eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Facultad de Ciencia y Tecnología
Departamento de Genética, Antropología Física y Fisiología Animal
Programa de Doctorado de Biología Molecular y Biomedicina

Molecular genetic diagnosis of patients from the Basque Country with Inherited Retinal Dystrophies

Author:
Maitane Ezquerra Inchausti
Licenciada en Biología

Supervisors:
Javier Ruiz Ederra
David Otaegui Bichot

Donostia, Enero 2019

Familiakoei, Iñakiri

Gure ondotik joandako familiakoei; bi aitxitxe, amama eta bereziki izeko Garbiñeri.

ACKNOWLEDGEMENTS/ ESKERRAK

During this work, I was supported by different institution´s grants that I would like to acknowledge:

Personal grants:

- Grant from Basque Government's Department of Education: Ayudas a la Formación de Personal Investigador No Doctor (DEDUC14/309). Caracterización genetico molecular de pacientes de Gipuzkoa con distrofias de retina.

- EMBO short stay fellow: Identification of novel causal genes in Inherited Retinal Dystrophies among patients from the Basque country with challenging diagnosis. (April 2017)

Research projects

- National Institute of Health Carlos III (PI18/00507): Síndrome Pseudoexfoliativo: estudio clínico, epidemiológico y molecular en muestras procedentes de cirugía de catarata y glaucoma**.** Researchers: Javier Ruiz Ederra (IP); Javier Mendicute (CoIP); **Maitane Ezquerra Inchausti**; Olatz Barandika Fernandez; Txomin Alberdi Ibarloza; Leire Juaristi Eizmendi.

- National Institute of Health Carlos III (PI17/01413): Caracterización genético-molecular de los pacientes con distrofias hereditarias de la retina de la C.A. del País Vasco**.** Researchers: Cristina Irigoyen Laborra (IP); **Maitane Ezquerra Inchausti**; Olatz Barandika Fernandez; Marta Galdós Iztueta; Gonzaga Garay Aramburu.

- Mutua Madrileña Foundation: Caracterización genético-molecular de los pacientes de retinosis pigmentaria de la C.A de Euskadi, mediante un nuevo método basado en el secuenciado masivo de última generación. Cristina Irigoyen Laborra (IP); Javier Ruiz Ederra (IP); **Maitane Ezquerra Inchausti**; Olatz Barandika Fernandez; Ander Anasagasti Viteri.

- National Institute of Health Carlos III (PI13/02621/): Estudio de los micrornas en pacientes y en modelos murinos de retinosis pigmentaria. Nuevas dianas terapéuticas. Javier Ruiz Ederra (IP); Ander Anasagasti Viteri; Olatz Barandika Fernandez; Cristina Irigoyen Laborra; **Maitane Ezquerra Inchausti**.

- BEGISARE-Gipuzkoa Association. Association of patients with retinitis pigmentosa.

Agradecimientos……

ARTICLES PUBLISHED DURING THE THESIS

Related to this thesis

Ezquerra Inchausti M., Anasagasti A., Barandika O., Garai-Aramburu G., Galdós M., López de Munain A., Irigoyen C., Ruiz-Ederra J. A new approach based on targeted pooled DNA sequencing identifies novel mutations in patients with Inherited Retinal Dystrophies. Scientific Reports 18;8(1):15457. doi: 10.1038/s41598-018-33810-3 (2018).

Ezquerra-Inchausti M., Barandika O, Anasagasti A, Irigoyen C, López de Munain A, Ruiz-Ederra J. High prevalence of mutations affecting the splicing process in a Spanish cohort with autosomal dominant retinitis pigmentosa. Scientific Reports 6, 39652; doi: 10.1038/srep39652 (2017).

Barandika O, Irigoyen C, -Anasagasti A, Egiguren G, **Ezquerra-Inchausti M**, López de Munain A, Ruiz-Ederra J. A Cost-Effective Mutation Screening Strategy for Inherited Retinal Dystrophies. Ophthalmic Res 56 (3), 123-31 (2016).

Not Related to this thesis

Moreno-Cugnon L., Anasagasti A., **Ezquerra-Inchausti M**., Izeta A., de la Villa P., Ruiz-Ederra J., Matheu A. SOX2 haploinsufficiency promotes impaired vision in aged mice. (Accepted Oncotarget).

Anasagasti A, **Ezquerra-Inchausti M**, Barandika O, Muñoz-Culla M, Caffarel MM, Otaegui D, López de Munain A, Ruiz-Ederra J. Expression Profiling Analysis Reveals Key MicroRNA-mRNA interactions in early retinal degeneration in Retinitis pigmentosa. Invest Ophthalmol Vis Sci. 1;59(6):2381-2392 (2018).

Barandika O [1], **Ezquerra-Inchausti M [1]**, Anasagasti A, Vallejo-Illarramendi A, Llarena I, Bascaran L, Alberdi T, De Benedetti G, Mendicute J, Ruiz-Ederra J. Aquaporin-1 protein levels are increased in the lens epithelium of cataract patients. Biochim Biophys Acta Molecular Bases of Disease 1862, 2015-2021 (2016). **[1] Both authors contributed equally.**

TABLE OF CONTENTS

ABBREVIATIONS

CGH array	Comparative genomic hybridization array
AD	Autosomal dominant
AR	Autosomal recessive
BBS	Bardet-Biedl syndrome
bp	Base pair
cGDP	cyclic guanosine diphosphate
cGMP	cyclic guanosine monophosphate
CNV	Copy number variation
COD	Cone Dystrophies
CORD	Cone-rod dystrophies
CSNB	Congenital stationary night blindness
CSVS	Ciberer Spanish variant server
CVD	Colour visual defects
DM	Myotonic dystrophy
DMD	Duchene muscular dystrophy
DNA	Deoxyribonucleic acid
ERG	Electroretinogram
GATK	Genome analysis toolkit
grch(37),(38)	Genome reference consortium
hg(19),(38)	Human genome assembly
HRM	High resolution melting
INL	Inner nuclear layer
IRD	Inherited retinal dystrophies
LCA	Leber congenital amaurosis
LOF	Loss of function
LOH	Lost of heterozygosity
MAF	Minor allele frequency
MD	Muscular Dystrophy
MLPA	Multiplex ligation-dependant probe amplification
NGS	Next generation sequencing
OD	Right eye
OI	Left eye
OMIM	Online mendelian inheritance in man
ONL	Outer nuclear layer
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RP	Retinitis pigmentosa
RPE	Retinal pigment epithelium
rs	Reference SNP
SNP	Single nucleotide polymorphism
USH	Usher syndrome

USH1	Usher syndrome type 1
USH2	Usher syndrome type 2
USH3	Usher syndrome type 3
UPD	Uniparental disomy
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing
WT	Wild type

RESUMEN

Las distrofias hereditarias de la retina, (DHRs), son un grupo heterogéneo de enfermedades responsables de distintos fenotipos clínicos, que afectan principalmente a la retina. Existen más de 250 genes ligados a más de 20 fenotipos distintos. Encontramos fenotipos sindrómicos, donde hay otros órganos o tejidos afectados aparte de la retina y fenotipos no sindrómicos donde la afección se confina únicamente a la retina. La patología más común de este último grupo es la Retinosis Pigmentaria (RP), con una prevalencia de entre 1 en 2000 o 1 en 3000 personas, dependiendo de la población. En el caso de la Retinosis Pigmentaria, se da una disfunción progresiva asociada a una pérdida de células fotorreceptoras (bastones y conos) donde al principio se comienza por una pérdida de bastones y finalmente se llega a la pérdida de los conos produciéndose así la atrofia de la retina. La alteración visual se suele manifestar como ceguera nocturna y pérdida de campo visual. El rango de trastornos visuales asociados a esta enfermedad va desde la pérdida del campo sectorial, prácticamente no percibida por el paciente, hasta una gran pérdida del campo visual periférico.

En el caso de la distrofia de conos y bastones (CORD/COD), la pérdida de los fotorreceptores comienza por los conos y puede progresar o no a la pérdida de bastones.

En cuanto a fenotipos sindrómicos, el Síndrome de Usher es el más común, y cursa con pérdida de audición aparte de la retinosis pigmentaria.

Atendiendo al patrón de herencia, en las distrofias de retina podemos encontrar casos de herencia autosómica dominante, autosómica recesiva, ligada al cromosoma X o incluso herencia mitocondrial.

La publicación del primer boceto del genoma humano en 2001 (Lander y cols., 2001; Venter y cols.,2001), ha promovido el aumento de técnicas de secuenciado que permiten obtener datos genómicos de una manera más rápida y precisa. Estas técnicas diagnósticas basadas en el secuenciado masivo, incluyen el secuenciado de paneles de genes candidatos, secuenciado del exoma completo (WES) o el secuenciado del genoma completo (WGS), cobran gran importancia en patologías de origen hereditario, como pueden ser las distrofias de la retina. Gracias a estas técnicas, se puede lograr: a) confirmar o modificar el diagnóstico clínico, b) ofrecer un pronóstico más ajustado, c) dar consejo genético a los pacientes, d) permitir a algunos pacientes poder beneficiarse de posibles terapias.

Partimos de la hipótesis de que, sabiendo que existe una gran heterogeneidad genética dentro de las distrofias de retina, es posible que, en un área geográfica con existencia documentada de haber tenido una alta tasa de consanguinidad, como es Euskadi, pudiera haber una representación más limitada del repertorio de mutaciones. Además de mutaciones conocidas, creemos que podemos encontrar mutaciones noveles o incluso mutaciones en genes que aún no se han asociado a distrofias hereditarias de retina.

A lo largo de la presente tesis, nos planteamos realizar la caracterización genético-molecular de la población de pacientes con DHR de Euskadi mediante distintas técnicas y poder implementar así un flujo de trabajo para poder llevarlo a cabo.

En primer lugar, se diseñó un panel que contenía 31 genes, relacionados con DHR de herencia autosómica dominante. 29 pacientes con posible herencia dominante y 3 controles positivos fueron analizados con este panel de genes. El análisis se realizó utilizando la metodología ion Torrent en el propio instituto Biodonostia. Tras el análisis de los resultados, 14 de los 29 pacientes analizados fueron diagnosticados molecularmente. Además, las variantes de los tres controles positivos introducidos, también fueron detectados, siendo la sensibilidad de la técnica del 100%. De los pacientes caracterizados, 9 contenían la mutación en genes relacionados con el spliceosoma. El spliceosoma, es un complejo para el corte de los intrones y empalme de los exones en los precursores del RNA mensajero (mRNA) y realizar así el *splicing*. Hay varios genes como *PRPF31*, *PRPF8* o *SNRNP200* entre otros, que son parte de este complejo. Mutaciones en estos genes se han asociado con retinosis pigmentaria autosómica dominante, y aunque su expresión se extienda por distintos tejidos, únicamente se observa afectación a nivel de retina (Ezquerra-Inchausti *y cols*., 2017).

En segundo lugar, se utilizó un panel prediseñado que incluye 316 genes relacionados con DHR y otras alteraciones oftalmológicas. Para poder llevar a cabo la secuenciación de este panel se utilizó la tecnología Ion Proton, con mayor capacidad que el Ion Torrent. Además, se empleó una metodología consistente en el secuenciado de las muestras mezcladas en *pooles* de DNA. La estrategia que se utilizó en este apartado es la siguiente: En primer lugar, se analizaron 3 *pooles*, con distinto número de pacientes, 4 8 y 16. En estos *pools* se introdujeron pacientes resueltos y pacientes no resueltos, pero en los que se habían

observado variantes genéticas con una frecuencia (MAF) inferior a 0,003. Tras observar que la sensibilidad era del 100% en los tres *pools*, se analizaron 9 pools de 16 pacientes cada uno, considerando por tanto las 16 muestras como 1 a efectos de secuenciado. Los *pooles* se realizaron de manera equimolar para minimizar el riesgo de sub o sobrerrepresentación de alguna de las muestras. De esta manera, se analizaron 160 pacientes diferentes, de los cuales 17 correspondían a controles positivos. Tras el análisis de las variantes, se filtraron las mutaciones de interés. Posteriormente, se utilizó como técnica de genotipado el análisis de alta resolución de fusión de la doble hebra de DNA (*high resolution melting analysis o* HRM), para identificar al paciente portador de la mutación de entre los 16 secuenciados (Anasagasti *y cols*., 2013). Finalmente, mediante el secuenciado de Sanger, confirmamos que la mutación seleccionada se encuentra en el paciente indicado por el análisis de los datos de HRM. De esta manera pudimos diagnosticar molecularmente a 60 pacientes de los 143 analizados (Ezquerra-Inchausti *y cols*., 2018). Además, para poder aumentar la tasa de hallazgos, empleamos la técnica de MLPA para la detección de cambios en el número de copias (CNV) en pacientes sin diagnóstico molecular tras en secuenciado masivo. Se analizaron por MLPA los genes *USH2A* y *EYS* en pacientes con mutación en heterocigosis en dichos genes o en pacientes con clínica de Usher tipo 2 en el primer caso. Además, se analizaron genes relacionado con retinosis pigmentaria dominante como *RP1, PRPF31 RHO* e *IMPDH1* en pacientes con herencia dominante, que no habían sido resueltos tras el secuenciado completo de los genes dominantes ni por el panel de 316 genes. Por último, se analizaron los genes *RP2, RPGR* y *CHM* ligados al cromosoma X en pacientes con herencia ligada a X o en varones que eran casos únicos en la familia. De esta manera detectamos una deleción no descrita hasta el momento en el gen *PRPF31* en una familia con posible herencia dominante.

Después se secuenciaron 10 pacientes mediante el análisis del exoma completo (WES). Mediante esta técnica tratamos de encontrar mutaciones en genes no asociados hasta el momento a DHR o que no estuvieran presentes en el panel de 316 genes empleado. De los 10 pacientes analizados 9 tenían herencia recesiva o eran casos únicos en la familia. De ellos, 8 tenían apellidos vascos con familia procedente de pequeños pueblos principalmente de Gipuzkoa. Al seleccionar a estos pacientes, tratamos de encontrar una mutación endémica de esta región. El décimo paciente seleccionado, presentaba un árbol genealógico sugerente

de herencia ligada al cromosoma X. El método utilizado fue el siguiente: los 9 pacientes con posible herencia recesiva se analizaron primero mediante *arrays* CGH. Mediante esta técnica se trata de encontrar, por un lado, regiones de pérdida de heterocigosidad (LOH), que nos indican que esas regiones del genoma son iguales en ambos alelos. Teniendo en cuenta que los pacientes seleccionados tienen una posible herencia recesiva, y provienen de pequeños pueblos donde es más posible la consanguinidad, esperamos encontrar mutaciones en homocigosis. De ese modo, anotamos las regiones de pérdida de heterocigosidad, detectadas en cada paciente y se analizaron en primer lugar las mutaciones presentes en los genes comprendidos en esas regiones en cada paciente. Por otro lado, los *arrays* CGH también nos permiten detectar cambios en el número de copias (CNV). Después, se secuenció el exoma completo en los 10 pacientes. Como se ha indicado anteriormente, en los 9 pacientes con herencia recesiva se analizaron en primer lugar únicamente los genes incluidos en las regiones de LOH. Al no tener ningún resultado concluyente con las alteraciones de estas regiones, se analizaron los resultados de todo el exoma. En el caso del paciente con herencia ligada al cromosoma X, únicamente se analizaron las mutaciones encontradas en dicho cromosoma. Una vez analizados los resultados de todos los exomas, se encontró una mutación en homocigosis en el gen *SAMD11* que, dada su reciente asociación a Retinosis pigmentaria recesiva (Corton *y cols*., 2016), no estaba incluido en nuestro panel de 316 genes.

Este hallazgo, es de gran importancia ya que refuerza la posibilidad de asociación de ese gen a Retinosis Pigmentaria al encontrar otro caso más con la misma mutación añadido a los descritos por (Corton *y cols*., 2016) Es interesante que esa alteración se encontrara también en población española.

Con el objeto de buscar mutaciones en este gen en toda nuestra muestra de pacientes sin resolver, analizamos el gen *SAMD11* a 83 pacientes mediante HRM. Sin embargo, tras analizar mutaciones en este gen en nuestros pacientes sin resultado, no encontramos ninguna mutación de interés en los mismos tras dicho análisis.

Por último y siguiendo nuestra metodología de trabajo, en este caso se analizaron 21 pacientes y 2 controles positivos mediante *array* CGH. Para ello se utilizaron dos tipos de *arrays* distintos. Por un lado, se analizaron 20 pacientes y dos controles positivos mediante el *array CytoScan XON* de Affymetrix. Este *array* permite analizar CNV en todo el exoma. Tras

analizar los resultados de este *array*, se observaron 4 posibles alteraciones en genes relacionados con DHR en 4 pacientes diferentes. Además, se pudieron detectar las alteraciones de los dos controles positivos introducidos, indicándonos una sensibilidad del 100%. Los 4 pacientes en los que se detectaron posibles alteraciones fueron analizados junto con los mismos dos controles positivos y otros dos pacientes (uno no analizado en el *array* anterior) mediante un *array* diseñado por nosotros de Agilent. En este *array* se incluyeron 123 genes relacionados con distintas DHR. Tras el análisis de este *array* se pudo detectar la alteración de uno de los controles positivos, pero no el otro. Se observó que la región donde se encontraba la alteración no detectada contenía muy pocas sondas, y eso hizo probablemente que no se detectara la alteración. Por otro lado, de los 4 pacientes introducidos con posibles alteraciones detectadas con el *array CytoScan XON*, solo se detectó una, la deleción que abarca los exones 2 y 3 del gen *PRPH2*. Para poder validar estas alteraciones, se realizó una qPCR con sondas comerciales TaqMan. Se diseñaron sondas en las regiones supuestamente alteradas y en las regiones adyacentes, como control negativo. El análisis se realizó en los casos índice como en los familiares disponibles. Tras el análisis de los resultados, solo se validó la deleción en *PRPH2*, encontrada en ambos tipos de *array*. Este resultado nos indica que el *array* de Agilent parece tener mayor especificidad que el *array CytoScan XON*. Sin embargo, el *array CytoScan XON* parece más sensible ya que detecta las mutaciones de los dos controles positivos introducidos.

Con todas las técnicas utilizadas en esta tesis, hemos podido diagnosticar molecularmente a 61 pacientes y 3 casos se han clasificado como VUS (Variantes de significado incierto) de los 157 pacientes analizados. La estrategia seguida en este trabajo consistió en primer lugar en el secuenciado de 31 genes asociados a adRP en pacientes analizados de manera individualizada. Uno de los avances que ha supuesto la presente tesis, ha sido el desarrollo de una estrategia de secuenciado masivo, que nos ha permitido analizar 316 genes de RP y otras DHR. Dicha estrategia se basa en la combinación del secuenciado en *pool* de grupos de 16 muestras con una técnica de genotipado de alta resolución (HRM) (Ezquerra-Inchausti *y cols*., 2018: https://rdcu.be/9wtv). Mediante esta estrategia hemos logrado reducir unas 6 veces los costes derivados del proceso de búsqueda de mutaciones, en relación con métodos de secuenciado masivo individualizado. Dado, además, que dicho abordaje nos ha permitido obtener una sensibilidad del 100%, decidimos emplear esta estrategia como primer paso en

nuestro flujo de trabajo. Esto nos permite analizar todos los pacientes mediante la misma técnica, sin sesgo por su posible tipo de herencia, facilitando la identificación de la mutación causal en todos los genes de DHR, no ciñéndonos únicamente a los genes de retinosis pigmentaria. Además, mediante el flujo de trabajo seguido, donde se han utilizado diversas técnicas y estrategias genéticas, como el análisis del exoma completo, MLPA o los *arrays* CGH, hemos podido diagnosticar molecularmente a pacientes que, mediante los métodos anteriormente descritos, no hubiera sido posible llevarlo a cabo. De todos modos, esta estrategia de flujo de trabajo no se ha podido aplicar en todos los pacientes dado su elevado coste. Hay que tener en cuenta que el coste de estas técnicas (exomas y *arrays* CGH) es elevado. Es esperable, por tanto, que a medida que se aumente el número de pacientes a analizar por cada técnica, irá aumentando el número de pacientes resueltos.

Por último, gracias a esta tesis 61 pacientes han podido ser diagnosticados molecularmente y esto permitirá que algunos de ellos puedan beneficiarse de posibles terapias o ensayos clínicos que se están llevando a cabo. Además, ha aportado luz sobre las variantes más comunes de nuestra población y ha descrito nuevas mutaciones no detectadas hasta el momento ayudando a futuros análisis de este tipo.

INTRODUCTION

1. THE RETINA

The retina is a light sensitive neural tissue located on the rear surface of the eye, between the choroid and the vitreous humour (Figure1). Light passes through the cornea and the pupil before it reaches the lens, by which it is focused onto the retina that converts it into nerve impulses. Subsequently, this information is transmitted to the rest of the visual system and to the brain. The transmission of the information is conducted through the optic nerve composed of axons of the ganglion cells. The interpretation of the information is performed in the visual primary brain cortex, located in the encephalon[1].

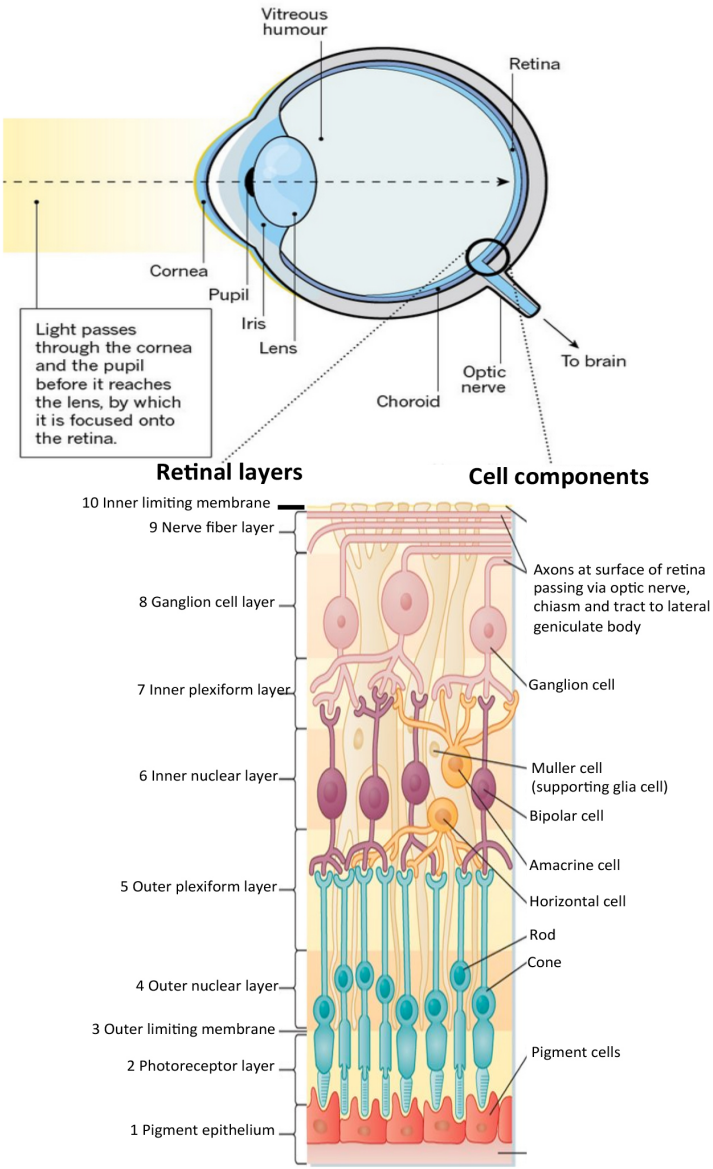


Figure1: Schema of the eye and the retinal structures. Adapted from https://www.nature.com/articles/d41586-018-06111-y and from Koeppen and Stanton (Berne and Levy Physiology, 6th Edition).

At the end of XIX century Dr. Santiago Ramon y Cajal, identified for the first time the main cellular types constituting the retina and the information flux trough it[2]. This work helped to understand the structure and function of this tissue. The vertebrate's retina contains ten different layers. Each layer and it´s components are explained below, in order from the closest layer to the choroid until the closest layer to the vitreous humour.

1. Retinal Pigment Epithelium (RPE): It is a monolayer of pigmented cells that forms the outer blood-retinal barrier. The cells conforming the RPE are connected with the outer segments of the photoreceptors with microvilli transporting nutrients, ions, and water between them. Other main functions of the RPE are: 1. Absorption of light and protection of the photoreceptors against photooxidation. 2. Reisomerization of all-*trans*-retinal into 11-*cis*-retinal, which is crucial for the visual cycle. 3. Phagocytosis of shed photoreceptor membranes, and 4. Secretion of essential factors for the structural integrity of the retina[3].
2. Photoreceptor layer (PR): It contains the outer and inner segments of cones and rods (Figure 2).

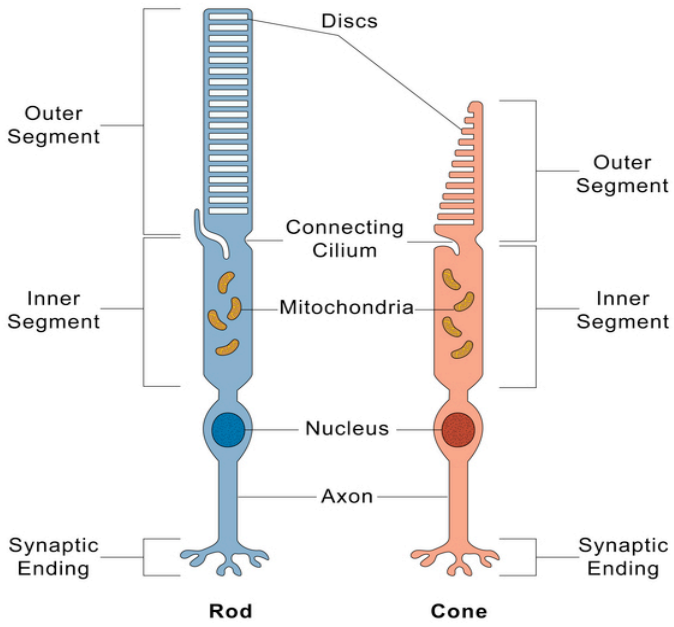


Figure 2. Rods and cones morphology. Adapted from https://ghr.nlm.nih.gov/condition/cone-rod-dystrophy

Photoreceptors are characterised for converting light stimulus in electric signals. However, each cell type has other different functions. Rods are more sensitive to light and are responsible for vision at low light levels (scotopic vision). Cones are

active at higher light levels (photopic vision) and are responsible for high spatial acuity and colour vision. There are three types of cones; red, green and blue. Each type is sensitive at different light wavelengths. There is an asymmetric distribution of cones and rods around the retina; the central area, containing the macula lutea and a yellowish area where the fovea is localized, is enriched in cones, for high acuity vision. The peripheral area, which takes part in periphery vision is enriched in rods.

3. Outer limiting membrane (OLM): This membrane separates the inner segment of the photoreceptors, from their nucleus.
4. Outer nuclear layer (ONL): The body and nucleus of photoreceptors are localised in this layer.
5. Outer plexiform layer (OPL): It contains the axons and the synaptic endings of the photoreceptors. These endings synapse with the bipolar cell dendrites.
6. Inner nuclear layer (INL): It contains the nucleus of the Bipolar cells, Horizontal cells, Amacrine cells and Müller cells.

   -Bipolar cells: They are a type of neuron conforming the retina. They act transmitting signals from the photoreceptors to the ganglion cells.

   -Horizontal cells: They are laterally interconnecting neurons. They modulate the information flow from photoreceptors to bipolar cells at the first synapse of the visual system[4].

   -Amacrine cells: They are interneurons that interact at the second synaptic level of the photoreceptor-bipolar-ganglion cell chain pathway. They serve to modulate and intervene a temporal domain to the visual message presented to the ganglion cells[5].

   -Müller cells: Are the major type of glial cells. These types of cells do not participate in the signal transduction. However they are responsible for the homeostatic and metabolic support of retinal neurons, mediating in the transport of transcellular ion, water and bicarbonates[6]. The Muller cell processes involve all retinal layers, from the ONL to the Ganglion cell layer (GCL).

7. Inner plexiform layer (IPL): It contains the axons and dendrites of bipolar cells, ganglion cells and amacrine cells for the synaptic process.
8. Ganglion cell layer (GCL): It contains the nucleus of Ganglion cells. Ganglion cells are the projection neurons that convey information from other retinal neurons to the rest of the

brain. Ganglion cells receive inputs from bipolar cells, which convey signals from photoreceptors and from amacrine cells[7].

9. Nerve fibre layer (NFL): Axons of ganglion cells are localised in this layer.
10. Inner limiting membrane (ILM): It is formed by astrocytes and processes of Müller cells. This is the limit between the retina and the vitreous.

1.1. The phototransduction and visual cycle

Phototransduction was discovered by George Wald (1906-1997) and is also called "Warld´s visual cycle". Visual phototransduction is the process by which a photon of light is absorbed by visual pigment molecules in the photoreceptor cells' outer segment and is converted into an electrical signal and therefore in the vision in our brain. This process involves the sequential activation of a series of signalling proteins, leading to the eventual opening or closing of ion channels in the photoreceptor cell membrane.

First of all, it is interesting to know that photoreceptors are in the depolarised state when they are not stimulated by light. In this state, voltage-gated Ca^{2+} channels are opened facilitating the process by which neurotransmitter (in this case glutamate), is released into the synaptic cleft. Thus, in the dark stage, the photoreceptor terminal is continually releasing glutamate[8]. The outer segment of the photoreceptor is permeable to Na^{+}. Intracellular levels of cyclic guanosine monophosphate (cGMP), regulates Na+ and Ca^{2+} ion channels. In the dark stage, high levels of cGMP maintain the channels open. In this state, the photoreceptor is depolarised with a membrane potential of approximately -40mV.

When light activates the visual pigment, a biochemical cascade occurs resulting in a decrease in the concentration of cGMP closing the Na^{+} and some $Ca2^{+}$ channels. The continuous Na^{+} loss increases the negativity inside the cell, which becomes hyperpolarized; the membrane potential approaches −75mV. The change in potential is graded and the level of hyperpolarization depends on the amount of light absorbed[8].

The visual photopigment molecules are composed by G protein-coupled receptors (GPCRs) called opsin (rhodopsin, in the case of rod photoreceptors) and a retinaldehyde chromophore. This chromophore is in *cis* during darkness and is called 11- *cis* retinal. When the light is absorbed, the conformation changes and it is converted into all-*trans* retinal. The conformational change results in the dissociation of the chromophore from the opsin.

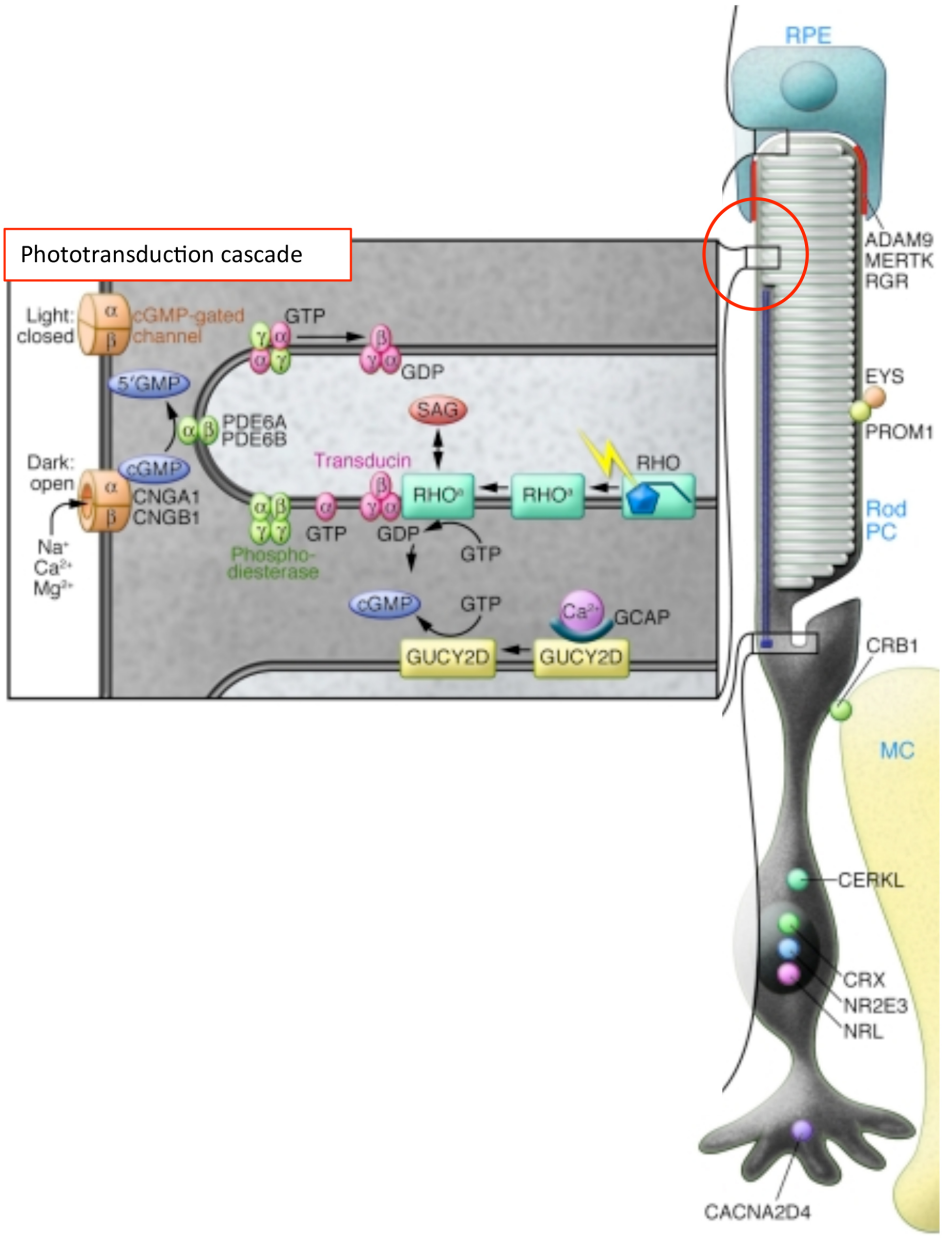


Figure 3: Phototransduction cascade. Marked with the red circle, the external segment with the discs, where the phototransduction occurs. Adapted from denHollander *et al*., 2010.

Briefly, the phototransduction occurs in four steps (Figure 3). First of all, the receptor protein, the opsin (rhodopsin in rods), is activated by the light (photon). This activation isomerizes the 11-*cis* retinal to all-*trans* retinal as explained before. Then, this conformational change stimulates the G-protein transducin, and GTP is converted to GDP. Afterwards, the activated transducin, activates the effector protein phosphodiesterase 6 (PDE6). This enzyme, catalyses the hydrolysis of cGMP to 5´ GMP. Finally, falling concentrations of cGMP cause the closure of transduction channels, decreasing intracellular Na^+ and the Ca^{2+} concentration.

After phototransduction, generated all-*trans* retinal is regenerated to 11-*cis* retinal through a series of steps. This cycle is known as the visual cycle (Figure 4). In rods, first step occurs in the outer segment, all-*trans* retinal is reduced to all-*trans* retinol (Vitamin A), by all-*trans* retinol dehydrogenase (atRDH). Then all-*trans* retinol exits the photoreceptor and enters into the RPE[9]. In the RPE, lecithin retinol acyl transferase (*LRAT*), links all-*trans* retinol to phosphatidyl choline to generate all-*trans* retinyl. The next step involves the simultaneous hydrolysis and isomerization of all-*trans* retinyl esters to yield 11-*cis* retinol. This isomerization and hydrolysis is facilitated by the RPE65 enzyme[10]. 11-*cis* retinol is then bound to CRALPB (Cellular retinaldehyde binding protein), which delivers it to 11-*cis* retinol dehydrogenase (11-*cis* RDH), that oxidizes 11-*cis* retinol to 11-*cis* retinal. Finally, 11-*cis* retinal is newly generated and enters into the rod outer segment[11].

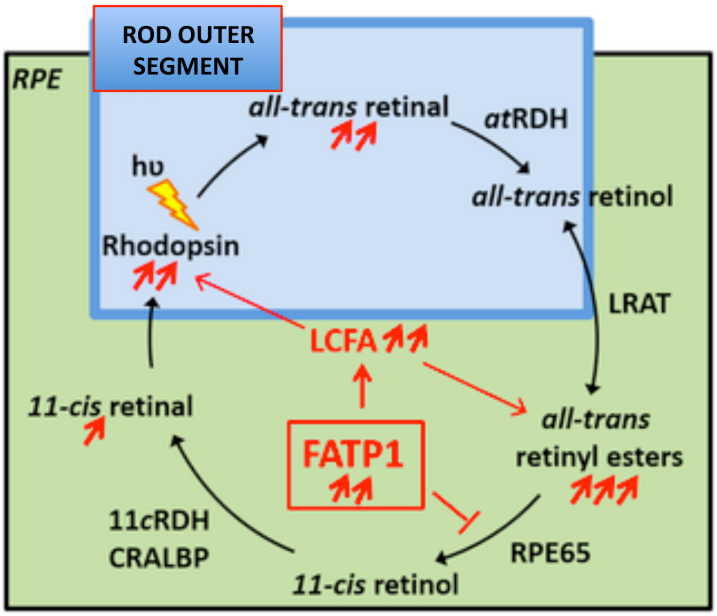


Figure 4: Visual cycle schema. Abbreviations; RPE, Retinal pigment epithleium. (Adapted from Cubizolle *et al.*, 2017).

2. INHERITED RETINAL DYSTROHIES (IRDs)

The retina provides visual information based on the correct function of all the structures implicated in this process. When any of these structures is affected we call it visual disability. The visual system, and especially the retina is one of the most energetically demanding systems of our organism. The correct retinal function requires a balance between cell proliferation, differentiation and apoptosis. This high renewal that occurs especially in photoreceptor cells, the active transport of ions against their concentration and the repolarization after depolarization consume high quantity of energy[12]. Some genetic mutations or epigenetic alterations modify the stability of the retina, inducing the apoptosis of photoreceptors, contributing to the development of different IRDs.

2.1. Epidemiology

Following World Health Organization´s data[13], there are around 39 million blind people, cataracts being the first cause in 51% of the cases. In the case of inherited retinal dystrophies which are a group of diseases characterised by a progressive photoreceptors affectation, a prevalence of 1 in 3,000 or 1 in 4,000[14,15] is estimated or more than 2 million[16] people affected worldwide. The prevalence can vary depending on the geographical area and type of population (ethnically heterogeneous or homogeneous). Indeed, the prevalence of some types of IRDs varies between 1 in 4,000 or 1 in 90,000 in some studies[17], and it can reach to be 1 in 230 in populations with high rates of consanguinity[18–20]. The prevalence data available on the Spanish population is only about retinitis pigmentosa (RP), the most common type of IRDs. In 2007 it was estimated that more than 15,000 people were affected and approximately 500,000 people were carriers of pathogenic mutations, able to transmit the disease[21].

2.2. Classification

IRDs can be categorised in three groups; 1) rod and rod-cone photoreceptor cell diseases, 2) cone-rod and cone diseases and 3) generalised photoreceptor diseases. At the same time,

the classification considers, the cases in which only the retina is affected (non-syndromic form) or those associated with pathologies also in other tissues (syndromic forms)[16].

2.2.1. Non-syndromic retinal diseases

2.2.1.1. Rod and rod-cone photoreceptor cell diseases.

Rods represent the majority of photoreceptor types in the retina. They are specialised in vision in low light conditions and are mainly located in the retinal periphery. retinitis pigmentosa (RP) is the principal and most common disease of this group with a prevalence of 1 in 4,000 individuals affected[22]. At cellular level, rods are the first photoreceptor type affected by apoptosis, producing night blindness at early stages, followed by tunnel vision (Figure 5), while rod photoreceptors apoptosis progresses towards the centre of the retina in later stages of the disease, causing complete blindness[22]. The age of onset varies depending on the mutated gene, but typically starts during the early teenage years and severe visual impairment occurs by 40-50 years old. However, there are also early onset, late onset and even non penetrant forms of RP[23]. To assess the disease status and progression, electroretinographic measurements are developed.

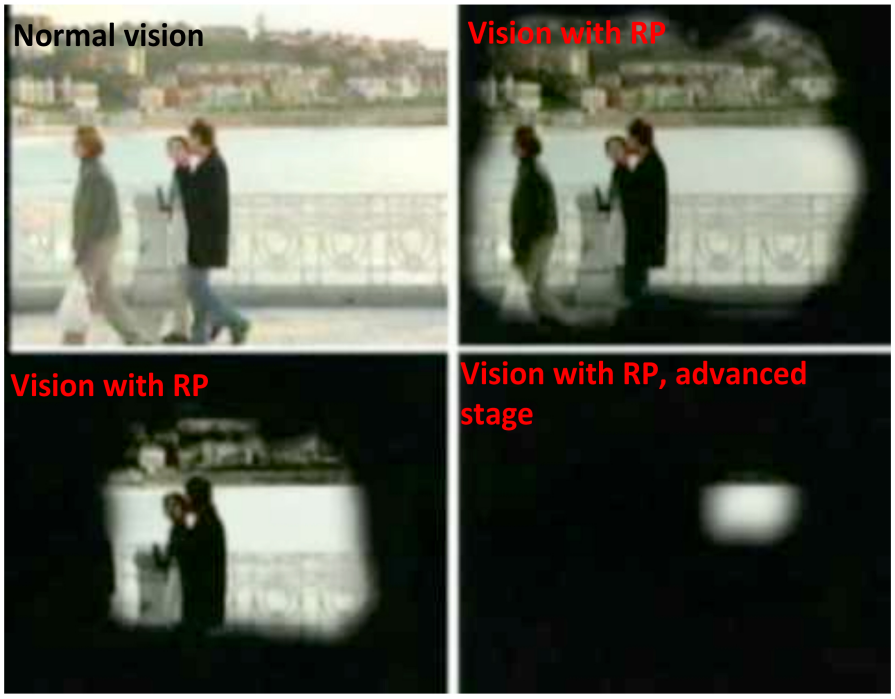


Figure 5: Comparison between normal vision and vision with RP. There are three images of tunnel vision of an RP patient in different stages of the disease. (http://www.begisare.org/).

2.2.1.2. Cone-rod and cone photoreceptor cells diseases.

As previously mentioned, cones are responsible for daylight vision, colour discrimination and high acuity vision. Only 5% of human photoreceptors are cones[24] and most of them are located in the fovea. Human diseases that affect the cone system lead to severe visual impairment. Clinically, the major features are photophobia, reduced visual acuity, nystagmus and colour vision abnormalities[25].

The cone dystrophies can be divided into two groups: stationary and progressive.

In the case of stationary cone dysfunctions, the major causes are complete and incomplete achromatopsia. Individuals affected by complete achromatopsia are unable to distinguish colours. However, patients with incomplete achromatopsia retain residual colour vision and visual acuity is more preserved[16].

Progressive cone diseases are principally cone and cone-rod dystrophies (CODs and CORDs). In both cases, the age of onset is usually during childhood and are generally more severe than RP and produce blindness earlier than in RP. In the case of CORDs, in contrast to CODs, a peripheral retinal involvement is observed and the electroretinogram (ERG) is characterised by a decrease in both cone and rod responses. Finally, night blindness occurs in later stages of the CORDs when rods also become affected.

2.2.1.3. Macular dystrophies

The most prevalent inherited macular dystrophy is the Stargardt disease. It is a monogenic disease with a prevalence of 1 in 8,000 – 10,000 individuals[26,27]. Cones are more affected in this disease and that is the reason of having affected principally the macula, the central area of the retina where the proportion of cones is high. Therefore, patients present central visual loss, and might notice grey, black or hazy spots in the centre of their vision field (NIH, National Eye Institute) (Figure 6). Moreover, there is a loss of retinal function and structure over the time[28]. Considering the age of onset, it varies but is usually in the early teens[27].

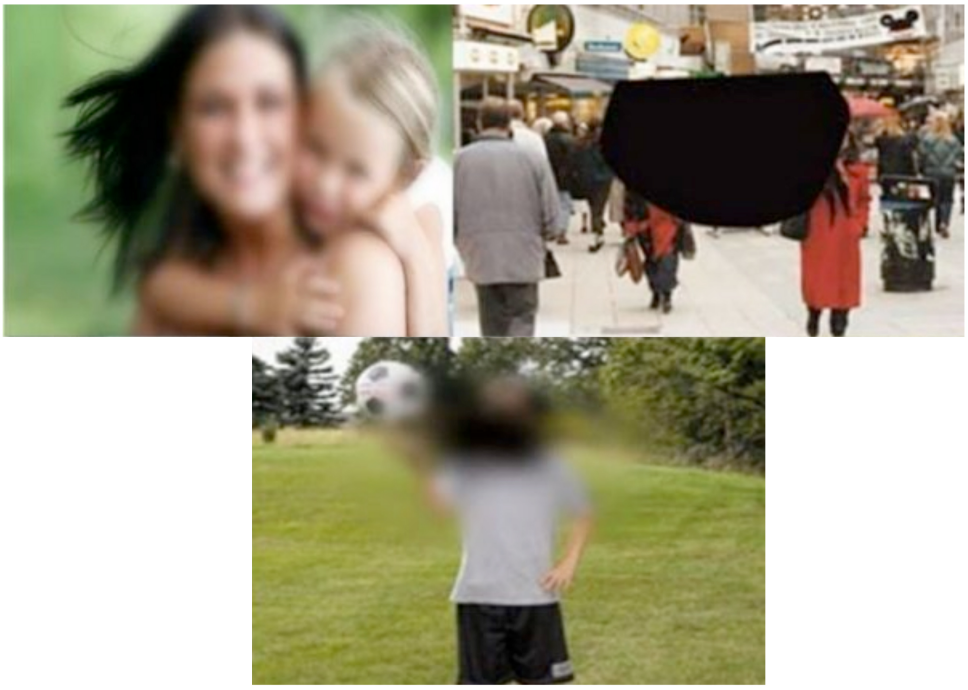


Figure 6: Representations of vision with Stargardt disease.

2.2.1.4. Generalised photoreceptor diseases

In this group of IRDs, diseases such as Leber congenital amaurosis (LCA) and Choroideremia (CHM) are included.

In the case of LCA, it is considered as the most severe non syndromic retinal dystrophy and causes blindness or severe visual impairment before the age of 1 year[29]. The major clinical features are severe and early visual loss, sensory nystagmus, amaurotic pupils and absence of electrical signals on ERG. Nevertheless, there is high phenotypic variability between patients. In fact, the appearance of the retina vary depending on the mutated gene and the phenotypic range of retinal aspects observed, still needs to be correlated with different genotypes[29]. The prevalence of LCA is estimated between 1 in 30,000 and 1 in 81,000[30,31].

In the case of CHM, which is an X-linked IRD, the disease is characterised by progressive degeneration of RPE, photoreceptors and finally choroid[32]. Symptoms begin with night blindness during teenage years, progresses with gradual loss of peripheral vision during 20s and/or 30s and can finally result in blindness by middle age[16]. Choroideremia has a prevalence of about 1 in 50,000 individuals[33]. As CHM is a X-linked disease, is more frequently observed in males. However, in the case of female carriers, although they usually maintain a good vision throughout their life, more severe phenotypes have also been

reported due to the effects of skewed X chromosome inactivation during early retinal development[34].

2.2.2. Syndromic retinal diseases

2.2.2.1. Usher syndrome (USH)

Patients with Usher syndrome are characterised by suffering from a combination of RP and sensorineural deafness or hearing impairment. This disease can be classified in three subtypes, depending on the severity of the phenotype and the mutated gene.

Usher syndrome type 1 (USH1) is the most severe of all three subtypes. Indeed, patients have profound and congenital deafness and vestibular dysfunction, leading to a delayed development and also suffer from adolescent onset RP[16]. The prevalence of USH1 is estimated between 1 in 100,000 and 1 in 60,000[35,36].

In the case of Usher Syndrome type 2 (USH2), it is less severe than USH1, and patients manifest moderate to severe hearing loss, absence of vestibular dysfunction and subsequent onset of RP[37]. Moreover, the degree of hearing loss can vary within and among families[38]. The prevalence of USH2 is higher than USH1 and it is estimated to be 1 in 45,000[35].

Usher syndrome type 3 (USH3), was defined later than USH1 and USH2. It is characterised by progressive hearing loss, variable vestibular abnormality and RP[39]. This form of USH is less frequent than the two ones described above and it has been estimated to comprise 2% of all Usher syndrome cases[40].

Finally, considering all three subtypes of USH, they affect between 1 in 12,000 or 1 in 30,000 people in different populations. Moreover, it is estimated that Usher cases may represent between 10% to 30% of all recessive cases of RP[14].

2.2.2.2. Bardet-Biedl syndrome and related syndromic ciliopathies

Syndromic ciliopathies are a group of diseases caused by alterations in primary cilia. Primary cilia are ubiquitously expressed in eukaryotic cells and play an important role as sensors, relaying information either from the extracellular environment or between two compartments of the same cell[41]. Photoreceptors are defined as ciliated cells with a primary cilium of modified structure and function. An alteration in genes implicated in primary cilia

biogenesis or maintenance, produces its dysfunction and frequently affects photoreceptors causing RP (there are cases where the visual impartment is due to cone-rod dystrophy, generalised severe dystrophy or macular dystrophy)[42]. All ciliopathies share the RP phenotype, but depending on the implicated gene, other tissues are also affected. In fact, there are 10 main target organs in ciliopathies: 1) Bones, with chondrodysplasia, 2) The limbs with polydactyly, 3) adipose tissue with obesity, 4) the kidney with nephronophthisis, 5) the liver and liver fibrosis, 6) the olfactory system with anosmia, 7) the retina with retinal degeneration, 8) the central nervous system with intellectual or cerebellar vermis hypoplasia 9) the gonads with infertility, 10) the heart with situs inversus.

Bardet-Biedl (BBS) is the most prevalent ciliopathy, although it is a very rare disease with a prevalence of 1 in 160,000 in Northern Europe[43], there are some regions with higher prevalence as 1 in 13,500 in the Bedouin population[44].

Indeed, BBS and USH are recognised as the major causes of syndromic retinal dystrophies. BBS is characterised by early onset leading to severe visual dystrophy before adulthood, renal failure, central nervous system failure (with cognitive impairment), obesity and polydactyly[41] (Figure 7).

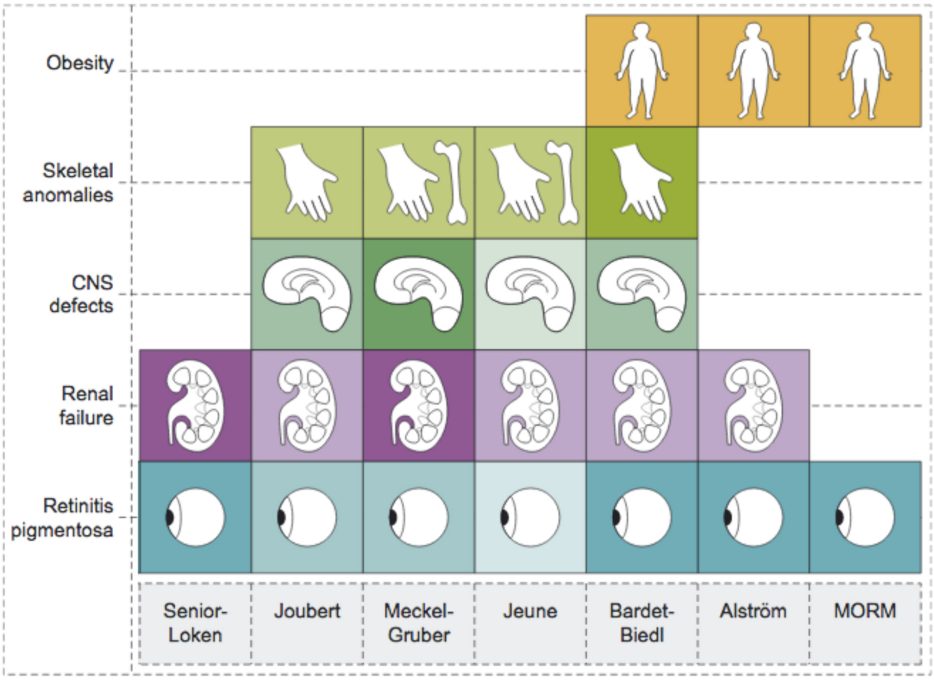


Figure 7: Different ciliopathies and the number of affected organs in each disease. The intensity of the colours indicates the severity of the phenotype in each tissue. Abbreviations: CNS, central nervous system. (Adapted from Moeckel *et al.,* 2011).

Patients with other ciliopathies also undergo retinal degeneration and alterations in different tissues with different (see Figure 7).

3. GENETICS AND GENE MECHANISMS OF IRDs

Inherited retinal dystrophies are a genetically heterogeneous group of diseases. Considering the inheritance pattern, they can be divided into autosomal dominant, autosomal recessive, X-linked and even into non-Mendelian inheritance patterns such as mitochondrial or digenic inheritance patterns[14,45]. There are over 250 genes associated to different IRDs (https://sph.uth.edu/retnet/home.htm). To further complicate the genetic basis of these diseases, different mutations in the same gene can produce different phenotypes[46] (Figure 8) or even the same mutation can produce different phenotypes[47]. Moreover, some genes inheritance patterns vary, and they can be related to autosomal dominant or autosomal recessive inheritance. For instance, mutations in RP1 gene have been related to autosomal dominant RP (adRP) and with autosomal recessive RP (arRP)[48]. On top of that, the repertoire of mutations and most prevalent mutated genes varies depending on the population. The frequency of mutations in unusual genes can be more frequent in isolated or consanguineous populations such as Finish or Ashkenazi Jewish in which a high prevalence of mutations in *USH3A* gene, which is very infrequent in most populations, has been described[49]. Most of our patients belong to the Spanish region of the Basque Country, mostly from the province of Gipuzkoa, which has been reported to be a genetically homogeneous region. Gipuzkoa has a high frequency of consanguinity, ranging between 5% to 30%[50], which highlights the interest to analyse the mutated gene spectrum of this region. In fact, in other diseases such as neuromuscular disorders and Parkinson disease it has been reported the presence of mutations not previously described, which are specific for this popilation[51].

Considering the types of mutations, a broad repertoire of mutations in IRD related genes have been described, such as missense, nonsense, splicing, frameshift or in-frame variants in exonic or splicing regions and also point mutations in deep intronic regions. Apart from this, large rearrangements such as deletions or duplications expanding all gene or just various exons[52,53] and chromosomal translocations have also been described[54]. Therefore, almost all types of mutations are described in this complex group of diseases.

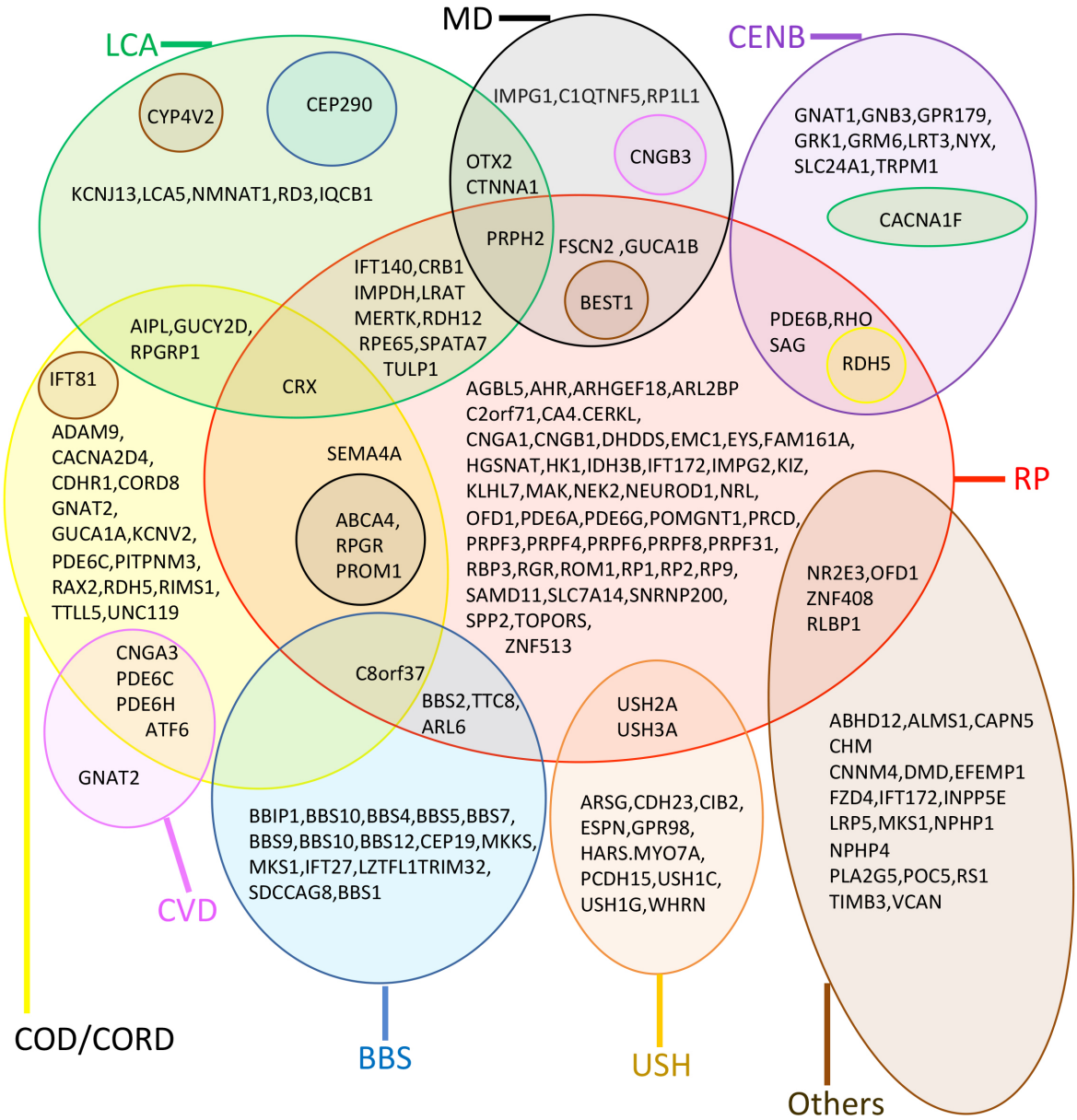


Figure 8: Genetic overlap of different IRD genes and phenotype. Clinical phenotypes are indicated by color circles and are grouped in: Abbreviations: BBS; Bardet-Biedl, RP; retinitis pigmentosa, CVD; Colour Visual defects, COD/CORD; Cone dystrophy/Cone-rod dystrophy, USH; Usher syndrome, LCA; Leber congenital amaurosis, Others; Other syndromic diseases where retinal degeneration is implicated, MD; Macular dystrophies, CENB; Congenital stationary night blindness. (Adapted from Berger *et al.*, 2010).

3.1. Inheritance patterns and most prevalent genes

All IRDs can also be divided according to their inheritance patterns. The principal three ones are: autosomal dominant, which encompasses between 30-40% of the cases, the autosomal

recessive which comprehends about 50-60% of the cases and X-linked that includes 10-15% of the cases[14].

3.1.1. Autosomal dominant inheritance pattern.

It is considered autosomal dominant inheritance pattern, when there is a causative mutation just in one allele in a gene located in an autosome (non-sex related chromosome). In most cases dominant inheritance patterns are observed when there are two or three consecutive generations in a family with affected members of both sexes. However, it has to be considered that in some cases, the mutation is first observed in the index case and the progenitors are not carriers of the alteration (*de novo* cases). It is also important to note that in some cases of dominant inheritance related genes, such as *PRPF31*[55], incomplete penetrance has been observed. This occurrence changes the inheritance pattern observed in family pedigree, making it more similar to a recessive pattern.

Most of genes causing IRDs with autosomal dominant inheritance pattern, are related to RP. Indeed, the most prevalent mutated gene is *RHO* found in 25-30% of the cases of adRP[56]. Other genes causing adRP with a high prevalence are *PRPF31, RP1, PRPF8* or *IMPDH*[56] (Figure 9). There are also genes related to other IRDs that are inherited in a dominant pattern such as *CRX*, which is responsible for LCA, CORD and RP, or *GUCA1A*, that causes COD or CORD (Figure 9 Figure 8).

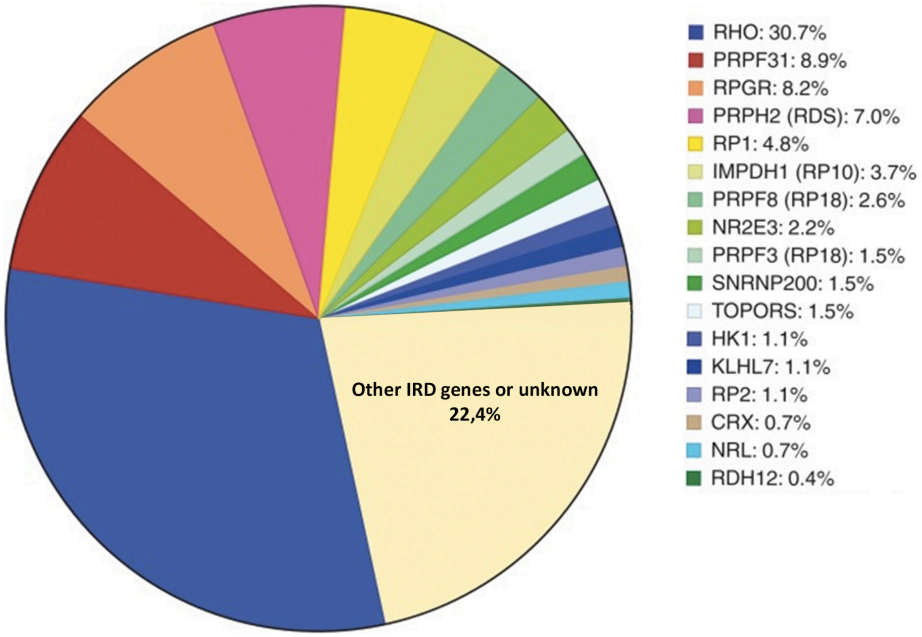


Figure 9: Prevalence of mutated genes in autosomal dominant and X-linked IRDs. (Adapted from Daiger *et al.*, 2014).

3.1.2. Autosomal recessive inheritance pattern

This is the most frequent inheritance pattern in IRD patients. In most cases, both progenitors are carriers of an altered allele. Sometimes the description of the inheritance pattern is difficult and is more difficult nowadays due to a reduced number of family members because of decreasing birth rates. Those cases are known as sporadic or single cases, in which the most prevalent inheritance pattern is the recessive. However, it is noteworthy that those cases could also be *de novo* dominant cases as described above.

The number of genes related to recessive IRDs is higher than those related to dominant ones. In fact, more than 200 genes[57] have been described (https://sph.uth.edu/retnet/home.htm). Most of them are very rare and cause less than 0.5% of recessive IRDs[49,58]. The most prevalent mutated gene in recessive IRDs is *USH2A*, mutations in this gene are linked to Usher syndrome type 2 and RP (Figure 10, Figure 8). Moreover, there are other genes also with relatively high frequencies such as *EYS*[59,60] ,*ABCA4*[14], *PDE6B* and *PDE6A*[14,49] (Figure 10).

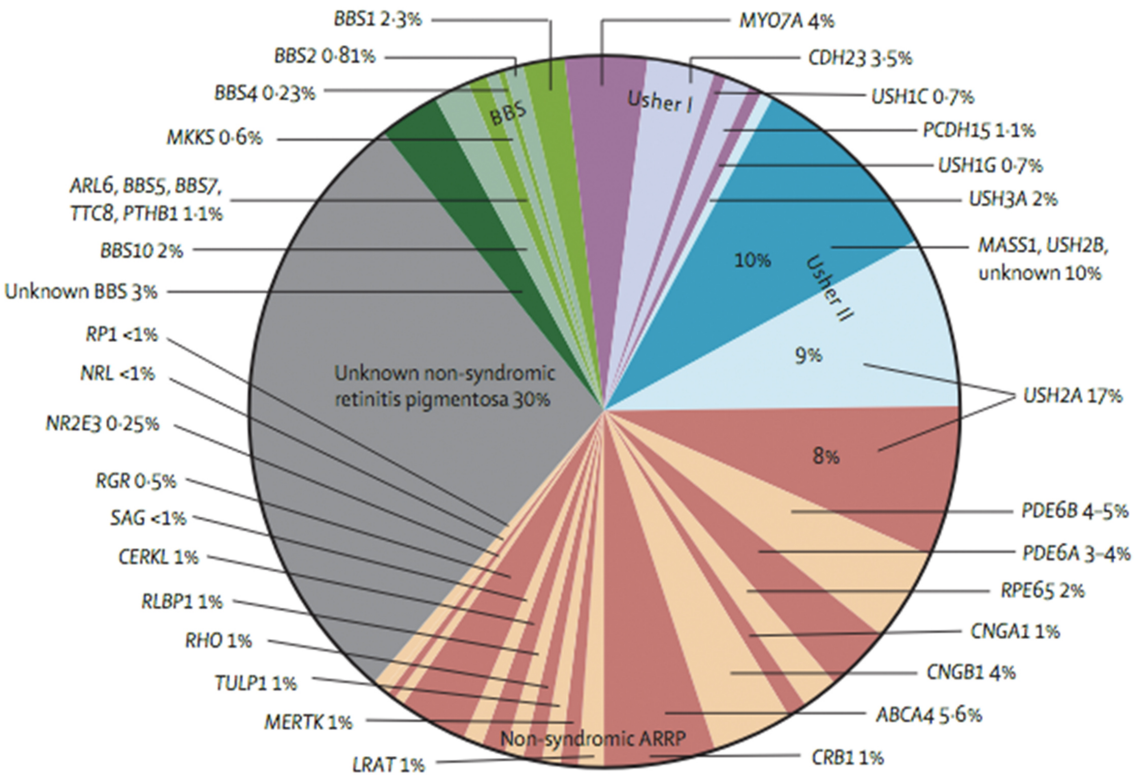


Figure 10: Prevalence of mutation in recessive inheritance pattern IRD genes. (Hartong *et al.*, 2006)

3.1.3. X-linked inheritance pattern.

This inheritance pattern consists on the transmission of the disease from women, to their offspring, and principally men are affected. However, cases in women have been described but in most of the cases affected women have variable phenotypes. This can occur due to skewed X chromosome inactivation[61] giving a variable number of incorrect transcripts[62,63]. Moreover, it is also important to take into account X-linked inheritance pattern in sporadic cases where the sole affected individual is male, indeed it has been observed that in 15% of simplex male cases mutations in x-linked genes are observed[64] as causative of the disease.
The most prevalent mutated genes with this inheritance pattern are *RPGR* and *RP2* (Figure 9). In the case of *RPGR* gene, there is a region called ORF15 which corresponds to the 15th exon of the gene where many pathogenic mutations are located. This region is highly repetitive, and it is considered challenging to amplify and the majority of mutations found in *RPGR* correspond to this region[65].

3.2. Molecular mechanisms of genes implicated in IRDs

The genetic basis that involves the development of IRDs is highly heterogeneous and complex. As it was mentioned above, mutations in the same genes cause different retinopathies (Figure 8). This genetic overlap suggests that there are similarities in the underlying molecular mechanisms of different IRDs.
The majority of genes mutated in RP and in most of IRDs encode for proteins that are expressed either in photoreceptor cells or in the RPE and are involved in several metabolic pathways related to physiologic functions of the retina. The alteration of any of those mechanisms produces the photoreceptor degeneration and their apoptosis.
Here we summarize the major mechanisms underlying photoreceptors degeneration (Figure 11).

snRNP/name of the protein	Function of the spliceosome complex	Related to RP
U4 snRNP		
7 Sm proteins	Stability of the particle	–
SNU13 (15.5K/NHPX)	Binding of U4 snRNA, U4 snRNP formation	–
PRPF31 (hPrp31/61K)	Interaction with PRPF6, formation of the tri-snRNP	+
U6 snRNP		
LSm proteins 2-8	U6 snRNA stability, U4/U6 annealing	–
SART3	U4/U6 snRNA annealing	–
U4/U6 snRNP		
PRPF3 (hPrp3/90K)	Binding of the U4/U6 duplex, tri-snRNP stability	+
PRPF4 (hPrp4/60K)	Tri-snRNP stability	+
PPIH (USA-Cyp/CypH)	Pre-mRNA splicing?	–
U5 snRNP		
7 Sm proteins	Stability of the particle	–
TXNL4A (hDib1/15K)	The thioredoxin fold superfamily	–
SNRNP40 (WDR57/40K)	Protein-protein interaction?	–
CD2BP2 (52K/Snu40/Lin1)	Interacts directly with Prpf6 and TXNL4A, not part of the U4/U6•U5 tri-snRNP	–
DDX23 (hPrp28/100K)	DEAD-box RNA helicase motif but the ATPase activity not confirmed	–
PRPF6 (hPrp6/102K)	Interaction with PRPF31, formation of the tri-snRNP	+
EFTUD2 (hSnu114/116K)	Regulation of SNRNP200 activity	–
SNRNP200 (hBrr2/200K)	Unwinding of U4/U6 snRNA duplex during splicing, activation of the spliceosome	+
PRPF8 (hPrp8/220K)	Formation of U5 snRNP, regulation of SNRNP200 activity, pre-mRNA splicing	+
U4/U6•U5 tri-snRNP		
SNRNP27 (27K)	?	–
USP39 (hSad1/65K)	Ubiquitin specific peptidase, recruitment of the tri-SNRNP to the spliceosome	–
SART1 (hSnu66/110K)	Recruitment of the tri-SNRNP to the spliceosome	–

Figure 11: Genes in which proteins form the spliceosome subunits. The function of each protein and its implication in retinitis pigmentosa is also shown. (Adapted from Růžičková *et al.*, 2017).

On one hand, there are genes that encode for proteins that are necessary for different parts of the visual cycle in rods. If any of these genes is altered, the recycling of all-*trans* retinol to 11-*cis* retinol is blocked resulting in the accumulation of toxic bis-retinoids that can undergo photo-oxidation. This process finally, forms lipofuscin, which is increased in some cases of photoreceptor degeneration[66,67]. For example, *RPE65*, *LRAT* and *RGR* genes are involved in retinol metabolism during visual cycle and *ABCA4* is necessary for all-trans retinal transportation to the photoreceptor's cytoplasm during visual cycle. Thus, if any of these genes is modified, the visual cycle is altered producing the final photoreceptor degeneration. On the other hand, there are alterations in genes that encode for proteins involved in the structure of the outer segment discs of photoreceptors (*PRPH2* and *FSCN2*)[68] or in proteins relevant for intracellular traffic (*RPGR, RP1* and *RP2*). Moreover, mutations in *MERTK* gene cause phagocytosis defects[69]. There are also genes involved in different compounds metabolic pathways such as *ABCA4* and *CERKL* in lipid metabolism, *IMPDH* in nucleotides metabolism and *TULP1, CRB1, MITS2, CA4* and *SEMA4A* in intermediary metabolism[67]. In addition to that, there are proteins which are involved in the photoreceptor's differentiation such as *NRL, NR2E3* or in the composition of extracellular matrix such as *USH2A.* Finally, there are also some genes related to the splicing process such as *PRPF3, PRPF8, PRPF31, PRPF4, PRPF6, RP9* or *SNRP200*. This aspect will be expanded in the next point.

3.2.1. Splicing process in adRP

The mutations in genes that are related to the splicing process are responsible for autosomal dominant retinitis pigmentosa. This process is particularly interesting as it involves genes that are widely expressed in different tissues, although the patients' phenotype is restricted to photoreceptors in the retina. *PRPF3, PRPF8, PRPF31, PRPF4, PRPF6, RP9* and *SNRP200* are involved in the assembly of the spliceosome, which is a large ribonucleoprotein (RNP) that carries out the removal of introns from pre-mRNAs. It comprises the U1,U2,U4/U6 and U5 small nuclear RNPs[70] and different genes encode for each subunit (Figure 11).

Considering that the correct removal of introns is essential in all cell transcription, it is intriguing that mutations in some spliceosome subunits affect only to the photoreceptors and produce RP. Possible explanations of this phenomenon are: 1) Photoreceptors have specialised splicing machineries which makes them to be more vulnerable to alterations in

the splice components than other cell types[70]. 2) The splicing of photoreceptor-specific genes is selectively affected. Studies suggest that, for instance, removal of intron 3 of *RHO* is inhibited by mutation in *PRPF31* gene[71] . This means that the depletion of core splicing proteins resulted in transcript specific splicing defects[72]. 3) Due to specialised requirements, such as high renovation rate of photoreceptors, they depend more strongly than other cell types on efficient splicing[16].

4. MOLECULAR DIAGNOSIS OF IRDs

4.1. Brief introduction to clinical genetics

Clinical genetics involves the study, counselling and treatment of individuals and families with inheritable disorders and disease predisposition. One if the most important aspects of human genetics is the quest to uncover the genetic basis of the disease. Defining the relationship between an alteration in a gene and the resulting disorder is essential for understanding the human biology and at the end, for giving a prognosis and/or therapeutic options to the patient. It is thus no surprising that a significant effort has been applied to the gene and mutation discovery processes. In fact, there are different diagnostic tools available for that purpose and during the last years new diagnostic approaches have been emerging for improving the correct diagnosis.

4.2. Diagnostic techniques

4.2.1. Evolution of the diagnostic techniques

In 1953, Watson and Crick (with the help of Maurice Wilkins and Rosalind Franklin´s X-Ray crystallography) suggested the structure of DNA as a double helix[73]. This discovery was the foundation of modern molecular genetics. The molecular methods developed since that moment utilize the complementary base-pairing postulated in that first paper and other papers[74]. During the next years, as previously suggested by George Gamow, it was proposed that three bases (codon) define an amino acid[75]. Later, in 1970, the central dogma of molecular genetics that consists on “DNA produces RNA, RNA produces protein”[76] was proposed. This knowledge has helped to develop new methods over the last 60 years. In the early 1970´s DNA sequencing techniques were developed permitting great advances in the field. In 1970 restriction enzymes were discovered after finding that an enzyme from *Haemophilus influenzae* split and cut at a specific sequence of DNA[77], this allowed the development of a specific mutation recognition tool based in the specific restriction pattern. In 1975, Southern blotting was described by Ed Sothern[78] enabling the visualization of DNA fragments in agarose gel after having transferred onto a membrane. Later, the polymerase chain reaction (PCR) was invented by Kary Mullis in 1985[79]. All these methods have been

developed for many uses during the last years and are the base for the new methods widely used nowadays for molecular genetics.

