



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

Year : 2019

## Gene discovery in inherited retinal diseases using whole genome sequencing and autozygome based analysis

Bedoni Nicola

Bedoni Nicola, 2019, Gene discovery in inherited retinal diseases using whole genome sequencing and autozygome based analysis

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB\_B4612FE1A5D53

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



**UNIL** | Université de Lausanne

Faculté de biologie  
et de médecine

**Département de Biologie Computationnelle**

**GENE DISCOVERY IN INHERITED RETINAL DISEASES USING WHOLE  
GENOME SEQUENCING AND AUTOZYGOME BASED ANALYSIS**

**Thèse de doctorat ès Sciences de la Vie (PhD)**

présentée à la

Faculté de biologie et de médecine  
de l'Université de Lausanne

par

**Nicola BEDONI**

Master en Sciences de la Vie de l'Université de Fribourg

**Jury**

Prof. Christine Sempoux, Présidente  
Prof. Carlo Rivolta, Directeur de thèse  
Prof. Hanno Jörn Bolz, expert  
Prof. Wolfgang Berger, expert

Lausanne, 2019

# Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

<b>Président·e</b>	Madame	Prof. Christine <b>Sempoux</b>
<b>Directeur·rice de thèse</b>	Monsieur	Prof. Carlo <b>Rivolta</b>
<b>Experts·es</b>	Monsieur	Prof. Hanno Jörg <b>Bolz</b>
	Monsieur	Prof. Wolfgang <b>Berger</b>

le Conseil de Faculté autorise l'impression de la thèse de

**Monsieur Nicola Bedoni**

Master of Science in Biology Université de Fribourg, Suisse

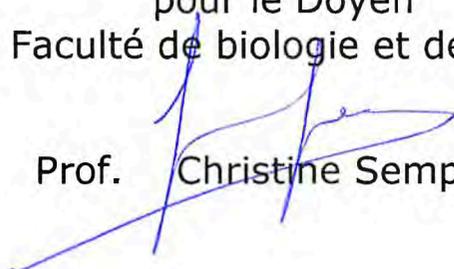
intitulée

**Gene discovery in inherited retinal diseases  
using whole genome sequencing and  
autozygome based analysis**

Lausanne, le 18 janvier 2019

pour le Doyen  
de la Faculté de biologie et de médecine

Prof. Christine Sempoux



## Acknowledgements

*I dedicate my thesis work to my parents, Agnese and Luigi, who always supported me unconditionally. Mamma, papà, grazie di cuore, siete la mia forza.*

*My deepest thank also goes to Prof. Carlo Rivolta, for welcoming me in his group and allowing me to participate to a number of captivating projects. I experienced a 5-years growth at both personal and professional levels. This wouldn't have been possible without your teaching and guidance, that I treasure.*

*I sincerely thank my Jury members Prof. Christine Sempoux, Prof. Hanno Jörn Bolz, and Prof. Wolfgang Berger for having invested their time in reading and evaluating my thesis, and for accepting to meet for my examination.*

*Thanks to Prof Andrea Superti-Furga, who has always been open to discussion, bringing precious advices from his extensive expertise in medical genetics. Thanks to the collaborators of the European Retinal Degeneration Consortium, the Ophthalmic Hospital Jules Gonin in Lausanne, the Instituto de Oftalmologia Dr Gama Pinto in Lisbon, the fertility clinic FERTAS in Lausanne, the Telethon Institute of Genetics and Medicine, and the collaborators from the West Virginia University.*

*A special thank goes to Beryl, who had the courage to guide me through my initiation to bioinformatics, as well as Mathieu and Katarina not only for their contributions with great programming skills but also for being such amazing persons. With no less regard, I would also like to thank Sara, Virginie, Rosanna, Kostas, Pietro, Adri, Suzanne, JJ, Nathalie and Alessandro, who, in their own way, all made the perfect working atmosphere.*

*Thanks to my brother Angelo, whose constant determination is a source of inspiration, and to my non-biological brothers Francisco and Dionisije, you guys are precious!*

*And last but not least, speaking of cherished non-biological family, I would like to express my deepest gratitude to Elena and Sergio. Thanks for always being there for me.*

## Résumé

Chez l'homme, la vision est le plus développé des sens. Nos yeux peuvent être considérés comme le portail à travers duquel on apprend, on communique, et on apprécie la beauté de notre monde.

Le but de ce travail est de mieux comprendre l'étiologie moléculaire de cas mendéliens de dégénérescence de la rétine, ainsi que de créer une base de connaissances des mécanismes conduisant à la maladie, en vue de futurs développements thérapeutiques. Notre stratégie consiste en l'utilisation des techniques de séquençage du génome entier, de l'exome et de l'ARN, ainsi que des techniques de laboratoires de recherche fondamentale, afin d'évaluer et de mieux comprendre, au niveau des processus physiologiques et moléculaires, les conséquences fonctionnelles induites par des variantes génétiques.

Depuis le début de ma thèse de doctorat, mon entraînement était concentré sur la bioinformatique et plus spécifiquement sur les méthodes d'analyse de données de séquençage à haut débit. Grâce à mon éducation en tant que biologiste moléculaire, j'ai pu mener la plupart des projets de A à Z, en touchant à la partie *in silico* ainsi qu'à la mise en œuvre d'une approche expérimentale

Les principaux résultats de mon travail sont de nouvelles associations gène-maladie, en particulier pour ce qui concerne des cas de ciliopathies syndromique avec une atteinte au niveau de la vue et de la fertilité chez l'homme. Dans les deux premiers projets, autour des variantes pathogéniques dans les gènes *TTL5* et *ARL2BP*, on associe deux organelles ciliées – le cil du photorécepteur et le flagelle du spermatozoïde – qui sont structurellement similaires, mais qui fonctionnent dans des processus complètement différents : la vision et le système reproductif. Une autre découverte originale est discutée dans le troisième projet, dans lequel on associe une variante de structure dans le gène *NMNAT1* avec une forme syndromique rare, sévère et inconnue d'amaurose congénitale de Leber. Pour finir, le quatrième projet a pour sujet l'investigation par moyen du séquençage du génome en entier de cas irrésolus par séquençage de l'exome, avec l'identification de deux nouvelles variantes de structure dans les gènes *EYS* et *CNGA1*, ainsi qu'un réarrangement chromosomique à la base d'une isodysomie uniparentale.

## Abstract

Vision is the most developed of the human senses. Our eyes can be considered as the portal through which we gain knowledge, we communicate, as well as we appreciate all the beauty of the world.

The goal of this work is to better understand the molecular etiology underlying rare Mendelian cases of retinal degeneration and create knowledge about this disease's mechanisms, in view of future developments of therapeutic interventions. In our approach, we couple the use of whole-genome, whole-exome, and RNA-sequencing, along with bench side basic research, in order to assess and better understand the functional consequences that identified mutations elicit on normal molecular and physiological processes. From the beginning of this PhD, my training focused on bioinformatics and specifically on methods of analysis of next generation sequencing data. Provided that my background is that of a molecular biologist, I could in fact develop all my projects from A to Z, by performing *in silico* genomic data analysis and by implementing it into experimental research approaches.

The main results of my work consist in novel disease-gene associations, in particular for syndromic ciliopathies causing loss of vision and male infertility. In the projects on the *TLL5* and *ARL2BP* genes we link two structurally similar, yet functionally different, ciliary organelles - the photoreceptor sensory cilium and the sperm flagellum - associating vision and the reproductive system. Another novel disease-gene association is discussed in the third project, where we bring evidences for the contribution of a disruptive structural variant in the *NMNAT1* gene to a rare, severe, and unique form of syndromic Leber congenital amaurosis. Finally, in the fourth project, we perform a whole-genome sequencing investigation of unresolved, negative exomes, and we identify new structural variants in the *EYS* and *CNGA1* genes, as well as a large chromosomal rearrangement leading to uniparental isodysomy.

# Index

<b>Acknowledgements</b> .....	<b>1</b>
<b>Résumé</b> .....	<b>2</b>
<b>Abstract</b> .....	<b>3</b>
<b>Index</b> .....	<b>4</b>
<b>Introduction</b> .....	<b>5</b>
Anatomy of the eye.....	5
The retina .....	6
The photoreceptor.....	7
Physiology of the retina.....	8
The primary cilium and motile cilia.....	9
Inherited retinal dystrophies.....	11
Ciliopathies .....	14
Next Generation Sequencing.....	15
<b>Projects with first-author contribution</b> .....	<b>17</b>
<b>PROJECT 1 - ARTICLE</b> .....	<b>17</b>
Mutations in the polyglutamylase gene <i>TLL5</i> , are associated with cone-rod degeneration and reduced male fertility.....	17
<b>PROJECT 2 - ARTICLE</b> .....	<b>28</b>
<i>ARL2BP</i> and primary ciliary dyskinesia: a mouse and human parallel study .....	28
<b>PROJECT 3</b> .....	<b>65</b>
<i>NMNAT1</i> variants: between LCA and lethality .....	65
<b>PROJECT 4</b> .....	<b>84</b>
NEXOME: A whole-genome study of Negative EXOMEs.....	84
<b>General discussion and conclusive remarks</b> .....	<b>93</b>
<b>References</b> .....	<b>96</b>
<b>APPENDIX</b> .....	<b>101</b>
<b>Contributions to other projects</b> .....	<b>101</b>
<b>PROJECT 5 - ARTICLE</b> .....	<b>101</b>
A homozygous founder missense variant in arylsulfatase G abolishes its enzymatic activity causing atypical Usher syndrome in humans .....	101
<b>PROJECT 6 - ARTICLE</b> .....	<b>111</b>
Analysis of the genetic basis of periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) syndrome.....	111
<b>PROJECT 7 - ARTICLE</b> .....	<b>122</b>
Mutations in <i>CEP78</i> Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects .....	122

# Introduction

## Anatomy of the eye

The eye is composed of three different layers, each with specific functions. The external layer is formed by the sclera and the cornea. The sclera, or the “white of the eye”, consists in dense connective tissue forming a supportive wall of the eyeball. It is continuous with the cornea, a transparent tissue that, together with the lens make the optical system of the eye, providing a sharp image at the level of the retina. The second, or intermediate layer, consists in an anterior part, with the iris and the ciliary body, which regulate the amount of light getting into the eye, and a posterior part, the choroid, consisting in a highly vascularized membrane nourishing the particularly-demanding cells of the retina. Finally, there is the retina, that makes the third and most internal layer, consisting in the sensory part of the eye. All ocular structures are important for proper vision, but the retina is probably the most essential because it harbors photoreceptors, where light is captured and translated into visual cues, which are then sent for processing to the brain via the optic nerve, in order to produce visual information.<sup>1,2</sup>

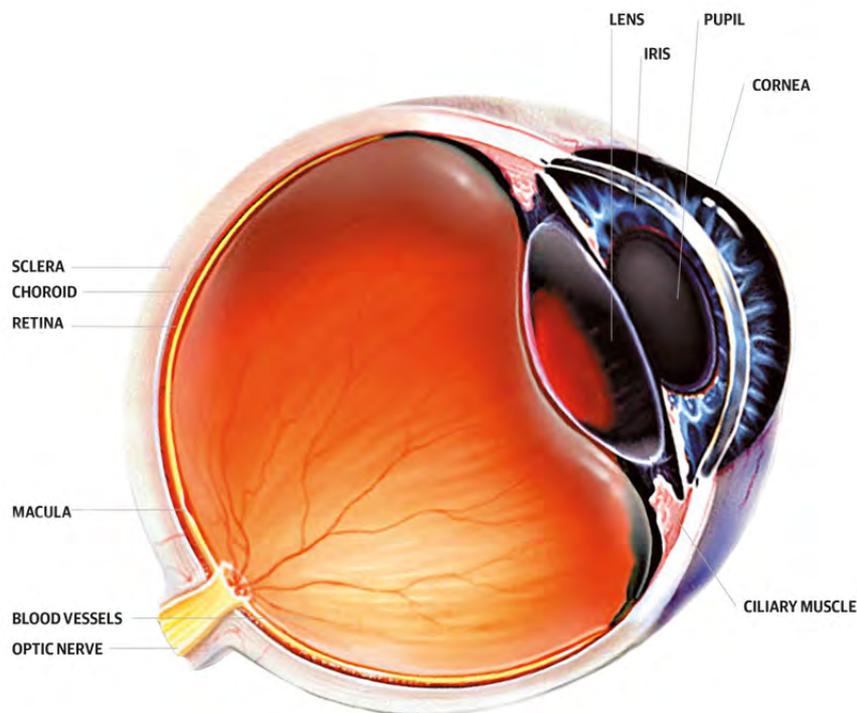
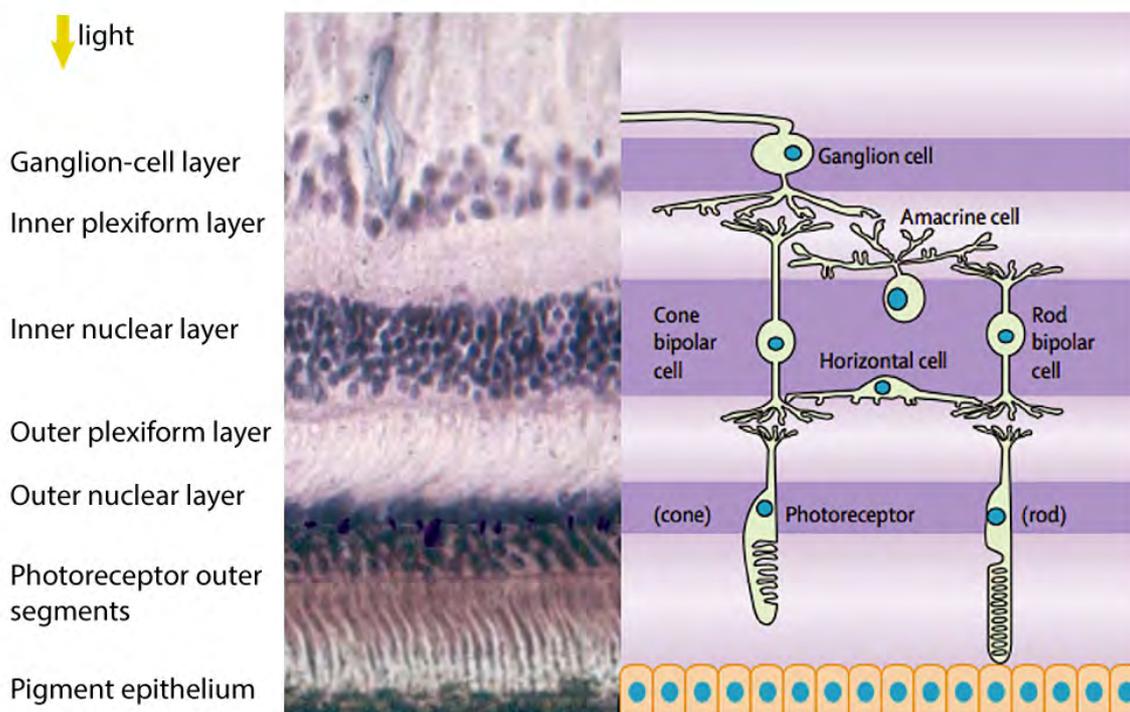


Figure 1 : A schematic view of the anatomy of eye. (Adapted, from <http://atlas.iapb.org/>)

## The retina

During development, the retina forms by invagination out from the neural tube, and therefore, at least from an embryonic point of view, it is considered as part of the central nervous system. It consists in a very thin layer of cells (approximately 0.4 mm) lining the back of the eye. In a typical ophthalmoscopic exam, we can notice two distinct anatomical regions: the macula and the optic disc. The macula is localized near the center of the retina and is responsible for giving us sharp central vision and fine details, with the very best vision provided by the tiny depression at its center, namely the fovea. The optic disc, also called “blind spot” because of its lack of photoreceptors, is where all axons of the ganglion cells meet to form the optic nerve. Millions of photoreceptors are capturing the light, generating action potentials that travel throughout the optic nerve, towards the brain’s visual centers. Before reaching the photoreceptor layer, light must go through more cell layers, i.e. blood vessels, nerve fibers, and four types of neurons (bipolar cells, ganglion cells, horizontal cells, and amacrine cells).<sup>1,3-5</sup> Figure 2 depicts a histological section of the different cell layers inside the retina, with its relative schematic representation.



**Figure 2:** Histological (left) and schematic (right) appearance of a healthy human retina.

Modified from Hartong et al<sup>6</sup>

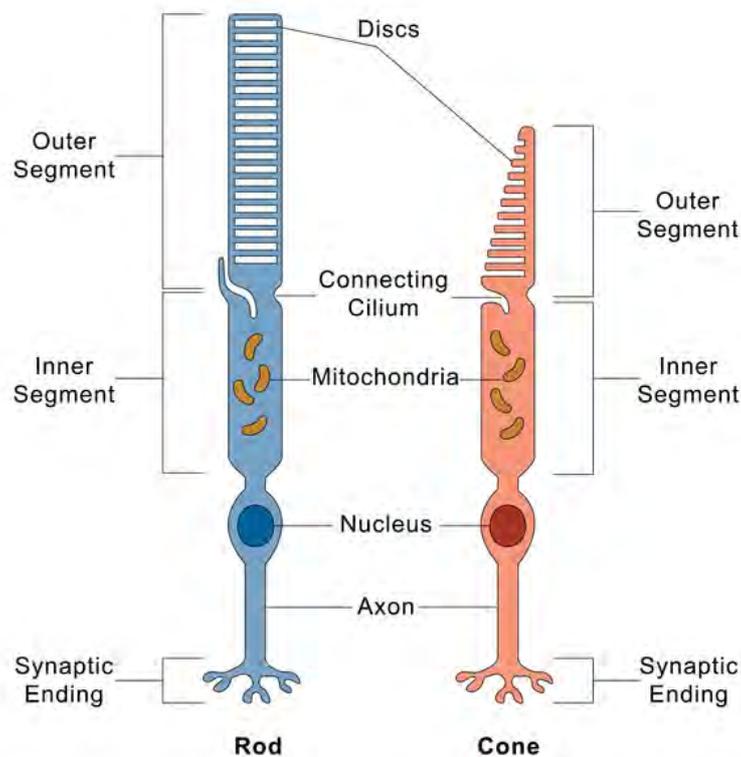
The outer nuclear layer is formed by cell bodies of photoreceptors, the inner nuclear layer comprises the cell bodies of amacrine, bipolar and horizontal cells, and the ganglion cell layer contains ganglion cells. The inner and outer plexiform layers are neuropils, where synaptic contacts between axons and dendrites from neuronal cells occur. The horizontal cells, placed at the first synapse of the visual system, modulate the information flow from photoreceptors to bipolar cells<sup>7</sup>. On the other hand, while bipolar cells have also a direct function in the transmission of the signal from the photoreceptor to ganglion cells, amacrine cells are interneuronal cells, and function more as “message collectors”. Bringing nutrients to all those layers, and phagocytizing the waste of their metabolism, the retinal pigment epithelium (RPE), firmly attached to the choroid, is placed in direct contact with the photoreceptor outer segment. Specifically, the RPE has two main roles: it is responsible for the phagocytosis of fragments of cones and rods that shed from their distal ends; and it is fundamental for the uptake, processing, transport and release of retinol (vitamin A) and some of its visual cycle intermediates<sup>8</sup>.

## **The photoreceptor**

As mentioned before, photoreceptors are responsible for performing the first step of the light signal transmission to the brain. Generally, in the vertebrate retina there are two types of photoreceptors: **rods** and **cones**. Despite their fundamentally similar structures, they accomplish distinct functions, with responses to different light intensities and with varying densities along the retina. Rods are very numerous (110-125 millions) and are specialized in vision at low light levels<sup>9</sup>. They are mainly present in the periphery of the retina, and their density decreases towards more central areas, reaching absence in the foveal region. On the other hand, cones (6.4 millions) provide sharp and color vision in daytime lighting conditions and are primarily located in the fovea and its surrounding area<sup>9</sup>.

This cone- or rod-specific light absorption is mediated by the presence of distinct **opsin** molecules that are capable of absorbing different wavelengths. Rhodopsin, making the photopigment of rods, has indeed a great sensitivity and it is able to detect very low amounts of light. Conversely, cones are constituted of less sensitive opsins, but they are much more discriminating for specific wavelengths. In human, three different opsins are responsible for capturing light at 430, 530 and 570 nm, allowing trichromatic vision (red, green and blue).

Structurally, the photoreceptor cell has evolved to produce a cell compartment packed with several membrane discs (the outer segment). This atypical structure increases the surface area for light capturing. The photopigment-containing outer segment is connected to the inner segment through a narrow section, the so-called connecting cilium (Figure 3). All proteins destined for maintenance of the outer segment must pass through this thin portion of the photoreceptor. The photoreceptor inner segment contains the endoplasmic reticulum, the Golgi apparatus, and mitochondria densely packed near the outer segment, in order to meet the high demand for metabolic energy associated with phototransduction.<sup>10-12</sup>



**Figure 3** : Structural scheme of rod and cone photoreceptor cells.

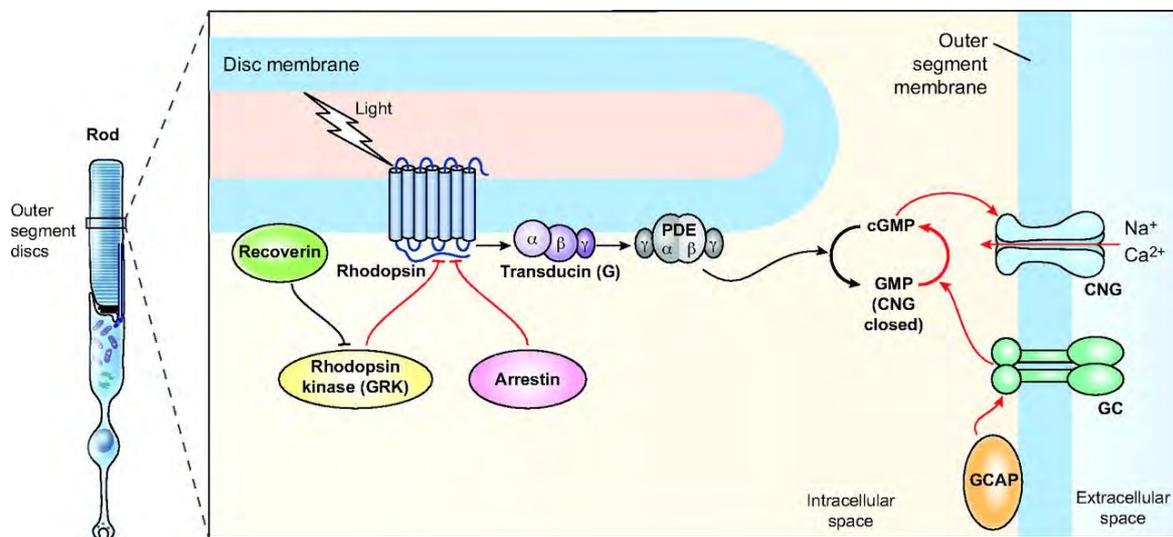
(<https://ghr.nlm.nih.gov>)

## Physiology of the retina

**Phototransduction** is the process of conversion of light into biological signals. This occurs thanks to the photopigments present in the outer segment of photoreceptors. Each photoreceptor type has specialized photopigments, consisting in a protein moiety, belonging to the class of opsins, conjugated to a small light-absorbing compound, called chromophore, and derived from vitamin A (also known as retinal). Throughout the animal kingdom there is variation in the structure of photoreceptors, nonetheless the

phototransduction cascade is highly conserved and is based almost exclusively on the function of opsins<sup>13</sup>.

When light reaches the photoreceptor, the chromophore isomerizes from 11-*cis*- to all-*trans*-retinal, leading to a conformational change of the opsin. As a consequence of this chemical shift, a specific G-protein called transducin is activated. The cascade follows with transducin activating a phosphodiesterase, which is responsible for the breakdown of cGMP, leading to closure of cGMP-gated cation channels and causing hyperpolarization of the membrane with the consequent decrease in release of the neurotransmitter glutamate. On the other hand, in darkness, cGMP is present in high concentration in the disk stacks, cation channels are open, the membrane depolarizes and glutamate is constantly released for transmission to bipolar cells<sup>14</sup>.



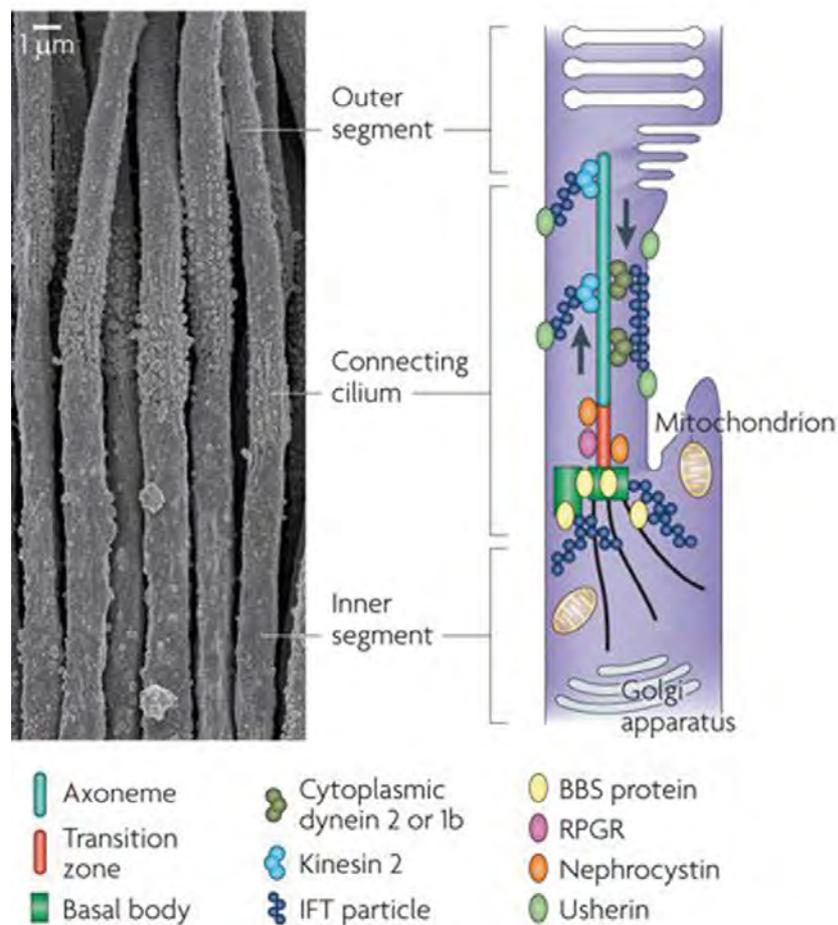
**Figure 4** : Schematic view of the major proteins involved in the phototransduction cascade<sup>15</sup>.

Another important step for the regeneration of photoreceptor light sensitivity is the so-called **visual cycle**. All-*trans* retinal is transported to the RPE for recycling and is then returned to the photoreceptor in its *cis* form, allowing to restore the photopigment functionality<sup>16</sup>.

## The primary cilium and motile cilia

Virtually all vertebrate cells have the ability of developing a primary cilium. In contrast to motile cilia and flagella, which share common structures, the majority of primary cilia are immotile and have evolved principally as sensory organelles<sup>17</sup>. Structurally, all primary cilia are composed of three main parts: the basal body (BB), the connecting-cilium (CC) and

the axoneme. The BB connects with structures of the inner segment, named rootlet, and as it is suggested by its name its function is to anchor the cilium to the endoplasmic reticulum<sup>18</sup>. The CC is the backbone of the axoneme and corresponds to the transition zone between outer and inner segments. Connecting axoneme and BB, it works as a bridge between these two structures. As depicted in Figure 5, there is a high flow of proteins from the inner to the outer segments. This flow is mediated through the intraflagellar transport (IFT), where kinesin 2 motors are responsible for the transport towards the tip of the axoneme, whereas dynein motors towards the basal body.

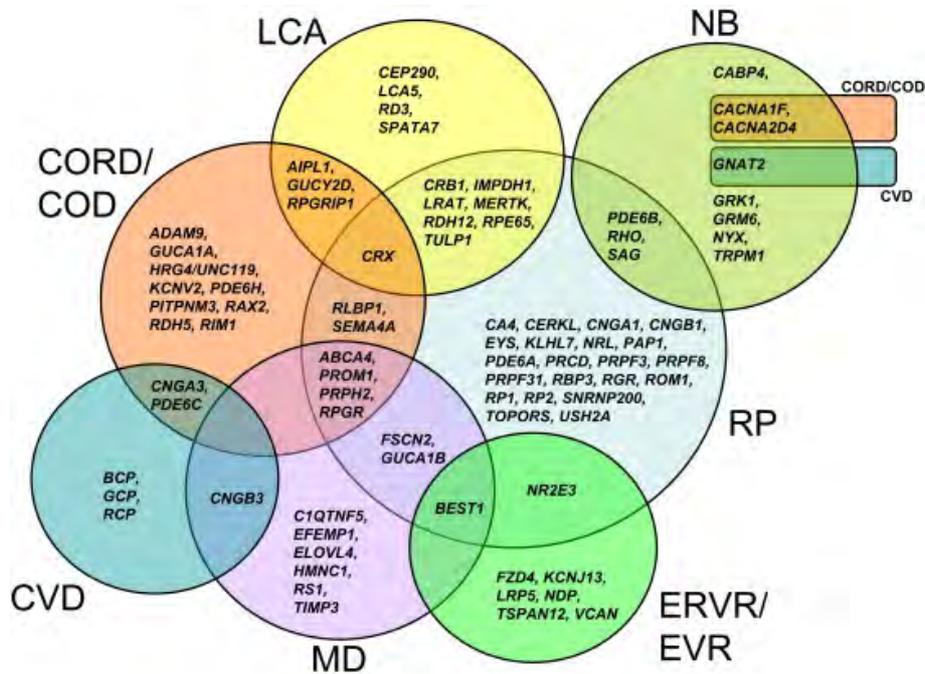


**Figure 5 :** The connecting cilium. The different segments of the photoreceptor sensory cilium are indicated both in the electron micrograph, on the left, and in the schematic representation of the connecting cilium on the right. Modified from Wright *et al* (2010)<sup>19</sup>.

Finally, axonemes, the part of primary cilia that protrudes from the basal body, are composed of specific microtubule arrangement (Figure 6), that largely determine the function of the ciliated cell. Indeed, the highly-organized microtubule patterns confer the functional specificity of both motile and primary cilia.



(Figure 7). IRDs are usually rare and monogenic with a prevalence in the population of about 1 in 3,000 individuals<sup>19</sup>. To date, a number of genes have been associated with retinal photoreceptor degeneration. Nevertheless, in both recessively and dominantly inherited forms, an estimated 30-40% of genetic background remains to be elucidated<sup>21-24</sup>.



**Figure 7 :** Diagnostic classes of non-syndromic monogenic retinal and vitreoretinal diseases and their causative genes. RP: retinitis pigmentosa; NB: night blindness; LCA: Leber congenital amaurosis; CORD/COD: cone rod and cone dystrophies; CVD: colour vision defects; MD: macular degeneration; ERVR/EVR: erosive and exudative vitreoretinopathies. From Berger *et al* (2010).<sup>23</sup>

The most prevalent IRD is **retinitis pigmentosa (RP)**, affecting approximately 1/4000 individuals worldwide<sup>6,25,26</sup>. It is primarily characterized by a poor night vision, due to degeneration of rod photoreceptors. The age of onset of RP is variable, developing in most of the cases during mid-adulthood. Due to its variable phenotype, the diagnosis is mainly done by electrophysiology (electroretinogram (ERG)) and imaging analysis (ophthalmoscopy or slit lamp funduscopy, and optical coherence tomography (OCT)). Typically, fundus imaging of RP patients reveals pigmented deposits in the periphery of the retina (also called bone spicule deposits). This pigmentation is due to the migration of the retinal pigment epithelium into the neural portion of the retina, as a consequence of

photoreceptor cell death<sup>27</sup>. With progression of the disease, cones and central vision are also affected, with cases that result in complete blindness. On the contrary, **cone** (CD) and **cone-rod degenerations** (CRD) consist in the primary loss of cones, followed by the progressive degeneration of rods and the peripheral retina in CRD. The disease progression is generally more severe than in RP. In CD and CRD, the first signs of the disease are disturbance of the central and colour vision, with retinal pigment deposits appearing in the macula. Subtypes of CD depend on the loss of specific cone photoreceptors and they can result in **achromatopsia**. The most severe form of IRD is considered to be **Leber congenital amaurosis** (LCA), with severe loss of photoreceptors in the first years of life. LCA often comprises other ocular features, such as nystagmus and amaurotic pupils. With a prevalence of 1 affected individual every 30,000 newborns, it is considered much rarer than RP.<sup>28</sup>

Further classification consists in non-syndromic forms, affecting only the retina or the RPE, and syndromic forms, when the blinding phenotype is associated with pathologies in other tissues<sup>23</sup>. A typical **syndromic** form associating with RP is **Usher syndrome**, which in humans is characterized by mild to severe deafness. Usher syndrome has a prevalence of 1 in 25,000 and represents the most common form of deafness-blindness. It has been demonstrated that genes mutated in Usher syndrome encode for structural proteins mostly implicated in the correct shaping of the auditory hair cell, with a role in the maintenance of the mechanoelectrical transduction machinery.<sup>29,30</sup>

Another example of syndromic RP is **Bardet-Biedl syndrome** (BBS). The primary features of BBS patients include retinal degeneration, obesity, polydactyly, renal and genital abnormalities, as well as intellectual disabilities. However, the disease can manifest with high variability, even within affected family members, and sometimes include additional phenotypes. This syndrome is very rare, and affects 1 person in 120,000 individuals of European ancestry.<sup>23,31</sup> To date, about a dozen genes have been associated with BBS. Affecting approximately 20 % of BBS cases, the most frequent are mutations in the *BBS1* gene, inherited in an autosomal recessive fashion<sup>32</sup>.

Another rare syndromic disorder, is the **Senior Løken syndrome**, combining phenotypes in the eyes and in renal function. Specifically, patients show nephronophthisis (a form of cystic kidney disease) and RP (or in some cases a diagnosis of LCA because of more severe phenotype). Genes causatives for Senior Løken have been found to participate in the structure and function of primary and motile cilia. An example of these is the *NPHP1* gene, mutated in approximately 20% of all cases. *NPHP1* (Nephrocystin1) has been shown to

localize to the transition zone of the renal, respiratory and photoreceptor connecting cilia<sup>33</sup>. *In vivo* and *in vitro* studies have demonstrated that NPHP1 interacts and co-localize with different other ciliary proteins that are known to cause Senior Løken syndrome, including nephrocystin-3, -4 and inversin<sup>34-36</sup>. However, individuals with defective NPHP1 have also been described suffering from Bardet-Biedl syndrome and another syndromic entity, Joubert syndrome, demonstration that the NPHP1-related disease is characterized by phenotypic heterogeneity.

The **Joubert syndrome** (1 affected in 100,000) is characterized by hypoplasia of the cerebellar vermis and accompanying neurologic symptoms, including dysregulation of breathing pattern and developmental delay. In some cases, patients also showed retinal dystrophy and renal anomalies (e.g.: NPHP1 defects<sup>37</sup>).

## **Ciliopathies**

As it is often the case, syndromic IRDs may be part of a broader phenotypic spectrum, consisting in ciliary dysfunction. This class of diseases is more frequently called “ciliopathy”. **Ciliopathies** involve deficiencies in ciliary and cilia-associated proteins and often affect a variety of tissues and organs<sup>38</sup>. Because of the presence of the connecting cilium in both rod and cones photoreceptors, many ciliopathies display a retinal phenotype, either as part of a syndromic condition (usually associated with hearing defects, e.g. Usher syndrome) or as the only pathological sign<sup>39,40</sup>.

Cilia and flagella, highly conserved in their core structure, are ancestral organelles composed of more than 650 proteins<sup>41</sup>. Cilia play crucial roles in normal function of most tissues of the human body, including development, tissue morphogenesis and homeostasis<sup>40</sup>. Some examples of diseases affecting primary cilia are polycystic kidney disease, Usher syndrome, retinitis pigmentosa, Bardet-Biedl syndrome and Joubert syndrome. On the other hand, motile cilia defects have been shown to be causative for Kartagener syndrome and other diseases grouped under the name of primary ciliary dyskinesia, with typical disease manifestations such as situs inversus, male infertility due to immotile spermatozoa, and chronic airway diseases<sup>20,42</sup>.

Two of the reports in the second chapter of this thesis – project 1 (TTLL5) and project 2 (ARL2BP) – are examples of evidences of new emerging ciliopathic entities in human, with an impact at the level of the eye photoreceptors and of the sperm flagellum.

## Next Generation Sequencing

Next Generation Sequencing (NGS), also called high-throughput sequencing, is a powerful tool to investigate many inherited human diseases. Since the appearance of the first NGS instruments on the market in the first decade of the year 2,000, impressive advances have enabled an increasing number of applications, in parallel to a decrease in costs associated with this technology.

The **whole-exome sequencing** (WES) technique, contemplating the sequencing of exonic DNA regions, has become the chosen 'routine' approach in most research laboratories; nonetheless it is used with an increasing extent in clinical diagnostics.

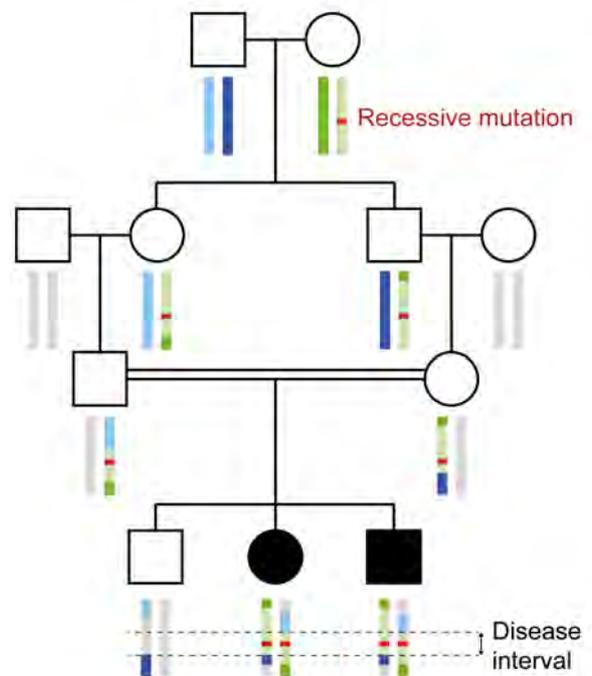
A more expensive but also more complete approach is called **whole-genome sequencing** (WGS). In contrast to WES, which investigates about 2 % of an individual's genetic information, WGS consists in determining the full sequence of an individual's genome. Moreover, this technique allows the identification of less "canonical" mutations due to structural variations such as large deletions, insertions, duplications, inversions, translocations and copy number variations. However, due to the generation of increasingly large amounts of raw files, analysis of any NGS data requires an adapted computational pipeline that is relevant for the phenotype studied.

The **general workflow of NGS** consists in an initial step named 'library construction'. During this first phase, the DNA is fragmented into smaller pieces that are then ligated to special adapters. In the case of WES, the fragmented DNA is captured by targeted hybridization, using probes specifically designed for exonic sequences, and it is then amplified. In contrast, for WGS the library construction does not require hybridization to probes, avoiding the risk of non-uniform coverage. Further comparison of WGS and WES techniques are discussed in project 4 – called NEXOME.

The simplest high-throughput sequencing strategy is through single-read sequencing, during which the sequencer reads DNA fragments only from one end. The length of the reads is variable, depending on the instrument and the technique used. On the other hand, in paired-end sequencing the DNA sequences are read in both directions – from both ends of the DNA fragment – resulting in two reads, separated by a variable distance, and harboring the same barcode identification for the DNA fragment that has been read. This peculiarity of the paired-end sequencing, allows higher quality alignments, and facilitates the detection of genomic rearrangements and repetitive sequence elements, as well as gene fusions and novel transcripts.

After sequencing, the workflow consists in performing **mapping and variant calling**. The short NGS reads are mapped to the reference genome using an appropriate software (e.g. Novoalign), and subsequently the variants, i.e. the differences between the reads and the reference genome, are listed in a standard file format named vcf (variant calling format). Roughly, each person harbors 4,000,000 variants genome wide, and 60,000 variants in the exons. Finally, in order to make sense out of this multitude of variants, we proceed with **annotation and filtering**. During this step, meaningful information is added to the called variant, e.g. gene name, genomic position, frequency in the population and in in-house databases, impact at the protein level, degree of conservation of the involved amino-acids, prediction of the effect on splicing, prediction of deleteriousness from different software, and other information relevant to the disease of interest. Additionally, low-quality calls are discarded based on genotype quality, depth of coverage, and other read quality metrics.

Following these procedures, we often find ourselves with too many, or too few candidate variants. Depending on the disease mode of inheritance different research approaches have to be considered for efficiently analyze the NGS data. In the case of rare diseases with Mendelian inheritance, consanguineous unions dramatically increase the probability for the offspring to be homozygous at any genetic locus (Figure 8)<sup>43</sup>. Indeed, very rare recessive disorders are predominantly observed in small isolated towns, where parents are closely related, and in regions with high levels of endogamy. In this typical case studies **homozygosity mapping** can provide substantial help in narrowing down the number of candidate variants, by looking at the changes occurring in the mapped autozygous regions. Other approaches include **family analysis**, intersecting data from multiple family members, both affected and non-affected.



**Figure 8 :** Model of inheritance of a homozygous (autozygous) region harbouring a mutation (red) from a common ancestor

## Projects with first-author contribution

### PROJECT 1 - ARTICLE

#### **Mutations in the polyglutamylase gene *TLL5*, are associated with cone-rod degeneration and reduced male fertility**

This article was published on August 22<sup>nd</sup> 2016 in the peer-reviewed journal *Human Molecular Genetics*.

#### **Contributions:**

For this project, I performed all the experiments and analyses, except for clinical examinations, done by the different collaborating ophthalmologists (16 of the 32 authors listed), and the WES on the proband of family F6, done by our collaborators in Leeds. I wrote the article under the guidance and expertise of my thesis director Carlo Rivolta. The colleagues from Lausanne supported my work in different ways. Because of their involvement, from clinical examination to blood sampling and DNA extraction. Several authors from the ERDC consortium were added for every recruited patient of the 6 families and for the 365 patient cohort that was screened for mutations in *TLL5*.

## ORIGINAL ARTICLE

# Mutations in the polyglutamylase gene *TLL5*, expressed in photoreceptor cells and spermatozoa, are associated with cone-rod degeneration and reduced male fertility

Nicola Bedoni<sup>1,†</sup>, Lonneke Haer-Wigman<sup>2,3,†</sup>, Veronika Vaclavik<sup>4,5</sup>, Viet H. Tran<sup>4</sup>, Pietro Farinelli<sup>1</sup>, Sara Balzano<sup>1</sup>, Beryl Royer-Bertrand<sup>1,6</sup>, Mohammed E. El-Asrag<sup>7</sup>, Olivier Bonny<sup>8</sup>, Christos Ikonomidis<sup>9</sup>, Yan Litzistorf<sup>9</sup>, Konstantinos Nikopoulos<sup>1</sup>, Georgia G. Yioti<sup>10</sup>, Maria I. Stefanidou<sup>10</sup>, Martin McKibbin<sup>11</sup>, Adam P. Booth<sup>12</sup>, Jamie M. Ellingford<sup>13</sup>, Graeme C. Black<sup>13</sup>, Carmel Toomes<sup>7</sup>, Chris F. Inglehearn<sup>7</sup>, Carel B. Hoyng<sup>14</sup>, Nathalie Bax<sup>14</sup>, Caroline C.W. Klaver<sup>14,15</sup>, Alberta A. Thiadens<sup>15</sup>, Fabien Murisier<sup>16</sup>, Daniel F. Schorderet<sup>17</sup>, Manir Ali<sup>7</sup>, Frans P.M. Cremers<sup>2,3</sup>, Sten Andréasson<sup>18</sup>, Francis L. Munier<sup>4,†</sup> and Carlo Rivolta<sup>1,†,\*</sup>

<sup>1</sup>Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Lausanne, Switzerland, <sup>2</sup>Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands, <sup>3</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands, <sup>4</sup>Jules Gonin Eye Hospital, Lausanne, Switzerland, <sup>5</sup>Department of Ophthalmology, Fribourg Hospital HFR, Fribourg, Switzerland, <sup>6</sup>Center for Molecular Diseases, Department of Pediatrics, Lausanne University Hospital (CHUV), Lausanne, Switzerland, <sup>7</sup>Section of Ophthalmology & Neuroscience, Leeds Institute of Biomedical & Clinical Sciences, University of Leeds, Leeds, UK, <sup>8</sup>Service of Nephrology, Lausanne University Hospital (CHUV), Lausanne, Switzerland, <sup>9</sup>Department of Otorhinolaryngology, Head and Neck Surgery, Lausanne University Hospital (CHUV), Lausanne, Switzerland, <sup>10</sup>Department of Ophthalmology, University of Ioannina School of Medicine, Ioannina, Greece, <sup>11</sup>The Eye Department, St. James's University Hospital, Leeds, UK, <sup>12</sup>Royal Eye Infirmary, Derriford Hospital, Plymouth, UK, <sup>13</sup>Centre for Genomic Medicine, St. Mary's Hospital, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK, <sup>14</sup>Department of Ophthalmology, Radboud University Medical Center, Nijmegen, the Netherlands, <sup>15</sup>Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands, <sup>16</sup>Fertas Andrology Laboratory, Lausanne, Switzerland, <sup>17</sup>Institute for Research in Ophthalmology, University of Lausanne and Ecole Polytechnique Federale de Lausanne, Switzerland and <sup>18</sup>Department of Ophthalmology, Lund University, Lund, Sweden

<sup>†</sup>The authors wish it to be known that, in their opinion, the first 2 authors and last 2 authors should be regarded as joint First and Last Authors, respectively. Received: July 8, 2016. Revised: August 19, 2016. Accepted: August 20, 2016

© The Author 2016. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

\*To whom correspondence should be addressed at: Carlo Rivolta, Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Rue du Bugnon 27, 1011 Lausanne, Switzerland. Tel: +41-21-6925451; Fax: +41-21-6925455; Email: carlo.rivolta@unil.ch

## Abstract

Hereditary retinal degenerations encompass a group of genetic diseases characterized by extreme clinical variability. Following next-generation sequencing and autozygome-based screening of patients presenting with a peculiar, recessive form of cone-dominated retinopathy, we identified five homozygous variants [p.(Asp594fs), p.(Gln117\*), p.(Met712fs), p.(Ile756Phe), and p.(Glu543Lys)] in the polyglutamylase-encoding gene *TLL5*, in eight patients from six families. The two male patients carrying truncating *TLL5* variants also displayed a substantial reduction in sperm motility and infertility, whereas those carrying missense changes were fertile. Defects in this polyglutamylase in humans have recently been associated with cone photoreceptor dystrophy, while mouse models carrying truncating mutations in the same gene also display reduced fertility in male animals. We examined the expression levels of *TLL5* in various human tissues and determined that this gene has multiple viable isoforms, being highly expressed in testis and retina. In addition, antibodies against *TLL5* stained the basal body of photoreceptor cells in rat and the centrosome of the spermatozoon flagellum in humans, suggesting a common mechanism of action in these two cell types. Taken together, our data indicate that mutations in *TLL5* delineate a novel, allele-specific syndrome causing defects in two as yet pathogenically unrelated functions, reproduction and vision.

## Introduction

Cone dystrophies (CDs) and cone-rod dystrophies (CRDs) are rare heterogeneous retinal disorders with an estimated prevalence of ~1:40,000 (1). They lead to severe visual impairment, primarily or exclusively due to the degeneration of cone photoreceptors. Patients experience progressive loss of visual acuity, defective colour vision, photophobia, and have central scotomas. Only later, as the disease progresses, in some cases loss of peripheral vision may also occur (2,3).

The progressive degeneration of retinal photoreceptors in CDs and CRDs is mostly nonsyndromic and has been associated with multiple genetic causes, with at least 20 associated disease genes (RetNet; <http://www.sph.uth.tmc.edu/RetNet/>; date last accessed July 8, 2016). However, more than 75% of cases presenting with dominant or recessive forms of these conditions are genetically unsolved (4). Recent discoveries in CD molecular genetics include the identification of pathogenic variants in the tubulin polyglutamylase *TLL5* (Tubulin Tyrosine Ligase-Like Protein 5) gene, found to cause retinal dystrophy in four British families (5). This gene, like the 12 other members of the *TLL* superfamily, is involved in post-translational modifications of  $\alpha$ - and  $\beta$ -tubulin, which are components of the axonemes of both cilia and flagella.

Interestingly, male mice with a defective *TLL5* display dramatically reduced fertility associated with defects in sperm motility (6). Most sperm tails of mutant mice were found to have disrupted axonemes with loss of tubulin doublets and a significantly decreased polyglutamylation in the upper and lower segments. No abnormal phenotype of retinal photoreceptors or of cochlear cells was initially observed, based on histologic examination (6). A second study with a more thorough characterization of the ocular phenotype in the same mouse model showed a decline of electroretinographic (ERG) amplitudes for both rods and cones in aged mice (20–22 mo). However, no microtubule defects were found after examination of electron micrographs (7).

Ciliopathies represent a class of hereditary disorders involving deficiencies in ciliary and cilia-associated proteins, often affecting a variety of tissues and organs (8). Due to the presence of an immotile cilium in both rods and cones photoreceptors, many ciliopathies display a retinal phenotype, either as part of a syndromic condition (associated with hearing defects, renal

nephronophthisis, liver fibrosis, bone and/or brain anomalies) or as the sole pathological sign (9–11).

Following the investigation of a cohort of patients displaying CD or CRD, we identified mutations in *TLL5* that are associated with both retinal degeneration and reduced sperm motility in humans, possibly defining a novel syndromic ciliopathy.

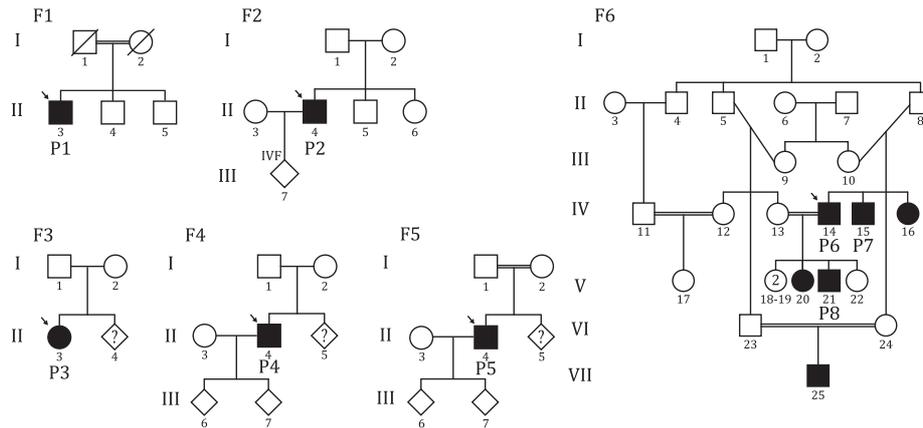
## Results

### Clinical and molecular findings

Our research started with the molecular characterization of a Swiss male patient (P1), aged 75 years, presenting with a late-onset cone dystrophy (CD). He was the eldest of three brothers and his parents were first cousins (Fig. 1, F1), without any history of ocular problems. The patient was first seen at the age of 33 years, when he first noticed blurred vision. His best corrected Snellen visual acuity (BCVA) at that time was 0.6 in the right eye and 0.8 in his left eye. Seven years later, his BCVA was still stable, but worsened when he was 53 years old, dropping dramatically to 0.05 in the right eye and 0.08 in the left eye. He was also complaining of a reduced dark adaptation. Twenty-two years later his vision remained stable, with the patient using more of his peripheral vision.

The clinical examination was typical of a cone dystrophy: the fundus examination showed central foveal atrophy with peripapillary hyperpigmentation and atrophy, while the peripheral retina was within normal limits (Fig. 2). The first ERGs (performed when the patient was 56 years old) showed normal scotopic responses, whereas photopic responses had severely reduced amplitude. The following ERG testing, when the patient was 70 years old, showed some rod involvement, with reduced rod-specific b-wave. The 30-Hz flicker was undetectable. Autofluorescent (AF) images at age 66 years (Fig. 2) showed central hypofluorescence corresponding to atrophy, surrounded by a large hyperfluorescent ring. AF imaging at age 72 indicated that the ring mildly increased in diameter, and so did the area of hypofluorescence (Fig. 2). Kinetic visual field tested at this later age showed mild constriction and central scotoma in both eyes.

The DNA of the patient was first screened for mutations in known disease genes. Following the negative output of a panel-



**Figure 1.** Pedigrees from families with pathogenic *TTLL5* variants. The probands are: subject P1 [III:3, family F1; c.1782del;p.(Asp594Gluufs\*29)], subject P2 [II:4, family F2; c.349C > T;p.(Gln117\*)], subject P3 [III:3, family F3; c.2132\_2133insGATA;p.(Met712IleAspfs\*15)], subject P4 [II:4, family F4; c.1627G > A;p.(Glu543Lys)], subject P5 [II:4, family F5; c.2266A > T;p.(Ile756Phe)], and subject P6 [IV:14, family F6; c.1627G > A;p.(Glu543Lys)].

