The causes of retinal dystrophy and the development of more comprehensive screening approach

Samar Yahya

MBBS, MSc Medical Genetics

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Medicine

February 2023

Jointly authored publications statement

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 4 of this thesis has entirely appeared in publication as follows:

YAHYA, S., SMITH, C. E. L., POULTER, J. A., MCKIBBIN, M., ARNO, G., ELLINGFORD, J., KÄMPJÄRVI, K., KHAN, M. I., CREMERS, F. P. M., HARDCASTLE, A. J., CASTLE, B., STEEL, D. H. W., WEBSTER, A. R., BLACK, G. C., EL-ASRAG, M. E., ALI, M., TOOMES, C., INGLEHEARN, C. for the UK inherited Retinal Dystrophy Consortium and the Genomics England Research Consortium. 2023. Late-Onset Autosomal Dominant Macular Degeneration Caused by Deletion of the CRX Gene. *Ophthalmology*, 130, 68-76.

Dr Smith initiated the work in this project by identifying the deletion in three samples. The author confirmed the CRX deletion in the additional cases using the PCR assay developed by Dr Smith. The author also performed a TA cloning and Sanger sequencing to characterize the deletion breakpoints, haplotype analysis by both microsatellite genotyping and SNPs analysis, and co-wrote the manuscript. The other authors provided additional cases and their clinical details, supervise, and co-wrote the paper.

The work in Chapter 5 of this thesis is partially appeared in a paper submitted in Journal Molecular Diagnosis and Therapy:

Yahya S, Watson CM, McKibbin M, Crinnion L, Bonin H, Fletcher T, El-Asrag ME, Ali M, Toomes C, Inglehearn CF. Long read nanopore sequencing of *RPGR-ORF15* is enhanced following DNase I treatment of MinION flow cells. Submitted (Under review).

Publication in preparation from Chapter 5 in this thesis:

Yahya S, Carr I, Watson CM, Chang L, Raynor M, McKibbin M, Mukherjee R, El-Asrag ME, Ali M, Toomes C, Inglehearn CF. Screening a UK IRD cohort for *RPGR-ORF15* variants by high fidelity PCR and PacBio long read sequencing. In preparation.

Acknowledgment

All gratitude belongs to the All-Mighty Allah, the Most Gracious and the Most Merciful, who granted me this opportunity, guided me, and gave me the strength to do this project effectively. I would like to express my gratitude to Prof Chris Inglehearn, who accepted me as a PhD student at the university of Leeds and gave me the honour of being one of his students, for his professional guidance, constructive feedback, endless support especially when I was expecting, and continuous motivation. My sincere appreciation goes to Dr Carmel Toomes for all of her assistance with lab experiments throughout the years, as well as for training me, refining me, and answering all of my questions.; her care has helped me get through the most stressful moments. I also like to thank Dr Manir Ali, from whom I learnt a lot, for his encouragement, guidance, and assistance with experiments and writing. Many thanks also to Dr Mohammed Elasrag for his assistance, valuable advice, and knowledge-sharing. The contribution of Dr Christopher Watson and Dr Ian Carr to this project cannot be underestimated since they assisted me with the long read sequencing and bioinformatics analysis while also training me. I would also like to thank Mrs Laura Crinnion and Mrs Morag Raynor for helping me with the experiments and answering my queries. Thank you to Dr. James Poulter, Dr. Claire Smith, Ummey Hany, Benjamin McClinton, Mohammed Derar, Dong Sun, Dr Georgios Nikolopoulos, and Ewa Jaworska from the department of Ophthalmology and Neuroscience at the Leeds Institute of Biomedical and Clinical Science (LIBAC). My dear friend Tomader, thanks for being always there for me.

I would like to acknowledge the financial support that I received from King Abdul-Aziz University and the Royal Embassy of Saudi Arabia Cultural Bureau.