4.2.2. Sanger Sequencing

Sanger sequencing was developed by Frederick Sanger in 1977 together with Alan Coulson[80]. It was based on the "plus and minus" sequencing systems developed in 1975. However, the drawback to this system was that it was difficult to determine the length of runs of the same nucleotide. So that in 1977 they redefined the technique[81], and radiolabeled dideoxynucleotides (ddNTPs) were added to the reaction mix. These ddNTPS are analogues of (deoxinucleotides (dNTPs) and lack the 3´hydroxyl group required for chain formation by the DNA polymerase. This time, for the reaction mix, a DNA polymerase was utilised for DNA synthesis, labelled primers that recognise the sequence, template DNA and dNTPs were required for chain synthesis. A proportion of the chains will "terminate" when ddNTP was added instead of dNTP[82]. This reaction was repeated once per 4 types of ddNTPs and run them out in a polyacrylamide gel followed by exposure to X-ray film. Finally the sequence was able to be read[82].

The next improvement in the technique was the development of cycle sequencing. It uses a thermostable DNA polymerase and the polymerase chain reaction (PCR). This combination of techniques reduced the amount of DNA needed for the sequencing[83].

A further refinement of the technique was the addition of different fluorescent dyes to each ddNTP replacing the radioactive labelling. This allowed the possible automation of the technique. In fact, Applied Biosystems (ABI), launched the first commercial sequencer ABI 370A, which could generate 1000bp per day. Finally, in 1996, ABI launched the capillary electrophoresis DNA sequencer. Using this method, each sanger reaction reads can reach up to 700-1000bp with a per-base raw accuracy of near 100%. This accuracy is higher than that achieved with NGS technologies but has a much higher cost per base and it is slower[84].

This method is considered to be sequencers of first-generation, with the latest technologies being denominated as Next-Generation Sequencing systems[84].

4.2.3. Screening techniques

4.2.3.1. Single-strand conformational polymorphism analysis (SSCP)

SSCP is a mutation detection technique based on the fact that single-stranded DNA (ssDNA) has a defined conformation. Single base changes alter the DNA conformation and can cause differences in migration under nondenaturing electrophoresis conditions[85,86]. After a PCR of the region of interest, the products are denatured before an electrophoresis running under non-denaturing conditions. Several methods have been developed to visualize the SSCP mobility shifts, such as silver staining of the electrophoresis gel, incorporation of radioisotope labelling or fluorescent dye-labeled PCR primers.

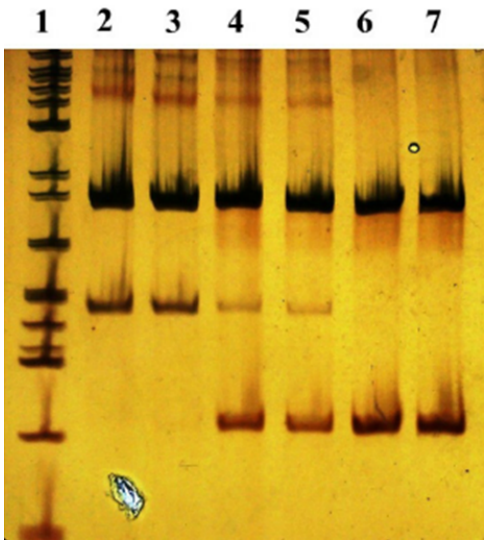


Figure 12: Representation of a silver stained SSCP gel. Modified from Fonseca P.A *et al.,* 2013.

4.2.3.2. High Resolution melting (HRM) analysis

HRM analysis is a post-PCR analysis method that provides rapid identification of genetic variations. This method is based on biophysical measurement of amplified DNA. The PCR product, with saturating dyes that fluoresce in the presence of double-stranded DNA, is dissociated from this conformation to its two single strands. The melting profile of a PCR product depends on its GC content, length, sequence and heterozygosity, so its denaturation allows the study of those conditions by measuring the change of fluorescence intensity per unit of time during the melting process. A real-time PCR instrument and specific analysis

software are necessary for the analysis[87,88]. This technique has been implemented for mutation screening and clinical diagnosis of many different genetic diseases[89,90]

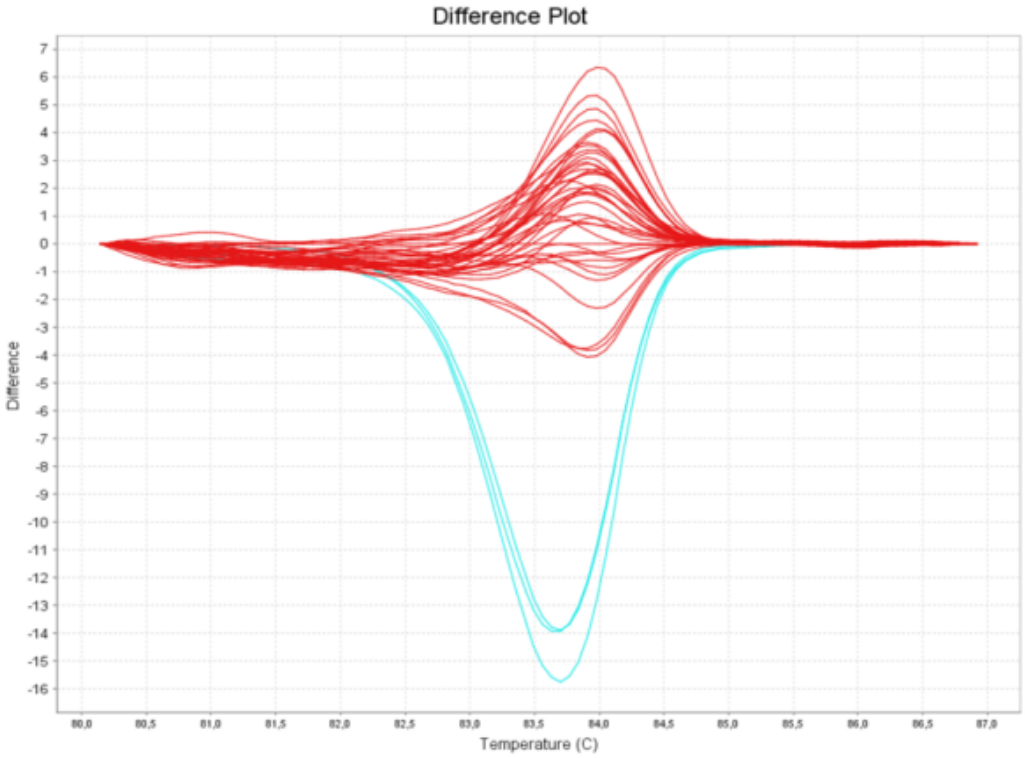


Figure 13: Representation of a melt curve.

4.2.3.3. Genotyping microarrays

Genotyping microarrays consist of chips that screen a limited number of genetic variants which have been previously reported as causative of the disease. The number of the variants analysed in each array depends on the design (between 200 and 1,500 variants). These arrays have been used in different pathologies such as IRDs. Custom designed microarrays have been analysed for different IRDs such as Stargardt disease[91], Usher syndrome[92], arRP[93] and adRP[94]. Moreover, there are other microarrays that simultaneously analyse over 300,000 genetic variants throughout the genome (Axiom Exome Array Plates; Affymetrix, Santa Clara, Calif., USA). Our group used these arrays in a recent study, analysing 5,000 IRD variants in 76 IRD families[95]

4.2.4. NGS Revolution

Next generation sequencing (NGS), describes a DNA sequencing technology, which has revolutionised genomic research. Using this technology, an entire human genome can be sequenced within a single day[96]. Although Sanger sequencing is considered to be the gold standard for accuracy with reported error rates below 1%[97] it has the major disadvantage of being slow and expensive. This process is time consuming for routine clinical diagnostics sequencing. The need for a less time consuming process has led to the development of NGS. There are different platforms using varied sequencing technologies but all of them share the common feature of being able to sequence a huge number of genes and samples at one time[82]. This allows obtaining higher molecular diagnostic rates and the possibility of finding new mutations. Moreover, over the last years, the cost of the technique has been reduced while the efficiency improved, making these techniques the first line approaches in clinical diagnosis.

4.2.4.1. Platforms based on NGS

4.2.4.1.1. Pyrosequencing

Pyrosequencing was first described in 1996 and is based on the detection of pyrophosphate release when a nucleotide is incorporated into a growing DNA strand[82]. When reaction is made in the presence of ATP sulphurylase and luciferase, each incorporation of nucleotide, results in the production of light which is detected by a camera. The addition of the dNTPs is performed in different cycles, adding only one type of dNTP in each cycle.

In 1999 pyrosequencing was first used for massively parallel sequencing, when Jonathan Rothberg set up the 454 Life Sciences company which was bought by Roche company in 2007. Interestingly, this technology was used for James Watson´s genome sequencing which was sequenced in a shorter period of time and was less expensive than the first individual genome published one year before[98,99]. This technique is based on sequencing longer reads (up to 500bp) than other methodologies. However, the drawback to pyrosequencing is its high error rate especially in homopolymer regions[100].

4.2.4.1.2. Illumina

The Illumina sequencing system (formerly known as Solexa), is considered the leader in the field. This technique has a resemblance to Sanger sequencing. In fact, it uses ddNTPs to terminate the synthesis of a strand, but the chain termination is reversible allowing synthesis of a complementary strand which is performed using one nucleotide type in each ligation cycle. The identification of the incorporate ddNTPs is determined by fluorescent labelling (each ddNTP is labelled with different colour and emits at different signal) and is detected using a laser. The sequencing process starts with the ligation of specific adapters to the DNA fragments, which are denatured and attached randomly to nearby primers that are already covalently connected to a solid surface called flow cell[101–103]. Thereafter, each single strand fragment creates a "bridge", hybridising its free end to the complementary adapter on the flow cell surface. This process is carried out in repeated cycles creating up to 1000 identical copies (polymerase colonies, called DNA "polonies") (Figure 14)[82,84,102]. Illumina technology is based on cyclic reversible termination, making it much less susceptible to the homopolymer errors than 454 Roche pyrosequencing or ion torrent sequencing[104] and has demonstrated it is very sensitive. However, this platform does display some under-representation in AT-rich[105] and GC rich[105] regions and come with a false-positive rate around 2,5%[106,107]. Finally, Illumina sequencing is much cheaper than the 454 and involves less time for the preparation of samples[82]. With all the characteristics, Illumina allows for a wide range of applications, but it is used specially in genome sequencing through whole genome sequencing (WGS) or exome sequencing (WES)[104].

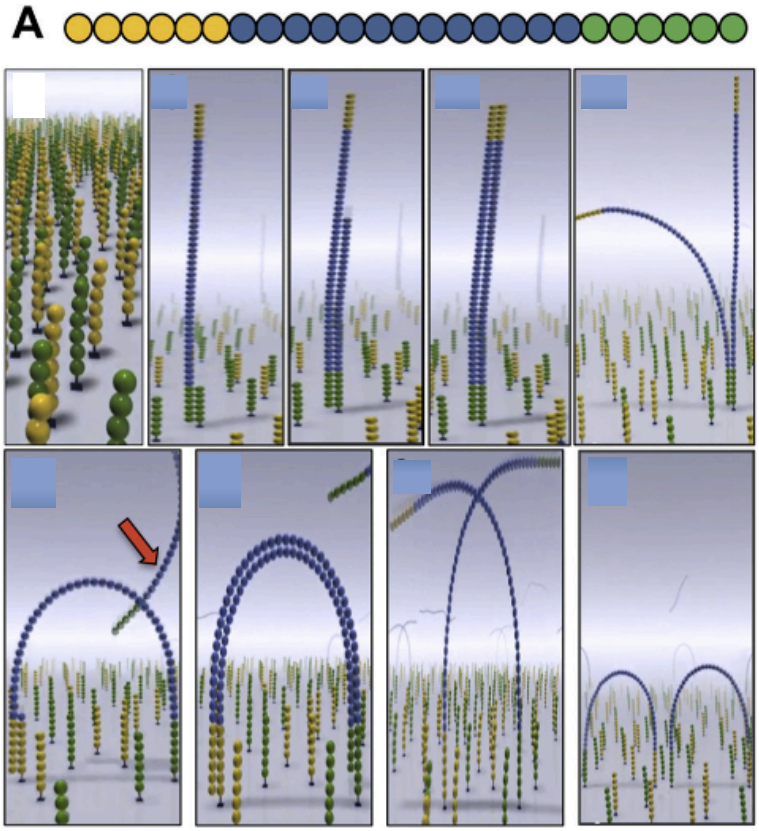


Figure 14: Schematic representation of the "bridge" PCR amplification process used in Illumina sequencing platform. A. DNA construct with adaptors in yellow and green and the sequence of interest in blue. Red arrow. Shows that *de novo* synthesized sequence binds to an adjacent primer, generating a bridge and the template strand is removed. Adapted from Anasagasti *et al.,* 2013.

4.2.4.1.3. Sequencing by Oligonucleotide ligation and Detection (SOLiD)

Applied Biosystems launched this technology in 2007, which was then bought by Thermo-Fisher Scientific. The fragment preparation for this methodology shares similarities with the one used for 454, in which sheared fragments are ligated to adapters, attached to beads and the amplification is performed using emulsion PCR. In this case, the sequencing is performed by ligation, which involves the hybridisation and ligation of labelled probes and primer sequences to the DNA strand to be sequenced[82]. The labelled probes encode two known bases followed by a series of degenerate five bases that anneals to the DNA. Sequences that anneal perfectly are ligated to the primer and in that moment the image is captured. Afterwards, the fluorophores are cleaved, and a new cycle starts (Figure 15). The process is repeated ten times until two of every five bases are identified.

This technology is highly accurate (99.9%)[104], as each base is probed multiple times. However, SOLiD method´s maximum read length is just 75bp which makes the alignment

difficult. Moreover, there is also evidence that shows some under-representation of AT-rich and GC-rich regions[105] and some substitution errors.

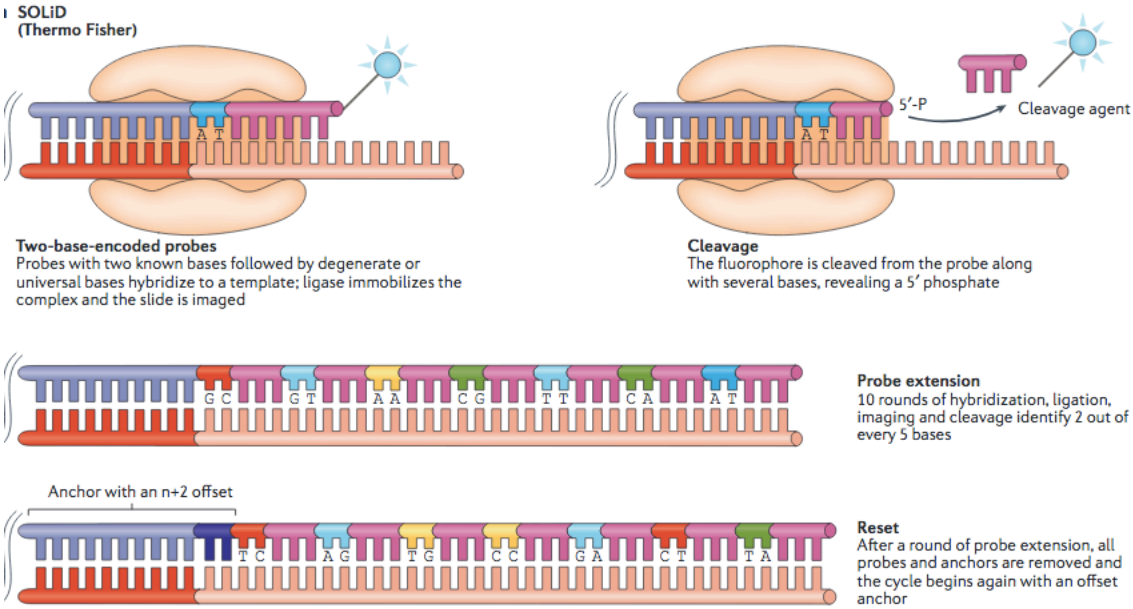


Figure 15: Sequencing by SOLiD method: In dark blue two-base-encoded probe, composed of known nucleotides followed by universal bases in pink. The two-base probe is ligated to the anchor in purple that is complementary to an adapter in red. The process is repeated 10 times. Adapted from Goodwin *et al*., 2016.

4.2.4.1.4. Ion semiconductor sequencing (Ion Torrent)

Ion Torrent PGM was the first manufactured benchtop sequence. Semiconductor sequencing is also known as pH-mediated sequencing. Part of the methodology is similar to that of 454, in this case the fragmented DNA is attached to adapters, amplified by emulsion PCR and linked to a nano-well on a chip. This chip is a metal-oxide semiconductor chip that contains nano-wells each one holding a different DNA fragment. Beneath the nano-wells is an ion sensitive layer, a pH meter, below which is a proprietary Ion sensitive Field Effect Transistor Sensor (ISFET) which transmits an electrical current[101]. Nucleotides are allowed to flow one at a time over the chip along with the required enzymes. When a complementary dNTP is incorporated, a positively charged hydrogen ion is generated. The charge from the ion changes the pH of the solution (a decrease in pH indicates a hydrogen ion has been released) which is detected by the ISFET leading to a shift in voltage allowing DNA sequencing without scanning cameras or light mediation[101] (Figure 16). Considering the chips above-mentioned, Ion Torrent offers several different types, to tune sequencer performance to the needs of the researcher. The throughput of these chips ranges from 50Mb to 15Gb, with running

times between 2 and 8 hours, making it faster than the other described platforms[104]. Moreover, it is considered the most economic technique of the ones described due to it obviates the need for optical methods for reading the sequence. This makes the device well suited for gene-panel sequencing and for point of care clinical applications. In fact, Ion Torrent is attempted to grow in clinical sequencing[104].

However, as this platform relies on single-nucleotide addition system, it has many drawbacks. Insertion and deletion errors dominate and homopolymer regions are problematic in measuring homopolymers larger than 6-8bp[108].

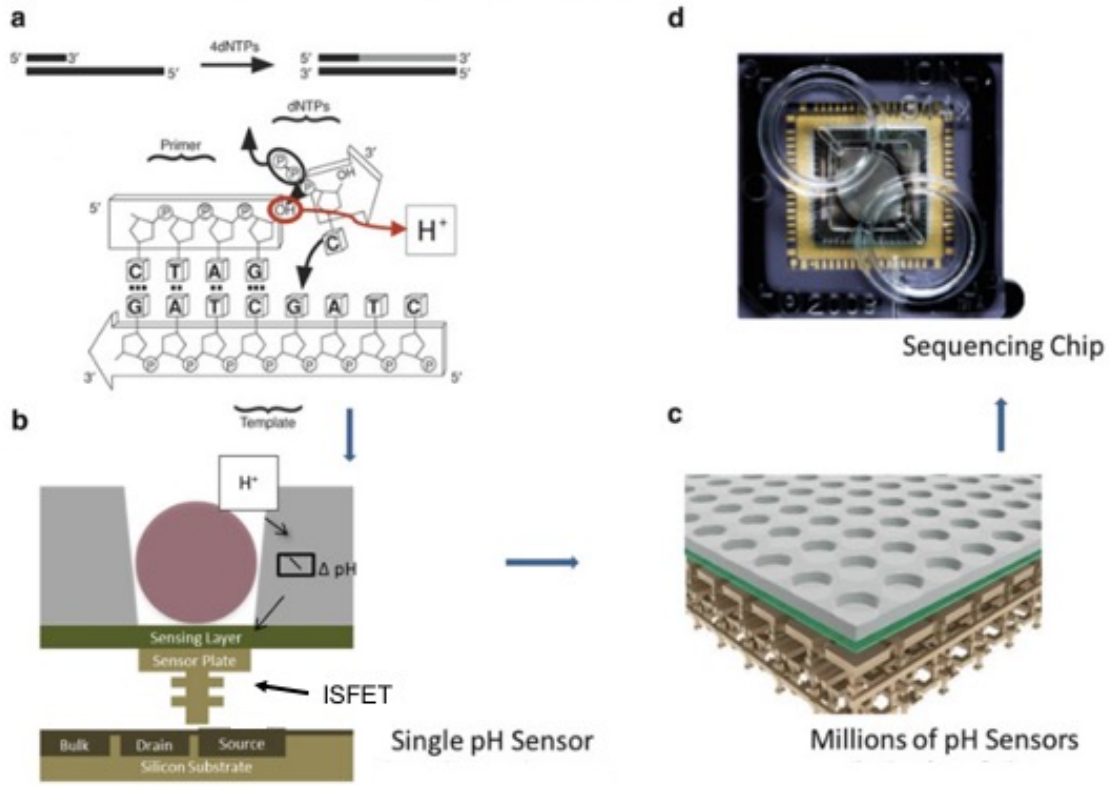


Figure 16: Principles and elements of semiconductor sequencing. a) Represents the mechanisms of semiconduction sequencing where every new nucleotide leads to a release of a H^+. b) shows the ISFET sensor in detail. c) Several million pH sensors. d) An image of the chip used for sequencing.

4.2.4.2. Application

4.2.4.2.1. Targeted Gene Panels

Targeted Gene panels are widely used in clinical genetics. These panels contain a concrete number of genes or regions associated with the disease of interest. On one hand, pre-designed panels with preselected content are available. In this case they are panels already

designed for a disease or groups of diseases. Although there are different companies, ThermoFisher Scientific is one of the main suppliers of this kind of panels (https://ampliseq.com/login/login.action). On the other hand, there are custom panels available. In this case, each researcher designs the panel with the regions of interest for the study. In the case of IRDs, different targeted gene panels have been used for diagnosis such as arIRDs[109,110], adIRDs[111,112] or most prevalent genes in IRDs[113–115].

An important advantage of this approach is that the data analysis is less complex than in WES or WGS approaches. Indeed, in the case of panel-based runs made with Ion Torrent technology, there is available a "user friendly" software called Ion reporter to facilitate the variant filtering is available. Moreover, focusing on individual genes or gene regions, much higher sequencing depth than WES is obtained, enabling identification of rare variants. However, targeted panels have some drawbacks. The most important one is that the analysis is limited to the genes introduced in the design. Regarding the promptness of the discovery of new genes, in some diseases the panels can rapidly become obsolete making it necessary to analyse new genes with other techniques or generating a new complementary panel. Furthermore, this methodology is less efficient than WES or WGS in the detection of structural variants such as CNVs[116,117].

4.2.4.2.2. Whole exome sequencing and whole genome sequencing (WES and WGS)

Whole exome sequencing, is an approach in which all exons of protein coding genes in the genome are covered, reaching only about 2% of all the genome. The first successful use of WES for human patient diagnosis was in the identification of the causal variant of a rare form of inflammatory bowel disease in 2011[118]. Since this initial diagnosis, exome sequencing has been used extensively specially in clinical research for new genes characterization in different diseases[119–121]. In the case of whole genome sequencing, all base pairs that form the genome are sequenced. Therefore, all intronic regions and regulatory regions are sequenced, which can increase the number of diagnosed patients since it has been observed that there is a high percentage of pathogenic variants in those regions[122].

As far as WES is concerned, it concentrates the sequencing power in the protein coding regions, which ensures a good coverage depth and achieves high quality genotyping of these regions. However, the remaining 98% of the genome is not covered which results in the

limitation of the discovery of regulatory region and intronic region changes that is not a drawback in the case of WGS. The second limitation is that as WES features an enrichment step that is primer based, this can introduce a bias against poorly annealed regions. These limitation can be overcome by WGS, since it covers more than 95% of the entire genome and it is not based on primer-based erinchment[82]. Nevertheless, as WGS covers almost all the genome, it also generates a huge amount of data which is laborious to process and interpret. On the one hand, special computing equipment is needed to be able to process all the data correctly. Moreover, there is not an established pipeline for data filtering and as in the case of WES, there are different strategies for it, causing the obtaining of different results depending on the algorithms used for the filtering process. On the other hand, the interpretation of variants in intronic regions or regulatory regions is still challenging due to the high number of variants found and the experiments needed for pathogenicity validation of the variant. However, despite being highly dependent on technical support, the use of whole genome sequencing is gaining momentum in clinical practice, and it seems plausible that it will become feasible in a near future, once a robust translational genomics workflow becomes an affordable option both in economic and technical terms[123].

4.2.5. MLPA

Multiplex Ligation-Dependent probe amplification (MLPA) is an alternative approach to identifying CNVs[124]. This technique includes the detection of deletions and duplications of different size, from all genes to one exon or even subtelomeric deletions. MLPA is a multiplex PCR-based screening method designed to determine in a single reaction tube the copy number of up to 50 DNA[125]. MLPA amplification uses specific oligonucleotide probe-pair for each fragment. Moreover, these oligonucleotides contain the sequence of a universal forward and reverse primer sequences where the forward primer is fluorescently labelled. The oligonucleotides recognize the sequence adjacent to the target DNA and when both hybridize to their target sequence they are ligated to form a complete probe. After ligation, probes are amplified in a PCR reaction (Figure 17). Each complete probe has a different length so it can be separated by capillary electrophoresis and by comparing the peak pattern obtained, deleted or duplicated regions can be identified comparing them to a control sample (Figure 17).

MRC-Holland is the only supplier of MLPA kits. In the case of IRDs they have different kits or salsas®, that cover genes of interest such as *ABCA4, EYS, USH2A, PRPF31, RP1, RHO, IMPDH, CHM, RP2* and *RPGR.*

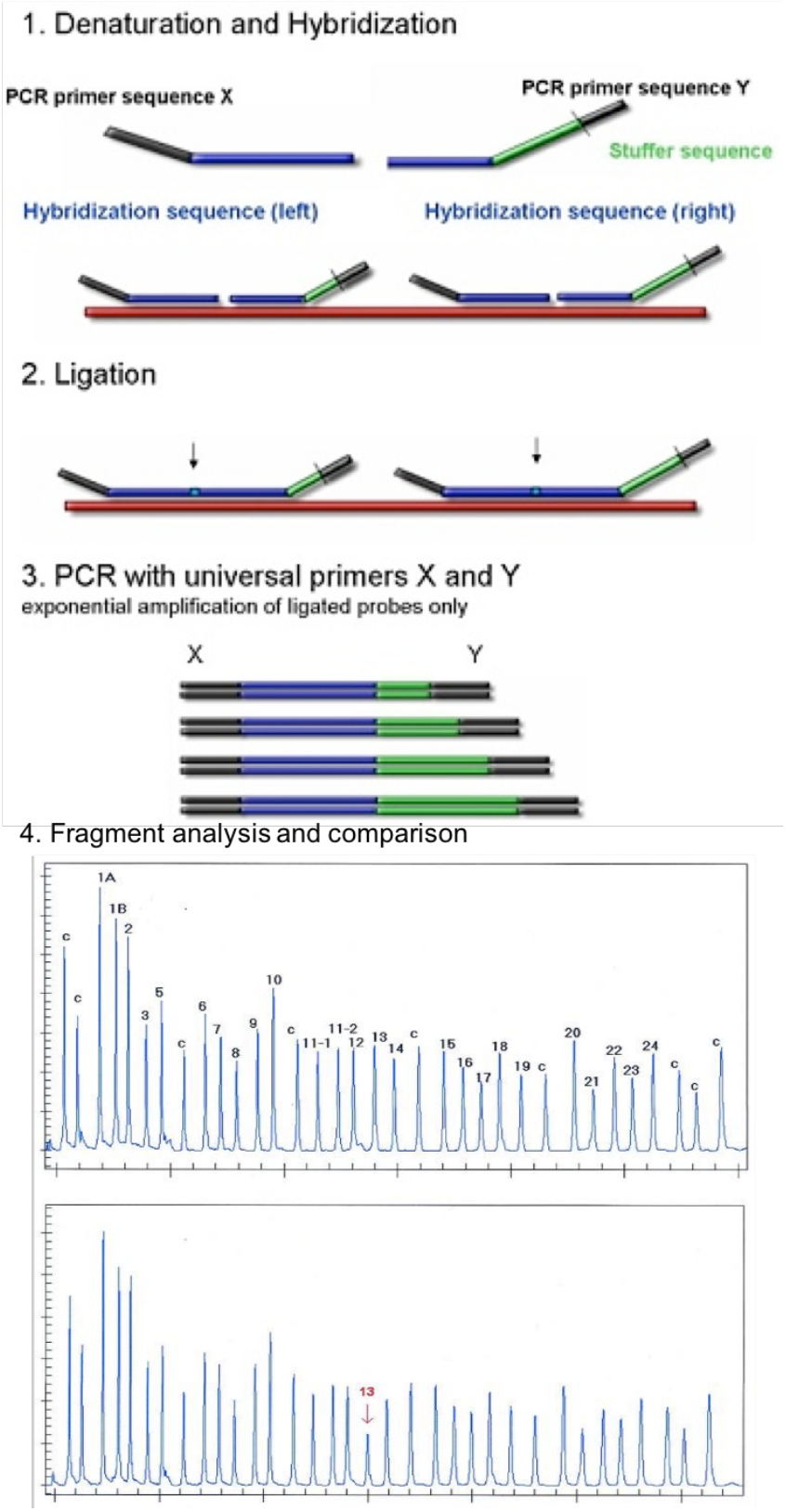


Figure 17: MLPA procedure schema: 4. The analysis shows a control sample above a patient sample with a deletion in BRCA1 gene. Adapted from MRC-Holland (https://www.mlpa.com/WebForms/WebFormMain.aspxand) and Schouten *et al*., 2002.

4.2.6. Comparative Genomic Hybridisation (CGH) array

This technique allows the detection of CNV and, depending on the array type used, it allows also the detection of Loss of Heterozygosity (LOH) regions in higher number of genes than MLPA or even in all exomes. There are different companies that offer distinct type of arrays such as customized arrays and arrays, which analyse all exome. In general, this methodology is based on denaturation and fragmentation of the DNA to prepare it for labelling with a specific fluorophore. In some arrays the patient genomic DNA is labelled in Cy5 and a reference genome is labelled with Cy3. In other arrays a reference DNA is not used. After labelling, samples are hybridised against the oligonucleotide probes adhered into a chip or a slide, for 24 hours. Finally, the different fluorescent intensity patterns are compared. In the case of the platforms where a reference DNA is also hybridised at the same time as the patient, both samples compete to hybridise to their target oligonucleotide and the analysis of the fluorescence shows equal expression when yellow colour (fusion of Cy5 and Cy3) is observed. The resolution of the technique depends on the distance between the consecutive oligonucleotide probes[82]. Moreover, in some arrays, SNP probes are also added for LOH region analysis.

LOH regions: Loss of heterozygosity is a common genetic event that indicates that part of the genome appears to be homozygous[126]. This phenomenon can occur due to defects in homologous recombination during meiosis producing uniparental disomy (UPD), due to consanguinity between the patient´s progenitors or due to similarities in the patient´s progenitors ancestors or even due to big deletions. In fact, it is known that the more closely the parents are related, the greater this effect of LOH is expected to be[127]. All these mechanisms produce two equal copies of a small region or even of a whole chromosome. This phenomenon is often seen in cancer, where a locus that is heterozygous in a normal cell becomes homozygous in a cancer cell derived from the normal cell. LOH region analysis has been used for mutation analysis in inherited recessive diseases[127] in order to find the candidate gene where the pathogenic mutation is located.

4.2.6.1. Customized arrays

As in the case of panel-based NGS, CGH arrays are also customizable. In this case, Agilent Technologies is the major supplier of this type of CGH arrays, that contain a concrete

number of genes. This number and the resolution obtained, depends on the number of oligonucleotides used in the design. In the case of Agilent technologies, there are different capacity options available; 60,000 probes, 180,000 probes, 400,000 probes and 1Million probe arrays (the number of probes is per patient analysed). The main advantage of this type of arrays is that the analysis is easier, due to the restricted number of CNVs to be analysed. However, it is not possible to find variations in new genes.

4.2.6.2. Exome arrays

This type of array is designed for CNV analysis of all the exons in the genome. The exons are densely covered and the intronic adjacent regions of the exons are less covered. There are different companies offering this type of arrays, being Affymetrix (Thermo-Fisher) the major supplier. As in customized arrays, the resolution also depends on the number of probes introduced in the array. In the case of Affymetrix, different capacity arrays are available; *CytoScan* 750, that contain 750,000 probes; *CytoScan HD* which contains 2.67 million markers and *CytoScan XON* containing 6.55 million probes. In contrast to customized array this technique is a better option for analysing other genes not previously associated with the disease. However, the analysis is by far more difficult due to the high number of variants obtained.

HYPOTHESIS

There is a high genetic heterogeneity between all different types of Inherited Retinal Dystrophies (IRDs). Both in patients with syndromic or non syndromic IRDs the most widely used approaches for variant analysis during last years have been targeted arrays and panel based NGS. With these techniques, a large amount of different mutations have been identified but there are still patients without molecular genetic diagnosis.
Using a combination of high throughput genetic mutation discovery approaches, we will be able to identify the genetic causal disease of most of our IRD patients.

It has been described that genetically homogeneous populations, have higher percentages of founder mutations. Since most of our patients have ancestors from the Basque Country, a genetically homogeneous region, we worked under the hypothesis that at least a fraction of our patients are carriers of founder mutation(s) in gene(s) previously related to IRDs and in genes not previously ascribed to any IRD, which might be prevalent in our region.

OBJECTIVES

General objectives

Our main objective was to advance in the molecular genetic characterization of IRD patients, using both currently available and novel methodology.

Specific Objectives

-To implement a methodological pipeline for the diagnosis of IRD patients.

-To develop and validate a new strategy based on targeted pooled DNA sequencing.

-To analyse by whole exome sequencing (WES) patients with no molecular characterization after analysing genes associated to IRDs.

-To study the frequency of patients with CNV mutations in patients not characterized by targeted NGS, using MLPA and CGH array techniques.

-To analyse deep intronic mutations in patients with *USH2A* gene monoallelic mutations or Usher syndrome type 2 phenotype by mRNA analysis.

CHAPTER 1

High prevalence of mutations affecting the splicing process in a Spanish cohort with autosomal dominant retinitis pigmentosa.

Ezquerra-Inchausti M., Barandika O., Anasagasti A., Irigoyen C.,
Lopez de Munain A., Ruiz-Ederra J.

Scientific Reports 6, 39652; doi:10.1038/srep39652(2017)

INTRODUCTION

Retinitis pigmentosa (RP; MIM# 268000) is the most frequent form of inherited retinal dystrophy (IRD), with a prevalence of 1 in 3000-4000 cases worldwide [1]. It is characterised by a progressive dysfunction associated with the death of rods and/or cones, which leads to retinal atrophy and loss of vision. The mode of inheritance of RP is complex, with autosomal dominant (ad), autosomal recessive (ar), X-linked (xl) Mendelian cases and some cases of digenism or mitochondrial forms having been reported[14,128,129]. From a genetic perspective, over 80 disease-causing genes are currently associated with RP, 27 of which have been associated with adRP (http://www.sph.uth.tmc.edu/retnet). However, to date, mutations in the known adRP genes account for only 50-75% of dominant cases, depending on the test and population used in the study[130]. This percentage is increasing, mainly due to the implementation of Next Generation Sequencing (NGS)-based techniques[45,56,111] and the discovery of new RP genes[131–134].

The majority of the pathogenic mutations in humans have been described inside the exons, the codificant part of the genes, however most human genes harbour introns that are removed during pre-mRNA splicing post-transcriptional modification[135]. The splicing reaction is catalysed by the spliceosome, a multisubunit complex comprising small noncoding nuclear RNAs (U1, U2, U4, U5, and U6) and several associated proteins[136]. The spliceosome orchestrates the two transesterification reactions needed to remove introns and to join the adjacent exons, and operates by step-wise formation of sub-complexes that recognise regulatory sequences and promote efficient splicing[95,135,136].

Mis-regulation of splicing is a common feature of many human diseases, including several retinal diseases[137]. These disorders can be caused by mutations that disrupt the splicing of specific genes or by mutations in genes coding for splicing factors, both of which lead to a general loss of spliceosomal function. Thousands of splice-site mutations have been identified in patients with retinal dystrophies. Although most of these mutations disrupt a consensus splice-site sequence and cause exon skipping, some result in intron inclusion, novel exon inclusion, or the usage of cryptic upstream or downstream splice sites. The resulting alteration in the protein sequence, which is often concomitant with frameshift and premature termination, unsettles the functional protein domains and leads to degeneration

of the retina[138]. For example, mutations in several genes coding for core spliceosomal proteins, such as pre-mRNA splicing factors (*PRPF3, PRPF4, PRPF6, PRPF8, PRPF31, RP9*) or RNA helicases (*SNRNP200*), are responsible for adRP[95,138,139]. However, given that these genes are expressed ubiquitously in all tissues and are highly conserved in all eukaryotes, it remains unclear why mutations in these genes are associated exclusively with adRP. Studies performed in rodent retina showed that *PRPF3, PRPF31, PRPC8* expression levels are higher in the retina than in other tissues in normal adult mice, thus suggesting that the retina may have a higher basal splicing demand than other tissues given that it is one of the most metabolically active tissues in the body[138,140,141]

In order to effectively identify adRP mutations, we have sequenced 31 genes associated with the autosomal dominant inheritance pattern using the Ion PGM platform (IPGM; Life Technologies), in combination with Sanger sequencing. We selected these genes as they have been linked to most of the cases of adRP reported. Remarkably, we found a high prevalence of mutations affecting the splicing process among our families, especially mutations affecting trans-acting splicing factors. This is of particular interest considering that several splicing-based therapeutic approaches, some of which are in clinical trials[137,139], are under active development for mutations affecting either core spliceosomal proteins or splice site mutations of individual genes.

The results of the present study will help in genetic counselling and will contribute to a better characterisation of the disease. Moreover, they may have a therapeutic impact in the near future in the light of analogous approaches used for other RNA mis-splicing diseases[139].

MATERIALS AND METHODS

Study subjects

RP patients were diagnosed at the Ophthalmology department of Donostia University Hospital (San Sebastian, Spain). Diagnostic criteria were night blindness, peripheral visual field loss, pigmentary deposits resembling bone spicules, attenuation of retinal vessels, pallor of the optic disc and diminution in a- and b-wave amplitudes in the electroretinogram[22]. A total of 29 Spanish probands with a family tree compatible with adRP were included. Samples from an additional four patients, three corresponding to patients with known mutations that we had detected in previous analysis and one from a non-affected individual, were included as positive and negative controls, respectively[88,95]. Family trees were generated from information obtained from probands. All procedures performed in studies involving human participants received approval from the institutional research ethics committee and were in accordance with the Declaration of Helsinki (2013) or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Human sample collection

High molecular weight DNA was extracted from blood samples from RP patients and their available family members. Total DNA from samples was extracted and isolated using and AutoGenFlex STAR instrument (AutoGen, Holliston, MA, USA) together with the FlexiGene DNA Kit (Qiagen, Hilden, Germany) following the manufacturer´s instructions. DNA concentrations were measured using a Nanodrop spectrophotometer (ND-1000, Thermo Scientific NanoDrop Products, Wilmington, DE, USA) and only those samples with 260/280 ratios ≥ 1.8 and 260/230 ratios ≥ 2 were used. DNA samples were stored at -80 ºC.

Amplicon Library preparation

A total of 663 primer pairs were designed and grouped in two Ion AmpliSeq Primer Pools to flank 31 IRD genes with a total coverage of 98.37% using the Ion AmpliSeq Designer software *(www.ampliseq.com)*. The regions excluded by the design represented only 1.63% of the total. Although most of the genes were related to adRP, representative genes associated with dominant forms of Leber congenital amaurosis and cone-rod dystrophies were also

included since the clinic symptoms associated with these genes are often hard to distinguish from those associated with RP (RetNet; https://sph.uth.edu/retnet/disease.htm) (see supplementary Table S1). The Ion AmpliSeq Library Preparation Kit v2.0 (Life Technologies, Foster City, CA, USA) was used to construct an amplicon library from genomic target regions with a maximum read length of approximately 200 base pairs (average length, 142 bp) for shotgun sequencing on the PGM. Briefly, target genomic regions were amplified by simple PCR using Ion AmpliSeq Primer Pools and 10 ng of each genomic DNA samples.

Sequencing Analysis

Ion Torrent Personal Genome Machine (PGM)

NGS was carried out on a PGM following the Ion PGM 200 Sequencing Kit protocol. Briefly, enriched Ion Sphere particles (ISPs) were annealed with the Ion Sequencing primer and mixed with the PGM200 Sequencing Polymerase. The polymerase-bound and primer-activated ISPs were then loaded into the previously checked and washed Ion 316 Chips (Life Technologies) and, after selecting the run plan on the Ion PGM System software, these chips were subjected to 500 cycles of sequencing with the standard nucleotide flow order. Signal processing and base calling for the data generated during the PGM runs were performed using the Ion Torrent platform-specific analysis software Torrent Suite version 4.0 to generate sequence reads. The sequences generated were aligned to the GRCh37/hg19 human genome for detection of genomic variants in the sequenced samples.

Sanger sequencing

Sanger sequencing was used to confirm those mutations detected by NGS and for co-segregation analysis. Primers were designed at least 60 bp upstream and downstream of the mutation. The amplicons were purified after PCR amplification, (ExoSAP-IT, USB Corporation*).* Sequencing was performed by dye termination DNA reaction on a 16-capillary ABI 3130xl platform (Applied Biosystems) according to the manufacturer's protocol. Sequences were analysed and compared with wild-type samples and reference sequences using the BioEdit Sequence Alignment Editor (Windows) and Ensembl and NCBI databases.

High resolution melting (HRM) analysis

HRM analysis was used to re-analyse those genomic regions with no or very low coverage in NGS platforms, following the previously described methodology[88].

Relevant variant identification and pathogenicity score

In order to determine genomic variants of relevance, we selected putative disease-causing variants using the following criteria: 1) variants previously reported as pathogenic, or 2) loss-of-function variants, such as stop gain, frameshift, small deletions or duplications (INDELS) and splice site variants, or 3) novel missense variants predicted to be damaging or highly pathogenic in at least four out of five web-based pathogenicity predictors, namely SIFT (<0.05), Polyphen2 (>0.750); PROVEAN[142]; GVGD[143]; MutationTaster[144]. Furthermore, all variants selected had to fulfil the criteria of having a Minor Allele Frequency (MAF) of less than 0.002, as obtained from human genome databases (see below), and being absent from Spanish in-house allele database with information from 578 unrelated Spanish individuals none of whom exhibited any IRD-related disease[145] *(*http://csvs.babelomics.org/*; s*ee Figure 18).

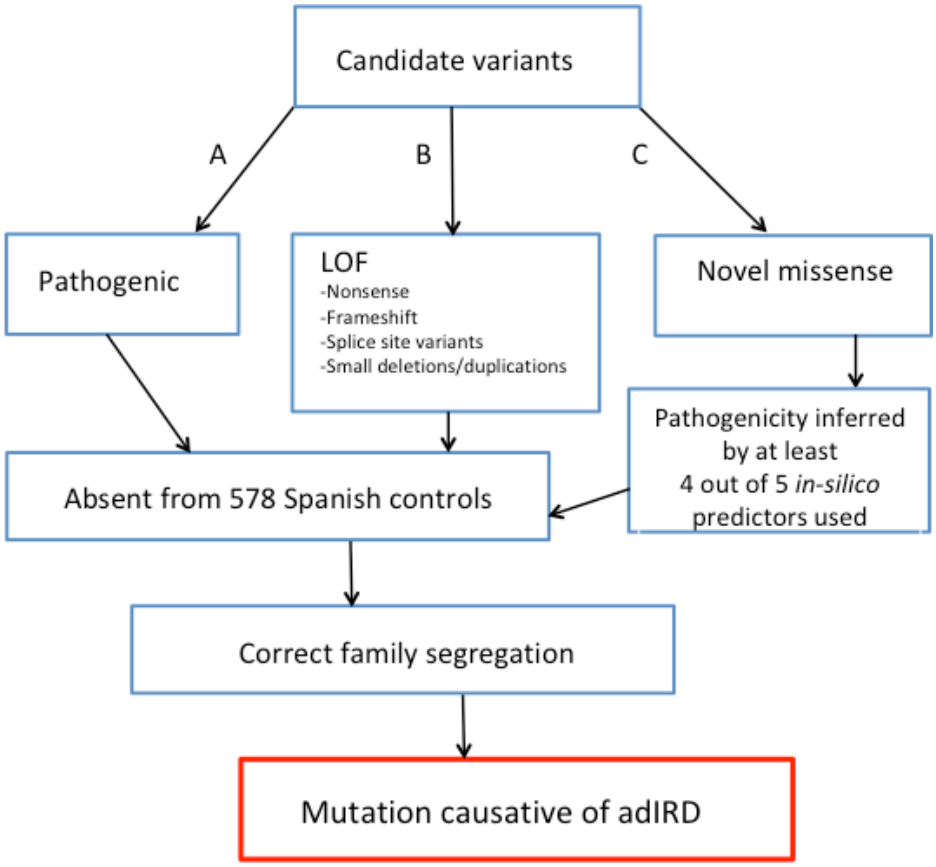


Figure 18 : **Schematic representation of the criteria used to select mutations responsible for autosomal dominant inherited retinal dystrophies**. adIRD: Autosomal dominant Inherited Retinal Dystrophies. LOF: loss of function.

Web sources

Ensembl, http://www.ensembl.org/

NCBI, http://www.ncbi.nlm.nih.gov/

Polyphen-2, http://www.genetics.bwh.harvard.edu/pph2/

RetNet, http://www.sph.uth.tmc.edu/Retnet/

SIFT, http://www.sift.bii.a-star.edu.sg/

SNPnexus, http://www.snp-nexus.org/

The Human Genome Variation Society (HGVS), http://www.hgvs.org/

1000 Genomes, http://www.1000genomes.org/ ENREF 48

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

Babelomics, http://csvs.babelomics.org

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

RESULTS

High variant detection coverage and sensitivity was achieved

An average of 3.3 million reads/chip was obtained. On average, each amplicon present in the panel was covered 658 times, with 95.92% of amplicons with >30x coverage and 94.27% of amplicons with >50x coverage. Those regions with no or low coverage (<30X), probably due to the presence of repetitive sequences or self-annealing of primers, were re-analysed. A highly sensitive, cost-effective method described recently by us that combines high resolution melting (HRM) analysis with direct sequencing was used for this re-analysis[88]. This allowed us to expand our analysis to 97% of target amplicons. Despite the implementation of HRM, no additional mutations were found within these re-analysed regions.

Variant identification

An average of 45 variants, including SNPs and INDELS, were initially identified for each sample in the targeted regions, including the negative control with 51 SNPs, none of which were putative disease-causing as expected (see Supplementary Table S2). After the clinically relevant variant identification screening described in the materials and methods section, we were able to identify putative disease-causing mutations in a total of 14 out of the 29 probands, which resulted in a ratio of clinically relevant genetic findings of 48.28%. A description of the main features of the genetic findings can be found in Table 1.

An average of 3.3 million reads/chip was obtained. On average, each amplicon present in the panel was covered 658 times, with 95.92% of amplicons with >30x coverage and 94.27% of amplicons with >50x coverage. Those regions with no or low coverage (<30X), probably due to the presence of repetitive sequences or self-annealing of primers, were re-analysed. A highly sensitive, cost-effective method described recently by us that combines high resolution melting (HRM) analysis with direct sequencing was used for this re-analysis[88]. This allowed us to expand our analysis to 97% of target amplicons. Despite the implementation of HRM, no additional mutations were found within these re-analysed regions.

FAMILY	GENE	MUTATION	TYPE	REF	HSF	PROV	SIFT	PH	MUT TASTER
RP19S	*PRPH2*	NM_000322 c.797G>A p.Gly266Asp	missense	[146]		D	0	0.99	Disease causing (0.999)
RP22 RP37 RP64,RP101 RP102,RP134 RP157	*SNRNP200*	NM_014014 c.3260C>T p.Ser1087Leu	missense	[147,148]		D	0	1	Disease causing (0.999)
RP90	*PRPF8*	NM_006445 c.6974_6994del p.Val2325_Glu 2330del	deletion	[149]			n/a	n/a	Disease causing (0.999)
RP113	*PRPF8*	NM_006445 c.6945del p.Leu2315 Leufs*2336 Aspext*21	frameshift	[149]			n/a	n/a	Disease causing (1)
RP133 RP146	*RHO*	NM_000539 c.937-1G>T	splice acceptor variant	[150]	Decrease 5' acceptor site of exon 5 (90.7>61.75)		n/a	n/a	Disease causing (1)
RP105	*RHO*	NM_000539 c.1045T>C p.Ter349Glu	stop loss	[151]			n/a	n/a	Polymorphism (0.999)
RP135	*RHO*	NM_000539 c.568G>A p.Asp190Asn	missense	[152]		D	0	0.431	Disease causing (0.999)

Table 1: Summary of mutations responsible for retinitis pigmentosa: Abbreviations: D: deleterious; HSF: human splicing finder; MUT TASTER: Mutation Taster; n/a: not available; PH: Polyphen; PROV: Provean; REF: bibliographic reference. All variants were absent in a Spanish in-house allele database containing information from 578 unrelated Spanish individuals (Spanish controls). See Materials and Methods section for detailed information.

An average of 3.3 million reads/chip was obtained. On average, each amplicon present in the panel was covered 658 times, with 95.92% of amplicons with >30x coverage and 94.27% of

amplicons with >50x coverage. Those regions with no or low coverage (<30X), probably due to the presence of repetitive sequences or self-annealing of primers, were re-analysed. A highly sensitive, cost-effective method described recently by us that combines high resolution melting (HRM) analysis with direct sequencing was used for this re-analysis[88]. This allowed us to expand our analysis to 97% of target amplicons. Despite the implementation of HRM, no additional mutations were found within these re-analysed regions.

Variant identification

An average of 45 variants, including SNPs and INDELS, were initially identified for each sample in the targeted regions, including the negative control with 51 SNPs, none of which were putative disease-causing as expected (see Supplementary Table S2). After the clinically relevant variant identification screening described in the materials and methods section, we were able to identify putative disease-causing mutations in a total of 14 out of the 29 probands, which resulted in a ratio of clinically relevant genetic findings of 48.28%. A description of the main features of the genetic findings can be found in Table 1.

A total of seven variants in four genes were found in 14 families. Two of the mutations, both in *PRPF8* gene, were deletions. One consisted in a loss of 21 nucleotides (p.Val2325_Glu2331del) and the other consisted of a frameshift deletion involving a single-point deletion (p.Leu2315Leufs*2336Aspext*21). Figure 19 shows colour fundus pictures of patients RP90 and RP113 bearing these two deletion mutations. Both variants were potentially pathogenic, co-segregated with the disease, were predicted as pathogenic by MutationTaster[149] and were previously described[149].

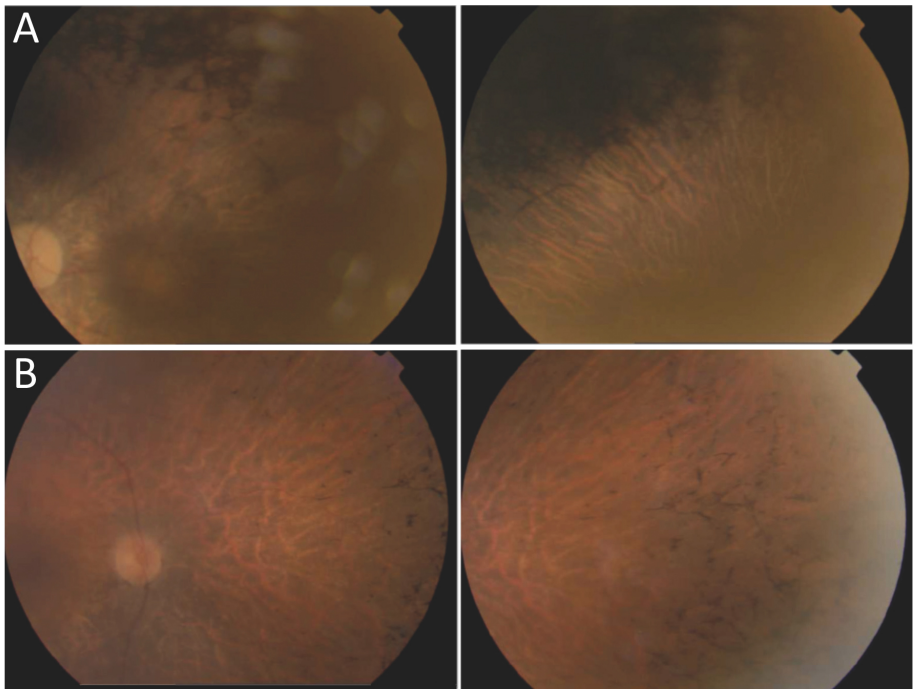


Figure 19: Fundus photographs of patients with deletion mutations in PRPF8. A. Patient RP90 (p.Val2325_Glu2330del) shows optical disc pallor, arteriolar attenuation and macular atrophy (right), with dense pigment in the mid-periphery (left). B. Patient RP148 (p.Leu2315Leufs*2336Aspext*21) shows optical disc pallor, arteriolar attenuation and bone spicule-shaped pigment deposits in the mid-periphery. The left and right pictures correspond to the left and right eyes, respectively.

Two genes were involved in 37.93% of our cohort of families, with *RHO* affecting four probands with three different mutations and *SNRNP200* affecting seven probands, all with the p.Ser1087Leu mutation[147,148].

The high prevalence of mutations affecting the splicing process among our families (12 out of 29 probands studied), represented the 38% of the probands in our adRP cohort. Most cases (9/29) were due to mutations affecting the genes *SNRNP200* (7) and *PRPF8* (2), which code for core spliceosomal proteins, although a splice site mutation in RHO[150] was also detected (2/29).

With respect to *SNRNP200*, after performing Sanger sequencing in all available family members we identified c.3260C>T (p. Ser1087Leu) mutation in a total of 12 cases from seven families (see representative family in Figure 20A). Co-segregation analysis showed that two out of seven healthy subjects analysed for this variant in these families were mutation carriers, which likely indicates cases of incomplete penetrance similar to what has recently been reported for this variant in a study also involving a Spanish cohort[111] (see Figure 20B*).* We also found a total of nine individuals in two families with c.937-1G>T mutations affecting *RHO* splicing. Interestingly, one of these nine patients is asymptomatic, probably due to the

disease being in an initial state given his young age (21 years old at the time of the study; *see* Figure 20C and Supplementary Fig. S1*).*

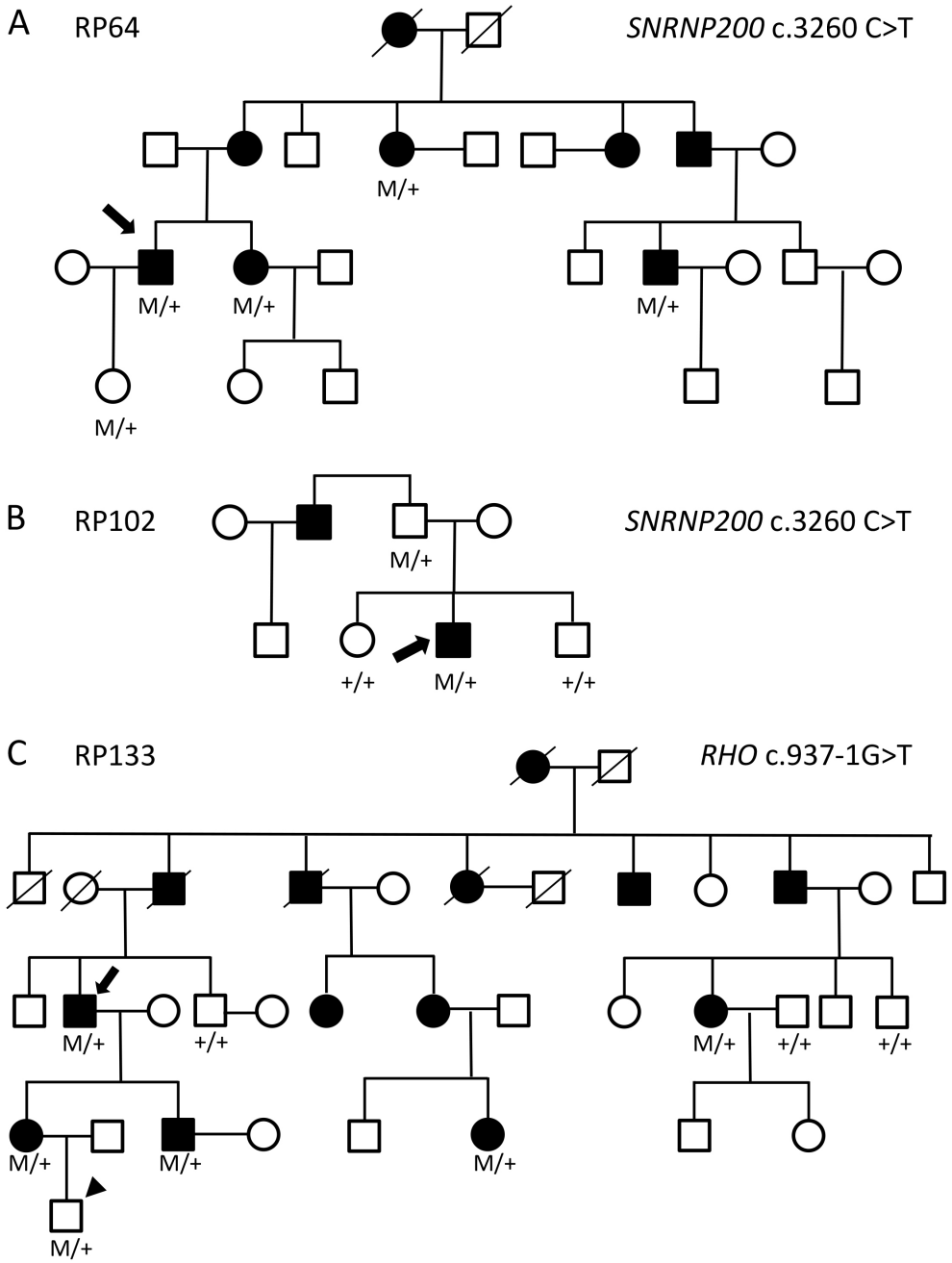


Figure 20: Representative trees for families with the two most prevalent mutations found in SNRNP200 and RHO genes. The c.3260C>T (p.Ser1087Leu) mutation in *SNRNP200* was found in families RP64 (**A**) and RP102 (**B**). **C.** The c.937-1G>T mutation in the *RHO* splice acceptor site in a total of six individuals from family RP133, one of whom is a young asymptomatic patient (arrowhead). Genotypes are annotated as M/+ (heterozygote) or +/+ (wild type). Arrows indicate proband patient.

Finally we also found mutations in both *RHO* and *PRPH2* genes that were not related to the splicing process: a stop loss in RP105[151] and a missense mutation in RP135[152], both in *RHO*, and a missense mutation in *PRPH2* (p.Gly266Asp) in patient RP19S[146]. Patient RP19S was included in this study since he is the son of a patient with a mutation in *PRPH2* that we had diagnosed previously[88]. Patient RP19S was asymptomatic at the initial diagnosis, when he was eight years old. However, two years later his molecular diagnosis confirmed the presence of the p.Gly266Asp mutation, therefore he was re-examined. This revealed a granular fundus and few bone spicules in the inferior periphery, with no signs of optical disc pallor or vascular attenuation. The visual field showed a concentric defect (preserving the central 18 degrees) with a hyperautofluorescent ring in the macula upon autofluorescence examination (see Table 2*).* Additional family trees of the rest of the patients recruited in the present study are included in supplementary Fig. S3.

PATIENT	AGE OF ONSET	SYMTONS AT DIAGNOSIS	VISUAL ACUITY IN 2015 (LogMAR VA RE)	VISUAL ACUITY IN 2015 (LogMAR VA LE)	ESPHERICAL EQUIVALENT RE	ESPHERICAL EQUIVALENT LE	CATARACT (YES, NO, PF -PSEUDOPHAKIA-, APHAKIA)RE	CATARACT (YES, NO, PF -PSEUDOPHAKIC-, AFAQUIC)LE	IOP RE	IOP LE	PALE OPTIC DISC	ARTERIOLAR ATTENUATION	BONE SPICULE PIGMENT	SECTORIAL RP(YES/NO)	EPIRETINAL MEMBRANE (ERM)	MACULAR EDEMA	MACULAR THICKNESS RE(STRATUS OCT)	MACULAR THICKNESS LE(STRATUS OCT)	MACULAR THICKNESS RE(CIRRUS OCT)	MACULAR THICKNESS LE(CIRRUS OCT)	VISUAL FIELD DEGREES	ERG
2	6	NYCTALOPIA	0.2	0.3	-0.25	-0.5	YES	YES	16	16	YES	YES	YES	NO	NO	YES	121	119	169	165	11	ND
16	43	DECREASE VA	0.2	0.1	0	0.125	PP	PP	19	16	YES	YES	YES	NO	YES	YES	146	230			4	NA
19S	13	ASYMPTOMATIC	0	0	0.5	0.875	NO	NO	13	14	NO	NO	YES	NO	NO	NO			266	249	18	NA
20	18	DECREASE VA	1	0.5	-5.25	-3.75	YES	YES	15	15	YES	YES	YES	NO	NO	NO					4	ND
22	45	NYCTALOPIA	0.5	0.8	1.625	2.25	YES	YES	15	17	YES	YES	YES	NO	NO	YES	114	140				ND
37	48	DECREASE VA	4	0.3	-5.375	-4.875	NO	YES	14	14	YES	YES	YES	NO	NO	NO	148	123			0	ND
39	27	DECREASE VA	0	0.2			PP	PP	16	16	YES	YES	YES	NO	NO	YES	244	324			8	ND
42	58	DECREASE VA	2	1.3			PP	YES	16	16	YES	YES	YES	NO	NO	NO	218	165			9	ND
43	49	VISUAL FIELD LOSS	0.2	0.5	-0.25	1.25	NO	YES	17	18	YES	YES	YES	NO	YES	YES	258	320				ND
48	58	NYCTALOPIA	0	0	0.625	-0.875	NO	NO	14	14	YES	YES	YES	NO	NO	NO	225	258	322	336	15	ND
64	22	VISUAL FIELD LOSS	0.1	0.3	-1.5	-1.75	YES	YES	12	12	YES	YES	YES	NO	YES	YES	328	345			10	ND
66	18	VISUAL FIELD LOSS	0.05	0.05	-11.125	-10.125	YES	YES	12	14	YES	YES	YES	NO	NO	NO	171	209			10	ND
69	29	NYCTALOPIA	0.18	0.2	-0.125	-5.75	PP	YES	14	14	YES	YES	YES	YES	NO	YES	191	234			5	ND
89	23	NYCTALOPIA	0.5	0.18	0.125	-1	YES	YES	10	10	YES	YES	YES	NO	NO	NO	121	212			11	ND
90	20	NYCTALOPIA	1.3	1.3	13.25	13.75	AP	AP	8	8	YES	YES	YES	NO	YES	NO	119	253				ND
105	12	NYCTALOPIA	3	2	-1	0.5	PP	PP	16	16	YES	YES	YES	NO								ND
101	37	NYCTALOPIA	0.3	0.3	-2.5	-2	YES	YES	14	14	YES	YES	YES	NO	NO	NO	226	296	312	330	15	ND
102	15	NYCTALOPIA	0.05	0	-5	-4.75	NO	NO	13	15	YES	YES	YES	NO	NO	NO			286	288		ND
113	14	NYCTALOPIA	0.18	0.18			YES	YES	14	14	YES	YES	YES	NO	YES	NO			260	269	5	ND
85	28	DECREASE VA	1	0.4	-8	-2.125	PP	NO	14	14	YES	YES	YES	NO	NO	NO			279	279		NA
80	29	DECREASE VA	0.2	0.5			YES	PP	18	18	YES	YES	YES	NO	SI	YES						NA
133	49	VISUAL FIELD LOSS	0.1	0.05	-0.5	-2.125	PP	YES	16	16	YES	YES	YES	NO	NO	NO			269	252	12	ND
134	44	DECREASE VA	0.2	0.4	-0.25	-0.25	YES	YES	18	18	YES	YES	YES	NO	YES	YES			287	272	10	ND
135	41	NYCTALOPIA	0	0	0.75	-0.625	NO	NO	14	14	NO	NO	YES	YES	NO	NO			267	276	SS	ND
146	37	NYCTALOPIA	0.05	0	0	0.25	NO	NO	17	15	NO	NO	YES	YES	NO	NO			281	294	12	ND
157	62	VISUAL FIELD LOSS	0	0	1.5	-0.625	NO	YES	15	18	YES	YES	YES	NO							5	ND
148	20	NYCTALOPIA	0,8	0,8	-0.25	-0.5	PP	PP	16	15	YES	YES	YES	NO								NA
70	49	PHOTOPHOBIA	0.00	0	0	0.125	NO	NO	14	14	NO	NO	NO	NO	NO	NO					20	NA
79	24	NYCTALOPIA	0.3	0.3	-1	0.5	YES	YES	13	15	YES	YES	YES	NO	NO	NO						

Table 2: Phenotypic aspects of the patients analysed. Abbreviations; AP: aphakia; N: normal; NA: not available; ND: not detectable; PP: pseudophakia; SS: superior scotoma.

DISCUSSION

In this work we have analysed the genotype and phenotype of a group of 29 adRP probands, using targeted NGS and Sanger sequencing to analyse 31 genes. We were able to detect putative disease-causing mutations in 14 out of the 29 probands analysed. This resulted in a clinically relevant genetic diagnosis ratio of 48.28%, which is comparable to values reported previously, ranging from about 24% to 88%,[56,111,113,116,153–157]. Several factors may be responsible for this wide range of diagnosis ratios reported, including the approach used or the nature of the cohort involved. In the present study, part of our cohort of adRP patients was already diagnosed in a previous study in which we screened some of the most prevalent adRP genes[88,95], therefore this might have contributed to the diagnostic ratio obtained.

Nevertheless, there is still a missing fraction of about 51% unsolved cases among our adRP cohort of 29 patients. One possible explanation is the presence of mutations in regions outside the 31 genes analysed, such as deep intronic regions. Another possibility is the presence of changes not detected by our analysis due to limitations in the design of our panel of target genes, such as large genomic rearrangements and mutations in novel genes. As such, it seems that the combination of NGS with other technologies, such as Multiplex Ligation-dependent Probe Amplification (MLPA) or Comparative Genomic Hybridisation arrays (aCGH), will be needed in order to address those genomic aberrations caused by copy number variations (CNV). Another possible explanation is the presence of novel RP genes among our patients, since most of them belong to the Basque province of Gipuzkoa, a well-known genetically homogeneous region[158,159]. Consequently, sequencing of the whole exome/genome could help in the discovery of novel RP genes.

A remarkable finding was the high prevalence of mutations affecting the splicing process among our families (11 out of 29 probands studied), representing 38% of the probands in our adRP cohort.

Most mutations were the p.Ser1087Leu mutation found in *SNRNP200*. This gene encodes for the 200-kDa helicase hBrr2. During splicing, the spliceosome undergoes structural rearrangements that are mediated by several RNA helicases including hBrr2, which is essential for unwinding of the U4/U6 snRNP duplex, a key step in the catalytic activation of the spliceosome complex[160,161]. hBrr2 comprises two helicase modules, one active and the other with regulatory activity.

All six mutations identified in *SNRNP200* to date, including the p.Ser1087Leu mutation, are located in the hBrr2 protein region containing the first DExD-helicase module, a key component for the U4-U6 unwinding function in vivo and in vitro and for cell survival[160–162]. The first of the two consecutive Hel308-like modules, which comprises a DExD/H domain and a Sec63 domain, shows the highest level of conservation among species, thus pointing to its functional relevance[163]. The p.Ser1087Leu mutation has been reported to reduce unwinding activity and to promote the use of cryptic splice sites, thus pointing to an influence of splicing fidelity[148,164].

Although most cases (9/29) were due to mutations affecting genes *SNRNP200* and *PRPF8* that code for spliceosomal proteins, splice-site mutations in *RHO* were also detected (2/29). The percentage of adRP probands with mutations affecting either spliceosome core factors or the splice site of several adRP genes accounted for 5–14.5% of all cases of adRP in previous studies[111,130,165,166]. With regard to mutations in the *SNRNP200* gene, although these were only initially described in two Chinese families[147,148], they have since been reported to contribute to a significant portion of cases of adRP in the Caucasian population, ranging from 1.5% to 4.2%[130,165,167,168].

The relatively high prevalence of splicing-related mutations found in our study is likely explained by the founder effect of two of the genes, which were present in very small and rather isolated Spanish populations.

Splicing modulation has been proposed as a therapeutic approach for several diseases. Two of the most advanced approaches in this regard are based on the use of modified antisense oligonucleotides (ASOs) to target specific RNA sequences and redirect splicing, and small molecules as modulators of the splicing process. A representative example of this approach is exon skipping for Duchenne muscular dystrophy (DMD), where the muscular protein dystrophin is prematurely truncated by mutations that disrupt the open reading frame, thus leading to a non-functional protein. Exon skipping creates an internally deleted and shorter than normal but partially functional protein, which leads to a much less severe phenotype in animal models of DMD. With respect to approaches based on small molecules and peptides, several splicing modulators have been shown to be effective in myotonic dystrophy (DM) and cancer[140,169].

As regards retinal dystrophies, most advanced therapeutic approaches that target splicing are aimed at correcting the splicing of individual genes using mutation-adapted U1 small nuclear RNA for the RPGR gene[170] or spliceosome-mediated RNA trans-splicing in *RHO*[171]. Both these approaches are based on cellular and animal models and have provided encouraging results. Once in the clinic, these promising approaches could be generalised and applied to other genes with splice donor site mutations[170] and to all adRP genes rather than only to *RPGR* and *RHO*, respectively[171].

With regard to therapeutic approaches targeting the splicing machinery, we are unaware of their use in retinal diseases. However, since the first steps towards the use of such therapeutic strategies have already been made for other diseases, it is plausible to imagine *a* broadening of the applications of small molecules to reverse aberrant splicing for other diseases, including retinal dystrophies, in the near future once our understanding of the mechanisms of the disease, and delivery systems and other technical issues, have been improved.

In summary, the combination of NGS with Sanger sequencing has allowed us to achieve a diagnostic rate of over 48%. As such, the methodology described herein exhibits a high diagnostic yield when applied to a well-defined adRP group and a relatively high number of genes. This will be of clinical relevance once ongoing studies on therapeutic options directed at manipulating splicing are completed.

CHAPTER 1 SUPPLEMENTARY INFORMATION

Supplementary Table S1

AIPL1, BEST1, CA4, CRX, FSCN2, GUCA1A, GUCA1B, GUCY2D, IMPDH1, KLHL7, NR2E3, NRL, OTX2, PITPNM3, PROM1, PRPF3, PRPF31, PRPF6, PRPF8, PRPH2, RDH12, RHO, RIMS1, ROM1, RP1, RP9, RPE65, SEMA4A, SNRNP200, TOPORS, UNC119.

Chapter 1 Supplementary Table S1. List of analysed genes.