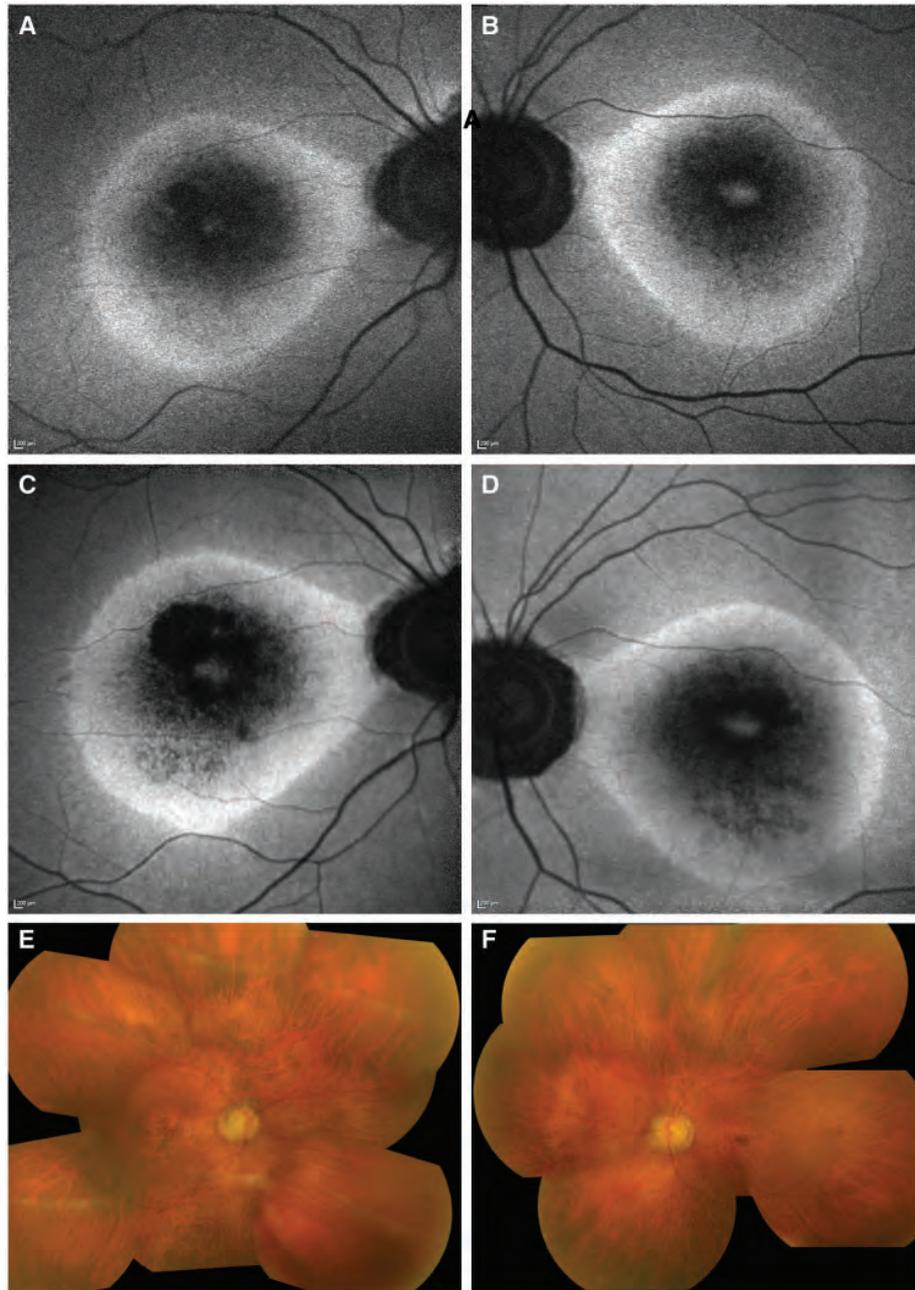
based Next-Generation Sequencing screening [the IROme (12)], we performed whole-genome sequencing (WGS) of the patient's DNA. This latter procedure resulted in more than 4 million DNA variants with respect to the human reference genome (Build hg19). These were evaluated by the use of an internal *in-silico* pipeline assessing their frequency in the general population, quality, etc. (Supplementary Material, Table S1), as well as their presence within autozygous regions (Supplementary Material, Fig. S1). At the end of this process, we were left with 19 variants (7 in autozygous regions), including c.1782delT in *TTLL5* (Fig. 3), a gene known to be involved in microtubule posttranslational modifications and associated with ciliary microtubule stabilization (13). The variant was in exon 20 and consisted of a 1-bp deletion causing a frameshift starting from codon 594 and terminating with the creation of a premature stop triplet 29 codons downstream [p.(Asp594Gluufs\*29)]. It was also not detected in the genome of 400 healthy controls from the same geographic region of Switzerland [data from the CoLaus study (14)], and from publicly available databases [ExAC (15), dbSNP (16)]. Importantly, this DNA change localized to a region of chromosome 14 (75,986,579-80,875,911, Build hg19) showing clear autozygosity (Supplementary Material, Fig. S1) in P1's genome.

Because of the possible involvement of *TTLL5* in extraocular ciliary functions, the patient underwent more detailed clinical examinations. To assess the presence of additional subtle abnormalities in organs known to be affected by ciliopathies, he was evaluated for both renal and otorhinolaryngological functions. He did not display the classical clinical features of renal ciliopathy (no polyuria/polydipsia), but chronic renal insufficiency without proteinuria (stage G3aA1) (17) was identified, associated with low-grade chronic anemia. By ultrasound, the size of the kidneys appeared to be preserved, but a small asymmetry was noted (9.6 cm on the left vs. 10.7 cm on the right side), without significant renal artery stenosis. No cysts were visible and no biopsy was performed. Altogether, the renal features were consistent with age-related decreased renal function, but low-grade and late appearance nephronoptosis cannot be fully excluded. The observed minor sensorineural hearing loss, and loss of some high frequencies, was compatible with the natural course of hearing about a person of the age of the subject (Supplementary Material, Fig. S2). The clinical examination and patient's history did not reveal any upper airway pathology necessitating further investigations for impaired

mucociliary function in the sinus or the bronchi. The patient reported a history of infertility due to reduced sperm motility, diagnosed when he was in his late 20's. He could not have offspring and adopted two children. Following our findings, a new semen analysis was performed at age 75, revealing azoospermia, a sign that nonetheless could simply be related to the difficulty in obtaining an ejaculate at his current age.

Based on these findings, we extended our analyses to a number of additional cohorts of 365 patients with CD and CRD from Switzerland, Sweden, Greece, The Netherlands, and Britain. In a Swedish male patient of Iraqi descent (P2), we identified the homozygous nonsense variant c.349C > T;p.(Gln117\*) by targeted Sanger sequencing of *TTLL5* (Figs. 1, F2; 3). Again, this variant was absent from controls and publicly available databases. Although the patient did not report any history of consanguinity, the occurrence of this extremely rare variant in a homozygous state suggests the presence of residual consanguinity or of a geographical founder effect, which was not tested at the genome level. The patient, aged 46 years, reported no family history of similar visual impairment. Fundus examination revealed degenerative changes, especially in the posterior pole, but more normal features in the periphery. Visual field analysis by Goldmann perimetry showed residual fields in the periphery, but a large central scotoma. Full-field ERG demonstrated residual cone and rod response, consistent with a diagnosis of cone-rod degeneration. In addition to these signs and symptoms typical of CRD, the patient had high myopia (Table 1). This subject was also infertile, but had been able to have a child by *in vitro* fertilization. His semen analysis revealed a normal spermatozoa count but, similar to P1, reduced sperm motility. He also had morphologically normal spermatozoa for 5% of the count, over three independent assays.

Following autozygome-based analysis (Supplementary Material, Fig. S1), we identified homozygous mutations in three additional patients from the Netherlands (Fig. 1, F3-5). All of them were diagnosed with CD, with full-field ERGs showing reduction of cone function but preserved rod responses (Table 1). Similar to the cases described above, the female patient P3 had a *TTLL5* truncating mutation, c.2132\_2133insGATA;p.(Met712IleAspfs\*15), resulting in a premature termination codon. The other two patients (P4, and P5) were males and both carried homozygous missense mutations, namely c.1627G > A;p.(Glu543Lys) and c.2266A > T;p.(Ile756Phe),

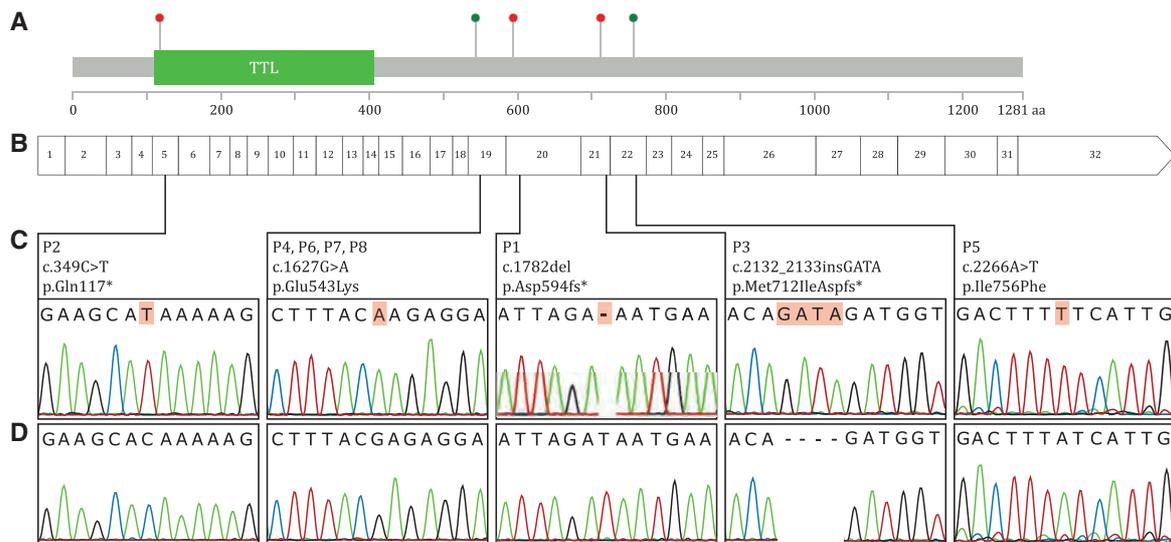


**Figure 2.** Fundi of index patient P1. Autofluorescence imaging of the right (A) and left (B) eye at age 66 years, showing a distinctive pattern of abnormalities, including a hyperfluorescent ring and hyperfluorescent area at the fovea surrounded by patchy hypofluorescence in both eyes. Hypofluorescence around the optic nerve was also present. The same images, obtained 6 years later (C,D), showed an increase of hypofluorescent areas within the ring and a mild enlargement of the hyperfluorescent ring in both eyes. Composite pictures of the fundi at age 72, showing atrophic areas around the fovea and around the optic nerve (E,F). Peripheral retina was within normal limits.

respectively (Fig. 3). Following colour vision testing, substantial mistakes were made in all three colour axes. All patients had a myopic refractive error, but none of them showed any additional extraocular symptoms. In particular, they did not report any fertility problems, and both male patients had offspring.

Finally, a male patient (P6) of Pakistani origin was screened by whole-exome sequencing (WES). He was a member of a consanguineous pedigree that included six additional individuals

with high myopia and an acquired CD or CRD with loss of corrected visual acuity from the second decade onwards (Fig. 1, F6). In this case as well, a homozygous mutation in *TTL5* was identified within an autozygous region in chromosome 14 (Supplementary Material, Fig. S1). It was the same missense detected above [c.1627G > A;p.(Glu543Lys), Fig. 3], which perfectly co-segregated with the disease in the three affected and two unaffected individuals for which DNA samples were available.



**Figure 3.** Mutation diagram of the TTL5 protein (A), and corresponding cDNA. Red and green dots indicate truncating and missense mutations, respectively. Exons are numbered and drawn to scale (B) with respect to the protein sequence. The TTL domain responsible for polyglutamylation activity is indicated. Chromatograms of the mutations identified (C), compared to their relevant wild type sequences (D).

Of note, in addition to retinal dystrophy, all patients from this pedigree reported high to very-high myopia (-5 to -22 diopters) (Table 1). None of the patients reported fertility problems, and indeed patient P6 had five children.

#### TTL5 RNA isoforms and expression in different tissues

According to publicly-available databases, *TTL5* produces six protein-coding and alternatively spliced isoforms (transcripts 001, 002, 003, 016, 017, 018 of the Ensembl database GRCh37, release 84), presenting a rather widespread expression throughout different tissues and organs (UniGene). To gain insights into this topic, we investigated the composition and abundance of *TTL5* transcripts in a panel of cadaveric organs and tissues (Fig. 4).

Our data confirmed that *TTL5* has an extremely variable expression pattern, both in terms of isoforms and of presence in various tissues. However, qualitative and quantitative assessment of all transcripts revealed that expression of the canonical isoform 001 was overwhelmingly more abundant (more than 40-fold higher than the average of the remainder) and that expression in the retina and testis represented ~64% of *TTL5* presence across all tissues and organs examined (31 and 33% in testis and retina, respectively; Fig. 4). Interestingly, all mutations identified in our cohort of patients affected isoform 001, and individuals with inactivating mutations showed reduced fertility (Table 1).

#### TTL5 mRNA level in the index patient P1

Since mutations leading to premature termination codons trigger nonsense-mediated mRNA decay (NMD) and result in no or in short-lived transcripts (18), we analysed expression of *TTL5* in skin fibroblasts from patient P1, displaying retinal degeneration and infertility. Quantitative real time PCR (q-PCR) resulted in a dramatically reduced detection of the transcript of interest as compared to a healthy control fibroblast mRNA (~10%; Fig. 4).

#### TTL5 protein in ciliated fibroblasts, retina, and spermatozoa

To better understand the role of *TTL5* with respect to the cellular cilium, we analysed control fibroblasts following serum starvation, a procedure that induces ciliogenesis. Immunofluorescence analysis revealed the clear localization of *TTL5* at both centrioles (Fig. 5, panels A–D).

Subsequently, we performed immunofluorescence analyses in the retina and sperm cells from rat and human, respectively (Fig. 5, panels E–G). In agreement with previous results in mouse and human (5), the anti-*TTL5* antibody decorated the inner segment of photoreceptors in proximity of the basal body and the connecting cilium, suggesting that *TTL5* may in fact play its functions at the base of the photoreceptor primary cilium.

Staining of mature human spermatozoa also indicated for the first time a clear centrosomal localization of *TTL5*, with no overlap with the polyglutamylated  $\alpha$ - and  $\beta$ -tubulin of the flagellum.

#### Discussion

Both primary (or immotile) and motile cilia play crucial roles in the normal function of most tissues of the human body. These tiny hair-like organelles participate in a wide range of cellular functions during development, tissue morphogenesis and homeostasis. It is therefore not surprising that mutations in ciliary genes are often associated with a broad range of conditions, classified as ciliopathies, either involving single organs or causing syndromic phenotypes (8). Some examples of diseases affecting primary cilia are polycystic kidney disease, Usher syndrome, retinitis pigmentosa, Bardet-Biedl and Joubert syndromes (19–22). On the other hand, motile cilia defects have been shown to be causative for Kartagener syndrome and allied diseases, collectively grouped under the disease spectrum of primary ciliary dyskinesias.

Cilia and flagella, highly conserved in their core structure, are ancestral organelles composed of more than 650 proteins (10,23). The building units of both ciliary and flagellar

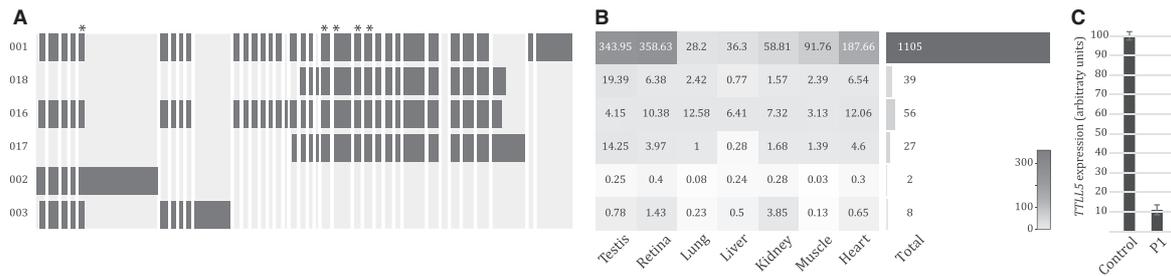
Table 1. Patients with TLL5 mutations and clinical features

Family	Patient	TLL5 protein change	Zygoty	Sex	Age at last examination	Visual acuity	Correction	Full field ERG at last examination			Macula	Periphery	Other features
								cone response	rod response	response			
F1	P1	c.1782del; p.Asp594Gluufs*29	hom	M	75	0.05	-3.6	absent residual	reduced residual	atrophy	normal	azoospermia	
	P2	c.349C>T; p.Gln117*	hom	M	46	0.05	-10.00	absent residual	reduced residual	atrophy	minor changes	reduced motility of sperm; normal anterior eye segment	
F3	P3	c.2132_2133insGATA; p.Met712IleAspfs*15	hom	F	58	0.16	-3	reduced	normal		normal		
F4	P4	c.1627G>A; p.Glu543Lys	hom	M	61	0.03	-5	absent	normal	atrophy	normal		
	P5	c.2266A>T; p.Ile756Phe	hom	M	38	0.33	-8	absent	normal	pigmentary changes	normal		
F6	P6	c.1627G>A; p.Glu543Lys	hom	M	53	NA	-8.00	NA	NA	atrophy	pigmentary changes	phthical right eye	
F7	P7	c.1627G>A; p.Glu543Lys	hom	M	38	NA	-5.00	NA	NA	NA	NA	posterior subcapsular cataract	
F8	P8	c.1627G>A; p.Glu543Lys	hom	M	18	NA	-22.00	NA	NA	NA	NA	fairly normal retina	

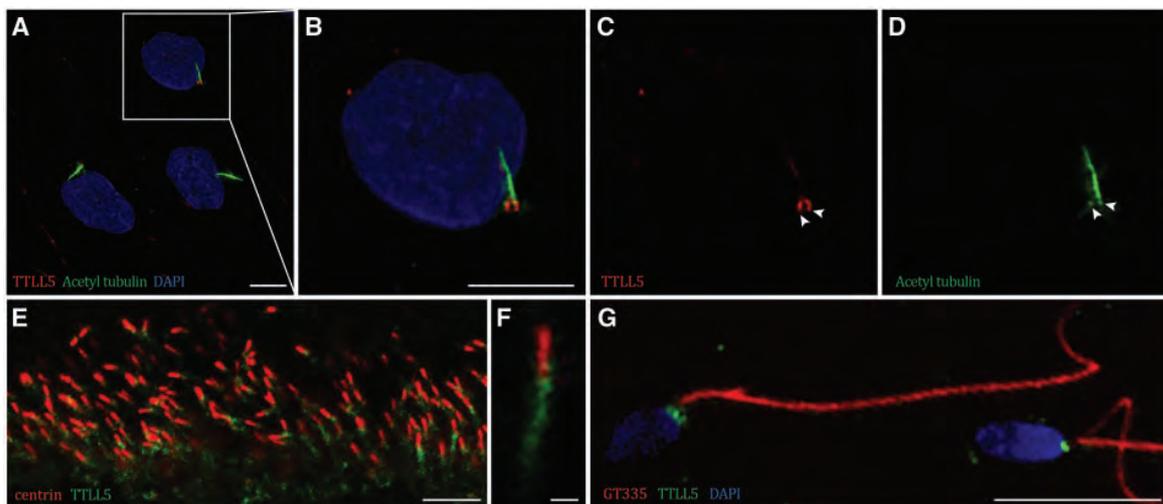
microtubules,  $\alpha$ - and  $\beta$ -tubulin, are subject to post-translational modifications, accomplished by enzymes catalyzing different reactions such as the generation of  $\Delta 2$ -tubulin, acetylation (24), tyrosination (25), polyglutamylation (26) and polyglycylation (27,28). Among members of the TLL superfamily there are glutamylases and glycolases (29–31). TLL5 initiates the formation of side chains within the C-terminal tail of  $\alpha$ - and  $\beta$ -tubulin, with a preference for  $\alpha$ -tubulin (32), and current models indicate that the role of polyglutamylation is to provide the necessary conditions for proper MT-MAPs (microtubule and microtubule-associated proteins) interactions. Studies have shown that polyglutamylation exerts differential regulation by selectively recruiting different MAPs: MAP1B, MAP2, tau, and neuronal kinesins have higher affinity for tubulins with 1-3 glutamyl units, whereas MAP1A has higher affinity for longer side chains (25,26,28). Moreover, it has been shown that masking polyglutamylated sites with a specific anti-polyglutamylated tubulin antibody (GT335) affects the amplitude of flagellar beating in sea urchin sperm axonemes, suggesting a key role of polyglutamylated sites for interaction with ciliary dyneins (33). Centriole stability was also shown to be influenced by the degree of polyglutamylation, and GT335 antibody-loaded HeLa cells showed a complete transient disappearance of the centriole pair (13). Finally, members of the TLL family, Tll3 and Tll6, play a role in cilia structure and motility in zebrafish (34). All TLL proteins have a preference for either  $\alpha$ - or  $\beta$ -tubulin and participate to either initiation or elongation of the polyglutamyl side chain. TLL5, together with TLL4 and TLL7, initiates polyglutamylation, while other members function in the elongation of the polyglutamyl side chain or in the initiation or the elongation of polyglycylation.

Specific patterns of modifications on microtubules might be responsible for various functions. In the case of polyglutamylation, the side chains are built within the carboxy-terminal tail of tubulin, where the binding sites of motor and MT-associated proteins (MAPs) are also found. Thus, it is plausible that the interaction of MTs with such proteins might depend on specific patterns of modifications (35). Additional studies highlighted the importance of polyglutamylation for the proper beating of airway cilia (36), as well as for providing a molecular traffic sign required by motor proteins in order to maintain continuous synaptic transmission (37). Major evidence of the implication of polyglutamylation in photoreceptor ciliary function was recently reported (7), showing that *Tll5*<sup>-/-</sup> mice developed a similar retinal phenotype to *Rpgr*<sup>-/-</sup> mice, a known mouse model for retinitis pigmentosa. In addition, *Tll5*<sup>-/-</sup> mice display strongly reduced glutamylation of RPGR<sup>ORF15</sup>, a retina-specific variant of RPGR (38). Altogether, current evidence strongly supports the notion that the presence and length of polyglutamyl side chains, not only on tubulin but also on other substrates, is crucial for proper functioning of both motile and immotile cilia.

In our work we show that mutations in a gene involved in the polyglutamylation of  $\alpha$ -tubulin is associated with defects in the retina and spermatozoa. Clinically, these molecular phenotypes translate into cone-first CRD and reduced sperm motility, likely due to the functional impairment of the primary cilium and the flagellum, respectively. Our assumption is supported by immunofluorescence data, demonstrating that TLL5 localizes at the basal body of the cilia in photoreceptors, as well as at the base of the spermatozoal axoneme and in ciliated skin fibroblasts. Moreover, we reveal that the highest levels of expression of the major TLL5 protein-coding isoform is in the retina and testis. It is also very interesting to note that, in terms of fertility and TLL5 pathogenic variants, there is an apparent



**Figure 4.** Relative expression of *TLL5* isoforms and their expression. *TLL5* has six known alternative transcripts (001, 002, 003, 016, 017, and 018), resulting from the splicing of the exons indicated here in dark grey (A). Their expression within seven different human tissues, measured by quantitative real time PCR, is indicated by both numerical values and shades of grey (B). Although all isoforms seem to be widely expressed, isoform 001 is the most prominent one, among all tissues considered ('Total' column). Asterisks show the position of the mutations identified in this work. *TLL5* isoform 001 expression in fibroblasts of the index patient P1 vs. a control, by quantitative PCR (C). *TLL5* mRNA amounts were normalized with respect to the housekeeping gene *HPRT1*.



**Figure 5.** Immunofluorescence staining of ciliated control human skin fibroblasts (A-D). *TLL5* co-localizes with acetylated tubulin at the centrioles, as indicated by arrowheads (C,D). Immunofluorescence in rat retina sections (E, and magnified cilium in F) and in control human spermatozoon (G). *TLL5* decorates the basal body in photoreceptors and the centrioles in spermatozoa. Scale bars: A-E, 5  $\mu$ m; F, 1  $\mu$ m; G, 20  $\mu$ m.

genotype/phenotype correlation, which seems to be irrelevant for retinal degeneration. In other words, the phenotype elicited by *TLL5* pathogenic changes appears to depend on mutation classes. Missense variants are seemingly associated with a non-syndromic phenotype that is limited to the retina, whereas inactivating mutations appear to disrupt the functions of both photoreceptors and spermatozoa, thus defining a novel allele-specific syndrome. Yet, three male patients with truncating *TLL5* mutations were previously reported having offspring (5), raising the possibility of variable expressivity or reduced penetrance of this class of mutations. In support of the latter hypothesis, *Tll5* knockout mice display extremely reduced but not completely abolished fertility (6). Alternatively, this genotype/phenotype correlation in our cohort could also be coincidental. The association of *TLL5* mutations with severe myopia is another intriguing hypothesis that warrants additional investigation in a larger cohort of patients, especially given the complex pattern of inheritance of nearsightedness in humans.

In conclusion, we show that mutations in *TLL5* are associated with a newly-defined syndrome affecting vision and the male reproductive system. Despite the fact that cilia and flagella have different morphologies and functions, they may share

similar physiological mechanisms, and the enzymatic reaction of polyglutamylation performed by *TLL5* may be one of these common elements.

## Materials and Methods

### Patients and controls

Patient P1 was recruited from the Jules Gonin Ophthalmic Hospital (Lausanne, Switzerland); and patient P2 from the Department of Ophthalmology of Lund University Hospital (Lund, Sweden). Patients P3, P4 and P5 were recruited from the Radboud University Medical Center (Nijmegen, The Netherlands) and the Erasmus University Medical Center (Rotterdam, The Netherlands). Patients P6-8 were sampled by author MMK, an ophthalmologist based at St James's University Hospital (Leeds, England), while on a field trip to Pakistan. DNA of all subjects was extracted from peripheral blood leukocytes. A control sperm sample was provided by a healthy donor. Our research has been conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Boards of our respective Organizations.

## Clinical evaluation

For patients P1-5, ophthalmologic examination included assessment of BCVA, slit-lamp examination, funduscopy, fundus photography, and optical coherence tomography. Full-field ERGs were also recorded, as prescribed by the International Society for Clinical Electrophysiology of Vision (ISCEV).

For patients P6-8, ophthalmologic assessment was limited to slit-lamp examination and fundus inspection using direct and indirect ophthalmoscopy through dilated pupils. Owing to the non-hospital setting, electrodiagnostic and other testing was not available. Visual acuity was recorded together with a history of nyctalopia or photoaversion.

Semen analysis was carried out for patients P1 and P2 by standard procedures of andrology laboratories, and according to WHO guidelines (39). For P1 patient, two Leja chambers (Leja) filled with 6  $\mu$ l of semen were entirely scanned under phase contrast microscopy to confirm the absence of spermatozoa. Additional clinical features were assessed only for patient P1, due to substantial problems in getting back to the other probands. These tests included a thorough otorhinolaryngologic examination assessing structure and function of the nasopharyngeal mucosa, the ear canal and the hearing (by pure-tone audiometry), as well as a full renal examination, including a complete checkup of renal function, urine and blood analysis, assessment of blood pressure, and ultrasonography. Analyses involving other tissues and organs known to be involved in other ciliopathies were not performed due to a negative clinical history: normal body mass index, no respiratory complaints, no metabolic disturbance and no skeletal abnormalities.

## Whole-genome and whole-exome sequencing

WGS in the Swiss index patient P1 was performed using 4  $\mu$ g of DNA. Sequencing was performed by Complete Genomics Inc. (Mountain View, CA, USA), as described previously (40). Genetic variants were identified using v2.0 of the Complete Genomics pipeline (41). WES was performed for proband P6 using 3  $\mu$ g of DNA. Protein-coding regions were captured using the SureSelect All Exon v4 kit (Agilent) and paired-end sequencing was performed using the Illumina HiSeq 2500 platform. Single nucleotide variants and small insertions or deletions were detected using the Genome Analysis Tool Kit (GATK v2.4.7) software package, using the Best Practice Guidelines identified by the developers (42). The pathogenicity of genetic variants detected through WGS and WES were assessed after functional annotation through ANNOVAR (43).

## Homozygosity mapping

Genomic regions with high homozygosity were determined using the free web-based tool HomozygosityMapper (44).

## Mutation screening

Primer pairs for *TLL5* exons and flanking intron boundaries were designed using the CLCbio Genomics Workbench (Qiagen, Supplementary Material, Table S2). PCR amplification was performed in a 20  $\mu$ l total volume containing 10 ng genomic DNA, 1x GoTaq buffer, 0.1 mM dNTPs, 10  $\mu$ M of each primer, and 5 U/ $\mu$ l of GoTag polymerase (Promega). PCR products were purified (ExoSAP-IT, USB) and a sequencing reaction was performed in a total volume of 5  $\mu$ l using 1  $\mu$ l primer 3.3  $\mu$ M, 0.5  $\mu$ l BigDye

Terminator v1.1, and 1  $\mu$ l of the provided Buffer (Applied Biosystems).

## Antibodies

Commercial goat polyclonal anti-TLL5 antibody (Santa Cruz Biotechnology Inc), raised against a peptide mapping near the C-terminus of TLL5 human origin, was used at a 1:100 dilution. Mouse monoclonal anti-centrin clone 20H5 antibody was purchased from Millipore and used at a dilution of 1:1000. Anti-polyglutamylated tubulin GT335 and monoclonal anti-acetylated tubulin antibodies were provided by Dr C. Janke (Institut Curie, Orsay, France) and purchased from Sigma-Aldrich, respectively. Secondary donkey anti-goat antibodies conjugated with Alexa Fluor 488 were purchased from Invitrogen and secondary goat anti-mouse antibodies (Life Technologies) were conjugated with Alexa Fluor 594 (1:1000).

## Fibroblast immortalization and culture

Primary skin fibroblasts were immortalized with exogenous hTERT by the use of pLOX-TERT-iresTK (45) and grown in DMEM(1x) + 1g/L D-glucose L-Glutamine Pyruvate (Gibco), supplemented with 10% FBS, 1% penicillin-streptomycin, and 1% fungizone, adapted from previously published protocols (46).

## Immunofluorescence

Immunofluorescence was performed for TLL5 localization in human control spermatozoa and rat retinal sections. The sperm staining procedure used was adapted from a previously published protocol (47). After washing five times with PBS 1X in a 15 ml Falcon tube and centrifugation steps of 5' at 800g, all at room temperature (RT), the semen pellet was resuspended and fixed with 4% (v/v) PFA in PBS, and incubated for 30 min on ice. The sample was then washed three times with PBS and stored at 4 °C for future use. The immunostaining procedure used 20  $\mu$ l aliquots of fixed sperm cells, transferred to a 1.5  $\mu$ l Eppendorf tube. Blocking was done for 30 min at RT in 100  $\mu$ l PBS containing 3% (w/v) BSA (PBSA). Primary antibody incubation was performed with specified dilutions in 100  $\mu$ l PBSA, overnight at 4 °C. Samples were washed three times with PBS containing 0.1% (v/v) TX100 (PBST). Secondary antibody incubation for 30 min was carried out at RT in 100  $\mu$ l PBSAT [PBST containing 3% (w/v) BSA and 0.1% (v/v) TX100]. After three washes with PBST and two washes with PBS the final pellet was resuspended in PBS and 5  $\mu$ l were placed on a slide. 5  $\mu$ l DAPI vectashield were added to the sample, which was then coverslipped and fixed with nail polish.

Unfixed Sprague-Dawley rat eyes and C57BL/6J mouse eyes were isolated and soaked for 3 h in PBS containing 30% sucrose. Eyes were embedded in Yazulla medium (30% egg albumen and 3% gelatin in water) and cryosectioned (12  $\mu$ m) onto Superfrost Plus slides (Thermo Scientific). Sections were washed three times with PBS and a stepwise procedure was followed similar to that described for sperm immunostaining, the only difference being that this was carried out on the slide and with lower volumes.

## Quantitative real time PCR

Primer pairs used for q-PCR are listed in Supplementary Material, Table S3. The q-PCR product was visualized on 1%

agarose gel to verify the primer's specificity. A standard curve using a control cDNA template prepared from human normal tissues total RNA (BioChain) was used to test the efficiency of each primer pair. HPRT1 was used as a normalization control, as described (48). Amplification was performed using the SYBR Green PCR Master Mix (Applied Biosystems).

## Supplementary Material

Supplementary Material is available at HMG online.

## Acknowledgements

We would like to thank the patients and their families who participated in this study.

Conflict of Interest statement. None declared.

## Funding

This work was supported by grants from the Swiss National Science Foundation to CR [grant numbers 138346, 156260], the Foundation Fighting Blindness USA to FPMC [grant number C-GE-0811-0545-RAD01], and RP Fighting Blindness and Fight For Sight [RP Genome Project GR586]. MEE was funded by an Egyptian Government Scholarship. Family F6 was analysed by the UK Inherited Retinal Disease Consortium.

## References

- Hamel, C.P. (2007) Cone rod dystrophies. *Orphanet J. Rare Dis.*, **2**, 7.
- Berson, E.L., Gouras, P. and Gunkel, R.D. (1968) Progressive cone-rod degeneration. *Arch. Ophthalmol.*, **80**, 68–76.
- Berson, E.L., Gouras, P. and Gunkel, R.D. (1968) Progressive cone degeneration, dominantly inherited. *Arch. Ophthalmol.*, **80**, 77–83.
- Roosing, S., Thiadens, A.A., Hoyng, C.B., Klaver, C.C., den Hollander, A.I. and Cremers, F.P. (2014) Causes and consequences of inherited cone disorders. *Prog. Retin. Eye Res.*, **42**, 1–26.
- Sergouniotis, P.I., Chakarova, C., Murphy, C., Becker, M., Lenassi, E., Arno, G., Lek, M., MacArthur, D.G., Consortium, U.C.E., Bhattacharya, S.S., et al. (2014) Biallelic variants in TLL5, encoding a tubulin glutamylase, cause retinal dystrophy. *Am. J. Hum. Genet.*, **94**, 760–769.
- Lee, G.S., He, Y., Dougherty, E.J., Jimenez-Movilla, M., Avella, M., Grullon, S., Sharlin, D.S., Guo, C., Blackford, J.A., Jr., Awasthi, S., et al. (2013) Disruption of Tll5/stamp gene (tubulin tyrosine ligase-like protein 5/SRC-1 and TIF2-associated modulatory protein gene) in male mice causes sperm malformation and infertility. *J. Biol. Chem.*, **288**, 15167–15180.
- Sun, X., Park, J.H., Gumerson, J., Wu, Z., Swaroop, A., Qian, H., Roll-Mecak, A. and Li, T. (2016) Loss of RPGR glutamylation underlies the pathogenic mechanism of retinal dystrophy caused by TLL5 mutations. *Proc. Natl Acad. Sci. U. S. A.*, **113**, E2925–E2934.
- Hildebrandt, F., Benzing, T. and Katsanis, N. (2011) Ciliopathies. *N. Engl. J. Med.*, **364**, 1533–1543.
- Waters, A.M. and Beales, P.L. (2011) Ciliopathies: an expanding disease spectrum. *Pediatr. Nephrol.*, **26**, 1039–1056.
- Fliegauf, M., Benzing, T. and Omran, H. (2007) When cilia go bad: cilia defects and ciliopathies. *Nat. Rev. Mol. Cell Biol.*, **8**, 880–893.
- Hildebrandt, F., Attanasio, M. and Otto, E. (2009) Nephronophthisis: disease mechanisms of a ciliopathy. *J. Am. Soc. Nephrol.*, **20**, 23–35.
- Schorderet, D.F., Bernasconi, M., Tiab, L., Favez, T. and Escher, P. (2014) IROme, a new high-throughput molecular tool for the diagnosis of inherited retinal dystrophies—a price comparison with Sanger sequencing. *Adv. Exp. Med. Biol.*, **801**, 171–176.
- Bobinnec, Y., Khodjakov, A., Mir, L.M., Rieder, C.L., Edde, B. and Bornens, M. (1998) Centriole disassembly in vivo and its effect on centrosome structure and function in vertebrate cells. *J. Cell Biol.*, **143**, 1575–1589.
- Firmann, M., Mayor, V., Vidal, P.M., Bochud, M., Pecoud, A., Hayoz, D., Paccaud, F., Preisig, M., Song, K.S., Yuan, X., et al. (2008) The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.*, **8**, 6.
- Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B., et al. (2015) Analysis of protein-coding genetic variation in 60,706 humans. *Nature.*, **536**, 285–291.
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
- KDIGO. (2013) Clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl.*, **3**, 150.
- Hentze, M.W. and Kulozik, A.E. (1999) A perfect message: RNA surveillance and nonsense-mediated decay. *Cell*, **96**, 307–310.
- Badano, J.L., Mitsuma, N., Beales, P.L. and Katsanis, N. (2006) The ciliopathies: an emerging class of human genetic disorders. *Annu. Rev. Genomics Hum. Genet.*, **7**, 125–148.
- Zaghloul, N.A. and Katsanis, N. (2009) Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy. *J. Clin. Invest.*, **119**, 428–437.
- Forsythe, E. and Beales, P.L. (2013) Bardet-Biedl syndrome. *Eur. J. Hum. Genet.*, **21**, 8–13.
- Falk, N., Losl, M., Schroder, N. and Giessel, A. (2015) Specialized Cilia in Mammalian Sensory Systems. *Cells*, **4**, 500–519.
- Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J.B. and Bettencourt-Dias, M. (2011) Evolution: Tracing the origins of centrioles, cilia, and flagella. *J. Cell Biol.*, **194**, 165–175.
- Piperno, G. and Fuller, M.T. (1985) Monoclonal antibodies specific for an acetylated form of alpha-tubulin recognize the antigen in cilia and flagella from a variety of organisms. *J. Cell Biol.*, **101**, 2085–2094.
- Gundersen, G.G. and Bulinski, J.C. (1986) Distribution of tyrosinated and nontyrosinated alpha-tubulin during mitosis. *J. Cell Biol.*, **102**, 1118–1126.
- Edde, B., Rossier, J., Le Caer, J.P., Desbruyeres, E., Gros, F. and Denoulet, P. (1990) Posttranslational glutamylation of alpha-tubulin. *Science*, **247**, 83–85.
- Westermann, S. and Weber, K. (2003) Post-translational modifications regulate microtubule function. *Nat. Rev. Mol. Cell Biol.*, **4**, 938–947.
- Janke, C. and Kneussel, M. (2010) Tubulin post-translational modifications: encoding functions on the neuronal microtubule cytoskeleton. *Trends Neurosci.*, **33**, 362–372.

29. Boucher, D., Larcher, J.C., Gros, F. and Denoulet, P. (1994) Polyglutamylation of tubulin as a progressive regulator of in vitro interactions between the microtubule-associated protein Tau and tubulin. *Biochemistry*, **33**, 12471–12477.
30. Larcher, J.C., Boucher, D., Lazereg, S., Gros, F. and Denoulet, P. (1996) Interaction of kinesin motor domains with alpha- and beta-tubulin subunits at a tau-independent binding site. Regulation by polyglutamylation. *J. Biol. Chem.*, **271**, 22117–22124.
31. Bonnet, C., Boucher, D., Lazereg, S., Pedrotti, B., Islam, K., Denoulet, P. and Larcher, J.C. (2001) Differential binding regulation of microtubule-associated proteins MAP1A, MAP1B, and MAP2 by tubulin polyglutamylation. *J. Biol. Chem.*, **276**, 12839–12848.
32. Janke, C., Rogowski, K., Wloga, D., Regnard, C., Kajava, A.V., Strub, J.M., Temurak, N., van Dijk, J., Boucher, D., van Dorsselaer, A., et al. (2005) Tubulin polyglutamylase enzymes are members of the TTL domain protein family. *Science*, **308**, 1758–1762.
33. Gagnon, C., White, D., Cosson, J., Huitorel, P., Edde, B., Desbruyeres, E., Paturle-Lafanechere, L., Multigner, L., Job, D. and Cibert, C. (1996) The polyglutamylated lateral chain of alpha-tubulin plays a key role in flagellar motility. *J. Cell Sci.*, **109**, 1545–1553.
34. Pathak, N., Austin, C.A. and Drummond, I.A. (2011) Tubulin tyrosine ligase-like genes *tll3* and *tll6* maintain zebrafish cilia structure and motility. *J. Biol. Chem.*, **286**, 11685–11695.
35. Janke, C., Rogowski, K. and van Dijk, J. (2008) Polyglutamylation: a fine-regulator of protein function? 'Protein Modifications: beyond the usual suspects' review series. *EMBO Rep.*, **9**, 636–641.
36. Ikegami, K., Sato, S., Nakamura, K., Ostrowski, L.E. and Setou, M. (2010) Tubulin polyglutamylation is essential for airway ciliary function through the regulation of beating asymmetry. *Proc. Natl Acad. Sci. U. S. A.*, **107**, 10490–10495.
37. Ikegami, K., Heier, R.L., Taruishi, M., Takagi, H., Mukai, M., Shimma, S., Taira, S., Hatanaka, K., Morone, N., Yao, I., et al. (2007) Loss of alpha-tubulin polyglutamylation in ROSA22 mice is associated with abnormal targeting of KIF1A and modulated synaptic function. *Proc. Natl Acad. Sci. U. S. A.*, **104**, 3213–3218.
38. Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Miano, M.G., Meindl, A., Meitinger, T., Ciccodicola, A. and Wright, A.F. (2000) Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat. Genet.*, **25**, 462–466.
39. WHO. (2010) WHO laboratory manual for the examination and processing of human semen.
40. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, **327**, 78–81.
41. Carnevali, P., Baccash, J., Halpern, A.L., Nazarenko, I., Nilsen, G.B., Pant, K.P., Ebert, J.C., Brownley, A., Morenzoni, M., Karpinchyk, V., et al. (2012) Computational techniques for human genome resequencing using mated gapped reads. *J. Comput. Biol.*, **19**, 279–292.
42. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, **43**, 491–498.
43. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
44. Seelow, D., Schuelke, M., Hildebrandt, F. and Nurnberg, P. (2009) HomozygosityMapper—an interactive approach to homozygosity mapping. *Nucleic Acids Res.*, **37**, W593–W599.
45. Salmon, P., Oberholzer, J., Occhiodoro, T., Morel, P., Lou, J. and Trono, D. (2000) Reversible immortalization of human primary cells by lentivector-mediated transfer of specific genes. *Mol. Ther.*, **2**, 404–414.
46. van Karnebeek, C.D., Bonafe, L., Wen, X.Y., Tarailo-Graovac, M., Balzano, S., Royer-Bertrand, B., Ashikov, A., Garavelli, L., Mammi, I., Turolla, L., et al. (2016) NANS-mediated synthesis of sialic acid is required for brain and skeletal development. *Nat. Genet.*, **48**, 777–784.
47. Nishimura, H., Gupta, S., Myles, D.G. and Primakoff, P. (2011) Characterization of mouse sperm TMEM190, a small transmembrane protein with the trefoil domain: evidence for colocalization with IZUMO1 and complex formation with other sperm proteins. *Reproduction*, **141**, 437–451.
48. Hoggart, C.J., Venturini, G., Mangino, M., Gomez, F., Ascari, G., Zhao, J.H., Teumer, A., Winkler, T.W., Tsernikova, N., Luan, J., et al. (2014) Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index. *PLoS Genet.*, **10**, e1004508.

## **PROJECT 2 - ARTICLE**

### ***ARL2BP* and primary ciliary dyskinesia: a mouse and human parallel study**

This manuscript is presently being finalized for submission in the *Journal of Clinical Investigation*.

#### **Contributions:**

I gained co-first authorship in this project with the characterization of the variants in human patients. I first linked *ARL2BP* to infertility in human. Thanks to the collaboration with Ms Abigail Moye we could reinforce this association by the use of a mouse model. My contribution in this paper is for all experiments related to human, except for the ophthalmological examination and the spermiogram. Also writing of the article was mainly done by Ms Abigail and myself, with the precious contribution of our respective thesis directors.

## **Title**

“Mutations in ARL2BP, a protein required for ciliary microtubule structure, cause syndromic infertility in human and mouse.”

## **Author List**

Nicola Bedoni<sup>1,\*</sup>, Abigail R. Moye<sup>2,3,\*</sup>, Jessica Cunningham<sup>3</sup>, Eric Tucker<sup>3</sup>, Peter Mathers<sup>3,4</sup>, Virginie Peter<sup>1</sup>, Liliana P. Paris<sup>5</sup>, Luísa Coutinho-Santos<sup>5</sup>, Peter Camacho<sup>5</sup>, Madeleine G. Purcell<sup>6</sup>, Abbie C. Winkelmann<sup>6</sup>, James A. Foster<sup>6</sup>, Carlo Rivolta<sup>2,7,\*</sup>, and Visvanathan Ramamurthy<sup>2,3,4,\*</sup>

## **Affiliations**

<sup>1</sup>Department of Computational Biology, University of Lausanne, Lausanne, Switzerland.

<sup>2</sup>Departments of Ophthalmology, <sup>3</sup>Department of Biochemistry, and <sup>4</sup>Rockefeller Neurosciences Institute, West Virginia University, Morgantown, WA 26506, USA. <sup>5</sup>Instituto de Oftalmologia Dr Gama Pinto, Lisbon, Portugal. <sup>6</sup>Biology Department, Randolph-Macon College, Ashland, VA 23005. <sup>7</sup>Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

## **Correspondence Emails**

Visvanathan Ramamurthy, Departments of Ophthalmology and Biochemistry; Eye Institute, One Medical Center Drive, West Virginia University, Morgantown, WV 26506-9193, USA. Phone: 304-598-6925; Email: [ramamurthyv@wvumedicine.org](mailto:ramamurthyv@wvumedicine.org)

Carlo Rivolta, Department of Computational Biology, University of Lausanne, Rue du Bugnon 27, 1011 Lausanne, Switzerland. Phone: +4121692 5451; Email: [carlo.rivolta@unil.ch](mailto:carlo.rivolta@unil.ch)

## **Additional Footnotes**

\*Equal contribution

## **Conflict of interest statement**

The authors have declared that no conflict of interest exists.

**Abstract**

Cilia are evolutionarily conserved hair-like cellular extensions with a wide spectrum of key biological roles. Studies have shown that cilia are not only involved in fluid movement, cell locomotion, and sexual reproduction, but also in sensation of the extracellular environment and signalling, directing embryonic development and organ function. Ciliary and axonemal dysfunction have been linked to the growing class of genetic disorders collectively known as ciliopathies. Many strides have been made towards deciphering the causes behind this group of diseases, which have in turn increased the knowledge base on cilia and cilia dysfunctions. One recently identified ciliopathy-associated gene is the ADP-Ribosylation Factor Like 2 Binding Partner (*ARL2BP*). In the present study, we report a novel disruptive variant in *ARL2BP* and we show that disruptive mutations in this gene, result not only in retinal degeneration but also in spermatogenesis defects, displayed by abnormally shaped sperm heads, shortened flagella, and mis-assembled flagellar structures. Moreover, we corroborate the human phenotypes in the mouse model, which additionally displays high incidence of situs inversus, suggesting a role of *ARL2BP* also in embryological nodal cilia. Finally, we propose *ARL2BP*-associated disease entities as an atypical primary ciliary dyskinesia with additional features of retinal degeneration.

## Introduction

Cilia are short, protruding organelles often referred to as “signaling hubs” of cells and are involved in diverse functional roles, including sensation of our environment (hair and photoreceptor cells), fluid movement (trachea, brain, and the embryonic node), and in signaling pathways (e.g. the Sonic Hedgehog cascade) (1-5). Single-gene mutations causing defects in proteins structurally and functionally present in cilia are often associated with a broad range of pleiotropic disorders classified as ciliopathies, involving single or multiple tissues or organs (6). Cilia have in fact a very important role for a broad spectrum of functions in the human body, as witnessed by the large body of literature reporting on ciliary dysfunctions and their associated diseases, such as polycystic kidney disease, hepatic fibrosis, retinal dystrophies, sensorineural hearing loss, skeletal dysplasia, endocrinopathies, neurodevelopmental defects, obesity, anomalies of the central nervous system, laterality defects, congenital heart disease, respiratory diseases) (6-10). Primary cilia dyskinesia (PCD) refer to conditions characterized by loss of motility of cilia and flagella. In particular, PCD is linked to defects in the cilia lining the trachea, the embryonic node, and the sperm flagellum, resulting in chronic respiratory infections, reversal of the internal organs (situs inversus), and male infertility, respectively. However, some PCD patients display additional symptoms, such as female infertility, chronic ear infections (otitis media), blinding diseases, and hydrocephaly (11).

Retinitis pigmentosa (RP), a progressive blinding condition leading to loss of retinal photoreceptor cells, is an example of a condition manifesting often as a result of a ciliopathy. Examples of ciliopathies leading to ciliary photoreceptor damage are the Joubert syndrome, Bardet-Biedl syndrome, and Senior-Loken syndrome. Moreover, a recent study has associated nonsense mutations in the polyglutamylase gene *TLL5* with cone-rod dystrophy and male infertility (12). In support of these findings, *Tll5* knockout mice, in addition to retinal degeneration, show flagella that are detached from the sperm head, disrupted axoneme patterns with loss of tubulin doublets, as well as a significantly decreased polyglutamylation in the upper and lower segments of the sperm flagellum, resulting in a severe loss of motility of sperm cells (13, 14). Similarly, defects in the Intraflagellar Transport Protein 27 (IFT27), associating with Bardet-Biedl syndrome, have also been associated with sperm malformations in mouse, with further experiments showing this protein to be required for both flagellum and photoreceptor protein trafficking (72, 73).

ARL2BP (ADP Ribosylation Factor Like GTPase 2 Binding Protein) is an effector protein of the small GTP-binding proteins ARL2, involved in the regulation of soluble tubulin dimers, and in the regulation of mitochondrial motility and fusion (17-22). In humans, defects in the *ARLBP2* gene have been linked to ciliopathic RP with or without situs inversus.

In this study we report the identification of two homozygous mutations in the gene *ARL2BP* in three Portuguese patients from two consanguineous families displaying RP and male infertility. The murine knockout models for the same gene, showed similar phenotypes, including retinal degeneration, immotile sperm cells and impaired spermatogenesis, as well as situs inversus and increased brain ventricular volume. Our data highlight a novel ciliopathic entity linking two structurally similar, yet functionally different, ciliary organelles - the eye photoreceptor and the spermatozoon, associating vision and the reproductive system.

## **Materials and Methods**

### **Patients and controls**

Patients were recruited from the Instituto de Oftalmologia Dr. Gama Pinto in Lisbon, Portugal. DNA was extracted from peripheral blood leukocytes (subjects P1 and P2), and from saliva (subject P3), following the signature of a written informed consent. RNA was extracted from a sperm sample from patient P1. A control sperm and DNA sample were provided by a healthy donor.

### **Clinical evaluation**

Ophthalmologic examination included assessment of BCVA, slit-lamp examination, dilated fundus examination, fundus photography, visual fields, and optical coherence tomography (OCT). Full-field ERGs were also recorded, following the International Society for Clinical Electrophysiology of Vision (ISCEV) protocol. Semen analysis (subject P1) was carried out by standard procedures by an andrology laboratory, and according to WHO guidelines (26). Additional clinical features were assessed based on patients' clinical history.

**Whole-exome sequencing (WES)**

WES was performed using 2 µg DNA derived from peripheral blood mononuclear cells. Protein-coding regions were captured using the HiSeq Rapid PE Cluster Kit v2, and an Illumina HiSeq 2500 instrument was used for paired-end sequencing. Single nucleotide variants and small insertions and deletions were detected using the Genome Analysis Tool Kit (GATK v4.0) software package, using the Best Practice Guidelines identified by the developers (27). The pathogenicity of the detected genetic variants was assessed after functional annotation through ANNOVAR (28) and with the mean of in-house scripts. Genomic regions with high homozygosity were determined using proprietary software.

**Sanger sequencing validation, PCR, and RT-PCR**

Primer pairs for *ARL2BP* exons and flanking intron boundaries were designed using the CLCbio Genomics Workbench (Qiagen). PCR amplification was performed in a 10 µl total volume containing 2 ng genomic DNA or cDNA, 1x GoTaq buffer, 0.1 mM dNTPs, 10 µM of each primer, and 5 U/µl of GoTaq polymerase (Promega). PCR products were purified using ExoSAP-IT, USB, or extracted from agarose gel using Nucleospin Gel and PCR Clean-up (Macherey-Nagel). Sanger sequencing was performed by a service provider (Fasteris,). Reverse transcription was performed on 1 µg RNA using a mix of random primers and the GoScript Reverse Transcription Protocol (Promega).

**Immunoblotting**

Mice were euthanized by CO<sub>2</sub> inhalation and testes were dissected and separated from epididymis and fat. Testis samples were weighed and immediately frozen in liquid nitrogen. Prior to analyses, the samples were homogenized in phosphate buffered saline (PBS, with protease inhibitor cocktail (Thermo Fisher #A32955)). Protein concentrations were measured using a NanoDrop spectrophotometer. Samples were analyzed by SDS-PAGE gel, then resolved and transferred onto polyvinylidene difluoride membranes. Membranes were blocked with western blot blocking buffer (LiCor 004864 Classic) for 30 min at room temperature and incubated with the primary antibodies overnight at 4°C. Blots were washed in PBST (PBS with 0.1% Tween-20) 3 times for 5 minutes (3 X 5 min) at room temperature and incubated in secondary antibody, goat anti-rabbit Alexa 680, goat anti-rat Alexa 680, or goat anti-mouse

Alexa 800 for 30 min at room temperature. After 3 x 5 minutes of washes with PBST, membranes were scanned using Odyssey Infrared Imaging System.

### **Immunofluorescence**

For testis, the anesthetized [??] animal was perfusion fixed with 4% PFA and then the testis was dissected out. This was followed by incubation in 30% sucrose/PBS overnight at 4°C. Afterward, testes were incubated in a 1:1 mixture of 30% sucrose in PBS and OCT (Cryo Optimal Cutting Temperature Compound, Sakura) for 1 hr, and flash frozen in OCT. Staining was performed following the same protocol as detailed below. For sperm, the animal was euthanized by CO<sub>2</sub> inhalation and the epididymis was collected. The tissue was then minced and sperm were allowed to swim out for 5-10 minutes (in the case of knockout, entire volume of liquid was collected) and collected into an Eppendorf tube. 100µl of sperm suspension was added to a Superfrost Plus slide and allowed to completely dry by placing the slides on a hot plate (setting 3, Corning Hot Plate). Cells were fixed in 4% PFA for 15 minutes followed by three 5 minute washes in PBS. Ice-cold methanol was then added for 2 minutes, followed by three 5 minute washes in PBS. The cells were then blocked overnight at 4°C (PBS with 5% goat sera, 0.5% TritonX-100, 0.05% sodium azide). The next day, the cells were incubated with primary antibody for 2 hours at RT, followed by three 5 minute washes in PBS. The cells were then incubated with secondary antibody for 1 hour, followed by three 5 minute washes in PBS. They were then mounted with ProLong Gold anti-fade reagent and viewed on a Nikon confocal microscope.

Immunofluorescence for human P1 and control sperm samples was performed as described in previously-published protocols (12, 29). The same anti-ARL2BP primary antibody described above was used in the human sperm staining experiments, in addition to a secondary goat anti-rabbit antibody conjugated with Alexa Fluor 488 (Invitrogen). Pictures were taken on a Zeiss LSM 780 confocal microscope.

### **Cell counting**

Mice were euthanized by CO<sub>2</sub> inhalation and epididymal sperm were dissected out and placed in 1ml PBS. Sperm were collected, spun at 1500 RPM for 3 minutes and re-suspended in 50µl of

PBS (knockout) or 4% paraformaldehyde (wild type; and further diluted 1:10 in PBS after 5 min). Sperm cells were counted using a hemocytometer.