I would like to thank my parents for their spiritual support. This thesis belongs to my beautiful family, especially to my wonderful husband Sultan. Dear Sultan, you have always placed my needs ahead of your own, and I will never forget the day you left everything and travelled with me to Aberdeen to achieve my goals. You have always been there for me, helping me with the kids, supporting me and believing in me even when I doubt myself. Without your care and encouragement, this thesis would not be possible. My darling kids, Nora, and Saif, you will always be the greatest achievement in my life, you have made my life momentous and added joy and brightness to it. I apologise for being absent and busy, but one day you'll realise that everything I do is for you because I love you.

Abstract

Inherited retinal diseases (IRDs) are a group of genetically and phenotypically heterogenous disorders caused by variants in around 280 genes. Additional loci have also been localised to chromosomal regions, though the causative genes remain unknown. Recent improvements in screening technologies have increased the detection of pathogenic variants in IRD. This thesis describes the use of next generation sequencing (second (short-read) and third (long-read) generation sequencing) to find missing or hard to find pathogenic variants in IRD patients.

The first results chapter describes use of whole exome sequencing to screen 24 individuals with syndromic and non-syndromic IRDs. This identified pathogenic variants in known genes in eight cases; *CDHR1* (c.1527T>G, p.Y509*), *RHO* (c.284T>C, p.L95P), *PRPF31* (c.797delC, p.S266*), *CNGA3* (c.1088T>C, p.L363P), *BBS10* (c.728-731delAAGA, p. K243Ifs*15), *USH2A* (c.252T>G, p.C84W), *ABCA4* (c.2588G>C, p.G863A and c.6089G>A, p.R2030Q), and *SLC25A46* (c.670A>G, p.T224A). In addition, several candidate variants were highlighted for further investigation.

In the second results chapter, seven patients with late onset macular dystrophy and one with age related macular degeneration were found to carry the same heterozygous ~126 kb deletion encompassing *CRX*, *TPRX1* and *SULT2A1*. This phenotype has already been documented in patients with heterozygous variants in the gene encoding retinal transcription factor CRX, while there is no known functional or phenotypic link with variants in *TPRX1* or *SULT2A1*. This therefore confirms that *CRX* haploinsufficiency is pathogenic, a finding that had previously been debated in the ophthalmic literature. The deletion was characterized using a PCR assay followed by cloning and Sanger sequencing or direct Sanger sequencing. Haplotype analysis was done by microsatellite genotyping.

The third results chapter describes use of SMRT PacBio and nanopore long-read sequencing to screen the hard-to-sequence mutation hotspot *RPGR-ORF15*. Both approaches were effective in reading throughout *ORF15* and allowed sequencing indexed pooled samples, and 218 IRD patients were screened, detecting known and new variants. Nanopore sequencing on the smaller Flongle flowcell allowed low-cost optimisation, but pores rapidly blocked, probably due to ORF15 secondary structures. Repeated DNase I washes reopened the pores but required use of the more expensive MinION flowcells. Ultimately, the PacBio sequencer proved simpler to use, cheaper, and more scalable.

Table of contents

List of figures

List of tables

List of abbreviation

11cRAL	11-cis-retinal
A1	11-cis-retinal
AAVs	Adeno-associated viruses
ACHM	Achromatopsia
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
ADD	Autosomal dominant drusen
ADOA	Autosomal dominant optic atrophy
ADOA+	Autosomal dominant optic atrophy plus
ADRP	Autosomal dominant retinitis pigmentosa
AF	Autofluorescence imaging
AIDS	Acquired immune deficiency syndrome
ALMS	Alstrom syndrome
AMD	Age Related Macular Degeneration
AONs	Antisense oligonucleotides
AR-CORD	Autosomal recessive cone rod dystrophy
AR-MD	Autosomal recessive macular dystrophy
ARMD	Age related macular dystrophy
ARRP	Autosomal recessive retinitis pigmentosa
ATP	Adenosine triphosphate
atRAL	All-trans-retinal
AutoMap	Autozygousity mapper
AVMD	Adult vitelliform macular dystrophy
BAM	Binary alignment and map
BBS	Bardet-Biedl syndrome
BCAMD	Benign concentric annular macular dystrophy
BD	Best disease
BEHRS	Behr syndrome
BRB	Blood Retina Barrier
BSA	Bovine serum albumin