Supplementary Table S2

A		DISCARDED VARIANTS				SELECTED VARIANTS						
PATIENT	NUMBER OF VARIANTS PER PATIENT	INTRONIC	SYNONIM	UTR	LOW PATHOGENICITY PREDICTION	PATHOGENIC	HIGH PATHOGENICITY PREDICTION	SPLICING VARIANTS	FRAMESHIFT VARIANTS	STOP CODON LOSS	DELETION	STOP CODON GAIN
5	40	21	12	1	5							1
9	49	24	13	4	7			1				
71	42	20	12	2	6		1		1			
98	51	24	17	3	7							
2	48	18	15	7	8							
16	44	16	14	2	10				2			
19S	48	23	14	4	6		1					
20	50	18	18	3	10				1			
22	38	15	14	3	5	1						
37	45	20	13	4	6	1			1			
39	48	24	15	3	6							
42	47	20	16	2	7		1		1			
43	45	19	15	1	8				2			
48	48	22	14	3	8				1			
64	44	21	11	1	8	1			2			
66	47	20	12	3	7		1		4			
69	50	22	15	4	7		1		1			
70	41	15	13	3	8	1			1			
79	43	16	12	5	7				3			
80	46	18	16	3	7				2			
85	44	19	15	3	6				1			
90	42	15	14	3	9						1	
99	44	20	15	2	5				2			
101	47	23	13	4	4	1			2			
102	45	20	11	4	5	1			4			
105	45	20	16	3	5					1		
113	44	19	13	3	7				2			
133	47	17	16	4	8			1	1			
134	46	23	11	5	6	1						
135	41	15	14	2	9		1					
146	49	23	14	4	6			1	1			
148	43	17	19	2	5							
157	49	19	18	4	6	1	1					

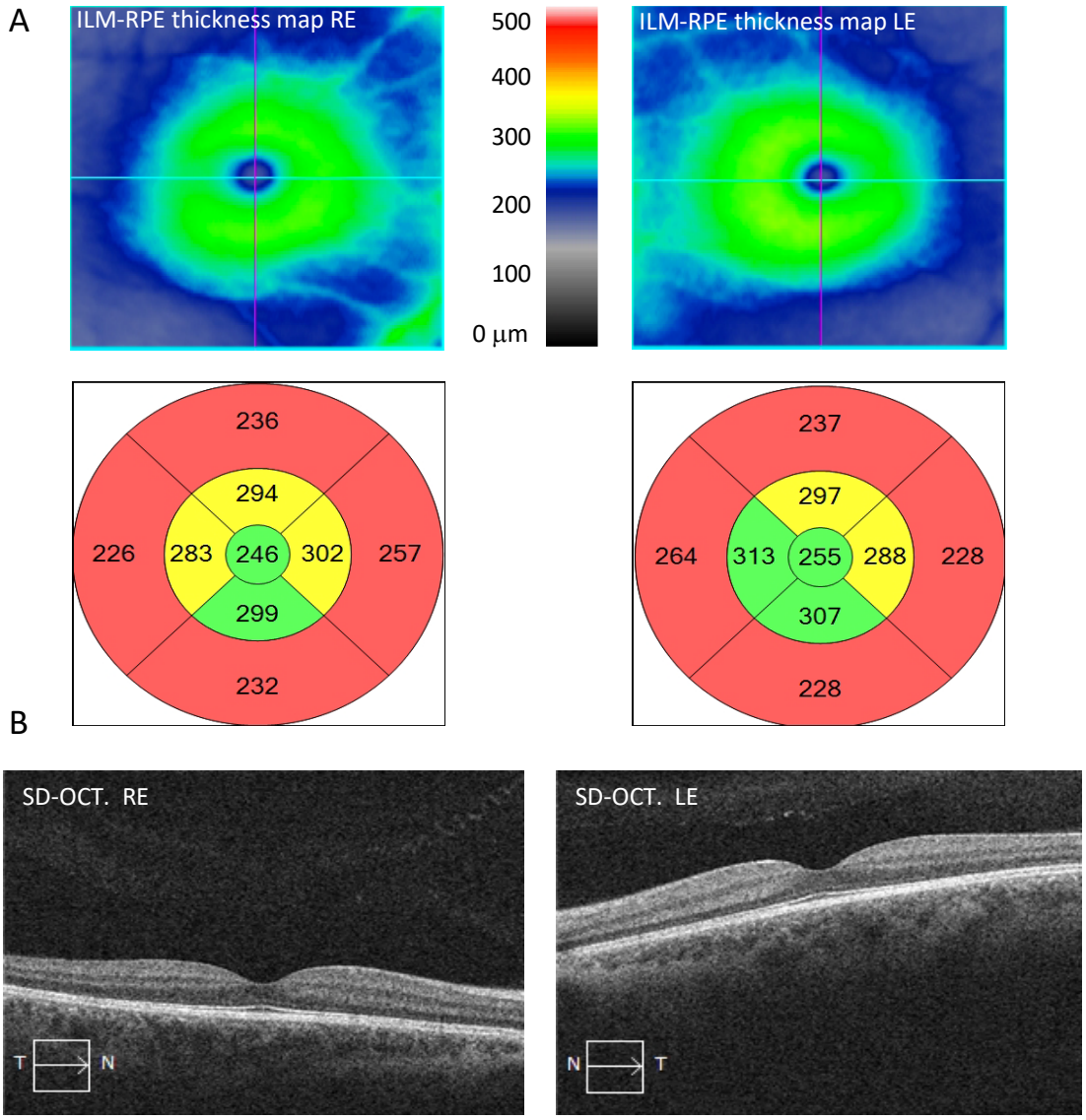
B

MUTATION	PATIENT	GENE	VARIANT TYPE	SANGER SEQUENCED	CONFIRMED	FAMILY SEGREGATION	CONCLUSION
c.1625C>G	5	RP1	stop codon gain	Yes	Yes		Confirmed control
c.937-1G>T	9	RHO	splicing variant	Yes	Yes		Confirmed control
c.259C>G	71	RHO	high pathogenicity prediction	Yes	Yes		Confirmed control
c.415delC	71	KLHL7	frameshift variant	No			False positive in other patient
c.324delA	16	SEMA4A	frameshift variant	Yes	No		False positive
c.650delG	16	AIPL1	frameshift variant	Yes	No		False positive
c.797C>T	19S	PRPH2	high pathogenicity prediction	Yes	Yes	Correct	Causative variant
c.415delC	20	KLHL7	frameshift variant	Yes	No		False positive
c.3260C>T	22	SNRNP200	pathogenic	Yes	Yes	Correct	Causative variant
c.3260C>T	37	SNRNP200	pathogenic	Yes	Yes	Correct	Causative variant
c.415delC	37	KLHL7	frameshift variant	No			False positive in other patient
c.415delC	42	KLHL7	frameshift variant	No			False positive in other patient
c.4555T>C	42	RIMS1	high pathogenicity prediction	Yes	Yes	Incorrect	Not causative variant
c.1596delT	43	RPE65	frameshift variant	No			False positive in other patient
c.1670delT	43	PROM1	frameshift variant	Yes	No		False positive
c.650delG	48	AIPL1	frameshift variant	No			False positive in other patient
c.1596delT	64	RPE65	frameshift variant	Yes	No		False positive
c.1670delT	64	PROM1	frameshift variant	No			False positive in other patient
c.3260C>T	64	SNRNP200	pathogenic	Yes	Yes	Incomplete penetrance	Causative variant
c.1596delT	66	RPE65	frameshift variant	No			False positive in other patient
c.1670delT	66	PROM1	frameshift variant	No			False positive in other patient
c.2088delT	66	SNRNP200	frameshift variant	Yes	No		False positive
c.415delC	66	KLHL7	frameshift variant	No			False positive in other patient
c.2044C>T	66	SEMA4A	high pathogenicity prediction	Yes	Yes	Incorrect	Not causative variant
c.2835A>C	69	PRPF8	high pathogenicity prediction	Yes	Yes	Incorrect	Not causative variant
c.324delA	69	SEMA4A	frameshift variant	No			False positive in other patient
c.324delA	70	SEMA4A	frameshift variant	No			False positive in other patient
c.149C>T	70	GUCA1A	pathogenic	Yes	Yes	Incorrect	Not causative variant
c.2088delT	79	SNRNP200	frameshift variant	No			False positive in other patient
c.324delA	79	SEMA4A	frameshift variant	No			False positive in other patient
c.650delG	79	AIPL1	frameshift variant	No			False positive in other patient
c.2088delT	80	SNRNP200	frameshift variant	No			False positive in other patient
c.324delA	80	SEMA4A	frameshift variant	No			False positive in other patient
c.324delA	85	SEMA4A	frameshift variant	No			False positive in other patient

c.6974_6994del	90	PRPF8	deletion	Yes	Yes	Correct	Causative variant
c.2088delT	99	SNRNP200	frameshift variant	No			False positive in other patient
c.650delG	99	AIPL1	frameshift variant	No			False positive in other patient
c.1336delA	101	PRPF3	frameshift variant	Yes	No		False positive
c.1596delT	101	RPE65	frameshift variant	No			False positive in other patient
c.3260C>T	101	SNRNP200	pathogenic	Yes	Yes	Correct	Causative variant
c.1336delA	102	PRPF3	frameshift variant	No			False positive in other patient
c.1596delT	102	RPE65	frameshift variant	No			False positive in other patient
c.1670delT	102	PROM1	frameshift variant	No			False positive in other patient
c.193delA	102	CA4	frameshift variant	Yes	No		False positive
c.3260C>T	102	SNRNP200	Pathogenic	Yes	Yes	Incomplete penetrance	Causative variant
c.1045T>C	105	RHO	codon stop loss	Yes	Yes	n/a	Causative variant
c.2088delT	113	SNRNP200	frameshift variant	No			False positive in other patient
c.6945delG	113	PRPF8	frameshift variant	Yes	Yes	Correct	Causative variant
c.937-1G>T	133	RHO	splicing variant	Yes	Yes	Correct	Causative variant
c.324delA	133	SEMA4A	frameshift variant	No			False positive in other patient
c.3260C>T	134	SNRNP200	pathogenic	Yes	Yes	Correct	Causative variant
c.568G>A	135	RHO	high pathogenicity prediction	Yes	Yes	n/a	Causative variant
c.666insG	146	RIMS1	frameshift variant	Yes	No		False positive
c.937-1G>T	146	RHO	splicing variant	Yes	Yes	Correct	Causative variant
c.1961G>T	157	SNRNP200	high pathogenicity prediction	Yes	Yes	n/a	Not causative variant
c.3260C>T	157	SNRNP200	pathogenic	Yes	Yes	n/a	Causative variant

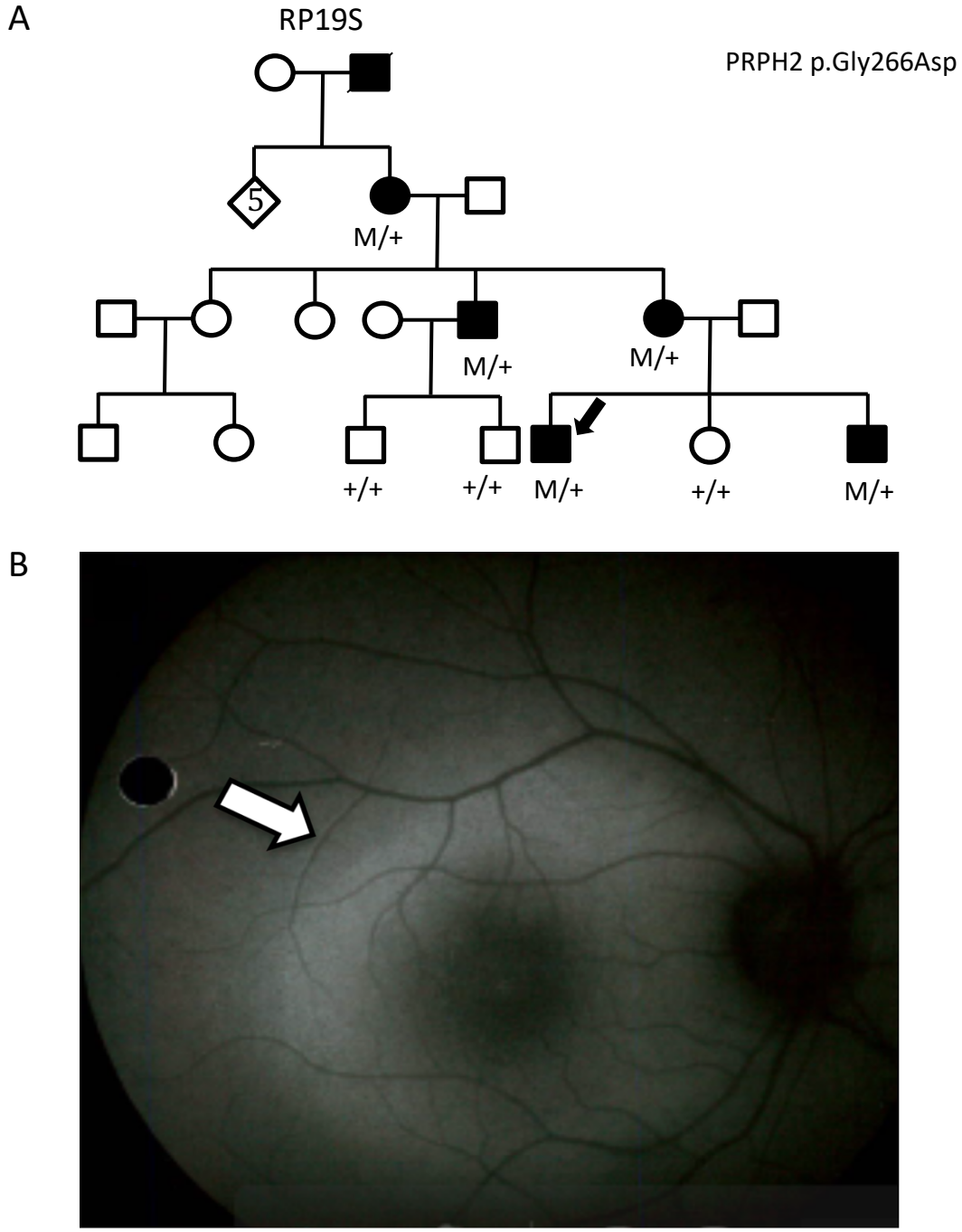
Chapter 1 Supplementary Table S2. Variant identification process in each patient analysed. (**A**) Classification of all variants detected in each patient. (**B**) Selection of variants likely involved in adIRD as determined by previous studies or by *in silico* predictors. Only those variants confirmed by Sanger were submitted to segregation analysis (for selection criteria used see Supplementary Figure 18.).

Supplementary Figure S1



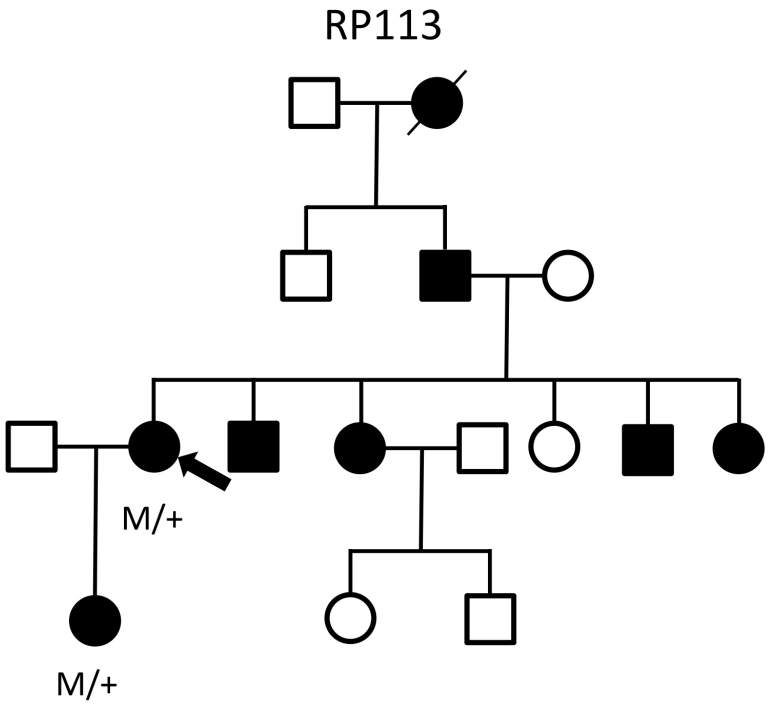
Chapter 1 Supplementary Figure S1. Optical coherence tomography (OCT). The c.937-1G>T mutation in RHO was found in a 21 year-old patient (indicated by an arrowhead in family tree in Figure 2C), prior to clinical diagnosis. Visual fields, fundoscopy and autofluorescence were normal. On OCT we can see what could be an early sign of RP: the thinning of the macula at the 6mm ring (red colour) (**A**). No disruption of the external limiting membrane or the photoreceptor layer was observed. No macular oedema or epiretinal membrane were seen in the OCT (**B**). Abbreviations: ILM-RPE: inner limiting membrane-retinal pigment epithelium; LE: left eye; RE: right eye; SD-OCT: spectral domain optical coherence tomography.

Supplementary Figure S2

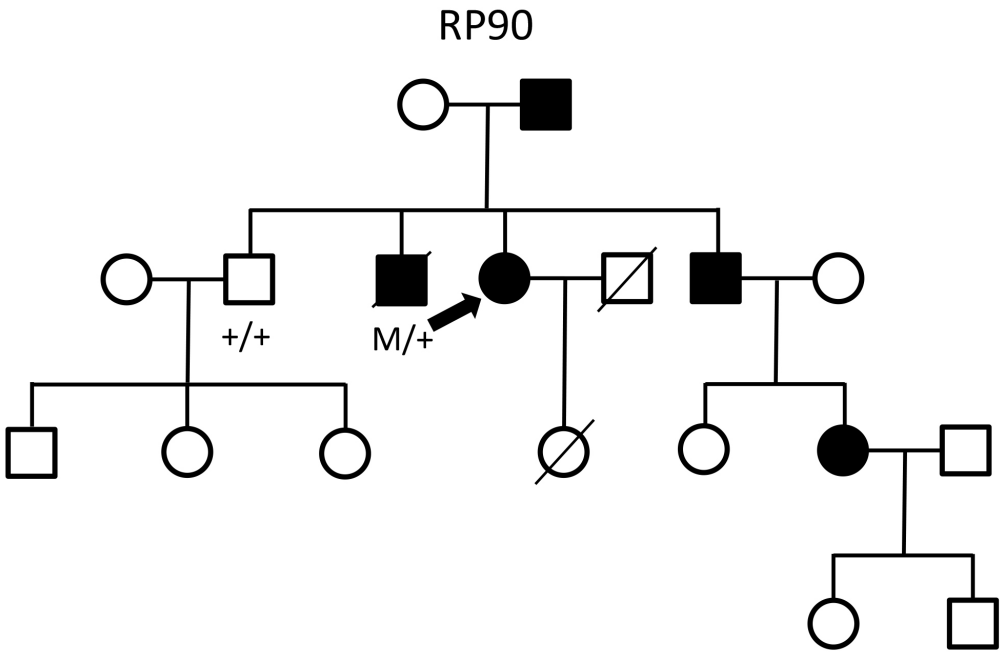


Chapter 1 Supplementary Figure S2. Autofluorescence retinography and family tree for proband RP19S. Family tree. Genotypes are annotated as M/-+(heterozygote); or +/+(wild type). Arrow indicate proband (**A**). Autofluorescence examination of the eye fundus shows a hyperautofluorescent ring in the macula (open arrow) (**B**).

Supplementary Figure S3

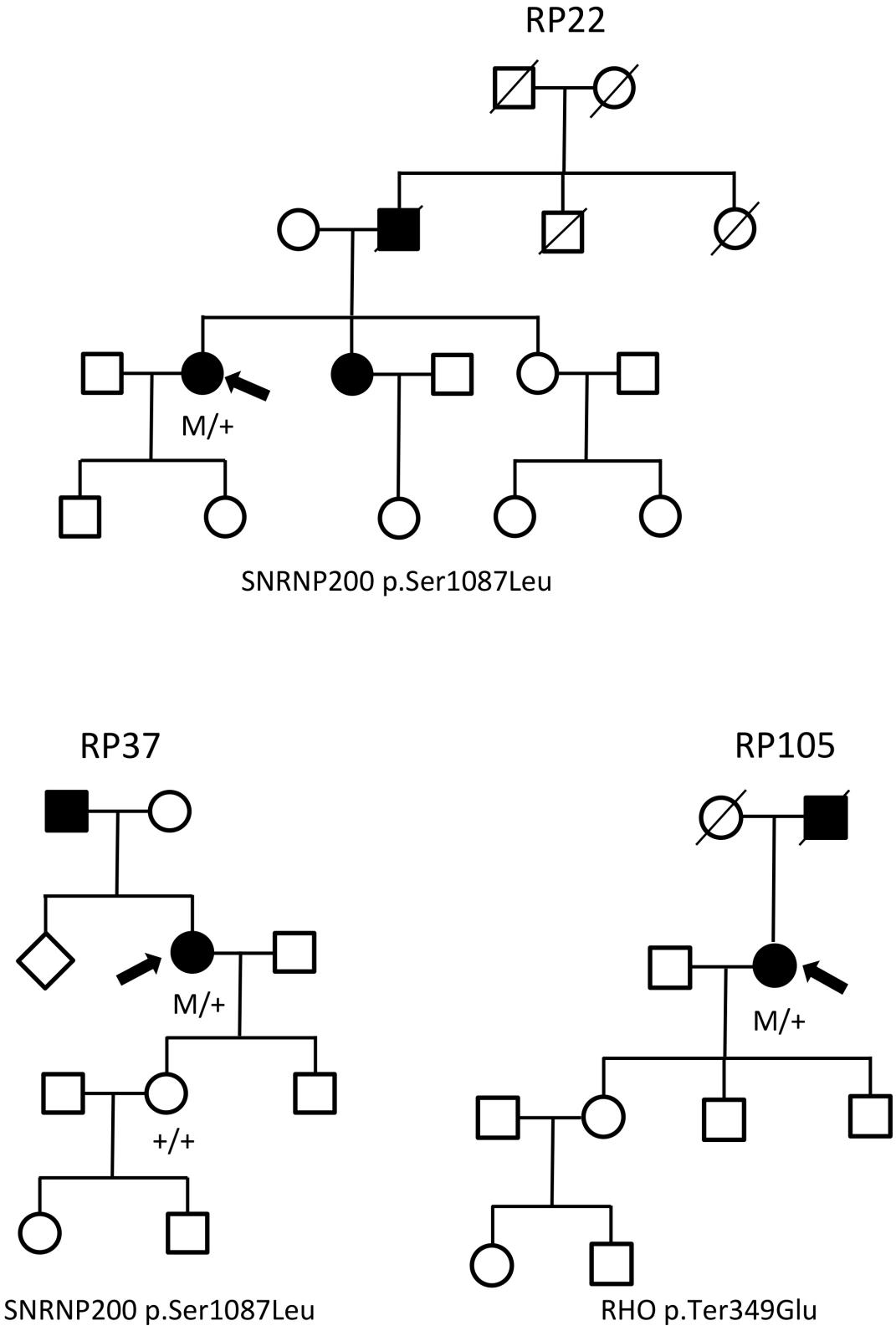


PRPF8 p.Leu2315Leufs*2336Aspext*21



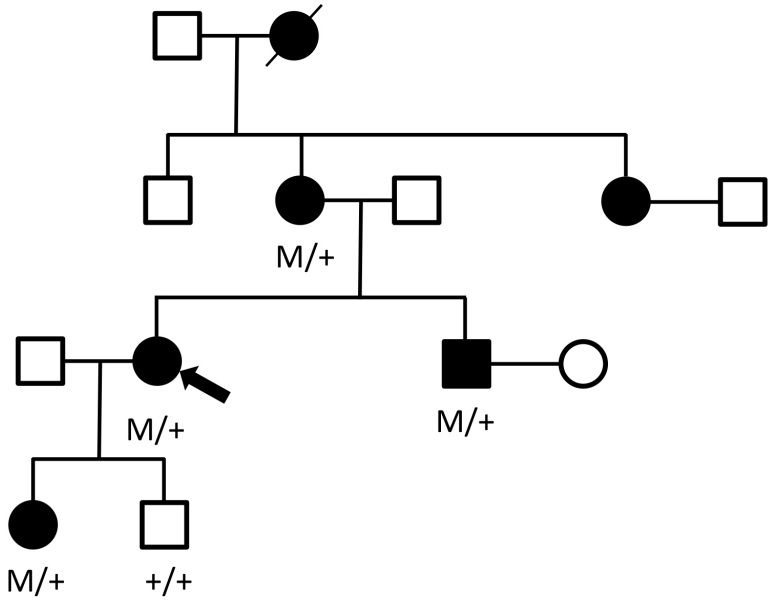
PRPF8 p.Val2325_Glu2330del

Supplementary Figure S3 (continued)



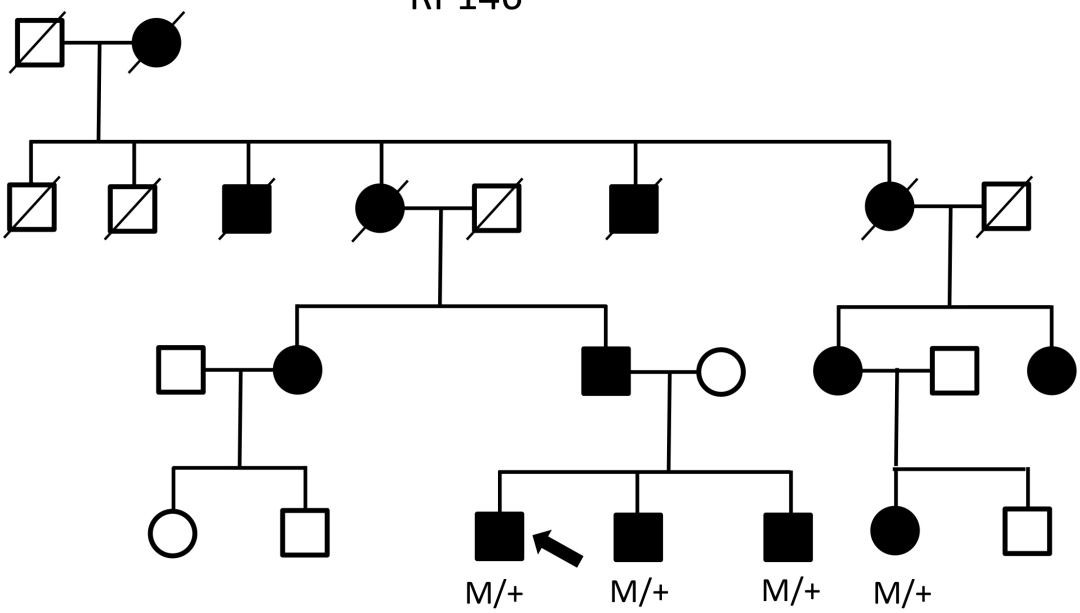
Supplementary Figure S3 (continued)

RP101



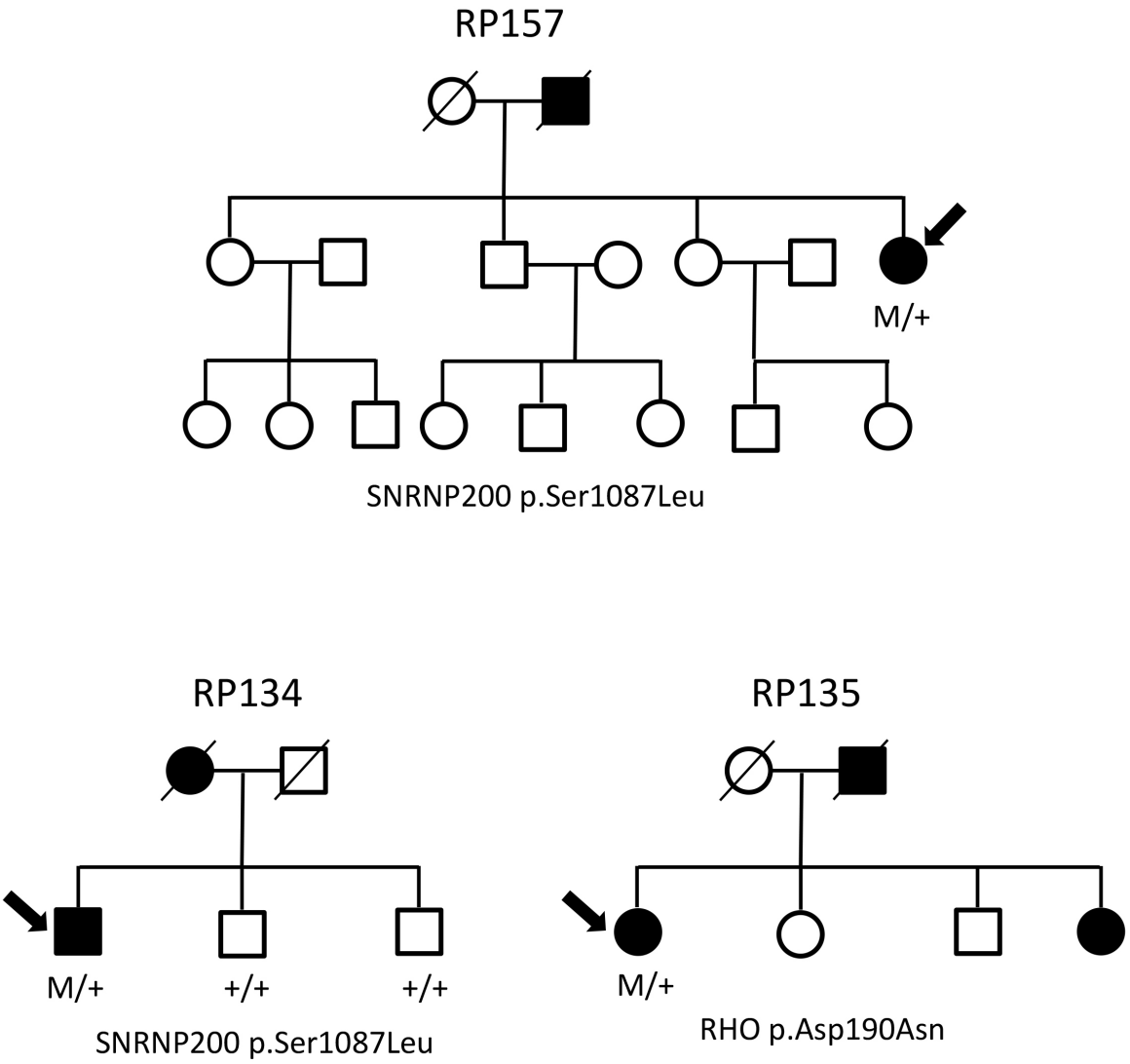
SNRNP200 p.Ser1087Leu

RP146



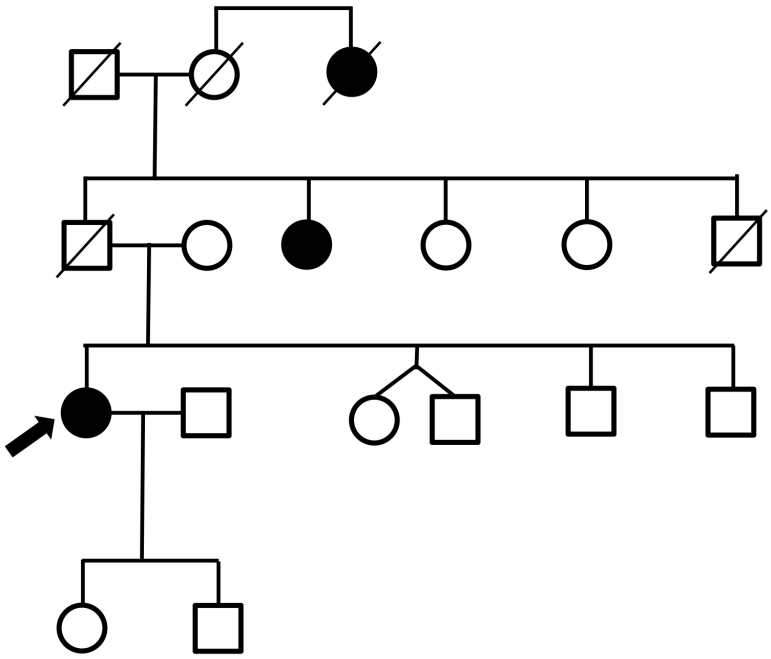
RHO c.937-1G>T

Supplementary Figure S3 (continued)

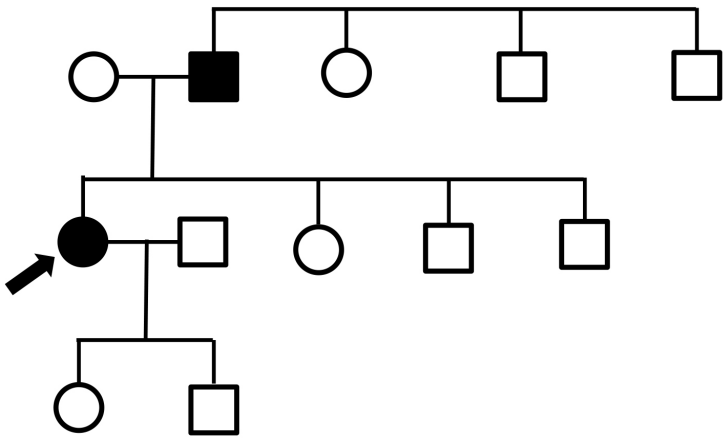


Supplementary Figure S3 (continued)

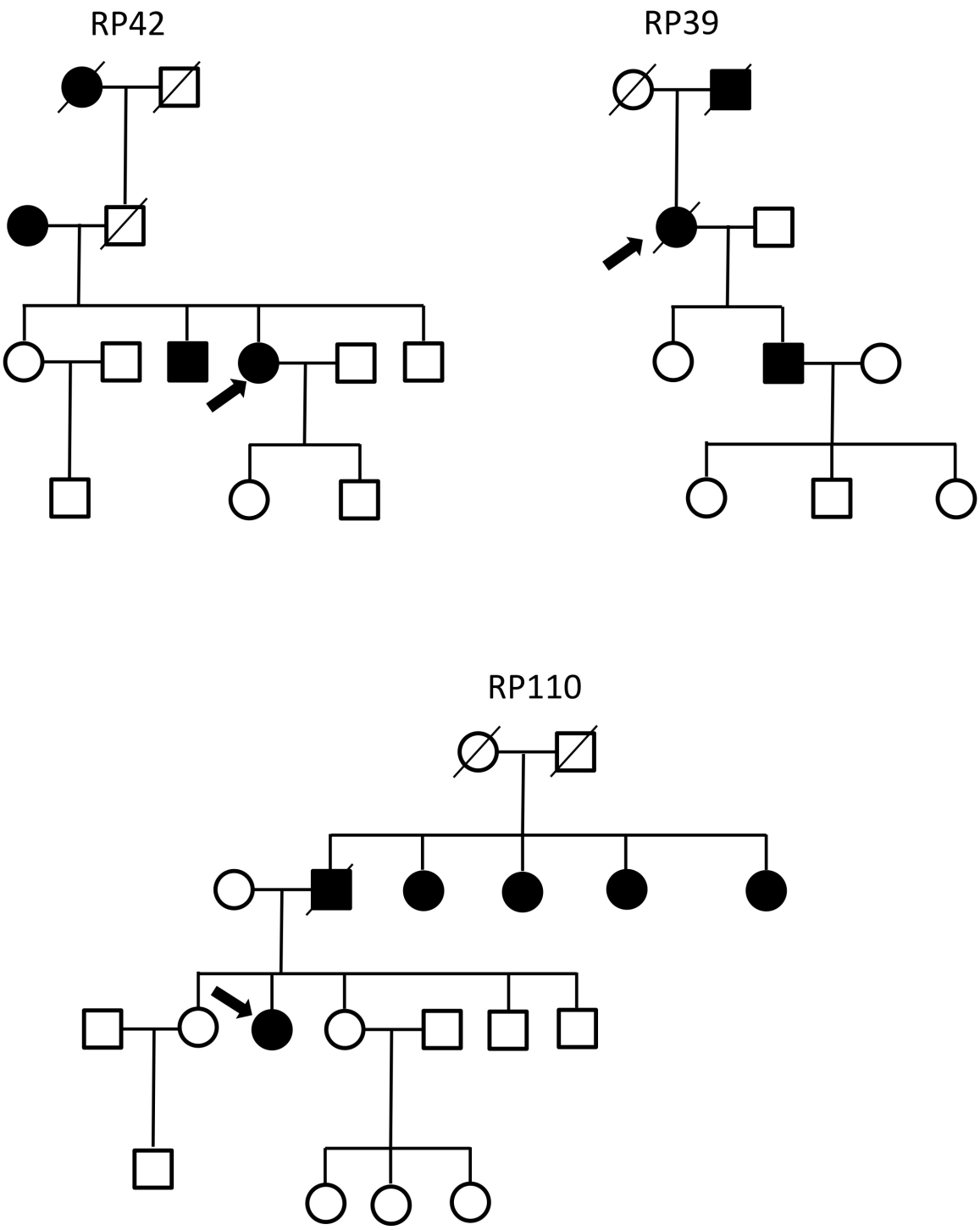
RP16



RP20

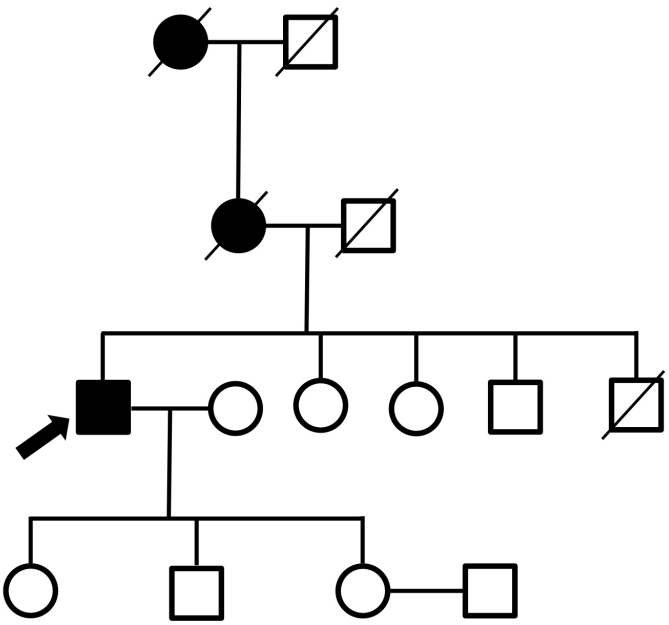


Supplementary Figure S3 (continued)

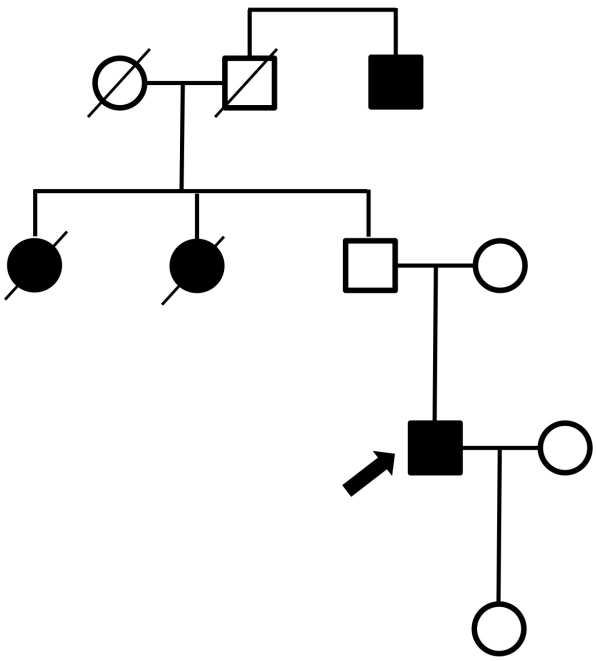


Supplementary Figure S3 (continued)

RP43

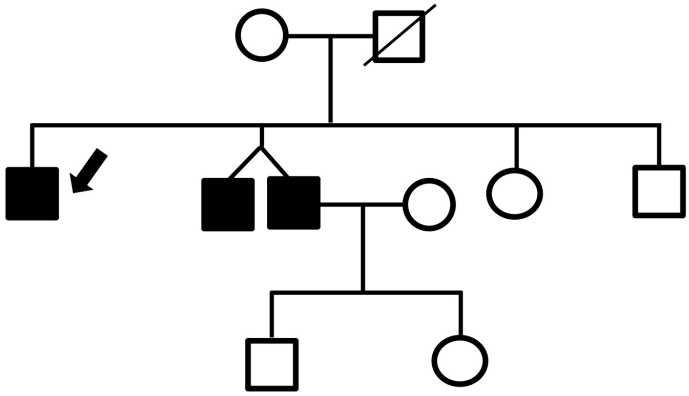


RP48

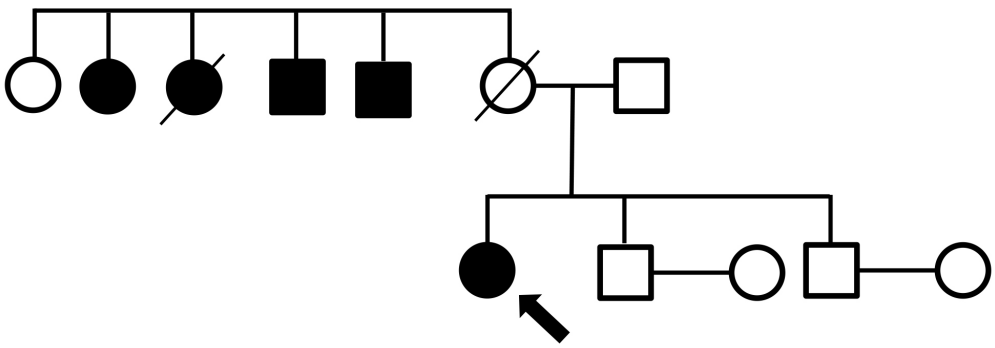


Supplementary Figure S3 (continued)

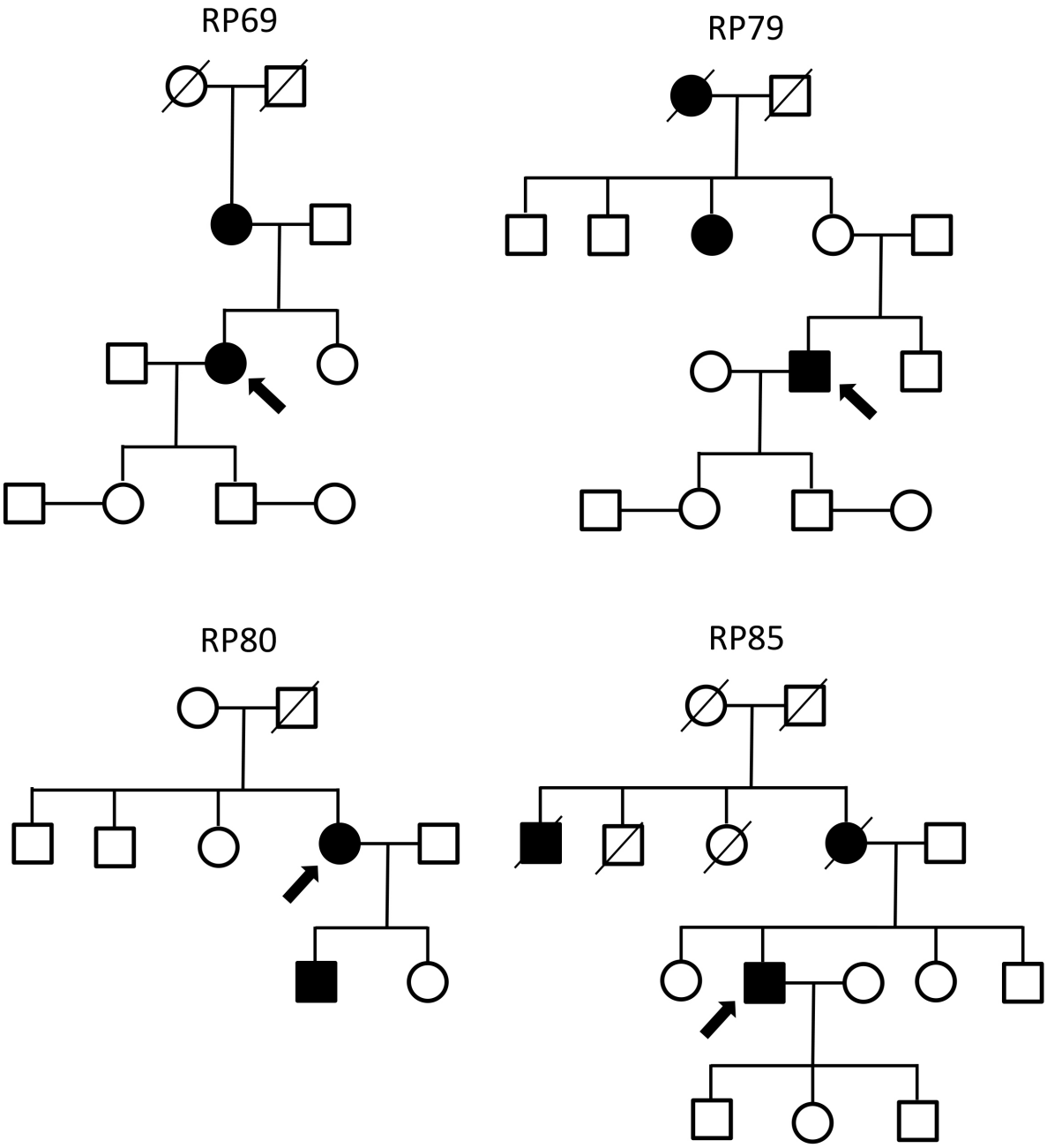
RP27



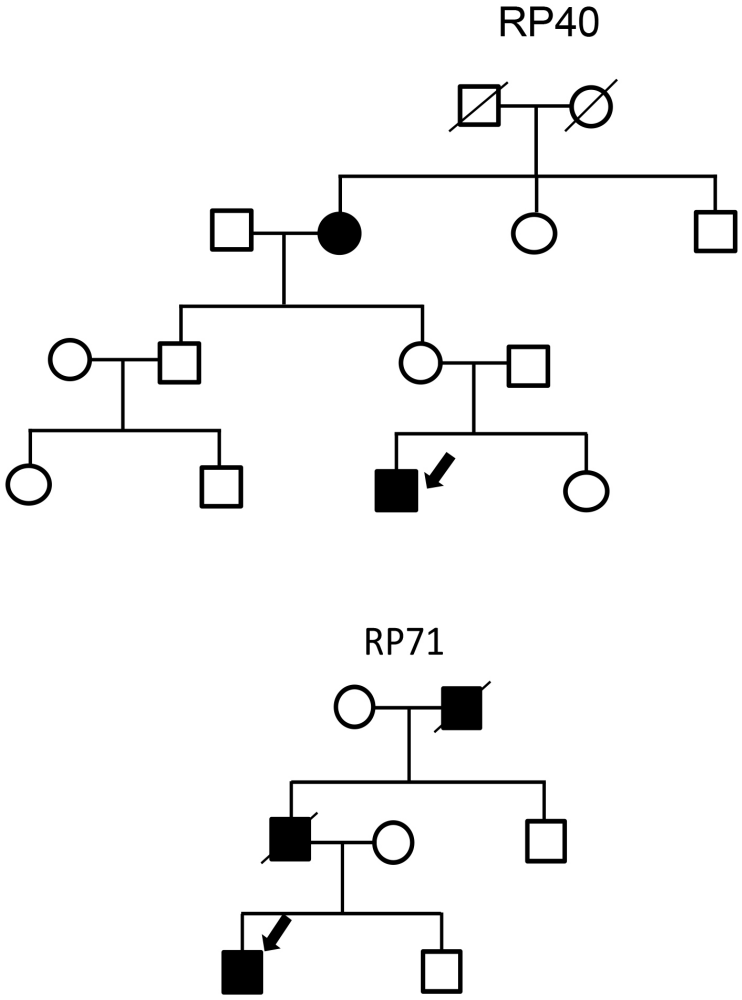
RP66



Supplementary Figure S3 (continued)



Supplementary Figure S3 (continued)



Supplementary Figure S3. Family trees for all probands recruited in the study, excluding those described in Figure 2 and Sup. Fig. 1. Genotypes are annotated as M/+ (heterozygote); or +/+ (wild type). Arrow indicates the proband.

CHAPTER 2

A new approach based on targeted pooled DNA sequencing identifies novel mutations in patients with Inherited Retinal Dystrophies.

Ezquerra-Inchausti M., Anasagasti A., Barandika O., Garai-Aramburu G., Galdós M., Lopez de Munain A.Irigoyen C., Ruiz-Ederra J.

Scientific Reports 18;8(1):15457. doi: 10.1038/s41598-018-33810-3 (2018)

INTRODUCTION

Inherited retinal dystrophies (IRDs) are a group of heterogeneous diseases responsible for different clinically distinctive phenotypes. The most common IRD is retinitis pigmentosa (RP) with a prevalence of 1 in 3,500 people. RP starts with night blindness and is followed by progressive loss of peripheral vision, leading to loss of central vision and blindness in most advanced cases. Although RP is clinically distinct from other IRDs, advanced stage of RP can be difficult to distinguish from other IRDs, including cone-rod or macular dystrophies[172]. Moreover, in some cases, clinical manifestations can differ among members of the same family. IRDs can be inherited in different traits including autosomal dominant (adRP), autosomal recessive (arRP) or X-linked (XlRP). The rate of inheritance has varied across populations studied. To date, over 250 genes have been related to various IRDs and some of them are responsible for the different phenotypes observed [173] (https://sph.uth.edu/retnet/sum-dis.htm, 3 July 2017).

Since the publication of the first draft of the human genome in 2001[174,175], we have seen an unprecedented flourishing of sequencing technologies that provide genomic information in an accurate, fast and cost-efficient way. Methods of massive parallel sequencing such as targeted NGS and WES are the most widely used methods for the diagnosis of IRD. These methods have contributed to an exponential reduction in time and costs for the execution of the sequencing[49,176]. Nevertheless, the use of whole genome sequencing for diagnostic purposes is limited, mainly by the amount of data generated, which demands high degree of expertise in terms of big data handling and interpretation of the results, and these factors complicate its transfer to the clinicians and to the patients. Comprehensive sequencing of the coding regions of all genes, WES, is more affordable, but still has high technical requirements that are an obstacle to its use as a diagnostic method in routine clinical practice. A more practical approach for clinical diagnosis may consist of an initial genetic screening of a subset of genes associated with a phenotype using targeted NGS, followed by a second more extensive genome analysis, such as WES[176], and the analysis of the copy number variations (CNVs)[172], for challenging cases for which the first strategy fails to indicate any genetic explanation.

In this study, we sequenced 316 genes associated with IRDs including several syndromic retinopathies. In order to simplify the sequencing process and to reduce the costs associated with individual labelling of DNA samples, we have developed a mutation detection approach based on targeted NGS in combination with high resolution melting (HRM) analysis. NGS was performed using pools of 16 DNA samples per pool, and identification of the sample/s carrying the mutation/s was performed using HRM analysis in individual samples, which allowed us to link mutations found in the pooled DNA samples to the DNA from individual patients. We sequenced samples from a total of 143 unrelated patients and 17 controls, 5 of which corresponded to samples from patients with IRD characterized by a third party laboratory. Information regarding mutations in these five controls was not revealed to us until completion of our analysis, to further test the sensitivity of our method in an objective way.

For those samples with negative results after the sequencing process, we used multiplex ligation-dependent probe amplification (MLPA) method for CNV analysis. After combining our sequencing strategy with MLPA, we were able to conclusively identify mutations in 45 patients, meaning that a genetic diagnosis rate was obtained in 31.5% of cases.

MATERIALS AND METHODS

Study subjects

IRD patients were clinically diagnosed by the Ophthalmology Service at Donostia University Hospital, San Sebastian, Spain. Most patients studied had been given a diagnosis of retinitis pigmentosa, though a few patients with an undetermined inherited retinal dystrophy (IRD) were also included, based on pedigrees and clinical criteria. The inclusion criteria used were night blindness, peripheral visual field loss, pigmentary deposits resembling bone spicules, retinal vessels attenuation, optic disc pallor and reduced rod and cone response amplitudes and a delay in their timing in the electroretinogram[14]. A total of 143 probands were selected. In addition, samples from 17 patients were included as characterized control patients. This control group was composed of 12/17 samples selected from our cohort of IRD patients with mutations identified in previous studies[88,95,177] and a further 5 control samples from IRD patients characterized by a third party laboratory, (those for which we were blinded to information regarding mutations until we had completed our analysis). Family pedigrees were generated from information obtained from probands. All procedures performed in studies involving human participants received approval from the ethical standards of the Clinical Research Ethics Committee of the Basque Country, Spain (CEIC-E) and were in accordance with the 2013 Helsinki declaration or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Human sample collection

High molecular weight DNA was extracted from blood samples from RP patients and their available family members. Total DNA from samples was extracted and isolated with the AutoGenFlex Star instrument (AutoGen, Holliston, MA, USA) using the FlexiGene DNA Kit (Qiagen, Hilden, Germany) following the manufacturer´s instructions. DNA concentrations were measured on the Qubit fluorometer using Quant-iT PicoGreen reagent (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Equimolar amounts of DNA samples were pooled (100ng/ul per sample). For a detailed description of the procedure see[88].

Pooled sequencing

In order to assess the sensitivity and cost-effectiveness of our method we performed a first experiment to compare the yield obtained after sequencing pools with increasing number of DNA samples and we estimated the differences in costs involved in individual *vs*. pooled sequencing. All pools were made up from samples from carriers of low-frequency variants, which corresponded to either causal, variants of uncertain significance (VUS) or non-pathogenic variants identified in previous studies[88,95,177] A total of 13 control samples were used in 3 sets of pools, with 4, 8 and 16 control samples in each. Of these control samples, 9 carried pathogenic variants (one provided by a third party laboratory), while 7 carried low frequency variants with a minor allele frequency (MAF) <0.003, and therefore we used these 7 samples both as controls and as test samples. Samples were prepared as follows: An initial pool of 4 samples was generated. This pool was used to generate the 3 pools, adding 0, 4 or 8 more samples to generate the pools with 4, 8 and 16 samples, respectively (Figure 21A and Supplementary Table S1A).

In order to further test the sensitivity of our method and to detect possible differences in the sequencing yield, inherent to each sequencing run, we conducted a complementary experiment. For this, we used a different set of controls, all from carriers of low-frequency, non-disease causing variants or individuals with recessive phenotypes with disease causing mutations present in only one allele. In this case, out of 143 patients analysed, a total of 136 test samples were interrogated: 16/136 corresponded to carriers of a total of 21 previously detected non disease causing variants with low MAF (<0.003) and were, therefore, used as both control and test samples (Supplementary Table S1B). 53/136 samples corresponded to patients that had been interrogated previously with negative results, and 67/136 corresponded to new samples interrogated in this study for the first time. As additional controls we used four samples from carriers of disease causing mutations provided by a third party laboratory (for which we were blinded to mutation-related information until after our analysis) and 4 controls from our cohort were used in the last 2 pools. For this experiment, patients were divided into 7 pools with 16 samples each. Control samples were distributed among each pool such as that each pool contained at least 2 control samples, and 4/9 pools had also control from a third party laboratory (Figure 21B). Finally, the rest 28 patients were analysed subsequently in 2 different pools. Two previously characterized patients were introduced in each pool as positive controls (Figure 21C).

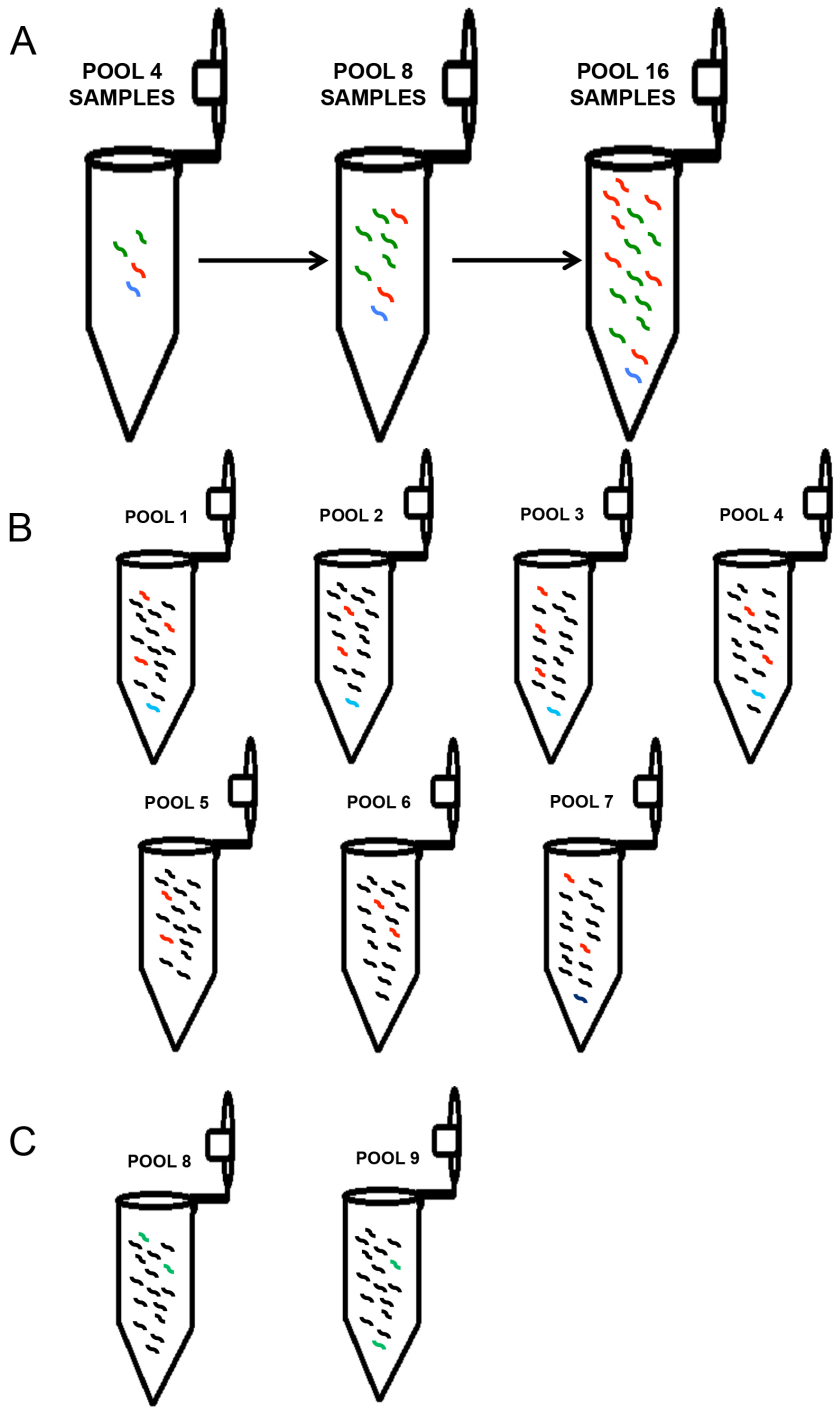


Figure 21: Schematic representation of sample preparation in two sets of experiments. A. DNA was pooled in groups of 4, 8 or 16. B. Seven pools with 16 samples each were prepared. In both cases lines represent DNA from 1 patient. Green and blue lines correspond to samples from patients previously characterized by our group (green) or by a third laboratory (blue). Red lines correspond to samples from unsolved patients, carriers of at least 1 low frequency variant (MAF<0,003), and therefore were used both as control and test samples. Black lines correspond to new samples or without variants with MAF<0,003 or no variants found before. Information from variants used as positive controls is described in Supplementary Table S1

Amplicon Library preparation

Ion AmpliSeq Library Preparation Kit v2.0 (Thermo Fisher Scientific) was used to construct an amplicon library from genomic target regions with a maximum read length of approximately 200 base pairs (average length, 142 bp) for shotgun sequencing on an Ion Proton system (Thermo Fisher Scientific). Briefly, target genomic regions were amplified by simple PCR using Ion Ampliseq primer pools and 10 ng of each DNA samples.

Sequencing Analysis.

Ion Proton Sequencing.

NGS was carried out on the Ion Proton system (Thermo Fisher Scientific). Briefly, enriched ion sphere particles (ISPs) were annealed with the sequencing primer and mixed with the sequencing polymerase from the Ion PGM_200 Sequencing Kit (Thermo Fisher Scientific). Then, the polymerase-bound and primer-activated ISPs were loaded into the previously checked and washed Ion PI Chips (Life Technologies) and having planned the run on the Ion Proton System software, chips were subjected to 500 cycles of sequencing with the standard nucleotide flow order. Signal processing and base calling of data generated from the Ion Proton runs were performed with the Ion Torrent platform-specific analysis software (Torrent Suite version 4.0).

Variant calling.

Using the Ion Reporter software, we performed the variant calling. First of all, GRCh37/hg19 was used as reference genome and alignment was performed against a bed file containing all regions corresponding to 316 genes sequenced. A key aspect in our mutation detection pipeline was to take into consideration the dilution effect of each variant due to our pooled sequencing approach. Therefore, we used the pipeline provided by the ion reporter program for the detection of somatic mutations with minor modifications. We used a somatic mutation detection approach, since this is the most suited for the detection of variants represented in very low frequency (1 in 32 alleles, in the lowest case). The only modification to the default parameters provided by the ion reporter program (5.0 version) consisted on the switch of 10 parameters within the Variant Filtering section in Parameters tab. All

parameters are described in detail in Supplementary Table S2. Finally, a Variant Caller File (VCF) was generated.

Genotyping by high resolution melting (HRM) analysis

Likely disease causing variants from each pool of 16 samples were selected from the VCF. Specific primers were designed to perform an HRM analysis generating amplicons ranging between 250 to 330 bp in length, in order to cover the mutation position. HRM analysis was used to identify which sample/s among 16 in the pool carried the mutation. We followed the methodology described in [88], with minor modification. Briefly, PCR amplification and HRM were performed in a single run on a 7900HT Fast Real-Time PCR System in 384-well plates (Applied Biosystems), each plate contained individual samples (in triplicates) from the 16 probands of the pool in which the variant was detected. We analysed up to 7 different variants in parallel in a single run. After HRM run, the analysis of post amplification fluorescent melting curves was performed using the HRM V2.0.1 software (ThermoFisher Scientific). Melting curves were normalized and difference plots were generated to compare the samples. Only samples showing a different melting curve (Figure 22) were Sanger sequenced.

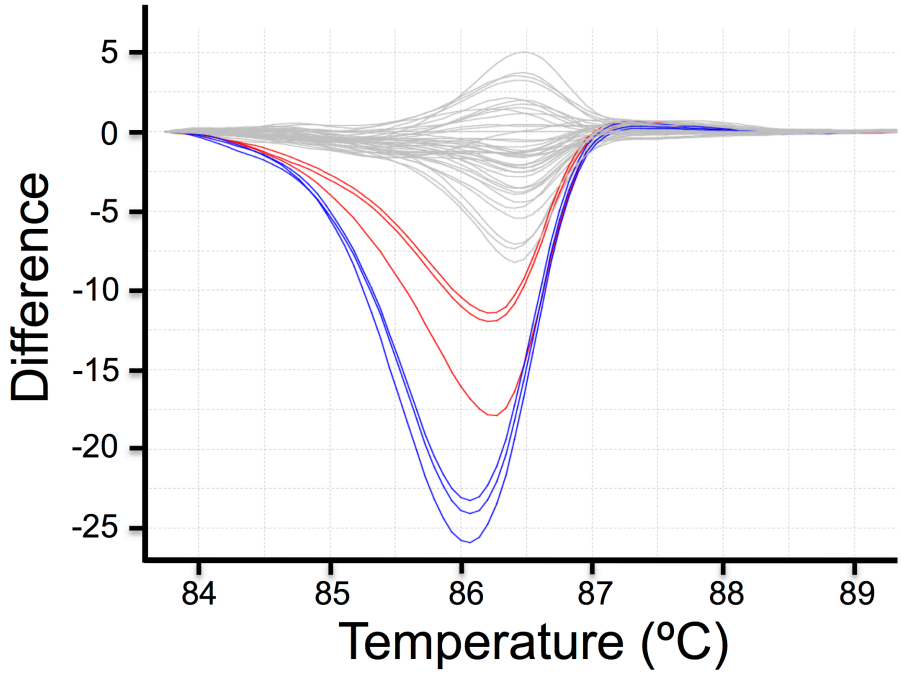


Figure 22 : HRM analysis of TULP1 gene. Difference plot shows c.1495+1G>C mutation in TULP1 gene, with 2 out of 16 samples that clearly differ from the non-carrier samples (grey lines). Sanger sequencing confirmed the presence of the mutation c.1495+1G>C in two patients, one in heterozygosis (blue lines) and the other one in homozygosis (red lines). Note that samples are in triplicates.

Sanger sequencing

Sanger sequencing was used to confirm those mutations detected by NGS and for co-segregation analysis using a 16-capillary ABI 3130xl platform (Applied Biosystems, Foster City, CA, USA) according to manufacturer´s protocol. Sequences were analysed and compared with wild-type samples and a reference sequences using BioEdit software (Ibis Biosciences, Carlsbad, CA, USA) and Ensembl and NCBI databases.

Relevant variant priorization and pathogenicity score

In order to determine genomic variants of relevance, we selected the potential disease causing variants according to the following pre-established criteria:

Variants previously reported as pathogenic.

Variants with a MAF < 0.001 for dominant genes or MAF < 0.003 for recessive genes obtained from genome aggregation database (gnomAD).

Novel Splicing variants and loss-of-function variants such as nonsense mutations, frameshift deletions or insertions.

Previously reported missense variants with pathogenicity scores assessed by *in silico* predictive software.

Novel missense variants predicted to be damaging by *in-silico* predictive software (as mentioned below).

Presence for all candidate variants was checked using the Spanish Variant Server Database (CSVS), (http://csvs.babelomics.org/)[145]. For dominant variants, only those absent from this database were considered further. With regard to recessive variants, only those variants with a MAF lower than 0.003 and only present in heterozygosis were considered further.

Multiplex Ligation-dependent Probe Amplification assay (MLPA)

MLPA was used to search for genomic copy number variations in 32 patients without causative mutations found after sequencing of 316 IRD genes. We selected 9 genes with high prevalence of reported rearrangements[53,178,179].

Patients with a dominant inheritance pattern were analysed using MLPA Retinitis Salsa® Probemix (P235). This probemix contains *PRPF31*, *RHO*, *RP1* and *IMPDH1* genes.

Patients with heterozygotic mutations in *USH2A* genes or *EYS* were also analysed for CNVs, in search of the second mutated allele within these genes (Salsa® Mixes P361/2 and P328, respectively).

In addition, patients with an X-linked inheritance pattern, clinically diagnosed with Choroideremia or families with only males affected, were analysed for *RP2*, *RPGR* and *CHM* genes (Salsa® probemix P366).

MLPA reactions were run according to the manufacturer's general recommendations (MRC-Holland, Amsterdam, Holland) as previously described[180]. The MLPA reaction products were separated by capillary electrophoresis on Abi Prism 3130XL Analyzer (Applied Biosystems) and the results obtained were analysed by GeneMapper software (Thermo Fisher Scientific).

Pathogenicity predictive software

SIFT (http://www.sift.bii.a-star.edu.sg).

Polyphen2 (http://www.genetics.bwh.harvard.edu/pph2/).

PROVEAN (http://provean.jcvi.org/seq_submit.php)[142].

GVGD (agvgd.iarc.fr/agvgd_input_php)[143].

MutationTaster (www.mutationtaster.org)[144].

Web sources

Ensembl, http://www.ensembl.org/

NCBI, http://www.ncbi.nlm.nih.gov/

Polyphen-2, http://www.genetics.bwh.harvard.edu/pph2/

RetNet, http://www.sph.uth.tmc.edu/Retnet/

SIFT, http://www.sift.bii.a-star.edu.sg/

SNPnexus, http://www.snp-nexus.org/

The Human Genome Variation Society (HGVS), http://www.hgvs.org/

1000 Genomes, http://www.1000genomes.org/_ENREF_48

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

Babelomics, http://csvs.babelomics.org

GnomAD browser, http://gnomad.broadinstitute.org/

HGMD, http://hgmd.biobase-international.com/hgmd/pro/all.php/

RESULTS

Targeted Sequencing

A total of 316 genes (Supplementary Table S3) divided into 7222 amplicons were analysed. A total of 2864 and 3350 genetic variants were found in the 4 and 8 sample pools, respectively, while 3997 +/-58 variants found in the 9 pools with 16 samples. Mean and median read depth obtained per sample were 196X and 193X, respectively. Less than 3.4% of targeted regions were covered less than 30X per pool, which we established as the cut off.

Sensitivity

In order to assess the sensitivity of our method we performed two independent experiments. In the first experiment, we included a set of 3 pools all containing an increasing number of control samples prepared from DNA from 16 patients (see methodology section and Figure 21 for a more detailed description). Each control sample carried at least one mutation that had been previously validated by Sanger sequencing. As a result, previously characterized mutations from all control samples were identified in the first set of samples, regardless of the size of the pool.

Following our method, one would expect a relative level of coverage of 1/32 in heterozygous variants and 2/32 in one homozygous or in two heterozygous variants. However, we found that the number did not fit exactly to these values when analysing variants among solved patients. Thus, in heterozygous variants the relative coverage ranged between 0.56 to 1.54/32 with 5 outliers with relative coverage of 1.75/32, 1.88/32, 1.99/32, 1.93/32 and 2/32, values more suggestive of mutations present in two alleles rather than in one. With respect to variants expected to be in two alleles (in homozygosis in one patient or in heterozygosis in two patients), the relative coverage ranged between 1.5-2.3/32. In this case we found 4 outliers with relative levels of coverage as low as 1.25/32 (2 cases), or as high as 2.98/32 and 3.13/32. In all cases with a higher relative coverage, in relation with the number of alleles found, all the pool was Sanger sequenced individually, in order to test for the presence of another allele with that variant and we found that there were no more alleles with the mutation among the pool.

Moreover, we tested 9 SNPs with higher MAFs in order to assess if the relative level of coverage was the same in the case of having more alleles with a specific SNP within the pool. All 16 samples from the pool in which the SNP was found, were directly Sanger sequenced. Similarly to what we observed in the candidate variants, we found some variability between expected *vs.* sequenced SNPs, with a slight mismatch of the variants present according to expected values (Table 3).

rs	Relative number of reads		Total relative number of reads	Expected number of alleles	Confirmed alleles according to zygosity		confirmed alleles
	WT	Mut			Het	Hom	
rs17821448	2109	1123	3232	11.1	7	4	11
rs1801555	3070	706	3776	6	3	2	7
rs11373	5041	1958	6999	9	7	0	7
rs61749605	966	368	1334	8.8	2	2	6
rs1801574	2489	401	2890	4	4	0	4
rs4916685	1016	686	1702	13	7	3	13
rs6666652	2281	178	2459	2.5	2	1	4
rs624851	1207	585	1792	10.5	9	0	9
rs17403955	928	103	1031	3	2	0	2

Table 3 : Relative level of coverage in variants with high MAF. We selected a set of 9 SNPs with relatively high MAF (ranging from 0.1 to 0.37) from the VCF, in order to assess the relative distribution of sequencing reads across samples. In the table are represented the frequency of both WT and mutated alleles (Relative number of reads), and the combination of both (Total relative number of reads). Since our methodology was based on sequencing DNA pools from 16 samples, we expected a relative level of coverage of 1/32 in samples from heterozygous patients and 2/32 in samples from one homozygous or from two heterozygous patients (Expected number of alleles). However, we found some variability between expected and confirmed number of alleles, as identified by Sanger sequencing (Confirmed alleles). Heterozygous and homozygous alleles are represented (confirmed alleles according to zygosity). See discussion section for possible explanations for this variability observed. Abbreviations: Het: heterozygote; Hom: homozygote; MUT: mutated allele; rs: reference SNP ID number; WT: wild-type allele.

Variant Identification

Once we established 16 as the most cost-effective sample size, we sequenced 7 pools of 16 samples/each, including a set of 19 different controls carrying a total of 21 previously detected rare (MAF <0.003), non-causative variants (control variants). All variants selected had a MAF <0.003 for genes mainly associated with a recessive inheritance pattern and were absent from the databases in the case of genes associated with a dominant inheritance

pattern (Supplementary Table S1). As a result, all 21 control variants were also redetected. Afterwards 2 new pools were analysed were 2 control samples (previously characterized samples), were introduced in each pool. In both sets of experiments our methodology yielded 100% sensitivity.

Furthermore, we included five samples from patients with IRD provided by a third party laboratory. As information about mutations within these samples was not initially disclosed to us, we were able to use these samples as an additional way to test the sensitivity of our method. We succeeded in identifying causal mutations in all of the samples. These were: a homozygous mutation c.1645G>T (p.Glu549Ter) in the *BBS1* gene; c.1040C>A (p.Pro347Gln) mutation in the *RHO* gene; c.1703T>A (p.Leu568Ter) mutation in the *CHM* gene; c.2888_2888del (p.Gly963fs) and c.3386G>T (p.Arg1129Leu) mutations in the *ABCA4* gene and a homozygous mutation, c.397C>T (p.His133Tyr) in *MYO7A* gene.

With regard to the 143 unrelated patients analysed, disease causing mutations were found in at least one allele in 76 patients. Nevertheless, since in some patients, mutations were found only in one allele in recessive genes, causal mutations were found in 45 patients, reaching a detection rate of 31.5% (Table 4, Table 5, Supplementary Figure S1). Most of the pathogenic mutations were found in the *USH2A* gene, although in many cases only in one allele without a second mutation, and therefore in these recessive cases, we could not determine the causal mutation. Among all mutations found in characterized patients, 21 were novel, 5 missense, 14 nonsense and frameshift mutations, 1 in-frame mutation and 1 splicing mutation. Novel missense and splicing variant mutations were potentially pathogenic, this being inferred from the score obtained from different *in-silico* tools and the fact that they co-segregated with the disease (Table 6).