### **MEF Cells**

Embryos were harvested at E13.5 and separated from each other and the placenta in PBS. The tissues were minced with sterile razor blades. The tissues were then trypsinized with 6ml total of 0.25% Trypsin/EDTA. 7ml of MEF media (catalog # 10-013-CV from Corning) DMEM containing glucose, pyruvate, and L-glut + 15%FBS, 1% Pen/Strep) was then added to each tube and spun down at 1200 RPM for 8 minutes. Cells were then added to a 100mm dish containing 25ml of MEF media and allowed to grow overnight at 37°C with 5% CO<sub>2</sub>. Cells were maintained for up to 4 passages. To induce cilia formation, cells were grown to 90% confluency and serum starved for 48 hours. Depolymerization of cilia was done by the addition of sera (after 48 hours of starvation), and cells were collected for staining after 2 hours, 6 hours, 12 hours, or 24 hours of serum addition. For immunocytochemistry, cells were fixed with 4% PFA for 15 min (or - 20°C Methanol for 2 minutes for centrosomal staining, acetylated tubulin and pericentrin), washed with PBS 3 x 5 minutes, and blocked for 30-60 minutes (PBS with 5% goat sera, 0.5% TritonX-100, 0.05% sodium azide). The remaining steps were performed in the same way as outlined in the immunofluorescence section.

### **Micro-CT Scans**

Transcardial perfusions were performed on anesthetized 60 day old mice with 4% paraformaldehyde. The brain was carefully dissected out of the skull and fixed in 4% paraformaldehyde for 2 days at 4°C. The brains were then transferred to stability buffer (4% w/v paraformaldehyde (pH 7.2), 4% w/v acrylamide, 0.05% w/v bis-acrylamide, 0.25% w/v VA044 initiator, 0.05% w/v Saponin, in 1xPBS) for 3 days at 4°C, and then underwent nitrogen desiccation, followed by a 3 hour incubation at 37°C. After a two-day staining in 0.1N iodine, the brains were embedded into 3% agarose for imaging. The brains were imaged on a Bruker SkyScan 1272 MicroCT scanner (Cu 0.11µm filter, 1500ms exposure, and 8µm resolution), and 3D reconstruction of ventricular volume was performed with Seg3D software.

### **Statistical Analysis**

All data are presented as mean  $\pm$  standard error margin. Immunoblots and ciliated cell counts were analyzed by unpaired, two-tailed *t* test ( $n=3$ ). For cilia measurements, staining was performed in triplicate with 100 cilia measured for each and data were visualized with the ggplot2 package in R version 3.3.2. Image and densitometry analysis were performed using ImageJ 1.50i. For ciliated cell counting, staining was performed in triplicate with 100 cells (and their cilia) counted for each. Mendelian ratios were analyzed using a chi square test with 1 degree of freedom. Chi square values of 3.84 or higher are statistically significant ( $p>0.05$ ).

### **Antibodies**

Information relative to antibody origin and working dilutions are presented in Supplementary Table 2.

### **Study approval**

Our research has been conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Boards of our respective Organizations. Informed consent was obtained from all patients prior to the sample collection.

## **Results**

### **Clinical evaluation**

The index subject P1 (ID: LL1) was initially evaluated at age 40 and diagnosed with retinitis pigmentosa. Born from a consanguineous union, and originally from Portugal, he was first diagnosed with myopia at age 8 and developed night blindness and photopsia by the age of 26. Over the following 10 years he developed progressive loss of vision and bilateral constriction of the peripheral visual fields. The patient noticed a more pronounced reduction in visual acuity at age 37, associated with development of onset photophobia (both eyes), and later, metamorphopsia in the right eye.

At age 42 his best corrected visual acuity (BCVA) was 0.2 in the right eye (OD) and 0.25 in the left eye (OS). Slit-lamp examination revealed moderate opacities of the crystalline lenses and generalized chorioretinal atrophy, including punched-out lesions, affecting the periphery and the

posterior pole, upon dilated fundus examination. (Figure 3A). Scarce pigment deposits were visible in the peripheral retina as well as bilateral optic disc pallor and attenuated retinal vessels (characteristic triad of retinitis pigmentosa). Fluorescence angiography showed generalized window defects and atrophy of the photoreceptor/pigment epithelium complex. Fundus autofluorescence showed spotty areas of hypo-autofluorescence outside of the vascular arcades and a central ring of hyper-autofluorescence in the posterior pole.

On the last examination, at age 43, marked progression of peripheral visual field constriction was noted in both eyes, which was more severe in the OD (limited to the central 10-15 degrees). Spectral domain optical coherence tomography (SD-OCT) revealed pronounced atrophy of the retinal layers, with consequent enhanced visualization of the choroidal vessels. Electroretinograms (ERG) showed no response to light stimuli, either from cone or from rod photoreceptor cells.

The patient's past medical history was remarkable for cardiac arrhythmia under bisoprolol treatment, as well as infertility. Due to unsuccessful conception, a full spermogram analysis was performed, the diagnosis resulting in a oligo-teratozoospermia (reduced sperm concentration and abnormal head shape) and severe asthenozoospermia (complete absence of motility). Table 1 summarizes this analysis, as compared to the reference normal values described by the World Health Organization (26). The causes of infertility were investigated by exploring possible occurrence of deletions in the Y chromosome, as described for the Sertoli cell-only syndrome (OMIM: 305700). The presence of 21 genetic markers (STS) on chromosome Yq and two markers on Yp (Supplemental table 1) was verified, therefore excluding this possibility. A fresh sperm sample, originally collected for immunofluorescence analysis for the present study, showed that most sperm heads were without flagellum, and approximately 10% of cells had a shorter flagellum.

Patient P2 (ID: LL89), a woman, was diagnosed with a classical form of Retinitis Pigmentosa at the age of 36 years. The patient was born from a consanguineous union between first degree cousins, of Portuguese origins. The patient's initial visual complaints started at the age 11 years, with photophobia, and progressed to night blindness and photopsia by the age of 25, as well as constriction of the peripheral visual field by the age of 28. Disease progression led to a sharp reduction in visual acuity at age 38 years and loss of color vision by the age of 40. She underwent cataract surgery at age 40 and 48. At 50 she developed metamorphopsia in the OS.

Her paternal aunt P3 and her uncle P4 were also diagnosed with RP, showing similar progression and manifestations. There is no clinical history of other ciliopathic disorders in this family, with the exception of P4, whose past medical history was positive for chronic bronchitis, infertility and multiple myeloma (which ultimately led to the patient's death prior to this study, preventing further phenotypical analysis). At the last ophthalmologic examination patient P2 showed BCVA of 0.3 in the OD, 0.16 in the OS, and severe constriction of the peripheral visual field, with bilateral tunnel-like vision, restricted to the 5 degrees central. Dilated fundus examination revealed attenuated retinal vessels, pale optic discs (Figure 3B) and a marked generalized chorioretinal atrophy in the posterior pole and in the periphery, with areas of complete atrophy in the inferior peripheral region of the RE. Mild to moderate mottled pigment deposition in the retinal periphery, along with some irregular white patches in the posterior pole near the superior vessels, were also observed. SD-OCT revealed pronounced atrophy of the outer retinal layers and the presence of a foveal cyst. Fundus autofluorescence was notable for generalized hypo-autofluorescence with a wide hyper-autofluorescent ring in the posterior pole.

None of the patients in this study had self-reported respiratory conditions or subjective complaints (except for P4), metabolic disturbances, situs inversus or skeletal abnormalities.

#### **Identification of pathogenic variants in the *ARL2BP* gene**

The DNA of P1 was initially screened for mutations in the *TLL5* gene, due to its association to a similar phenotype, but resulted to be negative. Following whole-exome sequencing (WES), the data were analyzed using an internal in-silico pipeline assessing variant frequency in the general population, quality, and predicted impact at the protein level. This filtering resulted in 7 homozygous changes, of which 6 residing in autozygous regions (Figure 1). Among the 6 variants, a single-nucleotide substitution disrupting the canonical consensus donor splice site downstream of exon 3 in *ARL2BP* (NM\_012106.3:c.207+1G>A) was identified.

P2 was studied according to the same methodology, and 20 of the 23 rare homozygous variants resulting from our pipeline were located in autozygous regions. These included a 4-bp deletion in the coding sequence of *ARL2BP* exon 1, resulting in a frameshift and in the creation of an early stop codon at the beginning of exon 2 (c.33\_36delGTCT:p.Phe13ProfsTer15).

### **Patient P1 aberrant splicing**

RT-PCR analysis was performed on cDNA derived from P1 sperm RNA, in comparison to RNA from a healthy control donor. Gel electrophoresis revealed the absence of a band corresponding to the expected PCR product in P1, and the appearance of a fragment of inferior molecular weight. Sanger sequencing of this latter fragment showed that in P1's sperm *ARL2BP* transcripts were abnormally spliced, and that c.207+1G>A caused the skipping of exon 3, the fusion of exon 2 with exon 4, and the shift of the canonical reading frame, leading in turn to a premature termination at the 9<sup>th</sup> codon of exon 4 (Figure 2).

### **ARL2BP localization in sperm**

To determine the role for ARL2BP in spermiogenesis, the final stages of spermatogenesis, we investigated the localization of ARL2BP within the sperm cell. In the human control sample, ARL2BP localized at the base of the flagellum, as well as at the so-called equatorial zone (Figure 4A). In P1's sample we observed most of the sperm heads separated from their tails. Some intact sperms had a shortened tail. Staining was negative for most of the cells (Figure 4A). In about 2-5% of the cells, staining was similar to that of the control sample, possibly reflecting the residual expression of ARL2BP present in this patient.

In agreement with this, ARL2BP was found in the head-tail connecting apparatus, the principal piece, and the sperm head in murine sperm (Figure 4B). In contrast, no signal was observed in KO murine sperm, demonstrating the specificity of the antibody used (Figure 4B).

### **Immotile sperm with stubby tails and a decreased sperm cell count in P1 and in *Arl2bp* KO mice**

Male *Arl2bp* KO mice were also infertile. Therefore, we examined sperm motility with live imaging and found that the KO sperm were immotile, in agreement with the human patient phenotypic description (Supplemental Video 1 and 2). Of note, testis size and weight were comparable between WT and KO (Figure 5A and B). Morphological analysis of the testes by H & E staining displayed normal spermatogenesis in KO animals, however, sperm release into the lumen appeared impaired, with a smaller lumen area, an absence of sperm tails, and an increase in residual bodies (RB) (Figure 5C). A closer inspection of the murine KO sperm revealed a drastically decreased sperm cell count (Figure 5D), and additional light microscopy images of

cauda epididymis sperm revealed that *Arl2bp* KO sperm had gross morphological defects, including numerous detached heads and tails, kinked tails at the connecting piece, stubby tails, abnormal heads, and cytoplasmic bulges attached to the tails (Figure 5E). These results are consistent with the oligo-teratozoospermia reported in the male patient and the morphology of his sperm, as detected by microscopy (Figure 4A).

### **Impaired development of sperm tails in the absence of ARL2BP**

The severity of azoospermia in P1 and KO mice spurred the investigation into the morphology and formation of the sperm tail and accessory structures using established markers. We first examined the sperm tail core (axoneme) using microtubule-associated markers. Despite severe loss in protein levels in *Arl2bp* KO animals (Figure 6A), Acetylated Tubulin (AcTu) and Glutamylated Tubulin (GluTu, GT335) localized to the sperm tail in WT and KO animals (Figure 4B and 6E). Moreover, testes cross-sections from KO mice revealed a further irregularity in tail shapes (spiraled) with GT335 staining (Figure 6D). In addition, Retinitis Pigmentosa GTPase Regulator (RPGR) and Sperm Flagellar Protein 2 (SPEF2), both axoneme associated proteins, were barely present in KO murine sperm, with diminished and spotty staining (Figure 6F). These findings show that loss of ARL2BP does not interrupt the initiation of microtubular axoneme growth. However, shortened axonemes and a decrease in axoneme-associated protein localization to the sperm tail indicates that axoneme formation is impaired. To determine if the assembly of the accessory structures is normal in the absence of ARL2BP, we assessed sperm tails using markers such as A-Kinase Anchoring Protein 82 (AKAP4, fibrous sheath) and Outer Dense Fiber Protein 1 (ODF1, outer dense fiber). Both markers were absent in murine KO sperm, though present in murine KO testes lysates (Figure 6A and F). Additionally, both forms of AKAP4 kept to the expected distribution in WT animals(30), with most of the soluble, non-assembled precursor to AKAP4 (pro-AKAP4, 82kDa) in the testes samples, while the majority present in the sperm lysates was the phosphorylated form of AKAP4 assembled into the FS (109kDa) (Figure 6A). In contrast, both forms of the protein were present in the KO testes, with neither present in the sperm lysates (Figure 6A). Supporting this, AKAP4 staining in KO testes sections revealed a retention of AKAP4 in the residual bodies shed during spermiogenesis, and an absence of AKAP4 from any sperm tails. In comparison, WT displayed AKAP4 in both residual bodies and luminal sperm tails (Figure 6C).

Conversely, there was presence (though minimal) of Heat-Shock Protein 60kDa (HSP60, mitochondrial sheath) in murine KO sperm (Figure 6F). Furthermore, staining of WT and KO testes cross-sections using Mitotracker showed that the development of mitochondria in elongating spermatids was present (Figure 6D), indicating the formation of a mitochondrial sheath. This was independently confirmed by the presence of pyruvate dehydrogenase protein (Pyr. Deh., mitochondrial sheath) in sperm lysates (Figure 6A). All together, these results demonstrate a failure to complete spermiogenesis by the inability to form the outer dense fiber layer or assemble the fibrous sheath in *Arl2bp* KO animals.

#### **Acrosome structure askew with lack of ARL2BP**

ARL2BP staining in the sperm head and the presence of abnormally shaped sperm heads in the male patient and KO mice prompted our investigation of acrosome development throughout spermatogenesis. It appeared that acrosome formation started to become irregular at stages IV-VI of spermatogenesis (Figure 7A and B). The acrosome starts to spread like a cap around the nucleus in stages VI-VII (cap phase), yet the KO displayed clusters of acrosomal granules instead of the cap structure (Figure 7B). Furthermore, at later stages in spermatogenesis, there were instances of irregularly shaped acrosomes surrounding regularly shaped nuclei. Figure 7C shows acrosome caps in KO testes at stages VIII-XI that are more oval instead of round, with a flimsier appearance. While at stage XII, there were a few misshapen acrosomes displaying a “molar tooth” shape (Figure 6D). Of note, most acrosomes on elongating spermatids at steps 14-16 (stage II-VIII) in KO testes appear normal, suggesting that ARL2BP is not crucial for acrosome formation.

#### **ARL2BP is required for proper sperm tail microtubule structure**

Extensive analysis of ultrastructural images produced from WT and KO murine testes and epididymis tissue corroborated the decreased sperm count and lack of sperm tail formation in the KO. Spermatogenesis appeared to be relatively normal through early tail formation in spermiogenesis, including formation of the manchette with centrally located centrioles; however, significant abnormalities in neck and tail ultrastructure were noted in later stages of spermiogenesis after step 10 (data not shown). In sperm from testis sections, the basal plate and capitulum were present in the neck, but segmented columns were either absent or severely

disrupted. The mitochondrial sheath contained centrally located mitochondria but they were not properly organized. The outer dense fibers and fibrous sheath were in disarray, though present, which is consistent with the protein levels observed for ODF1 and AKAP4. This contrasts with the WT sperm tails, which demonstrated the organized structure illustrated in Supplemental Figure 1 (data not shown). Furthermore, longitudinal and cross-sections of *Arl2bp* KO sperm tails revealed that microtubules were present in parallel arrays in the proximal tail region but they did not form the canonical 9 + 2 axoneme arrangement (data not shown). Microtubules were not paired at all and some seemed to be incomplete tubules. Nevertheless, most tubules were associated with electron-dense material that appears to be outer dense fibers and/or fibrous sheath. These accessory structures were not properly organized and were scattered in various portions of the tail.

#### **ARL2BP is necessary for left-right patterning in mice.**

Human patients, as well as KO animals, were further examined for additional phenotypes typically observed in ciliopathies. The human patients in this study did not report any symptoms associated with impaired cilia in the trachea or kidney. This finding is corroborated by our observations of normal tracheal cilia beat frequency and morphologically normal kidneys in *Arl2bp* KO mice (data not shown). However, we observed a high number of KO mice displaying situs inversus or heterotaxy, an asymmetric L/R positioning of the internal organs caused by defects in the nodal cilia of developing embryos (Figure 8A). The association of ARL2BP mutations in humans with situs inversus is variable, and the patients reported in this study did not possess situs inversus. However, a previous study identified one patient with a mutation in ARL2BP with situs inversus. In mice, this phenotype is much less variable, as most *Arl2bp* KO animals had situs inversus (55%), with 28% possessing heterotaxy of either the heart or stomach (Figure 8B). Furthermore, tracking and statistical analysis revealed that the number of KO's produced from Heterozygous x Knockout parents did not follow Mendelian ratios when examined from full-term litters or from embryonic day 13.5 (chi square value of 39.19,  $p > 0.001$ , and chi square value of 7,  $0.01 > p > 0.005$ , respectively). However, litters examined from embryonic day 7.5 did follow Mendelian ratios (chi square value of 0.96) (Figure 7C). Heterozygous x Heterozygous crosses followed a similar trend. These results establish ARL2BP's essential role in node-determined laterality during development.

**Cerebral ventricular volume affected with loss of ARL2BP.**

An additional ciliated tissue investigated in the *Arl2bp* KO mice was the cerebral ventricles, a region of the brain responsible for the circulation of cerebrospinal fluid. Remarkably, larger ventricular volume was observed and quantified in the brains of KO mice compared to WT littermates using ventricular volume comparisons from micro-CT scans (Figure 8). Of note, the third ventricle also appeared enlarged in *Arl2bp* KO mice, however we were not able to accurately quantify this. The gross morphology of the rest of the brain appears similar to WT in size and shape. Interestingly, all but one KO mouse did not appear hydrocephalic, a phenotype associated with many ciliopathies, with no obvious behavioral or motor deficits associated with this increase. Though the larger ventricular volume points to a role for ARL2BP in CSF flow regulation, a detailed behavioral analysis would need to be completed to confirm the observation that these structural changes do not cause any deficits.

**MEFs display shortened primary cilia in the absence of ARL2BP.**

To further investigate the role for ARL2BP, mouse embryonic fibroblasts (MEFs) were generated from *Arl2bp* KO mice and WT littermates. Primary cilia were induced in MEFs after removal of serum from the growth media. After 48 hours of serum starvation, the percentage of ciliated cells was comparable between WT and KO (Figure 9C and D). However, MEFs lacking ARL2BP possessed significantly shorter cilia (average of 2 $\mu$ m) than WT MEFs (average of 2.7 $\mu$ m) (Figure 9A and B). To determine if the cilia present in *Arl2bp* KO MEFs had difficulty in primary cilia depolymerization or re-entry into the cell cycle, cilia resorption was assessed throughout 24 hours of serum addition on cilia-induced cells. Interestingly, after 2 hours of serum addition, a significantly higher percentage of KO MEFs retained their cilia than what was observed in WT (Figure 9C and D). Further observation of KO and WT MEFs revealed that cilia resorption was comparable from 6 to 24 hours (Figure 9D). These results indicate that loss of ARL2BP affects the initial depolymerization of primary cilia.

## Discussion

Since its discovery, the function of ARL2BP has remained largely elusive. Initial characterization was established by the indication of its interaction with ARL2. ARL2BP was shown to bind with high affinity to ARL2-GTP but not to ARL2-GDP (33). Due to this GTP-dependent affinity, ARL2BP was determined as an ARL2-specific downstream effector, and not a GTPase activating protein or guanine nucleotide exchange factor. ARL2, together with its closest paralog ARL3, have been shown to allosterically regulate the release of farnesylated cargo between membranes, and specifically, in ciliated cells (34, 35). Moreover, research brought evidence for the role of ARL2 in  $\beta$ -tubulin conformation (17) and microtubule dynamics (36), and models describing ARL2 involvement in the promotion of  $\alpha\beta$ -tubulin biogenesis and degradation have been proposed (18, 37). Correspondingly, defects in ARF-related proteins were previously associated with retinal ciliary dysfunction. Examples are the Bardet Biedl associated gene *ARL6* (BBS3) (40), *ARL13B* causative for Joubert syndrome (41, 42), and the widely studied GTPase regulator *RPGR* causative for X-linked RP (43). Moreover, studies in mouse associate loss of function of ARL3 to RP and polycystic kidney disease (44, 45, 69). Past publications have clearly shown that dysfunctional ARL2BP impacts the photosensory cilium in humans and mice (38, 39). Experiments in human retina pigment epithelial cells showed that ARL2BP knockdown with small interfering RNA caused reduced cilia length (24), and similarly, the knockout mouse showed shortened photoreceptor cilia with impairment of ciliary doublet microtubule structure (25). The splice variant in our first index patient, as well as another nucleotide substitution in the same position, were recently published as causative for RP only (and situs inversus in one patient) (23, 24). Consistent with these findings, our previous murine studies show that loss of ARL2BP results in abnormal doublet microtubule structure of the axoneme and shortened cilia in photoreceptor cells, causing blindness (25).

### **ARL2BP is essential for sperm midpiece formation in humans and mice.**

The sperm tail flagellum is the longest in the body (~100 $\mu$ m in mice), and possesses accessory structures not present in other cilia. Spermatogenesis begins in the basal compartment at the outer edge of the seminiferous tubules and continues inward until fully formed spermatids are released into the lumen and on to the epididymal tissues. The last stage, spermiogenesis, is

characterized by sperm tail formation, beginning with axoneme growth and assisted by a transient microtubular structure, the manchette (46).

In mammals, proteins localizing at the equatorial segment are involved in the initiation of sperm oocyte fusion (70, 71). Similarly, we showed ARL2BP localizing in this same segment. If ARL2BP has any direct role in the egg fertilization process remains unclear. Certainly, without ARL2BP, sperm tails fail to elongate. Morphological analysis with TEM revealed that the 9+2 axoneme structure was disorganized, with no detection of doublet microtubules (DMT). It was also observed that ODF and FS still accumulate around the tail during spermatogenesis, but fail to assemble appropriately. This coincides with the loss of ODF1 and AKAP4 staining, and the increase of residual bodies present in testes of KO animals. Furthermore, the relative increase of phosphorylated AKAP4 present in the testes instead of the sperm of KO mice points to the inability to properly assemble the FS present throughout the principal piece. Its presence in the testes lysates is likely related to the fibrous sheath fragments seen in the cytoplasmic bulges present in TEM images of the testes, as well as the AKAP4-containing residual bodies. This is validated by the lack of AKAP4 staining on any sperm tails, and the increase of AKAP4 staining in residual bodies. The misassembled FS is also linked to the irregular axoneme structure, as the ODFs attach directly to their corresponding DMTs in the principal piece (47, 48). FS assembly occurs in a distal to proximal direction, following formation of the axoneme. Without the stable interaction and growth of the DMTs and ODFs, the FS cannot assemble correctly. Furthermore, AKAP4 processing and FS assembly is dependent on the formation and proper localization of the annulus. This is a septin ring-structure formed during spermiogenesis that travels to its position between the mid- and principal pieces after midpiece formation is complete (49). If the midpiece forms improperly, the annulus cannot localize appropriately, resulting in abnormal processing of AKAP4 and a failure to complete spermiogenesis. These defects in FS, and thereby principal piece formation, are accompanied by issues in midpiece formation. In the midpiece, the ODF does not directly bind to the microtubule doublets, but it is held in place by, and in early spermatogenesis attached to, the surrounding mitochondrial sheath (MS) (50, 51). The lack of a properly assembled MS could contribute to the poorly assembled ODFs. Therefore, it is possible that without ARL2BP, the malformation in the microtubule structure causes impaired assembly of all periaxonemal structures and a failure to complete spermiogenesis. This is evident by the lack of principal pieces in sperm tails from the KO model and in the majority of sperms from the

human sample. Relatedly, the detached heads and tails seen in the patient and murine model also fit with improperly formed ODFs, as these dense fibers morph into the striated columns of the connecting piece (50, 52), and the striated columns were not detected on any TEM images from KO tissue. Improper formation of the striated columns, presumably, would equal instability and breakage.

There seem to be an additional rationale for these spermiogenesis impairments with loss of ARL2BP. Manchette assembly, disassembly, or intra-manchette transport (IMT) could be affected. The manchette is a transient microtubule- and F-actin-based structure formed during spermiogenesis and is involved in the formation and assembly of the MS, ODF, FS, and acrosome (49). IMT is very similar to intraflagellar transport (IFT), even making use of IFT proteins such as IFT20 and IFT25 (15, 53, 54). Although the manchette forms in *Arl2bp* KO sperm (data not shown), it is possible that its function is impaired or that there is a delay in the manchette disassembly. This may account for the malformed MS, acrosome, FS, and ODFs. For instance, SPEF2, a protein required for the central axonemal pair in sperm flagellum, is also involved in IMT with dynein I. In SPEF2 KO mice, the manchette develops normally, but is elongated and is not removed on time, resulting in abnormally shaped heads and improperly assembled periaxonemal structures (55). Furthermore, though the acrosome formation relies on IMT during its “cap-phase,” it is a golgi-derived structure (golgi-phase) (56). SPEF2 was also shown to co-localize with Golgi in male germ cells along with IFT20, and to interact with GOLGA3 (53, 55). The hypothesis is that these proteins travel along manchette MTs with the Dynein I complex to form the golgi-derived acrosome. Furthermore, it was found that MEIG1 is required for spermiogenesis, with abnormal head and tail formation seen in KO mice, most likely due to a disrupted manchette (where MEIG1 was localized) (29-31). Though it is unlikely ARL2BP directly binds with SPEF2 or MEIG1, the abnormally shaped heads and impaired assembly of accessory structures with loss of these proteins is similar, indicating the ARL2BP may be involved in IMT or manchette assembly/disassembly.

Comparatively, the possibility must be kept open that the DMT/ODF disarray is linked to improperly formed MS, and FS disruption is a consequence. Loss of ARL2BP could impact the roles of ARL2, which as mentioned previously, has been localized to the mitochondrial IMS and is required for mitochondrial fusion and motility (20-22, 57). Additionally, ARL2 localization to mitochondria increases in *Mfn2* knockout MEFs, and expression of dominant active ARL2,

ARL2-Q70L, partially rescues mitochondrial fragmentation seen with loss of MFN2 (22). MNS1, a protein that interacts with MFN2, a mitogenic fusion protein that helps bind mitochondria to other cellular structures (like ODF) and is present in the midpiece of sperm (21, 59), is required for spermatogenesis and loss of MNS1 in mice causes similar phenotypes as seen with loss of ARL2BP (immotility, disorganized MT and ODF with short sperm tails, and situs inversus and mild hydrocephaly) (58). It could be hypothesized that ARL2BP binding to ARL2-GTP regulates its mitochondrial localization, and that ARL2-GTP is involved in mitochondrial motility and fusion with MFN2 in spermatogenesis. Examination of mitochondrial morphology demonstrated that the mitochondrial sheath helix is disrupted, paired with improperly assembled ODFs in the KO sperm, which could be due to a loss of fusion with the mitochondria. However, the mitochondria showed no signs of tubulation or fragmentation (data not shown). Therefore, it is difficult to interpret whether the mitochondrial defect seen is primary or secondary to the microtubule structural impairments. Many mouse models of mitochondria-related proteins show sperm phenotypes similar to the *Arl2bp* KO model. Aside from MNS1, loss of the cation channel VDAC3 results in loss of 1 DMT, slow motility, and enlarged and slightly disorganized mitochondria (21). However, there are many mouse models of non-mitochondrial proteins that display mitochondrial sheath disruptions as well. For instance, *Galnt3*<sup>-/-</sup> mice (an acrosome-related protein) had deformed acrosomes and slightly disturbed mitochondrial arrangement (22), and similarly, loss of *Iqcg* in a mouse model resulted in shortened sperm tails with misshapen heads and slightly disrupted mitochondrial arrangement (23). With these results in mind, we cannot conclude whether the sperm phenotype seen with loss of ARL2BP involves ARL2.

Taken together with the data observed from the WT and *Arl2bp* KO MEFs, the delay in cilia depolymerization most closely relates to the hypothesis that ARL2BP is in some way involved in manchette disassembly. Furthermore, we also did not observe mitochondrial tubulation or fragmentation in *Arl2bp* KO MEFs compared to WT MEFs (data not shown). Therefore, we hypothesize that ARL2BP is essential for depolymerization of microtubules in the manchette and MEF primary cilium.

### **ARL2BP is involved in organ laterality**

To date, only one patient with defective ARL2BP was reported to have situs inversus (24). This laterality defect was also observed in KO mice, but not in patients from the two families of

this study, who have normal organ position (situs solitus). L/R-asymmetry is determined by the mono-ciliated cells in the embryonic node during gastrulation. There is a leftward-fluid flow generated in this region by motile monocilia, which bend the immotile cilium of the crown cells and signals for asymmetrical protein expression related to eventual organ placement (61, 62). Therefore, if this fluid flow is disrupted in any way, the laterality of the organs will be affected. It was shown in an elegant study by Nonaka, et. al. that complete reversal of the flow (rightward) resulted in complete situs inversus (63). However, what determines situs inversus vs other forms of laterality placements is not known, and is thought to be random. Since the axonemes in photoreceptors and sperm tails are shorter with loss of *ARL2BP*, it is likely that nodal cilia are also shorter in the KO animal. Interestingly, there was an abnormally high incidence (55%) of situs inversus, compared with normal laterality or heterotaxia in the KO mice. Throughout the breeding process however, we noticed a significant decrease in the expected number of KO animals after laterality had been determined (e7.5). Therefore, we hypothesize that some KO embryos possessing heterotaxia died embryonically. Since different forms of heterotaxia are related to embryonic lethality (heart malformations, congenital heart disease), this could be possible (64, 65). This could also explain why only one patient identified with *ARL2BP* mutations possesses situs inversus, while the incidence in the mouse model is much higher.

#### ***ARL2BP* is not essential for the function of multi-ciliated cells**

Multi-ciliated cells are present in the ventricles of the brain, throughout the trachea, and in the fallopian tubes. Previously identified patients with mutations in *ARL2BP* were not reported to have any symptoms related to defects in these cilia (sinusitis, otitis media, hydrocephaly, or female infertility). Among patients of this study, only P4 was reported with chronic bronchitis. The late age of onset of this respiratory issue and the fact that the patient is presently deceased, makes it difficult to discern possible association of this to mutated *ARL2BP*-driven ciliary defects. To note, the mouse model also did not display any related symptoms. Interestingly however, the trachea of KO mice possessed tufts of normal cilia, immotile cilia, and cilia with uncoordinated beating (data not shown). Likewise, the ventricular volumes in KO mouse brains were significantly increased but neither of these defects were enough to cause an associated phenotype. Therefore, we consider that *ARL2BP* is not required for the function of multi-ciliated cells, though it does appear to be necessary for robust performance of these cilia.

Notably, is it is plausible that many adult ciliopathy patients with no reported hydrocephaly could possess asymptomatic hydrocephalus (symptoms are chronic headache, visual problems, clumsiness, and fogginess), which would go unnoticed as a disease symptom unless proper questions were asked and imaging was performed (66). Though this is based on the mouse model, it is possibly the case with *ARL2BP* patients as well. Furthermore, hydrocephalus is a more rarely reported phenotype of PCD, even when linked to a high incidence of hydrocephalus in PCD mouse models (10). Incidentally, there are many mouse models that report similar sperm tail deformations alongside symptoms of hydrocephaly (*Spef2*<sup>-/-</sup> mice, (67), and situs inversus (58). Interestingly, however, when trying to sonographically diagnose PCD patients, slightly enlarged ventricles were observed in an embryo with situs inversus (68). These observations point to two concerns in analyzing mouse models of ciliopathies vs human patients: mice ventricular ependymal cilia are more susceptible to disturbances, and therefore hydrocephaly, and human patients not presenting with overt childhood hydrocephalus might be overlooked in diagnosing ciliopathies.

Collaboratively, these results indicate *ARL2BP* as an essential cilia-related protein, loss of which results in a multi-syndromic ciliopathy, most closely resembling PCD. Loss of *ARL2BP* results in a defect in midpiece formation, most likely due to microtubule structural defects. Moreover, *ARL2BP* could be playing a role in mitochondrial motility (and therefore MS formation) with *ARL2*. These observations drive the hypotheses that *ARL2BP* is involved in multiple pathways related to ciliogenesis and DMT formation.

With our findings, we validate the past association of *ARL2BP* mutations with the blinding disease RP, and bring the first evidences of *ARL2BP* involvement in spermatogenesis. We thereby prove that additional cilia-related phenotypes might originate from *ARL2BP* deficiency, with manifestations that are similar in human and mouse. Furthermore, we provide a first insight into the disease mechanisms associated with *ARL2BP* mutations in relationship to defective ciliogenesis, pointing to an essential role for this protein in the maintenance of normal structure and homeostasis of cilia and flagella.

### Supplemental Data

Supplemental Data include two videos, two figures, and one table.

### Acknowledgements

The authors thank Dr. Peter Stoilov and Dr. Elena Pugacheva (West Virginia University) for help with data collection and analysis, the Imaging Core for use of video microscopes and Micro CT-scanner (West Virginia University), the Pathology Core for help with the preparation/processing of tissues (West Virginia University). This work was supported by USA National Institutes of Health Grants R01 EY028035 (to VR), R01 EY025536 (to VR and Peter Stoilov), R21 EY027707 (to VR), EY025291(to AFXG), and West Virginia Lions and Lions Club International Foundation, as well as grants from the Swiss National Science Foundation to CR [grant number:176097].

### Authors contributions

ARM and NB planned and conducted the experiments relative to the mouse and human samples, respectively. LCS, LPP, and PC are the medical scientists who visited the patients. ARM, NB, CR, VR, LCS, LPP, and VP wrote the manuscript. ARM, JC, ET, PM, MGP, ACW, and JAF performed mouse experiments. VR and CR supervised the research.

### References

1. Levin M, Johnson RL, Stern CD, Kuehn M, and Tabin C. A molecular pathway determining left-right asymmetry in chick embryogenesis. *Cell*. 1995;82(5):803-14.
2. Youn YH, and Han Y-G. Primary Cilia in Brain Development and Diseases. *The American Journal of Pathology*. 2018;188(1):11-22.
3. Bujakowska KM, Liu Q, and Pierce EA. Photoreceptor Cilia and Retinal Ciliopathies. *Cold Spring Harb Perspect Biol*. 2017;9(10).
4. Satir P. CILIA: before and after. *Cilia*. 2017;6(1).
5. Morthorst SK, Christensen ST, and Pedersen LB. Regulation of ciliary membrane protein trafficking and signalling by kinesin motor proteins. *The FEBS Journal*. 2018;0(0).
6. Hildebrandt F, Benzing T, and Katsanis N. Ciliopathies. *The New England journal of medicine*. 2011;364(16):1533-43.
7. Cardenas-Rodriguez M, and Badano JL. Ciliary biology: Understanding the cellular and genetic basis of human ciliopathies. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. 2009;151C(4):263-80.
8. Satir P, Pedersen LB, and Christensen ST. The primary cilium at a glance. *Journal of Cell Science*. 2010;123(4):499-503.
9. Badano JL, Mitsuma N, Beales PL, and Katsanis N. The Ciliopathies: An Emerging Class of Human Genetic Disorders. *Annual Review of Genomics and Human Genetics*. 2006;7(1):125-48.

10. Ferkol TW, and Leigh MW. Ciliopathies: the central role of cilia in a spectrum of pediatric disorders. *The Journal of Pediatrics*. 2012;160(3):366-71.
11. Leigh MW, Pittman JE, Carson JL, Ferkol TW, Dell SD, Davis SD, Knowles MR, and Zariwala MA. Clinical and Genetic Aspects of Primary Ciliary Dyskinesia / Kartagener Syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2009;11(7):473-87.
12. Bedoni N, H-WL, Vaclavik V., Tran V.H., Farinelli P., Balzano S., Royer-Bertrand B., El-Asrag M.E., Bonny O., Ikonomidis C., Litzistorf Y., Nikopoulos K., Yioti G.G., Stefanidou M.I., McKibbin M., Booth A.P., Ellingford J.M., Black G.C., Toomes C., Inglehearn C.F., Hoyng C.B., Bax N., Klaver C.C., Thiadens A.A., Murisier F., Schorderet D.F., Ali M., Cremers F.P., Andréasson S., Munier F.L., Rivolta C. Mutations in the polyglutamylase gene *TTL5*, expressed in photoreceptor cells and spermatozoa, are associated with cone-rod degeneration and reduced male fertility. *Hum Mol Genet*. 2016;25(45):46-55.
13. Lee G-S, He Y, Dougherty EJ, Jimenez-Movilla M, Avella M, Grullon S, Sharlin DS, Guo C, Blackford JA, Awasthi S, et al. Disruption of *Ttl5*/Stamp Gene (Tubulin Tyrosine Ligase-like Protein 5/SRC-1 and TIF2-associated Modulatory Protein Gene) in Male Mice Causes Sperm Malformation and Infertility. *The Journal of Biological Chemistry*. 2013;288(21):15167-80.
14. Sun X, Park JH, Gumerson J, Wu Z, Swaroop A, Qian H, Roll-Mecak A, and Li T. Loss of RPGR glutamylation underlies the pathogenic mechanism of retinal dystrophy caused by *TTL5* mutations. *Proc Natl Acad Sci U S A*. 2016;113(21):E2925-34.
15. Zhang Z, Li W, Zhang Y, Zhang L, Teves ME, Liu H, Strauss JF, Pazour GJ, Foster JA, Hess RA, et al. Intraflagellar transport protein IFT20 is essential for male fertility and spermiogenesis in mice. *Molecular Biology of the Cell*. 2016;27(23):3705-16.
16. Keady BT, Le YZ, and Pazour GJ. IFT20 is required for opsin trafficking and photoreceptor outer segment development. *Molecular Biology of the Cell*. 2011;22(7):921-30.
17. Francis JW, Goswami D, Novick SJ, Pascal BD, Weikum ER, Ortlund EA, Griffin PR, and Kahn RA. Nucleotide Binding to ARL2 in the TBCD·ARL2· $\beta$ -Tubulin Complex Drives Conformational Changes in  $\beta$ -Tubulin. *Journal of Molecular Biology*. 2017.
18. Nithianantham S, Le S, Seto E, Jia W, Leary J, Corbett KD, Moore JK, and Al-Bassam J. Tubulin cofactors and Arl2 are cage-like chaperones that regulate the soluble  $\alpha\beta$ -tubulin pool for microtubule dynamics. *eLife*. 2015;4(e08811).
19. Bhamidipati A, Lewis SA, and Cowan NJ. ADP ribosylation factor-like protein 2 (Arl2) regulates the interaction of tubulin-folding cofactor D with native tubulin. *J Cell Biol*. 2000;149(5):1087-96.
20. Sharer JD, Shern JF, Van Valkenburgh H, Wallace DC, and Kahn RA. ARL2 and BART enter mitochondria and bind the adenine nucleotide transporter. *Mol Biol Cell*. 2002;13(1):71-83.
21. Newman LE, Schiavon CR, Zhou C, and Kahn RA. The abundance of the ARL2 GTPase and its GAP, ELMOD2, at mitochondria are modulated by the fusogenic activity of mitofusins and stressors. *PLoS One*. 2017;12(4):e0175164.
22. Newman LE, Schiavon CR, Turn RE, and Kahn RA. The ARL2 GTPase regulates mitochondrial fusion from the intermembrane space. *Cellular logistics*. 2017;7(3):e1340104-e.
23. Fiorentino A, Yu J, Arno G, Pontikos N, Halford S, Broadgate S, Michaelides M, Carss KJ, Raymond FL, Cheetham ME, et al. Novel homozygous splicing mutations in *ARL2BP* cause autosomal recessive retinitis pigmentosa. *Molecular Vision*. 2018;24(6):3-12.
24. Davidson AE, Schwarz N, Zelinger L, Stern-Schneider G, Shoemark A, Spitzbarth B, Gross M, Laxer U, Sosna J, Sergouniotis PI, et al. Mutations in *ARL2BP*, encoding ADP-ribosylation-factor-like 2 binding protein, cause autosomal-recessive retinitis pigmentosa. *Am J Hum Genet*. 2013;93(2):321-9.
25. Moye AR, Singh R, Kimler VA, Dilan TL, Munezero D, Saravanan T, Goldberg AFX, and Ramamurthy V. *ARL2BP*, a protein linked to retinitis pigmentosa, is needed for normal photoreceptor cilia doublets and outer segment structure. *Mol Biol Cell*. 2018;29(13):1590-8.

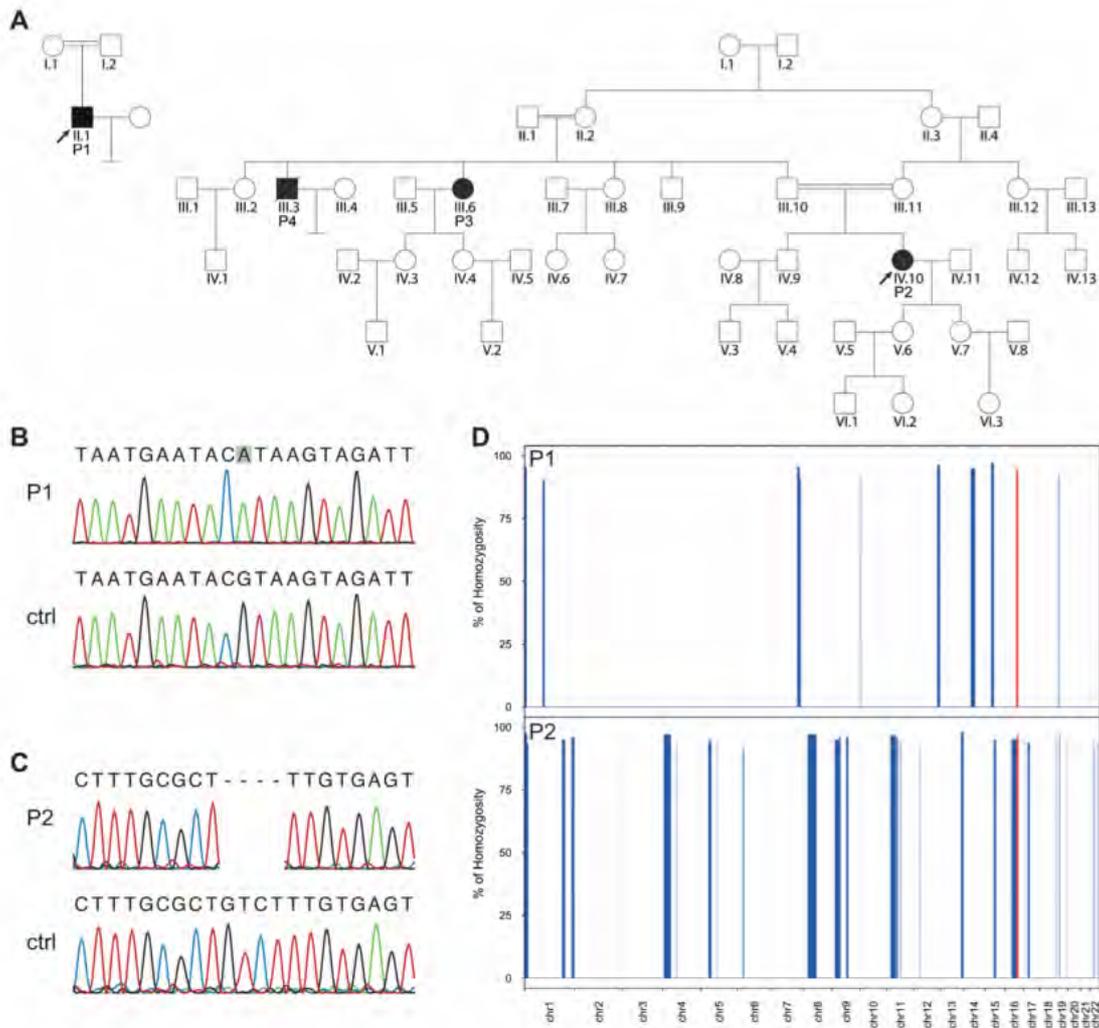
26. Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HWG, Behre HM, Haugen TB, Kruger T, Wang C, Mbizvo MT, et al. World Health Organization reference values for human semen characteristics\*. *Human Reproduction Update*. 2010;16(3):231-45.
27. DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011;43(5):491-8.
28. Wang K, Li M, and Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010;38(16):e164-e.
29. Nishimura H, Gupta S, Myles DG, and Primakoff P. Characterization of mouse sperm TMEM190, a small transmembrane protein with the trefoil domain: evidence for co-localization with IZUMO1 and complex formation with other sperm proteins. *Reproduction*. 2011;141(4):437-51.
30. Johnson LR, Foster JA, Haig-Ladewig L, Vanscoy H, Rubin CS, Moss SB, and Gerton GL. Assembly of AKAP82, a Protein Kinase A Anchor Protein, into the Fibrous Sheath of Mouse Sperm. *Developmental Biology*. 1997;192(2):340-50.
31. Sharer JD, and Kahn RA. The ARF-like 2 (ARL2)-binding protein, BART. Purification, cloning, and initial characterization. *J Biol Chem*. 1999;274(39):27553-61.
32. Clark J, Moore L, Krasinskas A, Way J, Battey J, Tamkun J, and Kahn RA. Selective amplification of additional members of the ADP-ribosylation factor (ARF) family: cloning of additional human and Drosophila ARF-like genes. *Proceedings of the National Academy of Sciences of the United States of America*. 1993;90(19):8952-6.
33. Zhang T, Li S, Zhang Y, Zhong C, Lai Z, and Ding J. Crystal structure of the ARL2-GTP-BART complex reveals a novel recognition and binding mode of small GTPase with effector. *Structure*. 2009;17(4):602-10.
34. Pearring JN, Salinas RY, Baker SA, and Arshavsky VY. Protein sorting, targeting and trafficking in photoreceptor cells. *Prog Retin Eye Res*. 2013;36(24-51).
35. Wätzlich D, Vetter I, Gotthardt K, Miertzschke M, Chen Y-X, Wittinghofer A, and Ismail S. The interplay between RPGR, PDE $\delta$  and Arl2/3 regulate the ciliary targeting of farnesylated cargo. *EMBO reports*. 2013;14(5):465-72.
36. Francis JW, Turn RE, Newman LE, Schiavon C, and Kahn RA. Higher order signaling: ARL2 as regulator of both mitochondrial fusion and microtubule dynamics allows integration of 2 essential cell functions. *Small GTPases*. 2016;7(4):188-96.
37. Al-Bassam J. Revisiting the tubulin cofactors and Arl2 in the regulation of soluble  $\alpha\beta$ -tubulin pools and their effect on microtubule dynamics. *Molecular biology of the cell*. 2017;28(3):359-63.
38. Audo I, El Shamieh S, Méjécasse C, Michiels C, Demontant V, Antonio A, Condroyer C, Boyard F, Letexier M, Saraiva JP, et al. ARL2BP mutations account for 0.1% of autosomal recessive rod-cone dystrophies with the report of a novel splice variant. *Clinical Genetics*. 2016;92(1):109-11.
39. Fiorentino A, Yu J, Arno G, Pontikos N, Halford S, Broadgate S, Michaelides M, Carss KJ, Raymond FL, Cheetham ME, et al. Novel homozygous splicing mutations in ARL2BP cause autosomal recessive retinitis pigmentosa. *Molecular vision*. 2018;24(603-12).
40. Fan Y, Esmail MA, Ansley SJ, Blacque OE, Boroevich K, Ross AJ, Moore SJ, Badano JL, May-Simera H, Compton DS, et al. Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome. *Nature Genetics*. 2004;36(989).
41. Hanke-Gogokhia C, Wu Z, Sharif A, Yazigi H, Frederick JM, and Baehr W. The guanine nucleotide exchange factor, Arf-like protein 13b, is essential for assembly of the mouse photoreceptor transition zone and outer segment. *Journal of Biological Chemistry*. 2017.
42. Wheway G, Parry DA, and Johnson CA. The role of primary cilia in the development and disease of the retina. *Organogenesis*. 2014;10(1):69-85.
43. Md SA, Ponjavic V, Abrahamson M, Ehinger B, Wu W, Fujita R, Buraczynska M, and Swaroop A. Phenotypes in Three Swedish Families With X-linked Retinitis Pigmentosa Caused by Different Mutations in the RPGR Gene. *American Journal of Ophthalmology*. 1997;124(1):95-102.

44. Hanke-Gogokhia C, Wu Z, Gerstner CD, Frederick JM, Zhang H, and Baehr W. Arf-like Protein 3 (ARL3) Regulates Protein Trafficking and Ciliogenesis in Mouse Photoreceptors. *The Journal of biological chemistry*. 2016;291(13):7142-55.
45. Strom SP, Clark MJ, Martinez A, Garcia S, Abelazeem AA, Matynia A, Parikh S, Sullivan LS, Bowne SJ, Daiger SP, et al. De Novo Occurrence of a Variant in ARL3 and Apparent Autosomal Dominant Transmission of Retinitis Pigmentosa. *PloS one*. 2016;11(3):e0150944-e.
46. Oakberg EF. A description of spermiogenesis in the mouse and its use in analysis of the cycle of the seminiferous epithelium and germ cell renewal. *American Journal of Anatomy*. 1956;99(3):391-413.
47. Olson GE, and Linck RW. Observations of the structural components of flagellar axonemes and central pair microtubules from rat sperm. *J Ultrastruct Res*. 1977;61(1):21-43.
48. Lindemann CB, and Gibbons IR. Adenosine triphosphate-induced motility and sliding of filaments in mammalian sperm extracted with Triton X-100. *J Cell Biol*. 1975;65(1):147-62.
49. Lehti MS, and Sironen A. Formation and function of sperm tail structures in association with sperm motility defects. *Biol Reprod*. 2017;97(4):522-36.
50. Lindemann CB, Lesich, Kathleen A. Functional Anatomy of the Mammalian Sperm Flagellum. *Cytoskeleton*. 2016;73(11):652-69.
51. Lesich KA, de Pinho TG, Dang L, and Lindemann CB. Ultrastructural evidence that motility changes caused by variations in ATP, Mg<sup>2+</sup>, and ADP correlate to conformational changes in reactivated bull sperm axonemes. *Cytoskeleton*. 2014;71(11):649-61.
52. Ounjai P, Kim KD, Lishko PV, and Downing KH. Three-dimensional structure of the bovine sperm connecting piece revealed by electron cryotomography. *Biol Reprod*. 2012;87(3):73.
53. Sironen A, Hansen J, Thomsen B, Andersson M, Vilkki J, Toppari J, and Kotaja N. Expression of SPEF2 during mouse spermatogenesis and identification of IFT20 as an interacting protein. *Biol Reprod*. 2010;82(3):580-90.
54. Liu H, Li W, Zhang Y, Zhang Z, Shang X, Zhang L, Zhang S, Li Y, Somoza AV, Delpi B, et al. IFT25, an intraflagellar transporter protein dispensable for ciliogenesis in somatic cells, is essential for sperm flagella formation. *Biol Reprod*. 2017;96(5):993-1006.
55. Lehti MS, Zhang F-P, Kotaja N, and Sironen A. SPEF2 functions in microtubule-mediated transport in elongating spermatids to ensure proper male germ cell differentiation. *Development*. 2017;144(14):2683-93.
56. Leblond CP, and Clermont Y. Spermiogenesis of rat, mouse, hamster and guinea pig as revealed by the "periodic acid-fuchsin sulfurous acid" technique. *American Journal of Anatomy*. 1952;90(2):167-215.
57. Newman LE, Zhou C-j, Mudigonda S, Mattheyses AL, Paradies E, Marobbio CMT, and Kahn RA. The ARL2 GTPase is required for mitochondrial morphology, motility, and maintenance of ATP levels. *PloS one*. 2014;9(6):e99270-e.
58. Zhou J, Yang F, Leu NA, and Wang PJ. MNS1 is essential for spermiogenesis and motile ciliary functions in mice. *PLoS Genet*. 2012;8(3):e1002516.
59. Vadnais ML, Lin AM, and Gerton GL. Mitochondrial fusion protein MFN2 interacts with the mitostatin-related protein MNS1 required for mouse sperm flagellar structure and function. *Cilia*. 2014;3(5-).
60. Lokaj M, K??sling SK, Koerner C, Lange SM, Van Beersum SEC, Van Reeuwijk J, Roepman R, Horn N, Ueffing M, Boldt K, et al. The Interaction of CCDC104/BARTL1 with Arl3 and Implications for Ciliary Function. *Structure*. 2015;23(11):2122-32.
61. McGrath J, Somlo S, Makova S, Tian X, and Brueckner M. Two populations of node monocilia initiate left-right asymmetry in the mouse. *Cell*. 2003;114(1):61-73.
62. Babu D, and Roy S. Left-right asymmetry: cilia stir up new surprises in the node. *Open Biol*. 2013;3(5):130052.
63. Nonaka S, Shiratori H, Saijoh Y, and Hamada H. Determination of left-right patterning of the mouse embryo by artificial nodal flow. *Nature*. 2002;418(6893):96-9.

64. Narasimhan V, and Roy S. Cilia: Organelles at the Heart of Heart Disease. *Curr Biol*. 2015;25(13):R559-62.
65. Li Y, Klena NT, Gabriel GC, Liu X, Kim AJ, Lemke K, Chen Y, Chatterjee B, Devine W, Damerla RR, et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature*. 2015;521(7553):520-4.
66. Cowan JA, McGirt MJ, Woodworth G, Rigamonti D, and Williams MA. The syndrome of hydrocephalus in young and middle-aged adults (SHYMA). *Neurological Research*. 2005;27(5):540-7.
67. Finn R, Evans CC, and Lee L. Strain-Dependent Brain Defects in Mouse Models of Primary Ciliary Dyskinesia with Mutations in *Pcdp1* and *Spef2*. *Neuroscience*. 2014;277(552-67).
68. Wessels MW, den Hollander NS, and Willems PJ. Mild fetal cerebral ventriculomegaly as a prenatal sonographic marker for Kartagener syndrome. *Prenatal Diagnosis*. 2003;23(3):239-42.
69. Schrick JJ, Vogel P, Abuin A, Hampton B, Rice DS. ADP-ribosylation factor-like 3 is involved in kidney and photoreceptor development. *Am J Pathol*. 2006;168(4):1288-98.
70. Yanagimachi R. Fertility of mammalian spermatozoa: its development and relativity. *Zygote*. 1994;2(4):371-2.
71. Wolkowicz MJ, Shetty J, Westbrook A, Klotz K, Jayes F, Mandal A, Flickinger CJ, Herr JC. Equatorial segment protein defines a discrete acrosomal subcompartment persisting throughout acrosomal biogenesis. *Biol Reprod*. 2003;69(3):735-45.
72. Aldahmesh MA, Li Y, Alhashem A, Anazi S, Alkuraya H, Hashem M, Awaji AA, Sogaty S, Alkharashi A, Alzahrani S, Al Hazzaa SA, Xiong Y, Kong S, Sun Z, Alkuraya FS. *IFT27*, encoding a small GTPase component of IFT particles, is mutated in a consanguineous family with Bardet-Biedl syndrome. *Hum Mol Genet*. 2014;23(12):3307-15.
73. Zhang Y, Liu H, Li W, Zhang Z, Shang X, Zhang D, Li Y, Zhang S, Liu J, Hess RA, Pazour GJ, Zhang Z. Intraflagellar transporter protein (IFT27), an IFT25 binding partner, is essential for male fertility and spermiogenesis in mice. *Dev Biol*. 2017;432(1):125-139.