BWA	Burrows-Wheeler Aligner
Ca+	Calcium ions
CADD	Combined annotation dependent depletion
CATCH	CAs9 Targeting of CHromosomal segment
CC	Connecting cilium
CCS	Circular consensus sequences
CGH	Comparative genomic hybridisation
cGMP	Cyclic guanosine monophosphate
CLR	Continuous long read
CMO	Cystoid macular oedema
CMT	Charcot-Marie-Tooth disease
CNGCs	Cyclic Nucleotide Gated Channels
CNVs	Copy number variations
COD	Cone dystrophy
CORD3	Cone rod dystrophy 3
CRISPR	Clustered regularly interspaced short palindromic repeats
CSC	Central serous chorioretinopathy
dATP	Deoxyadenosine triphosphate
dbSNP	Single nucleotide polymorphism database
ddNTPs	Dideoxy nucleotides
dH_2O	Distelled water
DHEAST	Dehydroepiandrosterone sulfotransferase
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribose nucleoside triphosphate
DSBs	Double stranded DNA breaks
dsDNA	Double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
eQTLs	Expression quantitative trait loci
ERG	Electroretinogram
ESR	Estrogen receptor mediated signalling

ExoSAP-IT	Exonuclease I and Shrimp Alkaline Phosphatase
FAF	Fundus autofluorescence
FAM	Fluorescein amidites
FATHMM	Functional analysis through Hidden Markov Models
FB	Flush buffer
FDA	Food and Drug Administration
FEVR	Familial exudative vitreoretinopathy
FFERG	Full field electroretinogram
FFM	Fundus flavimaculatus
FoSTes	Fork stalling and template switching
FT	Flush tether
G	Transducin
g.VCF	Genomic variant call format
G*	Activated Transducin
GA	Geographical atrophy
GATK	Genome analysis tool kit
GC	Guanylate Cyclase
GCL	Ganglion Cell Layer
gDNA	Genomic DNA
GDP	Guanosine diphosphate
gnomAD	Genome aggregation database
GPCRs	G-Protein Coupled Receptors
GRK	G-coupled receptor kinase
GTP	Guanosine triphosphate
Gα-GDP	inactive transducin
Gα-GTP	Activated Transducin
Gβ5	Complex of G protein beta 5
HDR	Homology-directed repair
hESCs	human embryonic stem cells
HGVS	Human Genome Variation Society
Hi-Di	Highly deionized

HM	Hand movement
IEMs	Inborn errors of metabolism
IGV	Integrative genomic viewer
IMS	Intermembrane space
INL	Inner nuclear Layer
IPL	Inner plexiform Layer
ipRGCs	Intrinsically Photosensitive Retinal Ganglion Cells
iPSCs	Induced pluripotent stem cells
IRBP	Interphotoreceptor retinoid binding protein
IRDs	Inherited Retinal Diseases
IS	Inner segment
JBTS	Joubert syndrome
KCl	Potassium chloride
$KHCO_3$	Potassium Bicarbonate
LB	Luria-Bertani agar
LB	Loading beads
LCA	Leber congenital amaurosis
LOD	Logarithm of odds
LogMar	Logarithm of the minimum angle of resolution
LOVD	Leiden open variation database
LRAT	Lecithin retinol acyl transferase
LVs	Lentiviral vectors
MAF	Minor allele frequency
MCLMR	Microcephaly with or without chorioretinopathy, lymphedema, or mental retardation
MD	Macular dystrophy
MFAT	Multifunctional O-acyltransferase
MfERG	Multifocal electroretinogram
MFSDs	Major facilitator superfamily domain containing proteins family
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulfate

MICOS	Mitochondrial contact site and cristae organising system
MIPs	Molecular inversion probes
miRNAs	Micro RNAs
MKKS	Mckusick-Kaufman syndrome
MMBIR	Microhomology mediated break-induced replication
MTDPS-14	Mitochondrial DNA depletion syndrome type 14
Na+	Sodium ions
NaCl	Sodium chloride
NAD+	Nicotinamide adenine dinucleotide
NAHR	Nonallelic homologous recombination
NaOH	Sodium hydroxide
NCBI	National center for biotechnology information
NDPK	Nucleotide diphosphate kinase like domain
NEB	New England biolabs
NGS	Next generation sequencing
NH_4Cl	Ammonium Chloride
NHEJ	Nonhomologous end-joining
NMD	Nonsense mediated decay
NPs	Nanoparticles
NR2E3	Nuclear Receptor subfamily 2 group E member 3
NRL	Neural Retina-specific Leucine zipper protein
NS-ARRP	Non syndromic autosomal recessive retinitis pigmentosa
nsSNPs	Non-synonymous single nucleotide polymorphisms
NTG	Normal tension glucoma
OA	Optic atrophy
OA1	Ocular albinism type 1
OCT	Optical coherence tomography scan
OD	Right eye
OMIM	Online mendelian inheritance in man website
ONL	Outer Nuclear Layer
ONT	Oxford nanopore technology

OPL	Outer Plexiform Layer
ORF15	Open reading frame 15
OS	Outer Segment
OS	Left eye
OTX1	Orthodenticle homeobox 1
OTX2	Orthodenticle homeobox 2
PCR	Polymerase chain reaction
PD	Pattern dystrophy
PDE	Phosphodiesterase
PDE6	Phosphodiesterase 6
PDE6*	Activated Phosphodiesterase 6
PE	Phosphatidylethanolamine
PED	Pigment epithelial detachment
pERG	Pattern electroretinogram
PHARC	Polyneuropathy, hearing loss, ataxia, retinitis pigmentosa and cataract
pLI	Probability of loss of function score
POS	Photoreceptor Outer Segment
R9AP	RGS9 anchor protein
RAX	Retinal homeobox protein
RCC1	Regulator of chromosomal condensation 1
RCS	Royal College of Surgeons
RD	Retinal detachment
RDH5	11-cis-retinol dehydrogenase 5
RDH8	all-trans retinol dehydrohenase 8
REs	Retinyl esters
RGCs	Retinal Ganglion Cells
RGS9	Regulator of G protein signalling isoform 9
RHO	Rhodopsin
RHO*	Activated Rhodopsin
RK	Rhodopsin kinase
RNA	Ribonucleic acid

ROH	Runs of homozygosity
ROI	Region of interest
ROX	Carboxy-X-rhodamine
RP	Retinitis Pigmentosa
RP19	Retinitis pigmentosa 19
RPE	Retinal Pigment Epithelium
rpm	Revolutions per minute
RR	Read ratio
RT	Room Temperature
S.O.C.	Super Optimal broth with Catabolite repression medium
SAM	Security accounts manager
SBS	Sequencing by synthesis
SCA5	Spinocerebellar ataxia-5
SDS	Sodium Dodecyl Sulfate
SFB	Short fragment buffer
SFs	Secondary findings
sgRNA	Single guide RNA
SGS	Second generation sequencing
SIFT	Sorting intolerant from tolerant
SLCs	Solute carriers
SLN	Senior-Løken syndrome
SMRT	Single molecule read technology
SNARs	Small NF-90 associated RNAs
SNPs	Single nucleotide polymorphisms
snRNAs	small nuclear RNAs
SNVs	Single nucleotide variants
SPG18	Spastic paraplegia-18
SQB	Sequencing buffer
sQTLs	Splicing quantitative trait loci
SRF	Subretinal fibrosis
ssDNA	Single stranded DNA