FAMILY	GENE	GENE TRANSCRIPT	Allele1			Allele2			Family segregation
			cDNA Change	Protein change	Reference	cDNA Change	Protein change	Reference	
RP1	*EYS*	NM_001142800	c.9405T>A	p.Tyr3135Ter	[181]	c.1830del	p.His610GlnfsTer26	This study	Yes
RP8	*CERKL*	NM_001030311.2	c.847C>T	p.Arg283Ter	[182]	c.847C>T	p.Arg283Ter	[182]	Yes
RP15	*USH2A*	NM_206933	c.12093del	p.Tyr4031Ter	[183]	c.11241C>G	p.Tyr3747Ter	This study	Yes
RP17	*CHM*	NM_000390	c.1272_1273delins CT	p.Gln425Ter	[184]				Yes
RP25	*CRB1*	NM_201253	c.2234C>G	p.Thr745Met	[185]	c.613_619del	p.Ile205AspfsTer13	[186]	Yes
RP27	*RPGR*	NM_001034853	c.2232_2235del	p.Asp744GlufsTer70	This study				Yes
RP30	*RP1*	NM_006269	c.1625C>G	p.Ser542Ter	[187]	c.227T>C	p.Leu76Pro	This study	Yes
RP34	*USH2A*	NM_206933	c.2276G>T	p.Cys759Phe	[188]	c.5278del	p.Asp1760MetfsTer10	[183]	Yes
RP35	*RP1*	NM_006269	c.4804C>T	p.Gln1602Ter	[189]	c.1837dup	p.Thr613AsnfsTer6	This study	Yes
RP49	*EYS*	NM_001142800	c.4045C>T	p.Arg1349Ter	[116]	c.4045C>T	p.Arg1349Ter	[116]	Yes
RP57	*TULP1*	NM_003322	c.1495+1G>C		[190]	c.1495+1G>C		[190]	Yes
RP59	*MYO7A*	NM_000260	c.1200G>T	p.Lys400Asn	[191]	c.5074C>T	p.Gln1692Ter	This study	N/A
RP67	*CERKL*	NM_001030311.2	c.847C>T	p.Arg283Ter	[182]	c.847C>T	p.Arg283Ter	[182]	Yes
RP77	*CNGA1*	NM_001142564	c.301C>T	p.Arg101Ter	[192]	c.1747C>T	p.Arg583Ter	This study	Yes
RP88	*MYO7A*	NM_000260	c.3763del	p.Lys1255ArgfsTer8	[193]	c.6_9dup	p.Leu4AspfsTer39	This study	Yes
RP91	*USH2A*	NM_206933	c.11754G>A	p.Trp3918Ter	[194]	c.3669del	p.Cys1223Ter	This study	Yes
RP106	*EYS*	NM_001142800	c.14C>A	p.Ser5Ter	This study	c.888del	p.Lys296AsnfsTer43	This study	Yes
RP109	*USH2A*	NM_206933	c.1570G>A	p.Ala524Val	This study	c.2276G>T	p.Cys759Phe	[188]	Yes

RP117	*EYS*	NM_001142800	c.4045C>T	p.Arg1349Ter	[116]	c.9405T>A	p.Tyr3135Ter	[181]	Yes
RP138	*CLN3*	NM_001042432	c.1213C>T	p.Arg405Trp	[195]	c.512C>T	p.Ser171Phe	This study	Yes
RP141	*USH2A*	NM_206933	c.2276G>T	p.Cys759Phe	[188]	c.2299del	p.Glu767SerfsTer21	[196]	Yes
RP153	*CERKL*	NM_001030311.2	c.847C>T	p.Arg283Ter	[182]	c.847C>T	p.Arg283Ter	[182]	Yes
RP154	*CNGA3*	NM_001298	c.162_163insT	p.Arg55Ter	This study	c.162_163insT	p.Arg55Ter	This study	Yes
RP165	*ABCA4*	NM_000350	c.3322C>T	p.Arg1108Cys	[197]	c.3322C>T	p.Arg1108Cys	[197]	Yes
RP166	*USH2A*	NM_206933	c.14091del	p.Phe4697LeufsTer2	[198]	c.12093del	p.Tyr4031Ter	[183]	N/A
RP169	*CERKL*	NM_001030311.2	c.847C>T	p.Arg283Ter	[182]	c.356G>A	p.Gly119Asp	[199]	Yes
RP173	*NR2E3*	NM_014249	c.932G>A	p.Arg311Gln	[200]	c.932G>A	p.Arg311Gln	[200]	N/A
RP174	*RGR*	NM_001012720	c.196A>C	p.Ser66Arg	[201]	c.196A>C	p.Ser66Arg	[201]	Yes
RP175	*CNGB3*	NM_019098	c.1148del	p.Thr383IlefsTer13	[202]	c.852+1G>C		This study	Yes
RP176	*CERKL*	NM_001030311.2	c.847C>T	p.Arg283Ter	[182]	c.847C>T	p.Arg283Ter	[182]	Yes
RP180	*USH2A*	NM_206933	c.14565del	p.Asn4856MetfsTer28	This study	c.14565del	p.Asn4856MetfsTer28	This study	Yes
RP182	*PDE6A*	NM_000440	c.1957C>T	p.Arg653Ter	[110]	c.1705C>A	p.Gln569Lys	[203]	Yes
RP185	*CNGA3*	NM_001298	c.1228C>T	p.Arg410Trp	[204]	c.829C>G	p.Arg277Gly	[205]	Yes
RP188	*CNGA3*	NM_001298	c.1228C>T	p.Arg410Trp	[204]	c.1706G>A	p.Arg569His	[205]	N/A
RP193	*ABCA4*	NM_000350	c.4577C>T	p.Thr1526Met	[195,206]	c.3386G>T	p.Arg1129Leu	[207]	N/A
RP196	*BBS1*	NM_024649	c.1220T>G	p.Met390Arg	[208]	c.1220T>G	p.Met390Arg	[208]	Yes
RP200	*CRB1*	NM_201253	c.444_452del	p.Asp148_Asp150del	[209]	c.2843G>A	p.Cys948Tyr	[210]	Yes
RP206	*USH2A*	NM_206933	c.1042_1044del	p.Asn348del	[211]	c.9799T>C	p.Cys3267Arg	[212]	N/A

RP208	*CERKL*	NM_001030311.2	c.847C>T	p.Arg283Ter	[182]	c.847C>T	p.Arg283Ter	[182]	Yes
RP211	*CRB1*	NM_201253	c.493_501del	p.Asp165_Ile167del	This study	c.493_501del	p.Asp165_Ile167del	This study	Yes
RP213	*CRB1*	NM_201253	c.3158T>C	p.Met1053Thr	This study	c.2843G>A	p.Cys948Tyr	[213]	Yes
RP215	*USH2A*	NM_206933	c.2276G>T	p.Cys759Phe	[188]	c.2276G>T	p.Cys759Phe	[188]	N/A
RP217	*ABCC6*	NM_001171	c.3421C>T	p.Arg1141Ter	[214]	c.3421C>T	p.Arg1141Ter	[214]	N/A
RP224	*CNGA1*	NM_001142564	c.301C>T	pArg101Ter	[192]	c.301C>T	pArg101Ter	[192]	N/A
RP40	PRPF31	NM_015629	exons9_13deletion		This study				Yes
RP148	*PRPF8*	*NM_006445*	*c.6835T>G*	*p.Trp2279Gly*	*This study*				*N/A*
RP181	*PRPF31*	*NM_015629*	*c.1165C>T*	*p.Gln389Ter*	*This study*				*N/A*
RP92	*PCDH15*	*NM_001142763/*	*c.733C>T*	*p.Arg245Ter*	[215]	*c.8326G>A*	*p.Gly2776Ser*	*This Study*	*Yes*
	CDH23	*NM_022124*							

Table 4 : Summary of all solved patients. Variants of uncertain significance (VUS) are in italics.

Family	Age at diagnosis	Symptoms at diagnosis	Visual Acuity in LogMAR RE	Visual Acuity in LogMAR LE	Spherical Equivalent RE	Spherical Equivalent LE	Subcapsular Cataract (Yes, No Pseudophakic)	Pale disc	Arteriolar Attenuation	Bone Spicule Retinal Pigment	Epiretinal Membrane	Macular Edema	Visual Fields (grades)	ERG (Electroretinogram)	Syndromic RP	Family member affected (including case study)
RP1	20	Photophobia	2	0.8	-2.2	-2.62	PP	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	1
RP8	17	Nyctalopia	5	5	N/A	N/A	PP	Yes	Yes	Yes	Yes	Yes	No, Low Vision	Ext	No	1
RP15	23	Nyctalopia	0.4	0.3	-0.12	-0.62	PP	Yes	Yes	Yes	Yes	No	4	N/A	No	1
RP17	26	Nyctalopia	0.7	0.1	-6.5	-5.37	No	Yes	Yes	No	Yes	Yes	4	Ext	No	2
RP25	13	Decrease VA	4	4	N/A	N/A	PP	Yes	Yes	Yes	N/A	N/A	No, Low Vision	Ext	No	2
RP27	8	Decrease VA	3	3	0.12	-0.5	PP	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	3
RP30	26	Nyctalopia	0.7	0.7	-5.5	-5.25	Yes	Yes	Yes	Yes	No	No	Altered	Ext	No	1
RP34	37	Visual Field Loss	0.3	0.8	-0.5	-0.62	PP	Yes	Yes	Yes	No	Yes	8	Ext	No	1
RP35	5	Decrease VA	0.8	1,3	0	-0.25	PP	Yes	Yes	Yes	Yes	Yes	Altered	Ext	No	1
RP40	8	Nyctalopia	0	0	0	-0.75	No	Yes	Yes	Yes	No	No	18	Ext	No	2
RP49	16	Nyctalopia	0.4	0.5	0.87	0.75	Yes	Yes	Yes	Yes	Yes	Yes	15	Ext	No	1
RP57	9	Nyctalopia	1.3	4	13	2	PP	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	1
RP59	12	Nyctalopia	0	0	1.625	-1.25	No	Yes	Yes	No	No	No	7	Ext	Usher type 1	1
RP67	50	Decrease VA	N/A	N/A	2	0.75	Yes	Yes	Yes	Yes	No	No	No, Low vision	Ext	No	2
RP77	40	Nyctalopia	0.3	0.2	0.75	0.62	PP	Yes	Yes	Yes	Yes	Yes	4	Ext	No	2
RP88	12	Visual Field Loss	1.3	1	N/A	N/A	PP	Yes	Yes	Yes	No	No	N/A, deafness	Ext	Usher type 1	2
RP91	16	Nyctalopia	0.3	0.4	-1.62	-1.87	Yes	Yes	Yes	Yes	No	No	8	Ext	Usher	1
RP106	45	Nyctalopia	4	4	-8.75	-9.5	Yes	Yes	Yes	Yes	Yes	NO	No, Low Vision	Ext	No	1
RP117	27	Decrease VA	0.5	0.4	1.12	-1.5	No	Yes	Yes	Yes	No	Yes	10	Ext	No	4
RP138	23	Nyctalopia	0	0.1	0	0.5	No	Yes	Yes	Yes	No	No	12	Ext	No	3
RP141	35	Nyctalopia	N/A	N/A	1	1	Yes	Yes	Yes	Yes	No	No	N/A	Ext	No	1
RP153	17	Decrease VA	3	1	-0.5	-0.25	Yes	Yes	Yes	Yes	No	No	No, Low Vision	N/A	No	2
RP154	1	Decrease VA	1	1	3	1	No	No	No	No	No	No	Central Scotoma	N/A	Achrom.	2
RP165	17	Decrease VA	3	3	N/A	N/A	Yes	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	5
RP166	N/A	Nyctalopia	0.2	0.3	-1	-1.75	Yes	Yes	Yes	Yes	No	No	7	Ext	Usher Type 2	1
RP169	31	Nyctalopia	5	4	N/A	N/A	Yes	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	2
RP173	1	Nyctalopia	1	1	-2	-0.25	No	No	No	Yes	No	Yes	No, Low Vision	Ext	No	2
RP174	38	Decrease VA	4	4	-3.37	-0.75	No	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	1
RP175	4	Decrease VA	1	1	-0.75	-0.125	No	No	No	No	No	No	No, Low Vision	*1	Achrom.	2
RP176	22	Decrease VA	0.3	0.4	-0.75	-1.5	Yes	Yes	Yes	Yes	No	No	Central scotoma	Ext	No	1
RP180	38	Nyctalopia	4	4	N/A	N/A	Yes	Yes	Yes	Yes	No	No	No, Low Vision	Ext	Usher Type 2	3

RP109	36	Nyctalopia	0.4	0.3	-0.5	0	Yes	Yes	Yes	No	Yes	Yes	7	Ext	No	1
RP182	10	Nyctalopia	0.05	0.05	-1.75	-1.25	Yes	Yes	Yes	Yes	No	No	5	Ext	No	1
RP185	1	Nystagmus	1.3	1.3	-5.37	-5.37	No	No	No	No	No	No	No, Low Vision	*1	Achrom.	1
RP188	49	Decrease VA	0.8	1	7.3	7.3	No	No	No	No	No	No	No, Low Vision	Ext	No	1
RP193	38	Decrease VA	1	1	2.62	2.61	No	No	No	No	No	No	No, central Scotoma	N/A	No	1
RP196	12	Decrease VA	1	1	-1.12	-2.12	Yes	Yes	Yes	Yes	No	No	4	Ext	No	1
RP200	31	Decrease VA	0.7	3	+0.75	+1.87	No	Yes	Yes	No	No	No	No, Low Vision	Ext	No	1
RP206	26	Nyctalopia	0.7	0.5	-2,5	-2,75	Yes	Yes	Yes	Yes	Yes	No	8	N/A	Usher Type 2	1
RP208	16	Decrease VA	0.2	0.2	2,5	3,25	No	No	No	No	No	No	15	*1	No	1
RP211	23	Decrease VA	0.2	0.3	-2,5	-3,5	No	No	No	No	No	No	Central Scotoma	*1	No	1
RP213	1	Nystagmus	0.001	0.001	N/A	N/A	Yes	Yes	Yes	Yes	N/A	N/A	No, Low Vision	N/A	No	2
RP215	65	Nyctalopia	0.4	0.4	2,625	3,25	Yes	Yes	Yes	Yes	No	No	9	Ext	No	1
RP217	10	Decrease VA	0.5	0.7	-1,75	-3	No	No	No	No	No	No	Altered	N/A	No	3
RP224	47	Decrease VA	0.1	0.2	-0,75	-0,62	Yes	Yes	Yes	Yes	Yes	No in LE	Altered	N/A	No	3

Table 5: Clinical features of characterized patients. Abbreviations; Achrom: Achromatopsia; LE: Left eye; NA: not available; PP: Pseudophakia; RE: Right Eye; VA: Visual Acuity. *ERG not detected either in photopic nor scotopic conditions.

GENE	cDNA CHANGE	PROTEIN CHANGE	TRANSCRIT	POLYPHEN	SIFT	PROVEAN	GVGD	MUT TASTER	HSF	MAF
GPR98	c.853C>T	p.Arg285Cys	NM_032119	1	0	Deleterious (-4,9)	C65	DC (0,99)		0,001
ABCA4	c.2023G>A	p.Val675Ile	NM_000350	1	0	Neutral	C25	DC (0,99)		Absent
PCDH15	c.2576G>A	p.Val861Met	NM_001142763	0,68	0,1	Neutral	C15	DC (0,90)		0,003
GPR143	c.609-1G>T		NM_000273						Decrease 5' aceptor site of exon 5 (79,18>78,96)	Absent
CEP290	c.4250A>G	p.Gln1417Arg	NM_025114	0,99	0,04	Neutral (-1,3)	C35	DC (0,99)		0,001
PCDH15	c.3817C>A	p.Arg1273Ser	NM_001142763	0,97	0	Deleterious (-2,8)	C65	DC (0,94)		Absent
CDH23	c.8326G>A	p.Gly2776Ser	NM_022124	1	0,01	Deleterious (-3,19)	C35	DC (0,99)		Absent
PRPF8	c.6835T>G	p.Trp2279Gly	NM_006445	1	0	Deleterious (-11)	C65	DC (0,99)		Absent
PRPH2	c.734T>C	P.Leu245Pro	NM_000322	1	0	Deleterious (-4,2)	C65	DC (0,99)		Absent
CNGA3	c.868C>T	p.Arg290Cys	NM_001298	0.98	0	Deletereous (-6.98)	C65	DC (0.99)		<0.0001t

CNGB3	**c.852+1C>G**		**NM_019098**						**Alteration of the WT donor site**	**Absent**
CLN3	**c.512C>T**	**p.Ser171Phe**	**NM_001042432**	**0.99**	**0**	**Deletereous (-3.21)**	**C65**	**DC(0.99)**		**Absent**
USH2A	**c.1570G>A**	**p.Ala524Val**	**NM_206933**	**1**	**0**	**Deleterious (-3,19)**	**C55**	**DC (1)**		**Absent**
CRB1	**c.3158T>C**	**p.Met1053Thr**	**NM_201253**			**Deletereous (-4.07)**	**C65**	**DC(0.99)**		**Absent**
RP1	**c.227T>C**	**p.Leu76Pro**	**NM_006269**	**0,9**	**0**	**Deleterious (-6,46)**	**C65**	**DC (0,99)**		**Absent**

Table 6: Summary of novel missense and Splice site mutations found: In-silico pathogenicity was scored using 6 predictors. All variants were absent in a Spanish in-house allele database containing information from 791 unrelated Spanish individuals (Spanish controls) or were present only in heterozygous form in recessive genes. Bold cases indicate mutations found in characterized patients. See Materials and Methods section for detailed information. Abbreviations; DC: disease causing; HSF: human splicing finder; MAF: Minor Allele Frequency.

Regarding the distribution of mutations among our cohort of patients, most findings were found among the following five genes:

USH2A. Mutations within this gene were responsible for most cases of arRP in our cohort. Most of the patients were carriers of biallelic mutations. Compound heterozygous mutations are frequently reported in this gene[183,198]. Five of the mutations found in *USH2A* were novel: c.11241C>G, in patient RP15, c.3669del in patient RP91, c.1570G>A in patient RP109, c.1042_1044del in patient RP206 and c.14565del in patient RP180. Except for patient RP180 and RP215, homozygote carriers of the mutations, the rest of the patients were carriers of mutations in compound heterozygosis with the previously reported pathogenic mutations c.12093del, c.11754G>A, c.2276G>T and c.9799T>C respectively (Table 4).

CERKL. This was the second most commonly mutated gene in our cohort. We characterized 6 patients with the same mutation c.847C>T in this gene. In 5 of the cases it was in homozygosis and in one case it was in compound heterozygosis with c.356G>A mutation. This nonsense mutation is relatively common in Spanish cohorts[95,182].

EYS. This was the third most commonly mutated gene together with *CRB1* in our cohort. Three out of four patients shared mutations, such as RP1 and RP117 with c.9405T>A[181] and RP49 and RP117 with c.4045T>A[116], probably indicating the sharing of a common ancestor. This finding is consistent with previous studies involving Spanish cohorts, in which *EYS* was one of the most commonly mutated genes in recessive retinitis pigmentosa[23,60]. In addition, we found three novel mutations in this gene: two frameshift mutations in compound heterozygosis c.1830del in patient RP1 and c.888del in patient RP106; and a nonsense mutation also in compound heterozygosis c.14C>A, in patient RP106.

RPGR. We were able to detect a novel mutation c.2232_2235del in patient RP27 in the ORF15 region of this gene. Mutations in this region are challenging to amplify due to a large segment of highly repetitive purine-rich sequences[65]. Nevertheless, the high coverage of this region we obtained using our pooled-based approach, allowed us to detect this variant (Supplementary Fig S2).

Variants of Uncertain Significance (VUS)

For the family RP92, two heterozygous variants were observed in *PCDH15* and *CDH23*. Despite the fact that this digenic inheritance pattern has previously been found to be causative of Usher Syndrome[216], and that the variants segregated correctly within our family, there is some controversy with the pathogenicity of this digenism and, as far as we know, the *CDH23* and *PCDH15* digenism has been only reported in one study[16]. Despite cochlear degeneration specific to hair cells was observed in this type of mice, USH mutant mice do not display visual defects. Based on ultrastructural analyses, it has been shown that the USH1 proteins localize at the level of microvilli-like structures, called calyceal processes, which form a collar around the base of photoreceptor outer segments. These structures have only been found in primate and other large mammals, but not in mouse photoreceptor cells[217]. This has led to propose that the absence of these structures in the mouse retina is responsible for the lack of a visual phenotype in mouse models of Usher syndrome. Regardless of this structural difference, we cannot confirm that this digenism is the causative mutation.

In the case of family RP148, a novel missense mutation c.6835T>G was found in *PRPF8* gene. The mutation was predicted to be damaging by at least 5 *in silico* predictors. Nevertheless, given the lack of a complete segregation analysis due to the unavailability of many of the samples required, we were unable to conclude that c.6835T>G is the causal adRP mutation in this family. Similarly, in family RP181, we found a novel nonsense mutation, c.1165C>T, in *PRPF31* gene. However, we were not able to validate this finding in a segregation analysis due to a lack of samples available. In fact, the only family sample we were able to study was a non-affected sister who was also a mutation carrier.

Multiplex Ligation-dependent Probe Amplification (MLPA)

Among the 32 families analysed by this method, we detected a large deletion in the *PRPF31* gene expanding from exon 9 to 13 in family RP40, previously unreported. The deletion was also detected in an affected grandmother and the asymptomatic mother. Confirmation of the deletion region was performed sequencing the deleted DNA fragment (Figure 23).

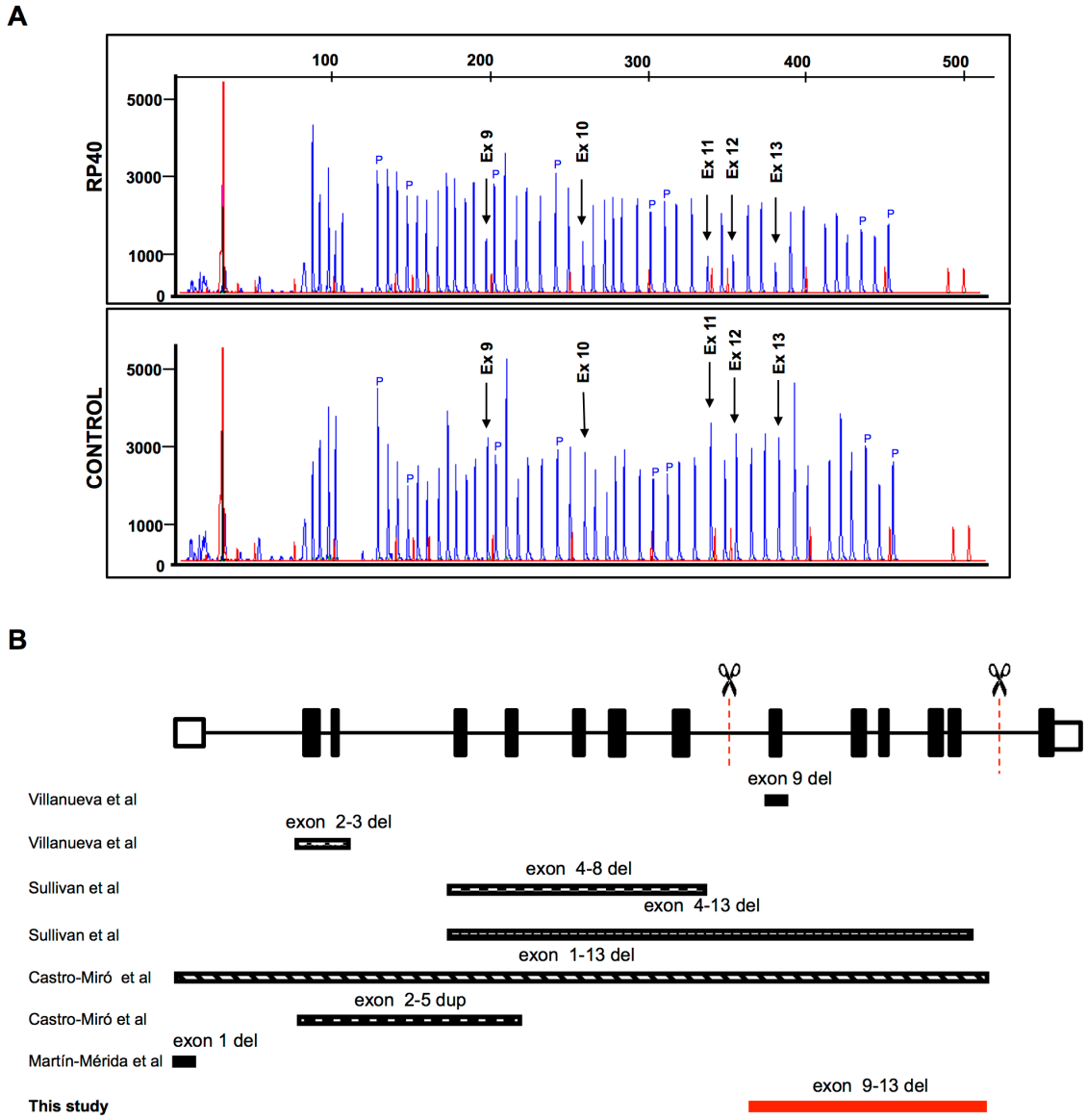


Figure 23: Novel deletion in PRPF31. A. Electropherogram showing a reduced dosage of exons 9-13 (arrows) in patient RP40. B. Schematic representation of PRPF31 deletions and/or duplications described in the literature, and the deletion of exons 9-13 we found in this study, represented by the red bar. Abbreviations: P: control probes; Ex: Exon.

DISCUSSION

In the present work, we have developed a cost-effective method for the diagnosis of IRDs based on pooled genomic DNA targeted NGS, in combination with HRM as a highly sensitive, versatile and affordable genotyping method. Following our methodology, we were able to find the causal mutation in 45 of our patients (31.5%) (Table 4).

Several studies have validated the feasibility of DNA sequencing pools to identify and quantify the genetic variants or single nucleotide polymorphisms (SNPs) in small genomes or small genomic regions of prokaryotes[218]; and single human genes[219,220]. Previous studies tested experimentally the accuracy in re-sequencing pools of strains of highly isogenic D. melanogaster, whose genome had been previously sequenced individually. They showed that the sequenced pool provides a correct estimate of the population allele frequency, enabling the discovery of new SNPs with a low rate of false positives[221].

Regarding clinical applications[222] evaluated the use of pooled DNA sequencing to accurately assess allele frequencies on transmitted and non-transmitted chromosomes in a set of families in an allelic association study[223] combined DNA samples from 1,111 individuals and sequenced 4 genes to identify rare germline variants. The main bottleneck in the use of a pooling strategy for genetic studies is related to the challenges of detecting rare and low-frequency variants reliably, allowing an accurate estimation of MAFs[224]. Moreover, pooled DNA sequencing was applied for the analysis of 3 genes of Gitelman's syndrome using semiconductor NGS in pooled DNA samples from 20 patients[225]. In a more recent study, 72 genes were analysed in pools consisting of samples from 12 individuals[226]. With respect to RP, pooled DNA NGS was used to search for mutations in the *SNRNP200* gene in a cohort of 96 unrelated patients from North America[167]. Pooled DNA sequencing has recently been used for population genetics studies (GWAS), in several different pathologies[227].

Compared to previous studies that limited to the sequencing of a restricted number of genes, this represents the first study based on the pooled sequencing of more than 300 genes. To estimate the reduction in costs derived from the use of our methodology we compared the costs per patient of our pooled method with an individual sequencing approach. The main source of cost savings was related to expenses involved in the preparation of DNA libraries. Specifically, there was a 10.6-fold reduction in sequencing costs with our methodology. Once we added costs associated with the HRM analysis-based genotyping method, the overall reduction in mutation detection/patient was 6.25-fold.

The choice of 16-sample pools was based, not only on terms of sensitivity, but also on the optimal number of samples for further analysis by HRM, which we found to be around 16 in a previous study[88]. One of the main advantages over previous pooled-NGS-based strategies for mutations detection is the genotyping method we used. HRM analysis is significantly

more affordable than other methods including TaqMan probes (Thermo Fisher Scientific) especially if used for a large cohort of patients and/or for a large number of genes[228]; or DNA arrays Sequenom IPLEX (CD Genomics), which requires specific equipment, making the applicability of the methodology highly dependent on the equipment available in each laboratory[228].

In order to test the sensitivity of our method we included a set of positive controls. Five of these positive controls were samples from IRD patients previously diagnosed elsewhere, for whom we only had access to their clinical data, but not to information on the causative mutations. Given that we obtained a sensitivity of 100%, the fact that our detection rate is not as high as in previous studies, ranging from 51 to 66%[229–232], might be explained, at least in part, by the nature of the cohort of patients included in our study, since part of our cohort of patients (60/143) were analysed in previous studies with no results, using a repertoire of different approaches[88,95,177].
Therefore, we believe that the great number of samples analysed in previous studies is the main factor for the relative low yield obtained. A similar observation was recently reported, where they found that the patients who were screened for the first time had a higher pathogenic variant detection rate than the overall rate, suggesting that their cohort was enriched for intractable cases giving a lower detection rate[195].

Another possibility is that the detection rate varies depending on the ethnicity of the individuals analysed[195]. In this regard, they reported a lower rate of homozygous variants detected in individuals of European origin, comparing with other populations, in recessive transmitted diseases[195]. Similarly, we found pathogenic heterozygous mutations in recessive genes in 27 patients, which therefore cannot be regarded as the causal mutation on their own. One possibility is that a fraction of our patients might be bearing large DNA re-arrangements, or mutations in deep intronic regions not covered by our approach, which would act in compound heterozygosis.

One limitation of the approach used in this work was that the relative level of coverage expected in validated variants (1/32 in heterozygous variants and 2/32 in one homozygous or in two heterozygous variants) did not fit exactly to expected values in some cases (see

Results section). This could be due to the fact that there is a pre-amplification step for library preparation. Despite great care was taken for preparing the pools using equimolar amounts of each DNA sample, we cannot discard the possibility of having some samples over or under-represented, offering higher or lower relative values, respectively. This might be reflecting an unequal sample bias, or that all DNAs of each pool were not amplified in all regions, which might be one of the potential explanations for the relative low diagnostic yield. However, we consider this possibility unlikely, considering that we were able to detect all control variants introduced in each pool.

Another limitation of pooled sequencing method is related to the lack of use of multiplex barcodes, which complicates CNV detection using NGS technology[233] .

There is increasing evidence of genomic rearrangements resulting in CNVs responsible for IRDs in several genes including *PRPF31*[178]; *EYS*[179]; *USH2A*[53] and X-linked *RPGR* and *CHM*[25,184]. Several recent studies have emphasized the importance of CNV analysis in IRD cases. For instance, Bujakowska *et al*., 2017[234] found mutations in 5 out of 28 IRD cases in *SNRNP200, PRPF31, EYS* and *OPN1LW* genes. Khateb *et al*., 2016[235]; identified rearrangements in 6 IRD patients out of 60 involving *EYS, MYO7A, NPHP4*, *RPGR* and *CHM*. In the case of the alteration in *CHM,* the deletion included other 6 adjacent genes. Van Cauwenbergh *et al*., 2016[172] identified CNVs in 3 patients out of 57 analysed, with mutations in *USH2A*, *HGSNAT* and *RCBTB1* genes. Interestingly, a recent paper has established a ranking of IRD genes according to genomic features and CNV occurrence. These authors recommend performing routinely a targeted CNV screening in the most prevalent 30 top-ranked IRD genes according to their genomic length[236].

Despite some authors have described the use of read depth methods for pooled multiple sequencing[237], we decided to select a group of 9 genes, most of which known to be prone to CNV formation[236] using MLPA. We analysed several patients with negative results after the sequencing of the 316 IRD genes, and we included some of the genes reported as the main contributors to CNV in different studies, such as *USH2A, EYS, CHM, PRPF31* and *RPGR* [52,55,172,178,234,235].

Using this approach, we were able to diagnose a patient with a deletion expanding from exon 9 to 13 in *PRPF31*. Rearrangements in this gene have been described to account for

around 2.5% in autosomal dominant cases[178]. Although different mutated regions have been described in *PRPF31*, the deletion of exons 9 to 13 has not been described before (Figure 23). The pattern of inheritance in family 40 is suggestive of an autosomal dominant pattern with incomplete penetrance. Segregation analysis was conducted in two family members, revealing the presence of an obligate carrier. Mutations in *PRPF31* have been mostly associated with cases of incomplete penetrance[238–240].

A limitation inherent to the technique employed, which is shared by WES, is the impossibility of finding mutations in deep intronic regions, not covered by the primer design. In this regard, in an attempt to find the second mutant allele, we analysed two commonly reported deep intronic mutations: c.2991+1655A>G in *CEP290*[241] and c.7595-2144A>G in *USH2A* genes[242,243], in patients with heterozygous mutations in those genes. We did not however, find the mutations that were likely causative of the disease within these regions.
Despite limitations inherent to NGS sequencing regarding its performance in repetitive or CG-rich regions of the genome, we were able to detect the mutation c.2232_2235del in ORF15 of the *RPGR* gene, a region regarded as challenging, with a poor sequencing performance, both in panel based NGS and whole exome sequencing[15]. Using our methodology, we were able to detect this mutation among one of the 16 samples of the pool, which further support the validity of our method in terms of sequencing capacity, genotyping and filtering methods.

Regarding the mutations found, *USH2A* represents the most commonly mutated gene within our cohort of patients, with thirteen different mutations found in this gene in nine patients characterized. Among *USH2* genes, *USH2A* is the most commonly mutated gene and it is responsible for approximately 74-90% of USH2 cases[183,188,244]. Mutations in *USH2A*, are responsible for Usher syndrome type 2 and non-syndromic RP[57]. *CERKL, EYS* and *CRB1* are the next most commonly mutated genes in our cohort, which is also in accordance with previous studies[245,246]. In case of mutations in *EYS* genes, high prevalence has also been observed among Spanish population[60], Americans with European origin[23] and among Japanese populations[59].

For those patients for whom we failed to identify putative disease-causing mutations, the use of alternative approaches will hopefully succeed in characterizing their disease, at the molecular level. For instance, WES aimed at the identification of mutations in genes not currently linked to IRDs; CGH arrays for the analysis of CNVs in other genes or regions not covered by our MLPA analysis; or whole genome sequencing to extend the analysis to the 99% of non-coding DNA. Despite being highly dependent on technical support, the use of whole genome sequencing is gaining momentum in clinical practice, and it seems plausible that it will become feasible in a near future, once a robust translational genomics workflow becomes an affordable option both in economic and technical terms, to allow feedback of potentially diagnostic findings to clinicians and research participants[123].

CHAPTER 2 SUPPLEMENTARY INFORMATION

Supplementary Table S1

A

GENE	ZYGOSITY	VARIANT	cDNA	PROTEIN CHANGE
BBS1	hom	chr11:66293652	c.1169T>G	p.Met390Arg
ABCA4	het	chr1:94473807	c.5882G>A	p.Gly1961Glu
RHO	het	chr3:129252450	c.937-1G>T	c.937-1G>T
USH2A	het	chr1:216052143	c.8521T>A	p.Trp2841Arg
RP1	het	chr8:55537560	c.1118C>T	p.Thr373Ile
CERKL	hom	chr2:182423344	c.847C>T	p.Arg283Ter
USH2A	het	chr1:216420460	c.2276G>T	p.Cys759Phe
CERKL	het	chr2:182423344	c.847C>T	p.Arg283Ter
PDE6A	hom	chr5:149263074	c.2053G>A	p.Val685Met
ROM1	het	chr11:62382123	c.868del	p.Gln290LysfsTer26
RHO	het	chr3:129247835	c.259C>G	p.Val87Leu
USH2A	hom	chr1:215847862	c.13388G>A	p.Thr4464Ter
PRPF31	het	chr19:54626832	c.770-1C>T	c.770-1C>T
PRPF8	het	chr17:1554160	c.6945del	p.Asn2316ThrfsTer43
RP2	hem	chrX:46736931	c.1073-9T>A	c.1073-9T>A

B

GENE	ZYGOSITY	VARIANT	cDNA	PROTEIN CHANGE
USH2A	het	chr1:215932085	c.11241C>A	p.Tyr3747Ter
PRPF31	het	chr19:54621969	c.194T>A	p.Met65Lys
PDE6A	het	chr5:149301194	c.933+4C>T	c.933+4C>T
GUCA1A	het	chr6:42141469	c.118C>T	p.Arg40Cys
PRPF3	het	chr1:150325252	c.2071-57ins	c.2071-57ins
USH2A	het	chr1:216420460	c.2279G>T	p.Cys759Phe
ABCA4	het	chr1:94526230	c.2023G>A	p.Val675Ile
ABCA4	het	chr1:94544977	c.1140T>A	p.Asn380Lys
BEST1	het	chr11:61722590	c.164C>T	p.Thr55Met
BBS10	het	chr12:76739848	c.1917C>G	p.Gly639Gly
RLBP1	hom	chr15:89754954	c.684+20C>T	c.684+20C>T
MERTK	hom	chr2:112740597	c.1296+27del	c.1296+27del
SNRNP200	het	chr2:96959129	c.1957C>A	p.Thr654Asn
RHO	het	chr3:129247887	c.311T>A	p.Val104Asp
RHO	het	chr3:129252535	c.1021G>A	p.Glu341Lys
CNGA1	het	chr4:47938971	c.1747C>T	p.Arg583Ter
CNGA1	het	Chr4:47939328	c.1519C>T	p.Ala459Val
CNGA1	het	chr4:47972953	c.165T>C	p.Ser55Ser
CNGA1	het	chr4:47973110	c.8C>T	p.Ser3Phe
PDE6A	het	chr5:149323876	c.367G>T	p.Asp123Tyr
RPGR	het	chrX:38158349	c.1105C>T	p.Arg369Cys

Chapter 2 Supplementary Table S1. **Variants used as positive control in two sets of experiments. A.** pooled DNA with 4, 8 and 16 samples. **B.** 7 pools with 16 samples each. Distribution of control variants among samples is depicted in Figure 21.

Supplementary Table S2

PARAMETER	VALUES
realignment threshold	0
position bias pvalue	0,05
position bias reference fraction	0,05
position bias	0,75
data quality stringency	10
downsample to coverage	7000
snp min cov each strand	4
snp min variant score	3
snp min allele freq	0,01
snp min coverage	10
hotspot strand bias pval	0,01
snp strand bias pval	0,01
indel strand bias pval	1
snp strand bias	0,95
indel min cov each strand	4
indel min varint score	4
indel min allele freq	0,02
indel min coverage	100
indel strand bias	0,9
hotspot min cov each strand	2
hotspot min variant score	3
hotspot min allele freq	0,01
hotspot min coverage	20
hotspot strand bias	0,95
prediction precision	1
outlier probability	0,005
heavy tailed	3
filter unusual predictions	0,3
filter insertion predictions	0,2
filter deletion predictions	0,2
hp max length	8
do snp realignment	false

do mnp realignment	false
indel as hp indel	false
use position bias	false
suppress recalibration	false
SSE probability threshold	1
mnp min cov each strand	4
mnp min variant score	6
mnp min allele freq	0,01
mnp min coverage	100
mnp strand bias	0,95
mnp strand bias pval	0,01

Chapter 2 Supplementary Table S2. Ion reporter workflow parameters.

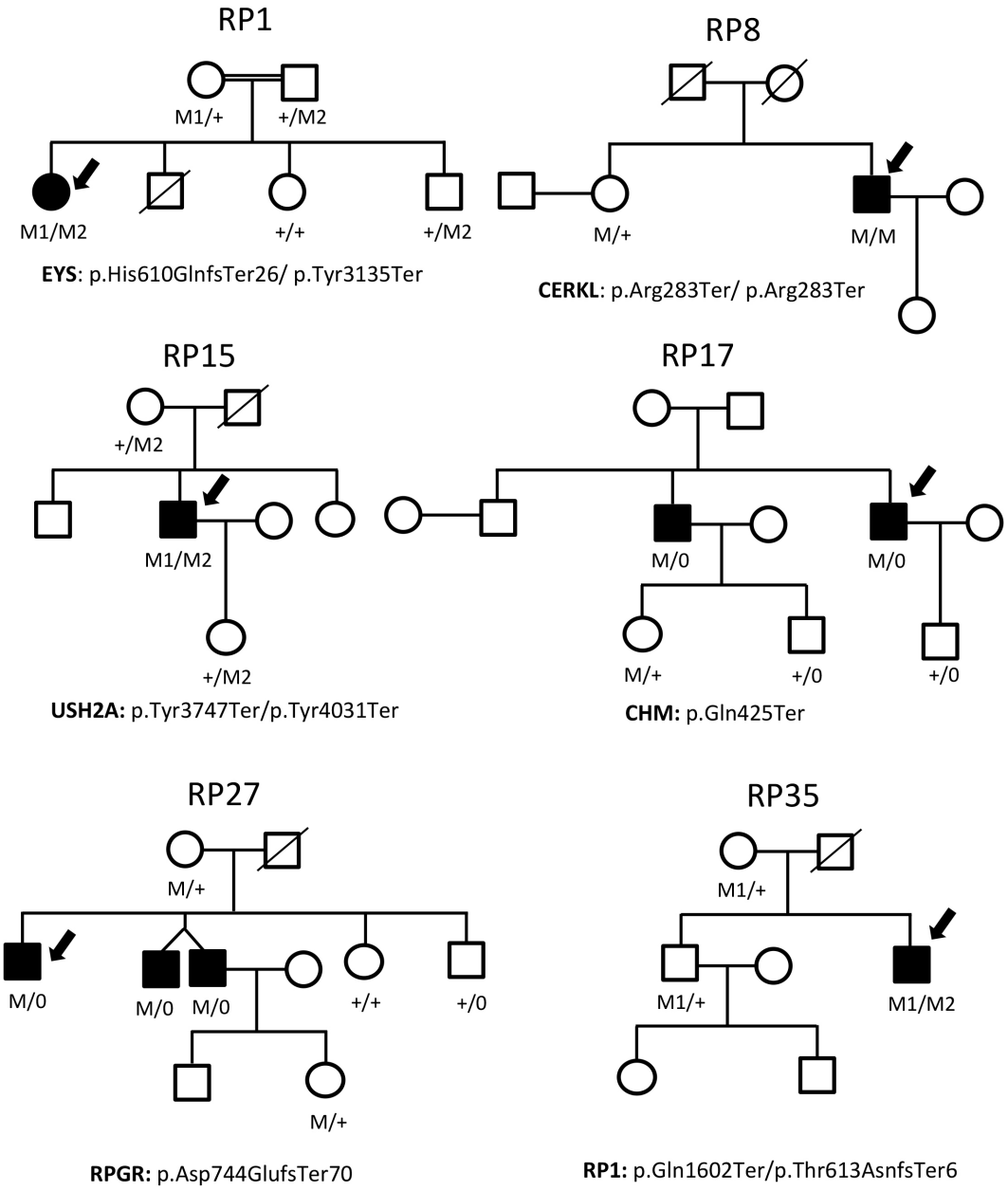
Supplementary Table S3

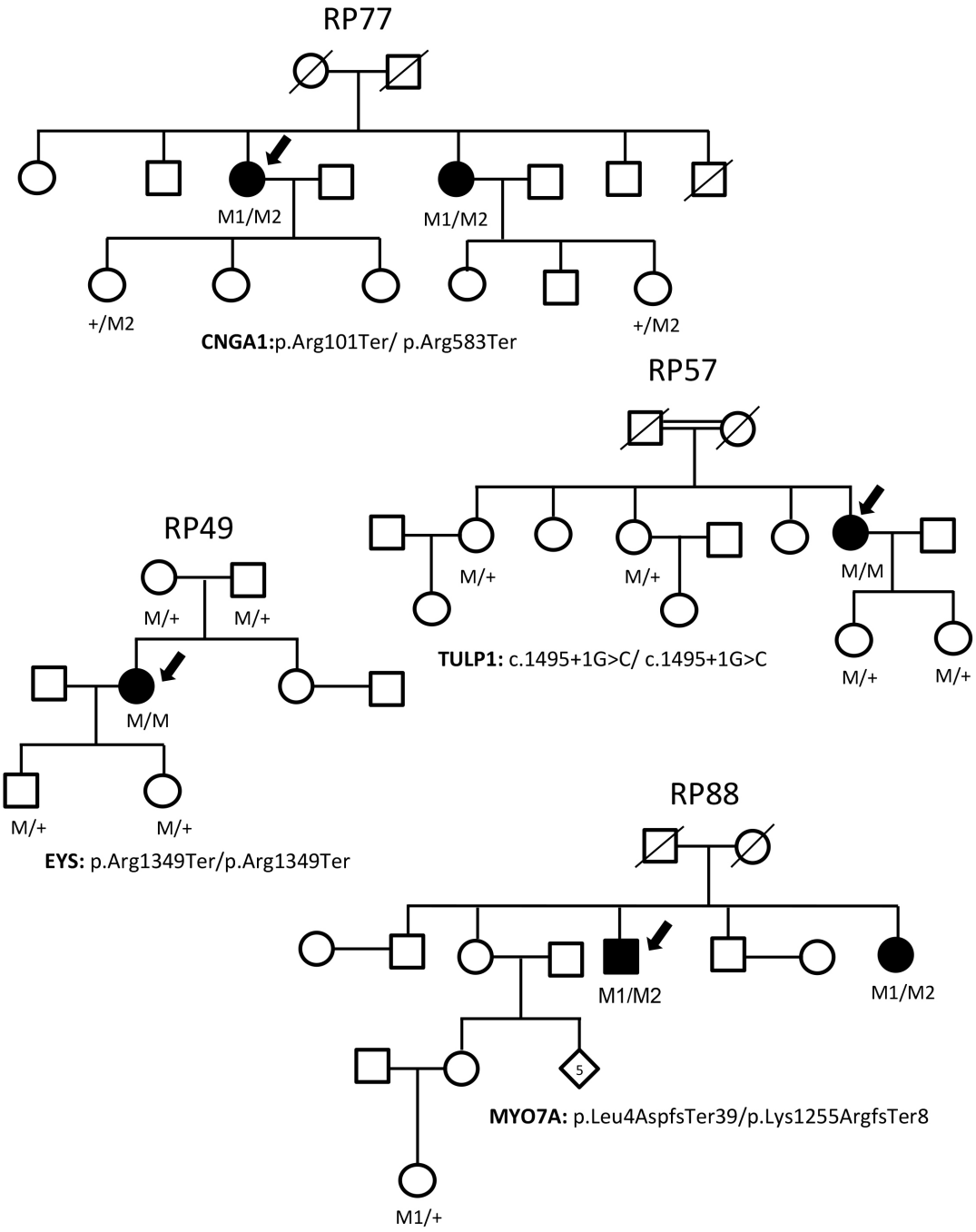
ABCA4, ABCC6, ABHD12, ACBD5, ACO2, ADAM9, ADAMTS10, ADAMTS18, ADAMTSL4, AGK, AHI1, AIPL1, AKR1E2, ALDH1A3, ALMS1, APOA1, ARL6, ATXN7, B3GALTL, BBS1, BBS10, BBS12, BBS2, BBS4, BBS5, BBS7, BBS9, BCOR, BEST1, BFSP1, BFSP2, BMP4, C10orf2, C12orf57, C1QTNF5, C21orf2, C2orf71, C8orf37, CA4, CABP4, CACNA1F, CACNA2D4, CAPN5, CC2D2A, CDH23, CDH3, CDHR1, CEP164, CEP290, CERKL, CHD7, CHM, CHMP4B, CHN1, CHRDL1, CHST6, CIB2, CLN3, CLRN1, CNBP, CNGA1, CNGA3, CNGB1, CNGB3, CNNM4, COL11A1, COL2A1, COL8A2, COL9A1, CRB1, CRX, CRYAA, CRYAB, CRYBA1, CRYBA4, CRYBB1, CRYBB2, CRYBB3, CRYGB, CRYGC, CRYGD, CRYGS, CSAD, CTDP1, CYP1B1, CYP27A1, CYP4V2, CYP51A1, DCN, DFNB31, DHDDS, DMD, DMPK, DTHD1, EFEMP1, ELOVL4, EMC1, EPHA2, EYS, FAM161A, FLVCR1, FOXC1, FOXE3, FRMD7, FSCN2, FYCO1, FZD4, GALK1, GALT, GCNT2, GDF3, GDF6, GJA3, GJA8, GNAT1, GNAT2, GNPTG, GPR125, GPR143, GPR179, GPR98, GRK1, GRM6KRT3, GUCA1A, GUCA1B, GUCY2D, HARS, HCCS, HDAC8, HMCN1, HMX1, HOXA1, HSF4, IDH3B, IFT140, IGBP1, IGFBP7, IMPDH1, IMPG2, INPP5E, INVS, IQCB1, IQSEC2, JAG1, KCNJ13, KCNV2, KERA, KIAA1549, KIF11, KIF21A, KLHL7, KRT12, KRT3, LCA5, LEPREL1, LIM2, LRAT, LRIT3, LRP5, LTBP2, LZTFL1, MAF, MAK, MERTK, MFN2, MFRP, MFSD6L, MIP, MIR184, MITF, MKKS, MKS1, MTTP, MYH9, MYO7A, MYOC, NBAS, NDP, NHS, NMNAT1, NPHP1, NPHP3, NPHP4, NR2E3, NRL, NTF4, NYX, OAT, OCRL, OFD1, OPA1, OPA3, OPN1LW, OPN1MW, OPN1SW, OPTN, OTX2, PANK2, PAX2, PAX6, PCDH15, PDE6A, PDE6B, PDE6C, PDE6G, PDE6H, PDZD7, PEX1, PEX2, PEX7, PGK1, PHGDH, PHOX2A, PHYH, PIKFYVE, PITPNM3, PITX2, PITX3, PLA2G5, PLOD3, POLG, POLG2, POMT1, PRCD, PRDM5, PROM1, PRPF3, PRPF31, PRPF6, PRPF8, PRPH2, PRSS56, RAB18, RAB3GAP1, RAB3GAP2, RAX, RAX2, RB1, RBP3, RBP4, RD3, RDH12, RDH5, RGR, RGS9, RGS9BP, RHO, RIMS1, RLBP1, RNLS, ROBO3, ROM1, RP1, RP1L1, RP2, RP9, RPE65, RPGR, RPGRIP1, RPGRIP1L, RRM2B, RS1, RYR1, SAG, SDCCAG8, SEMA4A, SETX, SIL1, SIX6, SLC16A12, SLC24A1, SLC25A4, SLC4A11, SMOC1, SNRNP200, SOX2, SPATA7, STRA6, TACSTD2, TDRD7, TEAD1, TENM3, TGFBI, TIMM8A, TIMP3, TMEM126A, TMEM237, TOPORS, TREX1, TRIM32, TRPM1, TSPAN12 , TTC8, TTPA, TUBB3, TULP1, UBIAD1, UNC119, USH1C, USH1G, USH2A, VAX1, VCAN, VIM, VSX1, VSX2, WDPCP, WDR19, WDR36, WFS1, WRN, ZEB1, ZNF423, ZNF469, ZNF513, ZNF644.

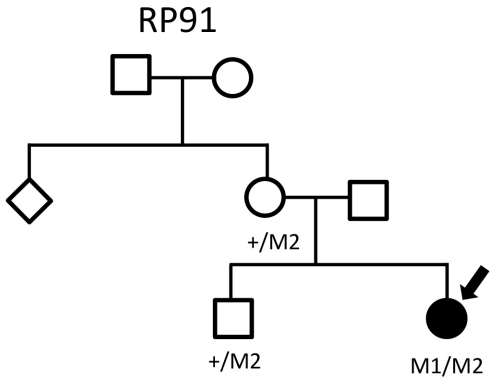
Chapter 2 Supplementary Table S3. List of genes analysed.

Supplementary Figure S1

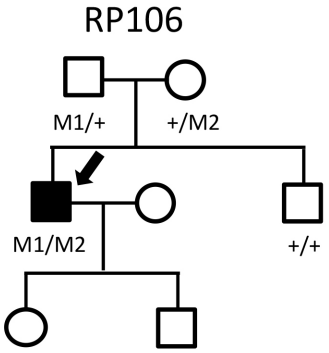
Family trees from all available probands with causal mutations and with variants of uncertain significance (VUS). Genotypes are annotated as M/M or M1/M2 (homozygotes or compound heterozygotes); M/+, M1/+ or +M2 (heterozygotes); or +/+ (wild type). Arrow indicates probands.



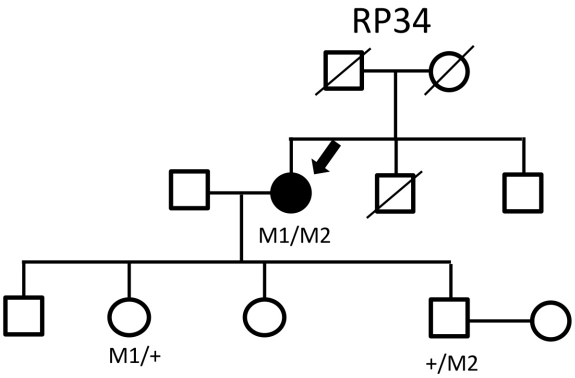

RP77
M1/M2
M1/M2
+/M2
+/M2
CNGA1:p.Arg101Ter/ p.Arg583Ter
RP57
M/+
M/+
M/M
M/+
M/+
TULP1: c.1495+1G>C/ c.1495+1G>C
RP49
M/+
M/+
M/M
M/+
M/+
EYS: p.Arg1349Ter/p.Arg1349Ter
RP88
M1/M2
M1/M2
5
MYO7A: p.Leu4AspfsTer39/p.Lys1255ArgfsTer8
M1/+



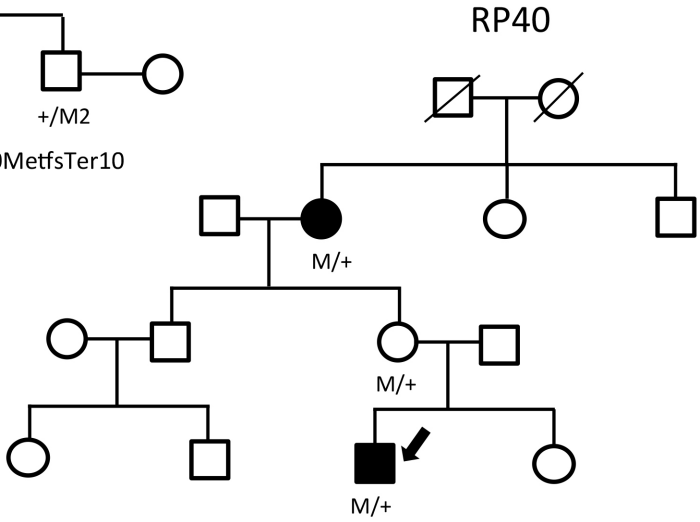
USH2A: p.Cys1223Ter / p.Trp3918Ter



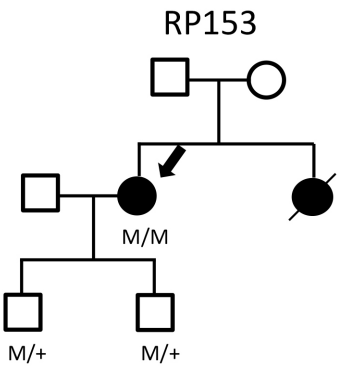
EYS: p.Lys296AsnfsTer43 / p.Ser5Ter



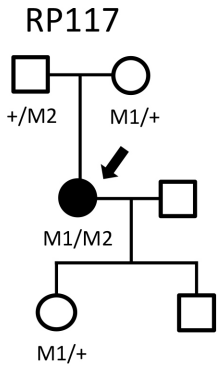
USH2A: p.Cys759Phe / p.Asp1760MetfsTer10



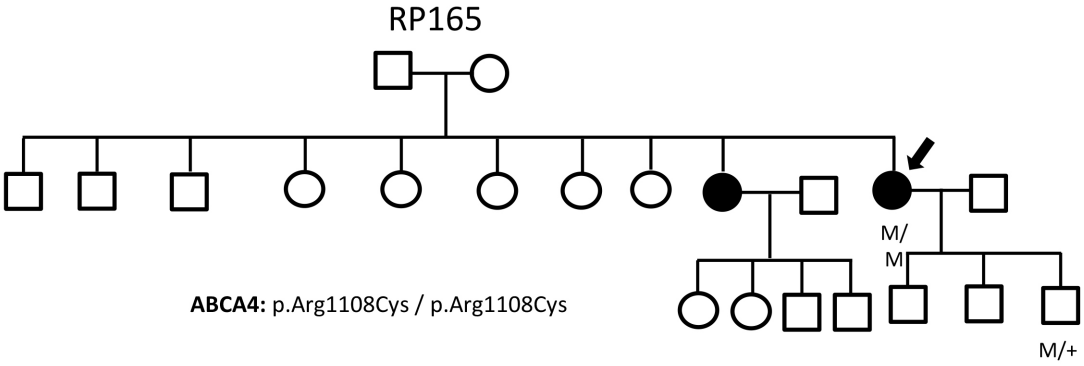
PRPF31: NM_015629 c.(801_856-43)_(1314_1500+80)del



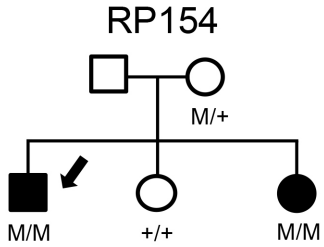
CERKL: p.Arg283Ter/ p.Arg283Ter



EYS: p.Arg1349Ter / p.Tyr3135Ter

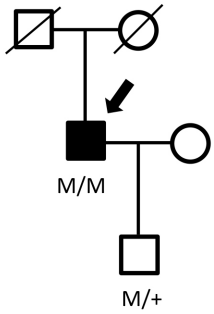


ABCA4: p.Arg1108Cys / p.Arg1108Cys



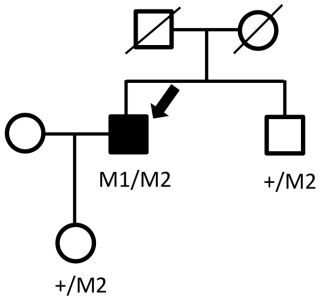
CNGA3:p.Arg55Ter / p.Arg55Ter

RP67



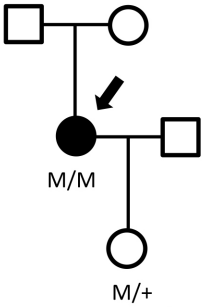
CERKL: p.Arg283Ter/ p.Arg283Ter

RP141



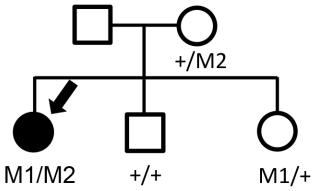
USH2A: p.Cys759Phe/ p.Glu767SerfsTer21

RP174



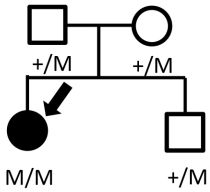
RGR: p.Ser66Arg/ p.Ser66Arg

RP175



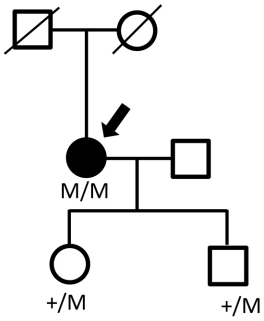
CNGB3: p.Thr383IlefsTer13/ c.852+1G>C

RP176

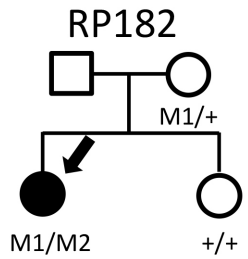


CERKL: p.Arg283Ter/ p.Arg283Ter

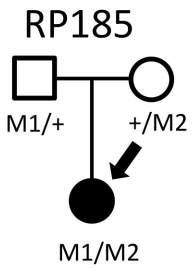
RP180



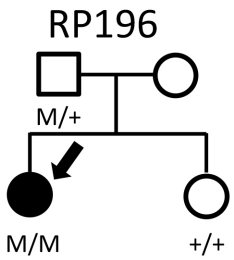
USH2A: p.Asn4856MetfsTer28/p.Asn4856MetfsTer28



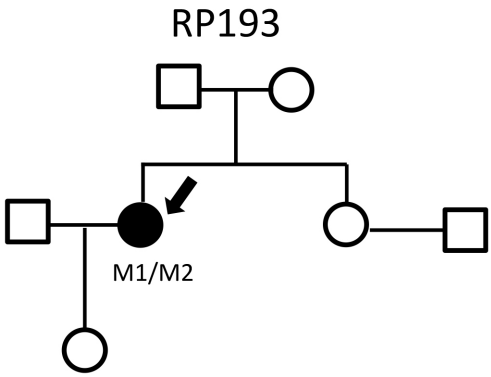
PDE6A: p.Arg653Ter/ p.Gln569Lys



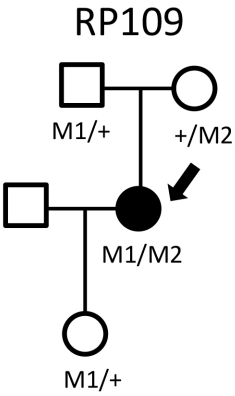
CNGA3: p.Arg410Trp/ p.Arg277Gly



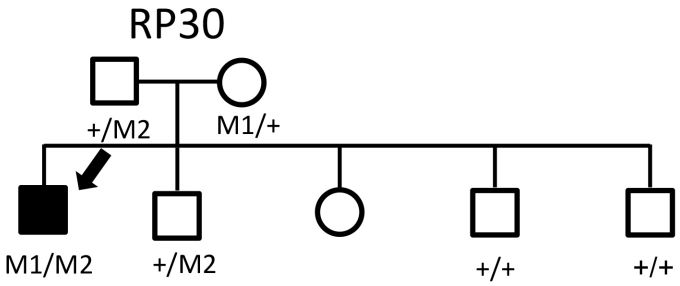
BBS1: p.Met390Arg/ p.Met390Arg



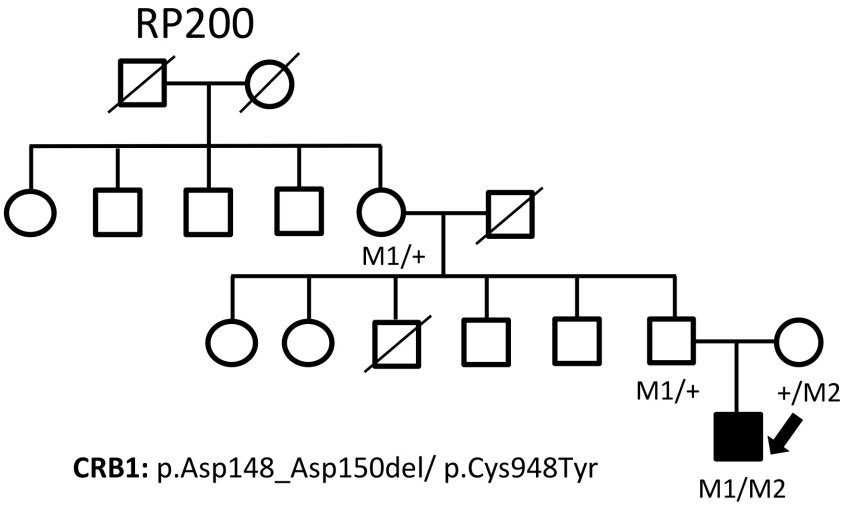
ABCA4: p.Thr1526Met/ p.Arg1129Leu



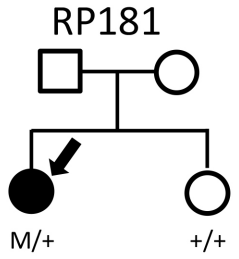
USH2A: p.Ala524Val/ p.Cys759Phe



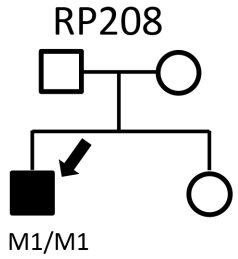
RP1: p.Ser542Ter/ p.Leu76Pro



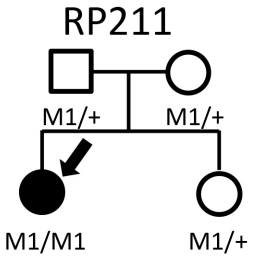
CRB1: p.Asp148_Asp150del/ p.Cys948Tyr



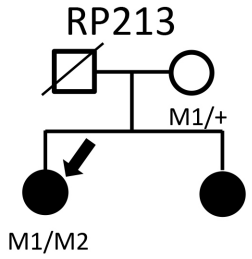
PRPF31: p.Gln389Ter



CERKL: p.Arg283Ter/ p.Arg283Ter

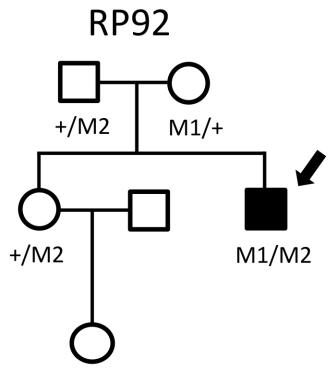


CRB1:p.Asp165_Ile167del

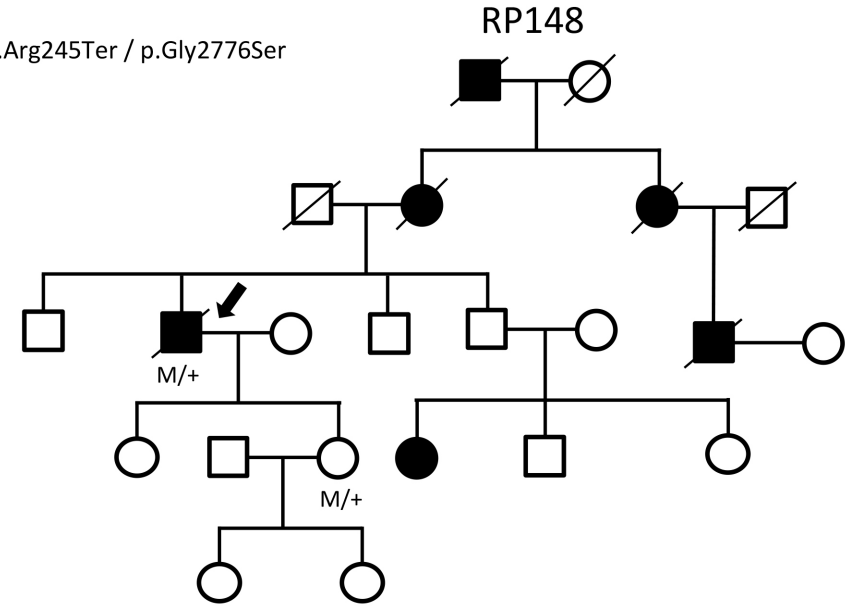


CRB1:p.Met1053Thr/ p.Cys948Tyr

VUS

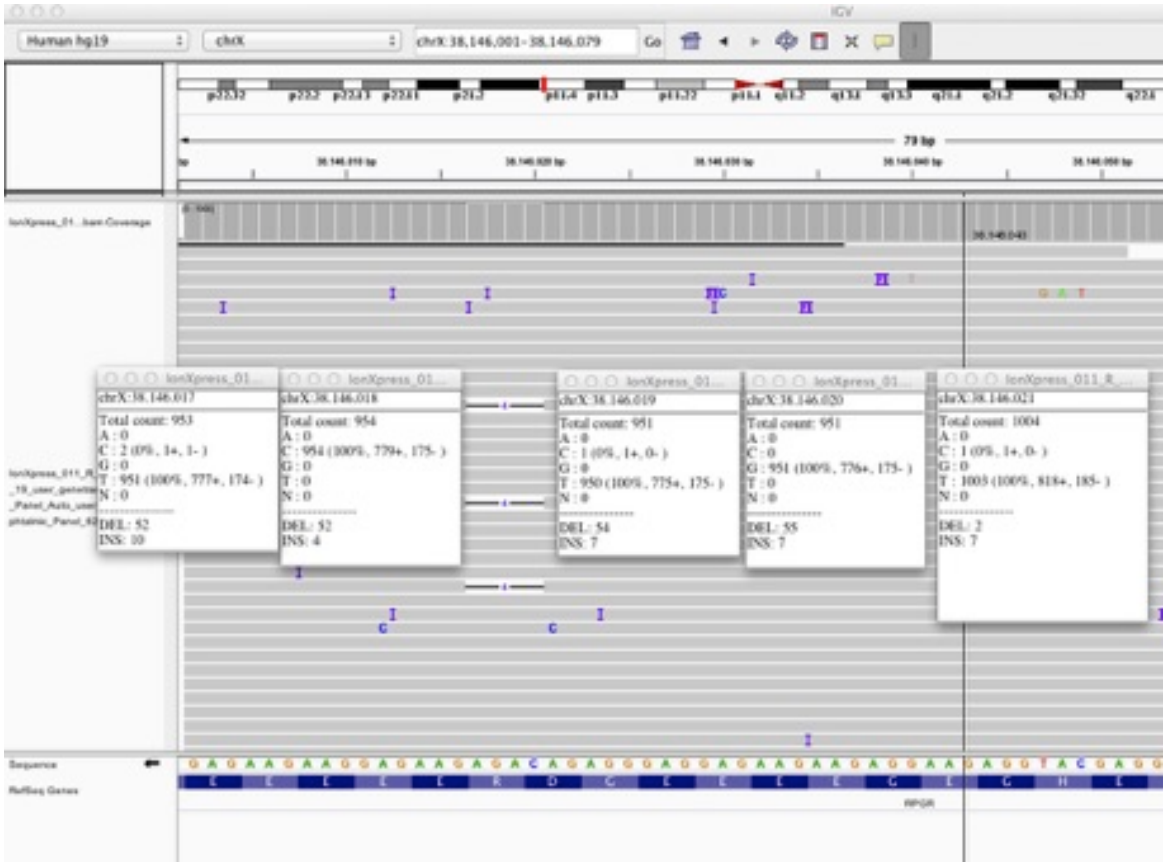


PCDH15/CDH23: p.Arg245Ter / p.Gly2776Ser



PRPF8: p.Trp2279Gly

Supplementary Figure S2



Chapter2 supplementary figure S2: IGViewer Screenshot of the RPGR ORF15 region: A deletion of 4bp was detected using our strategy. Total reads of each nucleotide and the number of reads were a deletion has been observed is annotated in each table. The fifth table is to show the difference in reads of the contiguous nucleotide, with no apparent deletions detected.

CHAPTER 3

LOH analysis followed by WES in ten patients with IRDs with no molecular genetic diagnosis after 316 genes panel based analysis

INTRODUCTION

Methods for molecular diagnosis, have developed greatly during the last years. The implementation of several techniques used in combination, such as panel based targeted-capture next-generation sequencing (NGS), genotyping microarrays of previously known mutations and Sanger sequencing, have permitted the identification of the causative mutations in 20–70% of IRD cases depending on the inheritance pattern and the selection criteria used[177,229,231,247,248]. The variability of unsolved cases implies that more robust diagnostic approaches are needed, and that new genes still remain undiscovered. Indeed in arRP or simplex cases, the genes identified to date and combining results from conventional Sanger sequencing and targeted-capture NGS, hardly explain the underlying pathogenic mutations or mutations in 20-30% of cases[247]. To increase this percentage WES has been widely used for novel candidate gene discovery not only in IRDs but also in other diseases [249–251]. In this study, WES was used to analyse 10 patients with no genetic characterization which had been studied with a panel based strategy where 316 IRD genes were sequenced. All patients, except one, were isolated cases which are traditionally predicted to be recessive, with unaffected carrier parents[247]. The other patient presented a typical X-linked inheritance pattern. A significant percentage of our patients have ancestors from the Basque Country, which is a Spanish region with a reported genetically homogeneous population. Several areas in Gipuzkoa, a province of the Basque Country, where most of our patients come from, have high frequency of consanguinity, ranging between 5 to 30%[50]. Therefore, we consider that it is quite likely to find homozygous carriers of mutations in genes not previously associated with any IRDs, which might be prevalent in our region. In fact this hypothesis has been tested in other diseases such as neuromuscular disorders and Parkinson disease with positive results[51]. To test this hypothesis, we first focused on the analysis of the Loss of Heterozygosity (LOH) regions in 9 out of 10 patients analysed by WES. LOH regions are fragments of the genome where the information of both alleles is the same. Possible causes include acquired uniparental disomy (UPD), gene conversion[252] and consanguinity. Individuals born to consanguineous parents have segments of their genomes that are homozygous as a result of inheriting identical ancestral genomic segments. A consequence of this type of union is an increased incidence of recessive diseases[127]. LOH regions were

tested using a probe and SNP based genome wide array platform *CytoScan XON* (Affymetrix, ThermoFisher, Santa Clara, California, USA). With this array we were able to detect not only LOH regions but also Copy Number Variations (CNV). After that, WES was performed in each index patient, where at first only variants in LOH regions were analysed. In the case of a negative result in this first approach, all variants from WES data were then analysed. Using this strategy, the patients were screened for CNV and for point mutations in genes not previously associated with IRDs.

MATERIAL AND METHODS

Sample Collection

10 families were selected for WES. Nine out of the 10 cases corresponded to isolated cases probably with autosomal recessive IRD inheritance pattern and the other one had a clear X-linked inheritance pattern. DNA was extracted from blood samples from IRD patients and their available family members (Family members did not undergo Homozygosity mapping nor WES). DNA was isolated with an AutoGenFlex STAR instrument and FlexiGene DNA Kit following the manufacturer´s instructions and stored at -80ºC in the DNA node of the Basque Biobank at Biodonostia Health Research Institute. The integrity of DNA was evaluated using 1% agarose gel and the quantity of DNA was measured using Qubit 2.0 Fluorometer (Invitrogen, ThermoFisher Scientific). Only DNAs with good integrity quality scores were used for WES analysis.

Inclusion criteria

The patients included in the study were patients with no characterization using the NGS panels with 316 genes analysed[248].

Other selection criteria included:

- Preference for patients or family members with surnames of Basque origin.
- Patients from Basque regions with high reported consanguinity index.
- Patients with a relatively large number of family members available for segregation analysis.

Homozygosity mapping

Considering that most of our families have recessive inheritance patterns and probably a high rate of endogamy, whole genome homozygosity mapping was performed. This technique allows the detection of loss of heterozygosity (LOH) regions. Whole genome homozygosity mapping was applied prior to WES analysis and was aimed at reducing costs and time derived from WES analysis. Affymetrix *Cytoscan HD* arrays (Affymetrix, Inc., Santa Clara, CA) were used for this approach. This procedure was performed in the index cases of each family and LOH regions with a size >1 Mb, which is the lowest limit of detection of this technique, were studied. All the LOH regions of each patient were annotated and a bed file

was generated to help in with the WES data filtering. Using this approach we first focus on analysing the variants located in those LOH regions[253].

Whole exome sequencing

Targeted exome sequencing

The capture of targeted sequences was performed at Leeds Institute of Molecular Medicine (University of Leeds, UK). The exome libraries were prepared starting with 200ng of DNA. The DNA was sheared to achieve a size between 200-300bp with a Covaris sonicator (ThermoFisher Scientific). After that, SureSelectXT Targeted Enrichment system for Illumina Paired-End Multiplexed Sequencing Library was used following the manufacturers protocol (Agilent Technologies). Figure 24 represents a brief schema of the protocol followed. Finally, the samples were sequenced in a HiSeq 3000 sequencer (Illumina) achieving mean exome coverage of 30X.