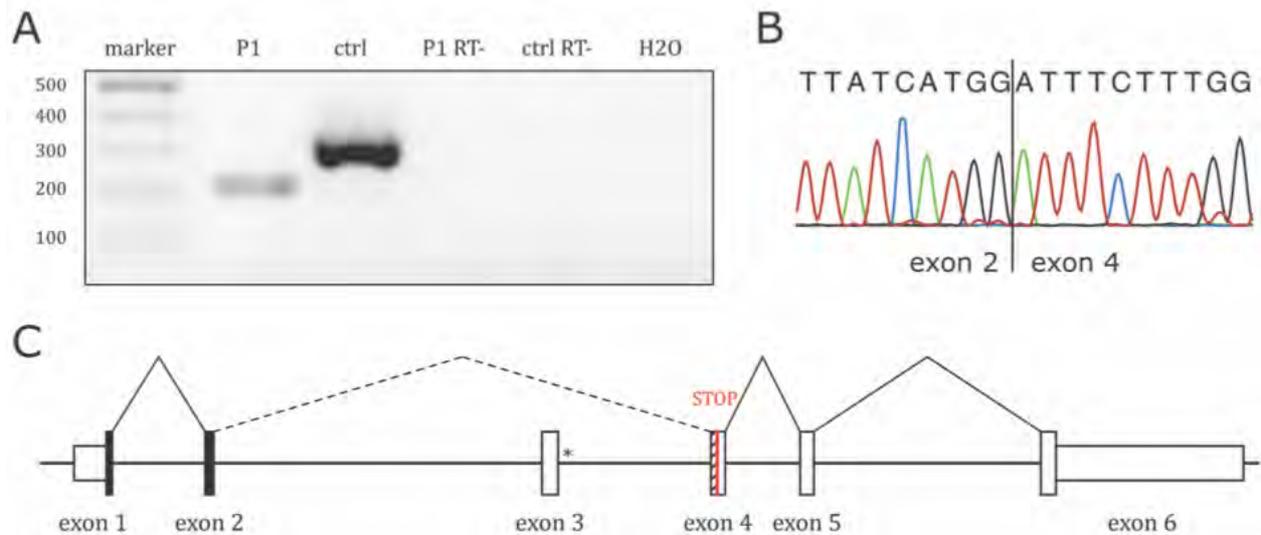
**Manuscript : Mutations in *ARL2BP*, a protein required for ciliary microtubule structure, cause syndromic infertility in human and mouse.**

**Figures and Legends**



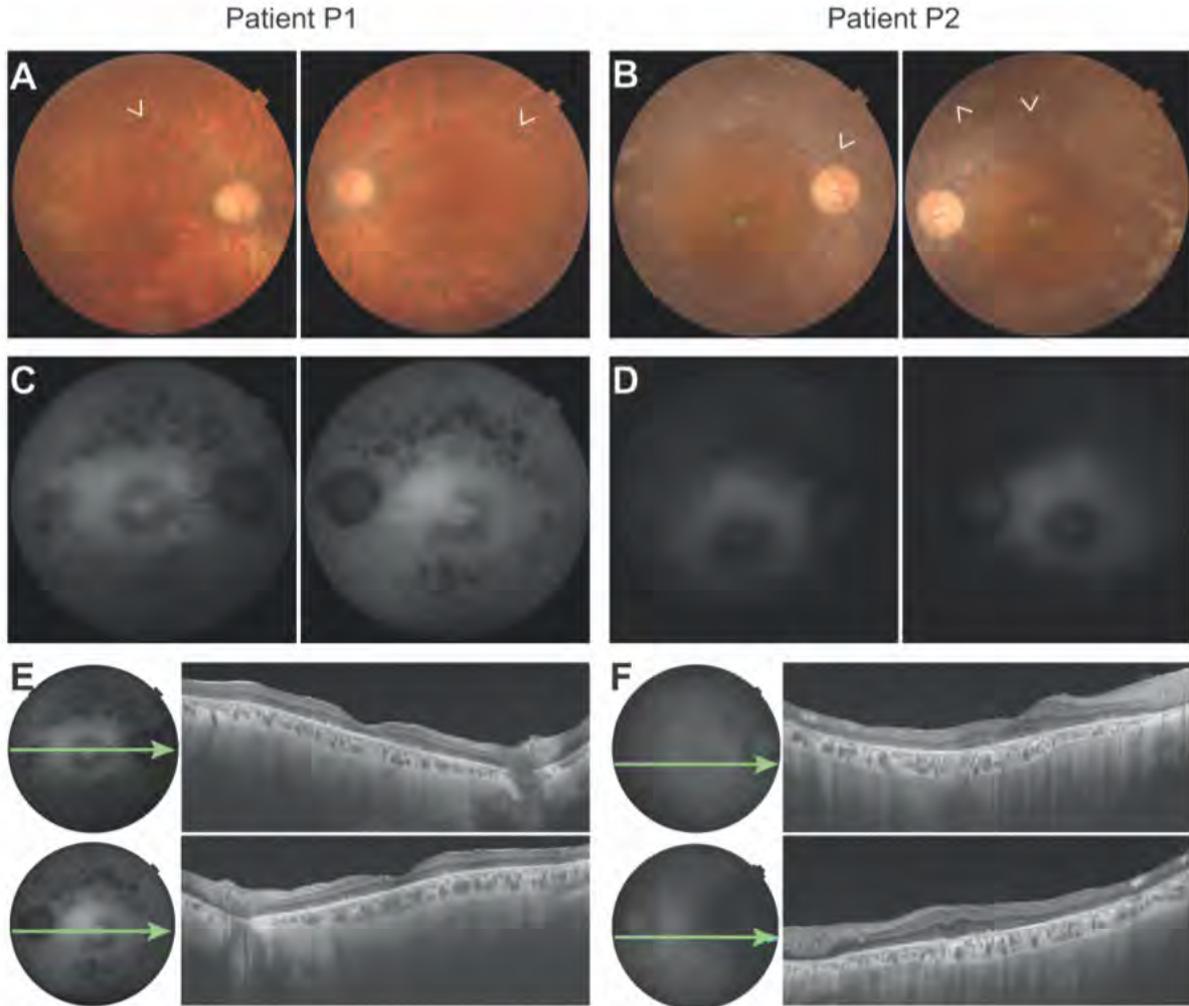
**Figure 1. Pedigrees and genetic findings.**

**A.** Pedigrees of the patients analysed in this study. DNA was available only for subjects P1, P2, and P3. **B and C.** Sanger validation of the WES findings, showing the presence of a homozygous splicing mutation in patient P1 (**B**, NM\_012106.3:c.207+1G>A, leading to p.Asp35PhefsTer8), and a frameshift deletion in patients P2 and P3 (**C**, NM\_012106.3:c.33\_36delGTCT: p.Phe13ProfsTer15), alongside with relevant control sequences (ctrl). **D.** Exome-wide homozygosity mapping for autosomal chromosomes, using the ARMS tool (unpublished). The autozygous region containing the gene *ARL2BP* is highlighted in red.



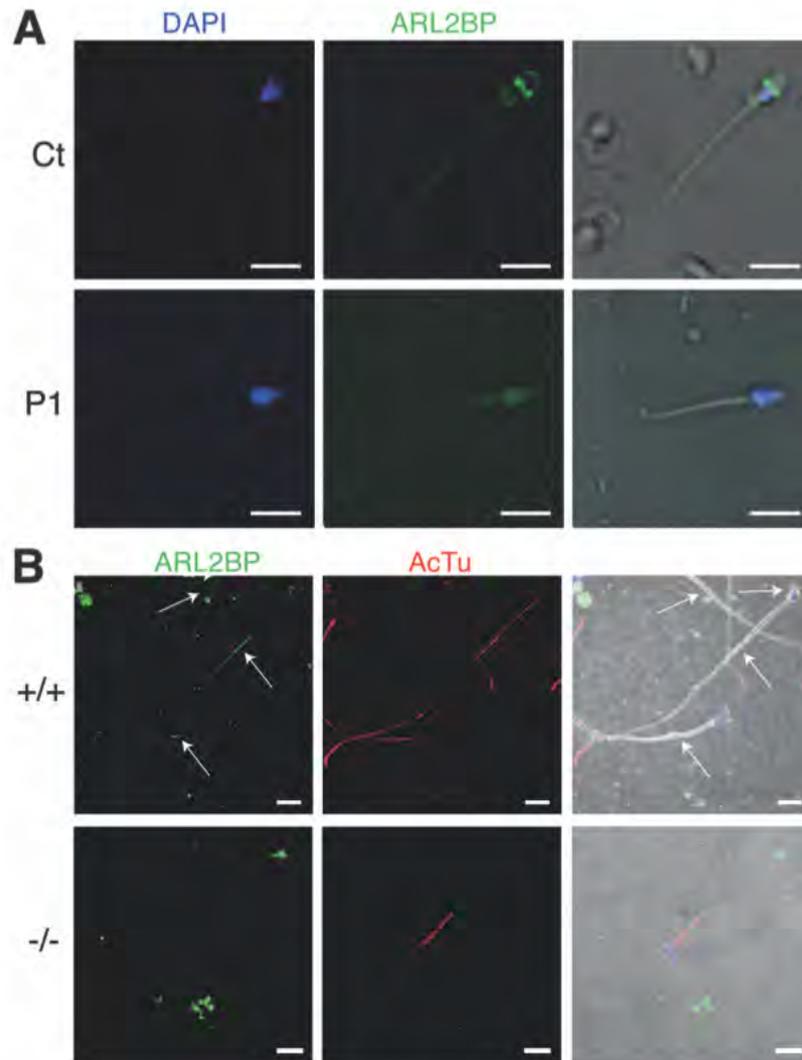
**Figure 2. Effect of the splicing mutation in patient P1.**

**A.** RT-PCR on cDNA obtained from sperm-derived RNA and Sanger sequencing of the same PCR product (P1, cDNA from patient; ctrl, cDNA from a healthy control; P1 RT- and ctrl RT-, reverse transcription controls) **(B).** **C.** In-scale schematic representation of the splicing pattern resulting from the presence of the mutation. c.207+1G>A leads to the skipping of exon 3 and the joining of exon 2 with exon 4, with subsequent shifting of the reading frame and creation of a premature stop codon early in exon 4 (red). The asterisk shows where the mutation is localized. The in-frame exons are black, whereas the striped portion of exon 4 is out-of-frame.



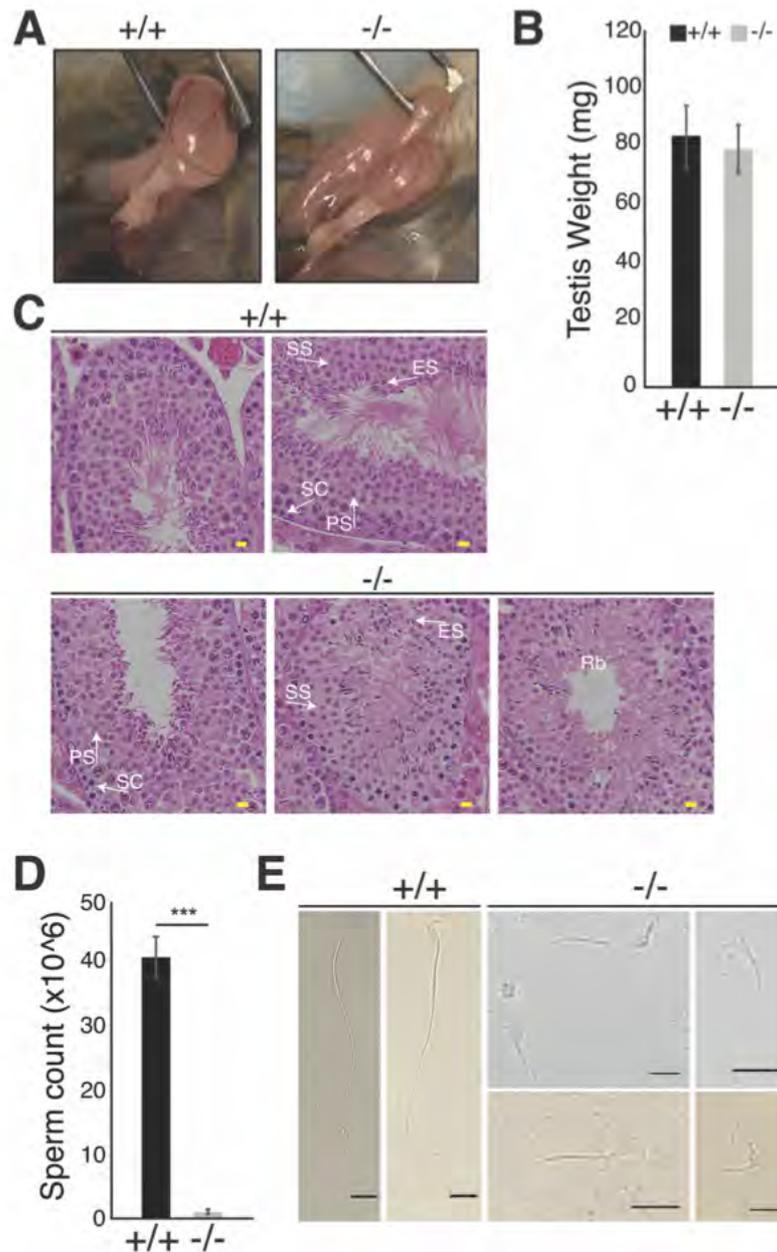
**Figure 3. Findings on retinal imaging of patients P1 and P2.**

**A and B.** Fundus photos of patients P1 and P2, respectively. P1 presents a pale optic disc, vascular thinning and retinal atrophy in the posterior pole and along the vascular arcades. Scarse pigment deposition can be found along the superior vascular arcade (as shown by the arrowheads). Similarly, P2 displays a pale optic disc, vascular thinning and marked retinal changes, with opalescent areas of the retinal tissue in the periphery and central atrophic lesion. Scarse pigment deposition can also be found adjacent to the optic nerve head and periphery (arrowheads) **C and D.** Fundus autofluorescence showing in P1 multiple hypoautofluorescent spots in the periphery corresponding to RPE atrophy, which in contrast is largely diffused in P2. Moreover, both patients present the typical central hyperautofluorescent ring. **E and F.** SD-OCT revealing in both patients diffuse retinal thinning, absent photoreceptors along with enhanced visualization of the choroidal vessels. Panels A-D: right eye followed by left eye. Panels E and F: right eye above and left eye below.



**Figure 4. ARL2BP localizes to the sperm tail in humans and mice.**

**A.** Most representative pictures: in the human control, staining is mainly at the equatorial segment of the sperm head whereas in the human patient there was very diminished staining. **B.** WT (+/+) and KO (-/-) murine sperm stained for ARL2BP to corroborate human sperm staining, co-stained with Acetylated Tubulin (AcTu, red). +/+ demonstrating localization in the sperm head, at the head-tail connecting apparatus (HTCA), the principal piece, and the annulus. Staining in the KO serves as a negative control. Scale bars = 10 $\mu$ m.



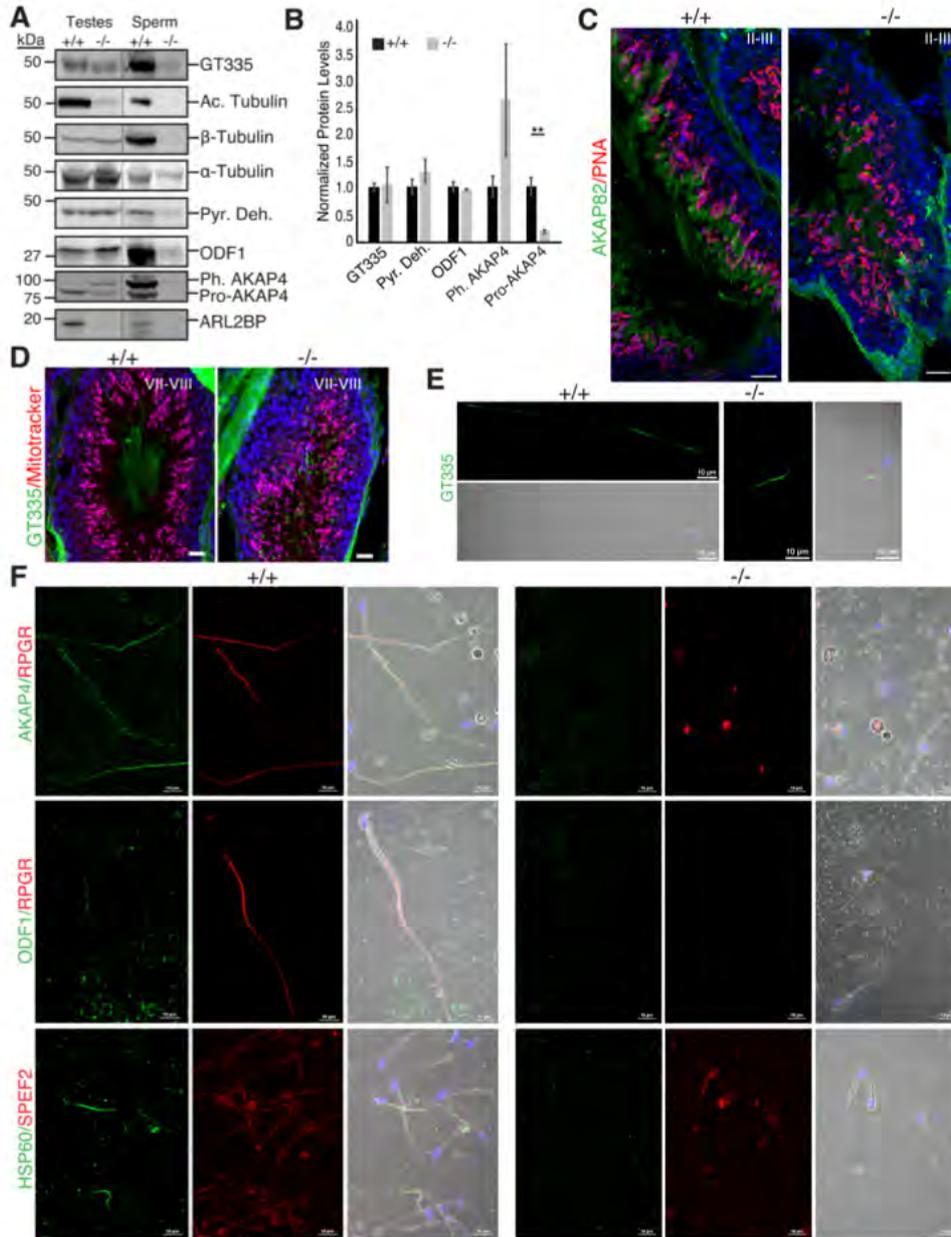
**Figure 5. Loss of ARL2BP leads to decreased sperm count and abnormal sperm structure in mice.**

**A and B.** Testis weight between WT (+/+) and KO (-/-) mice are comparable, represented in **A**) with average weights graphed in **B**). n=10. Data are represented as the mean  $\pm$  SEM.

**C.** H&E sections of WT (+/+) and KO (-/-) testis, with KO testis demonstrating lack of sperm tails in the lumen and an increase in residual bodies (RB). PS=primary spermatocyte, SS=secondary spermatocyte, ES=elongating spermatid, SC=sertoli cells. Scale bar = 20 $\mu$ m.

**D.** Graph presenting the sperm cell counts of WT (+/+) and KO (-/-) mice. n=3. Data are represented as the mean  $\pm$  SEM. P=0.0002.

**E.** Light images displaying WT sperm with normal sperm tail structure, while KO sperm show abnormal sperm head shapes, and sperm tails are shorter, detached, and retain cytoplasm. Scale bar = 10 $\mu$ m.

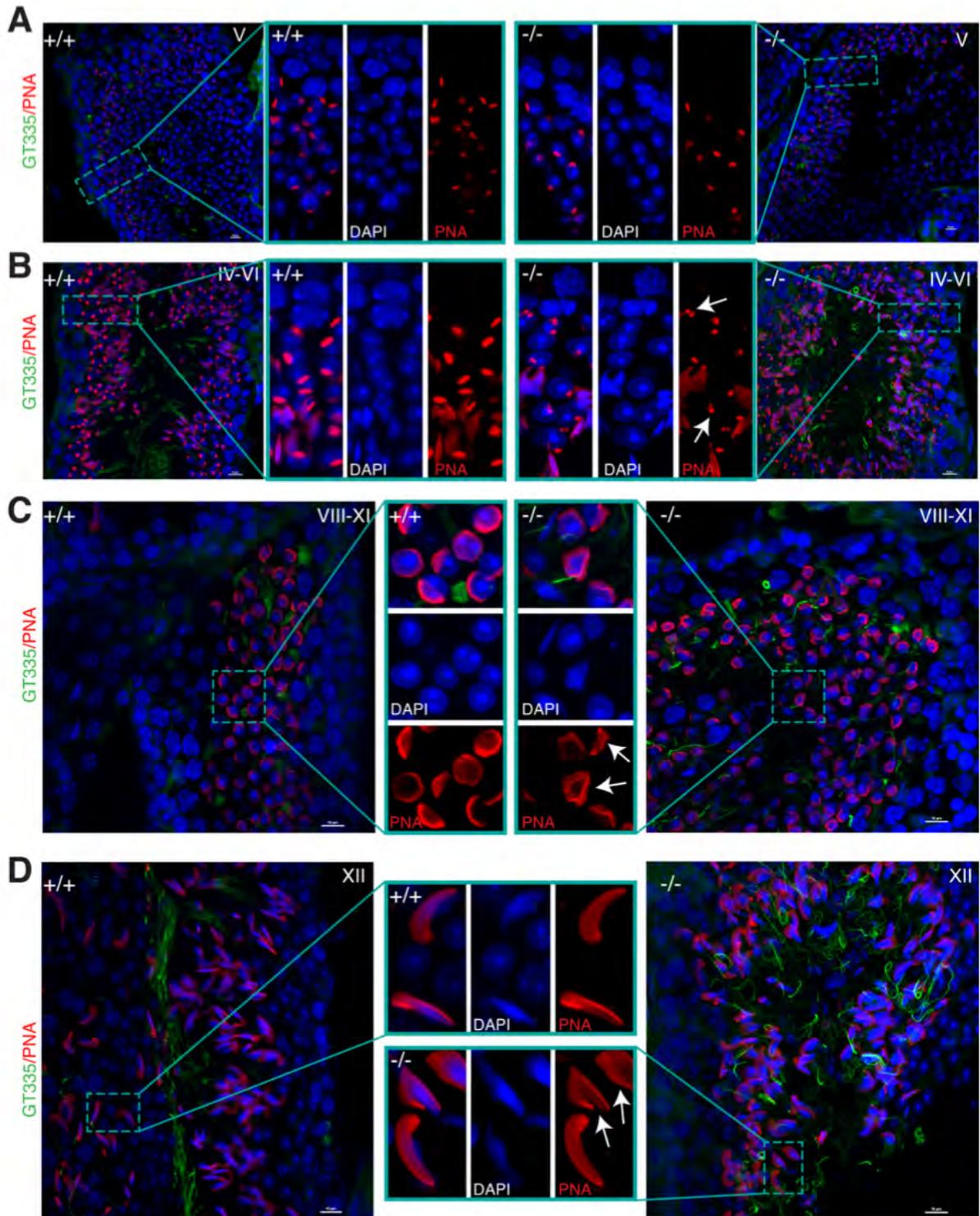


**Figure 6. ARL2BP loss results in impaired sperm tail development.**

**A.** Immunoblot of WT (+/+) and KO (-/-) testis and sperm lysates probed for the indicated sperm tail markers: axonemal markers (Glutamylated tubulin (GT335), Acetylated tubulin (Ac. Tubulin),  $\beta$ -tubulin, and  $\alpha$ -tubulin) and accessory structure markers (Pyruvate dehydrogenase (Pyr. Deh.), Outer Dense Fiber 1 (ODF1), and A-kinase anchoring protein 82 (pre-processed, phosphorylated = Pro-AKAP4 and processed = AKAP4).

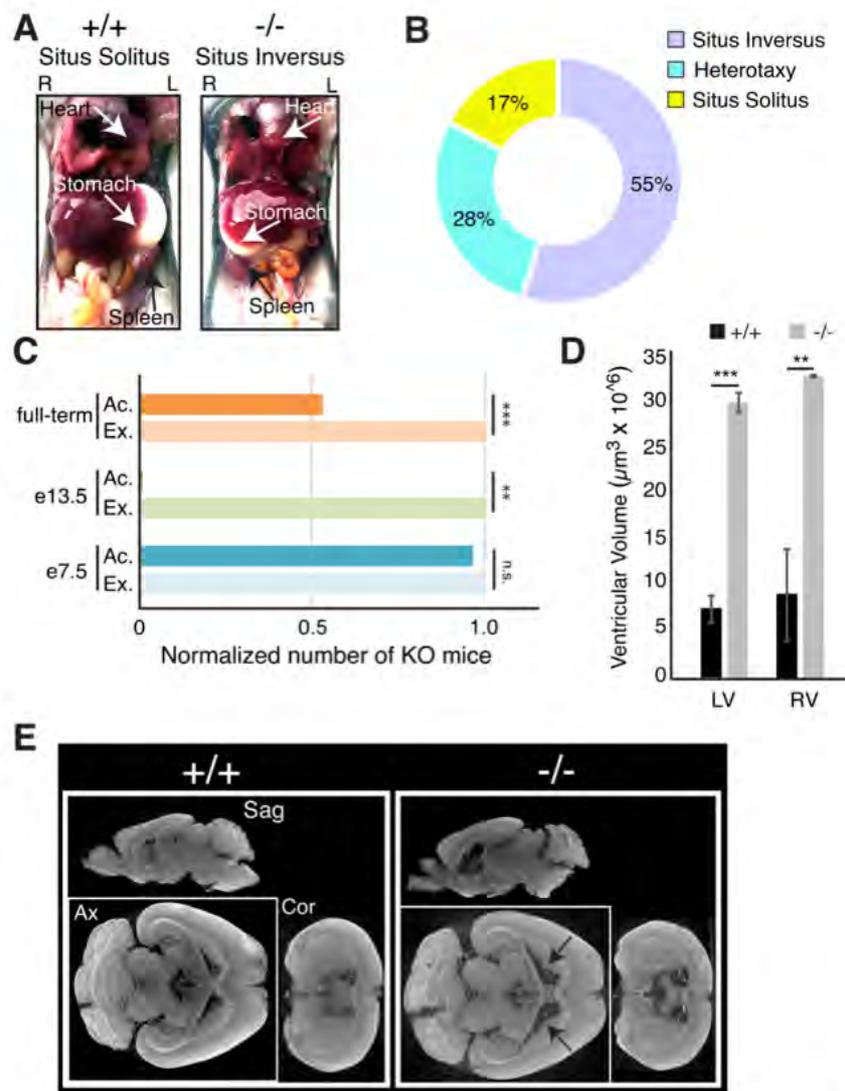
Molecular weight displayed on the left. **B.** Graph displaying the normalized protein levels corresponding to **A**). Pro-AKAP4  $p=0.2004$ , AKAP4  $p=0.0068$ .  $n=3$ . Data are represented as the mean  $\pm$  SEM. **C and D.** WT (+/+) and KO (-/-) murine testes sections displaying staining of DAPI (nuclei, blue) and **C**) PNA lectin (acrosomes, red) and AKAP4 (green), or **D**) mitotracker (mitochondria, red) and GT335 (glutamylated tubulin, green).

Scale bar = 20 $\mu$ m. **E and F.** Sperm stained with the indicated sperm tail markers in WT (+/+) and KO (-/-) murine sperm. Glutamylated Tubulin, (GT335 – axoneme, green), A-Kinase Associated Protein 4 (AKAP4 – fibrous sheath, green), Retinitis Pigmentosa GTPase Regulator (RPGR – axoneme, red), Outer Dense Fiber Protein 1, (ODF1 – outer dense fibers, green), Heat Shock Protein 60kDa (HSP60 – mitochondrial sheath, green), and Sperm Flagellar Protein 2 (SPEF2 – axoneme, red). Scale bar = 10 $\mu$ m.



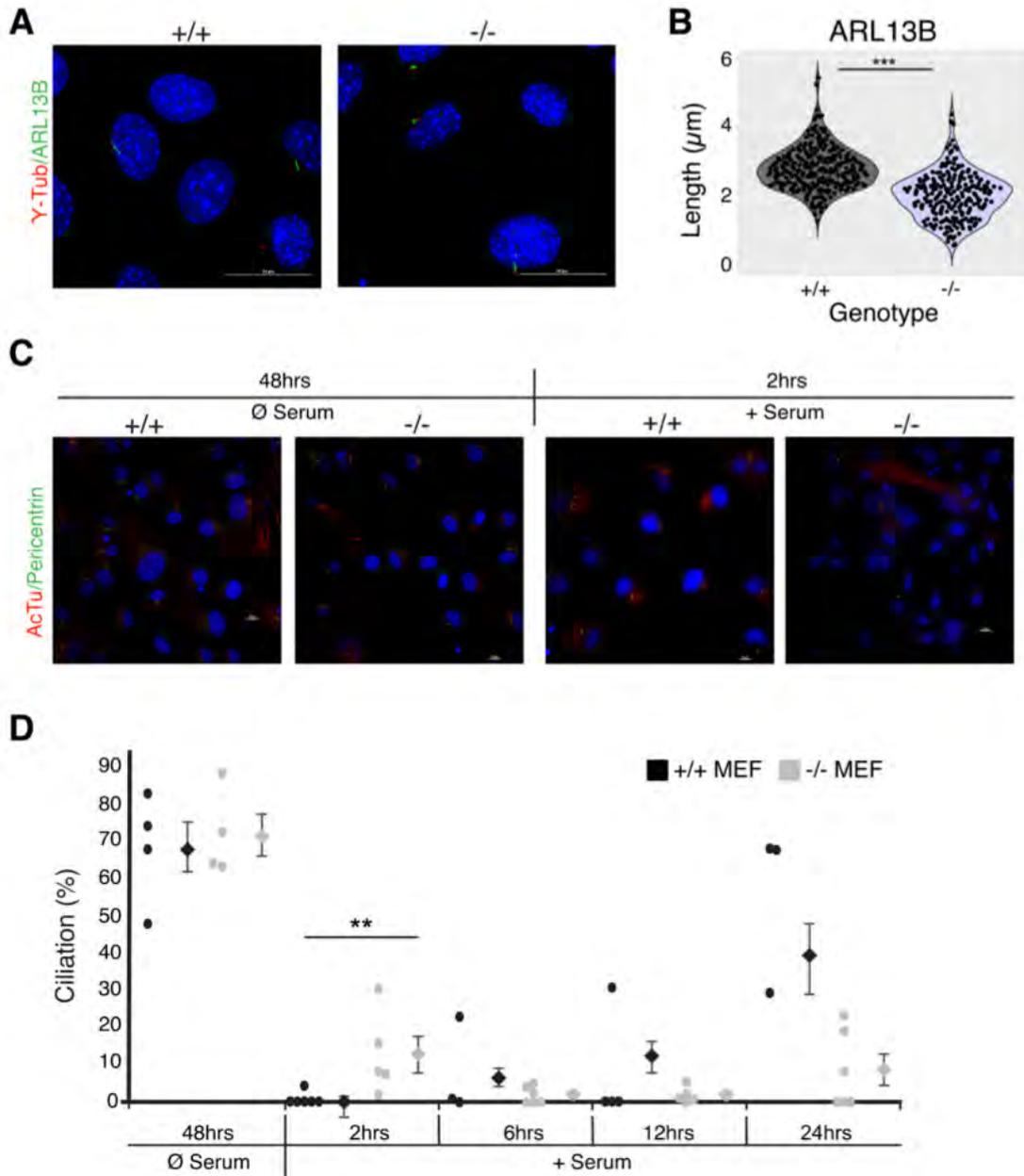
**Figure 7. Compromised spermiogenesis with loss of ARL2BP.**

**A-D.** WT (+/+) and KO (-/-) adult murine testes sections displaying staining of DAPI (nuclei, blue), PNA lectin (acrosomes, red) and GT335 (green, sperm tails in stages IV-VI and XII). Scale bar = 10 $\mu$ m. Inserts show zoomed in sections from each image. Arrows point to abnormal acrosomes.



**Figure 8. Laterality defects and enlarged lateral ventricles in the adult murine brain with absence of ARL2BP.**

**A.** WT (+/+) mouse displaying *situs solitus* (normal axis patterning), and KO (-/-) mouse exhibiting *situs inversus* (complete reversal). **B.** Percentages of KO animals with left/right-asymmetry abnormalities. **C.** Graph displaying the non-Mendelian distribution of KO animals identified from Het x KO crosses of full-term litters (total of 100 animals with 23 KO's, chi square value of 39.19,  $p > 0.001$ ), or collected at embryonic day 13.5 (total of 7 animals with 0 KO's, chi square value of 7,  $0.01 > p > 0.005$ ), compared to the Mendelian distribution seen at embryonic day 7.5 (total of 24 animals with 16 KO's, chi square value of 0.96). **D.** Ventricular volume comparison between WT (+/+) and KO (-/-) brains from 3D-reconstruction of CT images.  $n=3$ . Left  $p=0.0002$ , Right  $p=0.0087$ . Data are represented as the mean  $\pm$  SEM. **E.** Micro-CT scans of WT (+/+) and KO (-/-) murine brains at postnatal day 60, in Sagittal (Sag), Axial (Ax), and Coronal (Cor) planes. Arrows mark the enlarged lateral ventricles. Ac.=Actual #, Ex.=Expected #; LV=left ventricle, RV=right ventricle.



**Figure 9. Cells lacking ARL2BP undergo cilia depolymerization more slowly.**

**A.** Mouse embryonic fibroblasts (MEFs) from WT (+/+) and KO (-/-) littermates stained to show DAPI (nuclei, blue), ARL13B (cilia, green), and  $\gamma$ -tubulin (basal body, red). **B.** Violin plot displaying the length measurements from WT (+/+) and KO (-/-) MEFs. n=200. P=0.0001 **C.** MEFs from WT (+/+) and KO (-/-) littermates after serum starvation (cilia induction) for 48 hours, followed by serum addition (cilia depolymerization) for 2 hours stained for Acetylated tubulin (AcTu – cilia, red) and pericentrin (basal body, green). **D.** Dot plot displaying the percentage of ciliated cells between WT (+/+) and KO (-/-) MEFs after 48 hours of serum starvation and at indicated time points after re-addition of serum. Each dot represents 200 $\geq$ 100 cells. Average percentage of ciliated cells is to the right of each dot set. Data are represented as the mean  $\pm$  SEM. For 2hrs, p= 0.0174.

**Table 1**  
Spermiogram of patient P1

	P1	Lower reference limit <sup>A</sup>
Semen volume (ml)	1	1.5
Sperm concentration ( $10^6$ /ml)	50	15
Total number ( $10^6$ /ejaculate)	50	39
Total motility <sup>B</sup> (%)	0	40
Normal forms (%)	8	4
pH	9	7.2

<sup>A</sup> Reference data from WHO (PMID:19934213)

<sup>B</sup> Total motility includes progressive and non-progressive motility

## PROJECT 3

### ***NMNAT1* variants: between LCA and lethality**

A manuscript is under preparation for publication.

#### **Contributions:**

This work has been conducted in collaboration with another PhD student from our group, Mr. Mathieu Quinodoz. He initially found the genomic region that is common between the two families. Throughout the whole project, we widely exchanged thoughts and ideas, and he contributed with his bioinformatic skills, which were of great support. Supervision was provided by our thesis director Prof. Carlo Rivolta and by our collaborator Prof. Andrea Superti-Furga from the CHUV Lausanne hospital, who first detected the phenotypic similarities between the two families, and put us in contact with the respective clinical collaborators, Maja di Rocco, from the Institute Gaslini in Genoa, and Michele Pinelli and Sandro Banfi from the TIGEM (Telethon Institute of Genetics and Medicine) in Naples.

#### **Abstract**

We investigated the genetic cause for the phenotype of three children from two unrelated Italian families presenting with a previously-unrecognized, seemingly autosomal recessive disorder that included Leber congenital amaurosis (LCA), sensorineural hearing loss, intellectual disability, brain anomalies, and a distinct form of severe spondylo-epiphyseal dysplasia. Whole-exome sequencing and autozygome-based analysis of the two families, who share a common geographic origin, highlighted a 4.6 Mb homozygous region on the short arm of chromosome 1, common to all three patients. This genomic fragment comprises 40 genes, none of which harboured rare variants in coding regions. Expression analysis by RNA sequencing on fibroblasts from patients vs. controls revealed a 4-fold downregulation of the LCA-associated gene *NMNAT1*, residing in the shared autozygous interval. Semi-quantitative RT-PCR and Sanger sequencing of aberrant splice products highlighted a duplication that involves 2 out of the 5 *NMNAT1* exons. This rearrangement is predicted to alter splicing, leading to both longer transcripts with reduced half-life and to isoforms carrying aberrant coding sequences.

Mutations in *NMNAT1* have been previously shown to cause non-syndromic LCA. However, no patient with disruptive homozygous variants has ever been described, and murine *Nmnat1* knockouts show embryonic lethality. Because of the involvement of NMNAT1 in NAD<sup>+</sup> homeostasis and its crucial role in a wide range of cellular reactions, we hypothesize that complete absence of adenylyltransferase activity is not compatible with life, whereas milder (e.g. missense) mutations are responsible for LCA. The rearrangement found in our patients, presumably causing a strong but not complete reduction of enzymatic activity, may therefore result in an intermediate phenotype, characterized by LCA with skeletal dysplasia. We performed long read whole-genome sequencing to precisely define this rearrangement at the genomic level and exclude the presence of additional noncoding variants that may be associated with this new clinical entity.

## Introduction

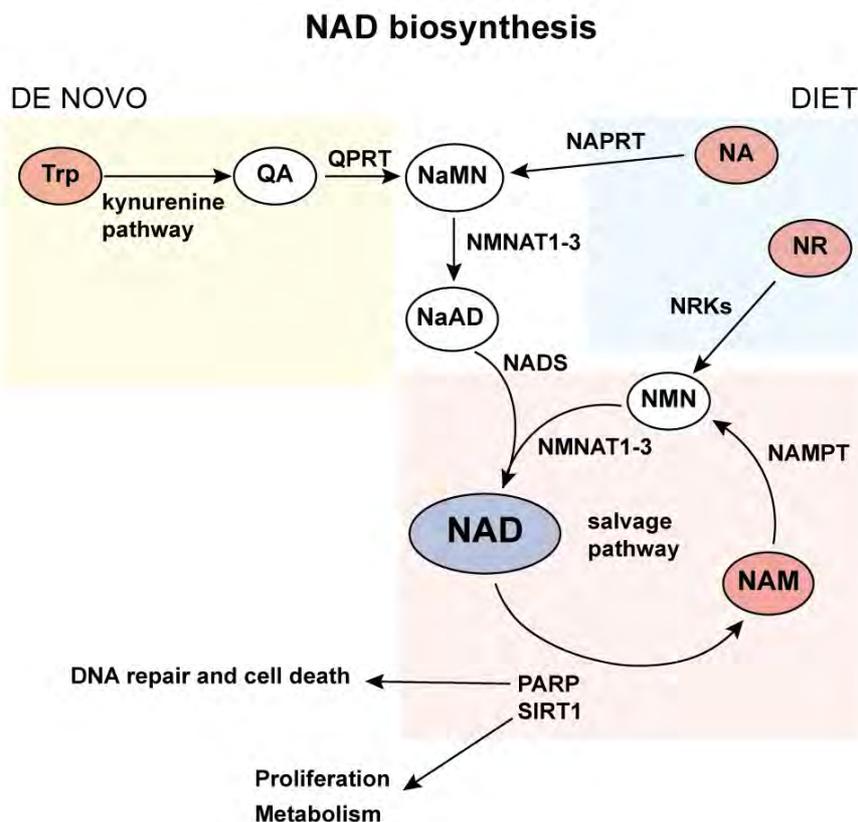
LCA is considered to be the most severe form of inherited neurodegenerative retinal disease, with an incidence of approximately 1 in 80,000<sup>44</sup>. Its distinctive feature is the early age of onset. LCA patients usually display in the first years of life a severe and rapidly progressive macular degeneration leading to central atrophy, with the appearance of congenital coloboma and atrophy of the optic nerve<sup>45</sup>.

Syndromic LCA is rare, and to date only two genes have been described as causative for LCA with additional manifestations. Autosomal dominant inheritance of loss-of-function variants in the *OTX2* gene were associated with early onset retinal degeneration and short stature. *OTX2* has been assigned with multiple roles, from development and function of both the RPE and the neural retina, as well as in the pituitary gland<sup>46</sup>. In the literature, two syndromic LCA cases were described in concomitance with renal features, these symptoms are collectively resumed under the name of Senior Loken syndrome (OMIM: 266900)<sup>47</sup>. Nevertheless, the ocular and renal phenotypes in Senior Loken syndrome are often very variable.

In the present study, we report syndromic LCA with a much broader phenotype range. Indeed, LCA was only a minor component of the signs and symptoms displayed by the three patients. Although many reports obviously exist about the genetic determinants leading to the individual phenotypes detected in these patients, to date no study has described a syndrome characterized by their simultaneous presence in the same individuals, highlighting the uniqueness and rarity of our patients' condition.

*NMNAT1* has been associated with LCA for the first time in 2012<sup>44,45,48,49</sup> and, since then, the genetic causes in numerous LCA patients were found in this same gene<sup>50-55</sup>, accounting for >5% of all inherited retinal degenerations<sup>28</sup>. NMNATs play a key role in the biosynthesis of nicotinamide adenine dinucleotide (NAD)<sup>56</sup> (Figure 1), catalysing the condensation of nicotinamide mononucleotide (NMN) or nicotinic acid mononucleotide (NaMN) with the AMP moiety of ATP to form NAD or Nicotinic acid dinucleotide (NaAD). The coenzyme NAD is consumed in hundreds of pathways, as for instance PARPs function in DNA repair and cell death, and for Sirtuin 1 (SIRT1) action in the regulation of proliferation and metabolism.

In this work, we applied an integrated approach, combining short- and long-read NGS, RNAseq, as well as functional studies, to determine the genetic elements leading to this previously-undescribed pathology. Our results led to the identification of a DNA structural rearrangement involving *NMNAT1* as the molecular cause for this disorder, thereby expanding the range of clinical consequences linked to mutations in this gene.



**Figure 1** : Schematic representation of the NAD biosynthesis. In mammals, NAD is produced *de novo* from transformation of tryptophan (Trp), or through the salvage pathway. In this latter, NAD is produced by transforming its precursors, nicotinic acid (NA) and nicotinamide riboside (NR), naturally occurring in food (vitamin B3), or by recycling performed components such as

nicotinamide (NAM) back into the NAD active form. Other abbreviations: quinolinic acid (QA), quinolinate phosphoribosyltransferase (QPRT), nicotinic acid adenine dinucleotide (NAAD), NAD synthase (NADS), nicotinic acid phosphoribosyltransferase (NAPT), nicotinamide phosphoribosyltransferase (NAMPT), nicotinamide riboside kinase (NRK).

## **Methods**

### **Patients and controls**

This study has been conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Boards of the University of Lausanne, of the Genoa Gaslini Institute and of TIGEM. Blood and skin biopsies have been collected by the patient's respective treating clinicians and written informed consent was signed by their legal tutors before donation.

### **Whole exome sequencing and autozygome analysis**

WES was performed on the three patients and two healthy parents from family pedigree 1 (Figure 2), using 2 µg of DNA derived from peripheral blood mononuclear cells. Protein-coding regions were captured using the HiSeq Rapid PE Cluster Kit v2 and an Illumina HiSeq 2500 instrument was used for paired-end sequencing. Single nucleotide variants and small insertions and deletions were detected using the Genome Analysis Tool Kit (GATK v4.0) software package, using the Best Practice Guidelines identified by the developers<sup>57</sup>. The pathogenicity of the detected genetic variants was assessed after functional annotation through ANNOVAR<sup>58</sup>. Shared regions of homozygosity and common haplotypes for each patient were analysed using an in-house built script (Mathieu Quinodoz, ARMS : Autozygosity Regions Mapping Software - unpublished).

### **RNA sequencing**

Total RNA was extracted from skin-derived fibroblasts of the three patients and of one control using trizol; kit and protocol from Direct-zol™ RNA MiniPrep (ZYMO Research). RNA libraries were prepared by a service provider (Fasteris, Geneva, Switzerland) and single-read RNA sequencing was performed using an Illumina HiSeq 2500 instrument. Fastq files were analysed with both the CLCbio Genomics Workbench (Qiagen) and the publicly available pipeline from RNAcocktail<sup>59</sup>. Raw Fastq files of 27 other control skin fibroblast cell lines were downloaded from ArrayExpress accession E-MTAB-4652 (ENA

study ERP015294)<sup>60</sup> and were used for more robust statistical analysis of our differentially expressed genes.

### **Quantitative real time PCR**

Relative expression levels were measured by quantitative real time PCR (q-PCR) using patient or control cDNA template prepared from primary skin fibroblast's RNA. Prior to sample measurements, efficiency of each primer pair was tested using a standard curve. Reference controls were selected according to their stability throughout the samples, the most stables being *TBP*, *GAPDH* and *ACTB*. Normalization was done using a geometric mean of the Ct values of the three reference genes. Amplification was performed using the SYBR Green PCR Master Mix (Applied Biosystems). The q-PCR product was visualized on 1% agarose gel to verify the primer's specificity.

### **High molecular weight DNA preparation and long reads WGS**

High molecular weight DNA was extracted from patient P1 primary fibroblasts using the Genomic tip 20/G kit (QIAGEN). At about 90% confluency, cells were scraped down from two 100 mm dishes, and the standard extraction protocol described in the providers kit was followed. DNA integrity was verified with a Fragment Analyzer automated system and analysed using the PROSize software version 3.0 (Advanced Analytical Technologies Inc). The DNA was calculated to sizes ranging an average of ~45.6 kbp.

Long reads WGS was performed with Pacific Bioscience (PacBio) technology. In order to get enough coverage 6 SMRT flow-cells were sequenced on the PacBio RS II platform reaching an approximate coverage of 10x. Reads were assembled to the reference human genome with the long-read specific aligner NGMLR, and variant calling was performed using the Sniffles software<sup>61</sup>.

### **NAD quantification**

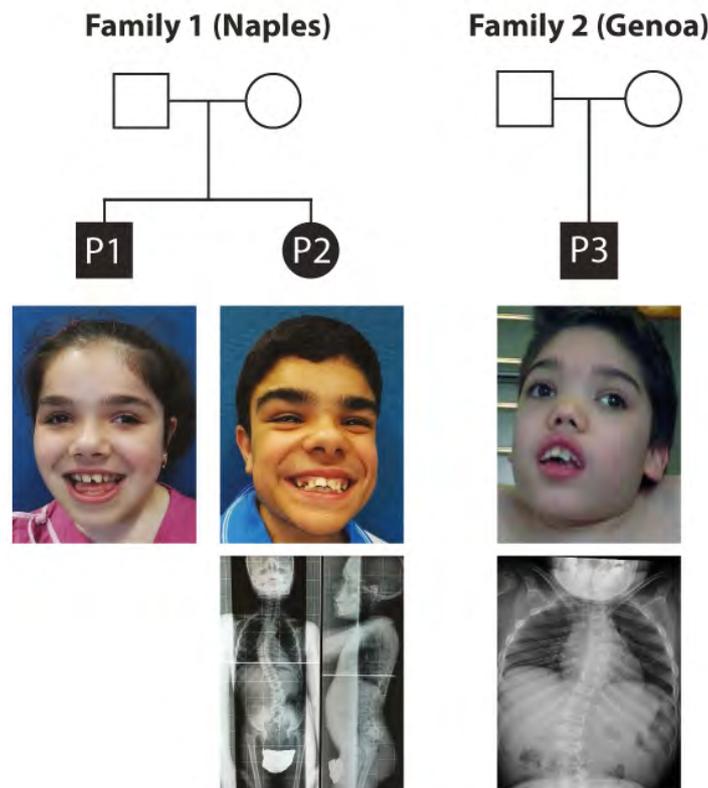
Cell lysates from cultured primary fibroblasts were prepared for each tested sample. After 3 washes with PBS, cells were scraped down from 100mm dishes at ~90% confluence, and were lysed using a the ReadUse™ mammalian cell lysis buffer 5X (AAT Bioquest). The total relative amount of both oxidized and reduced forms of nicotinamide adenine dinucleotides, NAD<sup>+</sup> and NADH, respectively, was measured using a luminescence assay, the NAD/NADH-Glo™ Assay kit (Promega). The values of detected luminescence were normalized to first to the blank samples of the same experiment, and then to the

protein content of each lysate, measured using the Pierce™ BCA Protein Assay Kit (Thermo Scientific). Luminescence and absorbance were measured using a Hidex Sense microplate reader.

## Results

### Clinical findings

Three patients from two families residing in two geographically distant cities, Naples and Genoa, presented with similar clinical features, altogether seemingly never previously described, and therefore representing a putatively novel clinical entity. There was no reported consanguinity for any of the two parental unions. All three patients showed early onset retinal degeneration, diagnosed as Leber congenital amaurosis (LCA), sensorineural hearing loss, short stature due to a spondylo-epiphyseal skeletal dysplasia, and additional neurodevelopmental delay, attributable to brain anomalies that were identified through magnetic resonance imaging, i.e. delayed myelinization, white matter hyperintensity, and cerebellar hypoplasia.



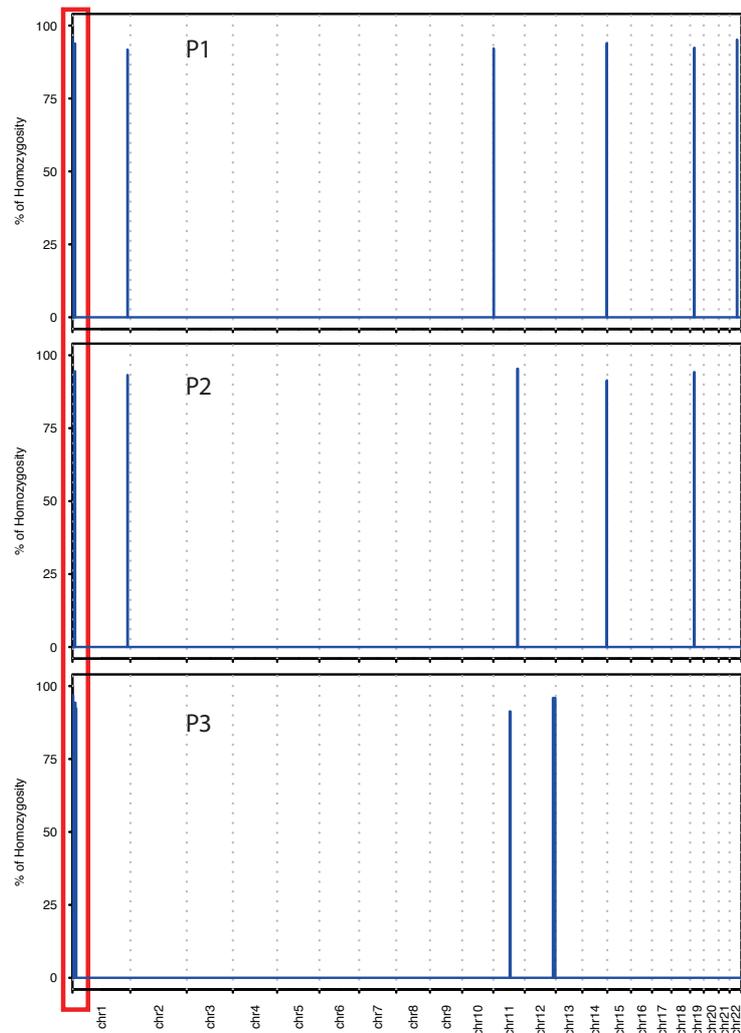
**Figure 2** : Pedigrees and phenotypes of the three affected children of this study.

The pictures show two characteristic phenotypes of these patients: dysmorphic facial features and scoliosis.

### Autozygome-based analysis reveals a shared homozygous interval

Analysis of the WES data resulted in no candidate rare variant shared by all three patients. Moreover, no rare insertion/deletion or change in copy number were found in any of the exome data from the two families. Interestingly, using autozygome-based analysis we could find a shared homozygous 4.6 Mb region on the short arm of chromosome 1 that was common to all three patients (Figure 3). Furthermore, we found that the haplotype was identical in the two families, indicating the presence of a shared ancestor. This genomic fragment comprises 40 genes, none of which harboured mutations or rare variants in coding regions. Among them we could find genes putatively relevant to the phenotype investigated: *CAMTA1*, associating with cerebellar ataxia with mental retardation (OMIM: 614756); *RERE*, associating with neurodevelopmental disorder with or without anomalies of the brain, eyes or heart (OMIM: 616975); *NMNAT1*, involved in leber congenital amaurosis (OMIM: 608553); and *CTNNBIP1*, with to date no human disease association, but known to cause short stature in mice when overexpressed<sup>62</sup>. Of those, only *NMNAT1* is known to associate with a recessive mode of inheritance, the mode of genetic transmission likely occurring in the two families of this study.

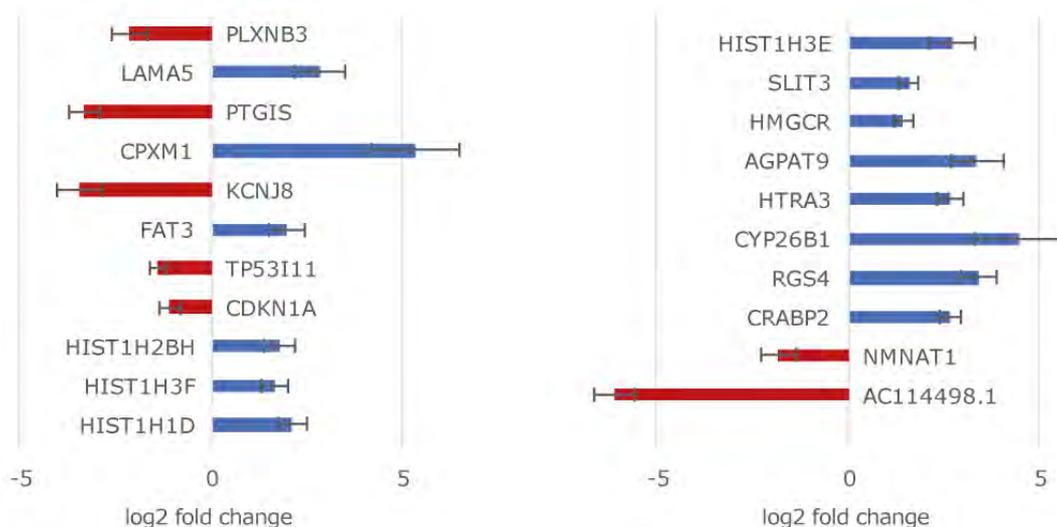
**Figure 3 :** Homozygosity mapping of the three patients. Total autozygous regions in P1, P2, and P3, span 8.51 Mb, 9.52 Mb, and 29.01 Mb, respectively. The region containing the shared interval of 4.6 Mb in chromosome 1 is highlighted in red.