STGD1	Stargardt disease 1
STRs	Short tandem repeats
TADs	Topologically associated domains
TAE	Tris-acetate-EDTA
TALENs	Transcription activator like effector nucleases
TBCC	Tubulin folding cofactor C-like domain
TE	Tris-EDTA
T_m	Melting temperature
TS	Targeted sequencing
UCSC	University of California Sana Cruz
UK IRDC	UK Inherited retinal disease consortium
USH	Usher syndrome
UTRs	Untranslated regions
UV	Ultraviolet light
VA	Visual acuity
VCF	Variant call format
VEP	Variant effect predictor
VMT	Vitreomacular traction
VNTRs	Variable number tandem repeats
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing
WSP	Tryptophan-serine-proline motif
XL	X-linked
XLRP	X-linked retinitis pigmentosa
XLRS	X-linked retinoschisis
ZFNs	Zinc-finger nucleases
ZMW	Zero mode waveguides

Chapter 1

Introduction

1.1 Human eye

The human eye is a specialised sense organ that connects us to the outside world by processing the light and allowing us to interpret shapes, colours, movements, and other features. It is a part of the sensory nervous system that can receive visual images and send the signal information as neural impulses to the brain.

1.1.1 Embryonic eye development

In humans, eye development commences at around the fourth week of intrauterine life, when the optic sulci start to appear as shallow grooves in the inner part of the neural plate (Forrester et al., 2020). The eye components originate from three embryonic tissue sources, the neuroectoderm, the surface ectoderm, and the periocular mesenchyme (Heavner and Pevny, 2012). The retina, retinal pigment epithelium (RPE), iris, ciliary body and optic nerves are derived from the optic vesicle and optic cup, which arise from the neuroectoderm of the diencephalon. The surface ectoderm gives rise to the lens placode, which eventually goes on and forms the lens vesicle and lens pit that contribute to the formation of the lens and corneal epithelium. The periocular mesenchyme is derived from a mixture of neural crest cells and mesoderm and generates the choroid, sclera, and fibrous and vascular coat of the eyeball (Forrester et al., 2020). (Figure 1.1)

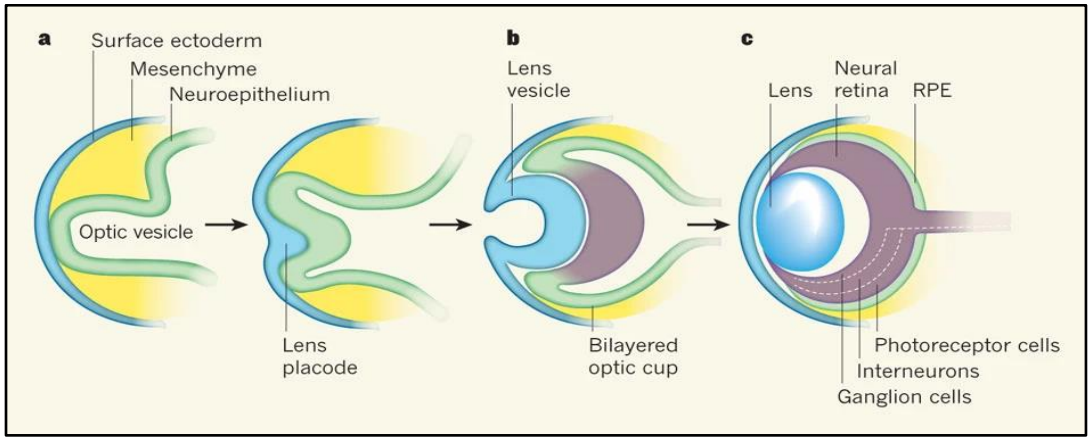


Figure 1.1: Eye embryonic development. a) The surface ectoderm thickens and invaginates with the underlying optic vesicle to form the lens placode. **b)** The lens vesicle develops from the lens placode, and the optic cup self-folds to produce a bi-layered optic cup. **c)** The neural retina (photoreceptors, interneurons, and retinal ganglion cells) emerges from the inner layer of the bi-layered optic cup, and the RPE emerges from the outer layer. (Reproduced from (Ali and Sowden,

2011) with permission of Nature (regenerative medicine) Copyright Clearance Centre, license number: 5393841126714).

