen woont.

Achter de schermen zijn er ook vele essentiële collega's altijd hard aan het werk. Zij zorgen dat alles blijft draaien en wij onze experimenten kunnen blijven doen. Heel veel dank aan iedereen die zich hiervoor inzet, maar vooral aan de dames van het secretariaat, de stamcelfaciliteit en van de celkweekfaciliteit. **Anita**, **Marlie** en **Saskia**, dankjewel voor het afleveren van de honderden DNA'tjes die ik de afgelopen jaren aangevraagd moet hebben en vooral **Saskia** bedankt voor alle (extra) uren die je hebt besteed aan het uitzoeken van de vele familiestambomen waarvan ik weer had besloten om mee aan de slag te gaan. Ik ben blij dat we veel van deze families nu als "solved" kunnen bestempelen!

Besides in the lab, I also spent many, many hours in the different offices at Q4. I really enjoyed the first years in the "kantoortuin", which allowed me to get to know many PhD-students from the different research groups within the department. During my final year I received a very warm welcome in the famous "Human Fanatics office". **Cenna**, **Evelien**, **Janine**, **Judith**, **Marije**, **Nikoleta** and **Siebren**, thank you for all those fun moments and

social gatherings! I am happy that I arrived just on time to join the “Achterhoek-tour”, I wouldn’t want to have missed that Abba-tribute band! Thank you for still allowing me as a Postdoc in your “flex” office.

Sarah, we haven’t shared offices or labs, but nevertheless we became good friends over the years. We started our PhDs around the same time, and since then we started to have coffee breaks, social outings and we had a lot of fun together. Oh, the joy we had in Vancouver! Even when really stressed and frustrated with work, I really enjoyed our conversations and moments together. I am a bit sad you are leaving Nijmegen, but I will for sure come and visit you and **Matze** in Tübingen.

Dan mijn lieve vriendinnen. **Laurie** en **Jasmijn**. Ik zeg het misschien niet zo vaak hardop, maar ik ben ontzettend dankbaar voor onze vriendschap en hoe jullie mij er de afgelopen jaren doorheen hebben gesleept. Jullie zorgden er altijd voordat ik zelfs tijdens mijn drukste periodes nog pauzes nam en een goed verzorgd bordje comfortfood voorgeschoteld kreeg (vooral dankjewel **Jasmijn**), natuurlijk vergezeld met een goed glas wijn (dankjewel **Laurie**). Dit allemaal met een grote dosis gezelligheid en vaak veel foute muziek. Ik ben blij dat jullie er zijn! Ook **Ilze**, **Jochem**, **Josje**, **Judith** en **Maud**, ik ben blij dat we elkaar nog niet uit het oog verloren zijn ondanks de grote afstand (**Ilze**, **Jochem** en **Maud** (ja, jij nu ook!)) of minder grote afstand (**Josje** en **Judith**). Heel veel dank voor de fijne afleiding in de vorm van de vele weekendtripjes, vakanties of gewoon gezellige avonden of spontane borrelmomenten die we de afgelopen jaren hebben gehad.

Ook de vriendengroep van “thuishuis” wil ik graag bedanken voor de welkome afleiding tijdens onder andere de gezellige weekendjes weg. Ook al zien we elkaar niet heel vaak, ik vind het heel bijzonder dat we nog steeds met zo’n grote groep contact hebben en dat het toch altijd meteen weer gezellig is. **Nathalie**; jou zie ik natuurlijk wel geregeld. Ondanks dat we ons met heel andere zaken en interesses bezighouden hebben wij altijd een goede klik gehad. Nu je wat verder weg woont zien we elkaar helaas minder vaak, maar laten we elkaar vooral snel weer opzoeken! Dankjewel voor alle gezelligheid, squash overwinningen, de vele lachbuien, een luisterend oor op zijn tijd en natuurlijk de vele reizen die we samen hebben gemaakt!

Lieve **papa** en **mama**, de deur is écht altijd bij jullie open blijven staan. Jullie hebben altijd in mij geloofd en mij gestimuleerd, ook wanneer ik het zelf even niet meer zag zitten. “Bij jou komt het toch wel weer goed” is een vaak gehoorde uitspraak. En wat blijkt, jullie hebben het ook nu weer bij het rechte eind gehad. Wat is het fijn om na een lange week weer terug naar Angeren te gaan en daar verwelkomd te worden met een bordje

warm eten en een luisterend oor, een knuffel, en helemaal uit te kunnen schakelen. Ook dank jullie wel voor alle hulp met mijn ontelbare verhuizingen, tuinverbouwingen en taxiritjes naar de bushalte. Zonder jullie hulp had ik dit niet kunnen doen en het betekent heel veel voor me dat jullie altijd voor mij klaar staan! Lieve **Joris**, ondanks onze verschillende interesses ben je bent altijd wel enorm geïnteresseerd geweest in mijn onderzoek maar ook oplettend dat ik mezelf niet in mijn werk verlies. Ik vind het altijd erg gezellig om samen een broer-zus activiteit te ondernemen en ik ben blij dat we het zo ontzettend goed met elkaar kunnen vinden. Het is ontzettend leuk dat je nu weer dichterbij bent komen wonen en dat we elkaar weer vaker kunnen opzoeken. Hopelijk vergeet je me niet wanneer je zelf heel succesvol en belangrijk bent geworden met je eigen bedrijf. Dankjewel voor alle steun en aanmoediging!