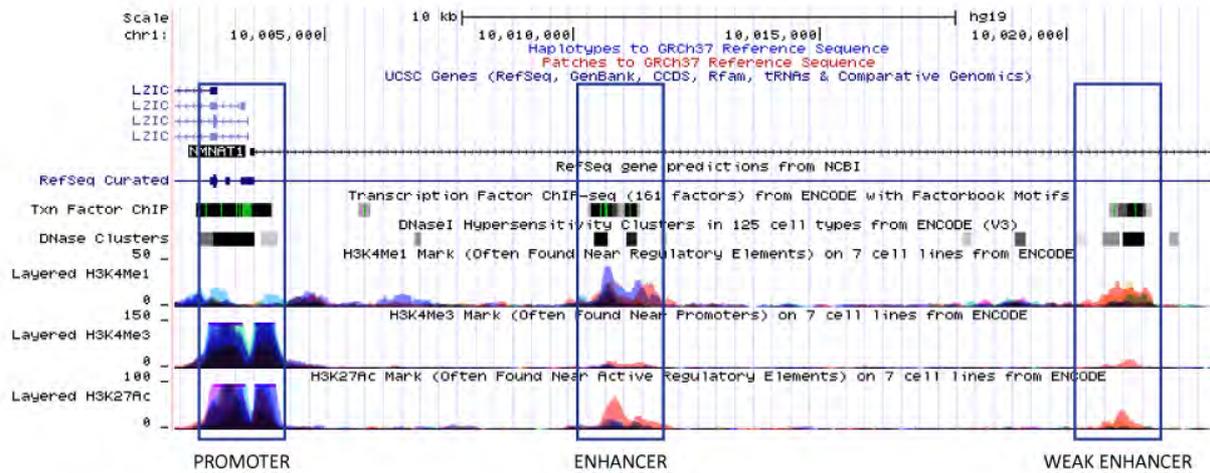
### ***NMNAT1* presents reduced expression levels compared to controls**

RNA sequencing analysis showed that 21 transcripts were significantly differentially expressed (Figure 4) in our three patients compared to 28 healthy controls. Among them the LCA-associated gene *NMNAT1*, the only one residing in the autozygous interval shared by the three patients. Statistical analysis proved significant downregulation by a 5.4-fold factor (p-value  $1.3 \times 10^{-4}$ ). Following this finding, we questioned ourselves about the reason of this reduced expression, given that the WES data did not reveal any candidate variant, nor we could see any evident abnormalities in coverage. Thus, we performed Sanger sequencing on *NMNAT1* regulatory regions (Figure 5), verifying sequences of the promoter and two upstream enhancer segments but, again, no anomalies were found. Similarly, we could attest that the close CTCF upstream insulator was intact, proving unlikely any disruption of the topologically associated domain (TAD) in this region.

Moreover, among the differentially regulated genes, we observed significant reduction in the expression of the gene encoding for a ATP-dependent potassium channel subunit, *KCNJ8*. This gene has been associated to Cantu syndrome (hypertrichotic osteochondrodysplasia, OMIM 239850) with, at a first glance, similar features to our patients' manifestations (skeletal dysplasia, coarse facial features) and additional symptoms such as a general increase in body hair, cardiomegaly, macrocephaly, and vascular features). Similarly to *NMNAT1*, this gene was negative for mutations, as observed in the WES data and confirmed using Sanger sequencing on coding regions.



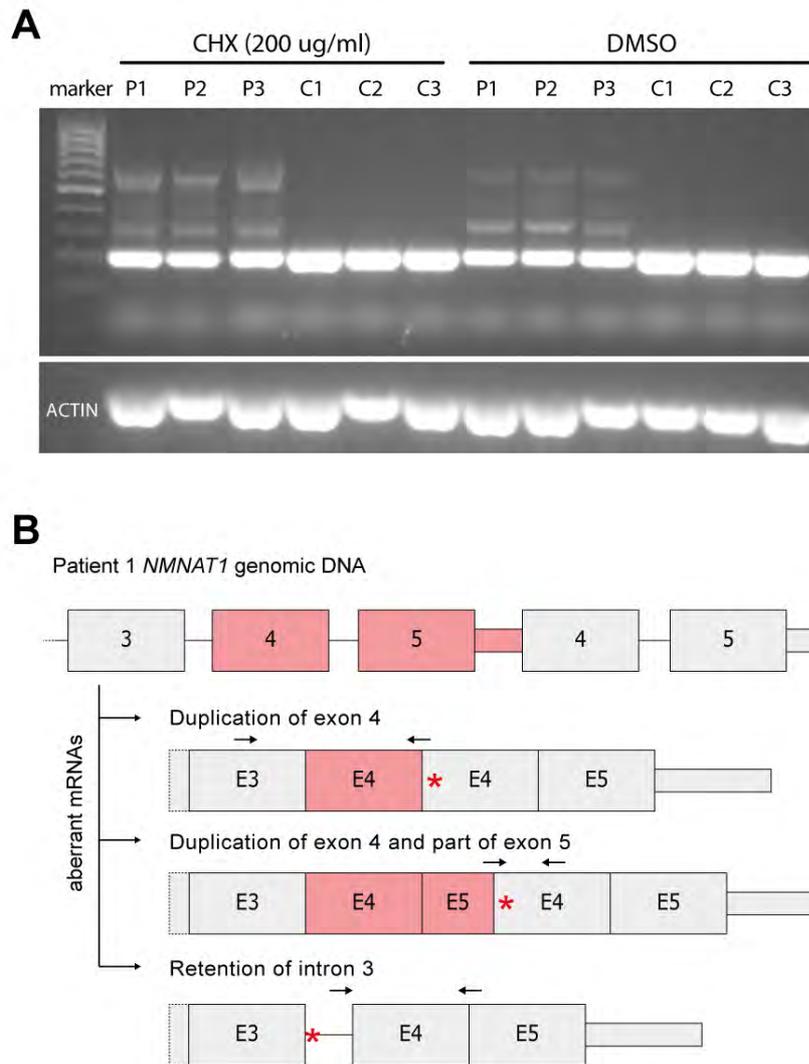
**Figure 4** : Results of the RNA sequencing. 21 significantly up and downregulated genes in the three patients compared to 28 healthy control primary fibroblasts. *NMNAT1* is the only gene resulting differentially regulated and residing in the shared homozygous interval.



**Figure 5** : UCSC visualization of the interval comprising three *NMNAT1* regulatory elements (genomic coordinates from human reference build 37: chr1:10,002,000-10,023,000). Chromatin immunoprecipitation and acylated or methylated lysine 4 and 27 of histone 3 show presence of transcription factors binding sites. The first region, comprising exon 1 of *NMNAT1*, clearly shows H3K4Me3 marks indicative for a promoter region, the second and third regions have higher occurrence for H3K4Me1 and H3K27Ac marks, suggestive for enhancer regions or other regulatory elements (data from the ENCODE project<sup>63</sup>).

### RT-PCR reveals patient-exclusive aberrant isoforms of *NMNAT1*

Validation of the RNA sequencing results using RT-PCR, brought to our evidence the presence of aberrant isoforms of *NMNAT1* mRNA. Sanger sequencing proved that these transcripts harboured either partial retention of intron 3, duplication of exon 4 or duplication both exon 4 and part of exon 5. The part of exon 5 retained consists in the first 278 bases, until the weak donor splice consensus “ttgGTACCA”. Human Splice Finder assigns a score of 70 (ranges between 0 and 100), and MAXEntScan calculates a score of 4.88 using the maximum dependence decomposition model (ranges between -10 and 10). Quantitative RT-PCR with wildtype- and abnormal isoform- specific primers served to validate the RNA sequencing results and to prove that the aberrant isoforms are a consequence of the duplication and are present in patients only (Figure 6B).

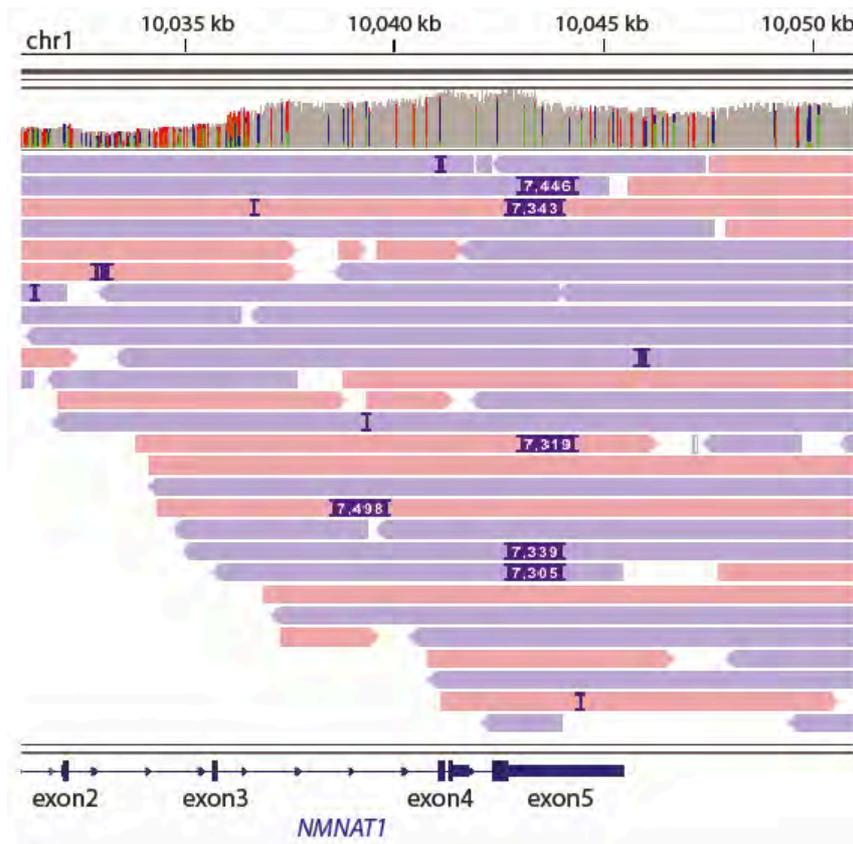


**Figure 6** : Patients present aberrant *NMNAT1* splice products. (A) Semi quantitative RT-PCR using cDNA from the three patients and three control fibroblast cell lines. On the left, cells treated with 200 ug/ml cycloheximide (CHX) on the right the control treatment with DMSO. PCR products were run on a 1% agarose gel. Actin was used as PCR control. (B) Model representing the rearrangement after interpretation of the Sanger results on patients' cDNA. Only the aberrant splice products verified by Sanger are shown.

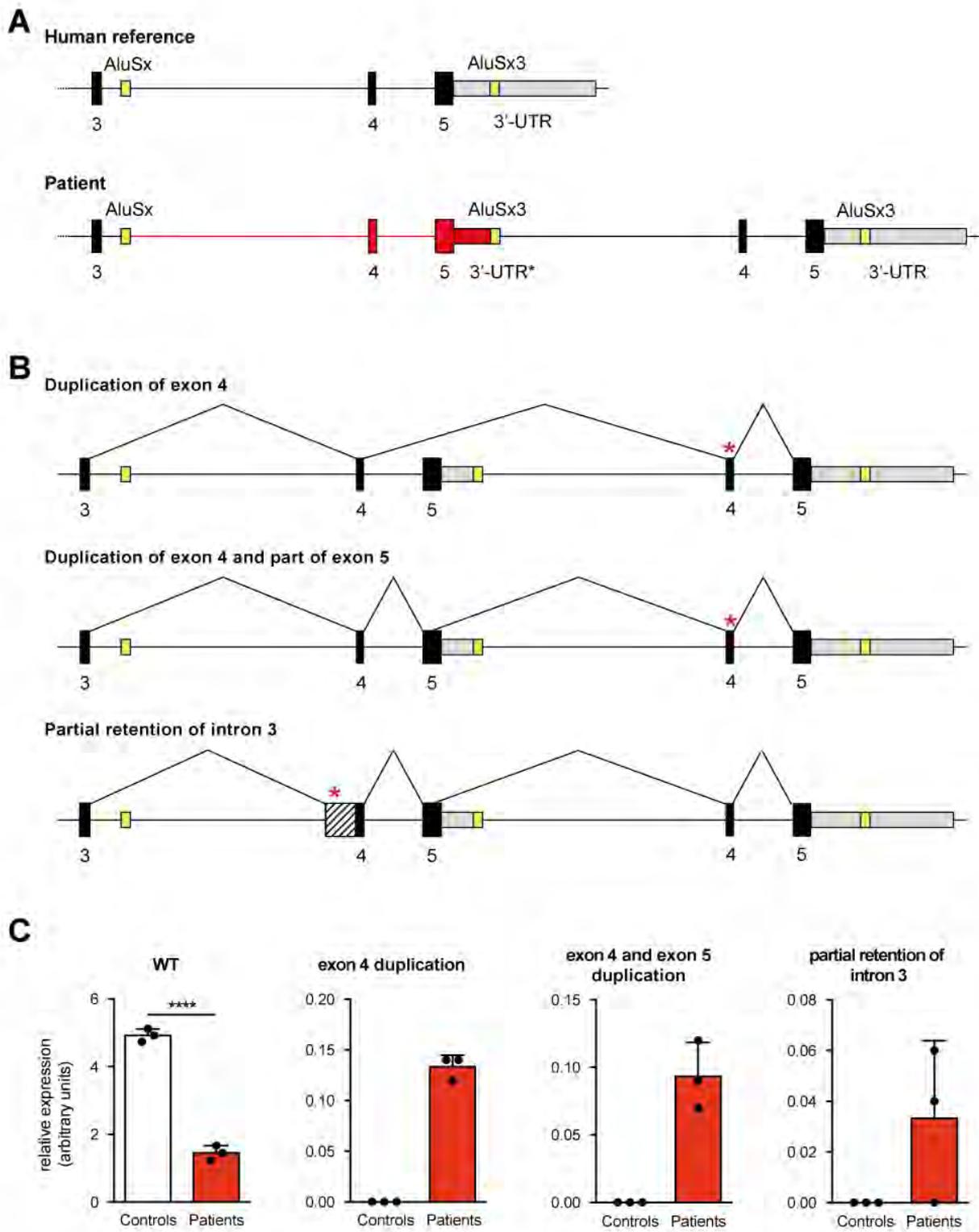
### Long reads whole-genome sequencing finds 7.4 kb duplication in *NMNAT1*

To precisely understand the rearrangement occurring in our patients, we decided to perform long-read WGS with Pacific Bioscience. WGS with long reads allowed us to define the exact boundary of the 7.4 kb duplication, involving the two last exons of *NMANT1*, and

spanning the beginning of intron 3 to the middle of the 3'-UTR. Interestingly, the duplicated fragment flanks two *Alu* elements : *AluSx* and *AluSx3* (genomic coordinates are 10,036,063-10,036,363 and 10,043,431-10,943,727, respectively). Their sequences, approximately 300 bp long, share about 88 % of homology. It is therefore likely that this tandem duplication event was mediated by recombination between these two *Alu* elements. Moreover, genome-wide single nucleotide variants (SNVs) and structural variants (SVs) analysis of the WGS data excluded the presence of additional rare pathogenic variants imputable as pathogenic mutations.



**Figure 7 :** IGV visualization of the long reads generated by PacBio. Only the reads entirely spanning the duplicated interval show a 7.4 kb insertion. Alignment of the ‘insert’ (dark purple) to the reference human sequence proved it consists in a duplication. The read orientation is color-coded with light red for the forward strand and purple for reads mapping to the reverse strand.

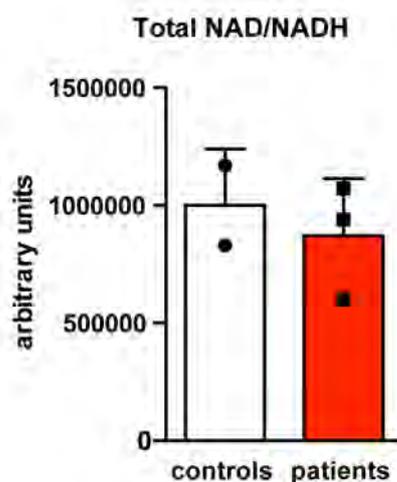


**Figure 8** : The structural variant in our patients and its consequences. (A) In-scale schematic diagrams of the reference vs patients' genomic *NMNAT1* sequences, with in red the duplicated fragment as confirmed by long read WGS. *NMNAT1* exons are numbered. (B) Schematic model of the aberrant *NMNAT1* splicing occurring in the patients. Red asterisks show approximate placement of the premature stop codon resulting from the sequence rearrangement. (C) Relative expression levels of the wildtype and of the aberrant mRNA

isoforms, after normalization in respect to three housekeeping genes. Each histogram was built with data from three controls and three patients. The first histogram, with values for the wild type isoform, validates the RNA sequencing results. The three other histograms clearly attest the patient-specific presence of three aberrant isoforms.

### **NAD production seems not to be impacted by the downregulation of *NMNAT1***

NMNAT1 involvement in NAD<sup>+</sup> biosynthesis brought us to investigate any change in NAD<sup>+</sup> or NADH levels. Using a luminescence assay, we could calculate the relative total cellular content of NAD and NADH in fibroblasts of patients *versus* controls. Measurements were done at different time points (20, 30, 45, 60 minutes), and the luminescence reaction resulted linear for both patients' and controls' cell extracts, as expected. The assay revealed 13% reduction of the total NAD<sup>+</sup>/NADH in patients. However, this trend was not significant using a standard t-test (p-value 0.6011).



**Figure 9 :** Total NAD<sup>+</sup>/NADH quantification assay in the three patients *versus* two control cell lines. On the y axis are represented values from the luminescence assay after normalization to the protein content of the lysate tested.

## Discussion and Perspectives

To date, the disease mechanism through which loss-of-function variants in *NMNAT1* cause LCA are not clear. In human, NMNAT1 together with NMNAT2 and NMNAT3 are responsible for the conversion of NaMN and NMN into NAD, an essential metabolite for a wide range of biological processes. Each of the NMNAT enzymes localizes to a specific subcellular compartment. NMNAT1 is a nuclear protein, whereas NMNAT2 and 3 are localized to the Golgi complex and the mitochondria, respectively<sup>64</sup>. The three isoforms are thought to have compartment-specific functions rather than functional redundancy, first of all because they present differences in their substrate specificities, and secondly because they are simultaneously expressed, as proven by their mRNA levels in two commonly used cell lines, HeLa and HEK293<sup>64</sup>.

Prior to its association with LCA, NMNAT1 has already been studied in *wld<sup>s</sup>* mice for its neuroprotective role in Wallerian degeneration. This latter, occurring after neuronal injury in the section of the axon distal to the lesion, is slowed down by overexpressing *Wld<sup>s</sup>*, a chimeric version of *Nmnat1*<sup>65-69</sup>. Therefore, NMNAT1 was considered of interest to study potential therapeutic targets in the NAD biosynthesis pathway to fight neurological disorders.

Interestingly, the *Nmnat1* knockout mice die before birth, indicating that the extranuclear NMNAT isoforms cannot compensate for its loss<sup>66,67,69</sup>. Nevertheless, mouse models carrying missense mutations recapitulate the human LCA phenotype<sup>70</sup>. Noteworthy, among the numerous reported human pathogenic *NMNAT1* variants, we never encounter homozygous disrupting mutations, confirming that the complete absence of NMNAT1 is not compatible with life. The results of our genetic analysis of the three affected children led us to hypothesize that an intermediate phenotype, between lethality and LCA, could be imputable to the structural disrupting variant detected.

### **A disrupting duplication in *NMNAT1* is associated with a novel syndrome**

The strong resemblance of phenotypes and the common haplotype in the three children are two robust arguments suggesting that the genetic cause of this novel syndrome is likely to be found in the shared interval of chromosome 1. The genomic rearrangement in the *NMNAT1* gene is certainly contributing to the disease manifestation. What needs to be clarified is whether this particular SV in *NMNAT1* is responsible for the whole syndromic symptoms, or if it is exclusively answering to the LCA phenotype. Being the only

significantly differentially regulated gene in this region, the *NMNAT1* duplication seems to be the only causative event. Another candidate resulting from the RNA sequencing analysis is *KCNJ8*. One particular variant in this gene was associated with the Cantu syndrome, a disorder partly resembling to the phenotype in our patients. However, *KCNJ8* was discarded for several valid genetic and clinical reasons. First, because the gene is not in the shared homozygous interval. Secondly, its association with Cantu syndrome consists in a single heterozygous variant responsible for autosomal dominant inheritance in one patient only, and none of the parents or other family members in our study seem to be affected. At the clinical level, the two phenotypes are dissociable. The osteochondro-dysplasia in Cantu syndrome and the spondylo-epiphyseal dysplasia in our patients, imply different impacts at the skeletal level. Children with Cantu syndrome have a rather tall stature compared to our patients. Also, intellectual development is almost normal or mildly affected in Cantu patients, which is not the case in the subjects of our study.

Furthermore, no additional rare SNV or SV were found in an exome-wide analysis of the WES and long-read WGS data. Altogether, these observations support the hypothesis that *NMNAT1* is the only causative gene.

### **The *NMNAT1* duplication results from an *Alu*-mediated recombination event**

The PacBio long-reads WGS allowed us to determine the precise breakpoint of the duplicated *NMNAT1* fragment, as well as the regions flanking this segment. The two *Alu* elements identified, *AluSx* and *AluSx3*, presenting with highly homologous sequences, are very likely involved in the mutational event at the origin of the duplication, creating an *AluSx-AluSx3* fusion element between the two copies of the duplicated fragment. Such rearrangement represents a non-allelic homologous recombination, consisting in a crossing-over occurring between non-equivalent sequences with high homology. This event can occur between sister chromatids in mitosis or between homologous chromosomes during meiosis, and can result in deletion or duplication events, or in more complex chromosomal rearrangements. To date, *Alu*-mediated recombination has been associated with numerous diseases<sup>71-74</sup>, including two LCA cases<sup>50</sup>, for whom were found *Alu*-mediated deletions, both coupled in compound heterozygosity with other missense changes.

Nowadays, with the use of high coverage WES, we can likely find rearrangements that have an impact on exonic regions, such as the duplication found in our patients.

Unfortunately, this was not the case for the WES data we could obtain, which did not reveal significant difference in coverage in patients *versus* controls.

### **NMNAT1 variants : lethality, LCA, and neural protection**

Several mutations in the *NMNAT1* gene have been shown to cause LCA<sup>44,45,49-51,53,54</sup>. However, no patient with disruptive homozygous variants has ever been described, except for one case: a LCA patient homozygous for a nonsense mutation (p.Trp169\*)<sup>45</sup>. The authors proposed that the resulting putatively truncated *NMNAT1* protein retains residual function, however, this hypothesis has never been experimentally validated. Concurrently, a recombinant *NMNAT1* with a substitution in this same codon (Trp169Ala) was reported to cause complete loss of NMNAT1 enzymatic activity *in vitro* in transformed Hela cell lines<sup>48,64,75</sup>. Yet, we do not know what phenotype would result in human with this same homozygous missense mutation.

Noteworthy, non-coding NMNAT1 variants were previously found in two LCA patients with homozygous single nucleotide substitutions in *NMNAT1* 5'UTR, causing decreased *NMNAT1* mRNA levels in patients' lymphocytes, and decreased enzymatic activity (luciferase assay) in human retinal pigment epithelial RPE-1 cells<sup>50</sup>. We hypothesised that, since in these patients the NMNAT1 protein sequence is preserved, residual wildtype function could be retained, limiting the phenotype to LCA only.

Since the retina is considered to be one of the most active metabolic tissues of the human body<sup>76</sup>, it could be that retinal degeneration takes place as a first manifestation of mild *NMNAT1* loss-of-function mutations, whereas hypothetically a stronger impact could be responsible for the appearance of additional, and more severe phenotypes (Figure 10). Another important consideration regarding the severity of the mutation, is that the *Nmnat1* KO mouse dies embryonically, proving that NMNAT1 is critical for viability<sup>68</sup>. In addition, this lethality demonstrates that NMNAT2 and NMNAT3 cannot compensate the loss of NMNAT1 function.

Both *in vitro* and *in vivo* mouse studies on the Wallerian degeneration mouse model *wld<sup>S</sup>*, overexpressing a fusion protein that incorporates *Nmnat1* and the ubiquitination factor *Ube4b*, proved NMNAT1 to have a beneficial effect in neuronal protection after injury<sup>75</sup>. Additionally, it has been demonstrated that *in vitro* inactivation of NMNAT1 enzymatic activity in the *wld<sup>S</sup>* mouse model (using the *Nmnat1* recombinant construct with the inactivating mutation Trp169Ala), causes the loss of the neuroprotective effect<sup>75,77</sup>.

In contrast, it has been reported that overexpression of *Nmnat1* in wildtype mouse has no neuroprotective effect<sup>69</sup>. Altogether, these results suggest that the beneficial outcome of *Nmnat1* expression might be simply linked with its function in the production of NAD, however it remains specific to the fusion protein expressed by the *Wld<sup>S</sup>* mice.

On the other hand, we showed that whole cell extracts of patients' *versus* controls' fibroblasts do not have significantly lower quantities of total NAD/NADH. On this same note, it has been recently shown in mouse that modulation of the enzymatic activity by overexpressing or downregulating *NMNAT1*, was not associated with significant alterations of NAD levels<sup>78</sup>. Altogether, these data open the possibility for the presence of additional functions of *NMNAT1*, other than in NAD biosynthesis, that might be implicated in the disease development of our patients. Nonetheless, we cannot exclude that fluctuations of nuclear NAD could also play a role in the disease mechanism.

Moreover, *in vitro* analysis of synthetic *NMNAT1* enzymatic activity resulting from LCA-associated variants showed large variability: some variants producing almost complete inactivation (99.5% p.Arg66Trp) and other variants milder inactivation (18.9 p.Arg237Cys) of *NMNAT1*<sup>48,79</sup>. Interestingly, in line with our hypothesis, we could not find among the LCA cases reported in the literature any homozygous carrier for strong inactivating variants. This suggests that disease phenotypes driven by a defective *NMNAT*, may directly depend from the variant harboured in its sequence (Figure 10).

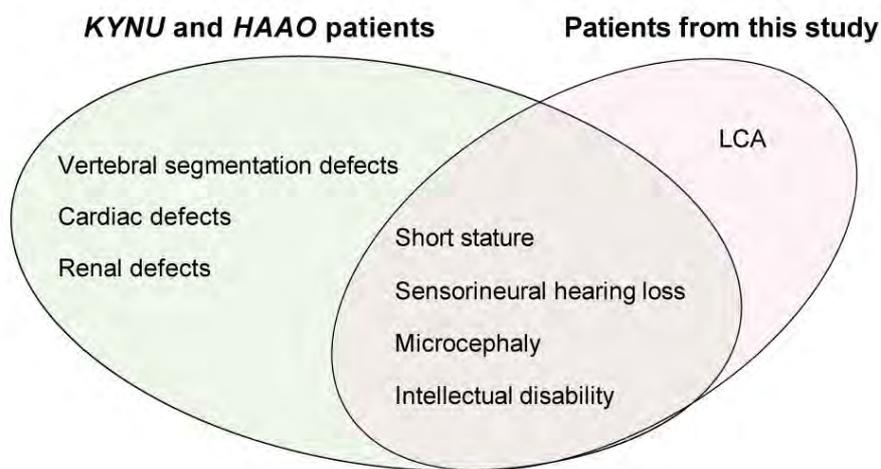
NMNAT1	wildtype	missense	downregulation	absence
protein function	normal	reduction	strong reduction	no activity
phenotype	healthy	LCA	syndromic disorder	non viable

**Figure 10** : Schematic representation of our working hypothesis on the genotype-phenotype relationship driven by different *NMNAT1* mutations.

### NAD deficiency and congenital malformations

In 2017, described three patients with homozygous variants in two genes from the Kynurenine pathway, *HAAO* and *KYNU*, genes responsible for the *de novo* synthesis of NAD from tryptophan (Figure 1). Patients showed phenotypes with variable impact at different levels, somehow similar to those identified by us: vertebral segmentation defects, short stature (skeletal dysplasia), sensorineural hearing loss, microcephaly, intellectual

disability, as well as cardiac and renal defects (Figure 11). Interestingly, they demonstrated that NAD levels in patients and in mouse models with loss-of-function *KYNU* and *HAAO* were reduced, concomitantly with a reduced *in vitro* activity of both enzymes. We therefore hypothesised that the phenotype resemblance with our patients may rely on the common participation of *KYNU*, *HAAO* and *NMNAT1* in the biosynthesis of NAD, and that the syndrome in our patients is likely due to the deficiency of *NMNAT1* activity. Remarkably, in the attempt to compensate the compromised *de novo* synthesis in the mice models with defective kynurenine enzymes, Shi *et al.*<sup>80</sup> could prove that dietary niacin supplementation during gestation prevented the appearance of some signs of the syndrome.



**Figure 11 :** Overlapping phenotypes between congenital malformations due to NAD deficiency observed in *KYNU* and *HAAO* loss-of-function patients and patients from our study.

## Perspectives

In human, the final step of NAD biosynthesis is performed by the three isozymes, *NMNAT1-3*. Considering their different subcellular localization and differences in structure and catalytic properties, these enzymes might be subject to differential regulation. Moreover, these enzymes could answer to the need of NAD in a compartment-specific manner. As we verified NAD production in our patients in whole cell lysates, it would be interesting to study the subcellular levels of NAD. In this way, we could have a better insight into the nuclear *NMNAT1*-dependent production of NAD in patients *versus* controls and possibly we would be able to clarify if the phenotype in our patients is due to the lack

of the nuclear production of NAD or to another yet unknown function of NMNAT1. Furthermore, we should evaluate NMNATs enzymatic activities by providing different quantities of NMN, the NAD precursor, in cell nuclear and cytoplasmic extracts. Comparatively, NAD levels and NMNAT1 activity of the homozygous Trp169\* patient would answer to the question whether the cause of LCA and/or syndromic features is related to a change in NAD production. Finally, it would be interesting to generate the same duplication in a mouse model in order to see if it phenocopies the human disease. Finding other families with *NMNAT1* variants with a similar impact to what we observe in our patients would be a final confirmation of our severe intermediate phenotype hypothesis. Nevertheless, we should consider that the identified genetic rearrangement in *NMNAT1* may contribute to the phenotype in our patients, due to the disturbance of biosynthesis of NAD<sup>+</sup>, although we ignore if NMNAT1 takes part to other to date unknown processes.

## **PROJECT 4**

### **NEXOME: A whole-genome study of Negative EXOMES**

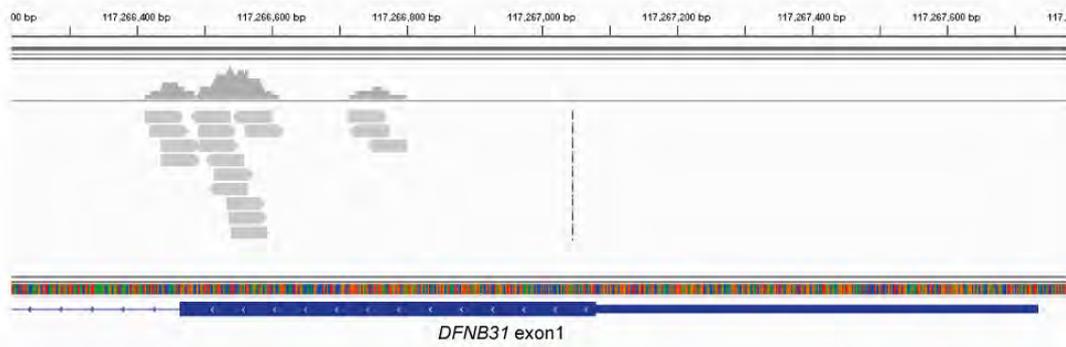
#### **Contributions**

This project consists in a large collaboration with numerous other research groups from the European Retinal Disease Consortium (ERDC), with important contributions also from other colleagues of Prof Rivolta's lab. I personally took part to the project from the beginning and later took charge of directing the analysis. Together with Beryl Royer-Bertrand we initially analysed all WES data of negative exomes. Afterwards, I carried out most of the analysis of WGS data, I analyzed all structural variants, performed and explored the autozygome, and discussed with the ERDC members about the possible candidate genes. Other lab members participated in the analysis of the output files of the WGS pipeline.

#### **Introduction**

This work aimed at analyzing whole genome datasets of patients originating from different countries across Europe and worldwide (ERDC consortium), initially screened and found to be negative following WES analysis. There are three main possible explanations for a WES investigation to miss disease-causative variants: i) insufficient coverage, ii) inefficient mapping and variant calling, and iii) presence of the primary mutation event in non-coding regions.

The coverage is defined by the average number of times that each nucleotide is expected to be sequenced given a certain number of reads mapping to a genomic region. Depending on the technique used and the properties of the DNA sequence (e.g: GC content) the coverage of WES might be insufficient.

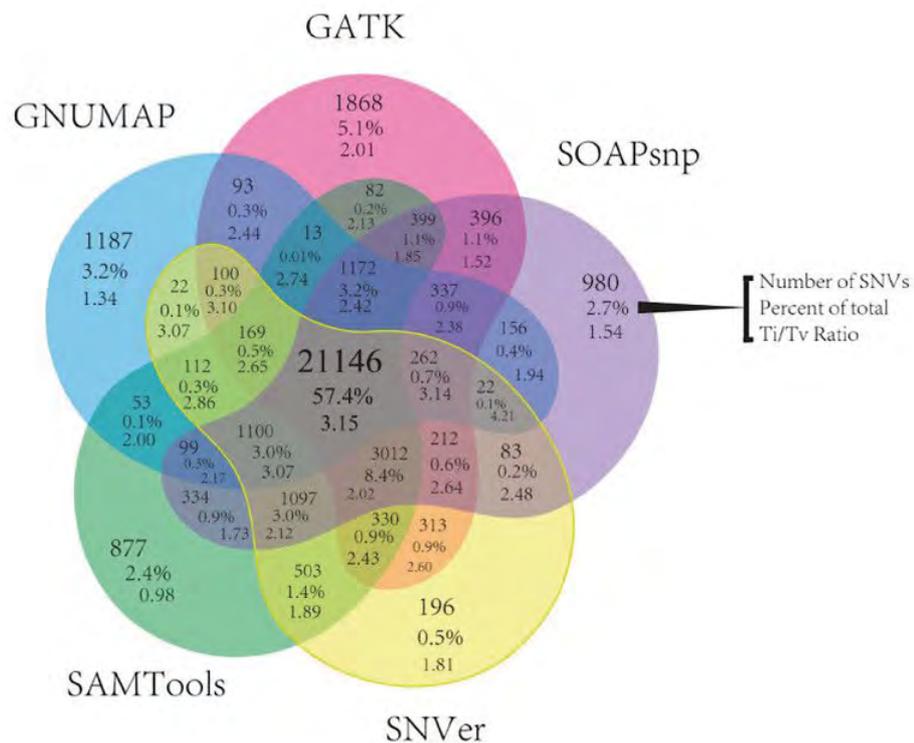


**Figure 1** : Example of insufficient coverage as visualized on the IGV software. A big portion of exon1 of the *DFNB31* gene has few or no reads (grey blocks) covering the sequence. In this specific example, a pathogenic variant was subsequently found in the portion of *DFNB31* that was not covered.

Another explanation for “lost” candidate variants was pointed out by O’Rawe and colleagues<sup>81</sup>, who tested five different mapping and variant calling pipelines and found that only about 57% of called variants were concordant (Figure 2). On the other hand, 0.5 to 5% of the variants were unique to each pipeline. Altogether these findings demonstrate that caution should be taken in interpreting negative findings, as variants might be missed depending on the mapping and variant calling software used, as well as false positive calls might be more or less retained.

Theoretically, if there are no mistakes due to mis-diagnosis or inaccuracy in DNA samples handling, there is one last reason for which no pathogenic variant is detected in any WES study: the variant can be located in the remaining 98% of the patient’s genome as a deep intronic, intergenic or structural variation.

An estimated 8% of the general population harbor in their genome a large (>500 kb) deletion or duplication with at an allele frequency of < 0.05%<sup>82</sup>. There is growing evidence that the missing heritability of genetic diseases (not only IRDs), might rely on yet undisclosed associations with structural or non-coding variations of the genome<sup>83-87</sup>.



**Figure 2** : Comparison of five mapping and variant calling pipelines and determination of the mean SNV concordance between each pipeline. The testing data consists in 15 WES. Strikingly, only about 57 % of the variants are common.<sup>81</sup>

## Methods

### Project design

This project was performed in collaboration with members of the European consortium ERDC. WES data from patients with retinal degeneration, in particular autosomal recessive and X-linked RP and Leber congenital amaurosis (LCA), were provided to us as raw fastq reads. After initial reanalysis using our internal WES pipeline, we considered patients with no exonic, disease-causative, candidate variants (Negative EXOME) for further investigation using sequencing of the whole genome.

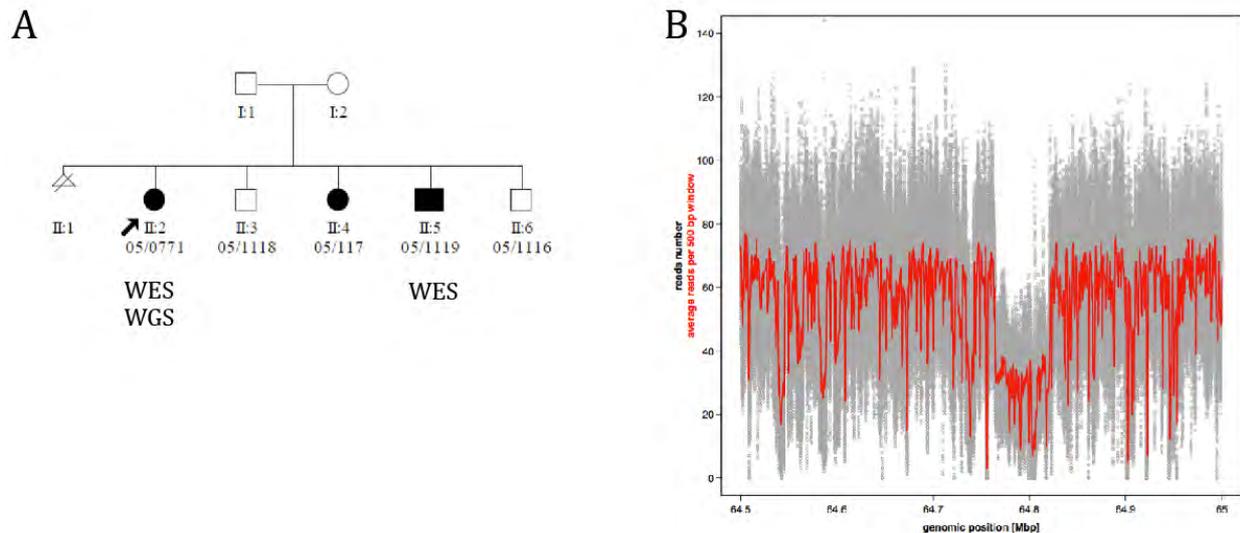
### Whole Genome Sequencing

WGS was performed by Complete Genomics Inc. (Mountain View, CA, USA), as described previously<sup>88</sup>. Genetic variants were identified using v2.0 of the Complete Genomics pipeline. The pathogenicity of genetic variants detected through WGS and WES was assessed after functional annotation through ANNOVAR<sup>58</sup>.

## Results

### Compound heterozygous non-canonical mutations in the *EYS* gene

We examined exomic and genomic data from a Spanish family composed of three siblings with arRP. Two of them were initially analyzed by WES with no genetic finding. What emerged from the subsequent WGS of one of the affected siblings is a 56kb heterozygous deletion (chr10:64'764'235-64'820'591; hg19), removing two exons of the *EYS* gene, a well known retinal degeneration associated gene.



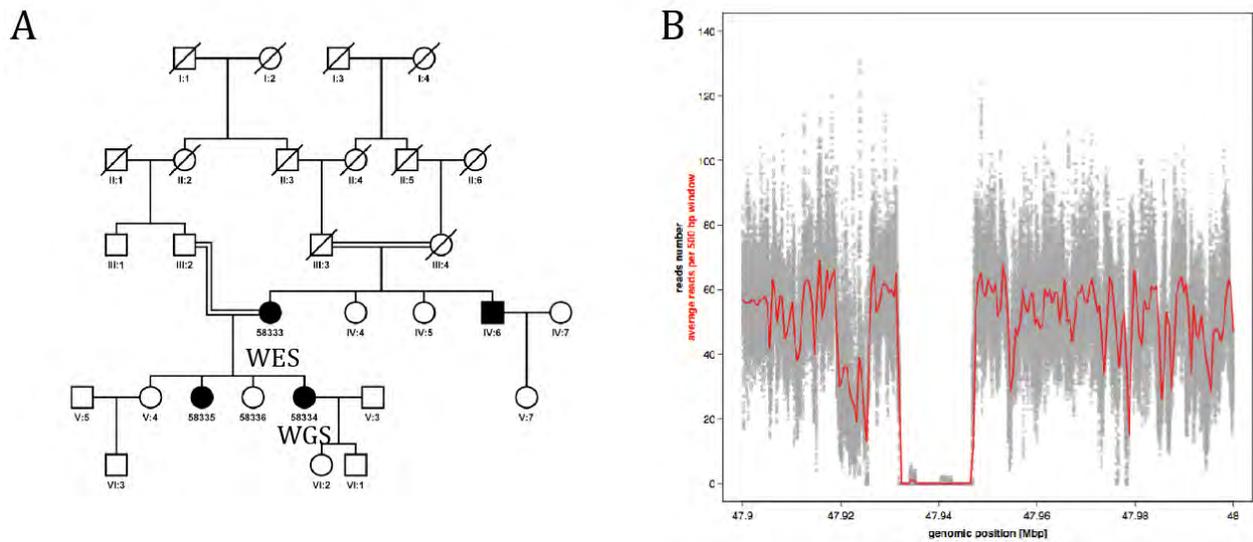
**Figure 3** : (A) pedigree of the family with arRP and (B) plotted coverage in the genomic area of *EYS*. In grey is the number of reads mapping to each nucleotide position, in red the average number in a window of 500 bp.

In addition, we found a rare non-synonymous heterozygous variant (p.Tyr2972Asp), that escaped the WES analysis because the gene was insufficiently covered. The two variants segregated correctly within the family members, confirming the recessive mode of inheritance of these compound heterozygous variants.

### *CNGA1* homozygous deletion

Another family with arRP (Figure 4) was found with a large homozygous deletion in the *CNGA1* gene (chr4: 47,931,964-47,946,798; hg19). The deletion of 44.8 kb fully removes five exons. Autozygome analysis revealed several regions of homozygosity, *CNGA1* lying in one of them. The affected mother's WES data were re-assessed and was indeed found to be sufficiently covered in this region. Although, supposedly, the mother is carrier of at least

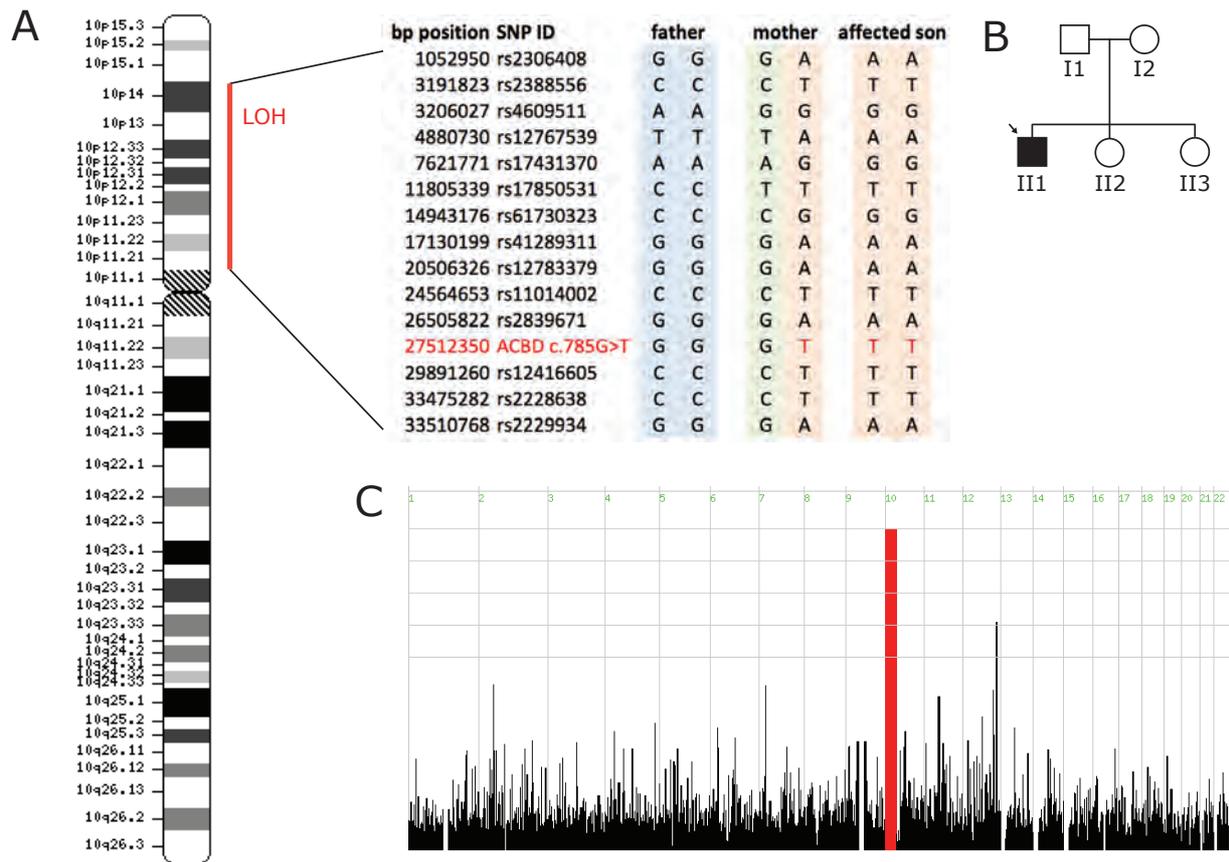
one healthy copy of the gene, she is affected with RP. By looking at the pedigree, and considering the consanguineous relatedness of two of the couples of this family, we might consider the hypothesis of a second gene involved in the disease. In fact, a recent study on a large Spanish cohort of retinal dystrophy cases estimated that about 1% of the families show intrafamilial heterogeneity of mutations contributing to the disease<sup>89</sup>. We therefore should consider reanalyzing the data looking for other possible pathogenic mutations.



**Figure 4 :** Similar to Figure 3, this family is also affected with arRP and presents consanguineous unions, as the underlying cause of the emergence of the disease.

### A case of uniparental isodysomy

We performed homozygosity mapping on the SNP data extracted from the WGS of this LCA case and found a large region of autozygosity (chr10:113'934-36'788'039). The parents were not declared as consanguineous. Deeper analysis of the SNP data in this region and flanking areas revealed that the patient was carrying a duplicated portion of chr10 of his mother, whereas the copy of the father was missing. In this autozygous region we found a rare non-synonymous variant in the *ACBD5* gene (p.Val134Phe), affecting an evolutionarily conserved amino acid. This genomic region was not covered in the WES. *ACBD5* was previously reported as a putative candidate for syndromic forms of retinal degeneration<sup>90</sup>. Moreover, pathogenic mutations in this gene have recently been found from our collaborators in Nijmegen, confirming our finding in *ACBD5* as the causative gene in this case of uniparental isodysomy.



**Figure 5:** (A) in red is depicted the region with the loss of heterozygosity in the patient due to the isodysomy. Genotypes in this region belonging to the father, mother and affected son are listed. (B) pedigree of the family and (C) autozygome analysis depicting a unique region of homozygosity.

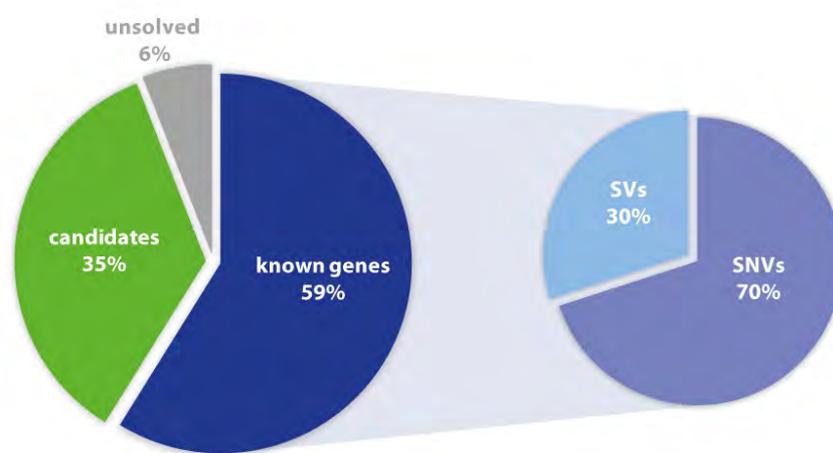
### Other findings

Several cases were solved with the use of WGS, while WES was not efficient enough because of bad coverage. Table 1 and Figure 6 show an overview of all additional findings in the examined samples. In total, we have 7 new candidate genes for novel disease-gene associations that need further investigation for validation through segregation and functional impact. Moreover, further confirmation of those candidate variants need to be obtained with other affected families that are carrier of mutations in these same genes or through functional biochemical validation *in vitro* and/ or *in vivo* in model animals.

**Table 1** : overview of the findings in the negative exome WGS study

WGScore	original_ID	diagnosis	finding	variant_type	category
NEX1	09/1325	CSNB	NYX	nonsense	known gene
NEX6	09/1168	arRP + hypoacusia	DFNB31	nonsense	known gene
NEX8	05/1117	arRP	EYS	missense + deletion	known gene
NEX9	04/0654	LCA	ACBD5	Missense + isodysomy	known gene
NEX12	3841	arRP	PDE6B	splice	known gene
NEX15	737-2	arUSH2A/RP	ARSG	missense	NEW GENE
NEX17	R220	arRP	CERKL	misstart	known gene
NEX18	R69	arRP xl?	NR2E3	missense + splice	known gene
NEX19	R240	arRP	TNXB	MEI	NEW GENE
NEX20	R240	arRP	CNGA3	missense	known gene
NEX28	46200	arRP	no candidates	-	unsolved
NEX29	46235	arRP	CACNA1S, EDEM3	missense	NEW GENE
NEX30	46238	microphthal	NDUFS3	missense	NEW GENE
NEX33	48774	arRP	RP1L1/TUB	missense	known gene
NEX34	58134	WMS	KCNQ5	missense	NEW GENE
NEX37	58334	arRP	CNGA1	deletion	known gene
NEX38	62494	arRP	CASKIN/CPN1	missense	NEW GENE

In table 1 are listed the findings in known gene, as well as candidate genes for novel disease-gene associations. Nevertheless, these latter group is still in the phase of being validated in collaboration with the ERDC research groups. To date, we could validate only one of the new disease gene candidates, the *ARSG* gene, with a rare, highly conserved variant in proband NEX15 (see Project 5 in the Appendix).



**Figure 6** : Graphical overview of the NEXOME findings. SV: structural variants, SNV: single nucleotide variants

## Discussion

Today's questions in genomic research have reached a wide complexity, making traditional DNA sequencing technologies (i.e. Sanger sequencing) inefficient and time-consuming. The NGS techniques, subject to constant improvements, have become an everyday research tool to address these questions.

A bias exists in NGS analyses, since often the main focus of these investigations concentrates on the coding regions of the genome, corresponding to only 2% of an individual's genetic makeup. With WGS we can gain a wider view and read every single base of the remaining 98% of the genome. Although there is still need for development of an effective pipeline for the analysis of all the genetic variations in a full genome, WGS can often 'easily' detect large structural events. This was the case for three of our families, described above.

Nevertheless, WGS helped finding the genetic causes of the disease in 20 of 32 cases (with known genes or candidate genes for novel disease-gene associations). From our results, we learned that quality control of WES data needs to be improved. Principally, we need to develop a systematic and automated validation of the WES coverage in all genes already known to be associated with retinal degeneration, a tool that is being developed presently from our bioinformaticians, and that will be integrated in our pipeline.

Even after WGS, a number of cases remain unsolved. Where is the causative variant? Possible answers might be in deep intronic or intergenic regions. To date, with the enormous amount of data collected with the WGS, it is nearly impossible for us to verify all the rare intronic variants that are called after WGS. An effective filtering pipeline for variants in non-coding regions is nowadays extremely needed in the community of WGS users. Moreover, a contribution in determining the impact of a given variant in a disease might be offered by the use of other omics techniques, such as transcriptomics.

Similarly to SNVs in non-coding regions, we face a too high amount of candidate variants when we handle data for dominantly inherited diseases. In this case we might improve our analysis by including more family members in the sequencing process, solution that might reveal to be very expensive. On the other hand, for the study of recessive diseases, especially for cases presenting consanguinity, we can exploit the power of homozygosity mapping to considerably reduce the number of candidate variants.

At the present-day, significant challenges in NGS techniques persist; improvements in data processing and analysis are required. In particular, WES may be prone to technical errors during sequencing, leading to bad coverage, while other types of errors may be met

during data processing following mapping and variant calling. Noise reduction due to false positive calls needs also to be improved. Despite the fact that WES, compared to WGS, has a weaker sensitivity and inferior coverage, it is preferred, because of its reduced costs, as a first line analysis tool to investigate the genetics of an inherited disease. Finally, technologic improvement is allowing structural variant detection in WES data as well. The research community is constantly producing new tools for NGS data analysis, even for the identification of more complex variants in high coverage WES, such as for example HMZdelfinder<sup>91</sup>.

## General discussion and conclusive remarks

Despite each project has been comprehensively discussed in its section, a few considerations can be done on the general findings composing my thesis.