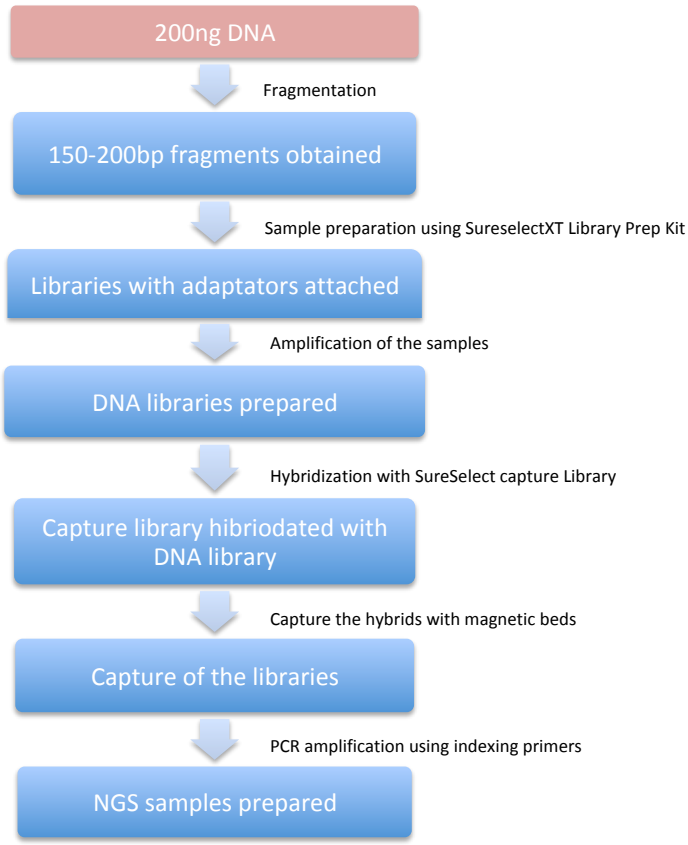


Figure 24: Schema of the protocol used for samples preparation before WES run.

Data analysis

Bioinformatic processing of the data was developed at the University College London (UCL) Institute of Ophthalmology. All sequenced data was aligned to the GRCh37 human reference genome using the Novoalign (3.02.08 version) alignment tool. Duplicates were marked and sorted using Picard tool. The variants were called according to Genome Analysis Toolkit (GATK) best practices (joint variant calling followed by variant quality score recalibration) incorporated into large-scale sequencing projects like the 1000 Genomes Project and The Cancer Genome Atlas [254]

Variants were then annotated using the Variant Effect Predictor (VEP)[255]. Variants with Minor Allele Frequency (MAF) < 0,005 were then filtered using the public control database gnomAD (http://gnomad.broadinstitute.org/). Computational prediction tools (PhyloP[256], CADD[257], SIFT[258], Polyphen2[259], and MutationTaster[144]) were used to predict the conservation and pathogenicity of candidate variants.

Variant filtering and analysis pipeline

Bed files were generated with LOH regions of each patient. All data was then filtered using the bed files generated. Among the remaining exome data, those regions or genes harbouring candidate variants previously linked to retinal dystrophies, or those genes expressed in the retina were prioritized[260]. Databases used for establishing a prioritized classification of variants discovered were RetNet; (https://sph.uth.edu/retnet/) and The Human Protein Atlas (http://www.proteinatlas.org). Single nucleotide Variants (SNVs), nonsense, frameshift, nonsynonim variants and splicing region variants in intronic regions (with the limitation in a WES dataset) were also included in the analysis.

Out of this first tier of variants identified we selected those with Minor Allele Frequency (MAF) < 0.003 in recessive cases and absent from databases in dominant cases. MAF values were not only checked in gnomAD, which was used for variant filtering, but also in the Spanish exome database (http://csvs.babelomics.org/).

All variants from WES data were analysed in patients without causative mutations found after analysing variants in LOH regions (

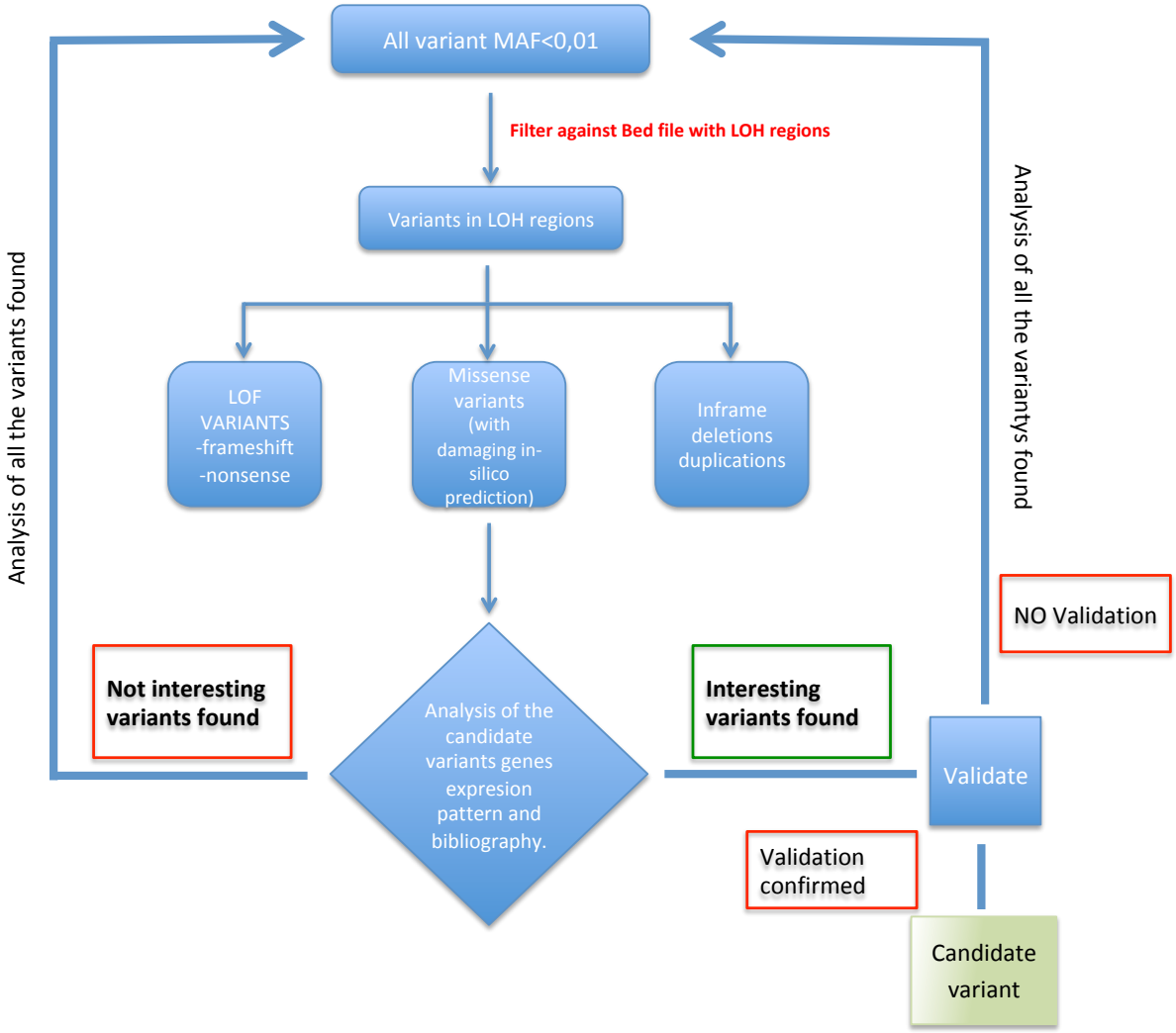


Figure **25**).

Figure 25: Schematic representation of the pipeline followed for variant analysis.

Molecular validation and family segregation

Sanger sequencing was used to validate all candidate variants. If the variant was validated in the index case the study was extended to the family members available for segregation analysis. In those cases that did not follow the expected segregation, other variants were selected. Candidate *novel* variants that were included in all filtering processes were also analysed in non-molecularly diagnosed patients from our cohort with the aim of finding the same candidate mutation in additional patients. These will significantly strengthen the likelihood of the implication of this potential *novel variant* in IRD.

RESULTS

LOH Regions Analysis and Identification of CNV

Each patient had an average of 106 LOH regions in the case of males and 130 regions in the case of females, probably due to the regions in the X chromosome. 63.7%+/-3.7 of the LOH regions length was between 1Kb and 1.5Kb and were widely represented across the genome (Figure 26). 3 LOH regions were observed consistently in different patients. 2 of them were repeated in 2 patients and the third region was repeated in 7 patients. Moreover, 16 similar regions were also repeated in some of the patients.

CNVs were also found in some of the patients, but in all cases the frequency of alteration in those regions in the population was high (when analysed against Affymetrix own database) or was shared by most of the patients analysed, and therefore regarded as neutral CNVs.

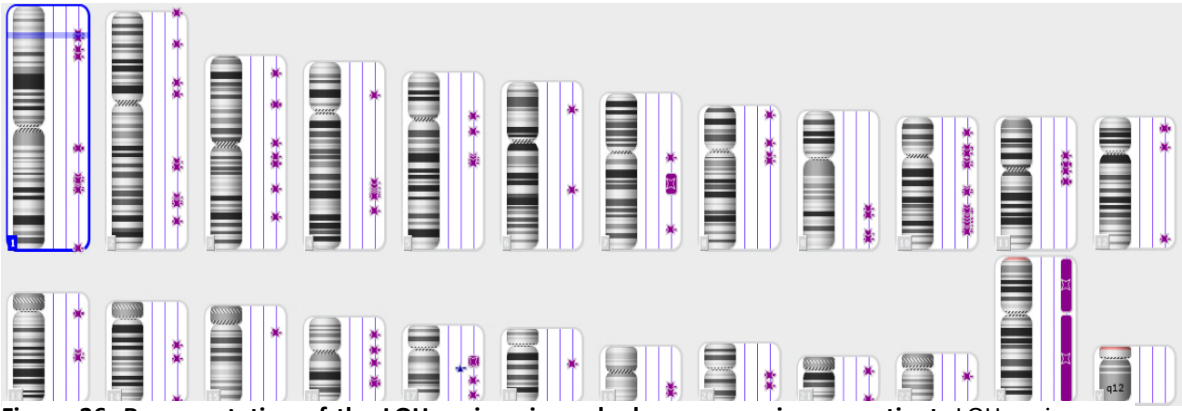


Figure 26: Representation of the LOH regions in each chromosome in one patient. LOH regions are represented in purple on the right side of each chromosome. Note that in chromosome X, the long purple area reflects that this sample is a male and the system detects all the X chromosome as LOH because there is just one copy.

Variant Identification

Due to the huge quantity of data generated, we needed to apply several filters to rescue the relevant list of variants. Firstly, and after all the quality filters, variants in LOH regions with MAF < 0,01 were filtered and analysed. As mentioned before, variants in LOH regions were first analysed in order to find mutations in the homozygosis state in genes not previously associated with IRDs, considering that most of our patients analysed come from areas with a high percentage of consanguinity. An average of 498+/-54 variants were found in each

patient with two outliers that had 296 and 833 variants. Variants which did not fit with the filtering criteria established. The majority of variants found in each patient were shared by more than three of the patients analysed. These variants were discarded given that we were looking for a very rare variant. After discarding those variants an average of 79+/-16 variants per patient were still present. After the analysis of these variants, we did not find any interesting mutations within LOH regions. Once we discarded the variants in LOH regions, we focused on the rest of the exome. With the applied filters and once variants repeated in more than 3 patients were discarded, the average of variants found per patient was 5324+/-417. As we expected very rare mutations with a recessive inheritance pattern (except in patient RP104 with X-linked inheritance pattern), variants with MAF< 0.003 were selected. Afterwards, inframe deletions and duplications, nonsense variants and missense variants predicted to be damaging by various software were selected. Moreover, the expression pattern and the function of the genes were also checked before the selection.

FAMILY	GENE	GENE TRANSCRIPT	cDNA Change	CADD Score	PhyloP	SIFT	POLIPHEN2	MUTATION TASTER
RP104	CNKSR2	NM_014927	c.524G>T	24,8	0.999	0.01	0.22	DC(0.99)
RP78	SAMD11	NM_152486.2	c.1888C>T	29,8	0.005	/	0.57	DC(0.99)
RP30	RP1	NM_006269	c.227T>C	22,9	0.96	0	0.99	DC(0.77)
RP109	USH2A	NM_206933	c.1571C>T	26,1	0	0	0.99	DC(0.99)

Table 7: In-silico pathogenicity score predictors of validated variants. Pathogenicity scores from the variants chosen for validation after WES analysis.

In patient RP104, who showed an X-linked inherited pattern, the X Chromosome was only analysed. Variant c.524G>T in gene *CNKSR2* (Table 7) was the sole interesting variant selected. This gene encodes a scaffold and adaptor protein that is part of the neuronal postsynaptic density (PSD) in the central nervous system[261]. It has been shown to produce X-linked mental retardation. Moreover, it has been reported that, Cnk (the homolog of mammalian CNKSR2) is a multidomain protein that participates in Ras (HRAS; OMIM: 190020) signalling in Drosophila eye development[262]. Unfortunately, Sanger sequencing analysis did not confirm correct segregation, so that the variant was discarded.

Regarding patient RP78, a nonsense mutation c.1888C>T with damaging prediction in *in-silico* predictive software was found in homozygosis in *SAMD11* gene. This gene is expressed

in the retina, it has been detected specifically in the developing murine photoreceptor layer at postnatal day 3 (P3) and the expression peaked in the photoreceptor layer at postnatal day 6 (P6). Protein-protein network analysis revealed a significant interaction of *SAMD11* with *CRX* (cone-rod homeobox containing gene), previously related to cone-rod dystrophy and Leber congenital amaurosis[263]

Moreover, we found that the same mutation, c.1888C>T, (Table 7) was recently described in two families of Spanish origin[119], supporting our finding.

In the segregation study we found that none of the members of the family were homozygous for this mutation, all of them being carriers resulting in a correct segregation pattern (Figure 27). Then we checked by HRM in our uncharacterized IRD patients trying to find more carriers of the mutation but with no results.

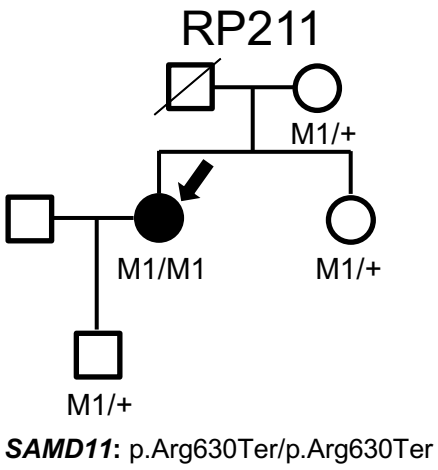


Figure 27: RP78 family tree

Patient RP30, previously described as a carrier of a nonsense mutation in *RP1* gene (c.1625C>G), in this analysis, a c.227T>C missense mutation (Table 7) was found in the same gene. Similarly, in patient RP109, who was a carrier of a missense mutation (c.2276G>T) in *USH2A* gene, a missense a second mutation (c.1570G>A) was found in heterozygosis state in the same gene (Table 7). Both second mutations found were also confirmed in parallel using a less strict reanalysis criteria of the 316 genes NGS panel, that was carried out in those samples (chapter 2).

The compound heterozygous mutations in RP30 and RP109 segregated properly in the families and agreed with their pathogenicity[248] so that we considered the mutations as the possible cause of the disease.

DISCUSSION

WES approach was used in ten patients, after LOH and CNV analysis, using a genome-wide array technology (CytoScan, Affymetrix). Eight out of ten patients were from Basque regions with high percentage of consanguinity and had Basque surnames and accessibility to family members. Following the approach described, three index cases were molecularly diagnosed.

Regarding the characterized RP78 patient with the c.1888C>T mutation in *SAMD11* gene, it is an important finding which strengthens the pathogenesis and the implication of this gene in retinitis pigmentosa.

These findings suggest the likely presence of a common ancestor and a mutation endemic to the Spanish population since it has not yet been reported in any other population.

Moreover, a heterozygous variant c.1814 T>A at 74bp from the nonsense mutation was observed indicating that as expected by LOH analysis, this mutation was not in a LOH region.

SAMD11 protein expression was found in the three nuclear layers of the retina but mainly in the Outer Nuclear Layer (ONL). In the case of the Inner Nuclear Layer (INL), expression was observed in a small population of amacrine cells. In the Ganglion Cell Layer (GCL), there is expression in the ganglion cells as well as in their axons. In the case of photoreceptors, localized in the ONL, there is a differential expression between cones and rods. While in rods the expression is observed in the cell body, as in inner and outer segments, in cones there is no expression found in the cell bodies. The protein expression observed in the rod cell bodies indicates the importance of the correct function of this protein in adult human retina rods meaning that the dysfunction could be involved in the RP pathogenesis[119].

The majority of genes mutated in retinal diseases and highly expressed in photoreceptors are under the control of the transcription factor Cone Rod Homeobox (*CRX*) the same as *SAMD11* and the more recently discovered *SAMD7* gene[263,264]. Both proteins contain a SAM domain which is located in the C-terminus region of the protein. This domain forms a protein-protein interaction module during transcriptional regulation[265]. Inoue *et al., 2006*[263] and Hlawatsch *et al., 2013*[264] state that both *SAMD11* and *SAMD7* have a CRX-mediated transcriptional repressor function. However, it seems that the repressor activity of *SAMD11*

is not due to SAM interactions but it resides in the conserved C-terminal region[263], where, interestingly the c.1814 T>A nonsense mutation described is located.

As indicated by Corton *et al*., 2016[119] the discovery of this causative mutation in *SAMD11* in three families, can indicate the importance of SAM related proteins, such as *SAMD7*, that share many common features with *SAMD11* in IRDs. All this data supports the idea that *SAMD7* could be a good novel candidate gene related to inherited retinal pathologies and should be studied thoroughly.

The other two characterized patients RP30 and RP109 were also characterized by the panel based NGS platform where 316 genes were sequenced (Chapter 2). In these cases, we were able to observe that both techniques demonstrated they were a good option for patient diagnosis.

In conclusion, in the case of the pooled panel-based NGS strategy WES helped learn that the first approach criteria for variant filtering was a little restrictive, so that with our present filtering criteria, the NGS panel have solved these cases without further experiments.

In this case, by WES analysis only one variant out of ten would have not been detected using panel based NGS strategies, since *SAMD11* was not included in our panel based NGS due to the recent discovery of this gene. However, it has to be taken into account that an important benefit of WES is that, in contrast to targeted panel based NGS sequencing, it is not limited to already known causative genes. Moreover, as novel disease-causing genes not previously ascribed to IRDs are discovered, a reanalysis of the data can be performed in order to search for mutations among those novel genes identified. However, WES shares some limitations with targeted sequencing. On one hand, neither can identify changes in non-coding regions, so that deep intronic variants are not analysed. On the other hand, the insertion and breakpoints of CNVs are often deeply intronic or intergenic, which results in difficulty in targeting by gene panel based NGS and WES, limiting variant detection algorithms[237].However, the improvements in the algorithms and quality assurance parameters have permitted the description of large structural variants, such as CNVs using these techniques[235,237]. In both cases, whole genome sequencing (WGS) would provide a solution to these problems since WGS is able to find mutations in deep intronic regions, it is less sensitive to high GC content and the coverage is more homogeneous, compared to targeted sequencing. This makes WGS more suitable for CNV detection. Nevertheless, the

WGS data is more complicated to implement and analyse; Firstly, because of the huge amount of data generated, the storage and the data processing capability is hard to handle and creates the need for computational power. Secondly, it is challenging to determine the pathogenicity and effect of many of the variants found in intronic regions [266]. Despite some promising on going initiatives such as the 100000 genome project in the United Kingdom (https://www.genomicsengland.co.uk/the-100000-genomes-project/understanding-genomics/genome-sequencing/) and the 1000 Arab genome project[267], the use of WGS as a molecular diagnosis technique is still challenging especially in economic terms, for most laboratories.

In summary, we have molecularly diagnosed 3 patients out of 10 analysed. In the case of the mutations found in genes related to IRDs we have seen that they can be observed easily using both strategies. However, it would have been enough with the panel based NGS strategy, if we had been less restrictive with the parameter thresholds in our first approaches to it. In the case of the mutation found in *SAMD11* gene, it is an interesting finding which strengthens the implication of this gene in retinitis pigmentosa pathology, adding a new case to the reported ones. To conclude, considering our results, WES can be contemplated as an appropriate second tier approach for novel candidate gene discovery. However, in the near future if affordability improves significantly, WGS will likely become a second tier technique, considering its advantages in comparison to WES.

CHAPTER 4

CNV and deep intronic variation analyses as a potential source of mutations for unsolved patients

INTRODUCTION

Inherited retinal degenerations (IRDs) are important causes of blindness that affect more than 2 million people worldwide[16]. There are several major clinical subtypes of IRDs, the most common being retinitis pigmentosa (RP). Other subtypes of IRDs are those predominantly affecting cones, such as achromatopsia or those firstly affecting cones and rods in more advanced stages (cone rod dystrophies) and/or affecting the macula and other pan-retinal degenerations such as Leber congenital amaurosis. Retinal degeneration is also one of the clinical manifestations of syndromic disorders such as different types of Usher syndrome and Bardet-Biedl syndrome[16,234]. Moreover, there is considerable phenotypic overlap between the different types of IRD. Apart from that, end-stage RP may be difficult to differentiate from late stages of some cone–rod dystrophies (CRD) and macular dystrophies (MDs) which becomes challenging for genetic testing[172,234]. In fact, over 280 different genes related to IRDs have been identified so far. This clinical and genetic heterogeneity hampers the efficiency and the promptness of molecular diagnosis of IRD[268].

Next-generation sequencing (NGS) enables the simultaneous parallel sequencing of numerous genes with high efficiency and is an efficient tool for molecular diagnosis of IRDs. On the other hand, WES is also widely used for molecular diagnosis and in this case, it also allows the possibility of describing a new gene associated to IRDs. Both strategies are aimed mainly at identifying single-nucleotide variants and small insertions and deletions in coding sequences of known and candidate genes. However, they are not able to find mutations in deep intronic regions and are less suitable than CGH arrays for finding, coding or noncoding small copy number variations (CNVs)[172] due to the fact that the insertion or deletion breakpoint positions are sometimes in deep intronic or intergenic regions[237]. Nevertheless, there is an increasing number of publications where they describe new filtering strategies for CNV discovery in WES[235] and panel based NGS[237] with remarkable results.

Recent studies have shown that copy number variations (CNVs) are frequent in higher eukaryotes and are associated with a substantial portion of inherited and acquired risk for various human diseases[269]. In the case of IRDs, the percentage of CNV cases among the patients is estimated to account for 3.5 to 10%, depending on the analysed cohort[172,235]. Moreover, it is estimated that CNVs are present within IRD genes in at least 1 in 13

individuals presenting IRDs[237]. To date, the majority of CNV cases in IRD have been described for a small number of genes. Most prevalent CNV mutations have been observed in *PRPF31*[116,178,270], *USH2A*[53], *EYS*[116,179] and *KCNV2*[271] genes, with other genes also bearing rearrangements reported by several studies. This includes *EYS, MYO7A*, *NPHP4, RPGR, CHM, HGSNAT* and *SNRNP200*[172,234,235] genes among others. Thus, these data highlight the importance of CNV screening.

In this regard, the screening of duplications, deletions and common deep intronic mutation in patients with *USH2A* mutation detects up to 35% of second mutations[53].

In fact, deep intronic pathogenic mutations have been described mainly in *USH2A*[272], *CEP290*[241]and *ABCA4*[273] genes among others.

In this study we aimed to increase the mutation detection yield obtained from panel based NGS, MLPA and WES strategies (Chapters 1, 2 and 3), analysing CNVs in 22 patients and 2 positive controls using two different Comparative Genomic Hybridization (CGH) arrays.

Moreover, in the case of patients with monoallelic mutations in *USH2A* or patients diagnosed as Usher syndrome type two, without second mutated allele after *USH2A* MLPA analysis, RNA from hair roots was amplified. It was used in order to detect mRNA length alterations caused by deep intronic mutations. Hair roots have been used alternatively as a source of *USH2A* mRNA, where both existing transcripts (both expressed also in the retina) are expressed. They have also been used in an to attempt to describe the mRNA expression modification of different mutations found[274] in deep intronic *USH2A* mutations.

MATERIAL AND METHODS

Study subjects

All patients were clinically diagnosed with different inherited retinal dystrophies, (IRDs), by the Ophthalmology Service at Donostia University Hospital. In the case of Comparative Genomic Hybridization array (CGH array) analysis, all the selected patients were previously analysed by panel passed Next Generation sequencing, but they did not carry any causative mutation or have monoallelic mutation in recessive genes not analysed previously by MLPA[177,248]. A total of 22 probands were analysed by *CytoScan XON* CGH array (Affimetrix, ThermoFisher, Santa Clara, California, USA), 2 of them were positive controls previously characterized by MLPA. From these probands, 8 (including the two controls) were analysed by Agilent Customized 8x60K CGH array (see supplementary Table 1) (Agilent Technologies, Santa Clara, California, USA).

For deep intronic mutations analysis, patients with a monoallelic mutation in USH2A and patients with Usher syndrome type 2 diagnosis were included. In the case of patients with monoallelic mutations in *USH2A* gene, MLPA analysis was previously performed in order to search for the second mutated allele.

All procedures performed in studies involving human participants received approval from the ethical standards of the Clinical Research Ethics Committee of the Basque Country, Spain (CEIC-E) and were in accordance with the 2013 Helsinki declaration or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Sample collection for DNA extraction

Genomic DNA was obtained from peripheral blood and isolated using AutoGenFlex Star instrument (AutoGen, Holliston, MA, USA) and FlexiGene DNA Kit (Qiagen, Hilden, Germany) following the manufacturer´s instructions. DNA concentrations were measured on the Qubit fluorometer using Quant-iT PicoGreen reagent (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). A260/280 purity ratio was measured in all DNA and the value required was between 1.7-2.1.

Array CGH analysis

Affymetrix CytoScan XON array

Samples were genotyped using 100ng DNA and the *CytoScan XON* array strictly following the manufacturer's instructions (Affymetrix). This part was performed in *Centro Nacional de Genotipado* (CeGen, Nodo Santiago). This array is designed for exon-level copy number variation analysis across the whole genome, containing 6.85M oligonucleotide probes (for the detection of CNV) and 300K SNP probes (for the detection of LOH regions). The copy number variation analysis was performed using the Chromosome Analysis Suite software (ChAS), version 3.3 (Affymetrix). All genomic rearrangements were annotated based on the GRCh37/hg19 Genome Build (February 2009).

All samples went through Quality Control (QC) filtering. The QC parameters were the Median Absolute Pairwise Difference (MAPD) that indicates the quality of Copy Number (CN) information; SNP-QC score, that indicates quality of allelic information and Waviness SD, that indicates the "oscillation" of the signal. Threshold values for QC pass for each sample were: MAPD< 0.21, SNP-QC>10 and Waviness SD< 0.08.

Agilent Custom CGH array

Agilent oligonucleotide-based custom microarray technology was used to analyse 8 samples. Six out of eight samples were first analysed using *CytoScan XON* array and were tested in this array in order to compare both array types. The remaining two patients corresponded to patients not previously analysed for CNV. SurePrint 8x60 format array was used. The microarray design was performed using Sure Design Web-based application (https://earray.chem.agilent.com/suredesign/) (Agilent Technologies). Human reference sequence GRCh37/hg19 genome build (February 2009) was used for probe selection. During the design, 14 genes were excluded considering that they were less than 60% covered. Finally, a total of 123 different IRD related genes were targeted with different coverage percentage (see supplementary Table 1).

Briefly, 200ng of genomic DNA was used to perform the experiment. In this case probe DNA and control DNA (Agilent Technologies) are simultaneously labelled using cyanine 3 (Cy3) and cyanine 5 (Cy5) respectively following the manufacturer´s protocol (Agilent Technologies). Both samples were then added to a slide and hybridized for 24 hours in an

oven using Cot-1 DNA (Invitrogen, part of Thermofisher Scientific, Carlsbad, California, USA). After washing, the slide was scanned using an Agilent DNA Microarray Scanner following the manufacturer´s protocol (Agilent Technologies). The signal was quantified using Agilent Feature Extraction Software (version 10.7.1.1). The data and the CNVs were then analysed using Agilent CytoGenomics software (Agilent Technologies). In this case the QC parameter was the derivative log ratio (DLR) that calculates the probe-to-probe log ratio noise. A DLR< 0.2 is considered the optimal threshold for aberration detection. DLR between 0.2 and 0.29 is considered borderline and a DLR> 0.3 is considered not appropriate for the study and was therefore disregarded.

CNV classification criteria

In order to classify and select the CNV variants, in the case of *CytoScan XON* array, we first selected variants in genes related to different IRDs. (https://sph.uth.edu/retnet/)

A database with 1855 healthy control samples, genotyped using *CytoScan XON*, was used as CNV frequencies assessment (This database was provided by Affymetrix, ThermoFisher Scientific). Variants with a prevalence higher than 0.5% in this database, were filtered out. Database of Genomic Variants (aDGV), was also used to compare CNV frequencies. After that, variants in genes not related to IRDs were analysed following the same criteria.

In the case of Agilent CGH array, variants were considered positive if the region altered encompassed more than three altered probes and the log2 ratio value was below -0.35 in the case of a possible deletion or above +0.35 in the case of duplication[275]. In the case of *CytoScan XON* array same threshold values were applied for the summarized Log2 ratio values (the average log2 ratio of the altered region).

CNVs were classified according to variant type (gain and loss), size, location and gene content. Available family members were also interrogated for CNVs. Finally, an attempt to match the patient´s phenotype with the finding was performed, in order to establish a phenotype-genotype correlation. Web databases that catalogue CNVs were also used, including Chromosomal Imbalance and Phenotype in Humans using ensembl resources (DECIPHER) (https://decipher.sanger.ac.uk/) and Online Mendelian Inheritance in Man (OMIM) (https://www.omim.org/).

qPCR analysis using TaqMan Assay

CNVs found in patient samples were then analysed by qPCR (CFX384 Touch; BioRad, Hercules, California, USA) in order to confirm the alteration. CNVs were also interrogated in available family members. The genomic positions of identified CNVs were used to find TaqMan (ThermoFisher Scientific) probes in the altered region and another probe at about 4Kb from the end or the start of the altered region. TaqPath ProAmp Master Mix (ThermoFisher Scientific) was used following the manufacturer´s protocol. *RNASEP* TaqMan Copy Number probe was used as reference assay (Thermofisher Scientific). In addition, 10 control samples were analysed in order to obtain robust results. These control samples were patients analysed by the microarrays with no relevant CNV detected.

Sample collection for RNA extraction

RNA from patients was isolated from the hair bulb. About forty hairs were obtained from each patient´s head. Hair bulbs were isolated under a dissecting microscope and rapidly transferred to RNA extraction mix, which consisted of 350ul of RLT from RNeasy Micro Kit protocol (Quiagen, Hilden, Germany) mixed with 35ul of 2-Mercaptoethanol (SigmaAldrich, San Luis, Missouri, USA). Bulb homogenization was achieved applying vortex agitation steps set at high intensity followed by various phases of syringe-based homogenization. A decreasing size of needle diameter was used, starting with 23g then 26g and finally 30g all adapted to 1ml syringes (Becton Dickinson, Franklin Lakes, New Jersey, USA). RNA extraction was performed using RNAeasy Micro Kit protocol (Quiagen) following the manufacturer´s instructions including a DNAse step, to eliminate the possible DNA remainders. RNA quantity and quality were measured in a Nanodrop spectrophotometer (ThermoFisher). All RNA samples were retrotranscribed to cDNA using SuperScript Vilo cDNA synthesis Kit (Invitrogen), following the manufacturer´s protocol. Finally, all the cDNA was diluted to 25ng/ul.

cDNA PCR

Primers in exon-exon junctions were designed for cDNA amplification. 39 different primer pairs were designed to cover all the exons of *USH2A* long transcript (NM_206933). The primer´s amplicons length ranged between 405bp and 574bp in order to have an equilibrium between the number of amplicons and their size that could lead to the discrimination in size

in case of alteration. Platinum SuperFI DNA polymerase (TheroFisher Scientific) was used for all PCRs, following the methods indicated by the manufacturer. The PCR products obtained were run in a 2% agarose gel using X174 RF DNA Hae III fragments as molecular weight control (Invitrogen). Finally, the amplicons obtained were observed in a UV transilluminator for amplicon size analysis.

RESULTS

CytoScan XON analysis

Samples were in optimal conditions for CNV analysis, since all of them successfully passed established quality control criteria, described in the material and methods section.

A mean of 43 different copy number variations were found in each patient. In the case of the two internal controls introduced with mutations in *PRPF31* gene, both mutations were confirmed by this array, suggesting a correct run and quality control parameters (see Table 8).

Following our CNV identification pipeline described in the material and methods section we were able to detect relevant rearrangements in 4 out of 20 patients analysed (see Table 8). All CNVs were located in IRD-related genes, matched with the likely inheritance pattern, and were present in heterozygous state.

Since mutations in *PRPH2* are involved in autosomal dominant retinitis pigmentosa, we consider it likely that this mutation is responsible for the disease in this patient. Moreover, this deletion encompassing exons 2 and 3 was previously described in a Sardinian family[276].

Regarding the two mutations in *EYS* genes and the mutation in *GPR98*, they would require a second mutated allele in the same gene as these genes are related to autosomal recessive retinitis pigmentosa. No relevant mutations were found in 16 out of 20 patients analysed, neither in IRD related genes nor in any other gene. Only a few rearrangements were observed which were discarded, since they were highly represented in the databases analysed or among several patients from our cohort, suggesting that they are probably neutral or rather common variants.

Agilent CGHarray

In this case, 8 patients were analysed: 4 samples corresponding to patients with mutations in IRD genes identified using *CytoScan XON* arrays, 2 positive controls used also in *CytoScan XON* array, and 2 patients not previously analysed for CNV. All patients passed the quality control. 6 patients with DLR value <0.2, one with DLR=0.2 and one with DLR=0.25.

Notably, the mutations carried by the 2 positive controls used were not detected by this array (see Table 8). Moreover, mutation in RP40 previously observed by MLPA and by the *CytoScan XON* array was not detected in this case. This array only analysed 123 IRD-related genes, so that the number of variants found per patient was very low. Indeed, there was one patient with no mutations detected and an average of 3 variants per patient were detected in the other patients.

FAMILY	GENE	VARIANT CYTOSCAN XON	VARIANT AGILENT CGH array	EXONS	REFERENCE
RP2 (POSITIVE CONTROL)	PRPF31	chr19:54621572-54626832	chr19:54621818-54625965	1 to 5	[277]
RP40 (POSITIVE CONTROL)	PRPF31	chr19:54629877-54633746	not detected	9 to 13	[248]
RP36	EYS	chr6:65611751-65612956	not detected	17 and 18	
RP71	PRPH2	chr6:42663861-42672787	chr6:42663740-42673060	2 and 3	[276]
RP114	EYS	chr6: 66111427-66112591	not detected	7	
RP142	GPR98	chr5:89932560-89934763	not detected	11	

Table 8: Variants found in two array types.

Therefore, from 6 patients analysed, the only interesting mutation found was the one detected by the *CytoScan XON* array in RP78, which corresponded to a deletion in *PRPH2* (see Table 8). It is worth noting the fact that we were not able to validate the other 3 mutations detected by *CytoScan XON* array.

Similar to the previous approach, this array allowed the detection of common variants, some of which were present in several patients. Moreover, alterations encompassing less than 4 probes or alterations in the genes used as control probes for the array were not considered.

qPCR TaqMan validation assay

As a second technique qPCR TaqMan probes were used for confirmation of the results provided by arrays. According to the result obtained with the Agilent CGH array, only the mutation in *PRPH2* in patient RP71 was redetected. In this case, family members available in each case were also analysed (Figure 28).

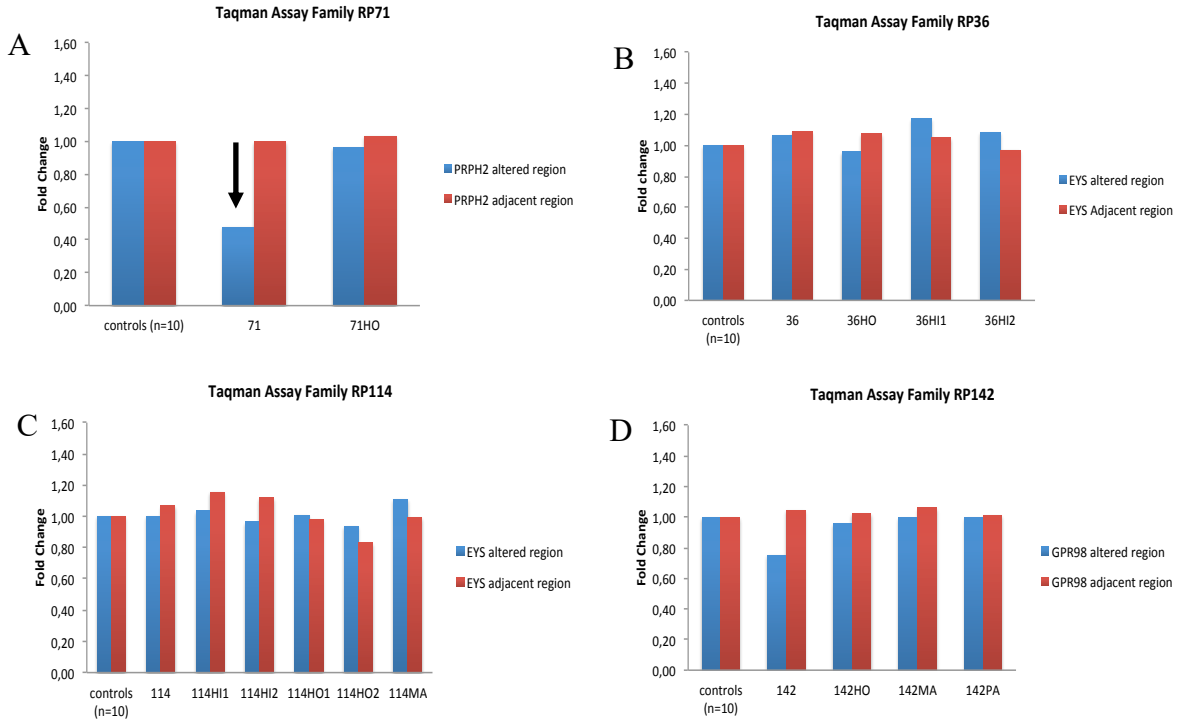
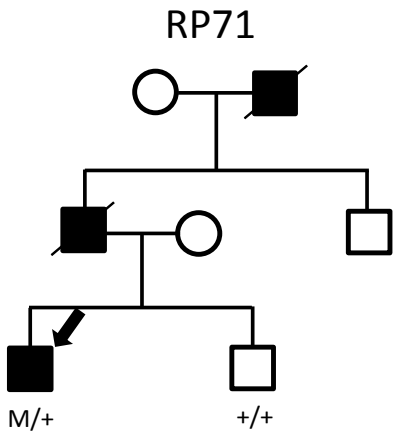


Figure 28: TaqMan results expressed in foldchange per family. Blue colour bars represent the fold change in the altered region, considering previous arrays results. Red colour bars indicate the fold change in a non-altered adjacent region, following previous arrays results. Alterations were considered when the fold change was inferior to 0.6 for monoallelic deletions and 1.4 for monoallelic duplications**.** A). The deletion was detected in *PRPH2* in patient RP71 (arrow) which was not present in the healthy brother (71HO). B) Family RP36 with negative result for alteration in *EYS*. C) Family RP114 with negative result for alteration in *EYS*. D) Family RP142 with negative result for the possible deletion in *GPR98* detected by *CytoScan XON* array.

Note that in patient RP71, there is half of dose (Fold change = 0,49) in *PRPH2,* compared with the adjacent control region and with his healthy brother, indicating a correct segregation (Figure 29).



PRPH2: NM_000322 (c.581+1_582-1)_(c.1041)del

Figure 29: RP71 family tree.

***USH2A* RNA level analysis**

Five patients with monoallelic mutations in *USH2A* gene or diagnosed as Usher syndrome type 2, were analysed using this methodology. RNA was extracted from hair roots in all patients, but the amount extracted varied depending on the patient. The amount of RNA extracted was lower in patients with thin hair and in patients with dyed hair. However, all RNA was retrotranscribed to cDNA. RNA was also extracted from hair roots of a control sample.

After cDNA PCR analysis, in none of the 39 amplicons designed, did we observe any differences in the length of the expected amplicons in any patient (see examples in Figure 30).

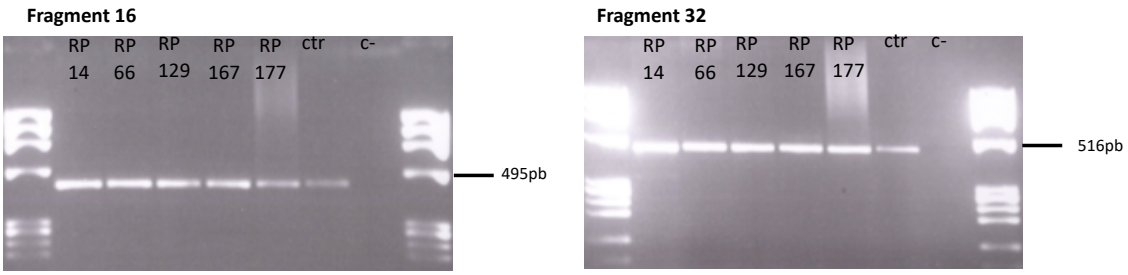


Figure 30: Examples of cDNA PCR of 2 fragments out of 39 amplicons analysed in all patients: Abbreviations: Crt. Control sample. c-. Negative control of the PCRs where water was added in the place of cDNA.

DISCUSSION

In this chapter, two different methodologies were used in order to increase the number of characterized patients after using the approaches previously mentioned. 27 different patients were analysed, 22 for CNV analysis and 5 for *USH2A* deep intronic mutations screening.

Molecular diagnosis was achieved in one patient where a *PRPH2* heterozygous deletion, encompassing exons 2 and 3 was detected. However, we did not find any relevant mutation in *USH2A* deep intronic analysis.

The two array based platforms used in this study and the TaqMan based qPCR uncovered a CNV variation in the *PRPH2* gene in one patient with a dominant inheritance pattern pedigree. *PRPH2* mutations have previously been associated with different IRD phenotypes such as choroidal dystrophy[278,279] and retinitis pigmentosa[280,281]. Interestingly, the mutation found in our study, which expands exons 2 and 3, was previously described in a Sardinian family with autosomal dominant Butterfly-Shaped Macular Dystrophy[276].

In both arrays, we detected a small CNV in *PRPH2* gene. A reliable detection of small CNVs using NGS data in poorly or highly covered exons can be challenging[172]. Based on current methodology, CGH arrays, seem to be more sensitive to detect small CNVs, especially in highly dense and equally distributed sets of probes arrays[282]. However, the customized Agilent CGH array, did not detect the positive control deletion in *PRPF3*1. This could be explained by the lower coverage achieved in the design of the array in this gene (see supplementary Table 1.).

It is also noteworthy that the 3 CNV variants found in the CytoScan array, were neither detected by the customized CGH array nor validated by qPCR-based TaqMan probes. These 3 alterations had the summarized log2 ratio above +0.35 or below -0.35, which were the established filtering criteria. However, after reanalysing the data when we realised that were false positives, some differences were observed in other thresholds used for each probe deviation assessment, such as log2 ratio and weighted log2 ratio. (see Figure 31)

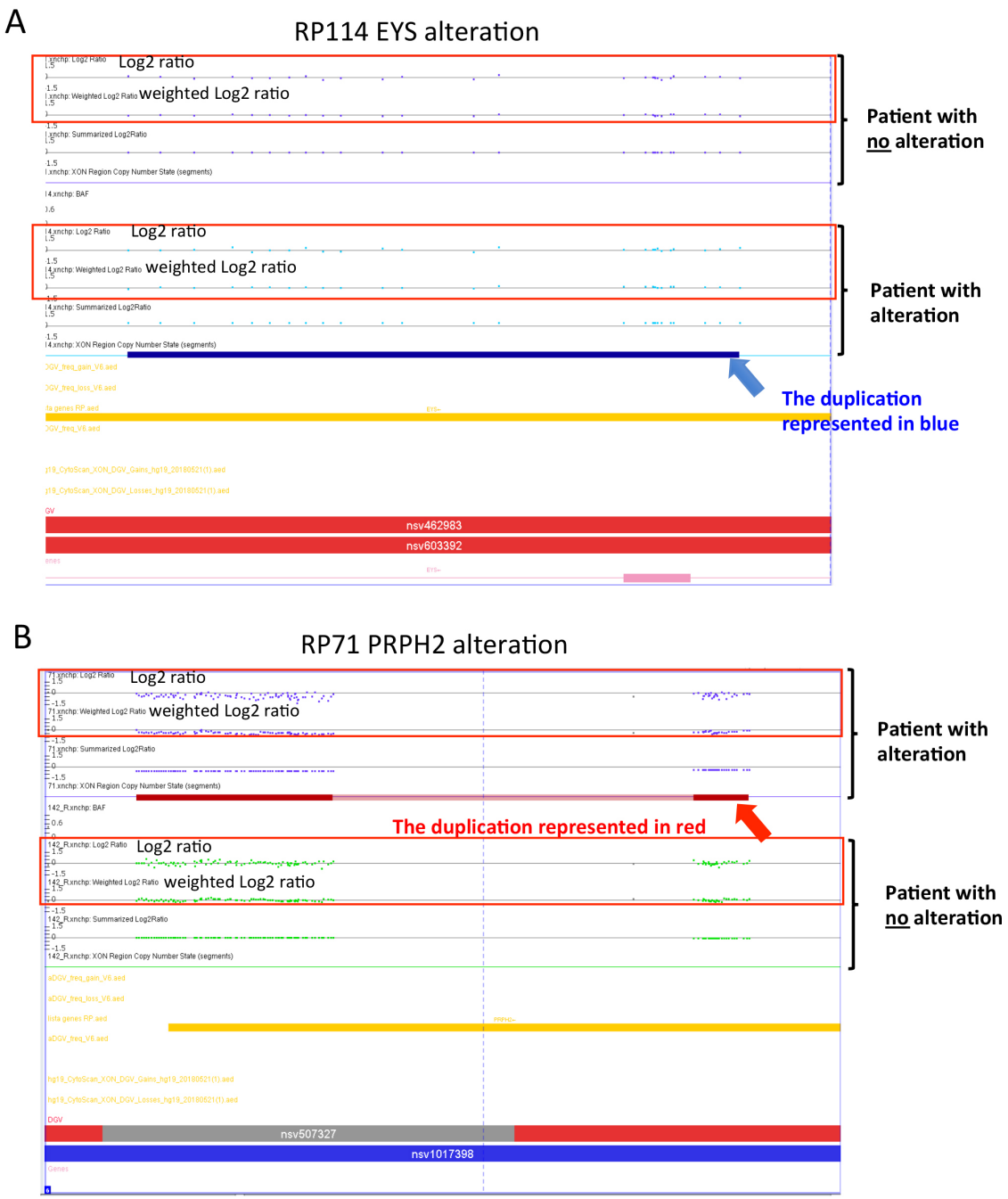


Figure 31: Summary of results provided by Chas analysis software (Affymetrix). A. Duplication of RP114 patient in EYS gene compared to non-carrier patient. B. Deletion of RP71 patient in PRPH2 gene compared to non-carrier patient.

Note that in Figure 31A (patient RP114 whose alteration was not confirmed) the CNV probes in log2 ratio threshold and the weighted log2 ratio threshold, shows no differences between the affected and the non-affected patients. In contrast, Figure 31B shows clear differences between affected and non-affected patients, as measured by log2 ratio threshold and the weighted log2 ratio threshold parameters. This CNV was further validated by PCR analysis.

Based on these results using two different CGH arrays and the new reanalysis, we were able to improve our selection criteria, which will allow us to reduce the number of false positives in future studies.

Regarding the number of patients molecularly diagnosed in the literature, several CNV analyses have been performed in different diseases[275,283,284], and also in IRDs[116,172,234] with different results ranging from 3.5 to 10% of diagnostic yield. In our study we were able to detect 4 alterations, with only one validated in patient RP71 in *PRPH2* gene.

One explanation for this relatively low diagnostic yield obtained is the small cohort of patients analysed. Nevertheless, this yield is in accordance with reported prevalence of CNV among IRD genes of 1 in 13 individuals analysed[237], considering that we searched for CNV in 20 patients for the first time in this work.

Increasing the number of patients in future experiments will hopefully increase our diagnosis yield. Another possible explanation could be that these patients carry mutations in genes not previously associated with IRDs or point mutations in deep intronic regions or regulatory regions not analysed in this study. Whole genome sequencing could be an option for analysing those regions and also whole-gene targeted sequencing, analysing the most prevalent genes where deep intronic mutations have been found[285]. The second case has the advantage of being more affordable and that the amount of data generated per patient is lower.

In the case of *USH2A* deep intronic regions analysed using RNA from hair roots, although the patients analysed had genetic findings in this gene or had clinical Usher type 2 syndrome, in none of the 5 patients analysed a positive result was obtained. Our purpose was to search for deep intronic mutations involved in changes at mRNA level, introducing abnormal inclusions of intronic sequences, leading to frameshift or introducing premature stop codons[286]. Following our methodology using 2% agarose gels, we expected to observe differences in the migration (indicative of size differences) of some of the cDNA samples analysed. This would have led to the discovery of the “second” mutation in *USH2*A, as reported by several studies[53,285]. Unfortunately, this was not our case. One reason for the negative result obtained could be due to the low number of patients analysed. Another reason could be that we missed some little changes in the length of some amplicons, which means that a possible deep intronic mutation can cause a subtle change in the length of the

region as small as 1 or few nucleotides. However, this would be enough to induce a change at the protein level, which we were not able to detect with the methodology used in our approach. A third possibility is that, the monoallelic mutation carried by the patient in *USH2A* gene, is not the mutation causing the disease. Despite USH2A being by far, the most prevalent gene for Usher syndrome type 2, other USH genes such as *GPR98* and *WHRN*[287] need to be taken into consideration.

One of the patients was included in the study despite not having been diagnosed as Usher type 2, since she is a carrier of a monoallelic mutation in *USH2A* gene. It is known that some *USH2A* mutations cause retinitis pigmentosa phenotype with or without hearing loss[288]. Therefore, some of the patients initially diagnosed with retinitis pigmentosa, are then re-diagnosed as Usher syndrome type 2 after genetic characterization.

About 40% of the causal variants are uncovered by sequencing coding region and splice site sequences[273]. Whole-genome sequencing can identify most of the non-coding variants, but their interpretation is still very challenging, especially when the relevant gene is expressed in a tissue-specific manner and this is difficult to obtain as in the case of the retina.

In this case, we used an alternative way to search for deep intronic mutations in *USH2A* gene using an accessible tissue with reported expression of the two main reported transcripts. In light of our results, we can conclude that this approach might not be the one of choice for finding the second mutated allele in deep intronic regions. In fact, this seems to be a better option as a method for characterising new deep intronic variants as it has been used by Nakanishi *et al*., 2010[274].

In summary, using these two approaches we were able to find a CNV in one patient with dominant inheritance pattern, but we were not able to find any mutations in deep intronic regions.

CHAPTER 4 SUPPLEMENTARY INFORMATION

Supplementary Table S1

TargetID	Interval	Coverage	Total Probes	Median Probe Spacing(bp)
RDH5	chr12:56113151-56119526	97,16%	21	263
OTX2	chr14:57266425-57278197	95,03%	39	249
ABCA4	chr1:94457393-94587705	94,58%	434	264
PDE6H	chr12:15124956-15135799	94,48%	36	253
ELOVL4	chr6:80623529-80658315	94,39%	115	269
EFEMP1	chr2:56092097-56152298	93,74%	200	254
TIMP3	chr22:33195802-33260030	93,43%	214	258
RP1	chr8:55527627-55544394	93,34%	55	266
CEP290	chr12:88441790-88536993	93,32%	317	254
RPGR	chrX:38127416-38187817	92,96%	201	259
CEP78	chr9:80849978-80895606	92,80%	152	243
RPE65	chr1:68893505-68916642	92,80%	77	268
OFD1	chrX:13751832-13788480	92,49%	122	251
FZD4	chr11:86655717-86667440	92,49%	39	249
VCAN	chr5:82766284-82879122	92,45%	376	246
MKS1	chr17:56281797-56297966	92,29%	53	255
PRPF4	chr9:116036623-116056466	92,19%	66	255
MKKS	chr20:10380657-10415887	91,96%	117	247
TTC8	chr14:89289497-89345340	91,83%	186	256
CDHR1	chr10:85953391-85980377	91,79%	89	253
RHO	chr3:129246482-129255187	91,64%	29	233
SNRNP200	chr2:96939074-96972307	91,54%	110	250
LZTFL1	chr3:45863808-45958534	91,12%	315	251
PROM1	chr4:15963699-16087001	90,72%	411	235
PDE6C	chr10:95371345-95426767	90,64%	184	238
GDF6	chr8:97153558-97174020	90,56%	68	237
DRAM2	chr1:111658954-111683838	90,21%	82	246
TMEM237	chr2:202483907-202509293	90,06%	84	258
BBS10	chr12:76737254-76743222	90,00%	19	268
TOPORS	chr9:32539542-32553626	89,89%	46	252
RPGRIP1L	chr16:53630595-53738850	89,57%	360	237
USH1C	chr11:17514442-17566963	89,56%	175	233
GNAT2	chr1:110144889-110156705	89,47%	39	228
BBS5	chr2:170334688-170383432	89,33%	162	227
LCA5	chr6:80193708-80248175	89,28%	181	231
RLBP1	chr15:89752098-89765982	89,22%	46	225
ATF6	chr1:161735034-161934860	89,20%	666	234
BBS2	chr16:56499748-56555195	88,80%	184	219
IDH3B	chr20:2638041-2645865	88,65%	26	253
C2ORF71	chr2:29282842-29298127	88,43%	50	243
SDCCAG8	chr1:243418307-243664394	88,34%	820	235
LRAT	chr4:155547097-155675271	88,17%	427	237
SLC7A14	chr3:170176342-170304863	88,13%	428	235
RD3	chr1:211648864-211667259	87,92%	61	237
KCNV2	chr9:2716502-2731037	87,75%	48	242
USH2A	chr1:215795236-216597738	87,46%	2675	236
CERKL	chr2:182400401-182546392	87,10%	486	239

RIMS1	chr6:72595406-73113845	87,01%	1728	235
CC2D2A	chr4:15470489-15604180	86,87%	445	222
NEK2	chr1:211830599-211849972	86,84%	64	223
BBS12	chr4:123652857-123667098	86,74%	47	250
BBS7	chr4:122744484-122792652	86,67%	160	230
TTLL5	chr14:76098968-76422425	85,86%	1083	223
AHI1	chr6:135603670-135819914	85,86%	720	228
UNC119	chr17:26872725-26880686	85,69%	26	234
KLHL7	chr7:23144353-23218533	85,63%	247	227
RP9	chr7:33133409-33150013	85,62%	55	202
RBP4	chr10:95350444-95362501	85,60%	40	218
CRB1	chr1:197169592-197448585	85,40%	929	230
IFT172	chr2:27666238-27713678	84,83%	158	211
BEST1	chr11:61716293-61733987	84,79%	58	217
MERTK	chr2:112655056-112788138	84,74%	443	209
PRPF8	chr17:1552923-1589176	84,68%	120	201
SIX6	chr14:60974669-60980568	83,95%	19	233
IFT43	chr14:76367479-76551928	83,87%	615	213
CWC27	chr5:64063745-64315590	83,53%	839	220
RP1L1	chr8:10462859-10570697	83,51%	359	210
BBS4	chr15:72977520-73031817	83,46%	180	197
NPHP4	chr1:5921868-6053533	83,29%	438	211
ARL2BP	chr16:57278010-57288547	83,19%	35	212
POC1B	chr12:89812495-89921039	83,09%	361	209
AGBL5	chr2:27264232-27294490	82,97%	100	220
GPR98	chr5:89824161-90461038	82,76%	2122	217
POC5	chr5:74968949-75014313	82,71%	151	211
RCBTB1	chr13:50105082-50160719	82,71%	185	210
MAK	chr6:10761956-10839788	82,66%	259	191
ZNF513	chr2:27599098-27604657	82,45%	18	199
PDE6A	chr5:149236519-149325356	82,44%	296	191
CDH23	chr10:73155691-73576704	82,32%	1403	212
PCDH15	chr10:55561531-57388702	82,22%	6090	214
CLRN1	chr3:150642950-150691786	82,17%	162	227
DHDDS	chr1:26757773-26798795	81,78%	136	190
NRL	chr14:24548316-24585223	81,61%	123	195
ARL6	chr3:97482365-97521086	81,37%	129	217
GUCA1B	chr6:42150022-42163694	81,17%	45	209
BBS9	chr7:33167856-33646680	81,15%	1596	210
FAM161A	chr2:62050983-62082278	80,75%	104	195
RS1	chrX:18656808-18691229	80,54%	114	196
HARS	chr5:140051758-140072609	80,35%	69	193
CNGB3	chr8:87565205-87756903	80,06%	638	210
SAG	chr2:234215309-234256701	79,92%	137	168
GUCY2D	chr17:7904912-7924658	79,76%	65	192
PITPNM3	chr17:6353583-6460877	79,66%	357	199
CIB2	chr15:78395948-78424886	79,61%	96	202
EYS	chr6:64428876-66418118	79,23%	6630	207
EMC1	chr1:19541158-19579053	78,84%	126	181
ZNF408	chr11:46721317-46728466	78,57%	23	219

CHM	chrX:85115185-85303566	78,36%	627	201
ADAM9	chr8:38853388-38963779	78,36%	367	189
CNGA3	chr2:98961618-99016064	78,31%	181	199
AIPL1	chr17:6296013-6339519	78,06%	145	173
IMPG2	chr3:100940390-101040419	78,03%	333	204
RBP3	chr10:48380487-48391991	77,57%	38	180
PHYH	chr10:13318796-13345412	77,52%	88	157
RP2	chrX:46695347-46742793	77,22%	158	177
HK1	chr10:71028740-71162638	76,79%	446	178
CACNA1F	chrX:49060523-49090833	76,18%	101	186
SEMA4A	chr1:156116157-156148543	76,13%	107	175
CNGB1	chr16:57915244-58006020	74,88%	302	163
GUCA1A	chr6:42122115-42148821	73,42%	89	157
CNGA1	chr4:47936994-48019689	73,04%	275	163
PRPH2	chr6:42663333-42691358	72,86%	93	138
RGR	chr10:86003809-86020716	72,32%	56	142
IMPDH1	chr7:128031331-128051306	72,12%	66	197
MYO7A	chr11:76838310-76927286	69,09%	296	149
TULP1	chr6:35464651-35481715	68,40%	56	168
CRX	chr19:48321703-48347587	67,79%	86	125
RDH12	chr14:68167603-68202169	67,75%	115	151
USH1G	chr17:72911176-72920358	67,34%	30	150
HGSNAT	chr8:42994556-43058998	64,71%	214	129
PDE6B	chr4:618363-665681	64,40%	157	107
PRPF6	chr20:62611431-62665453	63,87%	180	119
PRPF31	chr19:54617790-54636150	62,31%	61	105

Chapter 4 Supplementary Table S1. **Targeted genes in Agilent customized array**: Interval; Indicates the chromosomic region in where the genes are located. Coverage; Indicates the percentage of the gene covered by oligonucleotides. Total probes; Indicates the number of probes used for covering each gene. Median probe spacing; Indicates the median base pairs (bp), between one probe and the next one.

GENERAL RESULTS AND DISCUSSION

In this work, 174 patients with IRDs have been analysed using different strategies. 17 of these patients had undergone previous molecular diagnosis and were used as positive controls for sensitivity and reliability assessment. All mutations harboured by positive control samples were redetected by the methodology used in each case, giving 100% sensibility reading, with the exception of the deletion reported by us using MLPA, which was not then detected with the customized Agilent CGH array.

Out of 157 patients analysed without molecular diagnosis, 91 harboured a damaging mutation in at least one allele, which means that we found at least one damaging allele in 58% of the patients. However, complete molecular diagnosis was possible in 61 patients and another 3 cases were included as VUS, giving a diagnosis rate of 38.9% or 40.8% in the case of considering VUS variants. This percentage is not as high as in other studies where the detection rate is between 49% and 64% using panel based strategy and/or WES [114,154,199,231,289].

Considering the inheritance pattern, 32 cases of autosomal dominant patients were analysed in this work and 17 were diagnosed, giving a detection rate of 53.1% in this group of patients. Apart from the diagnosed patients, 2 other patients were classified as VUS. Out of the diagnosed patients, one was characterised for a mutation in an X-linked gene, in this case allowing a genetic reclassification. The percentage of solved patients is in accordance with other works. In fact, this percentage is higher than the 23% reported in a Spanish cohort[111] and 41% reported in a German cohort[113] and is slightly lower than the 56% of a Belgian cohort and 60% of a large Spanish cohort[290]. However, it is lower than the 78% described in a large cohort from the United States[130]. These percentage differences between different works shows the diagnosis rate variability between different populations.

On the other hand, 121 cases considered as possible recessive were analysed. In this case 110 out of 121 were simplex cases, which means that the index case was the only affected member of the family. In this group of patients, 45 out of 121 have been diagnosed in this work giving a diagnostic rate of 37.1%. Furthermore, another patient was classified as VUS. This percentage is lower than the diagnosis rate obtained in other works, such as in a Spanish cohort of recessive retinitis pigmentosa in which 57% of the patients were diagnosed[110].

Moreover, it is important to mention that, although some patients have been analysed using different techniques, there is another high percentage of the cohort, in which only 316

genes have been sequenced by panel-based NGS and probably if we had analysed all the patients with all the techniques described in this work, the percentage would have increased. In fact, as mentioned above, there are 20 patients from a total of 27 with pathogenic mutations in heterozygosis (Table 9, Table 10), in which a second mutated allele has not been found neither with WES, CGH arrays nor MLPA depending on the case. This can indicate that the second mutation can be located in an intronic or a regulatory region or that the gene in which a mutated allele has been found, is not the causative gene of the disease. This is the case of patients RP31, RP114, RP119 or RP143 among others, in which MLPA and/or CGH array has been performed with negative result for finding a second mutation allele. On the other hand, there are 7 patients with pathogenic mutations in one allele but in which other techniques for second mutation allele analysis has not been performed yet. In these cases, there is the possibility of finding the altered allele using other techniques as in patients RP184, RP210 or RP222 among others. Table 9 summarizes the techniques used for diagnosis in each patient.

Considering the lower diagnostic rate obtained in this work compared to other studies, one reason could be that our cohort of patients is more heterogeneous than others. In fact, it has been observed that levels of mutation detection rates achieved for IRDs that are less genetically heterogeneous, such as cohorts with high rate of patients with Choroideremia, involving solely the *CHM* gene, and/or Stargardt disease patients, predominantly caused by *ABCA4* mutations, were typically significantly higher[291]. Another reason could be that more than 1/3 (60 patients) of the 157 patients analysed, have been previously analysed using other diagnosis approaches such as HRM and microarrays[88,95] with negative results indicating that some cases are challenging. A third reason could be that we are faced with a cohort with high prevalence of mutations in intronic or regulatory regions not analysed in this work. Finally, it is noteworthy that detection rates are affected by the range of conditions under study, such as whether the rate is obtained based on previous screening of the same population or not, making precise comparison of detection rates between studies, challenging[291].

In this work 84 different mutations were found, 34 of which were described for the first time and the most common ones were missense followed by frameshift mutations. This result indicates the high variability and complexity of alterations found in IRDs and shows the challenge that supposes. These mutations have been observed in 28 different genes *USH2A*

being the most prevalent one with 9 patients solved and 2 with mutation in heterozygosis followed by *SNRNP200* and *CERKL* genes (Figure 32). *USH2A* has also been reported as the most prevalent recessive gene also in other studies with a prevalence of 17%[14], showing that mutations in this gene are responsible for both Usher syndrome type 2, and RP. It is seen that, mutation p.Cys759Phe is the most repeated in this work, found in homozygosis in one patient and in heterozygosis in 3 patients and is described as the second most prevalent *USH2A* mutation in the Spanish population[37]. Patients with homozygous p.Cys759Phe mutation have been related with arRP or arRP plus hypoacusis[47,188], as in the case of our patient. However, the pathogenic role of this mutation has been questioned since it was found in homozygosis in two healthy siblings[292]. Nevertheless, there is another recent publication where the pathogenicity of this mutation is supported[47]. In this work the authors show the pedigree of 14 homozygous families for this mutation and in 4 of them a correct segregation is proved. Moreover, checking the Babelomics Spanish database of exomes (http://csvs.babelomics.org/) we observe not a single individual with this mutation in homozygous state among healthy individuals or individuals with other diseases not related with IRDs.

Mutations in *CERKL* have been described as causative of 1% of the arRP cases world-wide. Moreover, this prevalence is especially high in Spanish population with a prevalence of 5%[293] of recessive IRD cases and is also commonly mutated in the Finnish population[294]. In the case of Spanish population, the most prevalent mutation is p.Arg257Ter which is present in 3.3% of recessive IRD patients and until 2008 was only detected in the Spanish population. The prevalence of this mutation is even higher in our cohort of patients where 5 cases in homozygosis and in 1 case in heterozygosis state were found, giving a prevalence of 4.2% of the recessive cases analysed.

Finally, the mutation p.Ser1087Leu in *SNRNP200* found in 7 families is likely founder effect of this mutation, which was present in very small and rather isolated Spanish populations. In fact, this is not the most prevalent mutated gene in the Spanish population neither in other populations, so that our proportion is probably overestimated.

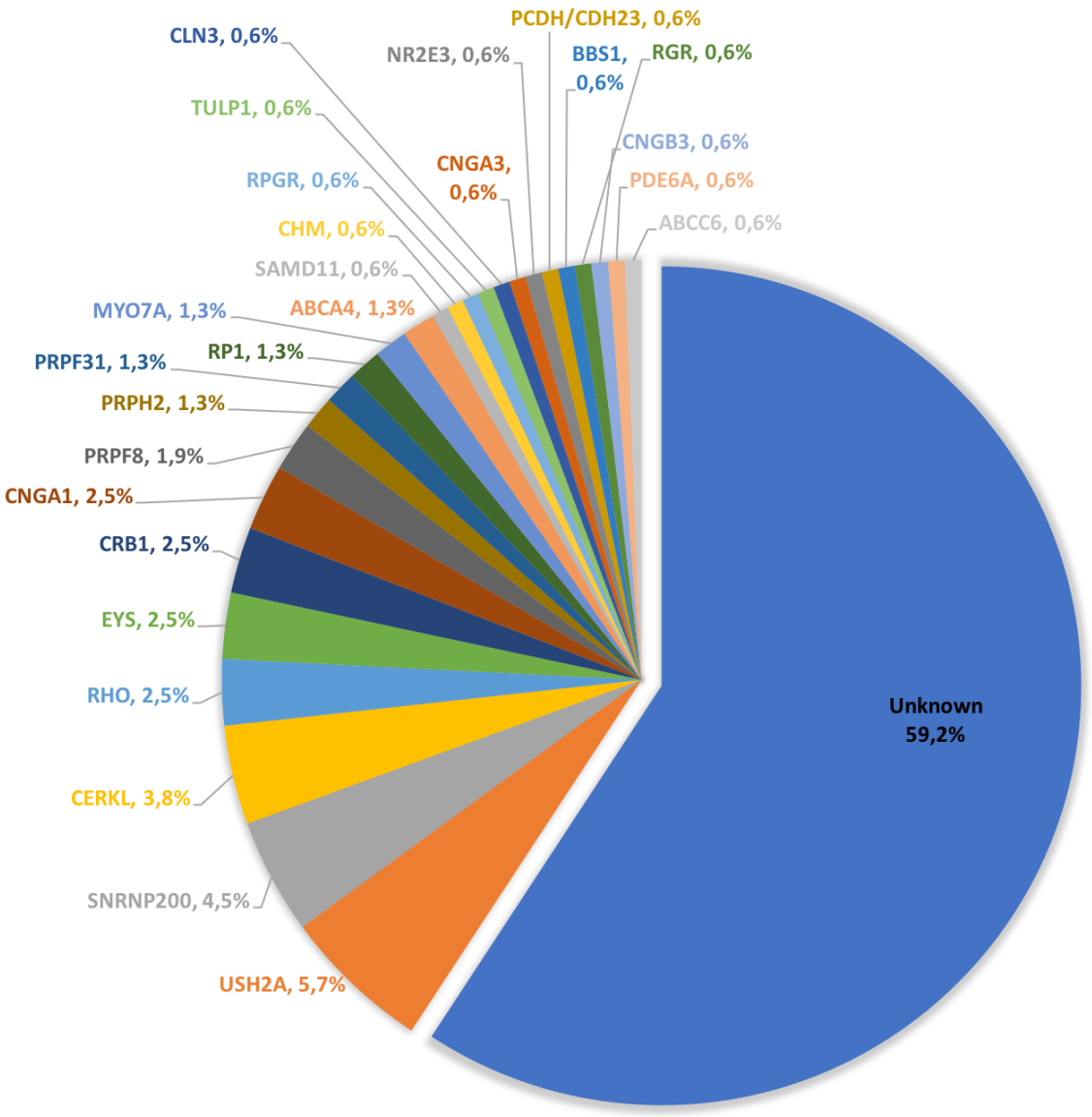


Figure 32: Representation of the percentage of genes in which mutations were found.

It is interesting to mention that 2 patients were diagnosed with mutations in *CLN3* and *ABCC6* genes. *CLN3* is associated with Batten disease or Ceroid lipofuscinosis, neuronal 3[295]

that is characterized by the intracellular accumulation of autoflorescent lipopigment and courses with progressive dementia in some cases, and progressive visual failure[296]. *ABCC6* is associated with pseudoxanthoma elasticum[297] and courses with characteristic lesions of the posterior segment of the eye, of the skin and of the cardiovascular system due to accumulation of mineralized and fragmented elastic fibres in Bruch´s membrane of the eye, in the skin and in vascular walls respectively. These two genes are not always included in different panels described in the literature[110,237,298,299], in our case, the analysis of high number of genes permitted the diagnosis of those two patients that in other cases would have been missed.

Moreover, the design of this panel in which all genes related to different IRDs (discovered before 2015) are included, allows the analysis of all patients using the same panel without considering the possible inheritance pattern or the clinical phenotype of each. Thus, the analysis does not depend on the previous inheritance pattern or phenotype, permitting the possible reclassification of the patients. Indeed, it is important to highlight that in this work one patient was reclassified considering the inheritance pattern. Patient RP27 was first analysed with the dominant IRDs panel, with negative result. After pooled panel based 316 IRD genes analysis, a pathogenic mutation was found in *RPGR* that segregated properly in the family (Table 9). Moreover, 5 patients from solved cases were reclassified. This result shows the high phenotypic overlap and highlights the challenges involved in the phenotypic classification of the patients[300]. Therefore, the correct molecular genetic diagnosis is important for the proper classification of the patients. Out of those 5 patients, 2 were first diagnosed as Stargardt, and after molecular genetic reclassification, both were harbouring mutation in *CERKL*, which is related to RP. The other three cases were first diagnosed as RP but after genetic analysis one patient was reclassified as Stargardt, the second patient harboured a mutation in *CLN3* gene associated to Batten disease and the last one had a homozygous mutation in *ABCC6* gene, associated with Pseudoxantoma elasticum.

The different techniques used in the working pipeline of this thesis (Figure 33) allowed us to diagnose 40.7% of our patients (including VUS cases). Although we initially used a panel with dominant IRD genes, once we validated our approach based on pooled sequencing of 316 IRD genes panel observing a 100% of sensibility in control patients, we decided to use this as

our first approach. This means that, all new patients incorporated in our cohort would be analysed with this panel regardless of their inheritance pattern or phenotype and as mentioned above, permitting a comprehensive analysis.

Our pipeline included WES sequencing after LOH and CNV analysis, using *CytoScan HD* array, in 10 of our patients with negative results after 316 genes panel analysis. The use of this methodology allowed on one hand, the diagnosis of one patient with a homozygous mutation in *SAMD11* gene, which was recently related to arRP, and the detection of a second mutated allele in two patients which were also detected in a less restrictive analysis of panel-based data.

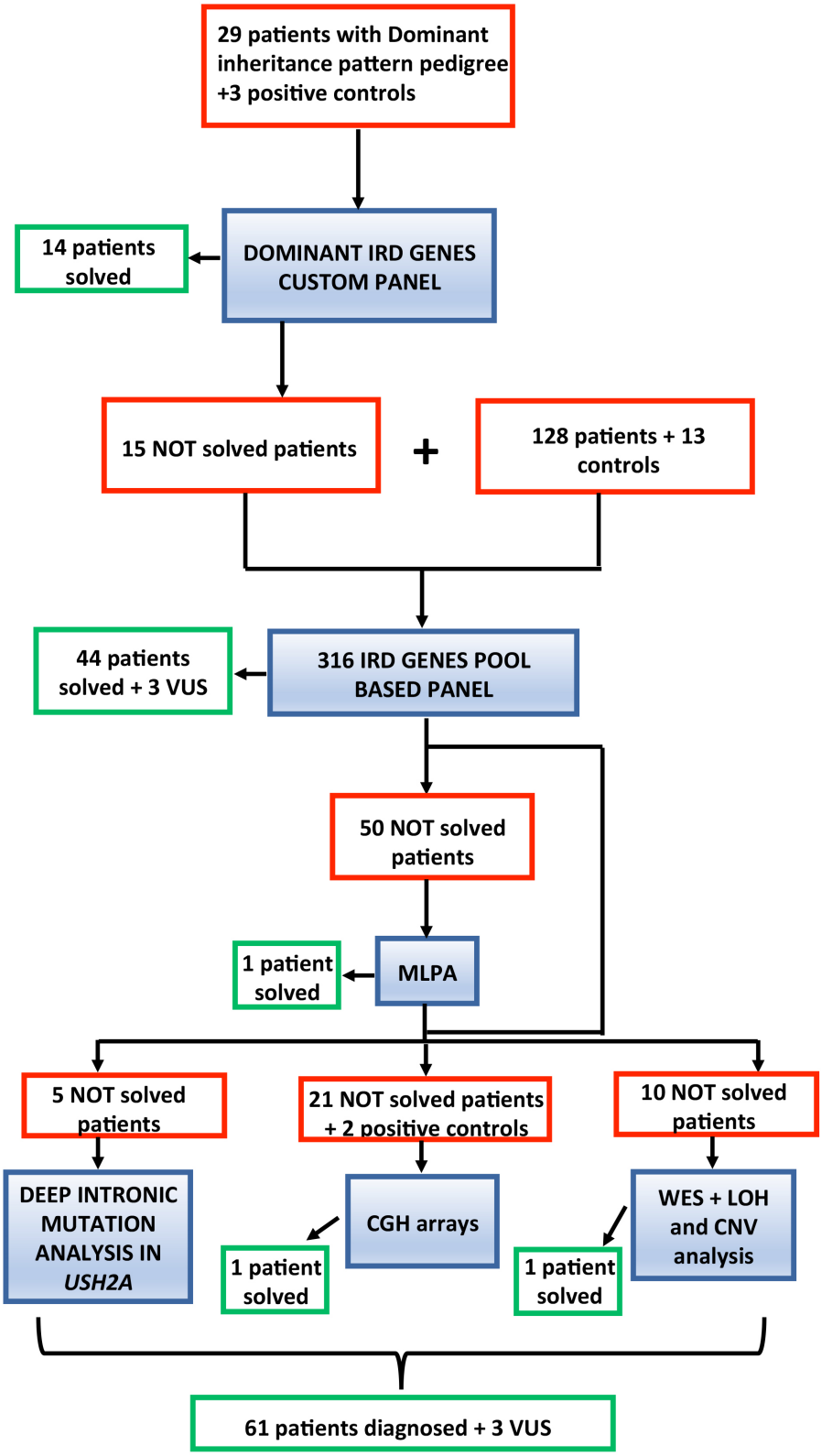


Figure 33: Schema of the analysis pipeline followed in this work.