Op naar de volgende avonturen!

Suzanne

ABBREVIATIONS

2'-MOE	2'-O-methoxyethyl
2'-OMe	2'-O-methyl
3D	three-dimensional
aa	amino acids
AAV	adeno-associated virus
ABR	auditory brainstem response
ACMG	American College of Medical Genetics
adNSHL	autosomal dominant nonsyndromic hearing loss
adRP	autosomal dominant retinitis pigmentosa
AF	allele frequency
ALS	amyotrophic lateral sclerosis
AON	antisense oligonucleotide
arHL	autosomal recessive hearing loss
arRP	autosomal recessive retinitis pigmentosa
ARTA	age related typical audiogram
ATAC-seq	assay for transposase accessible chromatin using sequencing
BSA	bovine serum albumin
CADD	Combined Annotation-Dependent Depletion
CAGE	Cap Analysis of Gene Expression
CCS	circular consensus sequencing
CEVA	Caucasian EVA
CGH	comparative genomic hybridization
ChIP-seq	chromatin immunoprecipitation sequencing
cM	centimorgan
CNV	copy number variant
Co-IP	co-immunoprecipitation
CRE	<i>cis</i> regulatory element
CUT&Tag	Cleavage Under Targets and Tagmentation
cVEMP	cervical vestibular-evoked myogenic potential
dB HL	decibel hearing level
DGV	Database of Genomic Variants
DNase-seq	DNase I hypersensitive sites sequencing
dsDNA	double-stranded template DNA molecule
DVD	Deafness Variation Database
ERDC	European Retinal Disease Consortium

ERG	electroretinography
EVA	enlarged vestibular aqueduct
FAF	fundus autofluorescent
FISH	fluorescent in situ hybridization
GO	Gene Ontology
GWAS	genome-wide association study
HBSS	Hank's balanced salt solution
HGMD	Human Gene Mutation Database
HL	hearing loss
Hz	hertz
IMPC	International Mouse Phenotyping Consortium
iPSC	induced pluripotent stem cell
IRD	inherited retinal diseases
ISO	international organization for standardization
IVF	<i>in vitro</i> fertilization
kb	kilobase
LINE	long interspersed nuclear element
LNA	locked-nucleic acid
lncRNA	long noncoding RNA
LOD	logarithm of the odds
LOVD	Leiden Open (source) Variation Database
LPA	lysophosphatidic acid
LRS	long-read sequencing
Lyso-PC	lysophosphatidylcholine
LysoPLD	lysophospholipase D
MAF	minor allele frequency
Mb	megabase
MIP	molecular inversion probe
miRNA	microRNA
MKS	Meckel syndrome
NGS	next generation sequencing
NIPT	non-invasive prenatal testing
nm	nanometer
NRV	neuronal retinal vesicle
nt	nucleotide
OCT	optical coherence tomography
OMIM	Online Mendelian Inheritance in Man
oVEMP	ocular vestibular-evoked myogenic potential
PCR	polymerase chain reaction

PEI	polyethyleneimine
PFA	paraformaldehyde
PGD	preimplantation genetic diagnosis
pLI	probability of loss of function intolerance
PPC	photoreceptor progenitor cell
PS	phosphorothioate
PTA	pure tone average
qPCR	quantitative real-time PCR
RD	retinal dystrophies
RIPOR2	RHO family interacting cell polarization regulator 2
RNA-seq	RNA sequencing
RO	retinal organoid
RP	retinitis pigmentosa
RPE	retinal pigment epithelium
scRNA-seq	single cell RNA sequencing
SD-OCT	spectral-domain optical tomography
SE-NL	Southeast Netherlands
SIFT	Sorting Intolerant From Tolerant
smMIP	single molecule molecular inversion probe
SMRT	single molecule real-time
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SRS	short-read sequencing
STGD1	Stargardt disease
SV	structural variant
TAD	topologically associating domain
TF	transcription factor
UPD	uniparental disomy
V1-CEVA	variant 1-CEVA
VarAFT	Variant Annotation and Filter Tool
VNTR	variable number of tandem repeats
WES	whole exome sequencing
WGS	whole genome sequencing
WHO	world health organization
ZMW	zero-mode waveguide