The main results of this work have led to novel disease-gene associations, in particular for syndromic ciliopathies. In the year 2000 was published the first landmark paper responsible for considering cilia more than just vestigial elements. In this paper, the authors described the intraflagellar transport core component *Ift88* to be essential for primary cilia formation in *Chlamydomonas reinhardtii* and in mouse<sup>92</sup>. This study provided the first evidence that intraflagellar transport was necessary for correct assembly and maintenance of the ciliary outer segment in vertebrates. Since then, a cascade of reports highlighted the importance of cilia function in development and tissue homeostasis. Given the broad expression and function attributed to this organelle<sup>42</sup>, it is not surprising that ciliary defects give rise to a multitude of pathologies with organ-specific functional defects. Ciliopathies are pathologically diverse, and can cause clinically mild to highly complex and severe multi-organ syndromes, which in some cases are incompatible with life, such as for example in Meckel-Gruber syndrome<sup>39,93,94</sup>. Most ciliopathies are linked to a dysfunctional primary cilium, leading for instance to defects in retina, renal function, hearing or bone development. On the other hand, there is a different category of ciliopathies that have no apparent defects in the primary cilium, whereas motile cilia are affected. This is the case of primary ciliary dyskinesia, manifesting with characteristic phenotypes such as situs inversus, male infertility due to immotile spermatozoa, and chronic airway diseases<sup>20,42</sup>.

It is interesting to note that our findings in *TLL5* and *ARL2BP* loss-of-function patients, with phenotypes in retina photoreceptors and spermatozoa, attest that some types of ciliopathies have an impact on both motile and immotile cilia. These results provide evidence of the existence of ciliary components, such as *TLL5* and *ARL2BP*, that play crucial roles in both subtypes of cilia. Accordingly, although cilia and flagella have a different morphology and functions, they may share a similar biochemistry.

We therefore contributed in broadening the spectrum of ciliopathic phenotypes, presenting diseases that are in between immotile- or motile-only ciliopathies. Regarding *TLL5* patients specifically, the infertility phenotype was not fully penetrant. After observation of the type of mutations present in our patient cohort, we could show that only male patients with truncating mutations had additional defects at the level of sperm. We therefore hypothesized an allele-specific effect, depending on the severity of the mutation at the

*TLL5* protein level. In order to consolidate our findings, it would be interesting to evaluate the spermogram of other patients with disruptive mutations in *TLL5*.

In the case of *ARL2BP*, we studied in parallel the phenotypes of human patients and of the mouse KO model. To date, only one human case was reported to have situs inversus as a consequence of a faulty *ARL2BP*, however, a high incidence of this laterality defect was seen in the mouse model, suggesting a role for *ARL2BP* also in embryological nodal cilia. Therefore, we might define this disease entity as an atypical primary ciliary dyskinesia with additional features of retinal degeneration.

Similarly, in the case of our third project, we studied a novel syndromic entity, characterized by severe phenotypes in multiple tissues and organs. We hypothesized that the structural variant found in the *NMNAT1* gene could be responsible alone for the unique phenotype. As previously discussed, this is a novel association, since in the past *NMNAT1* mutations were only associated with LCA.

In the fourth project, we performed a whole-genome sequencing investigation of unresolved, negative exomes, and we identified new structural variants in the *EYS* and *CNGA1* genes, as well as a large chromosomal rearrangement leading to uniparental isodysomy. In general, in our research approach we coupled different NGS techniques that allowed us to attribute definite pathogenic potential to small DNA variants as well as less canonical changes, such as structural variants.

The field of medical genetics is living a constant revolution thanks to both improvements in the NGS techniques, and the constant implementation of analytical pipelines. We are often biased in NGS analysis because it is more likely to find causative variants in coding regions. But there is growing interest in finding methods for assessing pathogenicity of genetic variations in the noncoding portion of the genome. Other than more complex, multifactorial inheritance, we could suppose that the missing heritability could be explained to a certain extent with variants in noncoding regions.

As a researcher, living in this generation is, using simple words, a thrilling moment. Thanks to NGS technologies we are indeed witnessing an unprecedented gain of knowledge. Gene therapy is maybe the best example of a groundbreaking outcome of the NGS era, providing potential treatment for diseases that have no other cures. Clinical trials

are presently being developed, with the aim of allowing blind patients to partly gain back their vision.

Generally, inherited diseases are numerous. Knowing the biological context of a genetic variation, has been proven to be a pivotal starting point for the development of better diagnosis and devising treatment modalities.

## References

1. Kolb, H., Nelson, R., Fernandez, E. & Jones, B. Webvision: The Organization of the Retina and Visual System. *Webvision* (2017).
2. Kolb, H. Gross Anatomy of the Eye. *Webvision Organ. Retin. Vis. Syst.* (1995). doi:NBK11534 [bookaccession]
3. Kolb, H. Glial Cells of the Retina. *Webvision Organ. Retin. Vis. Syst.* (2001). doi:NBK11516 [bookaccession]
4. Kolb. Simple Anatomy of the Retina. *Webvision Organ. Retin. Vis. Syst.* (1995). doi:10.1103/PhysRevE.74.031115
5. Kolb, H., Fernandez, E. & Nelson, R. Facts and Figures Concerning the Human Retina. *Webvision* (2007). doi:NBK11556 [bookaccession]
6. Hartong, D. T., Berson, E. L. & Dryja, T. P. Retinitis pigmentosa Prevalence and inheritance patterns. *Lancet* (2006). doi:10.1016/S0140-6736(06)69740-7
7. Masland, R. H. The Neuronal Organization of the Retina. *Neuron* (2012). doi:10.1016/j.neuron.2012.10.002
8. Bok, D. The retinal pigment epithelium: a versatile partner in vision. *J. Cell Sci.* (1993). doi:10.1242/jcs.1993.Supplement\_17.27
9. Osterberg, G. A. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmol.* (1935). doi:10.1212/WNL.57.4.632
10. Kolb, H. Cone Pathways through the Retina. *Webvision: The Organization of the Retina and Visual System* (2009). doi:10.1192/apt.10.3.171
11. Nelson, R. Visual Responses of Ganglion Cells. *Webvision Organ. Retin. Vis. Syst.* (2007). doi:10.1111/pers.12014
12. Kalloniatis, M. in *Webvision* (2012).
13. Yau, K. W. & Hardie, R. C. Phototransduction Motifs and Variations. *Cell* (2009). doi:10.1016/j.cell.2009.09.029
14. Yau, K. W. Phototransduction mechanism in retinal rods and cones. The Friedenwald lecture. *Investigative Ophthalmology and Visual Science* (1994).
15. Veleri, S. *et al.* Biology and therapy of inherited retinal degenerative disease: insights from mouse models. *Dis. Model. Mech.* (2015). doi:10.1242/dmm.017913
16. Travis, G. H., Golczak, M., Moise, A. R. & Palczewski, K. Diseases Caused by Defects in the Visual Cycle: Retinoids as Potential Therapeutic Agents. *Annu. Rev. Pharmacol. Toxicol.* (2007). doi:10.1146/annurev.pharmtox.47.120505.105225
17. Davis, E. E., Brueckner, M. & Katsanis, N. The Emerging Complexity of the Vertebrate Cilium: New Functional Roles for an Ancient Organelle. *Developmental Cell* (2006). doi:10.1016/j.devcel.2006.06.009
18. Yang, J. *et al.* Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *J. Cell Biol.* (2002). doi:10.1083/jcb.200207153
19. Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M. & Bhattacharya, S. S. Photoreceptor degeneration: Genetic and mechanistic dissection of a complex trait. *Nature Reviews Genetics* (2010). doi:10.1038/nrg2717
20. Leigh, M. W. *et al.* Clinical and genetic aspects of primary ciliary dyskinesia/kartagener syndrome. *Genetics in Medicine* (2009). doi:10.1097/GIM.0b013e3181a53562
21. 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. *Journal of Clinical Investigation* (2010). doi:10.1172/JCI42258
22. Chiang, J. P. W. & Trzuppek, K. The current status of molecular diagnosis of inherited retinal dystrophies. *Current Opinion in Ophthalmology* (2015).

- doi:10.1097/ICU.0000000000000185
23. Berger, W., Kloeckener-Gruissem, B. & Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Progress in Retinal and Eye Research* (2010). doi:10.1016/j.preteyeres.2010.03.004
  24. Roosing, S. *et al.* Causes and consequences of inherited cone disorders. *Progress in Retinal and Eye Research* (2014). doi:10.1016/j.preteyeres.2014.05.001
  25. Hamel, C., Hartong, D. T., Berson, E. L., Dryja, T. P. & Hamel, C. Retinitis pigmentosa. *Orphanet J. Rare Dis.* (2006). doi:10.1016/S0140-6736(06)69740-7
  26. Berson, E. L. Retinitis pigmentosa: The Friedenwald lecture. *Investigative Ophthalmology and Visual Science* (1993). doi:10.1016/j.materresbull.2013.06.075
  27. Li, Z. Y., Possin, D. E. & Milam, A. H. Histopathology of Bone Spicule Pigmentation in Retinitis Pigmentosa. *Ophthalmology* (1995). doi:10.1016/S0161-6420(95)30953-0
  28. Koenekeop, R. K. An overview of leber congenital amaurosis: A model to understand human retinal development. *Survey of Ophthalmology* (2004). doi:10.1016/j.survophthal.2004.04.003
  29. Kremer, H., van Wijk, E., Märker, T., Wolfrum, U. & Roepman, R. Usher syndrome: Molecular links of pathogenesis, proteins and pathways. *Human Molecular Genetics* (2006). doi:10.1093/hmg/ddl205
  30. Bonnet, C. & El-Amraoui, A. Usher syndrome (sensorineural deafness and retinitis pigmentosa): Pathogenesis, molecular diagnosis and therapeutic approaches. *Current Opinion in Neurology* (2012). doi:10.1097/WCO.0b013e32834ef8b2
  31. Katsanis, N., Lupski, J. R. & Beales, P. L. Exploring the molecular basis of Bardet-Biedl syndrome. *Hum Mol Genet* (2001). doi:10.1093/hmg/10.20.2293
  32. Katsanis, N. The oligogenic properties of Bardet-Biedl syndrome. *Hum. Mol. Genet.* (2004). doi:10.1093/hmg/ddh092
  33. Fliegau, M. Nephrocystin Specifically Localizes to the Transition Zone of Renal and Respiratory Cilia and Photoreceptor Connecting Cilia. *J. Am. Soc. Nephrol.* (2006). doi:10.1681/ASN.2005121351
  34. Otto, E. A. *et al.* Mutations in INVS encoding inversin cause nephronophthisis type 2, linking renal cystic disease to the function of primary cilia and left-right axis determination. *Nat. Genet.* (2003). doi:10.1038/ng1217
  35. Olbrich, H. *et al.* Mutations in a novel gene, NPHP3, cause adolescent nephronophthisis, tapeto-retinal degeneration and hepatic fibrosis. *Nat. Genet.* (2003). doi:10.1038/ng1216
  36. Hildebrandt, F. in *Genetic Diseases of the Kidney* (2009). doi:10.1016/B978-0-12-449851-8.00025-5
  37. Brooks, B. P. *et al.* Joubert Syndrome: Ophthalmological Findings in Correlation with Genotype and Hepatorenal Disease in 99 Patients Prospectively Evaluated at a Single Center. *Ophthalmology* (2018). doi:10.1016/j.ophtha.2018.05.026
  38. Hildebrandt, F., Benzing, T. & Katsanis, N. Ciliopathies. *N. Engl. J. Med.* (2011). doi:10.1056/NEJMra1010172
  39. Waters, A. M. & Beales, P. L. Ciliopathies: An expanding disease spectrum. *Pediatric Nephrology* (2011). doi:10.1007/s00467-010-1731-7
  40. Fliegau, M., Benzing, T. & Omran, H. When cilia go bad: Cilia defects and ciliopathies. *Nature Reviews Molecular Cell Biology* (2007). doi:10.1038/nrm2278
  41. Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J. B. & Bettencourt-Dias, M. Tracing the origins of centrioles, cilia, and flagella. *Journal of Cell Biology* (2011). doi:10.1083/jcb.201011152
  42. Basten, S. G. & Giles, R. H. Functional aspects of primary cilia in signaling, cell cycle and tumorigenesis. *Cilia* (2013). doi:10.1186/2046-2530-2-6
  43. Erzurumluoglu, A. M., Shihab, H. A., Rodriguez, S., Gaunt, T. R. & Day, I. N. M.

- Importance of Genetic Studies in Consanguineous Populations for the Characterization of Novel Human Gene Functions. *Ann. Hum. Genet.* (2016). doi:10.1111/ahg.12150
44. Koenekoop, R. K. *et al.* Mutations in NMNAT1 cause Leber congenital amaurosis and identify a new disease pathway for retinal degeneration. *Nat. Genet.* (2012). doi:10.1038/ng.2356
  45. Perrault, I. *et al.* Mutations in NMNAT1 cause Leber congenital amaurosis with early-onset severe macular and optic atrophy. *Nat. Genet.* (2012). doi:10.1038/ng.2357
  46. Henderson, R. H. *et al.* A rare de novo nonsense mutation in OTX2 causes early onset retinal dystrophy and pituitary dysfunction. *Mol. Vis.* (2009). doi:260 [pii]
  47. LØKEN, A. C., HANSSEN, O., HALVORSEN, S. & JØLSTER, N. J. Hereditary Renal Dysplasia and Blindness. *Acta Pædiatrica* (1961). doi:10.1111/j.1651-2227.1961.tb08037.x
  48. Falk, M. J. *et al.* NMNAT1 mutations cause Leber congenital amaurosis. *Nat. Genet.* (2012). doi:10.1038/ng.2361
  49. Chiang, P. W. *et al.* Exome sequencing identifies NMNAT1 mutations as a cause of Leber congenital amaurosis. *Nat. Genet.* (2012). doi:10.1038/ng.2370
  50. Coppieters, F. *et al.* Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5'UTR Mutations and Copy-Number Variations of NMNAT1. *Hum. Mutat.* (2015). doi:10.1002/humu.22899
  51. Nash, B. M. *et al.* NMNAT1 variants cause cone and cone-rod dystrophy. *Eur. J. Hum. Genet.* (2018). doi:10.1038/s41431-017-0029-7
  52. Corton, M. *et al.* Correction: Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. *PLoS ONE* (2016). doi:10.1371/journal.pone.0153121
  53. Wang, S., Zhang, Q., Zhang, X., Wang, Z. & Zhao, P. Clinical and genetic characteristics of Leber congenital amaurosis with novel mutations in known genes based on a Chinese eastern coast Han population. *Graefe's Arch. Clin. Exp. Ophthalmol.* (2016). doi:10.1007/s00417-016-3428-5
  54. Siemiatkowska, A. M. *et al.* Novel compound heterozygous NMNAT1 variants associated with Leber congenital amaurosis. *Mol Vis.* (2014).
  55. Jin, X., Qu, L. H., Meng, X. H., Xu, H. W. & Yin, Z. Q. Detecting genetic variations in hereditary retinal dystrophies with next-generation sequencing technology. *Mol. Vis.* (2014).
  56. Magni, G., Amici, A., Emanuelli, M., Raffaelli, N. & Ruggieri, S. Enzymology of NAD<sup>+</sup> synthesis. *Adv. Enzymol. Relat. Areas Mol. Biol.* (1999). doi:10.1002/9780470123195.ch5
  57. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* (2011). doi:10.1038/ng.806
  58. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* (2010). doi:10.1093/nar/gkq603
  59. Sahraeian, S. M. E. *et al.* Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat. Commun.* (2017). doi:10.1038/s41467-017-00050-4
  60. Kaisers, W. *et al.* Age, gender and UV-exposition related effects on gene expression in in vivo aged short term cultivated human dermal fibroblasts. *PLoS One* (2017). doi:10.1371/journal.pone.0175657
  61. Sedlazeck, F. J. *et al.* Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* (2018). doi:10.1038/s41592-018-0001-7
  62. Chen, M. *et al.* Inhibition of -catenin signaling causes defects in postnatal cartilage

- development. *J. Cell Sci.* (2008). doi:10.1242/jcs.020362
63. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012). doi:10.1038/nature11247
  64. Berger, F., Lau, C., Dahlmann, M. & Ziegler, M. Subcellular compartmentation and differential catalytic properties of the three human nicotinamide mononucleotide adenylyltransferase isoforms. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M508660200
  65. Coleman, M. P. *et al.* An 85-kb tandem triplication in the slow Wallerian degeneration (Wlds) mouse. *Proc. Natl. Acad. Sci.* (1998). doi:10.1073/pnas.95.17.9985
  66. Coleman, M. P. & Freeman, M. Wallerian degeneration, wld(s), and nmnat. *Annu. Rev. Neurosci.* (2010). doi:10.1146/annurev-neuro-060909-153248
  67. Avery, M. A., Sheehan, A. E., Kerr, K. S., Wang, J. & Freeman, M. R. Wlds requires Nmnatl enzymatic activity and N16- VCP interactions to suppress Wallerian degeneration. *J. Cell Biol.* (2009). doi:10.1083/jcb.200808042
  68. Conforti, L. *et al.* Reducing expression of NAD<sup>+</sup> synthesizing enzyme NMNAT1 does not affect the rate of Wallerian degeneration. *FEBS J.* (2011). doi:10.1111/j.1742-4658.2011.08193.x
  69. Conforti, L. *et al.* NAD<sup>+</sup> and axon degeneration revisited: Nmnat1 cannot substitute for WldSto delay Wallerian degeneration. *Cell Death Differ.* (2007). doi:10.1038/sj.cdd.4401944
  70. Greenwald, S. H. *et al.* Mouse Models of NMNAT1-Leber Congenital Amaurosis (LCA9) Recapitulate Key Features of the Human Disease. *Am. J. Pathol.* (2016). doi:10.1016/j.ajpath.2016.03.013
  71. Deininger, P. L. & Batzer, M. A. Alu repeats and human disease. *Molecular Genetics and Metabolism* (1999). doi:10.1006/mgme.1999.2864
  72. Garland, J. *et al.* Identification of an Alu element-mediated deletion in the promoter region of GNE in siblings with GNE myopathy. *Mol. Genet. Genomic Med.* (2017). doi:10.1002/mgg3.300
  73. Liccardo, R. *et al.* Characterization of novel, large duplications in the MSH2 gene of three unrelated Lynch syndrome patients. *Cancer Genet.* (2018). doi:10.1016/j.cancergen.2017.11.008
  74. Geoffroy, V. *et al.* Whole-genome sequencing in patients with ciliopathies uncovers a novel recurrent tandem duplication in IFT140. *Hum. Mutat.* (2018). doi:10.1002/humu.23539
  75. Araki, T., Sasaki, Y. & Milbrandt, J. Increased nuclear NAD biosynthesis and SIRT1 activation prevent axonal degeneration. *Science (80-. ).* (2004). doi:10.1126/science.1098014
  76. Niven, J. E. & Laughlin, S. B. Energy limitation as a selective pressure on the evolution of sensory systems. *J. Exp. Biol.* (2008). doi:10.1242/jeb.017574
  77. Jia, H. *et al.* Identification of a critical site in Wlds: Essential for Nmnat enzyme activity and axon-protective function. *Neurosci. Lett.* (2007). doi:10.1016/j.neulet.2006.11.067
  78. Rossi, F. *et al.* NAD-biosynthetic enzyme NMNAT1 reduces early behavioral impairment in the htau mouse model of tauopathy. *Behav. Brain Res.* (2018). doi:10.1016/j.bbr.2017.11.030
  79. Sasaki, Y., Margolin, Z., Borgo, B., Havranek, J. J. & Milbrandt, J. Characterization of Leber congenital amaurosis-associated NMNAT1 mutants. *J. Biol. Chem.* (2015). doi:10.1074/jbc.M115.637850
  80. Shi, H. *et al.* NAD Deficiency, Congenital Malformations, and Niacin Supplementation. *N. Engl. J. Med.* (2017). doi:10.1056/NEJMoa1616361
  81. O'Rawe, J. *et al.* Low concordance of multiple variant-calling pipelines: Practical implications for exome and genome sequencing. *Genome Med.* (2013).

doi:10.1186/gm432

82. Itsara, A. *et al.* Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* (2008). doi:10.1016/j.ajhg.2008.12.014
83. Van Schil, K. *et al.* Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genet. Med.* (2018). doi:10.1038/gim.2017.97
84. Jamshidi, F. *et al.* Contribution of noncoding pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. *Genetics in Medicine* (2018). doi:10.1038/s41436-018-0104-7
85. Gao, L. *et al.* Identifying noncoding risk variants using disease-relevant gene regulatory networks. *Nat. Commun.* (2018). doi:10.1038/s41467-018-03133-y
86. Stankiewicz, P. & Lupski, J. R. Structural Variation in the Human Genome and its Role in Disease. *Annu. Rev. Med.* (2010). doi:10.1146/annurev-med-100708-204735
87. Spielmann, M. & Mundlos, S. Looking beyond the genes: The role of non-coding variants in human disease. *Human Molecular Genetics* (2016). doi:10.1093/hmg/ddw205
88. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* (80-. ). (2010). doi:10.1126/science.1181498
89. Sánchez-Alcudia, R. *et al.* Contribution of mutation load to the intrafamilial genetic heterogeneity in a large cohort of spanish retinal dystrophies families. *Investig. Ophthalmol. Vis. Sci.* (2014). doi:10.1167/iovs.14-14938
90. Abu-Safieh, L. *et al.* Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res.* (2013). doi:10.1101/gr.144105.112
91. Gambin, T. *et al.* Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkw1237
92. Pazour, G. J. *et al.* Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene Tg737, are required for assembly of cilia and flagella. *J. Cell Biol.* (2000). doi:10.1083/jcb.151.3.709
93. Ramachandran, U., Malla, T. & Joshi, K. S. Meckel-Gruber syndrome. *Kathmandu Univ. Med. J.* (2006). doi:10.1007/s003810050198
94. Badano, J. L., Mitsuma, N., Beales, P. L. & Katsanis, N. The Ciliopathies: An Emerging Class of Human Genetic Disorders. *Annu. Rev. Genomics Hum. Genet.* (2006). doi:10.1146/annurev.genom.7.080505.115610

### Contributions to other projects

#### PROJECT 5 – ARTICLE

##### **A homozygous founder missense variant in arylsulfatase G abolishes its enzymatic activity causing atypical Usher syndrome in humans**

This article was published on January 4<sup>th</sup> 2018 in the Journal *Genetics in Medicine*.

##### **Contribution**

I initially analysed the WES and WGS data of the first proband of the study, and I first proposed the *ARSG* gene as candidate for the Usher syndrome in this patient. Thanks to this finding, the project evolved in Israel and Germany with more functional validations, also finding additional patients with mutations in this same gene.

# A homozygous founder missense variant in arylsulfatase G abolishes its enzymatic activity causing atypical Usher syndrome in humans

Samer Khateb, MD, PhD<sup>1</sup>, Björn Kowalewski, PhD<sup>2</sup>, Nicola Bedoni, MSc<sup>3</sup>, Markus Damme, PhD<sup>4</sup>, Netta Pollack, BSc<sup>1</sup>, Ann Saada, PhD<sup>5</sup>, Alexey Obolensky, PhD<sup>1</sup>, Tamar Ben-Yosef, PhD<sup>6</sup>, Menachem Gross, MD<sup>7</sup>, Thomas Dierks, PhD<sup>2</sup>, Eyal Banin, MD, PhD<sup>1</sup>, Carlo Rivolta, PhD<sup>3,8</sup> and Dror Sharon, PhD<sup>1</sup>

**Purpose:** We aimed to identify the cause of disease in patients suffering from a distinctive, atypical form of Usher syndrome.

**Methods:** Whole-exome and genome sequencing were performed in five patients from three families of Yemenite Jewish origin, suffering from distinctive retinal degeneration phenotype and sensorineural hearing loss. Functional analysis of the wild-type and mutant proteins was performed in human fibrosarcoma cells.

**Results:** We identified a homozygous founder missense variant, c.133G > T (p.D45Y) in arylsulfatase G (ARSG). All patients shared a distinctive retinal phenotype with ring-shaped atrophy along the arcades engirdling the fovea, resulting in ring scotoma. In addition, patients developed moderate to severe sensorineural hearing loss. Both vision and hearing loss appeared around the age of 40 years.

The identified variant affected a fully conserved amino acid that is part of the catalytic site of the enzyme. Functional analysis of the wild-type and mutant proteins showed no basal activity of p.D45Y.

**Conclusion:** Homozygosity for ARSG-p.D45Y in humans leads to protein dysfunction, causing an atypical combination of late-onset Usher syndrome. Although there is no evidence for generalized clinical manifestations of lysosomal storage diseases in this set of patients, we cannot rule out the possibility that mild and late-onset symptoms may appear.

*Genet Med* advance online publication 4 January 2018

**Key Words:** arylsulfatase G; lysosomal storage disease; retinitis pigmentosa; Usher syndrome; whole-exome sequencing

## INTRODUCTION

Usher syndrome (USH) is the leading cause of inherited deaf-blindness (MIM 268000). It is an autosomal recessive disease with symptoms including visual disturbance due to retinal degeneration and sensorineural hearing loss (SNHL) accompanied in some cases by vestibular areflexia. The prevalence of USH has been estimated as 3.5–16.6 in 100,000 in several populations.<sup>1–3</sup> USH has historically been classified into three subtypes according to the age of onset of the visual and auditory symptoms, the severity of hearing loss, and the presence of vestibular involvement: Usher type 1 (USH1) is characterized by congenital SNHL with vestibular areflexia and retinitis pigmentosa (RP) onset in the first decade of life; USH2 by mild to moderate nonprogressive SNHL with RP onset in the first or second decade of life and normal vestibular function; and USH3 by progressive SNHL, sometimes accompanied by vestibular involvement, with a variable time of onset of both SNHL and RP. However, cumulative data have identified a large number of families with

phenotypes that are not compatible with any of the three established types and are therefore considered as having atypical forms of USH.<sup>4–6</sup>

USH is highly heterogeneous, not only clinically but also genetically. Thirteen causative genes have been identified to date, and three more loci are suspected to include additional genes (RETN database at <https://sph.uth.edu/RetNet/>). The development of high-throughput sequencing techniques, including whole-exome sequencing (WES) and whole-genome sequencing (WGS), has contributed tremendously to the identification of novel causative mutations and genes for rare inherited diseases in general and retinal degeneration in particular.<sup>7,8</sup> Nevertheless, despite these advances, the causative gene remains unknown in a large number of cases with inherited retinal disease (IRD).<sup>9,10</sup> Some of the proteins encoded by USH-associated genes interact in complexes that are localized to hair cells of the inner ear and in retinal photoreceptors.<sup>11–13</sup> USH-related proteins perform a variety of cellular functions, including myosin-based motor transport

<sup>1</sup>Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; <sup>2</sup>Department of Chemistry, Biochemistry I, Bielefeld University, Bielefeld, Germany; <sup>3</sup>Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Lausanne, Switzerland; <sup>4</sup>Department of Biochemistry, University of Kiel, Kiel, Germany; <sup>5</sup>Monique and Jacques Roboh Department of Genetic Research and the Department of Genetic and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; <sup>6</sup>Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel; <sup>7</sup>Department of Otolaryngology—Head and Neck Surgery, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; <sup>8</sup>Department of Genetics and Genome Biology, University of Leicester, Leicester, UK. Correspondence: Thomas Dierks or Eyal Banin or Carlo Rivolta or Dror Sharon (thomas.dierks@uni-bielefeld.de or banine@cc.huji.ac.il or carlo.rivolta@unil.ch or dror.sharon1@gmail.com) The last four authors contributed equally to this work.

Submitted 1 September 2017; accepted 6 November 2017; advance online publication 4 January 2018. doi:10.1038/gim.2017.227

(MYO7A); cell–cell adhesion (CDH23 and PCDH15), scaffold proteins (USH1C, ADGRV1, and SANS), and ciliary function (ADGRV1 and C-Nap1).<sup>4,12,14</sup> In the present study, we used a combination of WES and WGS in a set of patients who share the same ethnicity and the same distinctive phenotype in order to identify the molecular cause of atypical USH.

## MATERIALS AND METHODS

### Patient recruitment

Three Yemenite Jewish families including five individuals with USH (two patients in family MOL0120, two in family MOL0737, and one in family TB55) were recruited for this study. The tenets of the Declaration of Helsinki were followed, the study was approved by the institutional review boards of our respective institutions, and before donation of a blood sample, written informed consent was obtained from every individual who participated in this study.

### Ocular examination

Ocular evaluation included a comprehensive ophthalmologic exam, Goldmann perimetry, full-field electroretinography (ffERG), electrooculography, color vision testing using the Ishihara 38-panel and Farnsworth-Munsell D-15 tests and color and pseudocolor fundus photos, optical coherence tomography, and fundus autofluorescence (FAF) imaging.

### Electrophysiological testing

International Society for Clinical Electrophysiology of Vision standard ffERG was performed using monopolar corneal electrodes (Henkes type; Medical Workshop, Groningen, The Netherlands) and a computerized system (UTAS 3000; LKC, Gaithersburg, MD). Cone responses to 30-Hz flashes of white light were acquired under a background light of 21 cd/m<sup>2</sup>. Scotopic responses, including a rod response to a dim blue flash and a mixed cone–rod response to a standard white flash, were acquired after 30–45 min of dark adaptation. Between two and four sets of responses were recorded in each condition for reproducibility. All ffERG responses were filtered at 0.3 to 500 Hz, and signal averaging was applied.

### Audiometry testing

Audiological screening included a comprehensive questionnaire (collecting information regarding any history of exposure to noise, ototoxic agents, and genetic factors related to hearing impairment), as well as physical and audiometric examinations. An age-appropriate audiological examination was performed, including pure-tone audiometry (250 Hz to 12,000 Hz), tympanometry, and transient-evoked otoacoustic emission (TEOAE) for each ear. We used the following scale to grade the degree of SNHL: slight, 16–25 dB hearing loss; mild, 26–40 dB; moderate, 41–55 dB; moderately severe, 56–70 dB; severe, 71–90 dB, and profound, over 90 dB. Specific types of SNHL were determined by audiometric curve patterns: ascending (hearing loss greater at the lower frequencies), flat, descending (hearing loss greater at the higher frequencies), and U-shaped (hearing loss at

mid-frequencies) curves. Otoacoustic emissions have been suggested as a sensitive measure of cochlear function with the potential for preclinical detection of damage. TEOAE tests were conducted on the same day as the pure tone test. A response at three frequencies of 3 dB or greater above the background with a minimum of 70% reproducibility at each frequency and 90% or greater stability was required for passing the TEOAE test. TEOAE examination was categorized as either pass or fail for each ear.

### Genetic analyses

Homozygosity mapping was performed using whole-genome single-nucleotide polymorphism microarrays (Affymetrix 6.0, USA and Illumina, San Diego, CA) and analyzed using HomozygosityMapper (<http://www.homozygositymapper.org>). WES analysis of individuals MOL0120 III:2 and MOL0737 II:2 was performed at Otogenetics (Norcross, GA) using Agilent V4 (Santa Clara, CA, USA) and Illumina HiSeq 2000 with 30× coverage. WES analysis of individual TB55 II:1 was performed at the Technion Genome Center, using the Nextera Expanded Exome Enrichment Kit (Illumina, USA) and Illumina HiSeq2500 at 30× coverage. Sequence reads were aligned to the human genome reference sequence (hg38), and variant filtering analysis was performed.

### WGS

WGS was performed using the Complete Genomics platform (format version 2.5), and the data were analyzed as described previously.<sup>9</sup>

### Primer design

Primers for mutation screening were designed using the Primer3 software and Sanger sequencing of polymerase chain reaction (PCR) products was used to verify the mutation and to screen additional patients and controls.

### Reverse-transcriptase PCR

Retinal RNA was isolated from human retinas using TRI Reagent (Sigma-Aldrich, St Louis, MO), and RNA derived from different human tissues was purchased (Clontech, Mountain View, CA, USA; cat. 636643, lot 8101369A). Complementary DNA was synthesized using the Verso cDNA kit (Thermo Fisher Scientific, Waltham, MA, USA) in accordance with the manufacturer's protocol. The PCR-specific primers for arylsulfatase G (*ARSG*) and an internal control (*PGM1*) are listed in **Supplementary Table S3** online.

### Lysosomal enzymatic activities

The lysosomal activities in plasma and leukocytes were determined using standard fluorimetric methods (as described by Z. Lukacs in *Laboratory Guide to the Methods in Biochemical Genetics* (N. Blau, M. Duran, and K.M. Gibson, eds), Springer: Berlin, Germany, 2008:287–324).

### Mutagenesis and cloning

The ARSG-D45Y mutant was generated by site-directed mutagenesis using a pCI-neo hARSG-WT construct as template encoding human wild-type ARSG equipped with a C-terminal RGS-His6 tag.<sup>15</sup> PCR was conducted using Kapa HiFi polymerase (Kapa Biosystems, Wilmington, MA, USA) and the primers 5'-gtgattatttggccgattacatgggggtggggtg-3' and 5'-caccccccctgtaatcgccaaaataacac-3', respectively. All plasmids were controlled by Sanger sequencing of the ARSG insert.

### Generation and cultivation of stable cell lines

HT1080 fibrosarcoma cells were transfected with the pCI-neo hARSG-D45Y plasmid using PEI (Polysciences Europe GmbH, Hirschberg am der Bergstrasse, Germany) in Opti-MEM (Gibco, Thermo Fisher Scientific, Waltham, MA, USA). Stably transfected cell clones were selected by adding 800 µg/ml Geneticin (G-418) for 14 days of culture. Clones were screened for ARSG expression by immunoblotting using anti-RGS-His6 (Qiagen, Hilden, Germany), anti-ARSG (R&D Systems, Minneapolis, MN, USA) and anti-GAPDH Santa Cruz, Santa Cruz, Dallas, TX, USA; Invitrogen, Carlsbad, CA, USA) antibodies together with corresponding HRP-coupled secondary antibodies (Invitrogen, Carlsbad, CA, and Dianova, Hamburg, Germany). The cell lines expressing ARSG wild type and the ARSG-C84A mutant had been generated in a previous study.<sup>15</sup> Cells were grown in Dulbecco's Modified Eagle Medium (Invitrogen) containing 10% FCS (PAN), 2 mM glutamine (Lonza, Basel, Switzerland), and penicillin/streptomycin (Lonza, Basel, Switzerland) under standard culture conditions (37 °C, 5% CO<sub>2</sub>).

### Sulfatase activity quantification

Cell lysates were prepared in ice-cold lysis buffer (TBS with 1 mM PMSF, 5 mM IAA, 1 mM EDTA and 0.1% (v/v) Triton X-100). For this purpose, cell pellets were resolved in lysis buffer and sonified (3 × 20 s, 40% intensity, 4 °C, Branson Sonifier, Slough, UK). The lysates were cleared by centrifugation at 18,000g at 4 °C for 15 min, and protein concentration was determined using the DC Protein Assay (Bio-Rad, Hercules, CA) with bovine serum albumen as a standard. Sulfatase activity was measured using the pseudosubstrate paranitrocatechol sulfate (Sigma-Aldrich Chemie GmbH, Munich, Germany) as described previously.<sup>16</sup> Specific activities were calculated and normalized according to western blot signals, i.e., after ARSG quantification by SDS-PAGE and immunoblotting using polyvinylidene fluoride membranes and the antibodies listed above. Band intensities were determined using ImageJ. Statistical analyses were conducted with GraphPad Prism (GraphPad Software, La Jolla, CA).

### Molecular modeling and structural analysis

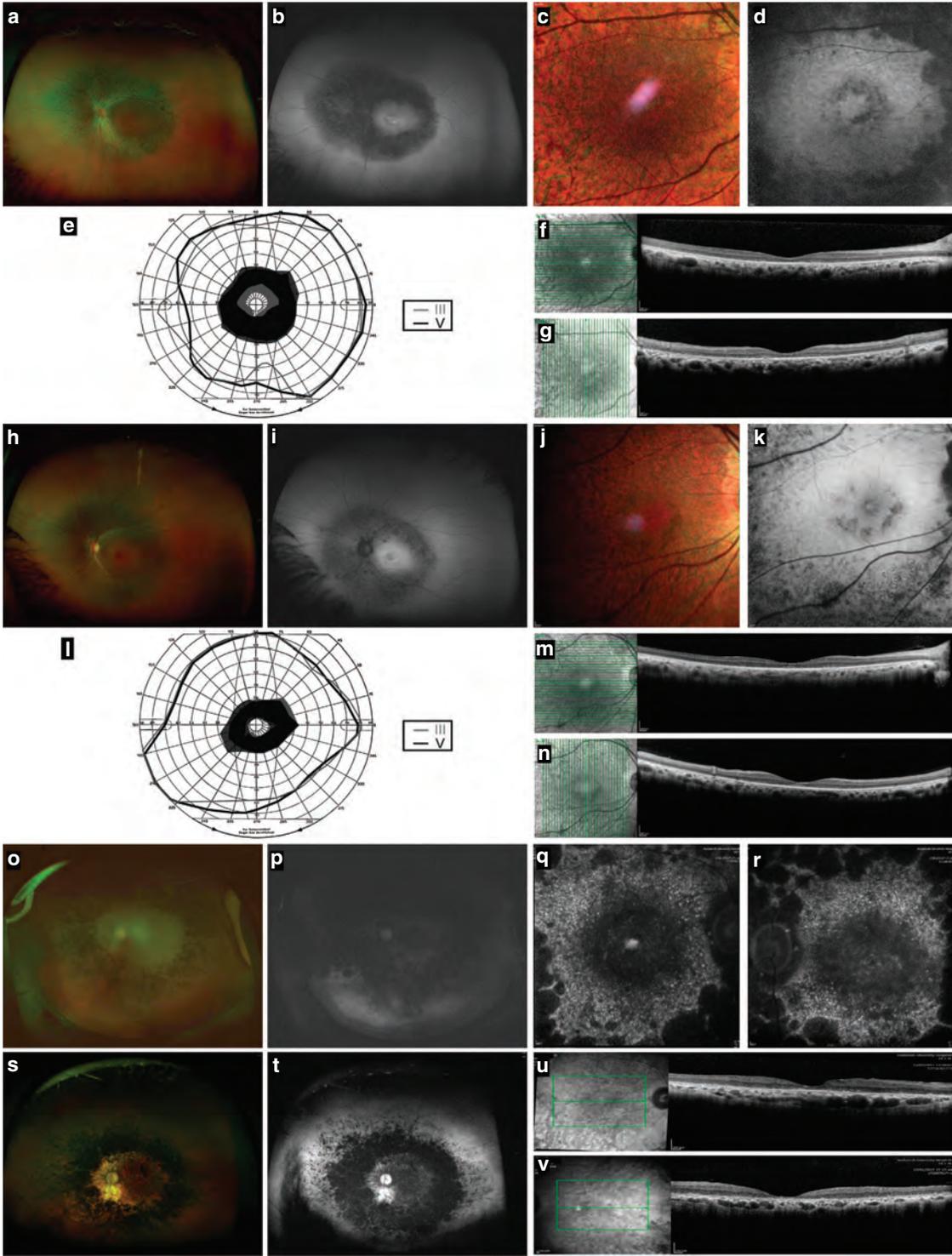
The ARSG structure was modeled using SwissModel with the structures of human arylsulfatase A (ARSA) (PDB 1e1z) or a catalytic formylglycine (FGly)-ester of human ARSA (PDB 1n2k) as templates. The ARSG and ARSA structures were

visualized and manipulated using PyMol (DeLano Scientific, Palo Alto, CA).

## RESULTS

### Clinical description of recruited patients

In our cohort of more than 1,500 Israeli and Palestinian families with IRDs, five affected individuals from three families shared the same ethnicity (Yemenite Jewish) and presented with a particular combination of late-onset IRD and SNHL. While some variability was present in the clinical manifestations among subjects (detailed case reports are provided in **Supplementary Information Appendix**), a characteristic pattern could be identified: all affected individuals presented with a distinctive retinal phenotype, including ring-shaped retinal atrophy delimiting the vascular arcades temporally and extending beyond the optic nerve nasally, with relative preservation of the mid- and far-periphery. Over time, pigment migration occurred within the atrophic areas, forming bone spicule-like pigmentary changes as well as pigment clumps, and the central macula also became involved (**Figure 1**). Fundus autofluorescence imaging using the Optos ultrawide-field retinal camera emphasized the areas of involvement, showing the hypo-fluorescent ring of atrophy surrounded by hyperfluorescent borders (**Figure 1**). On vertical and horizontal optical coherence tomography imaging, retinal thinning with loss of the outer retinal layers was evident also within the macular area, including the fovea at later ages (**Figure 1**). Functionally, Goldmann visual fields correlated with the structural changes, showing a central ring scotoma with relative preservation of central vision at earlier stages of disease. Later on, this progressed to involve the macular area with development of central scotomas, while mid- and far-peripheral visual fields remained relatively preserved (**Supplementary Figure S1**). Electroretinographic testing showed severely decreased rod and mixed cone-rod responses (**Table 1**). Electrooculography testing was performed in three affected individuals and the Arden ratio was found to be reduced (**Table 1**). These findings suggest injury to the retinal-pigmented epithelium, either as a primary event or secondary to photoreceptor degeneration. With regard to involvement of the auditory system, all affected subjects presented with progressive moderate to severe SNHL at a relatively late age (usually after the age of 40 years—**Supplementary Figure S2** online and **Supplementary Table S1**). None of the affected individuals reported significant abnormalities of the vestibular system (detailed description in **Supplementary Information**). In addition, the following clinical evaluations and tests performed in the affected individuals were all within normal limits: lack of neurological abnormalities (all affected individuals), normal brain magnetic resonance imaging (performed on one patient), normal abdominal ultrasound exam (performed in two of the patients), and normal biochemical blood and urine tests (four patients). Osteoporosis, as determined by bone density measurements, is present in one of the patients and osteopenia in the two



sisters of family MOL0737. It should be noted that at the time of all the above testing the patients were of middle age or beyond, between the ages of 50 and 72.

### Genetic analysis of study participants

Aiming to identify the cause of disease in these families, we initially performed homozygosity mapping on four individuals (MOL0737 II:1, MOL0737 II:2, MOL0120 III:1, and TB55 II:1—**Figure 2a**), using single-nucleotide polymorphism arrays. Following analysis with HomozygosityMapper, family MOL0737 revealed two large homozygous regions (a 9.3 Mb region on chr. 5: 126.2–135.5 Mb and a 4-Mb region on chr. 17: 66–70 Mb), shared by the two affected sisters, neither of which included a known USH gene or a clear-cut candidate for the disease. Homozygosity mapping of MOL0120 III:1 revealed four large homozygous regions (more than 20 Mb), one of which on chr. 17 (50.4–76.4 Mb) showed overlap with the one identified in family MOL0737. Patient TB55 II:1 harbored eight homozygous regions of over 10 Mb each, one of which was a 12.58-Mb homozygous region on chr. 17 (56.8–69.4 Mb). All four patients had a 3.4-Mb identical haplotype on chr. 17 (66–69.4 Mb). We subsequently performed WES analysis on three samples (MOL0737 II:2,

MOL0120 III:1, and TB55 II:1, **Figure 2a**) and WGS on one sample (MOL0737 II:1), and all were negative for suspected disease-causing mutations in any of the 226 known IRD genes.<sup>17</sup> WES and WGS data were also analyzed for homozygosity and the combined information revealed a single shared homozygous region on chr. 17 (66–69.4 Mb). Variant analysis revealed a single shared missense transversion, c.133G > T, p.D45Y (chr.17: 66303768G > T), in exon 2 of ARSG (NM\_014960.4), which was validated by Sanger sequencing (**Supplementary Figure S3**) and found to be in a homozygous state in all five affected individuals. All haplotypes surrounding the variant site were identical and this variant is therefore a founder mutation.

To rule out the possibility that the ARSG c.133G > T variant is a polymorphism in the Yemenite Jewish population, we screened 101 controls and identified one heterozygous individual (this corresponding to a minor allele frequency of 0.005). In addition, the variant does not appear in any of the available databases, including gnomAD (with 246,228 sequenced alleles of various origins). We subsequently used Sanger sequencing of exon 2 to screen the remaining unsolved IRD index cases of Yemenite origin ( $n = 19$ ) for p.D45Y, as well as other mutations in this exon. The analysis revealed

**Table 1** Clinical data of patients who are homozygous for the p.D45Y ARSG mutation

Patient no. (age in years) <sup>a</sup>	Best corrected visual acuity	Refraction	Full field ERG results			EOG (%)
			Cone flicker—30 Hz, IT in msec	Mixed cone–rod response (μV)	Rod response—blue light (b, μV)	
<b>MOL0120 III:1</b>						
(55)	ND	ND	25, 37.2	a-18, b-40	20	100
(62)	ND	-2.68	ND	ND	ND	ND
(65)	0.2	ND	ND	ND	ND	ND
<b>MOL0120 III:2</b>						
(75)	0.0001	NA	NA	NA	NA	NA
<b>MOL0737 II:1</b>						
(58)	0.3	+5.18	56, 34	a-85, b-129	98	133
<b>MOL0737 II:2</b>						
(47)	0.8	+1.50	63, 31	a-118, b-341	227	146
(55)	0.7	+2.50	84, 32	a-104, b-209	141	143
<b>TB55 II:1</b>						
(51)	NA	NA	24, 35	a-52, b-70	45	NA

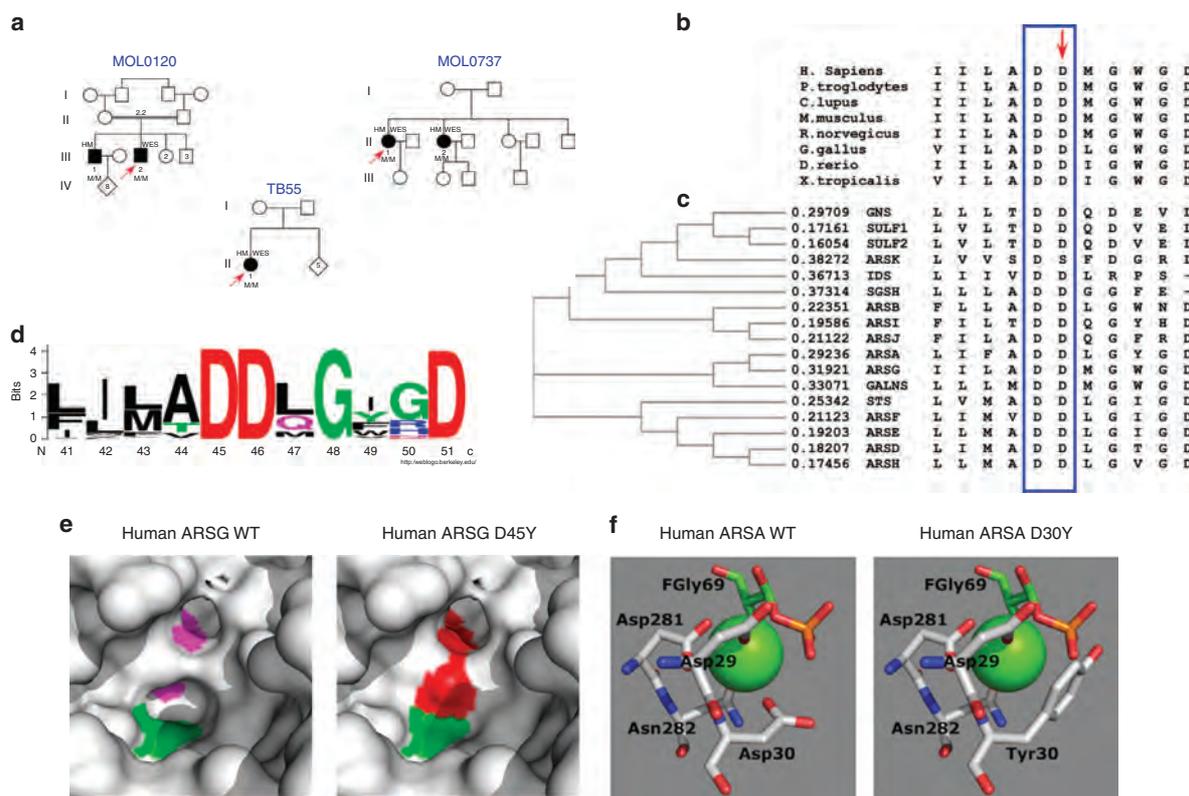
EOG, electroretinography; IT, implicit time; NA, not available; ND, not done.

Best corrected visual acuity is presented in decimal values as an average of the two eyes.

Cone Flicker 30 Hz—lower normal threshold: amplitude—60 μV; response interval—33 msec; mixed cone–rod response—normal lower threshold of b wave: 400 μV; blue-light rod response—lower normal threshold: 200 μV; EOG—normal average 250%, lower threshold—185%.

<sup>a</sup>Age at testing is indicated in parentheses.

**Figure 1** Retinal imaging of patients homozygous for the p.D45Y ARSG mutation. (a–g) MOL0737 II:1 (56 years), (h–n) MOL0737 II:2 (55 years), (o and p) MOL0120 III:2 (72 years), (q–v) MOL0120 III:1 (69 years). a, h, o, s and b, i, p, t represent ultrawide-field pseudocolor and FAF photos, respectively, taken using the Optos Panoramic 200 Optomap Fundus Camera. Characteristic ring-shaped atrophic changes accompanied by intraretinal pigment migration along the major vascular arcades and extending beyond the optic disc can be seen. (e and l) Goldmann visual field of the parallel fundus photos (a and h), respectively, each showing the corresponding ring-shaped scotoma. (c, d, j, and k) Heidelberg Spectralis multicolor fundus photos and FAF of the posterior pole of subjects MOL0737 II:1 and MOL0737 II:2, respectively. (q and r) Spectralis FAF images of the posterior pole of subject MOL0120 III:1, demonstrating macular involvement. (f, g, m, n, u, and v) are optical coherence tomography sections showing retinal thinning and loss of the outer retinal layers in the macular area of subjects MOL0737 II:1, MOL0737 II:2 and MOL0120 III:1, respectively.



**Figure 2** ARSG mutation impact on the ARSG active site structure. (a) Pedigrees of families with the ARSG mutation. Filled symbols designate affected individuals. Red arrows represent index cases. Family number is depicted above each family tree and individual number below each symbol. The ARSG genotype of recruited individuals is depicted below the individual number: M, c.133G>T. Applied sequencing techniques are presented as acronyms above each subject. (b) Multiple alignment of the amino acid sequence flanking the ARSG-p.D45 mutated amino acid shows complete preservation of the aspartate in different species. (c) Multiple alignment of the amino acid sequence flanking the mutated amino acid shows preservation of the relevant aspartate in 16 of the 17 human members of the sulfatase family. (d) A WebLogo image showing the relative preservation of amino acids around ARSG-p.D45 in 74 related sequences. (e) Modeling of the human ARSG active site. The structure of human ARSG was modeled using human arylsulfatase A (ARSA), the closest homologue of ARSG, as a template (PDB code 1e1z). The model is presented as surface structure. The green color indicates the catalytic  $\alpha$ -formylglycine (FGly) residue, which is located next to a cavity in the ARSG surface. Mutation of aspartate-45 (magenta) to tyrosine (red) positions the phenolic ring of the tyrosine within this cavity, thereby altering the accessibility of the FGly and interfering with substrate desulfation. (f) Molecular structure of the calcium-binding site of ARSA. (Left) The four complexing aspartate and asparagine residues are shown in white with heteroatoms shown in red and blue. The FGly residue is shown in green. The complexed metal ion is represented by the green sphere. In this structure (PDB code 1n2k) a critical intermediate of the catalytic cycle was trapped using 4-methylumbelliferyl phosphate as a substrate surrogate which leads to FGly phosphorylation, thereby mimicking the physiological FGly sulfation. The phosphor atom is shown in orange. (Right) Upon D30Y mutation the tyrosine side chain clashes with the phosphate group, indicating steric hindrance of substrate binding and, consequently, of covalent FGly-ester formation. In addition, the coordination of the metal ion will be massively distorted.

a single heterozygous case with early-onset RP and normal hearing at the age of 46 years. Sequencing analysis of all ARSG exons failed to detect a second possible pathogenic variant in this individual, and we therefore predict that she carries p. D45Y by chance, while the disease-causing mutation lies in a yet-to-be-identified RP-causing gene. The genotype difference between the cohort of IRD families (three homozygous: one heterozygous: 18 wt) and controls (none homozygous: one heterozygous: 100 wt) of Yemenite Jewish origin is statistically significant ( $p < 4.0 \times 10^{-4}$ ). Aiming to expand the ARSG tissue expression profile,<sup>18</sup> we performed reverse-transcriptase PCR analysis of 17 different human tissues, which revealed

a ubiquitous pattern, with ARSG being expressed in the majority of the studied tissues, including the retina (Supplementary Figure S4).

**p.D45 is critical for ARSG function**

In the encoded arylsulfatase G (ARSG) enzyme, an ensemble of ten conserved amino acid residues forms the catalytic pocket, as holds true for type I sulfatases in general.<sup>18,19</sup> For ARSG, one of these ten residues is aspartate at position 45. This amino acid is therefore extremely important and did not tolerate any substitution in the course of evolution (Figure 2b–d). It is perfectly conserved not only among all known

ARSG orthologues (Figure 2d, which is based on 74 ARSG-related sequences) but also among 16 of the 17 members of the human sulfatase family (Figure 2c). Of note, mutations in the corresponding aspartate residue in *N*-sulfoglucosamine sulfohydrolase (encoded by *SGSH*) and *N*-acetylgalactosamine-6-sulfatase (encoded by *GALNS*) cause loss-of-function alleles resulting in different forms of mucopolysaccharidosis (MPS type IIIA and IVA, respectively).<sup>20,21</sup>

Arylsulfatase G was identified in 2002 as a novel sulfatase gene<sup>18</sup> and has been further characterized biochemically as a lysosomal sulfatase.<sup>15</sup> As mentioned above, aspartate-45 is part of the ARSG active site, where it coordinates a calcium ion next to the catalytic formylglycine (FGly) residue that is essential for all human sulfatases.<sup>22</sup> Figure 2e shows the surface structure of the ARSG active site, as modeled on the known structure of human ARSA, i.e., the closest relative of ARSG (see Figure 2c). A tyrosine (Figure 2e, red) replacing the aspartate (magenta), as found in the patients described here, fills part of the active site cavity next to the FGly (green). The bulky tyrosine side-chain leaves no space for formation of the FGly-sulfate ester, which is an essential intermediate of the catalytic cycle,<sup>23</sup> as shown in Figure 2f on the basis of a published crystal structure with a trapped ester intermediate in ARSA.<sup>24</sup> In addition, a loss of the calcium ion may be predicted as a consequence of a massively distorted coordination, because the tyrosine hydroxyl at a distance of 4.0 Å is too remote (2.4 Å for the wild-type aspartate carboxyl group). Loss of the calcium ion in turn may also affect protein stability.