SAMD11 gene was not included in the panel as it was first described at the end of 2016, some months before our result after WES. This finding is interesting because it strengthens the relation of this gene to RP and it is also noteworthy that the same mutation was detected before in the Spanish population, indicating that this mutation can be a founder mutation from Spain. We found no patients with the same mutation nor other relevant mutations in *SAMD11* gene after analysing our non-solved cohort of patients using HRM analysis. This is in line with a trend in which the more recently discovered IRD associated genes are responsible for a very small percentage of cases compared to the genes previously discovered[301] as it was proved in studies in Saudi-Arabian and Dutch populations. In these studies they observed that only 3-4% of IRD cases that are still unsolved, carry variants in novel IRD associated genes[302–304]. In fact, we were not able to find any mutation prevalent in our cohort, which could explain the lower diagnostic rate obtained compared to other studies. Following the argument explained above, we find it more likely that a high number of patients are carriers of mutations in deep intronic regions or a regulatory region of a previously IRD related gene which will need to be analysed by WGS in future studies. On the other hand, we were not able to detect a possible prevalent mutation in a gene not previously associated to IRDs in our patients, probably due to the low number of patients in which WES was performed.

WES has become the most commonly used technology for finding the causative mutation in patients without diagnosis after panel based analysis. However, it has also been widely used as first tier approach. In this case, it has been seen that studies employing WES rather than panel based NGS, have resulted in the identification of the underlying genetic defect yielding similar results as in panel based NGS, about 60%[232,277,291]. Taking into account the similarities in the detection rate of both techniques, we consider that panel based NGS is the best option as a first-tier approach. Moreover, the variant analysis is often more complex in WES studies than in panel based NGS, considering that the number of variants obtained in the last case is lower than in WES because there are a limited number of genes studied. Nevertheless, we consider that WES can be adequate as a second tier approach, although it does not allow deep intronic region or regulatory region analysis. In this sense, WGS is a powerful technique that avoids those limitations and although currently restricted to some research laboratories, its use is being considered as the first tier sequencing method of choice, once it becomes an affordable option for clinical applications[305].

Apart from WES, CNV analysis was also used as a second tier approach in our patients. MLPA and two types of CGH arrays were used for this analysis and a total of 62 patients were analysed with one or both techniques. Among all patients analysed, a CNV was found in two cases, one was detected by MLPA in *PRPF31* gene and the other one was detected using CGH array in *PRPH2* gene giving a detection rate of 3.2%. This percentage is similar to that obtained by Van Cauwenbergh *et al*., 2017[172], with a 3.5% of detection rate in a cohort of 57 patients. However, it is lower than the 10% shown by Khateb *et al*., 2016[235], in a cohort of 60 patients. In any case, it is noteworthy that in our case all IRD genes were only analysed in only 21 patients in which CGH array was performed. The number of genes analysed by MLPA is very limited, so that it is possible to have patients with CNV alterations in other genes not analysed by MLPA. In fact, patient RP71 was analysed for dominant genes available with MLPA, with a negative result. After CGH array analysis, a deletion in *PRPH2* gene was detected which could not be found by MLPA. Thus, this percentage is probably underestimated considering that not all patients have been analysed with the most powerful techniques.

Finally, this work helped to molecularly diagnose 61 patients, some of which will hopefully be able to benefit from ongoing therapies and from genetic counselling. Moreover, novel genetic variants detected in this study increase the mutation spectrum of IRDs and have contributed in the advancement of the knowledge of the mutation spectrum of Basque and Spanish cohorts. In fact, common mutations from our population have been described, which will hopefully be useful for future studies made in this cohort. On the other hand, we describe a novel methodology based on pool-based panel gene analysis in which we observed a 100% of sensitivity and a reduced expense. Although used for IRDs analysis, this methodology could be applied in other genetic diseases and could be an affordable option for many research laboratories.

FAMILY	USED METHODOLOGIES								
	Dominants panel	Pooled 316 IRD genes	MLPA				WES+CytoScan HD	CGH Arrays	
			Dominants	X-linked	USH2A	EYS		CytoXON	Custom made
RP1		✓							
RP8		✓							
RP11		✗					✗		
RP14		✗			✗				
RP15		✓							
RP16	✗	✗	✗						
RP17		✓							
RP20	✗	✗	✗						
RP22	✓								
RP23		✗						✗	
RP25		✓							
RP27	✗	✓							
RP28		✗		✗					
RP29		✗		✗					
RP30		✓					✓		
RP31		✗						✗	
RP34		✓							
RP35		✓							
RP36		✗	✗	✗				✗	✗
RP37	✓								
RP38		✗						✗	
RP39	✗	✗	✗					✗	
RP40	✗	✗	✓						
RP41		✗	✗						
RP42	✗	✗	✗						
RP43	✗	✗	✗						
RP48	✗	✗	✗						
RP49		✓							
RP50		✗		✗					
RP54		✗				✗			
RP55		✗						✗	
RP57		✓							
RP58		✗						✗	
RP59		✓							
RP60		✗						✗	
RP62		✗							
RP64	✓								
RP66	✗	✗			✗				
RP67		✓							
RP68		✗		✗					
RP69	✗	✗	✗						
RP71	✗	✗	✗					✓	✓
RP74		✗							
RP77		✓							
RP78		✗					✓		
RP79	✗	✗	✗						
RP80	✗	✗	✗						
RP83		✗		✗					
RP84		✗		✗					
RP85	✗	✗	✗						
RP87		✗							
RP88		✓							
RP90	✓								
RP91		✓							
RP92		✓							
RP93		✗		✗				✗	

RP94		✗			✗				
RP95		✗		✗			✗		
RP96		✗							
RP101	✓								
RP102	✓								
RP104		✗		✗			✗	✗	✗
RP105	✓								
RP106		✓							
RP107		✗		✗					
RP108		✗	✗				✗	✗	
RP109		✓					✓		
RP110	✗	✗	✗	✗					
RP111		✗							
RP112		✗		✗					
RP113	✓								
RP114		✗				✗		✗	✗
RP117		✓							
RP124		✗							
RP125		✗							
RP126		✗					✗	✗	
RP129		✗			✗			✗	
RP130		✗		✗			✗	✗	
RP132		✗							
RP133	✓								
RP134	✓								
RP135	✓								
RP136		✗							
RP138		✓							
RP139		✗							
RP141		✓							
RP142		✗			✗			✗	✗
RP143		✗						✗	
RP144		✗				✗	✗		
RP145		✗						✗	
RP146	✓								
RP147		✗							
RP148		✓							
RP149		✗		✗				✗	✗
RP152		✗							
RP153		✓							
RP154		✓							
RP156		✗							
RP157	✓								
RP159		✗							
RP161		✗							
RP162		✗			✗			✗	
RP163		✗						✗	
RP164		✗				✗			
RP165		✓							
RP166		✓							
RP167		✗			✗			✗	
RP169		✓							
RP170		✗		✗					
RP172		✗				✗			
RP173		✓							
RP174		✓							
RP175		✓							
RP176		✓							
RP177		✗		✗	✗				
RP178		✗				✗			
RP180		✓							
RP181		✓							
RP182		✓							

RP183		✗							
RP184		✗							
RP185		✓							
RP186		✗							
RP188		✓							
RP190		✗		✗					
RP191		✗							
RP192		✗							
RP193		✓							
RP194		✗							
RP195		✗		✗					
RP196		✓							
RP197		✗						✗	
RP198		✗							
RP199		✗							
RP200		✓							
RP201		✗							
RP202		✗							
RP203		✗		✗					
RP204		✗							
RP206		✓							
RP207		✗		✗					
RP208		✓							
RP209		✗							
RP210		✗							
RP211		✓							
RP212		✗	✗	✗					
RP213		✓							
RP214		✗							
RP215		✓							
RP216		✗	✗						
RP217		✓							
RP218		✗							
RP220		✗							
RP222		✗							
RP223		✗							
RP224		✓							
RP19S	✓								

Table 9: All patients analysed in this thesis: ✓ indicates that the patient was diagnosed with that technique. ✗ Indicates that the patient was analysed using that technique, but it was not solved. Solved patients are in bold and VUS cases are in italics.

FAMILY	GENE	TRANSCRIPT	cDNA CHANGE	PROTEIN CHANGE	REFERENCE
RP14	USH2A	NM_206933	c.11754G>A	p.Trp3918Ter	194
RP31	ABCA4	NM_000350	c.3113C>T	p.Ala1038Val	306
RP41	GPR98	NM_032119	c.853C>T	p.Arg285Cys	This study
RP54	EYS	NM_001142800	c.9405T>A	p.Tyr3135Ter	181
RP66	USH2A	NM_206933	c.2276G>T	p.Cys759Phe	
RP68	MYO7A	NM_000260	c.1325A>G	p.Glu442Gly	307
RP95	ABCA4	NM_000350	c.2023G>A	p.Val675Ile	308
RP96	CRB1	NM_201253	c.2843G>A	p.Cys948Tyr	309
RP104	ABCA4	NM_000350	c.3113C>T	p.Ala1038Val	306
RP106	ABCA4	NM_000350	c.3386G>T	p.Arg1129Leu	310
RP114	ABCA4	NM_000350	c.6148G>C	p.Val2050Leu	311
RP129	USH1C	NM_153676	c.1466G>A	p.Trp489Ter	This study
RP142	MYO7A	NM_000260	c.397C>T	p.His133Tyr	191
RP143	ABCA4	NM_000350	c.5908C>T	p.Leu1970Phe	312
RP162	EYS	NM_001142800	c.1830del	p.Asn611ThrfsTer32	This study
RP164	CNGA3	NM_001298	c.868C>T	p.Arg290Cys	This study
RP167	USH2A	NM_206933	c.908G>A	p.Arg303His	183
RP172	EYS	NM_001142800	c.6050G>T	p.Gly2017Val	60
RP177	USH2A	NM_206933	c.2276G>T	p.Cys759Phe	
RP178	EYS	NM_001142800	c.9405T>A	p.Tyr3135Ter	181
RP184	ABCA4	NM_000350	c.3386G>T	p.Arg1129Leu	310
RP192	CRB1	NM_201253	c.613_619del	p.Ile205AspfsTer13	309
RP201	ABCA4	NM_000350	c.3113C>T	p.Ala1038Val	306
RP203	PDE6A	NM_000440	c.1957C>T	p.Arg653Ter	195
RP209	ABCA4	NM_000350	c.3113C>T	p.Ala1038Val	306
RP210	ABCA4	NM_000350	c.455G>A	p.Arg152Gln	313
RP222	CNGA3	NM_001298	c.829C>T	p,Arg277Cys	205

Table 10: Patients with mutation in one allele in recessive genes.

CONCLUSIONS

1. Following our genetic mutation detection pipeline, we have been able to find the causal mutation of inherited retinal dystrophies in 61 out of 157 patients investigated.

2. Panel-based pooled NGS proved to be a powerful method for the molecular diagnosis of inherited retinal dystrophies and permitted the characterization of 95% of all patients diagnosed in this work.

3. Pool based NGS strategy demonstrated high sensitivity as it detected all control variants introduced.

4. NGS analysis of almost all genes known to cause inherited retinal dystrophies, allowed us to analyse all patients in parallel, regardless of their inheritance pattern, and to reclassify a group of our patients.

5. The analysis of LOH regions did not help in the reduction of the number of variants to be investigated in the WES analysis.

6. WES is a reliable method for mutations analysis and identification of new genes related to inherited retinal dystrophies.

7. Detection of p. Arg630Ter mutation in *SAMD11* gene strengthens the implication of this gene in retinitis pigmentosa disease.

8. *CytoScanXON* array has more sensibility, but less specificity than Agilent customized CGH array.

9. CNV analysis using MLPA and/or CGH array analysis permitted the diagnosis of two patients from the 62 patients analysed with at least one of these techniques giving a frequency of 3.2%.

10. The analysis of cDNA from hair roots might be a better option for the validation of mutations detected in deep intronic regions of *USH2A* gene, rather than for the discovery of mutations in these regions.

REFERENCES

1. Barber Cárcamo, A. M. & Ponz Piedrafita, F. *Neurofisiología*. (Síntesis, 1989). at <https://www.casadellibro.com/libro-neurofisiologia/9788477380702/37037>
2. Ramon y Cajal S. *Manual de histología normal y técnica micrográfica : Ramón y Cajal, Santiago, 1852-1934. n 50056869 : Free Download, Borrow, and Streaming : Internet Archive*. at <https://archive.org/details/b21779995>
3. Simó, R., Villarroel, M., Corraliza, L., Hernández, C. & Garcia-Ramírez, M. The retinal pigment epithelium: something more than a constituent of the blood-retinal barrier--implications for the pathogenesis of diabetic retinopathy. *J. Biomed. Biotechnol.* **2010,** 190724 (2010).
4. Chaya, T. *et al.* Versatile functional roles of horizontal cells in the retinal circuit. *Sci. Rep.* **7,** 5540 (2017).
5. Kolb, H. *Roles of Amacrine Cells*. *Webvision: The Organization of the Retina and Visual System* (University of Utah Health Sciences Center, 1995). at <http://www.ncbi.nlm.nih.gov/pubmed/21413397>
6. Reichenbach, A. & Bringmann, A. New functions of Müller cells. *Glia* **61,** 651–678 (2013).
7. Marshak, D. W. Retinal Ganglion Cells: Anatomy. *Encycl. Neurosci.* 211–218 (2009). doi:10.1016/B978-008045046-9.00897-4
8. Remington, L. A. *Clinical Anatomy and Physiology of the Visual System*. (Elsevier, 2012). doi:10.1016/C2009-0-56108-9
9. Bunt-Milam, A. H. & Saari, J. C. Immunocytochemical localization of two retinoid-binding proteins in vertebrate retina. *J. Cell Biol.* **97,** 703–12 (1983).
10. Jin, M., Li, S., Moghrabi, W. N., Sun, H. & Travis, G. H. Rpe65 Is the Retinoid Isomerase in Bovine Retinal Pigment Epithelium. *Cell* **122,** 449–459 (2005).
11. Ala-Laurila, P. *et al.* Visual Cycle: Dependence of Retinol Production and Removal on Photoproduct Decay and Cell Morphology. *J. Gen. Physiol.* **128,** 153–169 (2006).
12. Wong-Riley, M. T. T. Energy metabolism of the visual system. *Eye Brain* **2,** 99–116 (2010).
13. Abner G, Mario SP & Myhill C et al. EFA Global Monitoring Report Regional Overview-Monitoring the Education for All goals:Sub-saharan Africa. *J. Vis. Impair. Blind.* **1,** 1–14 (2002).
14. Hartong, D. T., Berson, E. L. & Dryja, T. P. Retinitis pigmentosa. *Lancet (London, England)* **368,** 1795–809 (2006).
15. Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M. & Bhattacharya, S. S. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.* **11,** 273–284 (2010).
16. Berger, W., Kloeckener-Gruissem, B. & Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.* **29,** 335–375 (2010).
17. Na, K.-H. *et al.* Prevalence, Age at Diagnosis, Mortality, and Cause of Death in Retinitis Pigmentosa in Korea-A Nationwide Population-based Study. *Am. J. Ophthalmol.* **176,** 157–165 (2017).
18. Xu, L., Hu, L., Ma, K., Li, J. & Jonas, J. B. Prevalence of retinitis pigmentosa in urban and rural adult Chinese: The Beijing Eye Study. *Eur. J. Ophthalmol.* **16,** 865–6
19. Sen, P. *et al.* Prevalence of retinitis pigmentosa in South Indian population aged above 40 years. *Ophthalmic Epidemiol.* **15,** 279–81 (2008).
20. Nangia, V., Jonas, J. B., Khare, A. & Sinha, A. Prevalence of retinitis pigmentosa in India: the Central India Eye and Medical Study. *Acta Ophthalmol.* **90,** e649-50 (2012).

21. Delgado Pelayo Sari. *Retinosis Pigmentaria. Preguntas y Respuestas*. *Revista Medica MD* **3,** (2007).
22. Hamel, C. Retinitis pigmentosa. *Orphanet J. Rare Dis.* **1,** 40 (2006).
23. Daiger, S. P., Bowne, S. J. & Sullivan, L. S. Perspective on genes and mutations causing retinitis pigmentosa. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **125,** 151–8 (2007).
24. Mustafi, D., Engel, A. H. & Palczewski, K. Structure of cone photoreceptors. *Prog. Retin. Eye Res.* **28,** 289–302 (2009).
25. Simunovic, M. P. *et al.* The Spectrum of CHM Gene Mutations in Choroideremia and Their Relationship to Clinical Phenotype. *Invest. Ophthalmol. Vis. Sci.* **57,** 6033–6039 (2016).
26. Michaelides, M., Hunt, D. M. & Moore, A. T. The genetics of inherited macular dystrophies. *J. Med. Genet.* **40,** 641–50 (2003).
27. Fujinami, K. *et al.* A longitudinal study of stargardt disease: clinical and electrophysiologic assessment, progression, and genotype correlations. *Am. J. Ophthalmol.* **155,** 1075–1088.e13 (2013).
28. Tanna, P., Strauss, R. W., Fujinami, K. & Michaelides, M. Stargardt disease: clinical features, molecular genetics, animal models and therapeutic options. *Br. J. Ophthalmol.* **101,** 25–30 (2017).
29. den Hollander, A. I., Roepman, R., Koenekoop, R. K. & Cremers, F. P. M. Leber congenital amaurosis: Genes, proteins and disease mechanisms. *Prog. Retin. Eye Res.* **27,** 391–419 (2008).
30. Koenekoop, R. K. An overview of Leber congenital amaurosis: a model to understand human retinal development. *Surv. Ophthalmol.* **49,** 379–98 (2004).
31. Stone, E. M. Leber congenital amaurosis - a model for efficient genetic testing of heterogeneous disorders: LXIV Edward Jackson Memorial Lecture. *Am. J. Ophthalmol.* **144,** 791–811 (2007).
32. Patrício, M. I., Barnard, A. R., Xue, K. & MacLaren, R. E. Choroideremia: molecular mechanisms and development of AAV gene therapy. *Expert Opin. Biol. Ther.* **18,** 807–820 (2018).
33. Barnard, A. R., Groppe, M. & MacLaren, R. E. Gene Therapy for Choroideremia Using an Adeno-Associated Viral (AAV) Vector. *Cold Spring Harb. Perspect. Med.* **5,** a017293–a017293 (2015).
34. Edwards, T. L., Groppe, M., Jolly, J. K., Downes, S. M. & MacLaren, R. E. Correlation of Retinal Structure and Function in Choroideremia Carriers. *Ophthalmology* **122,** 1274–1276 (2015).
35. Rosenberg, T., Haim, M., Hauch, A. M. & Parving, A. The prevalence of Usher syndrome and other retinal dystrophy-hearing impairment associations. *Clin. Genet.* **51,** 314–21 (1997).
36. Spandau, U. H. M. & Rohrschneider, K. Prevalence and geographical distribution of Usher syndrome in Germany. *Graefes Arch. Clin. Exp. Ophthalmol.* **240,** 495–8 (2002).
37. Blanco-Kelly, F. *et al.* Clinical aspects of Usher syndrome and the USH2A gene in a cohort of 433 patients. *JAMA Ophthalmol.* **133,** 157–64 (2015).
38. Lentz, J. & Keats, B. *Usher Syndrome Type II*. *GeneReviews®* (1993). at <http://www.ncbi.nlm.nih.gov/pubmed/20301515>
39. Herrera, W. *et al.* Retinal disease in Usher syndrome III caused by mutations in the clarin-1 gene. *Invest. Ophthalmol. Vis. Sci.* **49,** 2651–60 (2008).
40. Sankila, E. M. *et al.* Assignment of an Usher syndrome type III (USH3) gene to

chromosome 3q. *Hum. Mol. Genet.* **4,** 93–8 (1995).
41. Mockel, A. *et al.* Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies. *Prog. Retin. Eye Res.* **30,** 258–74 (2011).
42. Beales, P. L., Elcioglu, N., Woolf, A. S., Parker, D. & Flinter, F. A. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J. Med. Genet.* **36,** 437–46 (1999).
43. Klein, D. & Ammann, F. The syndrome of Laurence-Moon-Bardet-Biedl and allied diseases in Switzerland. Clinical, genetic and epidemiological studies. *J. Neurol. Sci.* **9,** 479–513
44. Farag, T. I. & Teebi, A. S. High incidence of Bardet Biedl syndrome among the Bedouin. *Clin. Genet.* **36,** 463–4 (1989).
45. Anasagasti, A., Irigoyen, C., Barandika, O., López de Munain, A. & Ruiz-Ederra, J. Current mutation discovery approaches in Retinitis Pigmentosa. *Vision Res.* **75,** 117–129 (2012).
46. Huang, X.-F. *et al.* Genotype-phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing. *Genet. Med.* **17,** 271–8 (2015).
47. Pérez-Carro, R. *et al.* Unravelling the pathogenic role and genotype-phenotype correlation of the USH2A p.(Cys759Phe) variant among Spanish families. *PLoS One* **13,** e0199048 (2018).
48. Kabir, F. *et al.* Loss of function mutations in RP1 are responsible for retinitis pigmentosa in consanguineous familial cases. *Mol. Vis.* **22,** 610–25 (2016).
49. Ferrari, S. *et al.* Retinitis pigmentosa: genes and disease mechanisms. *Curr. Genomics* **12,** 238–49 (2011).
50. Zudaire, C. Coeficiente de consanguinidad en zonas de Alava, Guipúzcoa y Vizcaya. *Aranzadi* 245–254 (1981). at <http://www.aranzadi-zientziak.org/fileadmin/docs/Munibe/1981245254.pdf>
51. Massó, J. F. M., Zarranz, J. J., Otaegui, D. & López de Munain, A. Neurogenetic Disorders in the Basque Population. *Ann. Hum. Genet.* **79,** 57–75 (2015).
52. Abu-Safieh, L. *et al.* A large deletion in the adRP gene PRPF31: evidence that haploinsufficiency is the cause of disease. *Mol. Vis.* **12,** 384–8 (2006).
53. Steele-Stallard, H. B. *et al.* Screening for duplications, deletions and a common intronic mutation detects 35% of second mutations in patients with USH2A monoallelic mutations on Sanger sequencing. *Orphanet J. Rare Dis.* **8,** 122 (2013).
54. Mukkamala, K., Gentile, R. C., Willner, J. & Tsang, S. Choroideremia in a woman with ectodermal dysplasia and complex translocations involving chromosomes X, 1, and 3. *Ophthalmic Genet.* **31,** 178–82 (2010).
55. Martin-Merida, I. *et al.* Analysis of the PRPF31 Gene in Spanish Autosomal Dominant Retinitis Pigmentosa Patients: A Novel Genomic Rearrangement. *Invest. Ophthalmol. Vis. Sci.* **58,** 1045–1053 (2017).
56. Daiger, S. P. *et al.* in *Advances in experimental medicine and biology* **801,** 123–129 (2014).
57. Francis, P. J. Genetics of inherited retinal disease. *J. R. Soc. Med.* **99,** 189–91 (2006).
58. Ali, M. U., Rahman, M. S. U., Cao, J. & Yuan, P. X. Genetic characterization and disease mechanism of retinitis pigmentosa; current scenario. *3 Biotech* **7,** 251 (2017).
59. Iwanami, M., Oshikawa, M., Nishida, T., Nakadomari, S. & Kato, S. High Prevalence of Mutations in the *EYS* Gene in Japanese Patients with Autosomal Recessive Retinitis

Pigmentosa. *Investig. Opthalmology Vis. Sci.* **53,** 1033 (2012).
60. Barragán, I. *et al.* Mutation spectrum of EYS in Spanish patients with autosomal recessive retinitis pigmentosa. *Hum. Mutat.* **31,** E1772-800 (2010).
61. Pelletier, V. *et al.* Comprehensive survey of mutations in RP2 and RPGR in patients affected with distinct retinal dystrophies: genotype-phenotype correlations and impact on genetic counseling. *Hum. Mutat.* **28,** 81–91 (2007).
62. Talib, M. *et al.* The Spectrum of Structural and Functional Abnormalities in Female Carriers of Pathogenic Variants in the *RPGR* Gene. *Investig. Opthalmology Vis. Sci.* **59,** 4123 (2018).
63. Pomares, E. *et al.* Identification of an Intronic Single-Point Mutation in *RP2* as the Cause of Semidominant X-linked Retinitis Pigmentosa. *Investig. Opthalmology Vis. Sci.* **50,** 5107 (2009).
64. Branham, K. *et al.* Mutations in RPGR and RP2 account for 15% of males with simplex retinal degenerative disease. *Invest. Ophthalmol. Vis. Sci.* **53,** 8232–7 (2012).
65. Li, J. *et al.* Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. *J. Mol. Diagn.* **18,** 817–824 (2016).
66. Paskowitz, D. M., LaVail, M. M. & Duncan, J. L. Light and inherited retinal degeneration. *Br. J. Ophthalmol.* **90,** 1060–6 (2006).
67. Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M. & Bhattacharya, S. S. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.* **11,** 273–84 (2010).
68. Renner, A. B. *et al.* Phenotypic variability and long-term follow-up of patients with known and novel PRPH2/RDS gene mutations. *Am. J. Ophthalmol.* **147,** 518–530.e1 (2009).
69. Tschernutter, M. *et al.* Clinical characterisation of a family with retinal dystrophy caused by mutation in the Mertk gene. *Br. J. Ophthalmol.* **90,** 718–23 (2006).
70. Saltzman, A. L., Pan, Q. & Blencowe, B. J. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* **25,** 373–84 (2011).
71. Yuan, L., Kawada, M., Havlioglu, N., Tang, H. & Wu, J. Y. Mutations in PRPF31 inhibit pre-mRNA splicing of rhodopsin gene and cause apoptosis of retinal cells. *J. Neurosci.* **25,** 748–57 (2005).
72. Pleiss, J. A., Whitworth, G. B., Bergkessel, M. & Guthrie, C. Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol.* **5,** e90 (2007).
73. WATSON, J. D. & CRICK, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171,** 964–7 (1953).
74. WATSON, J. D. & CRICK, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171,** 737–8 (1953).
75. NIRENBERG, M. W. & MATTHAEI, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **47,** 1588–602 (1961).
76. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–3 (1970).
77. Kelly, T. J. & Smith, H. O. A restriction enzyme from Hemophilus influenzae. II. *J. Mol. Biol.* **51,** 393–409 (1970).
78. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. *Biotechnology* **24,** 122–39 (1992).
79. Saiki, R. K. *et al.* Enzymatic amplification of beta-globin genomic sequences and

restriction site analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–4 (1985).
80. Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94,** 441–8 (1975).
81. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–7 (1977).
82. Broadgate, S., Yu, J., Downes, S. M. & Halford, S. Unravelling the genetics of inherited retinal dystrophies: Past, present and future. *Prog. Retin. Eye Res.* **59,** 53–96 (2017).
83. Carothers, A. M., Urlaub, G., Mucha, J., Grunberger, D. & Chasin, L. A. Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method. *Biotechniques* **7,** 494–6, 498–9 (1989).
84. Metzker, M. L. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11,** 31–46 (2010).
85. Dong, Y. & Zhu, H. Single-strand conformational polymorphism analysis: basic principles and routine practice. *Methods Mol. Med.* **108,** 149–57 (2005).
86. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2766–70 (1989).
87. Reed, G. H., Kent, J. O. & Wittwer, C. T. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics* **8,** 597–608 (2007).
88. Anasagasti, A. *et al.* Genetic high throughput screening in Retinitis Pigmentosa based on high resolution melting (HRM) analysis. *Exp. Eye Res.* **116,** 386–394 (2013).
89. Cui, G., Ding, H., Xu, Y., Li, B. & Wang, D. W. Applications of the method of high resolution melting analysis for diagnosis of Leber's disease and the three primary mutation spectrum of LHON in the Han Chinese population. *Gene* **512,** 108–112 (2013).
90. HARLÉ, A. *et al.* Analysis of PIK3CA exon 9 and 20 mutations in breast cancers using PCR-HRM and PCR-ARMS: Correlation with clinicopathological criteria. *Oncol. Rep.* **29,** 1043–1052 (2013).
91. Jaakson, K. *et al.* Genotyping microarray (gene chip) for the *ABCR* (*ABCA4*) gene. *Hum. Mutat.* **22,** 395–403 (2003).
92. Cremers, F. P. M. *et al.* Development of a genotyping microarray for Usher syndrome. *J. Med. Genet.* **44,** 153–160 (2006).
93. Ávila-Fernández, A. *et al.* Mutation analysis of 272 Spanish families affected by autosomal recessive retinitis pigmentosa using a genotyping microarray. *Mol. Vis.* **16,** 2550–8 (2010).
94. Blanco-Kelly, F. *et al.* Genotyping microarray: mutation screening in Spanish families with autosomal dominant retinitis pigmentosa. *Mol. Vis.* (2012).
95. Barandika, O. *et al.* A Cost-Effective Mutation Screening Strategy for Inherited Retinal Dystrophies. *Ophthalmic Res.* **56,** 123–131 (2016).
96. Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Arch. Dis. Child. Educ. Pract. Ed.* **98,** 236–8 (2013).
97. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8,** 175–85 (1998).
98. Levy, S. *et al.* The Diploid Genome Sequence of an Individual Human. *PLoS Biol.* **5,** e254 (2007).
99. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452,** 872–876 (2008).

100. Balzer, S., Malde, K. & Jonassen, I. Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* **27,** i304-9 (2011).
101. Anasagasti, A., Irigoyen, C., Barandika, O., López de Munain, A. & Ruiz-Ederra, J. Current mutation discovery approaches in Retinitis Pigmentosa. *Vision Res.* **75,** 117–129 (2012).
102. Pettersson, E., Lundeberg, J. & Ahmadian, A. Generations of sequencing technologies. *Genomics* **93,** 105–111 (2009).
103. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).
104. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17,** 333–351 (2016).
105. Harismendy, O. *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10,** R32 (2009).
106. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–9 (2008).
107. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456,** 66–72 (2008).
108. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30,** 434–439 (2012).
109. Fu, Q. *et al.* Next-generation sequencing-based molecular diagnosis of a Chinese patient cohort with autosomal recessive retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **54,** 4158–66 (2013).
110. Perez-Carro, R. *et al.* Erratum: Corrigendum: Panel-based NGS Reveals Novel Pathogenic Mutations in Autosomal Recessive Retinitis Pigmentosa. *Sci. Rep.* **6,** 24843 (2016).
111. Fernandez-San Jose, P. *et al.* Targeted Next-Generation Sequencing Improves the Diagnosis of Autosomal Dominant Retinitis Pigmentosa in Spanish Patients. *Invest. Ophthalmol. Vis. Sci.* **56,** 2173–82 (2015).
112. Ezquerra-Inchausti, M. *et al.* High prevalence of mutations affecting the splicing process in a Spanish cohort with autosomal dominant retinitis pigmentosa. *Sci. Rep.* **7,** (2017).
113. Glöckle, N. *et al.* Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur. J. Hum. Genet.* **22,** 99–104 (2014).
114. Ge, Z. *et al.* NGS-based Molecular diagnosis of 105 eyeGENE(®) probands with Retinitis Pigmentosa. *Sci. Rep.* **5,** 18287 (2015).
115. Jones, K. D. *et al.* Next-generation sequencing to solve complex inherited retinal dystrophy: A case series of multiple genes contributing to disease in extended families. *Mol. Vis.* **23,** 470–481 (2017).
116. Eisenberger, T. *et al.* Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One* **8,** e78496 (2013).
117. 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.* **20,** 553–60 (2014).
118. Worthey, E. A. *et al.* Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease.

Genet. Med. **13,** 255–62 (2011).
119. Corton, M. *et al.* Identification of the Photoreceptor Transcriptional Co-Repressor SAMD11 as Novel Cause of Autosomal Recessive Retinitis Pigmentosa. *Sci. Rep.* **6,** 35370 (2016).
120. Weisz Hubshman, M. *et al.* Whole-exome sequencing reveals POC5 as a novel gene associated with autosomal recessive retinitis pigmentosa. *Hum. Mol. Genet.* **27,** 614–624 (2018).
121. de Bruijn, S. E. *et al.* Homozygous variants in *KIAA1549* , encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa. *J. Med. Genet.* **55,** 705–712 (2018).
122. Chiang, J. P. & Trzupek, K. The current status of molecular diagnosis of inherited retinal dystrophies. *Curr. Opin. Ophthalmol.* **26,** 346–51 (2015).
123. Wright, C. F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet (London, England)* **385,** 1305–14 (2015).
124. Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30,** e57 (2002).
125. Stuppia, L., Antonucci, I., Palka, G. & Gatta, V. Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases. *Int. J. Mol. Sci.* **13,** 3245–3276 (2012).
126. Gorringe, K. L. in *eLS* 1–8 (John Wiley & Sons, Ltd, 2016). doi:10.1002/9780470015902.a0026643
127. Woods, C. G. *et al.* Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am. J. Hum. Genet.* **78,** 889–896 (2006).
128. Dryja, T. P., Hahn, L. B., Kajiwara, K. & Berson, E. L. Dominant and digenic mutations in the peripherin/RDS and ROM1 genes in retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **38,** 1972–82 (1997).
129. Mansergh, F. C. *et al.* Retinitis pigmentosa and progressive sensorineural hearing loss caused by a C12258A mutation in the mitochondrial MTTS2 gene. *Am. J. Hum. Genet.* **64,** 971–85 (1999).
130. Daiger, S. P., Bowne, S. J. & Sullivan, L. S. Genes and Mutations Causing Autosomal Dominant Retinitis Pigmentosa. *Cold Spring Harb. Perspect. Med.* **5,** a017129 (2015).
131. Wang, F. *et al.* A missense mutation in HK1 leads to autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **55,** 7159–64 (2014).
132. Ma, X. *et al.* Whole-exome sequencing identifies OR2W3 mutation as a cause of autosomal dominant retinitis pigmentosa. *Sci. Rep.* **5,** 9236 (2015).
133. Liu, Y. *et al.* SPP2 Mutations Cause Autosomal Dominant Retinitis Pigmentosa. *Sci. Rep.* **5,** 14867 (2015).
134. Chen, X. *et al.* PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **23,** 2926–2939 (2014).
135. Dewey, F. E. *et al.* Clinical interpretation and implications of whole-genome sequencing. *JAMA* **311,** 1035–45 (2014).
136. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–18 (2009).
137. Havens, M. A., Duelli, D. M. & Hastings, M. L. Targeting RNA splicing for disease therapy. *Wiley Interdiscip. Rev. RNA* **4,** 247–266 (2013).
138. Liu, M. M. & Zack, D. J. Alternative splicing and retinal degeneration. *Clin. Genet.* **84,**

142–149 (2013).
139. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17,** 19–32 (2016).
140. Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* **18,** 472–482 (2012).
141. Cao, H. *et al.* Temporal and tissue specific regulation of RP-associated splicing factor genes PRPF3, PRPF31 and PRPC8--implications in the pathogenesis of RP. *PLoS One* **6,** e15860 (2011).
142. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7,** e46688 (2012).
143. Mathe, E. *et al.* Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res.* **34,** 1317–25 (2006).
144. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7,** 575–6 (2010).
145. Alonso, R. *et al.* Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic Acids Res.* **43,** W117-21 (2015).
146. Sohocki, M. M. *et al.* Prevalence of mutations causing retinitis pigmentosa and other inherited retinopathies. *Hum. Mutat.* **17,** 42–51 (2001).
147. Zhao, C. *et al.* A novel locus (RP33) for autosomal dominant retinitis pigmentosa mapping to chromosomal region 2cen-q12.1. *Hum. Genet.* **119,** 617–23 (2006).
148. Zhao, C. *et al.* Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs. *Am. J. Hum. Genet.* **85,** 617–27 (2009).
149. Martínez-Gimeno, M. *et al.* Mutations in the pre-mRNA splicing-factor genes PRPF3, PRPF8, and PRPF31 in Spanish families with autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **44,** 2171–7 (2003).
150. Bell, C., Converse, C. A., Hammer, H. M., Osborne, A. & Haites, N. E. Rhodopsin mutations in a Scottish retinitis pigmentosa population, including a novel splice site mutation in intron four. *Br. J. Ophthalmol.* **78,** 933–8 (1994).
151. Hollingsworth, T. J. & Gross, A. K. The severe autosomal dominant retinitis pigmentosa rhodopsin mutant Ter349Glu mislocalizes and induces rapid rod cell death. *J. Biol. Chem.* **288,** 29047–55 (2013).
152. Keen, T. J. *et al.* Autosomal dominant retinitis pigmentosa: four new mutations in rhodopsin, one of them in the retinal attachment site. *Genomics* **11,** 199–205 (1991).
153. Bowne, S. J. *et al.* Identification of Disease-Causing Mutations in Autosomal Dominant Retinitis Pigmentosa (adRP) Using Next-Generation DNA Sequencing. *Investig. Opthalmology Vis. Sci.* **52,** 494 (2011).
154. Audo, I. *et al.* Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J. Rare Dis.* **7,** 8 (2012).
155. Sullivan, L. S. *et al.* A Dominant Mutation in Hexokinase 1 (*HK1*) Causes Retinitis Pigmentosa. *Investig. Opthalmology Vis. Sci.* **55,** 7147 (2014).
156. Oishi, M. *et al.* Comprehensive molecular diagnosis of a large cohort of Japanese retinitis pigmentosa and Usher syndrome patients by next-generation sequencing. *Invest. Ophthalmol. Vis. Sci.* **55,** 7369–75 (2014).
157. González-del Pozo, M. *et al.* Mutation screening of multiple genes in Spanish patients

with autosomal recessive retinitis pigmentosa by targeted resequencing. *PLoS One* **6,** e27894 (2011).
158. Alfonso-Sanchez, M. A., Peña, J. A., Aresti, U. & Calderón, R. An insight into recent consanguinity within the Basque area in Spain. Effects of autochthony, industrialization and demographic changes. *Ann. Hum. Biol.* **28,** 505–21
159. Alfonso-Sánchez, M. A., Aresti, U., Peña, J. A. & Calderón, R. Inbreeding levels and consanguinity structure in the Basque province of Guipúzcoa (1862-1980). *Am. J. Phys. Anthropol.* **127,** 240–252 (2005).
160. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8,** 847–55 (1998).
161. Růžičková, Š. & Staněk, D. Mutations in spliceosomal proteins and retina degeneration. *RNA Biol.* **14,** 544–552 (2017).
162. Kim, D. H. & Rossi, J. J. The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. *RNA* **5,** 959–71 (1999).
163. Zhang, L. *et al.* Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. *Nat. Struct. Mol. Biol.* **16,** 731–9 (2009).
164. Cvačková, Z., Matějů, D. & Staněk, D. Retinitis pigmentosa mutations of SNRNP200 enhance cryptic splice-site recognition. *Hum. Mutat.* **35,** 308–17 (2014).
165. Coussa, R. G. *et al.* Genotype and Phenotype Studies in Autosomal Dominant Retinitis Pigmentosa (adRP) of the French Canadian Founder Population. *Invest. Ophthalmol. Vis. Sci.* **56,** 8297–305 (2015).
166. Sullivan, L. S. *et al.* Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. *Invest. Ophthalmol. Vis. Sci.* **47,** 3052–64 (2006).
167. Benaglio, P. *et al.* Next generation sequencing of pooled samples reveals new SNRNP200 mutations associated with retinitis pigmentosa. *Hum. Mutat.* **32,** E2246-58 (2011).
168. Bowne, S. J. *et al.* Mutations in the small nuclear riboprotein 200 kDa gene (SNRNP200) cause 1.6% of autosomal dominant retinitis pigmentosa. *Mol. Vis.* **19,** 2407–17 (2013).
169. Nakajima, H. *et al.* New antitumor substances, FR901463, FR901464 and FR901465. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. *J. Antibiot. (Tokyo).* **49,** 1196–203 (1996).
170. Glaus, E., Schmid, F., Da Costa, R., Berger, W. & Neidhardt, J. Gene therapeutic approach using mutation-adapted U1 snRNA to correct a RPGR splice defect in patient-derived cells. *Mol. Ther.* **19,** 936–41 (2011).
171. Berger, A. *et al.* Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: a new approach for autosomal dominant retinitis pigmentosa. *Mol. Ther.* **23,** 918–930 (2015).
172. Van Cauwenbergh, C. *et al.* arrEYE: a customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genet. Med.* **19,** 457–466 (2017).
173. Sohocki, M. M. *et al.* Prevalence of AIPL1 Mutations in Inherited Retinal Degenerative Disease. *Mol. Genet. Metab.* **70,** 142–150 (2000).
174. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
175. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science (80-.).* **291,** 1304–1351 (2001).

176. Lazaridis, K. N. *et al.* Outcome of Whole Exome Sequencing for Diagnostic Odyssey Cases of an Individualized Medicine Clinic: The Mayo Clinic Experience. *Mayo Clin. Proc.* **91,** 297–307 (2016).
177. Ezquerra-Inchausti, M. *et al.* High prevalence of mutations affecting the splicing process in a Spanish cohort with autosomal dominant retinitis pigmentosa. *Sci. Rep.* **7,** 39652 (2017).
178. Sullivan, L. S. *et al.* Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **47,** 4579–88 (2006).
179. Pieras, J. I. *et al.* Copy-number variations in EYS: a significant event in the appearance of arRP. *Invest. Ophthalmol. Vis. Sci.* **52,** 5625–31 (2011).
180. Kozlowski, P. *et al.* Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. *Hum. Genet.* **121,** 389–400 (2007).
181. Collin, R. W. J. *et al.* Identification of a 2 Mb human ortholog of Drosophila eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. *Am. J. Hum. Genet.* **83,** 594–603 (2008).
182. Tuson, M., Marfany, G. & Gonzàlez-Duarte, R. Mutation of CERKL, a Novel Human Ceramide Kinase Gene, Causes Autosomal Recessive Retinitis Pigmentosa (RP26). *Am. J. Hum. Genet.* **74,** 128–138 (2004).
183. Garcia-Garcia, G. *et al.* Mutational screening of the USH2A gene in Spanish USH patients reveals 23 novel pathogenic mutations. *Orphanet J. Rare Dis.* **6,** 65 (2011).
184. Sanchez-Alcudia, R. *et al.* A Comprehensive Analysis of Choroideremia: From Genetic Characterization to Clinical Practice. *PLoS One* **11,** e0151943 (2016).
185. den Hollander, A. I. *et al.* Mutations in a human homologue of Drosophila crumbs cause retinitis pigmentosa (RP12). *Nat. Genet.* **23,** 217–21 (1999).
186. Bujakowska, K. *et al.* CRB1 mutations in inherited retinal dystrophies. *Hum. Mutat.* **33,** 306–15 (2012).
187. Avila-Fernandez, A. *et al.* Identification of an RP1 prevalent founder mutation and related phenotype in Spanish patients with early-onset autosomal recessive retinitis. *Ophthalmology* **119,** 2616–21 (2012).
188. Rivolta, C., Sweklo, E. A., Berson, E. L. & Dryja, T. P. Missense Mutation in the USH2A Gene: Association with Recessive Retinitis Pigmentosa without Hearing Loss. *Am. J. Hum. Genet.* **66,** 1975–1978 (2000).
189. Corton, M. *et al.* Correction: Exome Sequencing of Index Patients with Retinal Dystrophies as a Tool for Molecular Diagnosis. *PLoS One* **11,** e0153121 (2016).
190. Banerjee, P. *et al.* TULP1 mutation in two extended Dominican kindreds with autosomal recessive retinitis pigmentosa. *Nat. Genet.* **18,** 177–9 (1998).
191. Roux, A.-F. *et al.* Four-year follow-up of diagnostic service in USH1 patients. *Invest. Ophthalmol. Vis. Sci.* **52,** 4063–71 (2011).
192. Paloma, E. *et al.* Novel homozygous mutation in the alpha subunit of the rod cGMP gated channel (CNGA1) in two Spanish sibs affected with autosomal recessive retinitis pigmentosa. *J. Med. Genet.* **39,** E66 (2002).
193. Jaijo, T. *et al.* MYO7A mutation screening in Usher syndrome type I patients from diverse origins. *J. Med. Genet.* **44,** e71 (2007).
194. Baux, D. *et al.* Enrichment of LOVD-USHbases with 152 USH2A genotypes defines an extensive mutational spectrum and highlights missense hotspots. *Hum. Mutat.* **35,** 1179–86 (2014).

195. Carss, K. J. *et al.* Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am. J. Hum. Genet.* **100,** 75–90 (2017).
196. Aller, E. *et al.* Genetic analysis of 2299delG and C759F mutations (USH2A) in patients with visual and/or auditory impairments. *Eur. J. Hum. Genet.* **12,** 407–410 (2004).
197. Fujinami, K. *et al.* ABCA4 gene screening by next-generation sequencing in a British cohort. *Invest. Ophthalmol. Vis. Sci.* **54,** 6662–74 (2013).
198. Bonnet, C. *et al.* An innovative strategy for the molecular diagnosis of Usher syndrome identifies causal biallelic mutations in 93% of European patients. *Eur. J. Hum. Genet.* **24,** 1730–1738 (2016).
199. Weisschuh, N. *et al.* Mutation Detection in Patients with Retinal Dystrophies Using Targeted Next Generation Sequencing. *PLoS One* **11,** e0145951 (2016).
200. Gerber, S. *et al.* The photoreceptor cell-specific nuclear receptor gene (PNR) accounts for retinitis pigmentosa in the Crypto-Jews from Portugal (Marranos), survivors from the Spanish Inquisition. *Hum. Genet.* **107,** 276–84 (2000).
201. Morimura, H., Saindelle-Ribeaudeau, F., Berson, E. L. & Dryja, T. P. Mutations in RGR, encoding a light-sensitive opsin homologue, in patients with retinitis pigmentosa. *Nat. Genet.* **23,** 393–4 (1999).
202. Peng, C., Rich, E. D. & Varnum, M. D. Achromatopsia-associated mutation in the human cone photoreceptor cyclic nucleotide-gated channel CNGB3 subunit alters the ligand sensitivity and pore properties of heteromeric channels. *J. Biol. Chem.* **278,** 34533–40 (2003).
203. Dryja, T. P., Rucinski, D. E., Chen, S. H. & Berson, E. L. Frequency of mutations in the gene encoding the alpha subunit of rod cGMP-phosphodiesterase in autosomal recessive retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **40,** 1859–65 (1999).
204. Kohl, S. *et al.* Total colourblindness is caused by mutations in the gene encoding the α-subunit of the cone photoreceptor cGMP-gated cation channel. *Nat. Genet.* **19,** 257–259 (1998).
205. Wissinger, B. *et al.* CNGA3 Mutations in Hereditary Cone Photoreceptor Disorders. *Am. J. Hum. Genet.* **69,** 722–737 (2001).
206. Lewis, R. A. *et al.* Genotype/Phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *Am. J. Hum. Genet.* **64,** 422–34 (1999).
207. Riveiro-Alvarez, R. *et al.* Frequency of ABCA4 mutations in 278 Spanish controls: an insight into the prevalence of autosomal recessive Stargardt disease. *Br. J. Ophthalmol.* **93,** 1359–64 (2009).
208. Mykytyn, K. *et al.* Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat. Genet.* **31,** 435–438 (2002).
209. Vallespin, E. *et al.* Gene symbol: CRB1. Disease: Leber congenital amaurosis. Accession #Hm0540. *Hum. Genet.* **118,** 778 (2006).
210. Corton, M. *et al.* High frequency of CRB1 mutations as cause of Early-Onset Retinal Dystrophies in the Spanish population. *Orphanet J. Rare Dis.* **8,** 20 (2013).
211. Pierrache, L. H. M. *et al.* Visual Prognosis in USH2A-Associated Retinitis Pigmentosa Is Worse for Patients with Usher Syndrome Type IIa Than for Those with Nonsyndromic Retinitis Pigmentosa. *Ophthalmology* **123,** 1151–60 (2016).
212. Aller, E. *et al.* Identification of 14 novel mutations in the long isoform of USH2A in

Spanish patients with Usher syndrome type II. *J. Med. Genet.* **43,** e55 (2006).
213. Weleber, R. G., Francis, P. J., Trzupek, K. M. & Beattie, C. *Leber Congenital Amaurosis*. *GeneReviews®* (1993). at <http://www.ncbi.nlm.nih.gov/pubmed/20301475>
214. Bergen, A. A. *et al.* Mutations in ABCC6 cause pseudoxanthoma elasticum. *Nat. Genet.* **25,** 228–31 (2000).
215. Ben-Yosef, T. *et al.* A Mutation of *PCDH15* among Ashkenazi Jews with the Type 1 Usher Syndrome. *N. Engl. J. Med.* **348,** 1664–1670 (2003).
216. Zheng, Q. Y. *et al.* Digenic inheritance of deafness caused by mutations in genes encoding cadherin 23 and protocadherin 15 in mice and humans. *Hum. Mol. Genet.* **14,** 103–11 (2005).
217. Sahly, I. *et al.* Localization of Usher 1 proteins to the photoreceptor calyceal processes, which are absent from mice. *J. Cell Biol.* **199,** 381–99 (2012).
218. Van Tassell, C. P. *et al.* SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* **5,** 247–52 (2008).
219. Margraf, R. L. *et al.* Variant identification in multi-sample pools by illumina genome analyzer sequencing. *J. Biomol. Tech.* **22,** 74–84 (2011).
220. Out, A. A. *et al.* Deep sequencing to reveal new variants in pooled DNA samples. *Hum. Mutat.* **30,** 1703–12 (2009).
221. Zhu, Y., Bergland, A. O., González, J. & Petrov, D. A. Empirical validation of pooled whole genome population re-sequencing in Drosophila melanogaster. *PLoS One* **7,** e41901 (2012).
222. Shaw, S. H., Carrasquillo, M. M., Kashuk, C., Puffenberger, E. G. & Chakravarti, A. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. *Genome Res.* **8,** 111–23 (1998).
223. Druley, T. E. *et al.* Quantification of rare allelic variants from pooled genomic DNA. *Nat. Methods* **6,** 263–5 (2009).
224. Wang, J. *et al.* Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. *Sci. Rep.* **6,** 33256 (2016).
225. Tavira, B. *et al.* A labor- and cost-effective non-optical semiconductor (Ion Torrent) next-generation sequencing of the SLC12A3 and CLCNKA/B genes in Gitelman's syndrome patients. *J. Hum. Genet.* **59,** 376–80 (2014).
226. Anand, S. *et al.* Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. *Sci. Rep.* **6,** 33735 (2016).
227. Jin, S. C., Benitez, B. A., Deming, Y. & Cruchaga, C. Pooled-DNA Sequencing for Elucidating New Genomic Risk Factors, Rare Variants Underlying Alzheimer's Disease. *Methods Mol. Biol.* **1303,** 299–314 (2016).
228. Jin, S. C. *et al.* Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. *Alzheimers. Res. Ther.* **4,** 34 (2012).
229. Consugar, M. B. *et al.* Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.* **17,** 253–261 (2015).
230. Zhang, Q. *et al.* Next-generation sequencing-based molecular diagnosis of 35 Hispanic retinitis pigmentosa probands. *Sci. Rep.* **6,** 32792 (2016).
231. Carrigan, M. *et al.* Panel-Based Population Next-Generation Sequencing for Inherited Retinal Degenerations. *Sci. Rep.* **6,** 33248 (2016).
232. Tiwari, A. *et al.* Next generation sequencing based identification of disease-associated

mutations in Swiss patients with retinal dystrophies. *Sci. Rep.* **6,** 28755 (2016).
233. Marelli, C. *et al.* Mini-Exome Coupled to Read-Depth Based Copy Number Variation Analysis in Patients with Inherited Ataxias. *Hum. Mutat.* **37,** 1340–1353 (2016).
234. Bujakowska, K. M. *et al.* Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. *Genet. Med.* **19,** 643–651 (2017).
235. Khateb, S. *et al.* Identification of genomic deletions causing inherited retinal degenerations by coverage analysis of whole exome sequencing data. *J. Med. Genet.* **53,** 600–7 (2016).
236. 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.* **20,** 202–213 (2018).
237. Ellingford, J. M. *et al.* Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *J. Med. Genet.* **55,** 114–121 (2018).
238. Saini, S., Robinson, P. N., Singh, J. R. & Vanita, V. A novel 7 bp deletion in PRPF31 associated with autosomal dominant retinitis pigmentosa with incomplete penetrance in an Indian family. *Exp. Eye Res.* **104,** 82–8 (2012).
239. Rose, A. M. & Bhattacharya, S. S. Variant haploinsufficiency and phenotypic non-penetrance in PRPF31-associated retinitis pigmentosa. *Clin. Genet.* **90,** 118–26 (2016).
240. Rose, A. M. *et al.* Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. *Sci. Rep.* **6,** 19450 (2016).
241. den Hollander, A. I. *et al.* Mutations in the CEP290 (NPHP6) Gene Are a Frequent Cause of Leber Congenital Amaurosis. *Am. J. Hum. Genet.* **79,** 556–561 (2006).
242. Vaché, C. *et al.* Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. *Hum. Mutat.* **33,** 104–8 (2012).
243. Slijkerman, R. W. *et al.* Antisense Oligonucleotide-based Splice Correction for USH2A-associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. *Mol. Ther. Nucleic Acids* **5,** e381 (2016).
244. Espinós, C., Millán, J. M., Beneyto, M. & Nájera, C. Epidemiology of Usher Syndrome in Valencia and Spain. *Public Health Genomics* **1,** 223–228 (1998).
245. Bocquet, B. *et al.* Relative frequencies of inherited retinal dystrophies and optic neuropathies in Southern France: assessment of 21-year data management. *Ophthalmic Epidemiol.* **20,** 13–25 (2013).
246. Yan, D. *et al.* Spectrum of DNA variants for non-syndromic deafness in a large cohort from multiple continents. *Hum. Genet.* **135,** 953–61 (2016).
247. Daiger, S. P., Sullivan, L. S. & Bowne, S. J. Genes and mutations causing retinitis pigmentosa. *Clinical Genetics* **84,** 132–141 (2013).
248. Ezquerra-Inchausti, M. *et al.* A new approach based on targeted pooled DNA sequencing identifies novel mutations in patients with Inherited Retinal Dystrophies. *Sci. Rep.* **8,** 15457 (2018).
249. Zhu, F. *et al.* Mutations in PMFBP1 Cause Acephalic Spermatozoa Syndrome. *Am. J. Hum. Genet.* **103,** 188–199 (2018).
250. Roosing, S. *et al.* Disruption of the basal body protein POC1B results in autosomal-recessive cone-rod dystrophy. *Am. J. Hum. Genet.* **95,** 131–42 (2014).
251. Tatour, Y. *et al.* Mutations in SCAPER cause autosomal recessive retinitis pigmentosa with intellectual disability. *J. Med. Genet.* **54,** 698–704 (2017).

252. Lo, K. C. *et al.* Comprehensive analysis of loss of heterozygosity events in glioblastoma using the 100K SNP mapping arrays and comparison with copy number abnormalities defined by BAC array comparative genomic hybridization. *Genes. Chromosomes Cancer* **47,** 221–37 (2008).
253. El-Asrag, M. E. *et al.* Biallelic Mutations in the Autophagy Regulator DRAM2 Cause Retinal Dystrophy with Early Macular Involvement. *Am. J. Hum. Genet.* **96,** 948–954 (2015).
254. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–303 (2010).
255. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** 122 (2016).
256. Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat. Rev. Genet.* **12,** 628–640 (2011).
257. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).
258. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11,** 1–9 (2015).
259. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).
260. Almoguera, B. *et al.* Application of Whole Exome Sequencing in Six Families with an Initial Diagnosis of Autosomal Dominant Retinitis Pigmentosa: Lessons Learned. *PLoS One* **10,** e0133624 (2015).
261. Vaags, A. K. *et al.* Absent CNKSR2 causes seizures and intellectual, attention, and language deficits. *Ann. Neurol.* **76,** 758–64 (2014).
262. Lanigan, T. M. *et al.* Human homologue of Drosophila CNK interacts with Ras effector proteins Raf and Rlf. *FASEB J.* **17,** 2048–60 (2003).
263. Inoue, T. *et al.* Cloning and characterization of mr-s, a novel SAM domain protein, predominantly expressed in retinal photoreceptor cells. *BMC Dev. Biol.* **6,** 15 (2006).
264. Hlawatsch, J. *et al.* Sterile Alpha Motif Containing 7 (Samd7) Is a Novel Crx-Regulated Transcriptional Repressor in the Retina. *PLoS One* **8,** e60633 (2013).
265. Schultz, J., Ponting, C. P., Hofmann, K. & Bork, P. SAM as a protein interaction domain involved in developmental regulation. *Protein Sci.* **6,** 249–53 (1997).
266. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17,** 405–24 (2015).
267. Al-Ali, M., Osman, W., Tay, G. K. & AlSafar, H. S. A 1000 Arab genome project to study the Emirati population. *J. Hum. Genet.* **63,** 533–536 (2018).
268. Wang, L. *et al.* Application of Whole Exome and Targeted Panel Sequencing in the Clinical Molecular Diagnosis of 319 Chinese Families with Inherited Retinal Dystrophy and Comparison Study. *Genes (Basel).* **9,** (2018).
269. Gai, X. *et al.* CNV Workshop: an integrated platform for high-throughput copy number variation discovery and clinical diagnostics. *BMC Bioinformatics* **11,** 74 (2010).
270. Villanueva, A. *et al.* Whole exome sequencing of a dominant retinitis pigmentosa family identifies a novel deletion in PRPF31. *Invest. Ophthalmol. Vis. Sci.* **55,** 2121–9 (2014).
271. Wissinger, B. *et al.* Large deletions of the KCNV2 gene are common in patients with cone dystrophy with supernormal rod response. *Hum. Mutat.* **32,** 1398–406 (2011).

272. Liquori, A. *et al.* Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. *Hum. Mutat.* **37,** 184–93 (2016).
273. Albert, S. *et al.* Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. *Am. J. Hum. Genet.* **102,** 517–527 (2018).
274. Nakanishi, H. *et al.* Hair roots as an mRNA source for mutation analysis of Usher syndrome-causing genes. *J. Hum. Genet.* **55,** 701–703 (2010).
275. Quintela, I. *et al.* Copy number variation analysis of patients with intellectual disability from North-West Spain. *Gene* **626,** 189–199 (2017).
276. Fossarello, M. *et al.* Deletion in the peripherin/RDS gene in two unrelated Sardinian families with autosomal dominant butterfly-shaped macular dystrophy. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **114,** 448–56 (1996).
277. de Castro-Miró, M. *et al.* Novel Candidate Genes and a Wide Spectrum of Structural and Point Mutations Responsible for Inherited Retinal Dystrophies Revealed by Exome Sequencing. *PLoS One* **11,** e0168966 (2016).
278. Boon, C. J. F. *et al.* Central areolar choroidal dystrophy. *Ophthalmology* **116,** 771–82, 782.e1 (2009).
279. Wells, J. *et al.* Mutations in the human retinal degeneration slow (RDS) gene can cause either retinitis pigmentosa or macular dystrophy. *Nat. Genet.* **3,** 213–218 (1993).
280. Farrar, G. J. *et al.* A three-base-pair deletion in the peripherin–RDS gene in one form of retinitis pigmentosa. *Nature* **354,** 478–480 (1991).
281. Apfelstedt-Sylla, E. *et al.* Extensive intrafamilial and interfamilial phenotypic variation among patients with autosomal dominant retinal dystrophy and mutations in the human RDS/peripherin gene. *Br. J. Ophthalmol.* **79,** 28–34 (1995).
282. Vasson, A. *et al.* Custom oligonucleotide array-based CGH: a reliable diagnostic tool for detection of exonic copy-number changes in multiple targeted genes. *Eur. J. Hum. Genet.* **21,** 977–87 (2013).
283. Overwater, E. *et al.* Results of next-generation sequencing gene panel diagnostics including copy-number variation analysis in 810 patients suspected of heritable thoracic aortic disorders. *Hum. Mutat.* **39,** 1173–1192 (2018).
284. Xia, Y. *et al.* Clinical application of chromosomal microarray analysis for the prenatal diagnosis of chromosomal abnormalities and copy number variations in fetuses with congenital heart disease. *Prenat. Diagn.* **38,** 406–413 (2018).
285. González-del Pozo, M. *et al.* Searching the second hit in patients with inherited retinal dystrophies and monoallelic variants in ABCA4, USH2A and CEP290 by whole-gene targeted sequencing. *Sci. Rep.* **8,** 13312 (2018).
286. Lo, Y.-F. *et al.* Recurrent deep intronic mutations in the SLC12A3 gene responsible for Gitelman's syndrome. *Clin. J. Am. Soc. Nephrol.* **6,** 630–9 (2011).
287. Le Quesne Stabej, P. *et al.* Comprehensive sequence analysis of nine Usher syndrome genes in the UK National Collaborative Usher Study. *J. Med. Genet.* **49,** 27–36 (2012).
288. Huang, L. *et al.* Mutation screening of the USH2A gene in retinitis pigmentosa and USHER patients in a Han Chinese population. *Eye* 1 (2018). doi:10.1038/s41433-018-0130-3
289. Consugar, M. B. *et al.* Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.* **17,** 253–261 (2015).
290. Martin-Merida, I. *et al.* Toward the Mutational Landscape of Autosomal Dominant

Retinitis Pigmentosa: A Comprehensive Analysis of 258 Spanish Families. *Investig. Opthalmology Vis. Sci.* **59,** 2345 (2018).
291. Farrar, G. J. *et al.* Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum. Mol. Genet.* **26,** R2–R11 (2017).
292. Bernal, S. *et al.* Mutations in USH2A in Spanish patients with autosomal recessive retinitis pigmentosa: high prevalence and phenotypic variation. *J. Med. Genet.* **40,** e8 (2003).
293. Avila-Fernandez, A. *et al. CERKL* Mutations and Associated Phenotypes in Seven Spanish Families with Autosomal Recessive Retinitis Pigmentosa. *Investig. Opthalmology Vis. Sci.* **49,** 2709 (2008).
294. Avela, K. *et al.* A founder mutation in *CERKL* is a major cause of retinal dystrophy in Finland. *Acta Ophthalmol.* **96,** 183–191 (2018).
295. Eksandh, L. *et al.* Full-field ERG in patients with Batten/Spielmeyer-Vogt disease caused by mutations in the CLN3 gene. *Ophthalmic Genet.* **21,** 69–77 (2000).
296. Mole, S. E., Mitchison, H. M. & Munroe, P. B. Molecular basis of the neuronal ceroid lipofuscinoses: Mutations inCLN1,CLN2,CLN3, andCLN5. *Hum. Mutat.* **14,** 199–215 (1999).
297. Bergen, A. A. B. *et al.* Mutations in ABCC6 cause pseudoxanthoma elasticum. *Nat. Genet.* **25,** 228–231 (2000).
298. Bernardis, I. *et al.* Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing: Added Value of Targeted Next-Generation Sequencing. *Biomed Res. Int.* **2016,** 6341870 (2016).
299. O'Sullivan, J. *et al.* A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J. Med. Genet.* **49,** 322–6 (2012).
300. Chang, S., Vaccarella, L., Olatunji, S., Cebulla, C. & Christoforidis, J. Diagnostic challenges in retinitis pigmentosa: genotypic multiplicity and phenotypic variability. *Curr. Genomics* **12,** 267–75 (2011).
301. Astuti, G. D. N. *et al.* Identification of Inherited Retinal Disease-Associated Genetic Variants in 11 Candidate Genes. *Genes (Basel).* **9,** (2018).
302. Patel, N. *et al.* Expanding the clinical, allelic, and locus heterogeneity of retinal dystrophies. *Genet. Med.* **18,** 554–62 (2016).
303. Abu-Safieh, L. *et al.* Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res.* **23,** 236–47 (2013).
304. Beryozkin, 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.* **5,** 13187 (2015).
305. Ellingford, J. M. *et al.* Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology* **123,** 1143–1150 (2016).
306. Allikmets, R. *et al.* A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat. Genet.* **15,** 236–46 (1997).
307. Cremers, F. P. M. *et al.* Development of a genotyping microarray for Usher syndrome. *J. Med. Genet.* **44,** 153–60 (2007).
308. Ścieżyńska, A. *et al.* Next-generation sequencing of ABCA4: High frequency of complex alleles and novel mutations in patients with retinal dystrophies from Central Europe. *Exp. Eye Res.* **145,** 93–99 (2016).

309. Lotery, A. J. *et al.* Mutations in the CRB1 gene cause Leber congenital amaurosis. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **119,** 415–20 (2001).
310. Riveiro-Alvarez, R. *et al.* Outcome of ABCA4 disease-associated alleles in autosomal recessive retinal dystrophies: retrospective analysis in 420 Spanish families. *Ophthalmology* **120,** 2332–7 (2013).
311. Corton, M. *et al.* Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. *PLoS One* **8,** e65574 (2013).
312. Allikmets, R. *et al.* Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science* **277,** 1805–7 (1997).
313. Paloma, E., Martínez-Mir, A., Vilageliu, L., Gonzàlez-Duarte, R. & Balcells, S. Spectrum of ABCA4 (ABCR) gene mutations in Spanish patients with autosomal recessive macular dystrophies. *Hum. Mutat.* **17,** 504–10 (2001).

APPENDIX

SCIENTIFIC REPORTS

OPEN

A new approach based on targeted pooled DNA sequencing identifies novel mutations in patients with Inherited Retinal Dystrophies

Received: 19 February 2018
Accepted: 4 October 2018
Published online: 18 October 2018

Maitane Ezquerra-Inchausti[1,2], Ander Anasagasti[1], Olatz Barandika[1], Gonzaga Garai-Aramburu[3], Marta Galdós[4], Adolfo López de Munain[1,5,6,7], Cristina Irigoyen[1,8] & Javier Ruiz-Ederra[1,2]

Inherited retinal diseases (IRD) are a heterogeneous group of diseases that mainly affect the retina; more than 250 genes have been linked to the disease and more than 20 different clinical phenotypes have been described. This heterogeneity both at the clinical and genetic levels complicates the identification of causative mutations. Therefore, a detailed genetic characterization is important for genetic counselling and decisions regarding treatment. In this study, we developed a method consisting on pooled targeted next generation sequencing (NGS) that we applied to 316 eye disease related genes, followed by High Resolution Melting and copy number variation analysis. DNA from 115 unrelated test samples was pooled and samples with known mutations were used as positive controls to assess the sensitivity of our approach. Causal mutations for IRDs were found in 36 patients achieving a detection rate of 31.3%. Overall, 49 likely causative mutations were identified in characterized patients, 14 of which were first described in this study (28.6%). Our study shows that this new approach is a cost-effective tool for detection of causative mutations in patients with inherited retinopathies.

Inherited retinal dystrophies (IRDs) are a group of heterogeneous diseases responsible for different clinically distinctive phenotypes. The most common IRD is Retinitis Pigmentosa (RP) with a prevalence of 1 in 3500 people. RP starts with night blindness and is followed by progressive loss of peripheral vision, leading to loss of central vision and blindness in most advanced cases. Although RP is clinically distinct from other IRDs, advanced stage of RP can be difficult to distinguish from other IRDs, including cone-rod or macular dystrophies[1]. Moreover, in some cases, clinical manifestations can differ among members of the same family. IRDs can be inherited in different traits including autosomal dominant (adRP), autosomal recessive (arRP) or X-linked (XlRP). The rate of inheritance has varied across populations studied. To date, over 250 genes have been related to various IRDs and some of them are responsible for the different phenotypes observed[2] (https://sph.uth.edu/retnet/sum-dis.htm, 3 July 2017).

Since the publication of the first draft of the human genome in 2001[3,4], we have seen an unprecedented flourishing of sequencing technologies that provide genomic information in an accurate, fast and cost-efficient way. Methods of massive parallel sequencing such as targeted Next Generation Sequencing technologies (NGS) and Whole Exome Sequencing (WES) are the most widely used methods for the diagnosis of IRD. These methods have contributed to an exponential reduction in time and costs for the execution of the sequencing[5,6]. Nevertheless, the use of whole genome sequencing for diagnostic purposes is limited, mainly by the amount of data generated, which demands high degree of expertise in terms of big data handling and interpretation of the results, and these factors

[1]Division of Neurosciences, Biodonostia Health Research Institute, San Sebastián, Spain. [2]RETICS OFTARED, National Institute of Health Carlos III, Ministry of Economy and Competitiveness, Madrid, Spain. [3]Department of Ophthalmology, Araba University Hospital, Vitoria-Gasteiz, Spain. [4]Department of Ophthalmology, Cruces University Hospital, Bilbao, Spain. [5]Department of Neurology, Donostia University Hospital, San Sebastián, Spain. [6]CIBERNED, Center for Networked Biomedical Research on Neurodegenerative Diseases, National Institute of Health Carlos III, Ministry of Economy and Competitiveness, Madrid, Spain. [7]Department of Neuroscience, University of the Basque Country UPV-EHU, San Sebastián, Spain. [8]Department of Ophthalmology, Donostia University Hospital, San Sebastián, Spain. Correspondence and requests for materials should be addressed to C.I. (email: cristina.irigoyenlaborra@osakidetza.eus) or J.R.-E. (email: javier.ruizederra@osakidetza.eus)

complicate its transfer to the clinicians and to the patients. Comprehensive sequencing of the coding regions of all genes (Whole Exome Sequencing or WES) is more affordable, but still has high technical requirements that are an obstacle to its use as a diagnostic method in routine clinical practice. A more practical approach for clinical diagnosis may consist of an initial genetic screening of a subset of genes associated with a phenotype using targeted NGS, followed by a second more extensive genome analysis, such as WES[6], and the analysis of the copy number variations (CNVs)[1], for challenging cases for which the first strategy fails to indicate any genetic explanation.

In this study, we sequenced 316 genes associated with IRDs including several syndromic retinopathies. Targeted NGS typically involves a DNA-barcode labelling of each of the individuals to be sequenced for genotyping purposes, this processing being a bottleneck process in terms of consumables, equipment and human resources. In order to simplify the sequencing process and to reduce the costs associated with individual labelling of DNA samples, we have developed a mutation detection approach based on targeted NGS in combination with high resolution melting (HRM) analysis. NGS was performed using pools of 16 DNA samples per pool, and identification of the sample/s carrying the mutation/s was performed using HRM analysis in individual samples, which allowed us to link mutations found in the pooled DNA samples to the DNA from individual patients. We sequenced samples from a total of 115 unrelated patients and 13 controls, 5 of which corresponded to samples from patients with IRD characterized by a third party laboratory. Information regarding mutations in these five controls was not revealed to us until completion of our analysis, to further test the sensitivity of our method in an objective way.

For those samples with negative results after the sequencing process, we used multiplex ligation-dependent probe amplification (MLPA) method for CNV analysis. After combining our sequencing strategy with MLPA, we were able to conclusively identify mutations in 36 patients, meaning that a genetic diagnosis rate was obtained in 31.3% of cases.

Results

Targeted Sequencing. A total of 316 genes (Supplementary Table S1) divided into 7222 amplicons were analysed. A total of 2864 and 3350 genetic variants were found in the 4 and 8 sample pools, respectively, while 3997 +/− 58 variants found in the 7 pools with 16 samples. Mean and median read depth obtained per sample were 196X and 193X, respectively. Less than 3.4% of targeted regions were covered less than 30X per pool, which we established as the cut off.

Sensitivity. In order to assess the sensitivity of our method we performed two independent experiments. In the first experiment, we included a set of 3 pools all containing an increasing number of control samples prepared from DNA from 16 patients (see methodology section and Supplementary Fig. S1 for a more detailed description). Each control sample carried at least one mutation that had been previously validated by Sanger sequencing (see methodology section). As a result, previously characterized mutations from all control samples were identified in the first set of samples, regardless of the size of the pool.

Following our method, one would expect a relative level of coverage of 1/32 in heterozygous variants and 2/32 in one homozygous or in two heterozygous variants. However, we found that the number did not fit exactly to these values when analysing variants among solved patients (see variants in Table 1). Thus, in heterozygous variants the relative coverage ranged between 0.56 to 1.54/32 with 5 outliers with relative coverage of 1.75/32, 1.88/32, 1.99/32, 1.93/32 and 2/32, with values more suggestive of mutations present in two alleles rather than in one.

With respect to variants expected to be in two alleles (in homozygosis in one patient or in heterozygosis in two patients), the relative coverage ranged between 1.5–2.3/32. In this case we found 4 outliers with relative levels of coverage as low as 1.25/32 (2 cases), or as high as 2.98/32 and 3.13/32. In all cases with a higher relative coverage, in relation with the number of alleles found, all the pool was Sanger sequenced individually, in order to test for the presence of another allele with that variant and we found that there were no more alleles with the mutation among the pool.