#### ARSG-p.D45Y abolishes enzyme activity but does not lead to significant glycosaminoglycan excretion in patients

To test these predictions experimentally, ARSG was stably expressed in human fibrosarcoma HT1080 cells, one of the most efficient cell lines used for production of human sulfatases.<sup>15,25</sup> Three independent clones expressing the p.D45Y-mutant were compared to wild-type ARSG, as well as to the null-mutant ARSG-p.C84A (Figure 3), which lacks the catalytic FGly-84 (deriving from the posttranslational modification of Cys84).<sup>26</sup> No significant sulfatase activity was detectable in the p.D45Y clones (Figure 3a, b), despite the fact that all ARSG variants were expressed at similar levels (Figure 3c). The measured residual activity (<20% of ARSG wild type) was in fact as low as that observed for the p.C84A null mutant (Figure 3a,b) and can be attributed to the endogenous sulfatase activity present in the HT1080 cells.<sup>15</sup> In addition to this marked impairment of enzymatic activity of ARSG-p.D45Y, we observed lack of processing of the ARSG precursor protein (63 kDa) into its mature three chains, of which the largest subunit (~30 kDa in humans),<sup>16</sup> could be detected only for the wild-type protein (Figure 3d). Lack of processing was observed earlier also for ARSG-p.C84A (unpublished data) and also for active site mutants of two other human sulfatases.<sup>27,28</sup> While this maturation of ARSG is dispensable for enzymatic activity in vitro, the observed lack

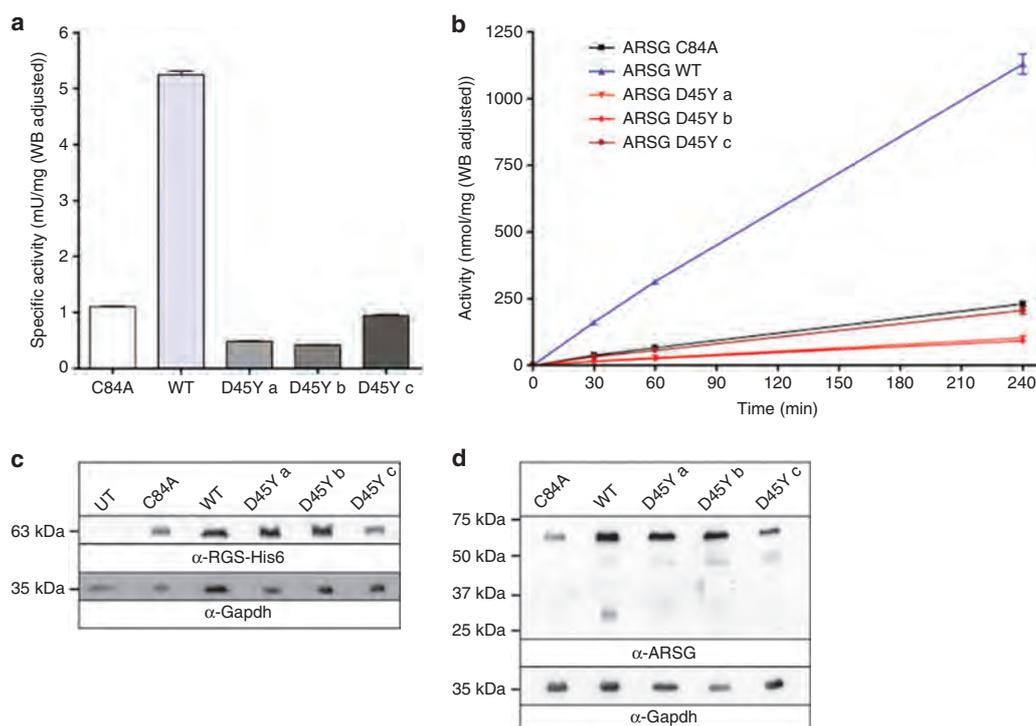
of in vivo processing, which relies on the lysosomal proteases cathepsin B and L,<sup>16</sup> suggests that ARSG-p.D45Y is improperly transported to the lysosomes. To conclude, the p.D45Y-mutation, apart from blocking the catalytic activity, clearly affects maturation and probably also trafficking of the enzyme to its target organelle. However, a drastic decrease in stability of the ARSG-p.D45Y precursor protein was not observed (Figure 3c,d).

Since the ARSG-KO in the mouse model and also mutations in genes that are paralogous to *ARSG* cause phenotypes affecting additional lysosomal enzymes and catabolic pathways,<sup>29,30</sup> we tested patient samples for biochemical parameters that might be affected by ARSG deficiency (Supplementary Table S2). However, the enzymatic activities of  $\beta$ -galactosidase, total hexosaminidase, and arylsulfatase A in plasma samples of two patients, as well as  $\alpha$ -L-iduronidase and  $\beta$ -galactosidase in leukocytes, were all within normal ranges. Of note, urinary glycosaminoglycans were at the upper limit of the normal range (Supplementary Table S2).

#### DISCUSSION

We report here that a mutation in a gene encoding a lysosomal protein, ARSG, causes atypical USH in humans. Patients homozygous for the p.D45Y variant manifest a late-onset rod-cone dystrophy with development of atrophy and pigmentary changes encroaching the posterior pole and causing a ring scotoma, which later progresses to involve the macula and fovea while the mid- and far-periphery of the retina are relatively preserved. This is accompanied by progressive moderate to severe SNHL.

In other species, mutated canine *ARSG* on the basis of its symptomatic presentation has been reported to cause late-onset neuronal ceroid lipofuscinosis (NCL),<sup>31</sup> a neurodegenerative lysosomal storage syndrome that includes static and dynamic ataxia and cerebellar atrophy. Although NCL in other models usually also triggers retinal degeneration (but not hearing loss), none of the 138 studied dogs showed visual impairment. In mice, *Arsg* knockout showed that the protein in fact is responsible for the degradation of 3-O-sulfated *N*-sulfoglucosamine residues of heparan sulfate glycosaminoglycans, causing a new form of lysosomal storage disease, which was termed MPS IIIIE.<sup>32,33</sup> These mice were shown to also have photoreceptor degeneration.<sup>34</sup> In this context, it should be mentioned that retinal degeneration is a very common finding in many human lysosomal storage diseases, including NCL and MPS. Remarkably, photoreceptor degeneration in *Arsg*-knockout mice appeared clearly before manifestation of symptoms affecting the central nervous system, including Purkinje cell loss and ataxia. The USH patients described here did not show any neurological symptoms or signs of peripheral lysosomal storage, even at a relatively old age (the oldest patient described here is 72 years old), but we cannot rule out the possibility that mild and very late-onset symptoms may appear. The phenotype in humans indicates that, as in mice, the retina is among the primary tissues



**Figure 3 Functional analysis of the ARSG-p.D45Y variant.** (a and b) Cell lysates of HT1080 cells, stably overexpressing ARSG wild type, the null-mutant C84A, or the D45Y mutant (independent cell clones D45Y a, D45Y b and D45Y c), were subjected to sulfatase activity assays using the chromogenic pseudosubstrate *para*-nitrocatechol sulfate. The enzyme reaction was terminated after 60 min (a) or after 30, 60, and 240 min (b), as indicated. The given specific activities were normalized according to western blot signals, using an anti-RGS-His<sub>6</sub> antibody (c) to correct for differences in mutant expression. These corrections explain the apparently lower activity for the high-expressing D45Y clones a and b, as compared to D45Y clone c and the C84A null-mutant, all of which in fact show only background activities owing to endogenous sulfatases. Error bars represent standard deviation from three experiments. (d) Analysis of ARSG processing. Cell lysates of overexpressing cells were blotted and ARSG was detected using an anti-ARSG antibody. The processed ~30 kDa form of mature ARSG was observed only for ARSG-WT. GAPDH detection served as loading control (c and d).

affected by ARSG deficiency, suggesting that 3-O-sulfated heparan sulfate has an important physiological function in the retina and the inner ear. Nevertheless, it should also be mentioned that in both NCL (type CLN7) and MPS (type IIIC), diseases that are generally fatal in late adolescence or early adulthood,<sup>35,36</sup> an increasing number of patients were recently described as expressing isolated RP, while the severe neurological and peripheral somatic features that are typically observed in both diseases were lacking.<sup>37-39</sup> In this respect, it should also be noted that the (almost) normal level of urinary glycosaminoglycans in our patients is not in accordance with findings among classical MPSs. Thus, ARSG patients are likely to be missed in the first-line screening for MPS diseases. Interestingly, also other MPS III cases were occasionally reported who presented with urinary glycosaminoglycan levels in the normal range.<sup>40</sup>

The phenotype which we describe here differs from the three accepted USH types that were defined historically, and adds to the large number of “atypical” USH cases that has emerged in recent years. This may further emphasize the need

to update the nomenclature of USH syndrome, in order to better distinguish between the different USH subtypes.

#### SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

#### ACKNOWLEDGMENTS

This study was financially supported by the Foundation Fighting Blindness USA (BR-GE-0214-0639 to D.S., E.B., and T.B.-Y.), Israel Science Foundation (2154/15 to S.K.), Chief Scientist Office of the Israeli Ministry of Health and the Lirot association (300009177 to S.K. and 300011893 to D.S. and T.B.-Y.), the Yedidut Research Grant (to E.B.), the Deutsche Forschungsgemeinschaft (DI 575/10-1 to T.D.), and the Swiss National Science Foundation (156260 to C.R.). The authors thank all patients and family members for their participation in this study, and Torben Lübke for critically reading the manuscript. Geto Mengisto, Kerstin Fentker and Mai-Britt Ilse, Inbar Erdinest, and Devora Marks-Ohana are acknowledged for expert technical assistance.

## DISCLOSURE

The authors declare no conflict of interest

## REFERENCES

- Boughman JA, Vernon M, Shaver KA. Usher syndrome: definition and estimate of prevalence from two high-risk populations. *J Chronic Dis*. 1983;36:595–603.
- Rosenberg T, Haim M, Hauch AM, Parving A. The prevalence of Usher syndrome and other retinal dystrophy-hearing impairment associations. *Clin Genet*. 1997;51:314–321.
- Kimberling WJ, Hildebrand MS, Shearer AE, et al. Frequency of Usher syndrome in two pediatric populations: implications for genetic screening of deaf and hard of hearing children. *Genet Med*. 2010;12:512–516.
- Khateb S, Zelinger L, Mizrahi-Meissonnier L, et al. A homozygous nonsense CEP250 mutation combined with a heterozygous nonsense C2orf71 mutation is associated with atypical Usher syndrome. *J Med Genet*. 2014;51:460–469.
- Bashir R, Fatima A, Naz S. A frameshift mutation in SANS results in atypical Usher syndrome. *Clin Genet*. 2010;78:601–603.
- Liu XZ, Hope C, Walsh J, et al. Mutations in the myosin VIIA gene cause a wide phenotypic spectrum, including atypical Usher syndrome. *Am J Hum Genet*. 1998;63:909–912.
- Pierrache LHM, Kimchi A, Ratnapriya R, et al. Whole-exome sequencing identifies biallelic IDH3A variants as a cause of retinitis pigmentosa accompanied by pseudocoloboma. *Ophthalmology* 2017;124:992–1003.
- Gong B, Wei B, Huang L, et al. Exome sequencing identified a recessive RDH12 mutation in a family with severe early-onset retinitis pigmentosa. *J Ophthalmol*. 2015;2015:942740.
- Nishiguchi KM, Tearle RG, Liu YP, et al. Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene. *Proc Natl Acad Sci USA*. 2013;110:16139–16144.
- Carrs KJ, Arno G, Erwood M, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am J Hum Genet*. 2017;100:75–90.
- Kremer H, van Wijk E, Marker T, Wolfrum U, Roepman R. Usher syndrome: molecular links of pathogenesis, proteins and pathways. *Hum Mol Genet*. 2006;15 Spec No 2:R262–270.
- Reiners J, Nagel-Wolfrum K, Jurgens K, Marker T, Wolfrum U. Molecular basis of human Usher syndrome: deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease. *Exp Eye Res*. 2006;83:97–119.
- El-Amraoui A, Petit C. The retinal phenotype of Usher syndrome: pathophysiological insights from animal models. *C R Biol*. 2014;337:167–177.
- Flanagan AM, Stavenschi E, Basavaraju S, Gaboriau D, Hoey DA, Morrison CG. Centriole splitting caused by loss of the centrosomal linker protein C-NAP1 reduces centriolar satellite density and impedes centrosome amplification. *Mol Biol Cell*. 2017;28:736–745.
- Frese MA, Schulz S, Dierks T. Arylsulfatase G, a novel lysosomal sulfatase. *J Biol Chem*. 2008;283:11388–11395.
- Kowalewski B, Lubke T, Kollmann K, et al. Molecular characterization of arylsulfatase G: expression, processing, glycosylation, transport, and activity. *J Biol Chem*. 2014;289:27992–28005.
- Beryozkin A, Shevah E, Kimchi A, et al. Whole exome sequencing reveals mutations in known retinal disease genes in 33 out of 68 Israeli families with inherited retinopathies. *Sci Rep*. 2015;5:13187.
- Ferrante P, Messali S, Meroni G, Ballabio A. Molecular and biochemical characterisation of a novel sulphatase gene: arylsulfatase G (ARSG). *Eur J Hum Genet*. 2002;10:813–818.
- Waldow A, Schmidt B, Dierks T, von Bulow R, von Figura K. Amino acid residues forming the active site of arylsulfatase A. Role in catalytic activity and substrate binding. *J Biol Chem*. 1999;274:12284–12288.
- Beesley CE, Young EP, Vellodi A, Winchester BG. Mutational analysis of Sanfilippo syndrome type A (MPS IIIA): identification of 13 novel mutations. *J Med Genet*. 2000;37:704–707.
- Morrone A, Tylee KL, Al-Sayed M, et al. Molecular testing of 163 patients with Morquio A (mucopolysaccharidosis IVA) identifies 39 novel GALNS mutations. *Mol Genet Metab*. 2014;112:160–170.
- Schmidt B, Selmer T, Ingendoh A, von Figura K. A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* 1995;82:271–278.
- Boltes I, Czapińska H, Kahnert A, et al. 1.3 A structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. *Structure* 2001;9:483–491.
- Chruszcz M, Laidler P, Monkiewicz M, Ortlund E, Lebioda L, Lewinski K. Crystal structure of a covalent intermediate of endogenous human arylsulfatase A. *J Inorg Biochem*. 2003;96:386–392.
- Dierks T, Schmidt B, Borissenko LV, et al. Multiple sulfatase deficiency is caused by mutations in the gene encoding the human C(alpha)-formylglycine generating enzyme. *Cell* 2003;113:435–444.
- Dierks T, Lecca MR, Schlotterhose P, Schmidt B, von Figura K. Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. *EMBO J*. 1999;18:2084–2091.
- Daniele A, Di Natale P. Heparan N-sulfatase: cysteine 70 plays a role in the enzyme catalysis and processing. *FEBS Lett*. 2001;505:445–448.
- Millat G, Froissart R, Maire I, Bozon D. Characterization of iduronate sulphatase mutants affecting N-glycosylation sites and the cysteine-84 residue. *Biochem J*. 1997;326 (Pt 1):243–247.
- Van Hoof F, Hers HG. The abnormalities of lysosomal enzymes in mucopolysaccharidoses. *Eur J Biochem*. 1968;7:34–44.
- Vitner EB, Platt FM, Futerman AH. Common and uncommon pathogenic cascades in lysosomal storage diseases. *J Biol Chem*. 2010;285:20423–20427.
- Abitol M, Thibaud JL, Olby NJ, et al. A canine arylsulfatase G (ARSG) mutation leading to a sulfatase deficiency is associated with neuronal ceroid lipofuscinosis. *Proc Natl Acad Sci USA*. 2010;107:14775–14780.
- Kowalewski B, Heimann P, Ortkras T, et al. Ataxia is the major neuropathological finding in arylsulfatase G-deficient mice: similarities and dissimilarities to Sanfilippo disease (mucopolysaccharidosis type III). *Hum Mol Genet*. 2015;24:1856–1868.
- Kowalewski B, Lamanna WC, Lawrence R, et al. Arylsulfatase G inactivation causes loss of heparan sulfate 3-O-sulfatase activity and mucopolysaccharidosis in mice. *Proc Natl Acad Sci USA*. 2012;109:10310–10315.
- Kruszewski K, Lullmann-Rauch R, Dierks T, Bartsch U, Damme M. Degeneration of photoreceptor cells in arylsulfatase G-deficient mice. *Invest Ophthalmol Vis Sci*. 2016;57:1120–1131.
- Kousi M, Siintola E, Dvorakova L, et al. Mutations in CLN7/MFSD8 are a common cause of variant late-infantile neuronal ceroid lipofuscinosis. *Brain* 2009;132(Pt 3):810–819.
- Wraith JE. Mucopolysaccharidoses and oligosaccharidoses. In: Fernandes J, Saudubray JM, van den Berghe G, Walter JH, (eds). *Inborn Metabolic Diseases. Diagnosis and Treatment*. Vol 4. Springer: Heidelberg, Germany, 2006:496–507.
- Khan KN, El-Asrag ME, Ku CA, et al. Specific alleles of CLN7/MFSD8, a protein that localizes to photoreceptor synaptic terminals, cause a spectrum of nonsyndromic retinal dystrophy. *Invest Ophthalmol Vis Sci*. 2017;58:2906–2914.
- Roosing S, van den Born LJ, Sangermano R, et al. Mutations in MFSD8, encoding a lysosomal membrane protein, are associated with nonsyndromic autosomal recessive macular dystrophy. *Ophthalmology* 2015;122:170–179.
- Haer-Wigman L, Newman H, Leibur R, et al. Non-syndromic retinitis pigmentosa due to mutations in the mucopolysaccharidosis type IIIC gene, heparan-alpha-glucosaminidase N-acetyltransferase (HGSNAT). *Hum Mol Genet*. 2015;24:3742–3751.
- Giugliani R. The mucopolysaccharidoses. In: Metha A, Winchester, B., (eds). *Lysosomal Storage Disorders: A Practical Guide*. John Wiley: Chichester, UK, 2012:94–100.

## **PROJECT 6 - ARTICLE**

### **Analysis of the genetic basis of periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) syndrome**

Published in *Scientific Reports*, May 19<sup>th</sup> 2015

#### **Contributions**

I took part to this project in the first months of my PhD contract. I contributed with screening and segregation analysis of candidate variants in genes involved in the inflammasome.

OPEN

## Analysis of the genetic basis of periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) syndrome

Received: 10 November 2014

Accepted: 07 April 2015

Published: 19 May 2015

Silvio Alessandro Di Gioia<sup>1</sup>, Nicola Bedoni<sup>2</sup>, Annette von Scheven-Gête<sup>2</sup>, Federica Vanoni<sup>2</sup>, Andrea Superti-Furga<sup>3</sup>, Michaël Hofer<sup>2,4,\*</sup> & Carlo Rivolta<sup>1,\*</sup>

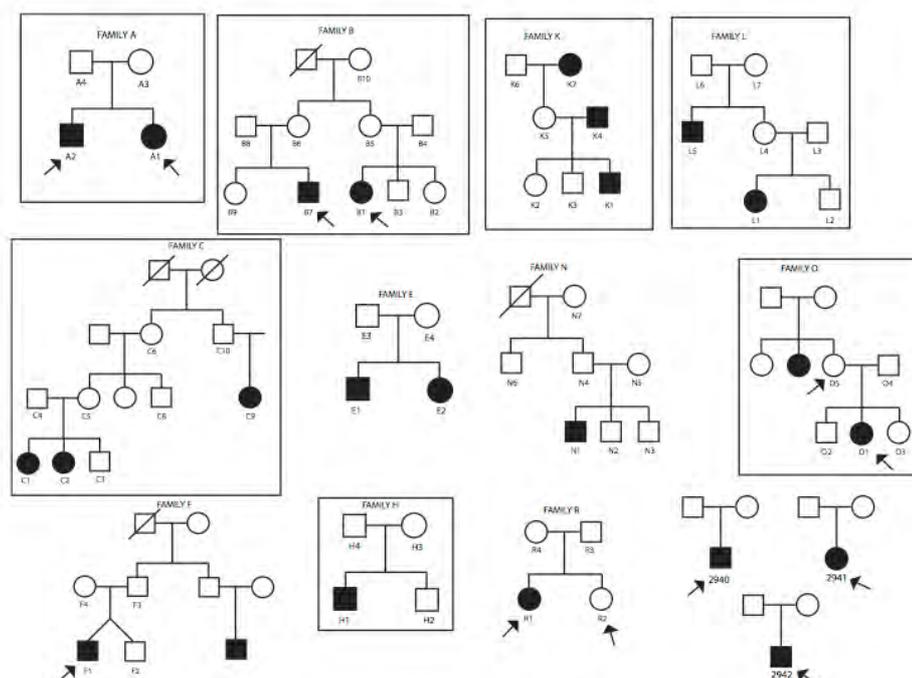
PFAPA syndrome is the most common autoinflammatory syndrome in children from Western countries. In spite of its strong familial clustering, its genetic basis and inheritance pattern are still unknown. We performed a comprehensive genetic study on 68 individuals from 14 families. Linkage analysis suggested a susceptibility locus on chromosome 8, but direct molecular sequencing did not support this initial statistical finding. Exome sequencing revealed the absence of any gene that was mutated in all patients. Exhaustive screening of genes involved in other autoinflammatory syndromes or encoding components of the human inflammasome showed no DNA variants that could be linked to PFAPA molecular pathology. Among these, the previously-reported missense mutation V198M in the *NLRP3* gene was clearly shown not to co-segregate with PFAPA. Our results on this relatively large cohort indicate that PFAPA syndrome is unlikely to be a monogenic condition. Moreover, none of the several genes known to be involved in inflammation or in autoinflammatory disorders seem to be relevant, alone, to its etiology, suggesting that PFAPA results from oligogenic or complex inheritance of variants in multiple disease genes and/or non-genetic factors.

Autoinflammatory syndromes (AIS) are a group of disorders characterized by attacks of inflammation that are not associated with the identification of autoreactive lymphocytes or of an external inflammation-triggering agent<sup>1</sup>. The majority of these syndromes are inherited as Mendelian traits, although a strong environmental influence has also been suggested for some forms<sup>2</sup>. Genes involved in AIS are linked to the inflammatory activation pathway and in particular to the activation of interleukin 1 $\beta$  (IL1 $\beta$ )<sup>3</sup>.

Periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) syndrome belongs to the AIS group, although the etiology of this disease is still unknown. Described for the first time by Marshall in 1987<sup>4</sup>, it is characterized by high fever (higher than 39°C) lasting three to seven days and reoccurring regularly every three to eight weeks, alongside with at least one of these other symptoms: aphthous stomatitis, pharyngitis, or cervical adenitis<sup>5</sup>. It is an early-onset disease (usually before the age of 5 years) and in general it completely resolves before adulthood. However, cases of PFAPA that were persistent after adolescence have also been reported<sup>6–9</sup>. Patients are asymptomatic between episodes and have no developmental or growth problems. During PFAPA febrile attacks, an increase of IL1 $\beta$  cytokines has been shown both at the gene<sup>10</sup> and protein levels<sup>11</sup>, suggesting the involvement of IL1 $\beta$  release in the etiology of the disease.

<sup>1</sup>Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. <sup>2</sup>Pediatric Rheumatology Unit of Western Switzerland, Department of Pediatrics, CHUV, University Hospital of Lausanne, Lausanne, Switzerland.

<sup>3</sup>Department of Pediatrics, CHUV, University Hospital of Lausanne, Lausanne, Switzerland. <sup>4</sup>Department of Pediatrics, HUG, Geneva, Switzerland. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.R. (email: carlo.rivolta@unil.ch)



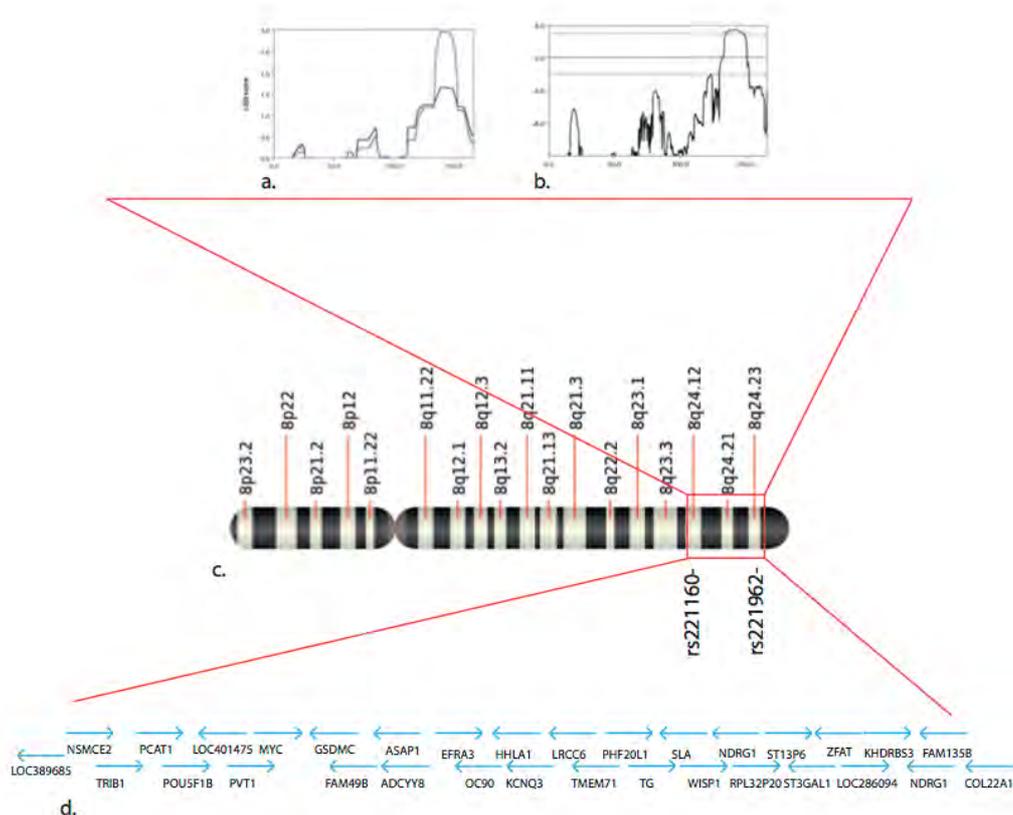
**Figure 1. Families analyzed.** Alphanumeric codes indicate individuals who were investigated by genome-wide SNP genotyping, while arrows indicate individuals whose DNA was processed by exome sequencing (some individuals were analyzed by both techniques). Boxes indicate families that underwent linkage studies.

The genetic origins of PFAPA are one of the most debated features of this syndrome. Although it is generally considered a sporadic disease<sup>12</sup>, familial clustering has been observed, suggesting the presence of a possible hereditary component<sup>13</sup>. It is not unusual to ascertain families having more than one member affected with PFAPA<sup>14,15</sup>, and positive family history for recurrent fevers can be detected in about half of PFAPA patients<sup>16</sup>. Several studies have investigated a possible involvement of the genes responsible for Familial Mediterranean Fever (FMF, gene *MEFV*), TNF-Receptor Associated Periodic Syndrome (TRAPS, gene *TNFRSF1A*) Hyper IgD Syndrome (HIDS, gene *MVK*) and Cryopyrin Associated Periodic Syndrome (CAPS, gene *NLRP3*) in PFAPA cohorts<sup>11,17–21</sup>. An interesting hypothesis is that PFAPA could represent an unspecific, milder form of other AIS for which DNA variants in known AIS genes and present in the general population could be involved. An example is the *NLRP3* variant V198M (or, more precisely, p.V200M, rs121908147), detected in the control population with an allele frequency of ~1%, and reported as a causative mutation for CAPS<sup>22</sup>, as well as to be associated with milder diseases or with other unspecified AIS<sup>23–25</sup>. V198M was also identified in PFAPA patients, suggesting a possible role of *NLRP3* hypomorphic alleles in the pathogenesis of the disease<sup>11</sup>, although a clear genotype-phenotype association is still missing. Recently, a heterozygous deletion of the gene *SPAG7* consequent to a chromosomal translocation was reported in a single individual affected with PFAPA and displaying other syndromic features such as growth retardation, facial dysmorphisms, and dysplastic brachymesophalangy<sup>26</sup>. However, these data could not be replicated and it is currently not clear whether the link between *SPAG7* and PFAPA is causal or coincidental.

In this work, we perform for the first time an unbiased, comprehensive genetic analysis on a collection of families with PFAPA syndrome. Our data indicate that PFAPA is neither a monogenic disease nor a condition that can be clearly associated with DNA changes in other AIS genes. In addition, we demonstrate that the *NLRP3* DNA variant V198M is not associated with this disease.

## Results

**PFAPA syndrome shows an apparent autosomal dominant with incomplete penetrance pattern of inheritance, if considered as a Mendelian disease.** We analyzed 68 individuals from 14 families segregating PFAPA by genome-wide SNP genotyping, whole-exome sequencing, or both. Pedigree analysis suggested clear inheritance, compatible with an autosomal dominant with incomplete penetrance model, if Mendelian inheritance with no genetic heterogeneity is assumed as a model (Fig. 1). Because of the lack of genetic data for this disease to infer the penetrance factor from the literature, we empirically set it to be around 50% following pedigree examination. Following these observations, we decided to perform a genome-wide linkage analysis to identify a possible locus carrying the causative



**Figure 2. Linkage analysis of PFAPA families.** **a**) Output of Merlin linkage analysis for linear (black line) and exponential (blue line) non-parametric models. For our study, the exponential model gives a better resolution, since it is indicated for a small number of families and large allelic sharing. **b**) Output for a parametric model (autosomal dominant with incomplete penetrance). The peak at chromosome 8 is still conserved. **c**) Location of the linkage interval on the long arm of chromosome 8, between markers rs221160 and rs221962. **d**) Protein coding genes present in the identified interval.

gene of PFAPA syndrome. We selected a subset of informative families (families A, B, C, H, K, L, O) based on their clinical information. Multifamily whole genome analysis considering non-parametric methods revealed a single and unique peak on chromosome 8 with a maximum LOD score of 2.9 (Fig. 2a). This peak was also conserved with a parametric method by considering a fully penetrant inheritance pattern with a LOD score of 3.4 (Fig. 2b). The identified region was of 11 Mb and spanned bands 8q21.1 to 8q24.4 of chromosome 8 (Fig. 2c). This region is scarcely populated, as it contains only 40 RefSeq genes, of which only 32 were protein coding (Fig. 2d – and Supplementary Table 1). We selected 3 affected individuals (B1, L1 and C2) to be screened for all exons of all genes that are present in the interval. Analysis of DNA variants failed to reveal any single gene that was mutated in all three affected individuals within the identified region, with the exception of two non-coding variants representing common polymorphisms (Supplementary Table 2).

**Exome capture and NGS show no common disease gene associated with PFAPA.** Despite the negative results from the screening of the chromosome interval, we could not completely exclude *a priori* a monogenic model of inheritance. Therefore we decided to select some key individuals to be whole-exome sequenced. Inclusion criteria were the availability and the quality of the extracted DNA, as well as the size and the quality of the clinical information regarding the family (Table 1). We sequenced, whenever possible, two affected individuals within the same family, or one obligated carrier and a patient, for a total of 11 disease-carrying subjects and one unaffected family member (Fig. 1, arrows).

Under the hypothesis that PFAPA is a monogenic condition with no genetic heterogeneity, we scanned the list of DNA variants present in all patients, in order to identify a gene carrying rare variants (minor allele frequency, or MAF < 2%) in more than 90% of the analyzed samples (to allow the presence of false negatives). In total, we detected 14 genes that satisfied these criteria (Table 2). Four of them belonged to the *MUC* family and represented false positives results, similar to those obtained by other exome sequencing projects<sup>27</sup>. The same was true for mutations in six other genes (*WNK1*, *ZNF384*, *TTN*,

Patient*	Age at onset (months)	m/f	Major symptoms	Other symptoms	Positive FH
A1	54	f	P		Brother A2: PFAPA
A2	60	m	P	AbdP, D	Sister A1: PFAPA
B1	6	f	AD		Cousin: PFAPA
F1	12	m	AD	AbdP, H	
O1	41	f	APH	AbdP	
R1	26	f	P, AD	AR	
PLA 6209	18	f	P, AD, APH		Both parents: TE, Father: recurrent P
PLA 6509	4	m	P, AD	AbdP	Both parents: TE
PLA 5008	16	m	P, AD, APH		Father: TE

**Table 1. Clinical characterization of patients analyzed by whole exome sequencing.** FH: family history, P: pharyngitis, AD: adenitis, APH: oral aphthosis, AbdP: abdominal pain, M: myalgia, AR: arthralgia, R: rash, H: headache, D: diarrhea, TE: tonsillectomy. \*Individuals O5, and R2 were not affected and thus were not examined. Individual B7 was not available for examination.

Gene name	Number of variants	Cases carrying one or more variants	Exclusion criteria
<i>WNK1</i>	54	100%	Alignment mistake
<i>MUC6</i>	40	100%	Mucin gene
<i>MUC2</i>	19	100%	Mucin gene
<i>NBPF15</i>	11	100%	Many novel variants, also present in controls
<i>ZNF384</i>	1	100%	Bad reference
<i>OR4C45</i>	1	100%	Many novel variants, also present in controls
<i>MUC17</i>	91	90%	Mucin gene
<i>MUC16</i>	26	90%	Mucin gene
<i>TTN</i>	18	90%	Low coverage regions
<i>ZNF492/ZNF98</i>	16	90%	Many novel variants, also present in controls
<i>GOLGAL6L1</i>	14	90%	Not well annotated gene
<i>PCMTD1</i>	8	90%	Many novel variants, also present in controls
<i>C11orf40</i>	3	90%	Suspected false positive change in dbSNP
<i>KIAA1751</i>	2	90%	Suspected false positive change in dbSNP

**Table 2. Variant-carrying genes that were common to all affected individuals.**

*GOLGAL6L1*, *C11orf40*, and *KIAA1751*) that, for various reasons, could be recognized as sequencing artifacts<sup>28</sup>. The remaining four genes (*NBPF15*, *OR4C45*, *ZNF492*, and *PCMTD1*) had many novel variants, which were however also present in controls. No common gene carrying rare variants that could be compatible with pathogenic mutations could therefore be identified in our cohort.

No rare or potentially abnormal DNA variants were found in the *SPAG7* gene, in any of the patients analyzed.

**Screening for mutations in genes for autoinflammatory recurrent fever conditions does not support their involvement in PFAPA etiology.** Since the presence of rare variants in other

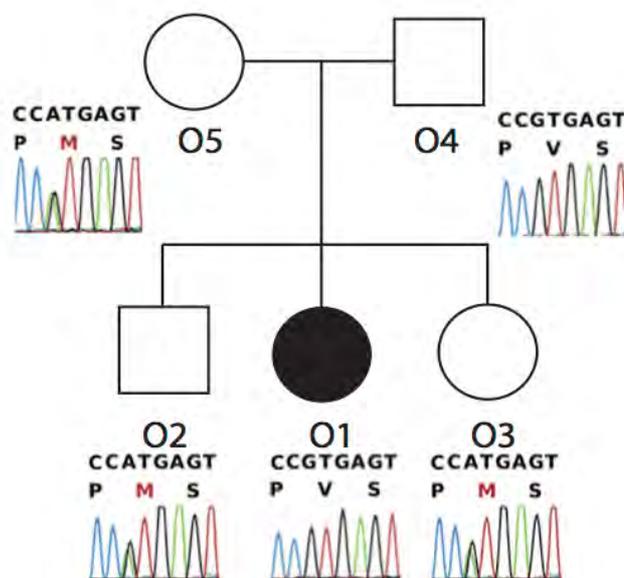
GENE	rs number	Samples	AA change	MAF	Domain
IPAF (NLRC4)	149451729	2941	p.G786V	0.004	LRR6
NAIP	61757629	B7	p.A161T	0.02	BIR2
NLRP1	61754791	2942	p.V939M	0.02	LLR5
NLRP2	142463014	2942	p.S4L	0.01	DAPIN
	novel	B7	p.G504D	0	NACHT
	182098487	2940	p.L607P	0.005	between NACHT and LRR1
NLRP3	121908147	O5	p.V198M	0.01	between DAPIN and NACHT
	35829419	A1,A2	p.Q703K	0.05	between NACHT and LRR1
NLRP4	117212164	O1,O5	p.T162M	0.004	NACHT
NLRP5	novel	F1	p.N255S	0.001	between DAPIN and NACHT
	novel	2940	p.R274Q	0.001	between DAPIN and NACHT
NLRP10	150112481	O1	p.I384T	0.001	NACHT
NLRP11	11671248	2940	p.S1025L	0.01	C-terminal
NLRP12	novel	R1,R2	p.R211C	0.001	NACHT
NOD1 (CARD4)	novel	2942	p.C241G	0	NACHT
NOD3 (NLRC3)	novel	2940	p.R1005Q	0	LRR15
NOD4	16965150	2941, F1	p.S210L	0.03	before NACHT domain
NOD5 (NLRX1)	150153921	B7	p.L193V	0.007	NACHT
	145779362	O1,O5	p.R547W	0.007	between NACHT and LRR1
NWD1	149694092	F1	p.S306L	0.006	between WD and NACHT

**Table 3. Variants in NLR and inflammasome-related genes, and protein domains affected by them. AA:** amino acid, MAF: minor allele frequency.

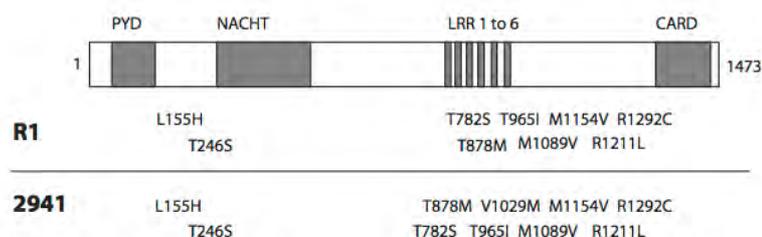
inflammatory genes could have an impact on the PFAPA phenotype, we decided to search specifically for variants in all known autoinflammatory genes in our cohort (Supplementary Table 3). Screening of the *MEFV* gene, associated with FMF, revealed the presence of polymorphisms in heterozygous state with unknown function for patients O1 (p.I591T, rs11466045, MAF 0.07) and R1 (p.F425Y, rs11466045, MAF 0.001). None of these were present in the other analyzed members of the family. Conversely, in family A, we identified a variant (p.E148Q, rs3732930, MAF 0.06) shared among the two affected individuals. Interestingly, both patients of family A also carried a heterozygous change in *NLRP3* related to CAPS (p.Q703K, rs35829419, MAF 0.04). In family B, we identified a rare variant in the gene *TNFRSF1A*, usually associated with milder forms of TRAPS (p.R121Q, rs4149584, MAF 0.018). This change cosegregated with the disease. We also identified a novel change in the gene *NLRP12*, associated with familial cold autoinflammatory syndrome-2, in patient R1 (p.R211C). This change, located in a well-conserved region of the protein (Table 3), was also present in R1's unaffected sister.

**The *NLRP3* variant p.V198M associated with CAPS does not cosegregate with PFAPA affected status.** Analysis of *NLRP3* in family O revealed the presence of the variant V198M (p.V200M, rs12908147, MAF 0.01), often associated with milder forms of CAPS<sup>23,29</sup>. This variant was found in heterozygous state only in the obligated carrier member O5, whereas it was not present in the proband. Cosegregation analysis on available members of the family revealed that this change is present also in the two healthy siblings and it is absent in the affected one, confirming that this change is not involved in the etiology of PFAPA in this family (Fig. 3).

**Several rare variants with uncertain function were present in inflammasome-related genes.** Since an impairment in IL-1 $\beta$  and NF- $\kappa$ B signaling and production has been reported in individuals with PFAPA syndrome<sup>10</sup>, we decided to look closely at all known genes belonging to the inflammasome complex, with particular attention to genes involved in the production of mature IL-1 $\beta$ . All identified variants in these genes are listed in Table 3. We identified two non-annotated changes in *NLRP5*, for patients 2940 and F1 (p.R274Q, and p.N255S, respectively). Again, patient 2940 presented another rare variant in the gene *NLRP2*, p.L607P (rs182098487, MAF 0.001). A novel change, p.G504D, was found in this same gene in patient B7, although it was not found in the affected cousin. Among



**Figure 3. Cosegregation analysis for the V198M variant in *NLRP3*, in family O.** Variant V198M (p.V200M) in gene *NLRP3* does not cosegregate with PFAPA in this pedigree (healthy members carry it, whereas the affected member does not), suggesting that this polymorphism is not associated with the disease.



**Figure 4. Graphical representation of the haplotype identified on *NLRP1* for two subjects.** The gene structure is depicted in the upper part. Amino acid changes identified in these individuals are reported. All these changes have a MAF of <0.1. PYD: pyrin domain; NACHT: Conserved NAIIP, CIITA, HET-E and TP-1 domains; LRR: leucine rich receptor domain; CARD: Caspase activator and recruitment domain.

the Nod-Like Receptor genes, *NLRP2* showed the highest number of variants within PFAPA subjects, but these were also rather common in the control CoLaus Lausanne cohort (p.S4L, rs142463014, for 2492, MAF 0.09; p.I331V, rs61735077, for R1 and R2, MAF 0.05). Another common variant in *NLRP4* was p.T162M (rs117212164, MAF 0.03), co-segregating with the condition in family O and affecting a conserved residue in the NACHT domain of the NLRP4 protein. Among other genes from the NLR families we identified a rare change in *NLRP10*, p.I384T (rs150112481, MAF 0.01) although only in patient O1, and the change p.S1025L (rs11671248, MAF 0.08) in *NLRP11* in patient 2940. Not many rare variants were identified in the coding region of the *NLRP1* gene, with the only exception being p.V939M (rs61754791, MAF 0.01), found only in patient 2942. Interestingly, however, both patient 2941 and family R shared a rare haplotype composed by different missense polymorphisms located at the C-terminal of the protein (Fig. 4). A similar haplotype has been recently associated with an impairment of IL-1 $\beta$  processing by the inflammasome. Common polymorphisms in *CARD8* were found in a heterozygous state in our cohort and these have been previously associated with other defects of the inflammatory system. These polymorphisms were p.C10X (rs2043211, MAF 0.30) found in patients A1, B7, F1, 2940 and p.V148fs (rs140826611, MAF 0.06) in A2, F1, 2941 and 2942. Among other inflammasome genes, we identified a heterozygous missense change in *NLRC4*, better known as IPAF, p.G786V (rs149451729, MAF 0.03) in patient 2941; a novel change in *NOD1* (p.C241G) in patient 2942, again in the NACHT domain of this protein. A summary of all analyzed variants identified in inflammasome genes and their affected functional domain is reported (Table 3).

## Discussion

PFAPA is a common auto-inflammatory disease with an overall good prognosis. Regular PFAPA fever flares have a significant impact on the quality of life of the patients and their families and are interfering with regular school attendance and normal daily activities. The delay to diagnosis may represent several months or years and children may be exposed to unnecessary diagnostic procedures. A better understanding of the pathogenesis of this disease would lead to more specific diagnostic tools and might relieve the suffering of patients and their families. Clearly the pathogenesis of the disease is linked to the immune system, but no convincing information about the etiology and in particular about the genetic basis of PFAPA has been obtained so far.

In our clinical practice, we identified a number of families for which PFAPA seems to have a genetic basis. Based on pedigrees analysis, we could infer an autosomal dominant model of inheritance, with a penetrance factor of almost 50%. However, some families showed an apparently dominant inheritance with full penetrance (e.g. family K), whereas for others a recessive status could not be excluded *a priori* (e.g. family F). According to these observations, we selected the seven most informative families of our collection to perform a genome-wide linkage study. This analysis revealed a clear peak on the long arm of chromosome 8, notably between 8q24.1 and 8q24.3. This region contained a small number of genes and was very close to the telomere of chromosome 8. The significance of the identified area was confirmed using non-parametric analysis, and by comparing haplotypes of affected individuals against non-affected ones. Interestingly, the identified region contained several lncRNAs and many miRNAs, as well as other unknown genes or genes without a clear function. Based on these data we therefore performed a custom sequence capture of all known exons for genes in the interval. The custom capture chip was designed based on annotation of build hg35.1 of the human genome. This implies that we could have missed some recently discovered genes, and all the hypothetical genes and lncRNAs that corresponded to the majority of elements annotated in the locus. The sequencing did not reveal any single mutated gene shared by all affected individuals.

To exclude the possibility that we did not detect a causative disease gene because of inconsistencies in the inferred inheritance pattern, we decided to sequence the entire exome of 11 affected individuals, one “obligate carrier” (i.e. a healthy individual connecting two branches of a pedigree with affected subjects), and one unaffected family member. Our analyses revealed the absence of a unique gene that was mutated in all the affected individuals, including genes and noncoding RNAs lying in the critical region on chromosome 8. Although we cannot exclude *a priori* that mutations that escape exome sequencing such as intronic variants and copy number variations could be the cause of the disease, our data disfavor genetic homogeneity for PFAPA.

We then decided to investigate the possibility that mild mutations or rare variants in other known autoinflammatory genes could cause the PFAPA phenotype. This hypothesis is currently a matter of debate, since the diagnosis of milder forms of autoinflammatory disease and PFAPA is often overlapping, leading to diagnostic uncertainties. In particular, low-frequency polymorphisms (common in 1–4% of the healthy population) have been associated with PFAPA or milder forms of autoinflammatory diseases, mostly because no other strong changes in candidate genes were identified<sup>17,19,30</sup>. Our analysis revealed that, although we identified a few rare missense changes in some inflammatory genes, the frequencies of these variants were not different from those that were present in the general population, indicating that they may not be sufficient to induce the disease. Moreover, some of these changes did not co-segregate with the phenotype, or were also present in healthy members of the family. This was for example observed for the *NLRP3* variant V198M (p.V200M). This change has been reported in several different forms of *NLRP3*-associated syndrome, such as classical CAPS<sup>22,23,29</sup>, or other autoinflammatory manifestation<sup>24,31</sup>, although it is classified as a low penetrant mutation, due to the fact that it is present in a number of healthy carriers (in the Lausanne control population it is present with a frequency of about 1%). The role of this variant is still debated, but it is regrettably reported as a *bona fide* pathogenic mutation in medical genetics diagnostic protocols. The presence of this variant only in the unaffected individuals of family O strongly contradicts its possible role in generating an inflammatory phenotype. A recent paper, probably the most complete clinical study of V198M, showed that 19 cases of CAPS out of 830 carried V198M, although 3 of these harbored other possible candidates in other inflammatory genes, and V198M was present in 7 asymptomatic individuals related to these 19 patients<sup>25</sup>. Furthermore, Valine at codon 200 is not at all conserved through evolution, suggesting that this amino acid residue is likely not detrimental for protein function. Other mammals such as rat, dog and mouse carry a methionine at the same position of their natural *NLRP3* orthologues. Altogether, these data indicate that V198M (p.V200M) is likely a neutral human polymorphism and that patients carrying this variant should probably be re-analyzed for other causative mutations.

Another example is the novel change found in patient R1, the *NLRP12* p.R211C, which is present in a well-conserved domain and has a predicted strong impact on protein function. Recently a similar missense mutation (p.R352C) in the same NACHT domain of this gene was associated with genetically unexplained periodic fevers<sup>32</sup>, with a PFAPA-like phenotype, and was shown to cause a functional defect in Caspase-1 signaling. The fact that R1's unaffected sister carries this allele likely indicates that this change has no effects (or has at best a hypomorphic role) at the clinical level. It is interesting to notice that the affected patient (R1) also carried another very rare change in the gene *MEFV* (p.F425Y) with function unknown. The intriguing hypothesis of digenic or multigenic mutations causing the impairment

of the inflammasome seemed to be reinforced by the identification of two polymorphisms often associated with the disease in family A. More specifically, we identified a change in the *MEFV* gene (p.E148Q) and in *NLRP3* (p.Q703K), both common polymorphisms with a possible functional effect on the protein. In particular, for p.Q703K, functional impairment of *NLRP3* has been clearly demonstrated<sup>33</sup>. Unfortunately this digenic pattern of inheritance was not identified in other affected subjects within known autoinflammatory-associated genes<sup>34</sup>. Nevertheless we cannot exclude, based on this observation, that other rare variants present in other inflammasome-associated genes could be involved in the etiology of PFAPA.

To investigate this possibility we carefully screened some of the genes that compose the inflammasome set (in particular the NLR family members), based on current literature<sup>35–37</sup>. We identified a few novel variants in *NLRP2*, *NLRP5* and *NOD1*, as well as some very rare variants in other inflammasome-associated genes, such as *NLRP10*, *NLRP4*, *NLRP11* and *NLRP12*. Interestingly, the majority of these genes were involved in IL-1 $\beta$  inflammasome activation, or in the NF- $\kappa$ B pathway. For example, new evidence demonstrated the high similarity and redundancy across the structure of *NLRP1*, *NLRP6*, *NLRP10*, *NLRP3* and *NLRP12*<sup>38</sup>. Although the function within the inflammasome for many of these members has not yet been clarified, structural similarities among NLRs suggest that all these components could somehow mediate inflammation in response to different stimuli and trigger IL-1 $\beta$  production<sup>37</sup>. Moreover, we identified two samples carrying a specific rare haplotype in *NLRP1* that has recently been associated with vitiligo and autoimmune diseases, and connected to an increase of IL-1 $\beta$  release in patients carrying the haplotype<sup>39</sup>. Dominant mutations in the same gene have been shown to cause systemic inflammation in mouse models<sup>40</sup>. Another observation from the analysis on the possible effect on protein structure was that the majority of identified variants seemed to be present in or in proximity of the NACHT domain. This domain has an ATPase activity and is crucial for the self-oligomerization (usually to form heptamers or hexamers) of NLRs to form the structure necessary for inflammasome assembly<sup>41</sup>. NACHT domain mutations have been associated with structural changes that lead to a continuous activation of the inflammasome complex. A link between mutations affecting NACHT and different autoinflammatory diseases has been observed in *NLRP3* and *NOD1* for example<sup>42,43</sup>, reinforcing the hypothesis of a gain-of-function effect of these mutations on the formation of the complex.

It is interesting to observe that every single affected individual in this cohort presented more than one rare variant in one of the inflammasome-composing genes. Unfortunately, due to large number of genes involved, it is impossible to perform a meaningful statistical analysis based on this observation, unless extremely large groups of PFAPA patients are recruited and analyzed, possibly as part of an extended international effort. However, it may not be unreasonable to start thinking of a “total inflammatory burden” similar to the concept of “total ciliary burden” proposed for disease such as the Bardet-Biedl syndrome, that has developed from a simple Mendelian disorder into a trigenic<sup>44,45</sup> and finally into an oligogenic disorder<sup>46</sup>.

In conclusion, the absence of a common gene harboring exonic mutations in all the affected individuals analyzed in this work suggests that PFAPA syndrome is transmitted either as Mendelian disease with high genetic heterogeneity, or as an oligogenic or complex trait. Mutations in noncoding or poorly covered regions of the genome that may have escaped our analysis are also possible, and further studies are needed to assess their potential presence in PFAPA patients. Oligogenic inheritance would be consistent with our observation that affected individuals may harbor multiple variants in inflammasome-associated genes. However, given the fact that such variants may be shared with a significant number of the general unaffected population, specific studies are needed to underpin or refute this hypothesis, and extreme caution should be used prior to indicating causality.

## Methods

**Patient selection and ethical commitment.** The families participating in this study were from a large collection of PFAPA patients who were ascertained at the Center for Pediatric Rheumatology of Western Switzerland at the Lausanne University Hospital and the Geneva University Hospital. The familial clustering and inheritance patterns were investigated by administering a specific survey to family members of the affected patients. The diagnosis of PFAPA in family members was based on the report, by the affected persons or their parents, of a stereotypical and recurrent pattern of illness during their own childhood. The PFAPA phenotype was defined according to previously published clinical criteria<sup>47</sup>; in addition a genetic screening for one or more monogenic periodic fever syndromes (FME, TRAPS, HIDS) was performed for selected individuals. Our study was designed in accordance with the tenets of the declaration on Helsinki and was approved by the Institutional Review Boards of the University of Lausanne and of the Lausanne University Hospital. Written informed consent was obtained from all patients who enrolled in the study. As controls for the screening of rare variants, data obtained by exome sequencing of 416 healthy anonymous individuals from a large population-based study done in Lausanne (the CoLaus cohort) were used<sup>48</sup>.

**DNA extraction and lymphocytes immortalization.** DNA was extracted from peripheral leukocytes (PBLs) of participants by using the NucleonBaCC3 DNA extraction kit (GE healthcare, Life Science). For some family members and for all probands, white mononucleate cells were isolated using Ficoll-Paque PLUS (GE healthcare Life Science), according to the manufacturer’s instruction. Once

isolated, cells were processed for EBV immortalization using classical protocols<sup>49</sup>, optimized for our requirements as follows. After three washes with PBS and 2% FCS,  $1 \times 10^6$  PBL cells were transferred to a 10 ml Falcon tube, diluted in 1 ml of LCC medium (RPMI 1640 medium, 20% serum, 2 µg/ml Cyclosporine A) and incubated with a EBV virus suspension for 1 h. Cells were then plated into single wells from a 12-well plate and kept in culture for ~1 month. Following this immortalization process, cells were frozen in RPMI 1640 with 20% serum and 10% DMSO for cryopreservation.

**Genotyping and linkage analysis.** Six out of 14 families were selected for whole genome genotyping, based on the availability of clinical information and the size of the family trees. DNA from all available members of each of these families was assessed by the Illumina Human Linkage12 array (Illumina, CA), containing more than 6000 highly heterozygous SNPs, with an average distance between SNPs of about 441 kb. Output was then analyzed using the Genotype Studio software (Illumina Inc., CA) and the SNPs calls were filtered based on quality and the presence of Mendelian inconsistencies. Genotype data were handled using the commercial software Progeny 7 (Progeny Software LLC) to generate appropriate files for linkage analysis software, carried out with the Merlin package and by using the Vital-IT high performance computer structure (the Swiss Institute of Bioinformatics, Lausanne, www.vital-it.ch). Both parametric and not-parametric analysis models were tested. Because of the lack of data on genetic background for this disease, we could not perform a classical computational approach to estimate the penetrance factor for the parametric analysis. Therefore, penetrance factor for parametric analysis was empirically calculated based on the observed families as a ratio between the number of affected patients and the total number of subjects predicted to be affected based on the model.

**Next generation sequencing.** In the first sequencing experiment, aimed at determining DNA variants on chromosome 8 interval, a custom made sequence capture chip, based on Agilent SureSelect technology, was designed by Genotypic (Bangalore, India). Captured DNA was then sequenced by FASTER (Geneva, Switzerland) by using a single lane of the Illumina Genome Analyzer II (Illumina, San Diego CA).