Moreover, we tested 9 SNPs with higher MAFs in order to assess if the relative level of coverage was the same in the case of having more alleles with a specific SNP within the pool. All 16 samples from the pool in which the SNP was found, were directly Sanger sequenced. Similarly to what we observed in the candidate variants, we found some variability between expected *vs.* sequenced SNPs, with a slight mismatch of the variants present according to expected values (Supplementary Table S2).

Variant Identification. Once we established 16 as the most cost-effective sample size, we sequenced 7 pools of 16 samples/each, including a set of 19 different controls carrying a total of 21 previously detected rare (MAF $<$ 0.003), non-causative variants (control variants). All variants selected had a MAF $<$ 0.003 for genes mainly associated with a recessive inheritance pattern and were absent from the databases in the case of genes associated with a dominant inheritance pattern (Supplementary Table S3). As a result, all 21 control variants were also redetected. In both sets of experiments our methodology yielded 100% sensitivity.

Furthermore, we included five samples from patients with IRD provided by a third party laboratory. As information about mutations within these samples was not initially disclosed to us, we were able to use these samples as an additional way to test the sensitivity of our method. We succeeded in identifying causal mutations in all of the samples. These were: a homozygous mutation c.1645G>T (p.Glu549Ter) in the *BBS1* gene; c.1040C>A (p.Pro347Gln) mutation in the *RHO* gene; c.1703TA (p.Leu568Ter) mutation in the *CHM* gene; c.2888_2888del (p.Gly963fs) and c.3386G>T (p.Arg1129Leu) mutations in the *ABCA4* gene and a homozygous mutation, c.397C>T (p.His133Tyr) in *MYO7A* gene.

With regard to the 115 unrelated patients analysed, disease causing mutations were found in at least one allele in 61 patients. Nevertheless, since in some patients, mutations were found only in one allele in recessive genes,

Family	Gene	Gene transcript	Allele1			Allele2			Family segregation
			cDNA Change	Protein change	Reference	cDNA Change	Protein change	Reference	
RP1	EYS	NM_001142800	c.9405T>A	p.Tyr3135Ter	11	c.1830del	p.His610GlnfsTer26	This study	Yes
RP8	CERKL	NM_001030311.2	c.847C>T	p.Arg283Ter	10	c.847C>T	p.Arg283Ter	10	Yes
RP15	USH2A	NM_206933	c.12093del	p.Tyr4031Ter	8	c.11241C>G	p.Tyr3747Ter	This study	Yes
RP17	CHM	NM_000390	c.1272_1273delinsCT	p.Gln425Ter	41				Yes
RP27	RPGR	NM_001034853	c.2232_2235del	p.Asp744GlufsTer70	This study				Yes
RP34	USH2A	NM_206933	c.2276G>T	p.Cys759Phe	56	c.5278del	p. Asp1760MetfsTer10	8	Yes
RP35	RP1	NM_006269	c.4804C>T	p.Gln1602Ter	67	c.1837dup	p.Thr613AsnfsTer6	This study	Yes
RP49	EYS	NM_001142800	c.4045C>T	p.Arg1349Ter	12	c.4045C>T	p.Arg1349Ter	12	Yes
RP57	TULP1	NM_003322	c.1495+1G>C		68	c.1495+1G>C		68	Yes
RP59	MYO7A	NM_000260	c.1200G>T	p.Lys400Asn	69	c.5074C>T	p.Gln1692Ter	This study	N/A
RP77	CNGA1	NM_001142564	c.301C>T	p.Arg101Ter	70	c.1747C>T	p.Arg583Ter	This study	Yes
RP88	MYO7A	NM_000260	c.3763del	p.Lys1255ArgfsTer8	71	c.6_9dup	p.Leu4AspfsTer39	This study	Yes
RP91	USH2A	NM_206933	c.11754G>A	p.Trp3918Ter	72	c.3669del	p.Cys1223Ter	This study	Yes
RP106	EYS	NM_001142800	c.14C>A	p.Ser5Ter	This study	c.888del	p.Lys296AsnfsTer43	This study	Yes
RP117	EYS	NM_001142800	c.4045C>T	p.Arg1349Ter	12	c.9405T>A	p.Tyr3135Ter	11	Yes
RP153	CERKL	NM_001030311.2	c.847C>T	p.Arg283Ter	10	c.847C>T	p.Arg283Ter	10	Yes
RP154	CNGA3	NM_001298	c.162_163insT	p.Arg55Ter	This study	c.162_163insT	p.Arg55Ter	This study	Yes
RP165	ABCA4	NM_000350	c.3322C>T	p.Arg1108Cys	73	c.3322C>T	p.Arg1108Cys	73	Yes
RP67	CERKL	NM_001030311.2	c.847C>T	p.Arg283Ter	10	c.847C>T	p.Arg283Ter	10	Yes
RP109	USH2A	NM_206933	c.1570G>A	p.Ala524Val	This study	c.2276G>T	p.Cys759Phe	56	Yes
RP141	USH2A	NM_206933	c.2276G>T	p.Cys759Phe	56	c.2299del	p.Glu767SerfsTer21	74	Yes
RP173	NR2E3	NM_014249	c.932G>A	p.Arg311Gln	75	c.932G>A	p.Arg311Gln	75	N/A
RP174	RGR	NM_001012720	c.196A>C	p.Ser66Arg	76	c.196A>C	p.Ser66Arg	76	Yes
RP175	CNGB3	NM_019098	c.1148del	p.Thr383IlefsTer13	77	c.852+1G>C		This study	Yes
RP176	CERKL	NM_001030311.2	c.847C>T	p.Arg283Ter	10	c.847C>T	p.Arg283Ter	10	Yes
RP180	USH2A	NM_206933	c.14565del	p. Asn4856MetfsTer28	This study	c.14565del	p. Asn4856MetfsTer28	This study	Yes
RP182	PDE6A	NM_000440	c.1957C>T	p.Arg653Ter	78	c.1705C>A	p.Gln569Lys	79	Yes
RP185	CNGA3	NM_001298	c.1228C>T	p.Arg410Trp	80	c.829C>G	p.Arg277Gly	81	Yes
RP196	BBS1	NM_024649	c.1220T>G	p.Met390Arg	82	c.1220T>G	p.Met390Arg	82	Yes
RP166	USH2A	NM_206933	c.14091del	p.Phe4697LeufsTer2	7	c.12093del	p.Tyr4031Ter	8	N/A
RP169	CERKL	NM_001030311.2	c.847C>T	p.Arg283Ter	10	c.356G>A	p.Gly119Asp	83	N/A
RP30	RP1	NM_006269	c.1625C>G	p.Ser542Ter	84	c.227T>C	p.Leu76Pro	This study	Yes
RP193	ABCA4	NM_000350	c.4577C>T	p.Thr1526Met	36,85	c.3386G>T	p.Arg1129Leu	86	N/A
RP200	*CRB1*	*NM_201253*	*c.444_452del*	*p.Asp148_Asp150del*	87	*c.2843G>A*	*p.Cys948Tyr*	88	*Yes*
RP188	CNGA3	NM_001298	c.1228C>T	p.Arg410Trp	80	c.1706G>A	p.Arg569His	81	N/A
RP40	PRPF31	NM_015629	exons9_13deletion		This study				Yes
RP148	*PRPF8*	*NM_006445*	*c.6835T>G*	*p.Trp2279Gly*	*This study*				
RP181	*PRPF31*	*NM_015629*	*c.1165C>T*	*p.Gln389Ter*	*This study*				*No*
RP92	*PCDH15 CDH23*	*NM_001142763/ NM_022124*	*c.733C>T*	*p.Arg245Ter*	89	*c.8326G>A*	*p.Gly2776Ser*	*This Study*	*Yes*

Table 1. Summary of all identified variants. Variants of uncertain significance (VUS) are in italics.

causal mutations were found in 36 patients, reaching a detection rate of 31.3% (Tables 1, 2 and Supplementary Fig. S2). Most of the pathogenic mutations were found in the *USH2A* gene, although in many cases only in one allele without a second mutation, and therefore in these recessive cases, we could not determine the causal mutation. Among all mutations found in characterized patients, 15 were novel, 2 missense and 13 loss-of-function (LOF) mutations. Novel missense and splicing variant mutations were potentially pathogenic, this being inferred from the score obtained from different *in-silico* tools and the fact that they co-segregated with the disease (Supplementary Table S4).

Regarding the distribution of mutations among our cohort of patients, most findings were found among the following five genes:

USH2A. Mutations within this gene were responsible for most cases of arRP in our cohort. Most of the patients were carriers of biallelic mutations. Compound heterozygous mutations are frequently reported in this gene[7,8]. Four of the mutations found in *USH2A* were novel: c.11241C>G, in patient RP15, c.3669del in patient RP91,

Family	Age at diagnosis	Symptoms at diagnosis	Visual Acuity in LogMAR RE	Visual Acuity in LogMAR LE	Spherical Equivalent RE	Spherical Equivalent LE	Subcapsular Cataract (Yes, No Pseudo-phakic)	Pale disc	Arteriolar Attenuation	Bone Spicule Retinal Pigment	Epiretinal Membrane	Macular Edema	Visual Fields (grades)	ERG (Elect-roretinogram)	Syndro-mic RP	Family member affected (including case study)
RP1	20	Photophobia	2	0.8	−2.2	−2.62	PP	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	1
RP8	17	Nyctalopia	5	5	N/A	N/A	PP	Yes	Yes	Yes	Yes	Yes	No, Low Vision	Ext	No	1
RP15	23	Nyctalopia	0.4	0.3	−0.12	−0.62	PP	Yes	Yes	Yes	Yes	No	4	N/A	No	1
RP17	26	Nyctalopia	0.7	0.1	−6.5	−5.37	No	Yes	Yes	No	Yes	Yes	4	Ext	No	2
RP27	8	Decrease VA	3	3	0.12	−0.5	PP	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	3
RP30	26	Nyctalopia	0.7	0.7	−5.5	−5.25	Yes	Yes	Yes	Yes	No	No	Altered	Ext	No	1
RP34	37	Visual Field Loss	0.3	0.8	−0.5	−0.62	PP	Yes	Yes	Yes	No	Yes	8	Ext	No	1
RP35	5	Decrease VA	0.8	1,3	0	−0.25	PP	Yes	Yes	Yes	Yes	Yes	Altered	Ext	No	1
RP40	8	Nyctalopia	0	0	0	−0.75	No	Yes	Yes	Yes	No	No	18	Ext	No	2
RP49	16	Nyctalopia	0.4	0.5	0.87	0.75	Yes	Yes	Yes	Yes	Yes	Yes	15	Ext	No	1
RP57	9	Nyctalopia	1.3	4	13	2	PP	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	1
RP59	12	Nyctalopia	0	0	1.625	−1.25	No	Yes	Yes	No	No	No	7	Ext	Usher type 1	1
RP67	50	Decrease VA	N/A	N/A	2	0.75	Yes	Yes	Yes	Yes	No	No	No, Low vision	Ext	No	2
RP77	40	Nyctalopia	0.3	0.2	0.75	0.62	PP	Yes	Yes	Yes	Yes	Yes	4	Ext	No	2
RP88	12	Visual Field Loss	1.3	1	N/A	N/A	PP	Yes	Yes	Yes	No	No	N/A, deafness	Ext	Usher type 1	2
RP91	16	Nyctalopia	0.3	0.4	−1.62	−1.87	Yes	Yes	Yes	Yes	No	No	8	Ext	Usher	1
RP106	45	Nyctalopia	4	4	−8.75	−9.5	Yes	Yes	Yes	Yes	Yes	NO	No, Low Vision	Ext	No	1
RP117	27	Decrease VA	0.5	0.4	1.12	−1.5	No	Yes	Yes	Yes	No	Yes	10	Ext	No	4
RP141	35	Nyctalopia	N/A	N/A	1	1	Yes	Yes	Yes	Yes	No	No	N/A	Ext	No	1
RP153	17	Decrease VA	3	1	−0.5	−0.25	Yes	Yes	Yes	Yes	No	No	No, Low Vision	N/A	No	2
RP154	1	Decrease VA	1	1	3	1	No	No	No	No	No	No	Central Scotoma	N/A	Achrom.	2
RP165	17	Decrease VA	3	3	N/A	N/A	Yes	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	5
RP166	N/A	Nyctalopia	0.2	0.3	−1	−1.75	Yes	Yes	Yes	Yes	No	No	7	Ext	Usher Type 2	1
RP169	31	Nyctalopia	5	4	N/A	N/A	Yes	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	2
RP173	1	Nyctalopia	1	1	−2	−0.25	No	No	No	Yes	No	Yes	No, Low Vision	Ext	No	2
RP174	38	Decrease VA	4	4	−3.37	−0.75	No	Yes	Yes	Yes	No	No	No, Low Vision	Ext	No	1
RP175	4	Decrease VA	1	1	−0.75	−0.125	No	No	No	No	No	No	No, Low Vision	*1	Achrom.	2
RP176	22	Decrease VA	0.3	0.4	−0.75	−1.5	Yes	Yes	Yes	Yes	No	No	Central scotoma	Ext	No	1
RP180	38	Nyctalopia	4	4	N/A	N/A	Yes	Yes	Yes	Yes	No	No	No, Low Vision	Ext	Usher Type 2	3
RP109	36	Nyctalopia	0.4	0.3	−0.5	0	Yes	Yes	Yes	No	Yes	Yes	7	Ext	No	1
RP182	10	Nyctalopia	0.05	0.05	−1.75	−1.25	Yes	Yes	Yes	Yes	No	No	5	Ext	No	1
RP185	1	Nystagmus	1.3	1.3	−5.37	−5.37	No	No	No	No	No	No	No, Low Vision	*1	Achrom.	1
RP196	12	Decrease VA	1	1	−1.12	−2.12	Yes	Yes	Yes	Yes	No	No	4	Ext	No	1
RP200	31	Decrease VA	0.7	3	+0.75	+1.87	No	Yes	Yes	No	No	No	No, Low Vision	Ext	No	1
RP188	49	Decrease VA	0.8	1	+7.3	+7.3	No	No	No	No	No	No	No, Low visión	C.R Ext	No	1
RP193	38	Decrease VA	1	1	+2.62	+2.61	No	No	No	No	No	No	No, Central Scotoma	N/A	No	1

Table 2. Clinical features of characterized patients. Abbreviations; LE: Left eye; NA: not available; PP: Pseudophakia; RE: Right Eye; VA: Visual Acuity**.** *ERG not detected either in photopic nor escotopic conditions.

c.1570G>A in patient RP109 and c.14565del in patient RP180. Except for patient RP180, homozygote carrier of the mutation, the rest of the patients were carriers of mutations in compound heterozygosis with the previously reported pathogenic mutations c.12093del, c.11754G>A and c.2276G>T respectively (Table 1).

CERKL. This was the second most commonly mutated gene in our cohort. We characterized 5 patients with the same mutation c.847C>T in this gene. In 4 of the cases it was in homozygosis and in one case it was in compound heterozygosis with c.356G>A mutation. This nonsense mutation is relatively common in Spanish cohorts[9,10].

EYS. This was the third most commonly mutated gene in our cohort. Three out of four patients shared mutations, such as RP1 and RP117 with c.9405T>A[11] and RP49 and RP117 with c.4045T>A[12], probably indicating the sharing of a common ancestor. This finding is consistent with previous studies involving Spanish cohorts, in which *EYS* was one of the most commonly mutated genes in recessive retinitis pigmentosa[13,14]. In addition, we found three novel mutations in this gene: two frameshift mutations in compound heterozygosis c.1830del in patient RP1 and c.888del in patient RP106; and a nonsense mutation also in compound heterozygosis c.14C>A, in patient RP106.

RPGR. We were able to detect a novel mutation c.2232_2235del in patient RP27 in the ORF15 region of this gene. Mutations in this region are challenging to amplify due to a large segment of highly repetitive purine-rich sequences[15]. Nevertheless, the high coverage of this region we obtained using our pooled-based approach, allowed us to detect this variant (Supplementary Fig. S3).

Variants of Uncertain Significance (VUS). For the family RP92, two heterozygous variants were observed in *PCDH15* and *CDH23*. Despite the fact that this digenic inheritance pattern has previously been found to be causative of Usher Syndrome[16], and that the variants segregated correctly within our family, there is some controversy with the pathogenicity of this digenism and, as far as we know, the *CDH23* and *PCDH15* digenism has been only reported in one study[16]. Despite cochlear degeneration specific to hair cells was observed in this type of mice, USH mutant mice do not display visual defects. Based on ultrastructural analyses, it has been shown that the *USH1* proteins localize at the level of microvilli-like structures, called calyceal processes, which form a collar around the base of photoreceptor outer segments. These structures have only been found in primate and other large mammals, but not in mouse photoreceptor cells[17]. This has led to propose that the absence of these structures in the mouse retina is responsible for the lack of a visual phenotype in mouse models of Usher syndrome. Regardless of this structural difference, we cannot confirm that this digenism is the causative mutation.

In the case of family RP148, a novel missense mutation c.6835T>G was found in *PRPF8* gene. The mutation was predicted to be damaging by at least 5 *in silico* predictors. Nevertheless, given the lack of a complete segregation analysis due to the unavailability of many of the samples required, we were unable to conclude that c.6835T>G is the causal adRP mutation in this family. Similarly, in family RP181, we found a novel nonsense mutation, c.1165C>T, in *PRPF31* gene. However we were not able to validate this finding in a segregation analysis due to a lack of samples available. In fact, the only family sample we were able to study was a non-affected sister who was also a mutation carrier.

Multiplex Ligation-dependent Probe Amplification (MLPA). Among the 32 families analysed by this method, we detected a large deletion in the *PRPF31* gene expanding from exon 9 to 13 in family RP40, previously unreported. The deletion was also detected in an affected grandmother and the asymptomatic mother. Confirmation of the deletion region was performed sequencing the deleted DNA fragment (Fig. 1A).

Discussion

In the present work, we have developed a cost-effective method for the diagnosis of IRDs based on pooled genomic DNA targeted NGS, in combination with HRM as a highly sensitive, versatile and affordable genotyping method. Following our methodology, we were able to find the causal mutation in 36 of our patients (31.3%) (Table 1).

Several studies have validated the feasibility of DNA sequencing pools to identify and quantify the genetic variants or single nucleotide polymorphisms (SNPs) in small genomes or small genomic regions of prokaryotes[18]; and single human genes[19,20]. Previous studies tested experimentally the accuracy in re-sequencing pools of strains of highly isogenic D. melanogaster, whose genome had been previously sequenced individually. They showed that the sequenced pool provides a correct estimate of the population allele frequency, enabling the discovery of new SNPs with a low rate of false positives[21].

Regarding clinical applications[22] evaluated the use of pooled DNA sequencing to accurately assess allele frequencies on transmitted and non-transmitted chromosomes in a set of families in an allelic association study[23] combined DNA samples from 1,111 individuals and sequenced 4 genes to identify rare germline variants. The main bottleneck in the use of a pooling strategy for genetic studies is related to the challenges of detecting rare and low-frequency variants reliably, allowing an accurate estimation of MAFs[24]. Moreover, pooled DNA sequencing was applied for the analysis of 3 genes of Gitelman's syndrome using semiconductor NGS in pooled DNA samples from 20 patients[25]. In a more recent study, 72 genes were analysed in pools consisting of samples from 12 individuals[26]. With respect to RP, pooled DNA NGS was used to search for mutations in the *SNRNP200* gene in a cohort of 96 unrelated patients from North America[27]. Pooled DNA sequencing has recently been used for population genetics studies (GWAS), in several different pathologies[28].

Compared to previous studies that limited to the sequencing of a restricted number of genes, this represents the first study based on the pooled sequencing of more than 300 genes. To estimate the reduction in costs derived from the use of our methodology we compared the costs per patient of our pooled method with an individual sequencing approach. The main source of cost savings was related to expenses involved in the preparation of DNA libraries. Specifically, there was a 10.6-fold reduction in sequencing costs with our methodology. Once we added costs associated with the HRM analysis-based genotyping method, the overall reduction in mutation detection/patient was 6.25-fold.

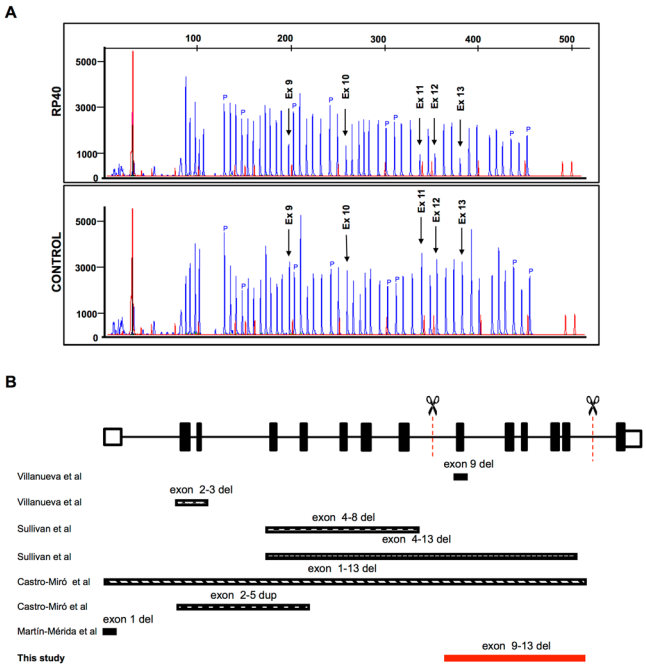


Figure 1. Novel deletion in *PRPF31*. (**A**) Electropherogram showing a reduced dosage of exons 9–13 (arrows) in patient RP40. (**B**) Schematic representation of *PRPF31* deletions described in the literature, and the deletion of exons 9–13 we found in this study, represented by the red bar. Abbreviations: P: control probes; Ex: Exon.

The choice of 16-sample pools was based, not only on terms of sensitivity, but also on the optimal number of samples for further analysis by HRM, which we found to be around 16 in a previous study[29]. One of the main advantages over previous pooled-NGS-based strategies for mutations detection is the genotyping method we used. HRM analysis is significantly more affordable than other methods including TaqMan probes (Thermo Fisher Scientific) especially if used for a large cohort of patients and/or for a large number of genes[30]; or DNA arrays Sequenom IPLEX (CD Genomics), which requires specific equipment, making the applicability of the methodology highly dependent on the equipment available in each laboratory[30].

In order to test the sensitivity of our method we included a set of positive controls. Five of these positive controls were samples from IRD patients previously diagnosed elsewhere, for whom we only had access to their clinical data, but not to information on the causative mutations. Given that we obtained a sensitivity of 100%, the fact that our detection rate is not as high as in previous studies, ranging from 51 to 66%[31–34], might be explained, at least in part, by the nature of the cohort of patients included in our study, since over half of our cohort of patients (69/115) were analysed in previous studies with no results, using a repertoire of different approaches[9,29,35].

Therefore, we believe that the great number of samples analysed in previous studies is the main factor for the relative low yield obtained. A similar observation was recently reported, where they found that the patients who were screened for the first time had a higher pathogenic variant detection rate than the overall rate, suggesting that their cohort was enriched for intractable cases giving a lower detection rate[36].

Another possibility is that the detection rate varies depending on the ethnicity of the individuals analysed[36]. In this regard, they reported a lower rate of homozygous variants detected in individuals of European origin, comparing with other populations, in recessive transmitted diseases[36]. Similarly, we found heterozygous mutations in recessive genes in 25 patients, which therefore cannot be regarded as the causal mutation on their own. One possibility is that a fraction of our patients might be bearing large DNA re-arrangements, or mutations in deep intronic regions not covered by our approach, which would act in compound heterozygosis.

One limitation of the approach used in this work was that the relative level of coverage expected in validated variants (1/32 in heterozygous variants and 2/32 in one homozygous or in two heterozygous variants) did not fit exactly to expected values in some cases (see Results section and supplementary Table S2). This could be due to the fact that there is a pre-amplification step for library preparation. Despite great care was taken for preparing the pools using equimolar amounts of each DNA sample, we cannot discard the possibility of having some samples over or under-represented, offering higher or lower relative values, respectively. This might be reflecting an unequal sample bias, or that all DNAs of each pool were not amplified in all regions, which might be one of the potential explanations for the relative low diagnostic yield. However, we consider this possibility unlikely, considering that we were able to detect all control variants introduced in each pool.

Another limitation of pooled sequencing method is related to the lack of use of multiplex barcodes, which complicates CNV detection using NGS technology[37].

There is increasing evidence of genomic rearrangements resulting in CNVs responsible for IRDs in several genes including *PRPF31*[38]; *EYS*[39]; *USH2A*[40] and X-linked *RPGR* and *CHM*[41,42]. Several recent studies have emphasized the importance of CNV analysis in IRD cases. For instance, Bujakowska *et al.*[43] found mutations in 5 out

of 28 IRD cases in *SNRNP200, PRPF31, EYS* and *OPN1LW* genes. Khateb *et al.*[44]; identified rearrangements in 6 IRD patients out of 60 involving *EYS, MYO7A, NPHP4*, *RPGR* and *CHM*. This last case *CHM* was deleted in conjunction with other 6 genes. Van Cauwenbergh *et al*., 2016[1] identified CNV in 3 patients out of 57 analysed, with mutations in *USH2A*, *HGSNAT* and *RCBTB1* genes. Interestingly, a recent paper has established a ranking of IRD genes according to genomic features and CNV occurrence. These authors recommend performing routinely a targeted CNV screening in the most prevalent 30 top-ranked IRD genes according to their genomic length[45].

Despite some authors have described the use of read depth methods for pooled multiple sequencing[46], we decided to select a group of 9 genes, most of which known to be prone to CNV formation[45] using MLPA. We analysed several patients with negative results after the sequencing of the 316 IRD genes, and we included some of the genes reported as the main contributors to CNV in different studies, such as *USH2A, EYS, CHM, PRPF31* and *RPGR*[1,38,43,44,47,48].

Using this approach, we were able to diagnose a patient with a deletion expanding from exon 9 to 13 in *PRPF31*. Rearrangements in this gene have been described to account for around 2.5% in autosomal dominant cases[38]. Although different mutated regions have been described in *PRPF31*, the deletion of exons 9 to 13 has not been described before (Fig. 1B).

The pattern of inheritance in family 40 is suggestive of an autosomal dominant pattern with incomplete penetrance. Segregation analysis was conducted in two family members, revealing the presence of an obligate carrier. Mutations in *PRPF31* have been mostly associated with cases of incomplete penetrance[49–51].

A limitation inherent to the technique employed, which is shared by WES, is the impossibility of finding mutations in deep intronic regions, not covered by the primer design. In this regard, in an attempt to find the second mutant allele, we analysed two commonly reported deep intronic mutations: c.2991+1655A>G in *CEP290*[52] and c.7595−2144A>G in *USH2A* genes[53,54], in patients with heterozygous mutations in those genes. We did not however, find the mutations that were likely causative of the disease within these regions.

Despite limitations inherent to NGS sequencing regarding its performance in repetitive or CG-rich regions of the genome, we were able to detect the mutation c.2232_2235del in ORF15 of the *RPGR* gene, a region regarded as challenging, with a poor sequencing performance, both in panel based NGS and Whole exome sequencing[15]. Using our methodology we were able to detect this mutation among one of the 16 samples of the pool, which further support the validity of our method in terms of sequencing capacity, genotyping and filtering methods (Supplementary Fig. S3).

Regarding the mutations found, *USH2A* represents the most commonly mutated gene within our cohort of patients, with eleven different mutations found in this gene in seven patients characterized. Among *USH2* genes, *USH2A* is the most commonly mutated gene and it is responsible for approximately 74–90% of USH2 cases[8,55,56]. Mutations in *USH2A*, are responsible for Usher syndrome type 2 and non-syndromic RP[57]. *CERKL* and *EYS* are the next most commonly mutated genes in our cohort, which is also in accordance with previous studies[58,59]. In case of mutations in *EYS* genes, high prevalence has also been observed among Spanish population[14], Americans with European origin[13] and among Japanese populations[60].

For those patients for whom we failed to identify putative disease-causing mutations, the use of alternative approaches will hopefully succeed in characterizing their disease, at the molecular level. For instance, WES aimed at the identification of mutations in genes not currently linked to IRDs; aCGH arrays for the analysis of CNVa in other genes or regions not covered by our MLPA analysis; or whole genome sequencing to extend the analysis to the 99% of non-coding DNA. Despite being highly dependent on technical support, the use of whole genome sequencing is gaining momentum in clinical practice, and it seems plausible that it will become feasible in a near future, once a robust translational genomics workflow becomes an affordable option both in economic and technical terms, to allow feedback of potentially diagnostic findings to clinicians and research participants[61].

Materials and Methods

Study subjects. IRD patients were clinically diagnosed by the Ophthalmology Service at Donostia University Hospital, San Sebastian, Spain. Most patients studied had been given a diagnosis of retinitis pigmentosa, though a few patients with an undetermined inherited retinal dystrophy (IRD) were also included, based on pedigrees and clinical criteria. The inclusion criteria used were night blindness, peripheral visual field loss, pigmentary deposits resembling bone spicules, retinal vessels attenuation, optic disc pallor and reduced rod and cone response amplitudes and a delay in their timing in the electroretinogram (Hartong, 2006). A total of 115 probands were selected. In addition, samples from 13 patients were included as characterized control patients. This control group was composed of 8/13 samples selected from our cohort of IRD patients with mutations identified in previous studies[9,29,35] and a further 5 control samples from IRD patients characterized by a third party laboratory, (those for which we were blinded to information regarding mutations until we had completed our analysis). Family pedigrees were generated from information obtained from probands. All procedures performed in studies involving human participants received approval from the ethical standards of the Clinical Research Ethics Committee of the Basque Country, Spain (CEIC-E) and were in accordance with the 2013 Helsinki declaration or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Human sample collection. High molecular weight DNA was extracted from blood samples from RP patients and their available family members. Total DNA from samples was extracted and isolated with the AutoGenFlex Star instrument (AutoGen, Holliston, MA, USA) using the FlexiGene DNA Kit (Qiagen, Hilden, Germany) following the manufacturer´s instructions. DNA concentrations were measured on the Qubit fluorometer using Quant-iT PicoGreen reagent (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Equimolar amounts of DNA samples were pooled (100 ng/ul per sample). For a detailed description of the procedure see[29].

Pooled sequencing. In order to assess the sensitivity and cost-effectiveness of our method we performed a first experiment to compare the yield obtained after sequencing pools with increasing number of DNA samples and we estimated the differences in costs involved in individual *vs.* pooled sequencing. All pools were made up from samples from carriers of low-frequency variants, which corresponded to either causal, variants of uncertain significance (VUS) or non-pathogenic variants identified in previous studies[9,29,35]. A total of 13 control samples were used in 3 sets of pools, with 4, 8 and 16 control samples in each. Of these control samples, 9 carried pathogenic variants (one provided by a third party laboratory), while 7 carried low frequency variants with a minor allele frequency (MAF) <0.003, and therefore we used these 7 samples both as controls and as test samples. Samples were prepared as follows: An initial pool of 4 samples was generated. This pool was used to generate the 3 pools, adding 0, 4 or 8 more samples to generate the pools with 4, 8 and 16 samples, respectively (Supplementary Fig. S1A and Supplementary Table S3A).

In order to further test the sensitivity of our method and to detect possible differences in the sequencing yield, inherent to each sequencing run, we conducted a complementary experiment. For this, we used a different set of controls, all from carriers of low-frequency, non-disease causing variants or individuals with recessive phenotypes with disease causing mutations present in only one allele. In this case, out of 115 patients analysed, a total of 108 test samples were interrogated: 16/108 corresponded to carriers of a total of 21 previously detected non disease causing variants with low MAF (<0.003) and were, therefore, used as both control and test samples (Supplementary Table S3B). 53/108 samples corresponded to patients that had been interrogated previously with negative results, and 39/108 corresponded to new samples interrogated in this study for the first time. As additional controls we used four samples from carriers of disease causing mutations provided by a third party laboratory (for which we were blinded to mutation-related information until after our analysis). For this experiment, patients were divided into 7 pools with 16 samples each. Control samples were distributed among each pool such as that each pool contained at least 2 control samples, and 4/7 pools had also control from a third party laboratory (Supplementary Fig. S1B).

Amplicon Library preparation. Ion AmpliSeq Library Preparation Kit v2.0 (Thermo Fisher Scientific) was used to construct an amplicon library from genomic target regions with a maximum read length of approximately 200 base pairs (average length, 142 bp) for shotgun sequencing on an Ion Proton system (Thermo Fisher Scientific). Briefly, target genomic regions were amplified by simple PCR using Ion Ampliseq primer pools and 10 ng of each DNA samples.

Sequencing Analysis. *Ion Proton Sequencing.* NGS was carried out on the Ion Proton system (Thermo Fisher Scientific). Briefly, enriched ion sphere particles (ISPs) were annealed with the sequencing primer and mixed with the sequencing polymerase from the Ion PGM_200 Sequencing Kit (Thermo Fisher Scientific). Then, the polymerase-bound and primer-activated ISPs were loaded into the previously checked and washed Ion PI Chips (Life Technologies) and having planned the run on the Ion Proton System software, chips were subjected to 500 cycles of sequencing with the standard nucleotide flow order. Signal processing and base calling of data generated from the Ion Proton runs were performed with the Ion Torrent platform-specific analysis software (Torrent Suite version 4.0).

Variant calling. Using the Ion Reporter software we performed the variant calling. First of all GRCh37/hg19 was used as reference genome and alignment was performed against a bed file containing all regions corresponding to 316 genes sequenced. A key aspect in our mutation detection pipeline was to take into consideration the *dilution* effect of each variant due to our pooled sequencing approach. Therefore we used the pipeline provided by the ion reporter program for the detection of somatic mutations with minor modifications. We used a somatic mutation detection approach, since this is the most suited for the detection of variants represented in very low frequency (1 in 32 alleles, in the lowest case). The only modification to the default parameters provided by the ion reporter program (5.0 version) consisted on the switch of 10 parameters within the Variant Filtering section in Parameters tab. All parameters are described in detail in Supplementary Table S5. Finally, a Variant Caller File (VCF) was generated.

Genotyping by high resolution melting (HRM) analysis. Likely disease causing variants from each pool of 16 samples were selected from the VCF. Specific primers were designed to perform a HRM analysis generating amplicons ranging between 250 to 330 bp in length, in order to cover the mutation position. HRM analysis was used to identify which sample/s among 16 in the pool carried the mutation. We followed the methodology described in[29], with minor modification. Briefly, PCR amplification and HRM were performed in a single run on a 7900HT Fast Real-Time PCR System in 384-well plates (Applied Biosystems), each plate contained individual samples (in triplicates) from the 16 probands of the pool in which the variant was detected. We analysed up to 7 different variants in parallel in a single run. After HRM run, the analysis of post amplification fluorescent melting curves was performed using the HRM V2.0.1 software (ThermoFisher Scientific). Melting curves were normalized and difference plots were generated to compare the samples. Only samples showing a different melting curve (Fig. 2) were Sanger sequenced.

Sanger sequencing. Sanger sequencing was used to confirm those mutations detected by NGS and for co-segregation analysis using a 16-capillary ABI 3130xl platform (Applied Biosystems, Foster City, CA, USA) according to manufacturer´s protocol. Sequences were analysed and compared with wild-type samples and a reference sequences using BioEdit software (Ibis Biosciences, Carlsbad, CA, USA) and Ensembl and NCBI databases.

Relevant variant priorization and pathogenicity score. In order to determine genomic variants of relevance, we selected the potential disease causing variants according to the following pre-established criteria:

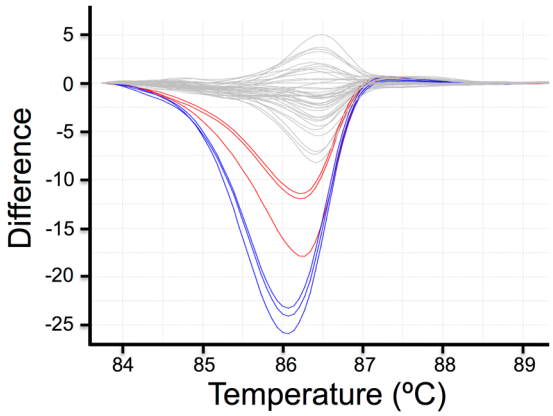


Figure 2. HRM analysis of *TULP1* gene. Difference plot shows c.1495+1G>C mutation in *TULP1* gene, with 2 out of 16 samples that clearly differ from the non-carrier samples (grey lines). Sanger sequencing confirmed the presence of the mutation c.1495+1G>C in two patients, one in heterozygosis (blue lines) and the other one in homozygosis (red lines). Note that samples are in triplicates.

(1) Variants previously reported as pathogenic.
(2) Variants with a MAF <0.001 for dominant genes or MAF <0.003 for recessive genes obtained from genome aggregation database (gnomAD).
(3) Novel Splicing variants and loss-of-function variants such as nonsense mutations, frameshift deletions or insertions.
(4) Previously reported missense variants with pathogenicity scores assessed by *in silico* predictive software.
(5) Novel missense variants predicted to be damaging by *in-silico* predictive software (as mentioned below).

Presence for all candidate variants was checked using the Spanish Variant Server Database (CSVS), (http://csvs.babelomics.org/)[62]. For dominant variants, only those absent from this database were considered further. With regard to recessive variants, only those variants with a MAF lower than 0.003 and only present in heterozygosis were considered further.

Multiplex Ligation-dependent Probe Amplification assay (MLPA). MLPA was used to search for genomic copy number variations in 32 patients without causative mutations found after sequencing of 316 IRD genes. We selected 9 genes with high prevalence of reported rearrangements[38–40].

Patients with a dominant inheritance pattern were analysed using MLPA Retinitis Probemix (P235). This probemix contains *PRPF31*, *RHO*, *RP1* and *IMPDH1* genes.

Patients with heterozygotic mutations in *USH2A* genes or *EYS* were also analysed for CNVs, in search of the second mutated allele within these genes (Salsa Mixes P361/2 and P328, respectively).

In addition, patients with an X-linked inheritance pattern, clinically diagnosed with choroideremia or families with only males affected, were analysed for *RP2*, *RPGR* and *CHM* genes (Salsa probemix P366).

MLPA reactions were run according to the manufacturer's general recommendations (MRC-Holland, Amsterdam, Holland) as previously described[63]. The MLPA reaction products were separated by capillary electrophoresis on Abi Prism 3130XL Analyzer (Applied Biosystems) and the results obtained were analysed by GeneMapper software (Thermo Fisher Scientific).

Pathogenicity predictive software. SIFT (http://www.sift.bii.a-star.edu.sg).
Polyphen2 (http://www.genetics.bwh.harvard.edu/pph2/).
PROVEAN (http://provean.jcvi.org/seq_submit.php)[64].
GVGD (agvgd.iarc.fr/agvgd_input_php)[65].
MutationTaster (www.mutationtaster.org)[66].

Web sources. Ensembl, http://www.ensembl.org/.
NCBI, http://www.ncbi.nlm.nih.gov/.
Polyphen-2, http://www.genetics.bwh.harvard.edu/pph2/.
RetNet, http://www.sph.uth.tmc.edu/Retnet/.
SIFT, http://www.sift.bii.a-star.edu.sg/.
SNPnexus, http://www.snp-nexus.org/.
The Human Genome Variation Society (HGVS), http://www.hgvs.org/.
1000 Genomes, http://www.1000genomes.org/_ENREF_48.
NHLBI Exome Sequencing Project (ESP), http://evs.gs.washington.edu/EVS/.
Babelomics, http://csvs.babelomics.org.
ExacBrowse, http://exac.broadinstitute.org/.
GnomAD browser, http://gnomad.broadinstitute.org/.

References

1. Van Cauwenbergh, C. *et al.* arrEYE: a customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genetics in medicine: official journal of the American College of Medical Genetics* **19**, 457–466, https://doi.org/10.1038/gim.2016.119 (2017).
2. Sohocki, M. M. *et al.* Prevalence of AIPL1 mutations in inherited retinal degenerative disease. *Mol Genet Metab* **70**, 142–150, https://doi.org/10.1006/mgme.2000.3001S1096-7192(00)93001-4 (2000).
3. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921, https://doi.org/10.1038/35057062 (2001).
4. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351, https://doi.org/10.1126/science.1058040 (2001).
5. Ferrari, S. *et al.* Retinitis pigmentosa: genes and disease mechanisms. *Curr Genomics* **12**, 238–249, https://doi.org/10.2174/138920211795860107CG-12-238 (2011).
6. Lazaridis, K. N. *et al.* Outcome of Whole Exome Sequencing for Diagnostic Odyssey Cases of an Individualized Medicine Clinic: The Mayo Clinic Experience. *Mayo Clinic proceedings* **91**, 297–307, https://doi.org/10.1016/j.mayocp.2015.12.018 (2016).
7. Bonnet, C. *et al.* An innovative strategy for the molecular diagnosis of Usher syndrome identifies causal biallelic mutations in 93% of European patients. *European journal of human genetics: EJHG* **24**, 1730–1738, https://doi.org/10.1038/ejhg.2016.99 (2016).
8. Garcia-Garcia, G. *et al.* Mutational screening of the USH2A gene in Spanish USH patients reveals 23 novel pathogenic mutations. *Orphanet journal of rare diseases* **6**, 65, https://doi.org/10.1186/1750-1172-6-65 (2011).
9. Barandika, O. *et al.* A Cost-Effective Mutation Screening Strategy for Inherited Retinal Dystrophies. *Ophthalmic research* **56**, 123–131, https://doi.org/10.1159/000445690 (2016).
10. Tuson, M., Marfany, G. & Gonzalez-Duarte, R. Mutation of CERKL, a novel human ceramide kinase gene, causes autosomal recessive retinitis pigmentosa (RP26). *American journal of human genetics* **74**, 128–138, https://doi.org/10.1086/381055 (2004).
11. Collin, R. W. *et al.* Identification of a 2 Mb human ortholog of Drosophila eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. *American journal of human genetics* **83**, 594–603, https://doi.org/10.1016/j.ajhg.2008.10.014 (2008).
12. Eisenberger, T. *et al.* Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PloS one* **8**, e78496, https://doi.org/10.1371/journal.pone.0078496 (2013).
13. Daiger, S. P., Bowne, S. J. & Sullivan, L. S. Perspective on genes and mutations causing retinitis pigmentosa. *Archives of ophthalmology* **125**, 151–158, https://doi.org/10.1001/archopht.125.2.151 (2007).
14. Barragan, I. *et al.* Mutation spectrum of EYS in Spanish patients with autosomal recessive retinitis pigmentosa. *Human mutation* **31**, E1772–1800 (2010).
15. Li, J. *et al.* Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. *The Journal of molecular diagnostics: JMD* **18**, 817–824, https://doi.org/10.1016/j.jmoldx.2016.06.007 (2016).
16. Zheng, Q. Y. *et al.* Digenic inheritance of deafness caused by mutations in genes encoding cadherin 23 and protocadherin 15 in mice and humans. *Human molecular genetics* **14**, 103–111, https://doi.org/10.1093/hmg/ddi010 (2005).
17. Sahly, I. *et al.* Localization of Usher 1 proteins to the photoreceptor calyceal processes, which are absent from mice. *The Journal of cell biology* **199**, 381–399, https://doi.org/10.1083/jcb.201202012 (2012).
18. Van Tassell, C. P. *et al.* SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature methods* **5**, 247–252, https://doi.org/10.1038/nmeth.1185 (2008).
19. Margraf, R. L. *et al.* Variant identification in multi-sample pools by illumina genome analyzer sequencing. *Journal of biomolecular techniques: JBT* **22**, 74–84 (2011).
20. Out, A. A. *et al.* Deep sequencing to reveal new variants in pooled DNA samples. *Human mutation* **30**, 1703–1712, https://doi.org/10.1002/humu.21122 (2009).
21. Zhu, Y., Bergland, A. O., Gonzalez, J. & Petrov, D. A. Empirical validation of pooled whole genome population re-sequencing in Drosophila melanogaster. *PloS one* **7**, e41901, https://doi.org/10.1371/journal.pone.0041901 (2012).
22. Shaw, S. H., Carrasquillo, M. M., Kashuk, C., Puffenberger, E. G. & Chakravarti, A. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. *Genome research* **8**, 111–123 (1998).
23. Druley, T. E. *et al.* Quantification of rare allelic variants from pooled genomic DNA. *Nature methods* **6**, 263–265, https://doi.org/10.1038/nmeth.1307 (2009).
24. Wang, J. *et al.* Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. *Scientific reports* **6**, 33256, https://doi.org/10.1038/srep33256 (2016).
25. Tavira, B. *et al.* A labor- and cost-effective non-optical semiconductor (Ion Torrent) next-generation sequencing of the SLC12A3 and CLCNKA/B genes in Gitelman's syndrome patients. *Journal of human genetics* **59**, 376–380, https://doi.org/10.1038/jhg.2014.37 (2014).
26. Anand, S. *et al.* Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. *Scientific reports* **6**, 33735, https://doi.org/10.1038/srep33735 (2016).
27. Benaglio, P. *et al.* Next generation sequencing of pooled samples reveals new SNRNP200 mutations associated with retinitis pigmentosa. *Human mutation* **32**, E2246–2258, https://doi.org/10.1002/humu.21485 (2011).
28. Jin, S. C., Benitez, B. A., Deming, Y. & Cruchaga, C. Pooled-DNA Sequencing for Elucidating New Genomic Risk Factors, Rare Variants Underlying Alzheimer's Disease. *Methods in molecular biology* **1303**, 299–314, https://doi.org/10.1007/978-1-4939-2627-5_18 (2016).
29. Anasagasti, A. *et al.* Genetic high throughput screening in Retinitis Pigmentosa based on high resolution melting (HRM) analysis. *Experimental eye research* **116**, 386–394 (2013).
30. Jin, S. C. *et al.* Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. *Alzheimer's research & therapy* **4**, 34, https://doi.org/10.1186/alzrt137 (2012).
31. Consugar, M. B. *et al.* Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genetics in medicine: official journal of the American College of Medical Genetics* **17**, 253–261, https://doi.org/10.1038/gim.2014.172 (2015).
32. Zhang, Q. *et al.* Next-generation sequencing-based molecular diagnosis of 35 Hispanic retinitis pigmentosa probands. *Scientific reports* **6**, 32792, https://doi.org/10.1038/srep32792 (2016).
33. Carrigan, M. *et al.* Panel-Based Population Next-Generation Sequencing for Inherited Retinal Degenerations. *Scientific reports* **6**, 33248, https://doi.org/10.1038/srep33248 (2016).
34. Tiwari, A. *et al.* Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Scientific reports* **6**, 28755, https://doi.org/10.1038/srep28755 (2016).
35. Ezquerra-Inchausti, M. *et al.* High prevalence of mutations affecting the splicing process in a Spanish cohort with autosomal dominant retinitis pigmentosa. *Scientific reports* **7**, 39652, https://doi.org/10.1038/srep39652 (2017).
36. Carss, K. J. *et al.* Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *American journal of human genetics* **100**, 75–90, https://doi.org/10.1016/j.ajhg.2016.12.003 (2017).
37. Marelli, C. *et al.* Mini-Exome Coupled to Read-Depth Based Copy Number Variation Analysis in Patients with Inherited Ataxias. *Human mutation* **37**, 1340–1353, https://doi.org/10.1002/humu.23063 (2016).
38. Sullivan, L. S. *et al.* Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. *Investigative ophthalmology & visual science* **47**, 4579–4588, https://doi.org/10.1167/iovs.06-0440 (2006).

39. Pieras, J. I. *et al*. Copy-number variations in EYS: a significant event in the appearance of arRP. *Investigative ophthalmology & visual science* **52**, 5625–5631, https://doi.org/10.1167/iovs.11-7292 (2011).
40. Steele-Stallard, H. B. *et al*. Screening for duplications, deletions and a common intronic mutation detects 35% of second mutations in patients with USH2A monoallelic mutations on Sanger sequencing. *Orphanet journal of rare diseases* **8**, 122, https://doi.org/10.1186/1750-1172-8-122 (2013).
41. Sanchez-Alcudia, R. *et al*. A Comprehensive Analysis of Choroideremia: From Genetic Characterization to Clinical Practice. *PloS one* **11**, e0151943, https://doi.org/10.1371/journal.pone.0151943 (2016).
42. Simunovic, M. P. *et al*. The Spectrum of CHM Gene Mutations in Choroideremia and Their Relationship to Clinical Phenotype. *Investigative ophthalmology & visual science* **57**, 6033–6039, https://doi.org/10.1167/iovs.16-20230 (2016).
43. Bujakowska, K. M. *et al*. Copy-number variation is an important contributorto the genetic causality of inherited retinal degenerations. *Genetics in medicine: official journal of the American College of Medical Genetics* **19**, 643–651, https://doi.org/10.1038/gim.2016.158 (2017).
44. Khateb, S. *et al*. Identification of genomic deletions causing inherited retinal degenerations by coverage analysis of whole exome sequencing data. *Journal of medical genetics* **53**, 600–607, https://doi.org/10.1136/jmedgenet-2016-103825 (2016).
45. Van Schil, K. *et al*. Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genetics in medicine: official journal of the American College of Medical Genetics* **20**, 202–213, https://doi.org/10.1038/gim.2017.97 (2018).
46. Ellingford, J. M. *et al*. Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *Journal of medical genetics* **55**, 114–121, https://doi.org/10.1136/jmedgenet-2017-104791 (2018).
47. Martin-Merida, I. *et al*. Analysis of the PRPF31 Gene in Spanish Autosomal Dominant Retinitis Pigmentosa Patients: A Novel Genomic Rearrangement. *Investigative ophthalmology & visual science* **58**, 1045–1053, https://doi.org/10.1167/iovs.16-20515 (2017).
48. Abu-Safieh, L. *et al*. A large deletion in the adRP gene PRPF31: evidence that haploinsufficiency is the cause of disease. *Molecular vision* **12**, 384–388 (2006).
49. Saini, S., Robinson, P. N., Singh, J. R. & Vanita, V. A novel 7 bp deletion in PRPF31 associated with autosomal dominant retinitis pigmentosa with incomplete penetrance in an Indian family. *Experimental eye research* **104**, 82–88, https://doi.org/10.1016/j.exer.2012.09.010 (2012).
50. Rose, A. M. & Bhattacharya, S. S. Variant haploinsufficiency and phenotypic non-penetrance in PRPF31-associated retinitis pigmentosa. *Clinical genetics* **90**, 118–126, https://doi.org/10.1111/cge.12758 (2016).
51. Rose, A. M. *et al*. Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. *Scientific reports* **6**, 19450, https://doi.org/10.1038/srep19450 (2016).
52. den Hollander, A. I. *et al*. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *American journal of human genetics* **79**, 556–561, https://doi.org/10.1086/507318 (2006).
53. Vache, C. *et al*. Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. *Human mutation* **33**, 104–108, https://doi.org/10.1002/humu.21634 (2012).
54. Slijkerman, R. W. *et al*. Antisense Oligonucleotide-based Splice Correction for USH2A-associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. *Molecular therapy. Nucleic acids* **5**, e381, https://doi.org/10.1038/mtna.2016.89 (2016).
55. Espinos, C., Millan, J. M., Beneyto, M. & Najera, C. Epidemiology of Usher syndrome in Valencia and Spain. *Community genetics* **1**, 223–228, https://doi.org/10.1159/000016167 (1998).
56. Rivolta, C., Sweklo, E. A., Berson, E. L. & Dryja, T. P. Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss. *American journal of human genetics* **66**, 1975–1978, https://doi.org/10.1086/302926 (2000).
57. Pierrache, L. H. *et al*. Visual Prognosis in USH2A-Associated Retinitis Pigmentosa Is Worse for Patients with Usher Syndrome Type IIa Than for Those with Nonsyndromic Retinitis Pigmentosa. *Ophthalmology* **123**, 1151–1160, https://doi.org/10.1016/j.ophtha.2016.01.021 (2016).
58. Bocquet, B. *et al*. Relative frequencies of inherited retinal dystrophies and optic neuropathies in Southern France: assessment of 21-year data management. *Ophthalmic epidemiology* **20**, 13–25, https://doi.org/10.3109/09286586.2012.737890 (2013).
59. Yan, D. *et al*. Spectrum of DNA variants for non-syndromic deafness in a large cohort from multiple continents. *Human genetics* **135**, 953–961, https://doi.org/10.1007/s00439-016-1697-z (2016).
60. Iwanami, M., Oshikawa, M., Nishida, T., Nakadomari, S. & Kato, S. High prevalence of mutations in the EYS gene in Japanese patients with autosomal recessive retinitis pigmentosa. *Investigative ophthalmology & visual science* **53**, 1033–1040, https://doi.org/10.1167/iovs.11-9048 (2012).
61. Wright, C. F. *et al*. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**, 1305–1314, https://doi.org/10.1016/S0140-6736(14)61705-0 (2015).
62. Alonso, R. *et al*. Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic acids research* **43**, W117–121, https://doi.org/10.1093/nar/gkv384 (2015).
63. Kozlowski, P. *et al*. Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. *Human genetics* **121**, 389–400, https://doi.org/10.1007/s00439-006-0308-9 (2007).
64. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PloS one* **7**, e46688, https://doi.org/10.1371/journal.pone.0046688 (2012).
65. Mathe, E. *et al*. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic acids research* **34**, 1317–1325, https://doi.org/10.1093/nar/gkj518 (2006).
66. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* **7**, 575–576, https://doi.org/10.1038/nmeth0810-575 (2010).
67. Corton, M. *et al*. Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. *PloS one* **8**, e65574, https://doi.org/10.1371/journal.pone.0065574 (2013).
68. Banerjee, P. *et al*. TULP1 mutation in two extended Dominican kindreds with autosomal recessive retinitis pigmentosa. *Nature genetics* **18**, 177–179, https://doi.org/10.1038/ng0298-177 (1998).
69. Roux, A. F. *et al*. Four-year follow-up of diagnostic service in USH1 patients. *Investigative ophthalmology & visual science* **52**, 4063–4071, https://doi.org/10.1167/iovs.10-6869 (2011).
70. Paloma, E. *et al*. Novel homozygous mutation in the alpha subunit of the rod cGMP gated channel (CNGA1) in two Spanish sibs affected with autosomal recessive retinitis pigmentosa. *Journal of medical genetics* **39**, E66 (2002).
71. Jaijo, T. *et al*. MYO7A mutation screening in Usher syndrome type I patients from diverse origins. *Journal of medical genetics* **44**, e71, https://doi.org/10.1136/jmg.2006.045377 (2007).
72. Baux, D. *et al*. Enrichment of LOVD-USHbases with 152 USH2A genotypes defines an extensive mutational spectrum and highlights missense hotspots. *Human mutation* **35**, 1179–1186, https://doi.org/10.1002/humu.22608 (2014).
73. Fujinami, K. *et al*. ABCA4 gene screening by next-generation sequencing in a British cohort. *Investigative ophthalmology & visual science* **54**, 6662–6674, https://doi.org/10.1167/iovs.13-12570 (2013).
74. Aller, E. *et al*. Genetic analysis of 2299delG and C759F mutations (USH2A) in patients with visual and/or auditory impairments. *European journal of human genetics: EJHG* **12**, 407–410, https://doi.org/10.1038/sj.ejhg.5201138 (2004).
75. Gerber, S. *et al*. The photoreceptor cell-specific nuclear receptor gene (PNR) accounts for retinitis pigmentosa in the Crypto-Jews from Portugal (Marranos), survivors from the Spanish Inquisition. *Human genetics* **107**, 276–284 (2000).

76. Morimura, H., Saindelle-Ribeaudeau, F., Berson, E. L. & Dryja, T. P. Mutations in RGR, encoding a light-sensitive opsin homologue, in patients with retinitis pigmentosa. *Nature genetics* **23**, 393–394, https://doi.org/10.1038/70496 (1999).
77. Peng, C., Rich, E. D. & Varnum, M. D. Achromatopsia-associated mutation in the human cone photoreceptor cyclic nucleotide-gated channel CNGB3 subunit alters the ligand sensitivity and pore properties of heteromeric channels. *The Journal of biological chemistry* **278**, 34533–34540, https://doi.org/10.1074/jbc.M305102200 (2003).
78. Perez-Carro, R. *et al.* Panel-based NGS Reveals Novel Pathogenic Mutations in Autosomal Recessive Retinitis Pigmentosa. *Scientific reports* **6**, 19531, https://doi.org/10.1038/srep19531 (2016).
79. Dryja, T. P., Rucinski, D. E., Chen, S. H. & Berson, E. L. Frequency of mutations in the gene encoding the alpha subunit of rod cGMP-phosphodiesterase in autosomal recessive retinitis pigmentosa. *Investigative ophthalmology & visual science* **40**, 1859–1865 (1999).
80. Kohl, S. *et al.* Total colourblindness is caused by mutations in the gene encoding the alpha-subunit of the cone photoreceptor cGMP-gated cation channel. *Nature genetics* **19**, 257–259, https://doi.org/10.1038/935 (1998).
81. Wissinger, B. *et al.* CNGA3 mutations in hereditary cone photoreceptor disorders. *American journal of human genetics* **69**, 722–737, https://doi.org/10.1086/323613 (2001).
82. Mykytyn, K. *et al.* Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nature genetics* **31**, 435–438, https://doi.org/10.1038/ng935 (2002).
83. Weisschuh, N. *et al.* Mutation Detection in Patients with Retinal Dystrophies Using Targeted Next Generation Sequencing. *PloS one* **11**, e0145951, https://doi.org/10.1371/journal.pone.0145951 (2016).
84. Avila-Fernandez, A. *et al.* Identification of an RP1 prevalent founder mutation and related phenotype in Spanish patients with early-onset autosomal recessive retinitis. *Ophthalmology* **119**, 2616–2621, https://doi.org/10.1016/j.ophtha.2012.06.033 (2012).
85. Lewis, R. A. *et al.* Genotype/Phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *American journal of human genetics* **64**, 422–434, https://doi.org/10.1086/302251 (1999).
86. Riveiro-Alvarez, R. *et al.* Frequency of ABCA4 mutations in 278 Spanish controls: an insight into the prevalence of autosomal recessive Stargardt disease. *The British journal of ophthalmology* **93**, 1359–1364, https://doi.org/10.1136/bjo.2008.148155 (2009).
87. Vallespin, E. *et al.* Gene symbol: CRB1. Disease: Leber congenital amaurosis. Accession #Hm0540. *Human genetics* **118**, 778 (2006).
88. Corton, M. *et al.* High frequency of CRB1 mutations as cause of Early-Onset Retinal Dystrophies in the Spanish population. *Orphanet journal of rare diseases* **8**, 20, https://doi.org/10.1186/1750-1172-8-20 (2013).
89. Ben-Yosef, T. *et al.* A mutation of PCDH15 among Ashkenazi Jews with the type 1 Usher syndrome. *The New England journal of medicine* **348**, 1664–1670, https://doi.org/10.1056/NEJMoa021502 (2003).

Acknowledgements

This work was supported by grants from the National Institute of Health Carlos III (Institute of Health Carlos III/ISCIII) (CP10/00572, PI13/02621 and RD16/0008/0027 to JRE, PI17/01413 to CI, and a Research Intensification Contract to ALdM); the Basque Government's Industry Department (SAIOTEK: SAIO11-PE11BN002; and SAIO12-PC12BN001 to JRE), a grant from the Mutua Madrileña Foundation and support from the Retinitis Pigmentosa Patients of Gipuzkoa Foundation (BEGISARE). JR-E is a Miguel Servet II Fellow, National Institute of Health Carlos III (ISCIII). MEI was supported by grants from the Basque Government's Department of Education (DEDUC14/309). OB is supported by funding from the Retinitis Pigmentosa Patients of Gipuzkoa Foundation (BEGISARE) and a grant from the Mutua Madrileña Foundation. AA was supported by grants from the Fundación Jesús de Gangoiti Barrera and from the Basque Government's Departments of Industry and Education (SAIOTEK-11BN002/PC12BN001/DEPLC13/002). CI is partially supported by a Research Intensification Contract (INTBIO15/001). The authors are grateful to Xabier Elcoroaristizabal and Marta Fernández-Mercado for their helpful advice on developing the base-calling setup. Maribel Gómez; Naiara Telletxea and Nahikari Pastoriza at the Basque Biobank for isolating DNA samples; and Dr. Carmen Ayuso for kindly providing control samples. We also give special thanks to all patients with IRD and their families involved in the study.

Author Contributions

M.E.-I. analysed most data and interpreted the results, generated all figures and drafted the manuscript. A.A. analysed part of the data and interpreted the results O.B. analysed part of the data and interpreted the results. G.G.A. selected a group of patients and collected the clinical data. M.G. selected a group of patients and collected the clinical data. A.L.d.M. interpreted the results and revised the manuscript. C.I. selected most patients collected the clinical data and obtained funding. J.R.-E. planned the experiments, interpreted the results, drafted the manuscript and obtained funding. All authors revised and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-33810-3.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SCIENTIFIC REPORTS

OPEN

High prevalence of mutations affecting the splicing process in a Spanish cohort with autosomal dominant retinitis pigmentosa

Received: 20 September 2016
Accepted: 24 November 2016
Published: 03 January 2017

Maitane Ezquerra-Inchausti[1,*], Olatz Barandika[1,*], Ander Anasagasti[1], Cristina Irigoyen[1,2], Adolfo López de Munain[1,3,4,5] & Javier Ruiz-Ederra[1]

Retinitis pigmentosa is the most frequent group of inherited retinal dystrophies. It is highly heterogeneous, with more than 80 disease-causing genes 27 of which are known to cause autosomal dominant RP (adRP), having been identified. In this study a total of 29 index cases were ascertained based on a family tree compatible with adRP. A custom panel of 31 adRP genes was analysed by targeted next-generation sequencing using the Ion PGM platform in combination with Sanger sequencing. This allowed us to detect putative disease-causing mutations in 14 out of the 29 (48.28%) families analysed. Remarkably, around 38% of all adRP cases analysed showed mutations affecting the splicing process, mainly due to mutations in genes coding for spliceosome factors (SNRNP200 and PRPF8) but also due to splice-site mutations in RHO. Twelve of the 14 mutations found had been reported previously and two were novel mutations found in PRPF8 in two unrelated patients. In conclusion, our results will lead to more accurate genetic counselling and will contribute to a better characterisation of the disease. In addition, they may have a therapeutic impact in the future given the large number of studies currently underway based on targeted RNA splicing for therapeutic purposes.

Retinitis pigmentosa (RP; MIM# 268000) is the most frequent form of inherited retinal dystrophy (IRD), with a prevalence of 1 in 3000–4000 cases worldwide[1]. It is characterised by a progressive dysfunction associated with the death of rods and/or cones, which leads to retinal atrophy and loss of vision. The mode of inheritance of RP is complex, with autosomal dominant (ad), autosomal recessive (ar), X-linked (xl) Mendelian cases and some cases of digenism or mitochondrial forms having been reported[1–3]. From a genetic perspective, over 80 disease-causing genes are currently associated with RP, 27 of which have been associated with adRP (http://www.sph.uth.tmc.edu/retnet). However, to date, mutations in the known adRP genes account for only 50–75% of dominant cases, depending on the test and population used in the study[4]. This percentage is increasing, mainly due to the implementation of Next Generation Sequencing (NGS)-based techniques[5–7] and the discovery of new RP genes[8–11].

Most human genes harbour introns that are removed during pre-mRNA splicing post-transcriptional modification[12]. The splicing reaction is catalysed by the spliceosome, a multisubunit complex comprising small non-coding nuclear RNAs (U1, U2, U4, U5, and U6) and several associated proteins[13]. The spliceosome orchestrates the two transesterification reactions needed to remove introns and to join the adjacent exons, and operates by step-wise formation of sub-complexes that recognise regulatory sequences and promote efficient splicing[12–14].

Mis-regulation of splicing is a common feature of many human diseases, including several retinal diseases[15–18]. These disorders can be caused by mutations that disrupt the splicing of specific genes or by mutations in genes coding for splicing factors, both of which lead to a general loss of spliceosomal function. Thousands of splice-site mutations have been identified in patients with retinal dystrophies. Although most of these mutations disrupt a consensus splice-site sequence and cause exon skipping, some result in intron inclusion, novel exon inclusion,

[1]Division of Neurosciences, Biodonostia Health Research Institute, San Sebastián, Spain. [2]Department of Ophthalmology, Donostia University Hospital, San Sebastián, Spain. [3]Department of Neurology, Donostia University Hospital, San Sebastián, Spain. [4]CIBERNED, Center for Networked Biomedical Research on Neurodegenerative Diseases, Institute of Health Carlos III, Ministry of Economy and Competitiveness, Spain. [5]Department of Neurosciences, University of the Basque Country UPV-EHU, Spain. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.R.-E. (email: javiruizederra@yahoo.es)

Family	Gene	Mutation	Type	Ref	HSF	Prov	Sift	Ph	Mut TASTER
RP19S	PRPH2	NM_000322c.797G > A p.Gly266Asp	missense	26		D	0	0.99	Disease causing (0.999)
RP22 RP37 RP64 RP101 RP102 RP134 RP157	SNRNP200	NM_014014c.3260C > T p.Ser1087Leu	missense	21, 22		D	0	1	Disease causing (0.999)
RP90	PRPF8	NM_006445c.6974_6994del p.Val2325_Glu2330del	deletion	novel			n/a	n/a	Disease causing (0.999)
RP113	PRPF8	NM_006445c.6945delG p.Leu2315 Leufs*2336 Aspext*21	frameshift	novel			n/a	n/a	Disease causing (1)
RP133 RP146	RHO	NM_000539c.937-1G>T	splice acceptor variant	23	Decrease 5′ acceptor site of exon 5 (90.7>61.75)		n/a	n/a	Disease causing (1)
RP105	RHO	NM_000539c.1045T>C p.Ter349Glu	stop loss	24			n/a	n/a	Polymorphism (0.999)
RP135	RHO	NM_000539c.568G>A p.Asp190Asn	missense	25		D	0	0.431	Disease causing (0.999)

Table 1. Summary of mutations responsible for Retinitis Pigmentosa. Abbreviations: D: deleterious; HSF: human splicing finder; MUT TASTER: Mutation Taster; n/a: not available; PH: Polyphen; PROV: Provean; REF: bibliographic reference. All variants were absent in a Spanish in-house allele database containing information from 578 unrelated Spanish individuals (Spanish controls). See *Materials and Methods* section for detailed information.

or the usage of cryptic upstream or downstream splice sites. The resulting alteration in the protein sequence, which is often concomitant with frameshift and premature termination, unsettles the functional protein domains and leads to degeneration of the retina[16]. For example, mutations in several genes coding for core spliceosomal proteins, such as pre-mRNA splicing factors (PRPF3, PRPF4, PRPF6, PRPF8, PRPF31, RP9) or RNA helicases (SNRNP200), are responsible for adRP[14,16,17]. However, given that these genes are expressed ubiquitously in all tissues and are highly conserved in all eukaryotes, it remains unclear why mutations in these genes are associated exclusively with adRP. Studies performed in rodent retina showed that PRPF3, PRPF31, PRPC8 expression levels are higher in the retina than in other tissues in normal adult mice, thus suggesting that the retina may have a higher basal splicing demand than other tissues given that it is one of the most metabolically active tissues in the body[16,19].

In order to effectively identify adRP mutations, we have sequenced 31 genes associated with the autosomal dominant inheritance pattern using the Ion PGM platform (IPGM; Life Technologies), in combination with Sanger sequencing. We selected these genes as they have been linked to most of the cases of adRP reported. Remarkably, we found a high prevalence of mutations affecting the splicing process among our families, especially mutations affecting *trans*-acting splicing factors. This is of particular interest considering that several splicing-based therapeutic approaches, some of which are in clinical trials[15,17], are under active development for mutations affecting either core spliceosomal proteins or splice site mutations of individual genes.

The results of the present study will help in genetic counselling and will contribute to a better characterisation of the disease. Moreover, they may have a therapeutic impact in the near future in the light of analogous approaches used for other RNA mis-splicing diseases.

Results

High variant detection coverage and sensitivity was achieved. An average of 3.3 million reads/chip was obtained. On average, each amplicon present in the panel was covered 658 times, with 95.92% of amplicons with >30x coverage and 94.27% of amplicons with >50x coverage. Those regions with no or low coverage (<30X), probably due to the presence of repetitive sequences or self-annealing of primers, were re-analysed. A highly sensitive, cost-effective method described recently by us that combines high resolution melting (HRM) analysis with direct sequencing was used for this re-analysis[20]. This allowed us to expand our analysis to 97% of target amplicons. Despite the implementation of HRM, no additional mutations were found within these re-analysed regions.

Variant identification. An average of 45 variants, including SNPs and INDELS, were initially identified for each sample in the targeted regions, including the negative control with 51 SNPs, none of which were putative disease-causing as expected (see Supplementary Table S1). After the clinically relevant variant identification screening described in the materials and methods section, we were able to identify putative disease-causing mutations in a total of 14 out of the 29 probands, which resulted in a ratio of clinically relevant genetic findings of 48.28%. A description of the main features of the genetic findings can be found in Table 1.

A total of seven variants in four genes were found in 14 families. Two of these mutations (both in PRPF8) were novel and were found in two families. One consisted in a loss of 21 nucleotides (p.Val2325_Glu2331del) and the other consisted of a frameshift deletion involving a single-point deletion (p.Leu2315Leufs*2336Aspext*21). Figure 1 shows colour fundus pictures of patients RP90 and RP113 bearing these two novel mutations. Both novel variants were potentially pathogenic, co-segregated with the disease, and were predicted as pathogenic by MutationTaster.

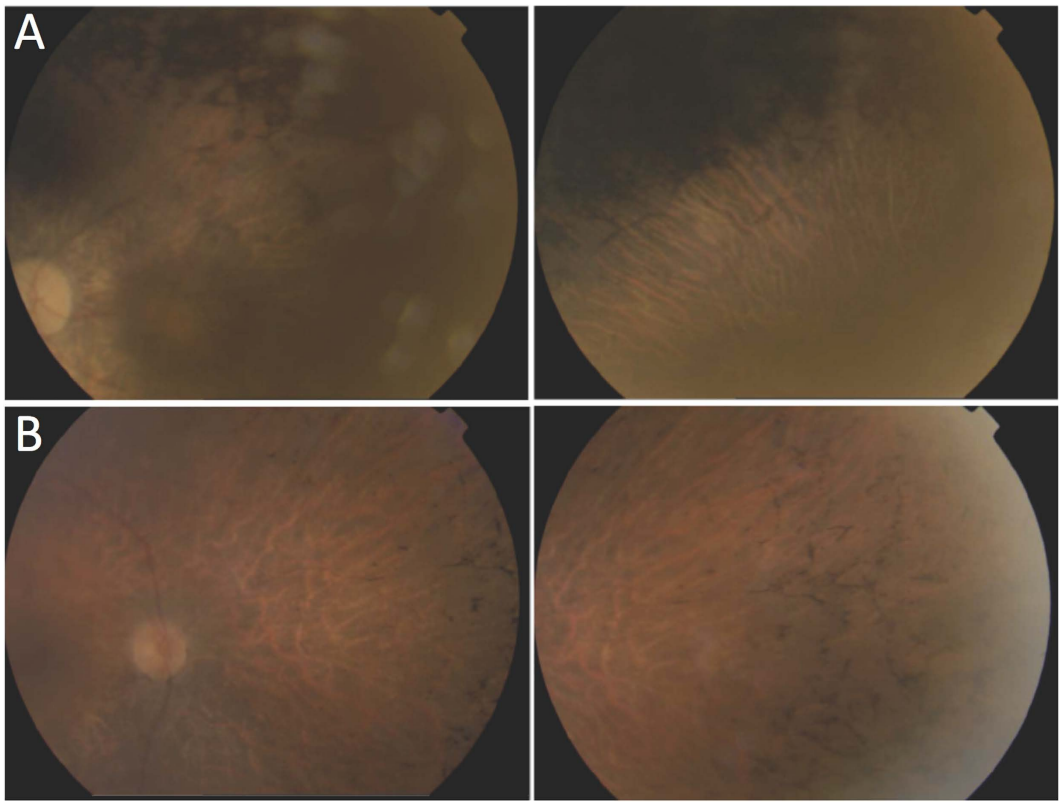


Figure 1. Fundus photographs of patients with novel mutations in PRPF8. (**A**) Patient RP90 (p.Val2325_Glu2330del) shows optical disc pallor, arteriolar attenuation and macular atrophy (right), with dense pigment in the mid-periphery (left). (**B**) Patient RP148 (p.Leu2315Leufs*2336Aspext*21) shows optical disc pallor, arteriolar attenuation and bone spicule-shaped pigment deposits in the mid-periphery. The left and right pictures correspond to the left and right eyes, respectively.

Two genes were involved in 37.93% of our cohort of families, with RHO affecting four probands with three different mutations and SNRNP200 affecting seven probands, all with the p.Ser1087Leu mutation[21,22].

The high prevalence of mutations affecting the splicing process among our families (11 out of 29 probands studied), representing 38% of the probands in our adRP cohort, was unexpected. Most cases (9/29) were due to mutations affecting the genes SNRNP200 (7) and PRPF8 (2), which code for core spliceosomal proteins, although a splice site mutation in RHO[23] was also detected (2/29).

With respect to SNRNP200, after performing Sanger sequencing in all available family members we identified c.3260C > T mutation in a total of 12 cases from seven families (see representative family in Fig. 2A). Co-segregation analysis showed that two out of seven healthy subjects analysed for this variant in these families were mutation carriers, which likely indicates cases of incomplete penetrance similar to what has recently been reported for this variant in a study also involving a Spanish cohort[7] (see Fig. 2B). We also found a total of nine individuals in two families with c.937-1G > T mutations affecting RHO splicing. Interestingly, one of these nine patients is asymptomatic, probably due to the disease being in an initial state given his young age (21 years old; see Fig. 2C and Supplementary Fig. S1).