For exomes, the entire capturing and sequencing procedures were performed at the Lausanne Genomic Facility (GTF). Genomic DNA samples were barcoded and exons were captured using the SureSelect Exome kit V5 (Agilent, Santa Clara, CA) according to the manufacturer's protocols. Sequencing was performed using again a single lane of the Illumina Genome Analyzer II (Illumina, San Diego, CA). Reads were aligned to the reference sequence (human genome reference 19 – hg19) using Novoalign 2.08 (Novocraft.com). The obtained coverage was 10x or higher for 92% of the targeted regions. Base quality score recalibration, indel realignment, duplicates removal and SNP and INDEL calling were performed by using Genome Analysis Tools kit (GATK)<sup>50</sup>. Variant annotation and filtering was performed with both Annovar<sup>51</sup> and the Ingenuity Variant Analyzer web interface (Ingenuity System, CA). A subset of the identified variants were confirmed and analyzed by classical Sanger sequencing according standard protocols. Statistical analysis on rare variants present in inflammasome related genes was performed using the Graphpad on-line calculator (<http://graphpad.com>). A two-tails  $\chi^2$  test adjusted for  $2 \times 2$  contingency table was used.

## References

1. Touitou, I. & Kone-Paut, I. Autoinflammatory diseases. *Best Pract. Res. Clin. Rheumatol.* **22**, 811–829 (2008).
2. Martinon, F., Petrilli, V., Mayor, A., Tardivel, A. & Tschopp, J. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440**, 237–241 (2006).
3. Masters, S. L., Simon, A., Aksentjevich, I. & Kastner, D. L. Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease (\*). *Annu. Rev. Immunol.* **27**, 621–668 (2009).
4. Marshall, G. S., Edwards, K. M., Butler, J. & Lawton, A. R. Syndrome of periodic fever, pharyngitis, and aphthous stomatitis. *J. Pediatr.* **110**, 43–46 (1987).
5. Padeh, S. *et al.* Periodic fever, aphthous stomatitis, pharyngitis, and adenopathy syndrome: clinical characteristics and outcome. *J. Pediatr.* **135**, 98–101 (1999).
6. Padeh, S., Stoffman, N. & Berkun, Y. Periodic fever accompanied by aphthous stomatitis, pharyngitis and cervical adenitis syndrome (PFAPA syndrome) in adults. *Isr. Med. Assoc. J.* **10**, 358–360 (2008).
7. Cantarini, L., Vitale, A., Galeazzi, M. & Frediani, B. A case of resistant adult-onset periodic fever, aphthous stomatitis, pharyngitis and cervical adenitis (PFAPA) syndrome responsive to anakinra. *Clin. Exp. Rheumatol.* **30**, 593 (2012).
8. Colotto, M. *et al.* PFAPA syndrome in a young adult with a history of tonsillectomy. *Intern. Med.* **50**, 223–225 (2011).
9. Cazzato, M., Neri, R., Possemato, N., Puccini, R. & Bombardieri, S. A case of adult periodic fever, aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) syndrome associated with endocapillary proliferative glomerulonephritis. *Clin. Rheumatol.* **32** (Suppl 1), 33–36 (2013).
10. Stojanov, S. *et al.* Periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) is a disorder of innate immunity and Th1 activation responsive to IL-1 blockade. *Proc. Natl. Acad. Sci. USA* **108**, 7148–7153 (2011).
11. Kolly, L. *et al.* Periodic fever, aphthous stomatitis, pharyngitis, cervical adenitis syndrome is linked to dysregulated monocyte IL-1beta production. *J. Allergy Clin. Immunol.* **131**, 1635–1643 (2013).
12. Hofer, M., Mahlaoui, N. & Prieur, A. M. A child with a systemic febrile illness - differential diagnosis and management. *Best Pract. Res. Clin. Rheumatol.* **20**, 627–640 (2006).
13. Akelma, A. Z. *et al.* Is PFAPA syndrome really a sporadic disorder or is it genetic? *Med. Hypotheses* **81**, 279–281 (2013).
14. Sampaio, I. C., Rodrigo, M. J. & Monteiro Marques, J. G. Two siblings with periodic fever, aphthous stomatitis, pharyngitis, adenitis (PFAPA) syndrome. *The Pediatric infectious disease journal* **28**, 254–255 (2009).
15. Valenzuela, P. M., Majerson, D., Tapia, J. L. & Talesnik, E. Syndrome of periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) in siblings. *Clin. Rheumatol.* **28**, 1235–1237 (2009).

16. Cochard, M. *et al.* PFAPA syndrome is not a sporadic disease. *Rheumatology* **49**, 1984–1987 (2010).
17. Dagan, E., Gershoni-Baruch, R., Khatib, I., Mori, A. & Brik, R. MEFV, TNF1 $\alpha$ , CARD15 and NLRP3 mutation analysis in PFAPA. *Rheumatol. Int.* **30**, 633–636 (2010).
18. Kovacs, L. *et al.* Elevated immunoglobulin D levels in children with PFAPA syndrome. *Neuro. Endocrinol. Lett.* **31**, 743–746 (2010).
19. Gattorno, M. *et al.* Differentiating PFAPA syndrome from monogenic periodic fevers. *Pediatrics* **124**, e721–728 (2009).
20. Toplak, N. *et al.* An international registry on autoinflammatory diseases: the Eurofever experience. *Annals of the rheumatic diseases* **71**, 1177–1182 (2012).
21. Taniuchi, S. *et al.* MEFV Variants in Patients with PFAPA Syndrome in Japan. *Open Rheumatol. J.* **7**, 22–25 (2013).
22. Aganna, E. *et al.* Association of mutations in the NALP3/CIAS1/PYPAF1 gene with a broad phenotype including recurrent fever, cold sensitivity, sensorineural deafness, and AA amyloidosis. *Arthritis Rheum.* **46**, 2445–2452 (2002).
23. Porksen, G. *et al.* Periodic fever, mild arthralgias and reversible moderate and severe organ inflammation associated with the V198M mutation in the CIAS1 gene in three German patients - expanding phenotype of CIAS1 related autoinflammatory syndrome. *Eur. J. Haematol.* **73**, 123–127 (2004).
24. Loock, J. *et al.* Genetic predisposition (NLRP3 V198M mutation) for IL-1-mediated inflammation in a patient with Schnitzler syndrome. *The Journal of allergy and clinical immunology* **125**, 500–502 (2010).
25. Rowczenio, D. M. *et al.* Clinical characteristics in subjects with NLRP3 V198M diagnosed at a single UK center and a review of the literature. *Arthritis Res. Ther.* **15**, R30 (2013).
26. Bens, S. *et al.* SPAG7 is a candidate gene for the periodic fever, aphthous stomatitis, pharyngitis and adenopathy (PFAPA) syndrome. *Genes. Immun.* **15**, 190–194 (2014).
27. Fuentes Fajardo, K. V. *et al.* Detecting false-positive signals in exome sequencing. *Human mutation* **33**, 609–613 (2012).
28. Musumeci, L. *et al.* Single nucleotide differences (SNDs) in the dbSNP database may lead to errors in genotyping and haplotyping studies. *Hum. Mutat.* **31**, 67–73 (2010).
29. Hoffman, H. M., Mueller, J. L., Broide, D. H., Wanderer, A. A. & Kolodner, R. D. Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome. *Nature genetics* **29**, 301–305 (2001).
30. Pelagatti, M. A. *et al.* Long-term clinical profile of children with the low-penetrance R92Q mutation of the TNFRSF1A gene. *Arthritis Rheum.* **63**, 1141–1150 (2011).
31. Touitou, I., Perez, C., Dumont, B., Federici, L. & Jorgensen, C. Refractory auto-inflammatory syndrome associated with digenic transmission of low-penetrance tumour necrosis factor receptor-associated periodic syndrome and cryopyrin-associated periodic syndrome mutations. *Annals of the rheumatic diseases* **65**, 1530–1531 (2006).
32. Jeru, I. *et al.* Identification and functional consequences of a recurrent NLRP12 missense mutation in periodic fever syndromes. *Arthritis Rheum.* **63**, 1459–1464 (2011).
33. Verma, D. *et al.* The Q705K polymorphism in NLRP3 is a gain-of-function alteration leading to excessive interleukin-1 $\beta$  and IL-18 production. *PLoS one* **7**, e34977 (2012).
34. Touitou, I. *et al.* Infefers: An evolving mutation database for auto-inflammatory syndromes. *Human mutation* **24**, 194–198 (2004).
35. Zambetti, L. P., Laudisi, F., Licandro, G., Ricciardi-Castagnoli, P. & Mortellaro, A. The rhapsody of NLRPs: master players of inflammation and a lot more. *Immunol. Res.* **53**, 78–90 (2012).
36. Correa, R. G., Milutinovic, S. & Reed, J. C. Roles of NOD1 (NLR1) and NOD2 (NLR2) in innate immunity and inflammatory diseases. *Bioscience reports* **32**, 597–608 (2012).
37. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
38. Anderson, J. P. *et al.* Structural, expression, and evolutionary analysis of mouse CIAS1. *Gene*. **338**, 25–34 (2004).
39. Levandowski, C. B. *et al.* NLRP1 haplotypes associated with vitiligo and autoimmunity increase interleukin-1 $\beta$  processing via the NLRP1 inflammasome. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2952–2956 (2013).
40. Masters, S. L. *et al.* NLRP1 inflammasome activation induces pyroptosis of hematopoietic progenitor cells. *Immunity* **37**, 1009–1023 (2012).
41. Proell, M., Riedl, S. J., Fritz, J. H., Rojas, A. M. & Schwarzenbacher, R. The Nod-like receptor (NLR) family: a tale of similarities and differences. *PLoS one* **3**, e2119 (2008).
42. Albrecht, M., Domingues, F. S., Schreiber, S. & Lengauer, T. Structural localization of disease-associated sequence variations in the NACHT and LRR domains of PYPAF1 and NOD2. *FEBS Lett.* **554**, 520–528 (2003).
43. Aksentjevich, I. *et al.* The clinical continuum of cryopyrinopathies: novel CIAS1 mutations in North American patients and a new cryopyrin model. *Arthritis Rheum.* **56**, 1273–1285 (2007).
44. Katsanis, N. *et al.* Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256–2259 (2001).
45. Eichers, E. R., Lewis, R. A., Katsanis, N. & Lupski, J. R. Triallelic inheritance: a bridge between Mendelian and multifactorial traits. *Ann. Med.* **36**, 262–272 (2004).
46. Davis, E. E. & Katsanis, N. The ciliopathies: a transitional model into systems biology of human genetic disease. *Curr. Opin. Genet. Dev.* **22**, 290–303 (2012).
47. Thomas, K. T., Feder, H. M., Jr., Lawton, A. R. & Edwards, K. M. Periodic fever syndrome in children. *The Journal of pediatrics* **135**, 15–21 (1999).
48. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
49. Brickell, P. M. Immortalization of human B-lymphocytes by Epstein-Barr virus. *Methods in molecular biology* **8**, 213–218 (1992).
50. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491–498 (2011).
51. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164 (2010).

## Acknowledgements

This work was supported by the Swiss National Science Foundation (Grant 310030\_138346), the Gebert R uf Foundation, Switzerland (Rare Diseases - New Technologies grant), and the Novartis Foundation.

## **PROJECT 7 - ARTICLE**

### **Mutations in *CEP78* Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects**

Published in *The American Journal of Human Genetics*, September 1<sup>st</sup> 2016

#### **Contribution**

My contribution to this paper consisted in the preparation of cell cultures of primary skin fibroblasts starting from skin biopsies of two patients. In addition I made Figure 2, using the raw data of the pure tone audiometry from patients and from publicly available measurements.

## Mutations in *CEP78* Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects

Konstantinos Nikopoulos,<sup>1,12</sup> Pietro Farinelli,<sup>1,12</sup> Basilio Giangreco,<sup>2</sup> Chrysanthi Tsika,<sup>3</sup> Beryl Royer-Bertrand,<sup>1,4</sup> Martial K. Mbefo,<sup>5</sup> Nicola Bedoni,<sup>1</sup> Ulrika Kjellström,<sup>6</sup> Ikram El Zaoui,<sup>1</sup> Silvio Alessandro Di Gioia,<sup>1</sup> Sara Balzano,<sup>1</sup> Katarina Cisarova,<sup>1</sup> Andrea Messina,<sup>7</sup> Sarah Decembrini,<sup>5</sup> Sotiris Plainis,<sup>3</sup> Styliani V. Blazaki,<sup>3</sup> Muhammad Imran Khan,<sup>8</sup> Shazia Micheal,<sup>8</sup> Karsten Boldt,<sup>9</sup> Marius Ueffing,<sup>9</sup> Alexandre P. Moulin,<sup>10</sup> Frans P.M. Cremers,<sup>8,11</sup> Ronald Roepman,<sup>8</sup> Yvan Arsenijevic,<sup>5</sup> Miltiadis K. Tsilimbaris,<sup>3</sup> Sten Andréasson,<sup>6</sup> and Carlo Rivolta<sup>1,\*</sup>

Cone-rod degeneration (CRD) belongs to the disease spectrum of retinal degenerations, a group of hereditary disorders characterized by an extreme clinical and genetic heterogeneity. It mainly differentiates from other retinal dystrophies, and in particular from the more frequent disease retinitis pigmentosa, because cone photoreceptors degenerate at a higher rate than rod photoreceptors, causing severe deficiency of central vision. After exome analysis of a cohort of individuals with CRD, we identified biallelic mutations in the orphan gene *CEP78* in three subjects from two families: one from Greece and another from Sweden. The Greek subject, from the island of Crete, was homozygous for the c.499+1G>T (IVS3+1G>T) mutation in intron 3. The Swedish subjects, two siblings, were compound heterozygotes for the nearby mutation c.499+5G>A (IVS3+5G>A) and for the frameshift-causing variant c.633delC (p.Trp212Glyfs\*18). In addition to CRD, these three individuals had hearing loss or hearing deficit. Immunostaining highlighted the presence of CEP78 in the inner segments of retinal photoreceptors, predominantly of cones, and at the base of the primary cilium of fibroblasts. Interaction studies also showed that CEP78 binds to FAM161A, another ciliary protein associated with retinal degeneration. Finally, analysis of skin fibroblasts derived from affected individuals revealed abnormal ciliary morphology, as compared to that of control cells. Altogether, our data strongly suggest that mutations in *CEP78* cause a previously undescribed clinical entity of a ciliary nature characterized by blindness and deafness but clearly distinct from Usher syndrome, a condition for which visual impairment is due to retinitis pigmentosa.

Cone-rod degeneration (CRD [MIM: 120970]) represents an extremely rare class of hereditary diseases that affect the light-sensing neurons of the retina, the cone and rod photoreceptors.<sup>1</sup> Cones are involved in daytime vision, providing the brain with color information and central, precise visual input. Conversely, rods are active in very dim light conditions, are more abundant in the retinal periphery, and produce achromatic information, typical for instance of the visual stimulation provided by a landscape on a moonless night. Individuals with CRD experience initial loss of visual acuity (central vision) and aberrant color vision as a result of the prominent loss of cones, whereas rod functions remain relatively preserved.<sup>2</sup> As the disease progresses, both cone and rod functions deteriorate and central vision is severely impaired or lost, but peripheral islands of the retina might retain some residual activity.<sup>3</sup> On the basis of these clinical parameters, CRD can be distinguished from retinitis pigmentosa (RP [MIM: 26800], also called rod-cone degeneration), the most common form of hereditary retinal degeneration. In retinitis pigmentosa, rods are more severely affected than cones;

initial symptoms include night blindness due to loss of rod function, and central vision (cone function) is often preserved until the very late stages of the disease.<sup>4</sup> CRD is almost invariably inherited as a Mendelian trait, predominantly according to a recessive pattern of transmission, and is characterized by an elevated genetic and allelic heterogeneity.<sup>5</sup> Although as many as 33 CRD-associated genes have been identified to date (RetNet; see [Web Resources](#)), they are found to be mutated in only ~25% of clinical cases, implying that a substantial percentage of affected people might carry mutations in yet-to-be-identified genes.<sup>6</sup>

According to this rationale, we performed whole-exome sequencing (WES) in 34 unrelated individuals with CRD (29 from Greece and five from Sweden). Genomic DNA was extracted from peripheral blood leukocytes according to standard procedures, and then exomic libraries (SureSelect V5 kit, Agilent) were sequenced on an Illumina HiSeq 2000. Raw sequence files were assessed, trimmed, and finally mapped back to the human genome reference sequence (build hg19); DNA variants were called

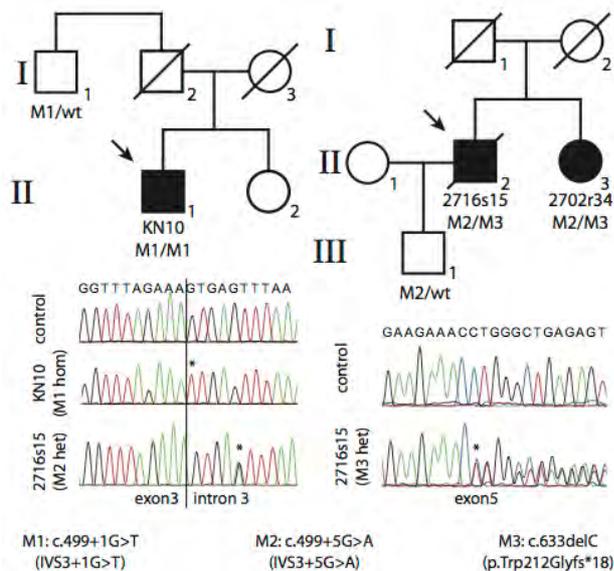
<sup>1</sup>Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, 1011 Lausanne, Switzerland; <sup>2</sup>Center for Psychiatric Neuroscience, Department of Psychiatry, Lausanne University Hospital, 1008 Lausanne, Switzerland; <sup>3</sup>Department of Ophthalmology, Medical School, University of Crete, 71003 Heraklion, Greece; <sup>4</sup>Center for Molecular Diseases, Lausanne University Hospital, 1011 Lausanne, Switzerland; <sup>5</sup>Unit of Gene Therapy and Stem Cell Biology, Jules Gonin Ophthalmic Hospital, 1004 Lausanne, Switzerland; <sup>6</sup>Lund University, Skane University Hospital, Department of Ophthalmology, 20502 Lund, Sweden; <sup>7</sup>Service of Endocrinology, Diabetes and Metabolism, Lausanne University Hospital, 1011 Lausanne, Switzerland; <sup>8</sup>Department of Human Genetics and Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525GA Nijmegen, The Netherlands; <sup>9</sup>Medical Proteome Center, Institute for Ophthalmic Research, University of Tuebingen, 72074 Tuebingen, Germany; <sup>10</sup>Department of Ophthalmology, Jules Gonin Ophthalmic Hospital, 1004 Lausanne, Switzerland; <sup>11</sup>Donders Center for Neuroscience, Radboud University, 6525EN Nijmegen, The Netherlands

<sup>12</sup>These authors contributed equally to this work

\*Correspondence: carlo.rivolta@unil.ch

<http://dx.doi.org/10.1016/j.ajhg.2016.07.009>

© 2016 American Society of Human Genetics.



**Figure 1. CEP78 Mutations**

Pedigrees and electropherograms of the identified DNA changes. Asterisks indicate the site of mutations. Abbreviations are as follows: Mx, mutant alleles; wt, wild-type allele.

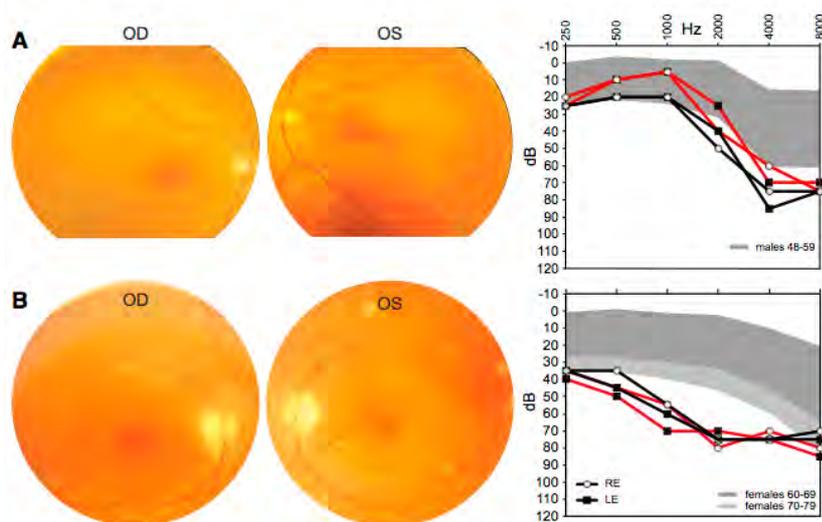
and scored according to a specific in silico pipeline, described previously.<sup>7</sup> Aggregate data analysis and variant filtering procedures (Tables S1 and S2 in the Supplemental Data available online) revealed biallelic mutations in two probands, one from Greece and another from Sweden, in *CEP78* (centrosomal protein 78, composed of 16 exons for its longest coding isoform [GenBank: NM\_001098802.1]). Both had classical signs and symptoms of CRD, clearly distinct from retinitis pigmentosa, as detailed below. Written informed consent was obtained from all individuals enrolled in this study, and approval for research on human subjects was obtained from the institutional review boards of all participating Institutions.

The Greek subject (KN10, individual II-1 on the left pedigree in Figure 1) was a 59-year-old male from the island of Crete and the eldest of two siblings of a non-consanguineous family. His sister was unaffected, and the family reported no history of retinal degeneration. Clinical history indicated hemarelopiia since early adulthood (18–20 years of age); the condition progressed to severe central vision loss at the age of 35–40 years and evolved into severe visual impairment, nystagmus, and photophobia. Dyschromatopsia was also reported. Fundus examination at first visit showed normal color and normal vessels but a small atrophic foveal area with subjacent ring-like glistening in one eye and bull's-eye maculopathy in the other eye. A few atrophic lesions were present in the inferior periphery in one eye (Figure 2). The 30° static automated perimetry revealed a diffuse suppression of the visual field in both eyes and a relative conservation of the peripapillary and superior periphery. Full-field electroretinography (ERG) showed flat cone responses but still some residual rod-

mediated signals in the left eye. This person also complained about minor hearing problems, and his audiogram exhibited relatively mild deficit; nonetheless, such deficit was clearly distinct and more severe than natural age-related hearing loss (presbycusis)<sup>9</sup> (Figure 2). KN10 carried a homozygous substitution in the first invariant base of intron 3 splice donor site c.499+1G>T (IVS3+1G>T) (Figure 1). This variant was located within a very small stretch of homozygosity that was not statistically significant for autozygosity, possibly indicating a mutational founder effect of geographic origin (not shown). The only relative who could be tested was his paternal uncle (individual I-1, left pedigree), who carried this DNA change heterozygously (Figure 1).

The Swedish proband (2716s15, individual II-2 on the right pedigree in Figure 1), now deceased, was last examined at 69 years of age. He was born from unaffected parents and was the first child of a kindred of two. His sister (2702r34, individual II-3, right pedigree), examined at age 65, also had retinal degeneration. Both had visual problems, including loss of color sensitivity and central vision, since childhood. Both also reported a hearing deficit since they were young, and both had hearing aids. Audiogram of the living Swedish subject at age 66 years revealed substantial sensorineural hearing loss, which did not seem to progress substantially over the following 11 years (Figure 2). Hospital records containing information on the hearing status of her deceased brother were destroyed upon his death. Fundus examination showed degenerative changes for both siblings in the macular region and some spicular pigment in the mid-periphery but fewer changes in the periphery (Figure 2). Progressive deterioration of the visual field was reported and documented as expanding from the center to the periphery. At last examinations, both siblings retained some residual vision at the periphery of the visual field, especially in dim-light conditions. Similar to the situation for the Greek subject, full-field ERG of both individuals highlighted almost no residual cone activity but still revealed some rod-mediated responses, even at these late ages. These siblings were compound heterozygotes for two *CEP78* mutations: a frameshift-causing single-nucleotide deletion (c.633delC; p.Trp212Glyfs\*18) in exon 5 and an intronic base substitution (c.499+5G>A; IVS3+5G>A) in the vicinity of the donor site for intron 3, just four nucleotides away from the mutation identified in the Greek subject. Genetic examination of the proband's son (individual III-1, right pedigree) uncovered the presence of this latter mutation in heterozygosity, confirming the biallelic nature of the changes detected in his father and his aunt (Figure 1).

Sanger sequencing of the entire reading frame of *CEP78* in a cohort of 99 unrelated CRD-affected individuals of Swedish, Swiss, Dutch, and Pakistani ethnic background failed to identify any additional causative variants. The three mutations present in our two families were not detected in the genome of an internal control cohort of 350 unrelated individuals or in any other public



**Figure 2. Clinical Features of the Analyzed Subjects**

(A) Individual KN10. Fundus pictures (at age 53 years) reveal macular coalescent hypochromatic lesions in the right eye and minor macular changes in the left eye. Pure tone audiograms at age 57 years (black lines) and 59 years (red lines) show mild hearing impairment at higher frequencies compared to the normal range for gender and age, indicated by the shaded area.<sup>8</sup>

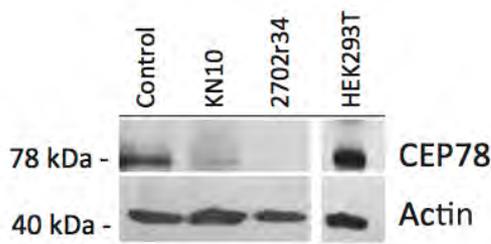
(B) Individual 2702r34. Fundus examination at age 65 years highlights attenuated vessels and degenerative changes in the posterior pole. Audiograms at age 66 years (black lines) and 77 years (red lines) show clear hearing loss at most frequencies. Abbreviations are as follows: OD, right eye; OS, left eye; RE, right ear; and LE, left ear.

database, including the 1000 Genomes Project, the Exome Variant Server (EVS), and the Exome Aggregation Consortium (ExAC) Browser, which reports sequencing data from more than 61,000 unrelated individuals. In addition, in silico assessment of the c.499+1G>T and c.499+5G>A mutations via two distinct web-based platforms, NNSPLICE 0.9<sup>10</sup> and Human Splicing Finder,<sup>11</sup> predicted for both variants the abolishment of the donor splicing site for intron 3.

To analyze the functional consequences of the three mutations, we obtained fresh blood samples and skin biopsies from the Greek proband and the living Swedish subject and performed the following experiments. We first retro-transcribed total RNA from immortalized lymphoblasts (GoScript Reverse Transcriptase, Promega). Then, after performing saturating RT-PCR of the region spanning all mutations (primers: 5'-TTTTGCAGAAGTCGTGTCCT-3' and 5'-TTCAAGGGCCTCTAGCAAAG-3'), we cloned the amplified products in *E. coli* (Zero Blunt PCR Cloning Kits, Invitrogen) and performed colony PCRs and capillary electrophoresis on 96 clones (48 clones per affected individual). Representative samples were Sanger sequenced, and relative numbers of splicing events were assessed and quantified. The c.499+1G>T mutation resulted invariably in the skipping of exon 3, leading to the production of an aberrant isoform, never reported in genomic databases, for which exons 2 and 4 were joined together. This event ablated 24 codons and altered the reading frame of *CEP78*, leading to the formation of a premature termination codon at nucleotides 16–18 of exon 4. Therefore, this non-canonical transcript was predicted to trigger nonsense-mediated mRNA decay (NMD)<sup>12</sup> and result in no viable mRNA. The same exon-skipping occurrence was observed for the other, nearby mutation, c.499+5G>A, that was present in the Swedish subjects. Finally, the frameshift mutation c.633delC resulted in reduced mRNA amounts, as deduced by the low number of *E. coli* colonies carrying this cDNA clone (4 out of 48), again,

probably as a result of the action of NMD. Immunoblot analysis in fibroblasts' extracts (antibodies: A301-799A, Bethyl Laboratories, epitope between residues 550 and 600 and A2066, Sigma, for beta actin) revealed the presence of *CEP78* in very reduced amounts in the Greek subject and the absence of any detectable band in the Swedish subject, in agreement with the mRNA findings described above (Figure 3). More specifically, the homozygous c.499+1G>T mutation probably resulted in a few canonical mRNA forms not detected by our cloning experiments, in turn producing small amounts of *CEP78*. Concerning the Swedish subject, it is likely that both the c.633delC and the c.499+5G>A alleles produced mostly non-viable mRNA and minimal quantities of wild-type mRNA and protein. In fact, overexposed films showed a faint band corresponding to *CEP78*, indicating that the protein was present in trace amounts (not shown).

To gain insights into the relationship between vision and *CEP78*, we analyzed its expression in a panel of human tissues of cadaveric origin (Human Total RNA Master Panel II, Takara, primers 5'-GTTTCCCATTAATCAAAACACG-3' and 5'-TCAACTTCAGAGGATGAAGGACT-3' for *CEP78* and 5'-AGAGTGGTGTGAGGATTGG-3' and 5'-CCCTC ATGCTCTAGCGTGTC-3' for the housekeeping gene *GUSB*). Although the number of *CEP78* transcripts in the retina was higher than in many other tissues and organs, retinal *CEP78* mRNA did not display the highest expression level (Figure 4). A time-course experiment on eyes of developing and postnatal mice (primers 5'-CTTCAGA AAGTGTCCAGGAAGC-3' and 5'-GATCACTCTCTCCTCC TTCAGC-3' for *Cep78* and 5'-CCTAAGATGAGCGCAAGTT GAA-3' and 5'-CCACAGGACTAGAACACCTGCTAA-3' for the housekeeping gene *Hprt1*) revealed high expression at embryonic stages, followed by a progressive decrease at perinatal stages and by a plateau at adulthood (Figure 4). This pattern is reminiscent of the expression of other genes involved in retinal degenerations and in early biogenesis and homeostasis of the centriole, notably



**Figure 3. Immunoblot Analysis of Endogenous CEP78 in Fibroblasts from Affected Individuals**

Numbers on the left refer to molecular-size markers. *Control* indicates human fibroblasts from a control individual; *HEK293T* indicates HEK293 cells containing SV40 large T antigen as a control for specificity of the anti-CEP78 antibody.

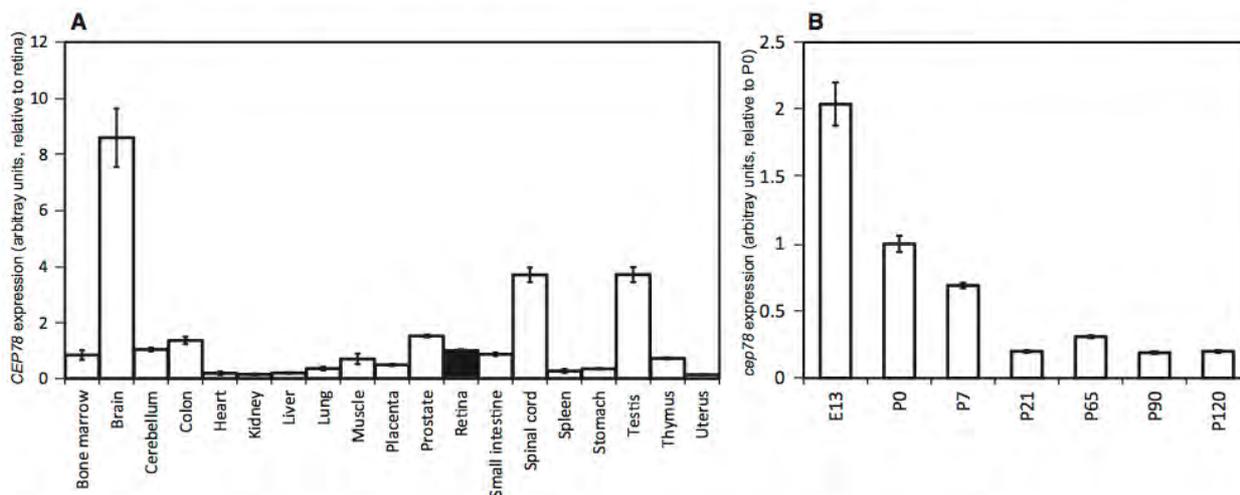
*CEP76*, *CEP110*, *CEP164* (MIM: 614848), *RAB8A* (MIM: 165040), *BBS4* (MIM: 600374), and *RPGR* (MIM: 312610).<sup>13</sup> Although we could not obtain primary data on *CEP78* expression in human cochlea, we assessed its RNaseq values from the only comprehensive transcriptome repository currently available for the inner ear.<sup>14</sup> *CEP78* displayed a FPKM (fragments per kilobase of exon per million reads mapped) value of 4.53 (average from three human cochleae), indicating moderate expression in this structure. Importantly, this FPKM value for *CEP78* appeared to be higher than that of most genes already known to be associated with hereditary deafness (40 out of 70, or 57%), as assessed in the same organs and in the same conditions (Table S3).

Little is known about the function of CEP78. Identified as a component of the centrosome by two independent proteomic screenings,<sup>15,16</sup> CEP78 is composed of five leucine-rich repeats located at the N-terminal half of the protein, as well as a coiled-coil domain at the C terminus. An important study using *Planaria* as the main experimental model revealed that miRNA-based knockdown of

*CEP78* resulted in defective primary cilia assembly in flatworms and human RPE1 cells.<sup>17</sup> Intriguingly, *CEP78* was also found upregulated more than 5-fold by noise stress in rat cochlea.<sup>18</sup> The function and impact of CEP78 in human physiology, however, remain largely elusive.

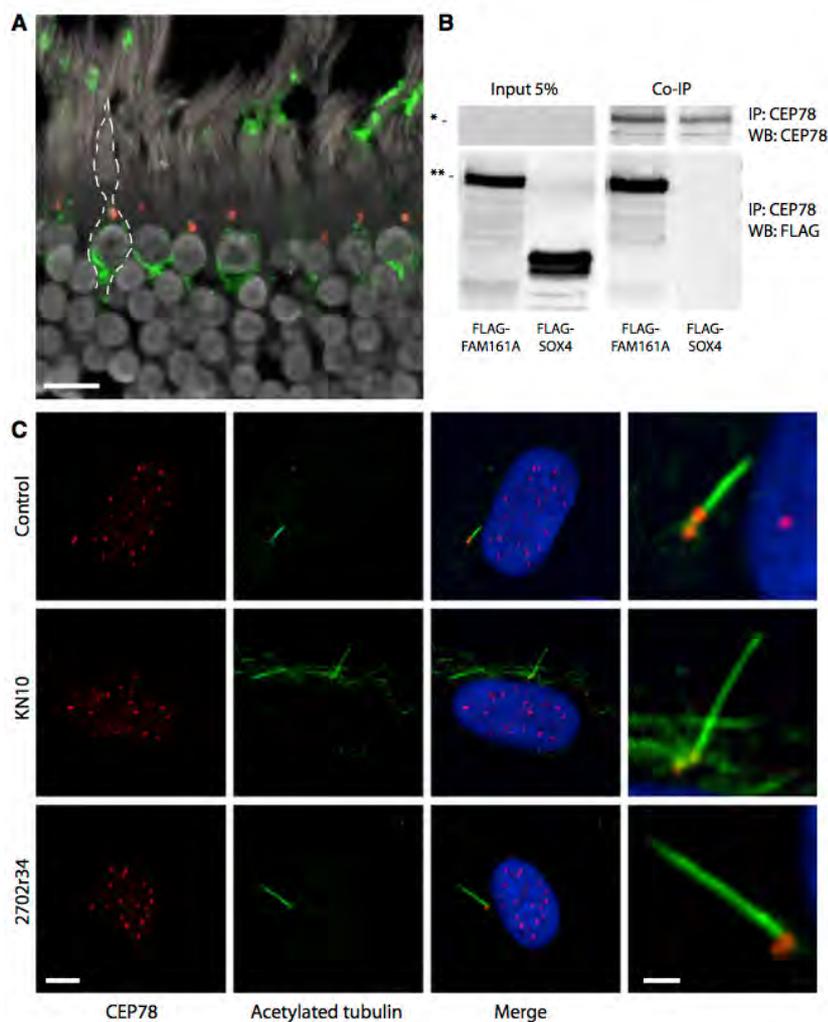
The reported centrosomal localization of CEP78 prompted us to investigate a possible role in relationship to the photoreceptor primary cilium. Immunofluorescence of human retinal sections with anti-CEP78 antibody (IHC-00364, Bethyl Laboratories, epitope between residues 550 and 600) and anti-cone arrestin (SC-54355, Santa Cruz) showed that CEP78 is present in dot-shaped foci in the inner segments, probably at the base of the primary cilium in retinal photoreceptors, predominantly cones (Figure 5 and Figure S1). This observation was confirmed when CEP78 was labeled together with acetylated tubulin, staining the primary cilium of human skin fibroblasts (Figure 5). Interestingly, positive staining was observed in fibroblasts from KN10 and 2702r34 as well, confirming that *CEP78* was in fact expressed at the protein level in these individuals, as inferred (Figure 5). The mild differences between these experiments and the protein-expression observations might be due to different sensitivities of the two analytical tools and the use of distinct antibodies (A301-799A for immunoblot and IHC-00364 for immunofluorescence). No specific differences concerning CEP78 subcellular localization were observed in cells from affected individuals versus cells from controls.

Presence at the base of the connecting cilium is a characteristic that is shared by other proteins associated with retinal degeneration, and in particular by FAM161A, the deficiency of which causes the RP28 form of retinitis pigmentosa (MIM: 606068).<sup>19–21</sup> Indeed, tandem-affinity purification analysis performed with full-length FAM161A showed a positive interaction with CEP78,<sup>22</sup> and co-immunoprecipitation with an anti-CEP78 antibody (A301-800A, Bethyl Laboratories, epitope between residues 639 and



**Figure 4. CEP78 mRNA Expression in Various Human Tissues and Organs and in the Developing Murine Eye**

Data are from real-time PCR relative expression analysis, for which *GUSB* and *Hprt1* were used as normalizing genes for (A) and (B), respectively.



**Figure 5. CEP78 in Human Cells and Its Interaction with FAM161A**

(A) Immunostaining of CEP78 (red dots) and of cone arrestin (green) in a section of human retina. Margins of a cone photoreceptor are highlighted by a dotted line. The scale bar represents 10  $\mu\text{m}$ .

(B) Co-immunoprecipitation of endogenous CEP78 and FLAG-tagged FAM161A in HEK293T cells. FLAG-SOX4 is a negative control. Abbreviations are as follows: IP:x, protein targeted by the antibody used in immunoprecipitation; and WB:x, protein or peptide targeted by the antibody used in immunoblots. The single and double asterisks indicate the expected molecular sizes for CEP78 (78 kDa) and the FLAG-FAM161A construct (90 kDa), respectively.

(C) Staining of CEP78 (red) and acetylated tubulin (green) in fibroblasts from a control individual, KN10, and 2702r34. CEP78 localizes at the centrioles and at the base of the induced primary cilium. The scale bars represent 5 and 1  $\mu\text{m}$  for regular and magnification panels, respectively.

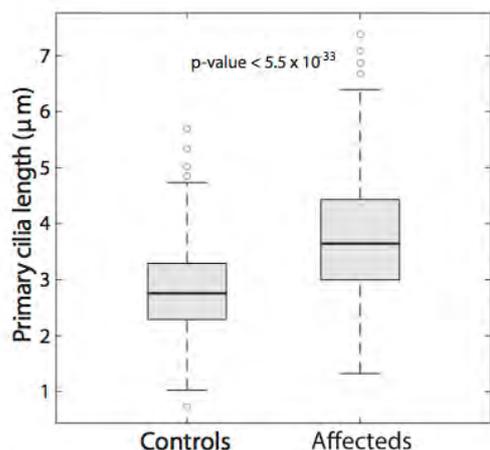
In recent years, a significant number of hereditary conditions have been recognized to be the consequence of abnormalities of the cellular cilium. These diseases, collectively called ciliopathies, form a genetically heterogeneous spectrum of disorders affecting various tissues and organs, for instance kidney, cochlea, brain, and retina.<sup>23,27</sup> Classical examples of ciliopathies involving retina and other tissues are Usher

syndrome (MIM: 276900, blindness and deafness) and Bardet-Biedl syndrome (BBS [MIM: 209900], blindness and multi-organ defects) for both of which vision loss is due to retinitis pigmentosa.<sup>28</sup> In addition, syndromic ciliopathies such as Senior-Løken syndrome (SLS [MIM: 266900]), Joubert syndrome (JBTS [MIM: 213300]), and Jeune syndrome (JATD [MIM: 616300]) can occasionally be accompanied by retinal dystrophy and, in particular, retinitis pigmentosa and/or Leber congenital amaurosis (LCA [MIM: 204000]).<sup>29–32</sup> Another multi-organ ciliopathy is Alström syndrome (ALMS [MIM: 203800]),<sup>33</sup> caused by null mutations in the gene *ALMS1* (MIM: 606844). This disease is characterized by cone-rod degeneration, dilated cardiomyopathy, obesity, type 2 diabetes, and short stature, which can be accompanied by hepatosteatorrhea and defects in the lungs, kidney, and bladder.<sup>33,34</sup> Most cases also display progressive sensorineural hearing loss.<sup>35</sup> Of interest, mutations in *ALMS1* have recently been suggested as being causative of non-syndromic CRD.<sup>36</sup>

689) in HEK293T cells transfected with full-length FLAG-FAM161A revealed direct binding between these two ciliary proteins (Figure 5). On the basis of these findings, we speculated that CRD due to mutations in *CEP78* could be a consequence of hindered ciliary function, similar to what occurs in many other retinal degenerations.<sup>23</sup> To test this hypothesis, we analyzed the morphology of primary cilia in fibroblasts derived from KN10 and 2702r34, with respect to four controls, after serum starvation. Unsupervised counting of at least 82 events per sample (207 events in affected individuals and 430 in controls) revealed that induced cilia in fibroblasts from KN10 and 2702r34 were significantly longer than those from control cells (Figure 6 and Figure S2), a phenomenon that has been previously associated with impaired function of this organelle. For instance, mutations in murine orthologs of *BBS4*, *ICK*, and *TSC1*, linked with ciliopathies such as Bardet-Biedl syndrome, endocrine-cerebro-osteodysplasia, and tuberous sclerosis, respectively, display kidney cells with elongated primary cilia.<sup>24–26</sup>

syndrome (MIM: 276900, blindness and deafness) and Bardet-Biedl syndrome (BBS [MIM: 209900], blindness and multi-organ defects) for both of which vision loss is due to retinitis pigmentosa.<sup>28</sup> In addition, syndromic ciliopathies such as Senior-Løken syndrome (SLS [MIM: 266900]), Joubert syndrome (JBTS [MIM: 213300]), and Jeune syndrome (JATD [MIM: 616300]) can occasionally be accompanied by retinal dystrophy and, in particular, retinitis pigmentosa and/or Leber congenital amaurosis (LCA [MIM: 204000]).<sup>29–32</sup> Another multi-organ ciliopathy is Alström syndrome (ALMS [MIM: 203800]),<sup>33</sup> caused by null mutations in the gene *ALMS1* (MIM: 606844). This disease is characterized by cone-rod degeneration, dilated cardiomyopathy, obesity, type 2 diabetes, and short stature, which can be accompanied by hepatosteatorrhea and defects in the lungs, kidney, and bladder.<sup>33,34</sup> Most cases also display progressive sensorineural hearing loss.<sup>35</sup> Of interest, mutations in *ALMS1* have recently been suggested as being causative of non-syndromic CRD.<sup>36</sup>

In this study, we show that mutations in *CEP78* result in cone-rod degeneration associated with hearing loss,



**Figure 6. Analysis of Ciliary Lengths**

Fibroblasts from affected subjects display significantly longer primary cilia than those from four unaffected controls, suggesting impaired functionality.

another hallmark of ciliopathy, but no other syndromic features. Interestingly, the two Swedish affected individuals had declared hearing loss, and the Greek subject had borderline hearing impairment. An intriguing possibility involves the presence of a genotype-phenotype correlation between *CEP78* alleles and hearing (but not vision), as it is the case for instance for mutations in *USH2A* and *ALMS1*.<sup>36–38</sup> Taken together, our data indicate that genetic defects in *CEP78* define a newly recognized ciliopathy, distinct from Usher and Alström syndromes, affecting both the visual and the hearing systems.

#### Accession Numbers

The mutations reported in this paper have been deposited in the Leiden Open Variation Database (LOVD) under variant ID numbers 0000119260, 0000119261, and 0000119262.

#### Supplemental Data

Supplemental Data include two figures and three tables and are available with this article online at <http://dx.doi.org/10.1016/j.ajhg.2016.07.009>.

#### Acknowledgments

This work was supported by the Swiss National Science Foundation (grant #156260, to C.R.), by the European Community's Seventh Framework Programmes FP7/2009 under grant agreement #241955 (SYSCILIA) to R.R. and M.U., by the Netherlands Organization for Scientific Research (NWO Vici-865.12.005 to R.R.), and by the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, the Stichting A.F. Deutman Researchfonds Oogheelkunde, and the Stichting voor Ooglijders (to F.P.M.C. and M.I.K.). F.P.M.C. and M.I.K. were also supported by the following foundations: the Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, the Landelijke Stichting voor Blinden en Slechtzienden, the Stichting Retina Nederland Fonds, and the Novartis fund, contributed through UitZicht. We would also like to

acknowledge Drs. Andrea Superti-Furga and Luisa Bonafé from the Center for Molecular Diseases of the Lausanne University Hospital, Dr. Frauke Coppieters from the Ghent University Hospital, Dr. Guy Van Camp from the University of Antwerp, and Dr. Isabelle Schrauwen from the Translational Genomics Research Institute in Phoenix.

Received: March 17, 2016

Accepted: July 7, 2016

Published: September 1, 2016

#### Web Resources

1000 Genomes, <http://browser.1000genomes.org>

ExAC Browser, <http://exac.broadinstitute.org/>

GenBank, <http://www.ncbi.nlm.nih.gov/genbank/>

Leiden Open Variation Database (LOVD), <http://www.lovd.nl/3.0/home>

NHLBI Exome Sequencing Project (ESP) Exome Variant Server, <http://evs.gs.washington.edu/EVS/>

OMIM, <http://www.omim.org/>

RefSeq, <http://www.ncbi.nlm.nih.gov/RefSeq>

RetNet, <https://sph.uth.edu/retnet/>

#### References

- Hamel, C.P. (2007). Cone rod dystrophies. *Orphanet J. Rare Dis.* 2, 7.
- Berson, E.L., Gouras, P., and Gunkel, R.D. (1968). Progressive cone-rod degeneration. *Arch. Ophthalmol.* 80, 68–76.
- Krauss, H.R., and Heckenlively, J.R. (1982). Visual field changes in cone-rod degenerations. *Arch. Ophthalmol.* 100, 1784–1790.
- Berson, E.L. (1993). Retinitis pigmentosa. The Friedenwald Lecture. *Invest. Ophthalmol. Vis. Sci.* 34, 1659–1676.
- Berger, W., Kloeckener-Gruissem, B., and Neidhardt, J. (2010). The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.* 29, 335–375.
- Roosing, S., Thiadens, A.A., Hoyng, C.B., Klaver, C.C., den Hollander, A.I., and Cremers, F.P. (2014). Causes and consequences of inherited cone disorders. *Prog. Retin. Eye Res.* 42, 1–26.
- Royer-Bertrand, B., Castillo-Taucher, S., Moreno-Salinas, R., Cho, T.J., Chae, J.H., Choi, M., Kim, O.H., Dikoglu, E., Campos-Xavier, B., Girardi, E., et al. (2015). Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. *Sci. Rep.* 5, 17154.
- Cruickshanks, K.J., Wiley, T.L., Tweed, T.S., Klein, B.E., Klein, R., Mares-Perlman, J.A., and Nondahl, D.M.; The Epidemiology of Hearing Loss Study (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. *Am. J. Epidemiol.* 148, 879–886.
- Jerger, J., Chmiel, R., Stach, B., and Spretnjak, M. (1993). Gender affects audiometric shape in presbycusis. *J. Am. Acad. Audiol.* 4, 42–49.
- Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* 4, 311–323.
- Desmet, F.O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., and Bérout, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37, e67.

12. Hentze, M.W., and Kulozik, A.E. (1999). A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* 96, 307–310.
13. Zhang, S.S., Xu, X., Liu, M.G., Zhao, H., Soares, M.B., Barnstable, C.J., and Fu, X.Y. (2006). A biphasic pattern of gene expression during mouse retina development. *BMC Dev. Biol.* 6, 48.
14. Schrauwen, I., Hasin-Brumshtein, Y., Corneveaux, J.J., Ohmen, J., White, C., Allen, A.N., Lusic, A.J., Van Camp, G., Huentelman, M.J., and Friedman, R.A. (2016). A comprehensive catalogue of the coding and non-coding transcripts of the human inner ear. *Hear. Res.* 333, 266–274.
15. Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* 426, 570–574.
16. Jakobsen, L., Vanselow, K., Skogs, M., Toyoda, Y., Lundberg, E., Poser, I., Falkenby, L.G., Bennetzen, M., Westendorf, J., Nigg, E.A., et al. (2011). Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *EMBO J.* 30, 1520–1535.
17. Azimzadeh, J., Wong, M.L., Downhour, D.M., Sánchez Alvarado, A., and Marshall, W.F. (2012). Centrosome loss in the evolution of planarians. *Science* 335, 461–463.
18. Han, Y., Hong, L., Zhong, C., Chen, Y., Wang, Y., Mao, X., Zhao, D., and Qiu, J. (2012). Identification of new altered genes in rat cochleae with noise-induced hearing loss. *Gene* 499, 318–322.
19. Bandah-Rozenfeld, D., Mizrahi-Meissonnier, L., Farhy, C., Obolensky, A., Chowers, I., Pe'er, J., Merin, S., Ben-Yosef, T., Ashery-Padan, R., Banin, E., and Sharon, D. (2010). Homozygosity mapping reveals null mutations in *FAM161A* as a cause of autosomal-recessive retinitis pigmentosa. *Am. J. Hum. Genet.* 87, 382–391.
20. Langmann, T., Di Gioia, S.A., Rau, I., Stöhr, H., Maksimovic, N.S., Corbo, J.C., Renner, A.B., Zrenner, E., Kumaramanickavel, G., Karlstetter, M., et al. (2010). Nonsense mutations in *FAM161A* cause RP28-associated recessive retinitis pigmentosa. *Am. J. Hum. Genet.* 87, 376–381.
21. Di Gioia, S.A., Letteboer, S.J., Kostic, C., Bandah-Rozenfeld, D., Hettterschijt, L., Sharon, D., Arsenijevic, Y., Roepman, R., and Rivolta, C. (2012). *FAM161A*, associated with retinitis pigmentosa, is a component of the cilia-basal body complex and interacts with proteins involved in ciliopathies. *Hum. Mol. Genet.* 21, 5174–5184.
22. Di Gioia, S.A. (2013). Targeted sequence capture and ultra high throughput sequencing for gene discovery in inherited diseases. PhD Thesis (University of Lausanne, Lausanne, Switzerland).
23. Wright, A.F., Chakarova, C.F., Abd El-Aziz, M.M., and Bhattacharya, S.S. (2010). Photoreceptor degeneration: Genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.* 11, 273–284.
24. Armour, E.A., Carson, R.P., and Ess, K.C. (2012). Cystogenesis and elongated primary cilia in *Tsc1*-deficient distal convoluted tubules. *Am. J. Physiol. Renal Physiol.* 303, F584–F592.
25. Moon, H., Song, J., Shin, J.O., Lee, H., Kim, H.K., Eggenschwiler, J.T., Bok, J., and Ko, H.W. (2014). Intestinal cell kinase, a protein associated with endocrine-cerebro-osteodysplasia syndrome, is a key regulator of cilia length and Hedgehog signaling. *Proc. Natl. Acad. Sci. USA* 111, 8541–8546.
26. Mokrzan, E.M., Lewis, J.S., and Mykytyn, K. (2007). Differences in renal tubule primary cilia length in a mouse model of Bardet-Biedl syndrome. *Nephron, Exp. Nephrol.* 106, e88–e96.
27. Waters, A.M., and Beales, P.L. (2011). Ciliopathies: An expanding disease spectrum. *Pediatr. Nephrol.* 26, 1039–1056.
28. Koenig, R. (2003). Bardet-Biedl syndrome and Usher syndrome. *Dev. Ophthalmol.* 37, 126–140.
29. Bachmann-Gagescu, R., Dempsey, J.C., Phelps, I.G., O’Roak, B.J., Knutzen, D.M., Rue, T.C., Ishak, G.E., Isabella, C.R., Gordon, N., Adkins, J., et al.; University of Washington Center for Mendelian Genomics (2015). Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. *J. Med. Genet.* 52, 514–522.
30. Bard, L.A., Bard, P.A., Owens, G.W., and Hall, B.D. (1978). Retinal involvement in thoracic-pelvic-phalangeal dystrophy. *Arch. Ophthalmol.* 96, 278–281.
31. Wilson, D.J., Weleber, R.G., and Beals, R.K. (1987). Retinal dystrophy in Jeune’s syndrome. *Arch. Ophthalmol.* 105, 651–657.
32. Hildebrandt, F., and Zhou, W. (2007). Nephronophthisis-associated ciliopathies. *J. Am. Soc. Nephrol.* 18, 1855–1871.
33. Marshall, J.D., Muller, J., Collin, G.B., Milan, G., Kingsmore, S.F., Dinwiddie, D., Farrow, E.G., Miller, N.A., Favaretto, F., Maffei, P., et al. (2015). Alström syndrome: Mutation spectrum of *ALMS1*. *Hum. Mutat.* 36, 660–668.
34. Marshall, J.D., Maffei, P., Collin, G.B., and Naggert, J.K. (2011). Alström syndrome: genetics and clinical overview. *Curr. Genomics* 12, 225–235.
35. Marshall, J.D., Bronson, R.T., Collin, G.B., Nordstrom, A.D., Maffei, P., Paisey, R.B., Carey, C., Macdermott, S., Russell-Eggitt, I., Shea, S.E., et al. (2005). New Alström syndrome phenotypes based on the evaluation of 182 cases. *Arch. Intern. Med.* 165, 675–683.
36. Lazar, C.H., Kimchi, A., Namburi, P., Mutsuddi, M., Zelinger, L., Beryozkin, A., Ben-Simhon, S., Obolensky, A., Ben-Neriah, Z., Argov, Z., et al. (2015). Nonsyndromic early-onset cone-rod dystrophy and limb-girdle muscular dystrophy in a consanguineous Israeli family are caused by two independent yet linked mutations in *ALMS1* and *DYSF*. *Hum. Mutat.* 36, 836–841.
37. Rivolta, C., Sweklo, E.A., Berson, E.L., and Dryja, T.P. (2000). Missense mutation in the *USH2A* gene: association with recessive retinitis pigmentosa without hearing loss. *Am. J. Hum. Genet.* 66, 1975–1978.
38. Lenassi, E., Vincent, A., Li, Z., Saihan, Z., Coffey, A.J., Steele-Stallard, H.B., Moore, A.T., Steel, K.P., Luxon, L.M., Héon, E., et al. (2015). A detailed clinical and molecular survey of subjects with nonsyndromic *USH2A* retinopathy reveals an allelic hierarchy of disease-causing variants. *Eur. J. Hum. Genet.* 23, 1318–1327.