Finally we also found mutations in both RHO and PRPH2 genes that were not related to the splicing process: a stop loss in RP105[24] and a missense mutation in RP135[25], both in RHO, and a missense mutation in PRPH2 (p.Gly266Asp) in patient RP19S[26]. Patient RP19S was included in this study since he is the son of a patient with a mutation in PRPH2 that we had diagnosed previously[20]. Patient RP19S was asymptomatic at the initial diagnosis, when he was eight years old. However, two years later his molecular diagnosis confirmed the presence of the p.Gly266Asp mutation, therefore he was re-examined. This revealed a granular fundus and few bone spicules in the inferior periphery, with no signs of optical disc pallor or vascular attenuation. The visual field showed a concentric defect (preserving the central 18 degrees) with a hyperautofluorescent ring in the macula upon autofluorescence examination (see Supplementary Fig. S2). Additional family trees of the rest of the patients recruited in the present study are included in Supplementary Fig. S3.

Discussion

In this work we have analysed the genotype and phenotype of a group of 29 adRP probands, using targeted NGS and Sanger sequencing to analyse 31 genes. We were able to detect putative disease-causing mutations in 14 out of the 29 probands analysed. This resulted in a clinically relevant genetic diagnosis ratio of 48.28%, which is comparable to values reported previously, ranging from about 24% to 88%[6,7,27–33]. Several factors may be responsible

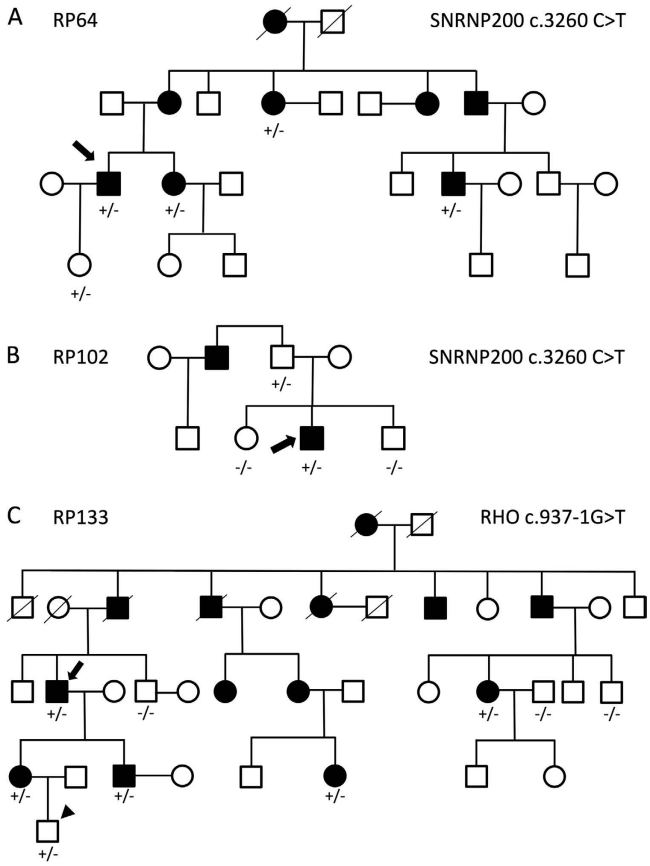


Figure 2. Representative trees for families with the two most prevalent mutations found in SNRNP200 and RHO genes. The p.Ser1087Leu mutation in SNRNP200 was found in families RP64 (**A**) and RP102 (**B**). (**C**) The c.937-1G > T mutation in the RHO splice acceptor site in a total of six individuals from family RP133, one of whom is a young asymptomatic patient (arrowhead). Genotypes are annotated as +/− (heterozygote) or −/− (wild type). Arrows indicate proband patients.

for this wide range of diagnosis ratios reported, including the approach used or the nature of the cohort involved. In the present study, part of our cohort of adRP patients was already diagnosed in a previous study in which we screened some of the most prevalent adRP genes[14,20], therefore this might have contributed to the diagnostic ratio obtained.

Nevertheless there is still a missing fraction of about 51% unsolved cases among our adRP cohort of 29 patients. One possible explanation is the presence of mutations in regions outside the 31 genes analysed, such as deep intronic regions. Another possibility is the presence of changes not detected by our analysis due to limitations in the design of our panel of target genes, such as large genomic rearrangements and mutations in novel genes. As such, it seems that the combination of NGS with other technologies, such as Multiplex Ligation-dependent Probe Amplification (MLPA) or Comparative Genomic Hybridisation arrays (aCGH), will be needed in order to address those genomic aberrations caused by copy number variations (CNV). Another possible explanation is the presence of novel RP genes among our patients, since most of them belong to the Basque province of Gipuzkoa, a well-known genetically homogeneous region[34]. Consequently, sequencing of the whole exome/genome could help in the discovery of novel RP genes.

A remarkable finding was the high prevalence of mutations affecting the splicing process among our families (11 out of 29 probands studied), representing 38% of the probands in our adRP cohort.

Most mutations were the Ser1087Leu mutation found in SNRNP200. This gene encodes for the 200-kDa helicase hBrr2. During splicing, the spliceosome undergoes structural rearrangements that are mediated by several RNA helicases including hBrr2, which is essential for unwinding of the U4/U6 snRNP duplex, a key step in the catalytic activation of the spliceosome complex[35,36]. hBrr2 comprises two helicase modules, one active and the other with regulatory activity.

All six mutations identified in SNRNP200 to date, including the Ser1087Leu mutation, are located in the hBrr2 protein region containing the first DExD-helicase module, a key component for the U4-U6 unwinding function *in vivo* and *in vitro* and for cell survival[35–37]. The first of the two consecutive Hel308-like modules, which comprises a DExD/H domain and a Sec63 domain, shows the highest level of conservation among species, thus pointing to its functional relevance[38]. The Ser1087Leu mutation has been reported to reduce unwinding activity and to promote the use of cryptic splice sites, thus pointing to an influence of splicing fidelity[22,39].

Although most cases (9/29) were due to mutations affecting genes SNRNP200 and PRPF8 that code for spliceosomal proteins, splice-site mutations in RHO were also detected (2/29). The percentage of adRP probands with mutations affecting either spliceosome core factors or the splice site of several adRP genes accounted for 5–14.5% of all cases of adRP in previous studies[4,7,40,41]. With regard to mutations in the SNRNP200 gene, although these were only initially described in two Chinese families[21,22], they have since been reported to contribute to a significant portion of cases of adRP in the Caucasian population, ranging from 1.5% to 4.2%[4,40,42,43].

The relatively high prevalence of splicing-related mutations found in our study is likely explained by the founder effect of two of the genes, which were present in very small and rather isolated Spanish populations.

Splicing modulation has been proposed as a therapeutic approach for several diseases. Two of the most advanced approaches in this regard are based on the use of modified antisense oligonucleotides (ASOs) to target specific RNA sequences and redirect splicing, and small molecules as modulators of the splicing process. A representative example of this approach is exon skipping for Duchenne muscular dystrophy (DMD), where the muscular protein dystrophin is prematurely truncated by mutations that disrupt the open reading frame, thus leading to a non-functional protein. Exon skipping creates an internally deleted and shorter than normal but partially functional protein, which leads to a much less severe phenotype in animal models of DMD. With respect to approaches based on small molecules and peptides, several splicing modulators have been shown to be effective in myotonic dystrophy (DM) and cancer[18,44].

As regards retinal dystrophies, most advanced therapeutic approaches that target splicing are aimed at correcting the splicing of individual genes using mutation-adapted U1 small nuclear RNA for the RPGR gene[45] or spliceosome-mediated RNA trans-splicing in RHO[46]. Both these approaches are based on cellular and animal models and have provided encouraging results. Once in the clinic, these promising approaches could be generalised and applied to other genes with splice donor site mutations[45] and to all adRP genes rather than only to RPGR and RHO, respectively[46].

With regard to therapeutic approaches targeting the splicing machinery, we are unaware of their use in retinal diseases. However, since the first steps towards the use of such therapeutic strategies have already been made for other diseases, it is plausible to imagine a broadening of the applications of small molecules to reverse aberrant splicing for other diseases, including retinal dystrophies, in the near future once our understanding of the mechanisms of the disease, and delivery systems and other technical issues, have been improved.

In summary, the combination of NGS with Sanger sequencing has allowed us to achieve a diagnostic rate of over 48%. As such, the methodology described herein exhibits a high diagnostic yield when applied to a well-defined adRP group and a relatively high number of genes. This will be of clinical relevance once ongoing studies on therapeutic options directed at manipulating splicing are completed.

Materials and Methods

Study subjects. RP patients were diagnosed at the Ophthalmology Department of Donostia University Hospital (San Sebastian, Spain). Diagnostic criteria were night blindness, peripheral visual field loss, pigmentary deposits resembling bone spicules, attenuation of retinal vessels, pallor of the optic disc and diminution in a- and b-wave amplitudes in the electroretinogram[47]. A total of 29 Spanish probands with a family tree compatible with adRP were included. Samples from an additional four patients, three corresponding to patients with known mutations that we had detected in previous analysis and one from a non-affected individual, were included as positive and negative controls, respectively[14,20]. Family trees were generated from information obtained from probands. All procedures performed in studies involving human participants received approval from the institutional research ethics committee and were in accordance with the Declaration of Helsinki (2013) or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. For a detailed description of clinical features of all patients recruited in the present study see Supplementary Table S2.

Human sample collection. High molecular weight DNA was extracted from blood samples from RP patients and their available family members. Total DNA from samples was extracted and isolated using an AutoGenFlex STAR instrument (AutoGen, Holliston, MA, USA) together with the FlexiGene DNA Kit (Qiagen, Hilden, Germany) following the manufacturer´s instructions. DNA concentrations were measured using a Nanodrop spectrophotometer (ND-1000, Thermo Scientific NanoDrop Products, Wilmington, DE, USA) an only those samples with 260/280 ratios $\geq$1.8 and 260/230 ratios $\geq$2 were used. DNA samples were stored at −80 °C.

Amplicon Library preparation. A total of 663 primer pairs were designed and grouped in two Ion AmpliSeq Primer Pools to flank 31 IRD genes with a total coverage of 98.37% using the Ion AmpliSeq Designer software (www.ampliseq.com). The regions excluded by the design represented only 1.63% of the total. Although most of the genes were related to adRP, representative genes associated with dominant forms of Leber congenital amaurosis and cone-rod dystrophies were also included since the clinic symptoms associated with these genes are often hard to distinguish from those associated with RP (RetNet; https://sph.uth.edu/retnet/disease.htm) (see Supplementary Table S3). The Ion AmpliSeq Library Preparation Kit v2.0 (Life Technologies, Foster City, CA, USA) was used to construct an amplicon library from genomic target regions with a maximum read length of approximately 200 base pairs (average length, 142 bp) for shotgun sequencing on the PGM. Briefly, target genomic regions were amplified by simple PCR using Ion AmpliSeq Primer Pools and 10 ng of each genomic DNA samples.

Sequencing Analysis. *Ion Torrent Personal Genome Machine (PGM).* NGS was carried out on a PGM following the Ion PGM 200 Sequencing Kit protocol. Briefly, enriched Ion Sphere particles (ISPs) were annealed with the Ion Sequencing primer and mixed with the PGM200 Sequencing Polymerase. The polymerase-bound and primer-activated ISPs were then loaded into the previously checked and washed Ion 316 Chips (Life

Technologies) and, after selecting the run plan on the Ion PGM System software, these chips were subjected to 500 cycles of sequencing with the standard nucleotide flow order. Signal processing and base calling for the data generated during the PGM runs were performed using the Ion Torrent platform-specific analysis software Torrent Suite version 4.0 to generate sequence reads. The sequences generated were aligned to the GRCh37/hg19 human genome for detection of genomic variants in the sequenced samples.

Sanger sequencing. Sanger sequencing was used to confirm those mutations detected by NGS and for co-segregation analysis. Primers were designed at least 60 bp upstream and downstream of the mutation. The amplicons were purified after PCR amplification, (ExoSAP-IT, USB Corporation). Sequencing was performed by dye termination DNA reaction on a 16-capillary ABI 3130xl platform (Applied Biosystems) according to the manufacturer's protocol. Sequences were analysed and compared with wild-type samples and reference sequences using the BioEdit Sequence Alignment Editor (Windows) and Ensembl and NCBI databases.

High resolution melting (HRM) analysis. HRM analysis was used to re-analyse those genomic regions with no or very low coverage in NGS platforms, following the previously described methodology[20].

Relevant variant identification and pathogenicity score. In order to determine genomic variants of relevance, we selected putative disease-causing variants using the following criteria: 1) variants previously reported as pathogenic, or 2) loss-of-function variants, such as stop gain, frameshift, small deletions or duplications (INDELS) and splice site variants, or 3) novel missense variants predicted to be damaging or highly pathogenic in at least four out of five web-based pathogenicity predictors, namely SIFT (<0.05), Polyphen2 (>0.750); PROVEAN[48]; GVGD[49]; MutationTaster[50]. Furthermore, all variants selected had to fulfil the criteria of having a Minor Allele Frequency (MAF) of less than 0.002, as obtained from human genome databases (see below), and being absent from Spanish in-house allele database with information from 578 unrelated Spanish individuals none of whom exhibited any IRD-related disease[51] (http://csvs.babelomics.org/; see Supplementary Fig. S4).

References

1. Hartong, D. T., Berson, E. L. & Dryja, T. P. Retinitis pigmentosa. *Lancet* **368,** 1795–1809, doi: S0140-6736(06)69740-7 [pii]10.1016/S0140-6736(06)69740-7 (2006).
2. Dryja, T. P., Hahn, L. B., Kajiwara, K. & Berson, E. L. Dominant and digenic mutations in the peripherin/RDS and ROM1 genes in retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **38,** 1972–1982 (1997).
3. Mansergh, F. C. *et al.* Retinitis pigmentosa and progressive sensorineural hearing loss caused by a C12258A mutation in the mitochondrial MTTS2 gene. *Am. J. Hum. Genet.* **64,** 971–985 (1999).
4. Daiger, S. P., Bowne, S. J. & Sullivan, L. S. Genes and Mutations Causing Autosomal Dominant Retinitis Pigmentosa. *Cold Spring Harb. Perspect. Med.* **5,** doi: 10.1101/cshperspect.a017129 (2014).
5. Anasagasti, A., Irigoyen, C., Barandika, O., Lopez de Munain, A. & Ruiz-Ederra, J. Current mutation discovery approaches in Retinitis Pigmentosa. *Vision Res.* **75,** 117–129, doi: S0042-6989(12)00303-3 [pii]10.1016/j.visres.2012.09.012 (2012).
6. Daiger, S. P. *et al.* Application of Next-Generation Sequencing to Identify Genes and Mutations Causing Autosomal Dominant Retinitis Pigmentosa (adRP). *Advances in experimental medicine and biology* **801,** 123–129, doi: 10.1007/978-1-4614-3209-8_16 (2014).
7. Fernandez-San Jose, P. *et al.* Targeted Next-Generation Sequencing Improves the Diagnosis of Autosomal Dominant Retinitis Pigmentosa in Spanish Patients. *Invest. Ophthalmol. Vis. Sci.* **56,** 2173–2182, doi: 10.1167/iovs.14-16178 (2015).
8. Wang, F. *et al.* A missense mutation in HK1 leads to autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **55,** 7159–7164, doi: 10.1167/iovs.14-15520 (2014).
9. Ma, X. *et al.* Whole-exome sequencing identifies OR2W3 mutation as a cause of autosomal dominant retinitis pigmentosa. *Scientific reports* **5,** 9236, doi: 10.1038/srep09236 (2015).
10. Liu, Y. *et al.* SPP2 Mutations Cause Autosomal Dominant Retinitis Pigmentosa. *Scientific reports* **5,** 14867, doi: 10.1038/srep14867 (2015).
11. Chen, X. *et al.* PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **23,** 2926–2939, doi: 10.1093/hmg/ddu005 (2014).
12. Dewey, F. E. *et al.* Clinical interpretation and implications of whole-genome sequencing. *JAMA*: *the journal of the American Medical Association* **311,** 1035–1045, doi: 10.1001/jama.2014.1717 (2014).
13. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–718, doi: 10.1016/j.cell.2009.02.009 (2009).
14. Barandika, O. *et al.* A Cost-Effective Mutation Screening Strategy for Inherited Retinal Dystrophies. *Ophthalmic research*, doi: 10.1159/000445690 (2016).
15. Havens, M. A., Duelli, D. M. & Hastings, M. L. Targeting RNA splicing for disease therapy. *Wiley interdisciplinary reviews. RNA* **4,** 247–266, doi: 10.1002/wrna.1158 (2013).
16. Liu, M. M. & Zack, D. J. Alternative splicing and retinal degeneration. *Clin. Genet.* **84,** 142–149, doi: 10.1111/cge.12181 (2013).
17. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17,** 19–32, doi: 10.1038/nrg.2015.3 (2016).
18. Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends in molecular medicine* **18,** 472–482, doi: 10.1016/j.molmed.2012.06.006 (2012).
19. Cao, H. *et al.* Temporal and tissue specific regulation of RP-associated splicing factor genes PRPF3, PRPF31 and PRPC8–implications in the pathogenesis of RP. *PLoS One* **6,** e15860, doi: 10.1371/journal.pone.0015860 (2011).
20. Anasagasti, A. *et al.* Genetic high throughput screening in Retinitis Pigmentosa based on high resolution melting (HRM) analysis. *Experimental eye research* **116,** 386–394 (2013).
21. Zhao, C. *et al.* A novel locus (RP33) for autosomal dominant retinitis pigmentosa mapping to chromosomal region 2cen-q12.1. *Hum. Genet.* **119,** 617–623, doi: 10.1007/s00439-006-0168-3 (2006).
22. Zhao, C. *et al.* Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs. *Am. J. Hum. Genet.* **85,** 617–627, doi: 10.1016/j.ajhg.2009.09.020 (2009).
23. Bell, C., Converse, C. A., Hammer, H. M., Osborne, A. & Haites, N. E. Rhodopsin mutations in a Scottish retinitis pigmentosa population, including a novel splice site mutation in intron four. *The British journal of ophthalmology* **78,** 933–938 (1994).
24. Hollingsworth, T. J. & Gross, A. K. The severe autosomal dominant retinitis pigmentosa rhodopsin mutant Ter349Glu mislocalizes and induces rapid rod cell death. *The Journal of biological chemistry* **288,** 29047–29055, doi: 10.1074/jbc.M113.495184 (2013).
25. Keen, T. J. *et al.* Autosomal dominant retinitis pigmentosa: four new mutations in rhodopsin, one of them in the retinal attachment site. *Genomics* **11,** 199–205 (1991).

26. Sohocki, M. M. *et al.* Prevalence of mutations causing retinitis pigmentosa and other inherited retinopathies. *Hum. Mutat.* **17,** 42–51, doi: 10.1002/1098-1004(2001)17:1<42::AID-HUMU5>3.0.CO;2-K[pii]10.1002/1098-1004(2001)17:1<42::AID-HUMU5>3.0.CO;2-K (2001).
27. Bowne, S. J. *et al.* Identification of disease-causing mutations in autosomal dominant retinitis pigmentosa (adRP) using next-generation DNA sequencing. *Invest. Ophthalmol. Vis. Sci.* **52,** 494–503, doi: 10.1167/iovs.10-6180 (2011).
28. Audo, I. *et al.* Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J. Rare Dis.* **7,** 8, doi: 10.1186/1750-1172-7-8 (2012).
29. Sullivan, L. S. *et al.* A dominant mutation in hexokinase 1 (HK1) causes retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **55,** 7147–7158, doi: 10.1167/iovs.14-15419 (2014).
30. Eisenberger, T. *et al.* Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One* **8,** e78496, doi: 10.1371/journal.pone.0078496 (2013).
31. Glockle, N. *et al.* Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur. J. Hum. Genet.* **22,** 99–104, doi: 10.1038/ejhg.2013.72 (2014).
32. Oishi, M. *et al.* Comprehensive molecular diagnosis of a large cohort of Japanese retinitis pigmentosa and Usher syndrome patients by next-generation sequencing. *Invest. Ophthalmol. Vis. Sci.* **55,** 7369–7375, doi: 10.1167/iovs.14-15458 (2014).
33. Gonzalez-del Pozo, M. *et al.* Mutation screening of multiple genes in Spanish patients with autosomal recessive retinitis pigmentosa by targeted resequencing. *PLoS One* **6,** e27894, doi: 10.1371/journal.pone.0027894PONE-D-11-14335 [pii] (2011).
34. Champe, S. P., Rao, P. & Chang, A. An endogenous inducer of sexual development in Aspergillus nidulans. *Journal of general microbiology* **133,** 1383–1387, doi: 10.1099/00221287-133-5-1383 (1987).
35. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8,** 847–855 (1998).
36. Ruzickova, S. & Stanek, D. Mutations in spliceosomal proteins and retina degeneration. *RNA Biol.* 1–9, doi: 10.1080/15476286.2016.1191735 (2016).
37. Kim, D. H. & Rossi, J. J. The first ATPase domain of the yeast 246-kDa protein is required for *in vivo* unwinding of the U4/U6 duplex. *RNA* **5,** 959–971 (1999).
38. Zhang, L. *et al.* Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. *Nat. Struct. Mol. Biol.* **16,** 731–739, doi: 10.1038/nsmb.1625 (2009).
39. Cvackova, Z., Mateju, D. & Stanek, D. Retinitis pigmentosa mutations of SNRNP200 enhance cryptic splice-site recognition. *Hum. Mutat.* **35,** 308–317, doi: 10.1002/humu.22481 (2014).
40. Coussa, R. G. *et al.* Genotype and Phenotype Studies in Autosomal Dominant Retinitis Pigmentosa (adRP) of the French Canadian Founder Population. *Invest. Ophthalmol. Vis. Sci.* **56,** 8297–8305, doi: 10.1167/iovs.15-17104 (2015).
41. Sullivan, L. S. *et al.* Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. *Invest. Ophthalmol. Vis. Sci.* **47,** 3052–3064, doi: 10.1167/iovs.05-1443 (2006).
42. Benaglio, P. *et al.* Next generation sequencing of pooled samples reveals new SNRNP200 mutations associated with retinitis pigmentosa. *Hum. Mutat.* **32,** E2246–2258, doi: 10.1002/humu.21485 (2011).
43. Bowne, S. J. *et al.* Mutations in the small nuclear riboprotein 200 kDa gene (SNRNP200) cause 1.6% of autosomal dominant retinitis pigmentosa. *Mol. Vis.* **19,** 2407–2417 (2013).
44. Nakajima, H. *et al.* New antitumor substances, FR901463, FR901464 and FR901465. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. *The Journal of antibiotics* **49,** 1196–1203 (1996).
45. Glaus, E., Schmid, F., Da Costa, R., Berger, W. & Neidhardt, J. Gene therapeutic approach using mutation-adapted U1 snRNA to correct a RPGR splice defect in patient-derived cells. *Molecular therapy*: *the journal of the American Society of Gene Therapy* **19,** 936–941, doi: 10.1038/mt.2011.7 (2011).
46. Berger, A. *et al.* Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: a new approach for autosomal dominant retinitis pigmentosa. *Molecular therapy*: *the journal of the American Society of Gene Therapy* **23,** 918–930, doi: 10.1038/mt.2015.11 (2015).
47. Hamel, C. Retinitis pigmentosa. *Orphanet J. Rare. Dis.* **1,** 40, doi: 1750-1172-1-40 [pii]10.1186/1750-1172-1-40 (2006).
48. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7,** e46688, doi: 10.1371/journal.pone.0046688 (2012).
49. Mathe, E. *et al.* Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic acids research* **34,** 1317–1325, doi: 10.1093/nar/gkj518 (2006).
50. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* **7,** 575–576, doi: 10.1038/nmeth0810-575 (2010).
51. Alonso, R. *et al.* Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic acids research* **43,** W117–121, doi: 10.1093/nar/gkv384 (2015).

Acknowledgements

This work was supported by grants from the National Institute of Health Carlos III (Institute of Health Carlos III/ISCIII) (CP10/00572 and PI13/02621 to JRE and a Research Intensification Contract to ALdM); the Basque Government's Industry Department (SAIOTEK: SAIO11-PE11BN002; and SAIO12-PC12BN001 to JRE), and an unrestricted grant from the Retinitis Pigmentosa Patients of Gipuzkoa Foundation (BEGISARE). JR-E is a Miguel Servet Fellow, National Institute of Health Carlos III (ISCIII). MEI was supported by grants from the Basque Government's Department of Education (DEDUC14/309). OB is supported by funding from the Retinitis Pigmentosa Patients of Gipuzkoa Foundation (BEGISARE) and a grant from the Mutua Madrileña Foundation. AA was supported by grants from the Fundación Jesús de Gangoiti Barrera and from the Basque Government's Departments of Industry and Education (SAIOTEK-11BN002/PC12BN001/DEPLC13/002). CI is partially supported by a Research Intensification Contract (INTBIO15/001). The authors thank David Otaegi for helpful advice on PGM use, Maribel Gomez, Nahiara Telletxea, Nahikari Pastoriza, and Aroa Zabalo for processing the samples, and Ana Gorostidi and Olaia Zurriarain for helping with DNA library preparation.

Author Contributions

M.E.I.: analysed part of the NGS data and drafted parts of the manuscript. O.B.: analysed most NGS data. A.A.: drafted parts of the manuscript and helped to interpret the results. C.I.: selected the patients and collected the clinical data. ALdM: interpreted the results and revised the manuscript. J.R.E.: planned the experiments, interpreted the results, wrote the manuscript and obtained funding for the project. All authors revised and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ezquerra-Inchausti, M. *et al.* High prevalence of mutations affecting the splicing process in a Spanish cohort with autosomal dominant retinitis pigmentosa. *Sci. Rep.* **7**, 39652; doi: 10.1038/srep39652 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Biochimica et Biophysica Acta 1862 (2016) 2015–2021

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbadis

Increased aquaporin 1 and 5 membrane expression in the lens epithelium of cataract patients

Olatz Barandika BSc [a,1], Maitane Ezquerra-Inchausti BSc [a,1], Ander Anasagasti BPharm [a], Ainara Vallejo-Illarramendi PhD [a], Irantzu Llarena [b], Lucia Bascaran MD [c], Txomin Alberdi MD [c], Giacomo De Benedetti MD [d], Javier Mendicute MD, PhD [c], Javier Ruiz-Ederra PhD [a,*]

[a] Division of Neuroscience, Biodonostia HRI, San Sebastián, Spain
[b] Optical Spectroscopy Platform, CIC biomaGUNE, San Sebastián, Spain
[c] Ophthalmology Service, Donostia University Hospital, San Sebastián, Spain
[d] Ophthalmology Service, Quirónsalud Donostia Hospital, San Sebastián, Spain

ARTICLE INFO

Article history:
Received 8 April 2016
Received in revised form 4 July 2016
Accepted 2 August 2016
Available online 4 August 2016

Keywords:
Cataract
Lens transparency
Aquaporins
AQP1
AQP5

ABSTRACT

In this work we have analyzed the expression levels of the main aquaporins (AQPs) expressed in human lens epithelial cells (HLECs) using 112 samples from patients treated with cataract surgery and 36 samples from individuals treated with refractive surgery, with transparent lenses as controls. Aquaporin-1 (AQP1) is the main AQP, representing 64.1% of total AQPs in HLECs, with aquaporin-5 (AQP5) representing 35.9% in controls. A similar proportion of each AQP in cataract was found. Although no differences were found at the mRNA level compared to controls, a significant 1.65-fold increase ($p = 0.001$) in AQP1protein expression was observed in HLECs from cataract patients, with the highest differences being found for nuclear cataracts (2.1-fold increase; $p < 0.001$). A similar trend was found for AQP5 (1.47-fold increase), although the difference was not significant ($p = 0.161$). Moreover we have shown increased membrane AQP5 protein expression in HLECs of patients with cataracts. No association of AQP1 or AQP5 expression levels with age or sex was observed in either group. Our results suggest regulation of AQP1 and AQP5 at the post-translational level and support previous observations on the implication of AQP1 and 5 in maintenance of lens transparency in animal models. Our results likely reflect a compensatory response of the crystalline lens to delay cataract formation by increasing the water removal rate.

1. Introduction

Cataract is a leading cause of blindness, affecting about 18 million people worldwide [1,2]. It is estimated that 1.3 million cataract operations are performed annually in the U.S. In the Spanish National Public Health System, cataract surgery is the most frequent outpatient surgery performed, with over 260,000 cataract operations each year [3]. In order to be able to develop new alternatives to cataract surgery to prevent, or at least delay, cataract, it is necessary to gain a more in-depth understanding of the pathogenic mechanisms involved in this ocular condition.

The lens is an avascular tissue composed of concentric layers of epithelial cells at various stages of differentiation [4,5]. An epithelial cell monolayer extends from the anterior pole of the lens to its equatorial surface, surrounding the elongated lens fibers, which are arranged with the oldest fibers in the lens nucleus. Upon maturation, lens fibers lose their attachment to the capsule, and cellular organelles are degraded in a synchronized manner [6]. Nourishment is provided to the lens by diffusion from the aqueous and vitreous humors. However, it is unlikely that simple diffusion can sustain the metabolic needs of the lens interior [7]. As such, a *circulatory system* in which an asymmetric distribution of ion pumps, transporters, channels and cell junctions drive ion-coupled fluid absorption, thereby facilitating the entry of nutrients and metabolites into the inner lens across the polar regions and exit through the lens equator, has been proposed [7–10]. The lens contains a uniquely high protein concentration and low water content. This tightly packed arrangement of fibers helps maintain an elevated refractive index for transparency, with lens water channels proposed to act by facilitating water removal [11].

The aquaporins (AQPs) are small integral membrane proteins (~30 kDa/monomer) expressed widely in both the animal and plant

Abbreviations: AQP, aquaporin; HLECs, human lens epithelial cells.

* Corresponding author at: Department of Neuroscience, Biodonostia HRI, Paseo Dr. Begiristain s/n, E-20014 San Sebastián, Spain.
E-mail addresses: javiruizederra@yahoo.es, javier.ruizederra@osakidetza.net (J. Ruiz-Ederra).

[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.bbadis.2016.08.001
0925-4439/

kingdoms, with 13 members in mammals. AQPs are expressed in epithelia and endothelia, where they are involved in fluid transport, but are also found in other cell types such as skin and fat cells, where they have other functions. In most cell types, the AQPs reside constitutively at the plasma membrane. One exception is kidney AQP2, which undergoes vasopressin-regulated trafficking between endosomes and the cell plasma membrane. Three AQPs are expressed in the lens: AQP0 (major intrinsic protein-MIP) found in the posterior pole and in nuclear fibers; AQP1 at the anterior pole in epithelial cells; and AQP5 recently described in both epithelial and fiber cells. Similar to AQP0, the distribution of AQP5 within lens cells has been reported to change as a function of fiber cell differentiation [12].

Mutations in AQP0 are associated with hereditary cataracts in mice and humans [13,14]. Indeed, cataract-producing AQP0 mutations are thought to produce endoplasmic reticulum-retained and non-functional AQP0 [15,16], although the mechanism linking AQP0 loss-of-function and cataracts remains unclear. Because of its low water permeability, it has been proposed that AQP0 might be involved in regulating the resistance of the paracellular pathway, rather than in cell membrane water permeability [5,17]. Therefore, proposed mechanisms for the implication of mutations in AQP0 in cataract include loss of AQP0-facilitated fiber-fiber adherence [14] and impaired fiber cell dehydration [18].

With respect to AQP1, cataracts were not reported in human subjects with AQP1 deficiency [19], and spontaneous cataracts are not seen grossly in AQP1 null mice [20]. Nevertheless, based on experimental observations of the role of AQP1 in the cornea or in the lens [21,22], we reported a functional implication of lens AQP1 in lens transparency [21], with epithelial cell water permeability being approximately three-fold lower in lenses from AQP1 null mice. Moreover, although AQP1 deletion did not alter baseline lens morphology or transparency, basal water content was significantly higher (by approx. 4%) in AQP1 null mice, and the involvement of AQP1 in cataract development was studied using in vitro and in vivo models, which showed that AQP1 facilitates the maintenance of lens transparency and opposes cataract formation, thereby suggesting the possibility of AQP1 induction to delay cataractogenesis [21].

With regard to AQP5, no cataract phenotype for AQP5 knockout animals has been reported [23], although AQP5 deficiency has recently been linked to cataractogenesis in an ex vivo hyperglycemic mouse model of cataract formation [24]. In humans, AQP5 deficiency has not been associated with cataract. Two recent studies have related AQP5 mutations with non-ocular phenotypes (palmoplantar keratoderma), but with an unclear implication of these mutations in the water permeability of AQP5 [25,26].

In the present work we have shown increased membrane AQP1 and 5 protein expression in the human lens epithelial cells (HLECs) of patients with cataracts, with no changes observed at the mRNA level. Considering the low water content of the lens that is required to maintain a high refractive index for transparency, our results might reflect a compensatory response in an attempt to increase the water removal rate. Furthermore, our results suggest the possibility of increasing AQP1 and/or AQP5 expression levels in the lens epithelial cell membrane by pharmacologic or genetic means as a treatment for cataracts, at least in their early stages, once AQP modulators become available.

2. Materials and methods

2.1. Study subjects

A total of 148 individual of both sexes from Donostia Teaching Hospital, the Begitek Ophthalmologic Hospital, and the Quironsalud Donostia Hospital were included in the study: 112 cataract samples were obtained from patients submitted to a cataract intervention and 36 samples from clear lens were obtained from patients undergoing refractive lens exchange surgery (these served as controls). The mean age of patients was 67.21 ± 6.04 years, ranging from 42 to 82 years for cataract patients, and 57.48 ± 5.47 years, ranging from 49 to 70 years, for controls. All patients gave written informed consent, and the research adhered to the tenets of the Declaration of Helsinki. Institutional Ethics Review Committee approval was obtained.

A total of 70/148 samples were used for analysis of mRNA expression levels by quantitative polymerase chain reaction (qPCR), with 64/148 samples being used for protein expression levels by immunoblot analysis, and 14/148 samples being used for immunolocalization analysis.

Cataracts were classified as cortical, nuclear, or posterior subcapsular according to clinical criteria. The cataract samples used included only cataracts with stage NC2-NC4, C2-C3, P2P3 in the LOCS III system [27], whereas all controls had clear lenses. The main clinical information is summarized in Supplementary Table 1.

2.2. Tissue collection

Femtosecond laser-assisted lens removal was performed using the Victus femtosecond laser platform (Bausch & Lomb, New York, USA). A 5.0 mm capsulotomy applying 7.0 μJ was performed, and circular sections of anterior lens capsules with attached anterior lens epithelial cells (HLECs) were isolated and stored at −80 °C until RNA/protein extraction.

2.3. RNA quantification by real-time quantitative PCR

Total RNA was extracted from single HLEC samples using the RNeasy Micro Kit following DNase Treatment (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Total RNA concentration was determined by spectrophotometry using a Nanodrop (Thermo Fisher Scientific) and cDNA was synthesized using the SuperScript VILO cDNA Synthesis Kit (Invitrogen, Thermo Fisher Scientific). Candidate genes were analyzed by qPCR as described previously [28]. Briefly, primers spanning exon-exon junctions were designed using the Primer Express Software (Applied Biosystems; see Supplementary Table 2) and specificity was verified by ePCR (http://www.ncbi.nlm.nih.gov/sutils/e-pcr/reverse.cgi). qPCR was carried out in triplicate using the 7900HT Fast Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific) according to the manufacturer's instructions. Gene expression was calculated in cataract versus control samples using the standard curve method. Target genes were normalized by means of a normalization factor, based on the geometric mean of three internal control genes [29]. This normalization factor was calculated using the expression levels of the three best-scoring genes, which were GAPDH, TUBA1B, and ACTB (see Supplementary Table 2). Data are expressed as fold change of gene expression in cataract versus control samples.

2.4. Immunoblot analysis

Isolated control and cataract HLECs were homogenized in 100 μL of loading buffer (62.5 mM of 1 M Tris pH 7.5, 5% glycerol, 1% bromophenol blue, 2% SDS and 5% β-mercaptoethanol). Homogenates were boiled for 5 min and supernatants were centrifuged at 5000*g* before loading in SDS-PAGE 4–20% polyacrylamide-gradient gel (Mini-Protean TGX; Bio-Rad, Hercules, CA, USA). One control sample was used in all gels for normalization purposes. Given the limited amount of sample from each patient, proteins were extracted directly with a loading sample buffer, in order to maximize protein extraction. To determine the range of linearity for GAPDH and AQPs, a standard curve was generated with 2-fold dilutions of a control sample that showed high expression levels of AQPs. Only those samples with GAPDH and AQP intensity signals falling within these values were considered for the analysis. Proteins were electrotransferred to PVDF membranes (Amersham Hybond LFP 0.2 PVDF; GE Healthcare Life Science, Little Chalfont, UK), blocked with 5% BSA (Bio-Rad) and 2% horse serum in TBST for 1 h, then incubated with the following primary antibodies: rabbit anti-

AQP1 (1:150; AB3272-Millipore, Billerica, MA, USA); rabbit anti-AQP5 (1:500; ab92320-Abcam, Cambridge, UK); mouse anti-GAPDH (1:500; MAB374-Millipore) mouse antiβ-Tubulin (at 2.5 μg/mL AB 2315513-DSHB, Iowa City, IA, USA) at 4 °C overnight; and with the secondary antibodies Alexa Fluor 647 Donkey Anti-rabbit IgG and Alexa Fluor 488 Donkey Anti-mouse IgG (both at 1:1000; Life Technologies, Thermo Fisher Scientific) for 1 h at room temperature. The fluorescence signal was detected using Typhoon (Amersham Typhoon Trio; GE Healthcare Life Science). Quantitative analysis was performed using Image Studio Lite (LI-COR, Lincoln, NE, USA) and GAPDH immunoreactivity was used to normalize AQP1 and AQP5 signals.

2.5. Immunostaining

AQP1 and AQP5 localization and expression patterns were determined in cataract HLECs. Samples were processed as whole-mounts and fixed in 4% paraformaldehyde at 4 °C overnight. For AQP5 distribution analysis, antibodies against the membrane N-cadherin were also used. HLEC specimens were blocked with 2% donkey serum and incubated with rabbit anti-AQP1 (1:500; Millipore); rabbit anti-AQP5 (1:200; Abcam) or mouse-anti N-cadherin (1:1000) in blocking solution, overnight at 4 °C and then for 1 h with Alexa Fluor 555 Donkey Anti-mouse IgG 555; Donkey Anti-rabbit IgG or Alexa Fluor 488 Donkey Anti-rabbit IgG (1:200; all from Life Technologies, Thermo Fisher Scientific). Nuclei were stained with Hoechst stain (1:5000; Sigma-Aldrich).

2.6. Confocal analysis

Stained HLECs were examined under a laser scanning confocal microscope (LSM 510 Meta confocal microscope Zeiss, Germany) equipped with a 63× oil objective lens. Images were acquired sequentially to avoid cross-talk using excitation wavelengths 405, 488 and 461 for Hoechst, Alexa 488 and Alexa 555 respectively. The images were acquired in the middle plane of the epithelial monolayer. Distribution analysis of AQP5 with N-cadherin was determined using LSM software (Zeiss) and Image J plug-in JACoP using Mander's M2 spatial colocalization coefficient. M2 represents the ratio of sum intensities from the red channel (N-cadherin) pixels for which the intensity in the green channel (AQP5) is above threshold to the total intensity in the red channel. Threshold values were calculated automatically by JacoP plug-in using Costes´ approach (open source, [30]). A total of 6 images per condition and 2 regions of interest per image were used to determine red and green fluorescence intensity values per distance (μm). Different color channels were overlaid using Adobe Photoshop software (version CS5, Adobe Inc., San Jose, CA). Intensity was graphed versus distance for both colors using Excel. Areas of overlap indicate colocalization.

2.7. Statistical analysis

All statistical analyses were performed using SPSS software (IBM, NY, USA). The normal data distribution was verified using the Kolmogorov–Smirnov test. Parametric (Student's *t*-test and ANOVA) or non-parametric analyses (Mann-Whitney *U* test) were applied accordingly. Comparisons between controls and different types of cataracts were performed using Post Hoc or Kruskal–Wallis analyses. Correlation analyses were performed using the Pearson or Spearman correlation coefficients.

3. Results

3.1. AQP0, AQP1, and AQP5 mRNA expression

AQP0, AQP1, and AQP5 expression levels were analyzed by qPCR in human lens epithelial cells (HLECs), since these are the main AQPs expressed in the lens. As expected, no AQP0 expression was detected in any of the samples analyzed. In control samples, AQP1 was the main AQP expressed, representing 64.1 ± 4.2% of AQPs, with AQP5 representing 35.9 ± 1.9%. An overall similar proportion of AQPs was observed in cataract samples, with a slight increase in the contribution of AQP1 (Fig. 1A). Likewise, no significant differences were observed in the mRNA expression levels of either AQP1 or AQP5 between the groups (Fig. 1B). We also found no differences in AQP mRNA levels between controls and cortical, nuclear, or posterior subcapsular cataracts.

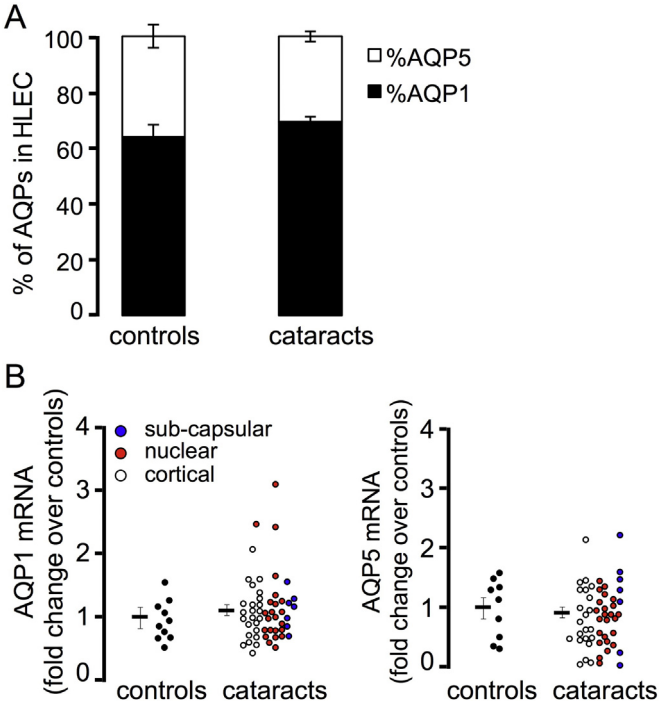


Fig. 1. Aquaporin mRNA expression in HLECs. A) Percentage of the contribution of AQP1 and AQP5 to total AQPs expressed in the HLECs. B) Summary of AQP1 (left) and AQP5 mRNA expression levels for individual HLEC samples normalized to the geometric mean of three house-keeping genes, expressed as the fold-change over controls. Black circles represent individual control samples; white circles represent individual samples from cortical cataracts; red circles represent individual samples from nuclear cataracts; and blue circles represent individual samples from posterior subcapsular cataracts. Solid lines are mean ± S.E.

3.2. AQP1 and AQP5 protein distribution

We then analyzed the localization and distribution of AQP1 and AQP5 proteins in HLECs from 7 controls patients and 7 nuclear cataract patients using whole mount immunofluorescence, and found a similar AQP expression pattern to that reported previously for controls [12,31]. AQP1 protein was expressed in both the apical and basolateral membranes of the anterior HLEC (Fig. 2, left), in both controls and cataracts. However, a differential AQP5 distribution pattern was found in samples from nuclear cataracts and controls (Fig. 2, right). As described previously, AQP5 shows a predominantly cytoplasmic labeling in the differentiating fiber cells in the cortex, suggesting that AQP5 is stored in intracellular vesicles or in organelle membranes, and its distribution shifts to a membrane labeling in the core lens, both in mouse, rat and human lens [24,32]. However in HLEC from cataract patients, we have observed increased membrane localization, compared to controls (Fig. 3A–B). Colocalization analysis of AQP5 with N-cadherin showed a significantly higher association of AQP5 with membranal N-cadherin, measured as the fraction of AQP5 labeling overlapping N-cadherin, with Manders'coefficient (M2) = 0,261 ± 0,02 for controls and 0,353 ± 0,01 for cataract HLEC ($p = 0{,}022$, Fig. 3C).

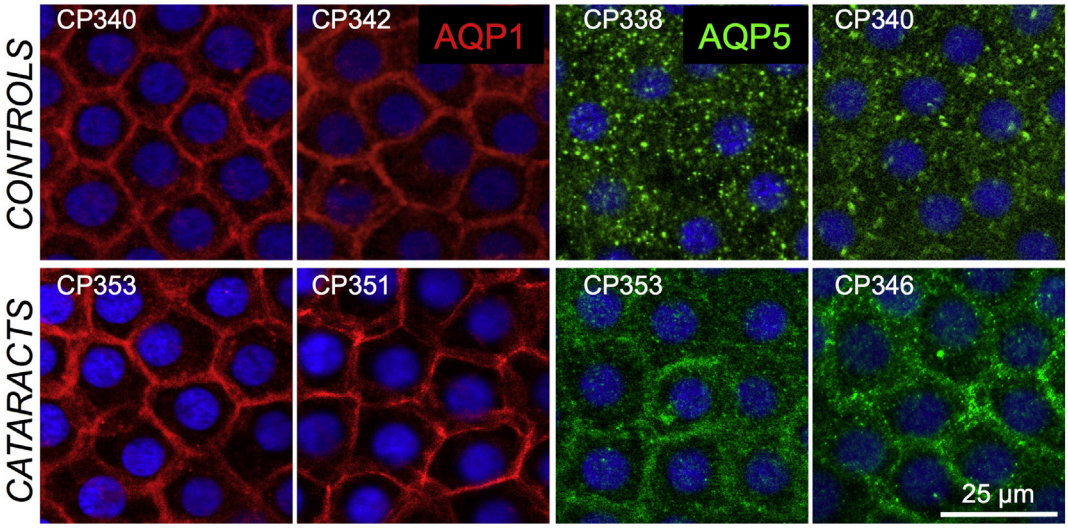


Fig. 2. Aquaporin localization in HLECs. Representative whole mount immunofluorescence images of AQP1 protein expression (red) and AQP5 (green) in HLECs from two cataract and control patients.

3.3. AQP1 and AQP5 protein quantification

Immunoblot analysis with either anti-AQP1 or -AQP5 antibodies showed a band at ~28 kDa, corresponding to non-glycosylated AQP. AQP1 immunoblots showed an additional diffuse band at ~34 kDa, corresponding to the glycosylated protein. No glycosylated AQP5 band was observed in the lens epithelial cells from either cataracts or control patients (Fig. 4A and Supplementary Fig. 1). When normalized to GAPDH expression, we found that AQP1 protein levels were significantly increased in the HLEC from cataract patients (1.65 ± 0–27-fold increase, $p = 0$–001), with 28% of the samples (12/43) showing a higher than twofold increase in AQP1 expression levels compared to controls. To verify proper use of GAPDH as an endogenous control in these samples, we also analyzed levels of β-tubulin, another commonly used loading control, and found that protein levels of GAPDH and β-Tubulin were comparable, with significant correlation in our samples ($p < 0.03$; $R_S = 0.353$). A subgroup analysis revealed that differences were highest among the nuclear cataracts (2.1 ± 0.27-fold increase, $p = 0.0008$), with 8 out of 20 nuclear cataracts (40%) showing a higher than twofold increase in AQP1 expression levels compared to controls. No significant differences in AQP1 expression levels were observed between cortical cataracts and control samples (Fig. 4B).

With respect to AQP5 protein expression, although no significant differences were observed, we found a similar increased pattern of AQP5 expression in cataract compared to control samples (1.47 ± 0.2-fold increase), which was more pronounced among nuclear cataracts (1.8 ± 0.39, $p = 0.068$; Fig. 4B).

3.4. AQP1 and AQP5 protein expression levels are not associated with age or sex

Since the incidence of cataract is known to increase with age, we tested whether there was a correlation between AQP1 and AQP5 expression levels and age in samples from both controls and patients, using Pearson's correlation coefficient formula. We found no correlation of AQP1 or AQP5 expression levels with age either among controls ($r = 0.195$) or cataract ($r = -0.224$). Likewise, no statistical association was observed between AQP1 or AQP5 expression and sex (Supplementary Figs. 2 and 3).

4. Discussion

In this study we have evaluated AQP expression and localization in human lens epithelial cells (HLECs). AQP1 is the major AQP in HLECs, representing around 64% of total AQP mRNA. This protein is expressed in both the apical and basolateral membranes of the anterior epithelial cells, whereas AQP5 is expressed mainly localized to the cytoplasm, representing around 36% of total AQP mRNAs. Interestingly, we found a significantly higher protein expression for AQP1 in samples from cataract patients, with this expression being highest (>twofold increase) in the sub-group of nuclear cataracts. Since mRNA levels showed no significant differences with respect to controls, the AQP1 upregulation observed in cataract patients is likely to be regulated at the post-translational level, either by increasing protein synthesis and/or by inhibiting protein degradation.

In this regard, previous reports have shown alterations in ubiquitination and stability of AQP1 in hypertonic stress in fibroblast from BALB/c mice. In particular, they showed that after exposure to hypertonic medium, there was a selective decrease in AQP1 ubiquitination, together with a marked increase in AQP1 protein stability. It has been proposed that reduction in ubiquitination and increase in protein stability under these conditions acts to facilitate protein induction at a time when the general pressure on the cell is to reduce protein synthesis [33,34]. With respect to AQP5, previous studies on hypertonic induction of this aquaporin on mouse lung epithelial cells suggest that both expression and degradation of this water channel is tightly controlled, since AQP5 mRNA and protein expression returned nearly to baseline levels within hours, after incubating the cells with hypertonic medium and then returned to isotonic conditions [35]. Therefore, it is possible that similar mechanisms leading to protein stabilization are taking place in HLECs from cataract patients.

It is not clear why the HLECs obtained from patients with nuclear cataracts expressed higher AQP1 protein levels than cortical cataracts, especially if we consider that the lens cortex is in contact with HLECs, which is where AQP1 is expressed in the lens. One explanation could be related to the location of nuclear cataracts in the inner layers of the lens. Fiber cells in this region have a more limited access to nutrients/metabolites via simple diffusion when compared with cortical fiber cells and are therefore more dependent on mechanisms implicating a driven ion-coupled fluid exchange, with AQP1 in HLECs possibly playing an active role. Another possibility could be age-related, since the

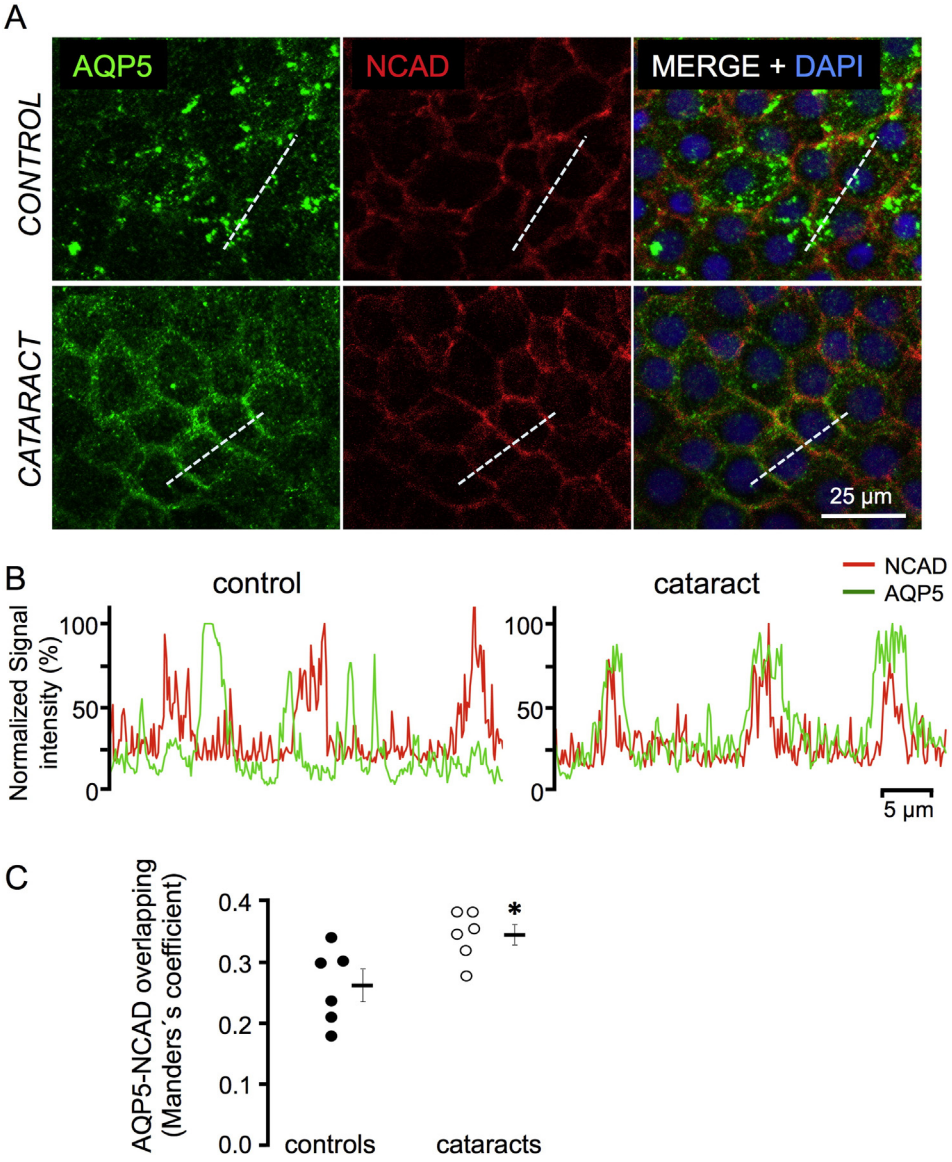


Fig. 3. Colocalization of AQP5 and NCAD in HLECs. A) Immunofluorescece imaging of control (top panel) and cataract HLECs (bottom panel) shows colocalization differences of AQP5 (green) with relation to the membrane protein NCAD (red). B) Red and green fluorescence intensity values per distance (μm) from the same section (dotted white lines in A). C) Significantly higher colocalization of AQP5 with NCAD was observed in HLECs from nuclear cataracts (Mander's coefficient $p < 0.05$). Black circles represent individual control samples and white circles represent individual samples from nuclear cataracts. Solid lines are mean ± S.E.

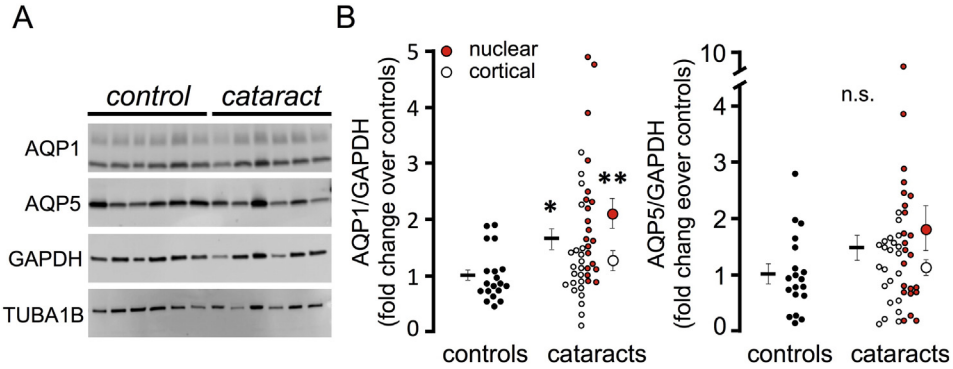


Fig. 4. Aquaporin protein levels in HLECs. A) Immunoblot analysis of AQP1 and AQP5 in HLEC homogenates from controls and cataract patients. GAPDH and TUBA1B were used as loading control. B) Quantification of AQP1 (left) and AQP5 (right) protein expression levels for individual HLEC samples normalized to GAPDH levels. Data are expressed as the fold-change over controls. Black circles represent individual control samples; white circles represent individual samples from cortical cataracts; and red circles represent individual samples from nuclear cataracts. Solid lines are mean ± S.E. Large white and red circles are mean ± S.E. for cortical and nuclear cataracts, respectively. *$p < 0.005$; **$p < 0.001$ (Student's *t*-test). N.S.: not significant.

patients from whom nuclear cataracts were extracted were about 7 years younger, on average, than the group of patients with cortical cataracts. This could explain a more active compensatory mechanism of AQP1 upregulation compared with elderly patients. However, further experiments will be required to test these possibilities.

One of the limitations of the present study was to match the age of control samples obtained from refractive surgery, which is typically performed in the sixth decade, with that for cataract samples, which is typically performed in the eighth decade. We were able to reduce age differences to about 10 years by selecting the *oldest* controls and *youngest* cataract patients among those recruited. Despite this limitation, we found no correlation between age and the levels of AQP1/5 expression in either the control or cataract groups, and there was no association with sex. Furthermore, the twofold higher AQP1 expression observed in the sub-group of nuclear cataracts corresponded to the group of patients with an average age of 66.5 years, compared with the group of cortical cataract patients, who had an average age of 73.7 years, for whom no significant differences in AQP1 expression levels were found. Indeed, the age gap between samples from the control and nuclear cataract groups was only 7.6 years. Altogether, in our cohort of individuals, AQP1 and AQP5 expression is independent of age.

Although the exact mechanisms linking AQP1 with cataract formation are yet to be fully determined, our study supports an involvement of AQP1, and to a lesser extent of AQP5, in the maintenance of lens transparency. In a previous study, we showed a functional role for AQP1 in the lens epithelium in maintaining lens transparency in a mouse model of cataractogenesis, both in vitro and in vivo [21]. Our observations supported an outward fluid flux at the anterior surface and/or the lens equator of AQP1-KO mice, since they revealed a concomitant approximately threefold reduction in anterior surface water permeability and increased basal water content [21]. This closely resembled the AQP1-dependent fluid transport reported in the cornea of AQP1-null mice [22], with a markedly reduced ability to expel excess fluid under stress conditions at the corneal endothelium level. This study proposed that an up-regulation of AQP1 in corneal endothelium could be particularly useful in reducing corneal edema and improving transparency under stress conditions. Following a similar rationale, the twofold increase in AQP1 in the lens epithelium from cataract patients observed in the present study could be a compensatory mechanism in an attempt to protect the lens from opacification by increasing water permeability in patients with cataract.

With respect to AQP5, intracellular labeling for this AQP has been described in rodent and in the human lens epithelial cells [24,32]. AQP5 does not seem to contribute significantly to the water permeability of lens epithelial cells in normal conditions, based on its cytosolic localization and its relatively lower expression levels compared with AQP1, according to this study and previously reported data [24]. Therefore AQP1 seems to be the main functional water channel in the lens epithelium, whereas AQP5 is more likely to be complementing the water permeability needs of the fiber cells along with AQP0 in the inner lens. AQP5 is predominately intracellular in the lens in normal conditions and is presumably located close to vesicular storage pools in HLECs and differentiating fiber cells of the outer cortex, while it is associated with the plasma membrane in terminally differentiated mature fiber cells in the inner lens [12] [32]. It has been proposed that this would change the membrane properties of cells in the lens core, which lack a protein synthesis machinery and are therefore unable to perform de novo protein synthesis [12,36].

The significant increased cell membrane expression of AQP5 we observed in HLEC from nuclear cataract might be contributing to retard lens opacification by increasing AQP5 trafficking to the cell membrane, which would fulfill the water permeability function in the HLECs of patients with initial stages of cataract. In support of this hypothesis, AQP5 has been shown to traffic to the apical cell membrane of salivary gland cells in response to a number of stimuli to increase saliva production [37]. Unlike AQP1, which is constitutively expressed in the plasma membrane, AQP5 is closely related to AQP2, which has a well-characterized mechanism of subcellular translocation in response to anti-diuretic hormone stimulation to increase water absorption in the renal collecting ducts [38]. With respect to the regulation of AQP5 localization, several investigators have demonstrated that AQP5 expression; intracellular trafficking, and localization in the epithelial cells are regulated by cAMP via the protein kinase A (PKA) pathway and osmotic triggers [39–44]. As such, dephosphorylation of AQP5 stored in the cytoplasmic pools using small molecules (e.g. stimulating protein kinase A via cAMP), would increase the availability of AQP5 water channels in the plasma membrane of lens epithelial cells and therefore water permeability. In this regard, due to its importance for fluid secretion in airway submucosal glands, AQP5 has been suggested to be a pharmacological target for treatment of the hyper-viscous and excessive gland secretions in cystic fibrosis and bronchitis/rhinitis, respectively [42].

In summary, we have observed an increased membrane expression of AQP1 and AQP5 in the lens epithelial cells from cataract patients, which was especially pronounced in patients with nuclear cataracts. Since the proportion of free to protein-bound water in the lens has been shown to increase with age [45–48] and further with cataract [49], it is likely that increasing AQP1 and/or AQP5 protein levels in the membrane of HLECs may be a compensatory mechanism to expel excess water by increasing water permeability in the cataract. Thus, further AQP1 upregulation and/or increased AQP5 membrane trafficking, either pharmacologically or by gene delivery, may help retard lens opacification and could be a potential therapeutic strategy for early cataracts. This might be achieved by increasing AQP1 expression, once AQP1 modulators became available or by increasing the translocation of AQP5 molecules stored in cytosolic reservoirs to the cell membrane, as has recently been proposed for the treatment of several human diseases, such as Sjögren's syndrome, bronchitis, and cystic fibrosis [42,50,51]. Further research on these possibilities may help develop novel therapies to treat cataracts.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbadis.2016.08.001.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgements

The authors thank Cristina Sarasqueta for assistance with SPSS statistical analysis and BIOEF for reviewing the English. The beta Tubulin monoclonal antibody, developed by Michael Klimkowsky, from University of Colorado, Boulder, CO, was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, USA (IA 52242).

JR-E is a Miguel Servet Fellow, National Institute of Health Carlos III (Instituto de Salud Carlos III/ISCIII). MEI was supported by grants from the Department of Education of the Basque Government (DEDUC14/309). OB was supported by funds from National Institute of Health Carlos III (CP10/00572). AA was supported by grants from Fundación Jesús de Gangoiti Barrera and from the Basque Government (SAIOTEK-PE11BN002/PC12BN001/DEPLC13/002). AV-I is a Ramón y Cajal Fellow. This work was supported by grants from the National Institute of Health Carlos III (Instituto de Salud Carlos III/ISCIII) (CP10/00572), from the Department of Industry, Basque Government (SAIOTEK: SAIO11-PE11BN002; and SAIO12-PC12BN001) and from the Foundation of Patients of Retinitis Pigmentosa of Gipuzkoa (BEGISARE) to JR-E.

References

[1] J.M. Petrash, Aging and age-related diseases of the ocular lens and vitreous body, Invest. Ophthalmol. Vis. Sci. 54 (2013) 54–59 (ORSF).
[2] G.N. Rao, R. Khanna, A. Payal, The global burden of cataract, Curr. Opin. Ophthalmol. 22 (2011) 4–9.
[3] Organisation for Economic Co-operation and Development. Health Division., European Commission., World Health Organization., Health at a glance. Europe, in, OECD Publishing, Paris, pp. v. ill.
[4] T. Kuwabara, The maturation of the lens cell: a morphologic study, Exp. Eye Res. 20 (1975) 427–443.
[5] G.A. Zampighi, S. Eskandari, M. Kreman, Epithelial organization of the mammalian lens, Exp. Eye Res. 71 (2000) 415–435.
[6] S. Bassnett, Lens organelle degradation, Exp. Eye Res. 74 (2002) 1–6.
[7] J. Fischbarg, F.P. Diecke, K. Kuang, B. Yu, F. Kang, P. Iserovich, Y. Li, H. Rosskothen, J.P. Koniarek, Transport of fluid by lens epithelium, Am. J. Phys. 276 (1999) C548–C557.
[8] O.A. Candia, Electrolyte and fluid transport across corneal, conjunctival and lens epithelia, Exp. Eye Res. 78 (2004) 527–535.
[9] R.T. Mathias, J. Kistler, P. Donaldson, The lens circulation, J. Membr. Biol. 216 (2007) 1–16.
[10] R.T. Mathias, J.L. Rae, G.J. Baldo, Physiological properties of the normal lens, Physiol. Rev. 77 (1997) 21–50.
[11] U.P. Andley, Crystallins in the eye: function and pathology, Prog. Retin. Eye Res. 26 (2007) 78–98.
[12] A.C. Grey, K.L. Walker, R.S. Petrova, J. Han, P.A. Wilmarth, L.L. David, P.J. Donaldson, K.L. Schey, Verification and spatial localization of aquaporin-5 in the ocular lens, Exp. Eye Res. 108 (2013) 94–102.
[13] V. Berry, P. Francis, S. Kaushal, A. Moore, S. Bhattacharya, Missense mutations in MIP underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q, Nat. Genet. 25 (2000) 15–17.
[14] A. Shiels, S. Bassnett, K. Varadaraj, R. Mathias, K. Al-Ghoul, J. Kuszak, D. Donoviel, S. Lilleberg, G. Friedrich, B. Zambrowicz, Optical dysfunction of the crystalline lens in aquaporin-0-deficient mice, Physiol. Genomics 7 (2001) 179–186.
[15] P. Francis, V. Berry, S. Bhattacharya, A. Moore, Congenital progressive polymorphic cataract caused by a mutation in the major intrinsic protein of the lens, MIP (AQP0), Br. J. Ophthalmol. 84 (2000) 1376–1379.
[16] D.D. Geyer, M.A. Spence, M. Johannes, P. Flodman, K.P. Clancy, R. Berry, R.S. Sparkes, M.D. Jonsen, S.J. Isenberg, J.B. Bateman, Novel single-base deletional mutation in major intrinsic protein (MIP) in autosomal dominant cataract, Am J. Ophthalmol. 141 (2006) 761–763.
[17] K.L. Nemeth-Cahalan, J.E. Hall, pH and calcium regulate the water permeability of aquaporin 0, J. Biol. Chem. 275 (2000) 6777–6782.
[18] D. Fotiadis, L. Hasler, D.J. Muller, H. Stahlberg, J. Kistler, A. Engel, Surface tongue-and-groove contours on lens MIP facilitate cell-to-cell adherence, J. Mol. Biol. 300 (2000) 779–789.
[19] G.M. Preston, B.L. Smith, M.L. Zeidel, J.J. Moulds, P. Agre, Mutations in aquaporin-1 in phenotypically normal humans without functional CHIP water channels, Science 265 (1994) 1585–1587.
[20] J. Ruiz-Ederra, A.S. Verkman, Mouse model of sustained elevation in intraocular pressure produced by episcleral vein occlusion, Exp. Eye Res. 82 (2006) 879–884.
[21] J. Ruiz-Ederra, A.S. Verkman, Accelerated cataract formation and reduced lens epithelial water permeability in aquaporin-1-deficient mice, Invest. Ophthalmol. Vis. Sci. 47 (2006) 3960–3967.
[22] J.R. Thiagarajah, A.S. Verkman, Aquaporin deletion in mice reduces corneal water permeability and delays restoration of transparency after swelling, J. Biol. Chem. 277 (2002) 19139–19144.
[23] J. Ruiz-Ederra, M.H. Levin, A.S. Verkman, In situ fluorescence measurement of tear film [Na+], [K+], [Cl-], and pH in mice shows marked hypertonicity in aquaporin-5 deficiency, Invest. Ophthalmol. Vis. Sci. 50 (2009) 2132–2138.
[24] S. Sindhu Kumari, K. Varadaraj, Aquaporin 5 knockout mouse lens develops hyperglycemic cataract, Biochem. Biophys. Res. Commun. 441 (2013) 333–338.
[25] D.C. Blaydon, L.K. Lind, V. Plagnol, K.J. Linton, F.J. Smith, N.J. Wilson, W.H. McLean, C.S. Munro, A.P. South, I.M. Leigh, E.A. O'Toole, A. Lundstrom, D.P. Kelsell, Mutations in AQP5, encoding a water-channel protein, cause autosomal-dominant diffuse nonepidermolytic palmoplantar keratoderma, Am. J. Hum. Genet. 93 (2013) 330–335.
[26] X. Cao, J. Yin, H. Wang, J. Zhao, J. Zhang, L. Dai, J. Zhang, H. Jiang, Z. Lin, Y. Yang, Mutation in AQP5, encoding aquaporin 5, causes palmoplantar keratoderma Bothnia type, J. Invest. Dermatol. 134 (2014) 284–287.
[27] L.T. Chylack Jr., J.K. Wolfe, D.M. Singer, M.C. Leske, M.A. Bullimore, I.L. Bailey, J. Friend, D. McCarthy, S.Y. Wu, The lens opacities classification system III. The longitudinal study of cataract study group, Arch. Ophthalmol. 111 (1993) 831–836.
[28] A. Vallejo-Illarramendi, M. Domercq, F. Perez-Cerda, R. Ravid, C. Matute, Increased expression and function of glutamate transporters in multiple sclerosis, Neurobiol. Dis. 21 (2006) 154–164.
[29] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002) (RESEARCH0034).
[30] S. Bolte, F.P. Cordelieres, A guided tour into subcellular colocalization analysis in light microscopy, J. Microsc. 224 (2006) 213–232.
[31] S. Hamann, T. Zeuthen, M. La Cour, E.A. Nagelhus, O.P. Ottersen, P. Agre, S. Nielsen, Aquaporins in complex tissues: distribution of aquaporins 1–5 in human and rat eye, Am. J. Phys. 274 (1998) C1332–C1345.
[32] S.S. Kumari, M. Varadaraj, V.S. Yerramilli, A.G. Menon, K. Varadaraj, Spatial expression of aquaporin 5 in mammalian cornea and lens, and regulation of its localization by phosphokinase A, Mol. Vis. 18 (2012) 957–967.
[33] V. Leitch, P. Agre, L.S. King, Altered ubiquitination and stability of aquaporin-1 in hypertonic stress, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 2894–2898.
[34] D.M. Cohen, J.C. Wasserman, S.R. Gullans, Immediate early gene and HSP70 expression in hyperosmotic stress in MDCK cells, Am. J. Phys. 261 (1991) C594–C601.
[35] J.D. Hoffert, V. Leitch, P. Agre, L.S. King, Hypertonic induction of aquaporin-5 expression through an ERK-dependent pathway, J. Biol. Chem. 275 (2000) 9070–9077.
[36] J. Tombran-Tink, C.J. Barnstable, Ocular Transporters in Ophthalmic Diseases and Drug Delivery, in: Ophthalmology Research, Humana Press, Totowa, N.J., 2008
[37] Y. Ishikawa, G. Cho, Z. Yuan, N. Inoue, Y. Nakae, Aquaporin-5 water channel in lipid rafts of rat parotid glands, Biochim. Biophys. Acta 1758 (2006) 1053–1060.
[38] A.S. Verkman, Dissecting the roles of aquaporins in renal pathophysiology using transgenic mice, Semin. Nephrol. 28 (2008) 217–226.
[39] A. Eckhard, A. Dos Santos, W. Liu, M. Bassiouni, H. Arnold, C. Gleiser, B. Hirt, C. Harteneck, M. Muller, H. Rask-Andersen, H. Lowenheim, Regulation of the perilymphatic-endolymphatic water shunt in the cochlea by membrane translocation of aquaporin-5, Pflugers Arch. - Eur. J. Physiol. 467 (2015) 2571–2588.
[40] T. Hasegawa, A. Azlina, P. Javkhlan, C. Yao, T. Akamatsu, K. Hosoi, Novel phosphorylation of aquaporin-5 at its threonine 259 through cAMP signaling in salivary gland cells, Am. J. Physiol. Cell Physiol. 301 (2011) C667–C678.
[41] M. Hollborn, S. Vogler, A. Reichenbach, P. Wiedemann, A. Bringmann, L. Kohen, Regulation of the hyperosmotic induction of aquaporin 5 and VEGF in retinal pigment epithelial cells: involvement of NFAT5, Mol. Vis. 21 (2015) 360–377.
[42] P. Kitchen, F. Oberg, J. Sjohamn, K. Hedfalk, R.M. Bill, A.C. Conner, M.T. Conner, S. Tornroth-Horsefield, Plasma membrane abundance of human aquaporin 5 is dynamically regulated by multiple pathways, PLoS One 10 (2015), e0143027.
[43] V. Sidhaye, J.D. Hoffert, L.S. King, cAMP has distinct acute and chronic effects on aquaporin-5 in lung epithelial cells, J. Biol. Chem. 280 (2005) 3590–3596.
[44] F. Yang, J.D. Kawedia, A.G. Menon, Cyclic AMP regulates aquaporin 5 expression at both transcriptional and post-transcriptional levels through a protein kinase A pathway, J. Biol. Chem. 278 (2003) 32173–32180.
[45] B.K. Pierscionek, J.W. Regini, The gradient index lens of the eye: an opto-biological synchrony, Prog. Retin. Eye Res. 31 (2012) 332–349.
[46] F.A. Bettelheim, M.J. Lizak, J.S. Zigler Jr., Relaxographic studies of aging normal human lenses, Exp. Eye Res. 75 (2002) 695–702.
[47] F.A. Bettelheim, M.J. Lizak, J.S. Zigler Jr., Syneretic response of aging normal human lens to pressure, Invest. Ophthalmol. Vis. Sci. 44 (2003) 258–263.
[48] D. Lahm, L.K. Lee, F.A. Bettelheim, Age dependence of freezable and nonfreezable water content of normal human lenses, Invest. Ophthalmol. Vis. Sci. 26 (1985) 1162–1165.
[49] P. Racz, C. Hargitai, B. Alfoldy, P. Banki, K. Tompa, 1H spin-spin relaxation in normal and cataractous human, normal fish and bird eye lenses, Exp. Eye Res. 70 (2000) 529–536.
[50] Y. Song, A.S. Verkman, Aquaporin-5 dependent fluid secretion in airway submucosal glands, J. Biol. Chem. 276 (2001) 41288–41292.
[51] K. Tsubota, S. Hirai, L.S. King, P. Agre, N. Ishida, Defective cellular trafficking of lacrimal gland aquaporin-5 in Sjogren's syndrome, Lancet 357 (2001) 688–689.