The formation of optic vesicles starts on day 22 of embryonic development when the optic sulci invaginate laterally in the surrounding mesoderm toward the surface ectoderm. Optic vesicles then continue growing laterally and remain connected to the forebrain by the optic stalk. Where they meet the surface ectoderm, it forms a localised thickening called a lens placode. The lens placode depresses and forms the lens pit which then gives rise to the lens vesicle. The optic vesicle folds on itself, creating a double layered goblet-shaped optic cup. The optic cup margins (optic cup rims) grow and cover the lens vesicle except on the inferior surfaces of the lens and the optic stalk, forming a gap called the choroidal fissure. The choroidal fissure contains the hyaloid blood vessels that supply the developing optic and lens vesicles. The distal end of the hyaloid vessels degenerates once the lens matures, whereas the proximal end forms the central vessels of the retina (Sontakke, 2018)

The two layered optic cup can be divided into two portions: the anterior (1/5) and posterior (4/5) portions. The anterior portion (the rim) gives rise to the iris, ciliary body, and pupillary muscles. The posterior portion forms the retina in the mature eye. At around four and half weeks, the thin outermost layer of the posterior portion is destined to form the RPE and, two weeks later, the inner layer forms the neural retina. The two layers are separated by the intraretinal space. A common retinal progenitor cell gives rise to all retinal cells (Nguyen-Ba-Charvet and Chédotal, 2014). The cells immediately next to the intraretinal space start to differentiate into the outer nuclear layer, which contains the photoreceptor cells (rods and cones), while the next layer of cells will form the inner nuclear layer (amacrine, bipolar, horizontal, and Müller cells), and finally the innermost superficial cells will give rise to the ganglion layer which contains the ganglion cells (Rosen and Mahabadi, 2021). The optic nerve is made up of the ganglion cell axons and its development is complete by week eight. All layers of the retina will be recognisable by eight months of gestation, but the photoreceptors continue maturing as the baby grows.

During the sixth and seventh week of intrauterine life, the mesenchyme condenses into two layers: a fibrous outer layer, the sclera, and a vascular pigmented inner layer, the choroid. The corneal stroma is derived from the surrounding mesenchyme whereas the corneal epithelium comes from the surface ectoderm and the endothelium from the neural crest cells. In the centre of the optic cup just behind the lens, the vitreous body is formed. It contains the vitreous humour, a gel-liked substance that originates from the neural crest

mesenchymal cells. Eyelids begin to form from neural crest cells and surface ectoderm at week six of intrauterine life. They form as two folds of skin and remain attached over the cornea until they separate at the 27th week of gestation.

1.1.2 Genetic regulation of ocular development

The development of the human eye (Section 1.1.1) is a complicated and strictly regulated process. It consists of multiple overlapping stages that can be categorised as eye field specification and splitting from the forebrain, evagination and patterning of the optic vesicle, regionalization of the optic cup into neural retina and RPE, and the growth and differentiation of the retinal cells. Genetic regulation (Table 1.1) of ocular development is mediated by both intrinsic factors (transcription factors of the homeobox and basic helix-loop helix families) and extrinsic factors (morphogens and growth factors) (Zagozewski et al., 2014).

Genes	Ocular development stage	Disease when mutated
RAX	Eye field specification and splitting Optic vesicle Retinal cell differentiation (photoreceptors)	Anophthalmia
PAX6	Optic sulcus, cup, and stalk Lens, corneal and conjunctival epithelium Mesenchymal cells	Anophthalmia Aniridia (Axenfeld-Rieger syndrome) Peter's anomaly Congenital glaucoma
SIX3	Eye field specification and splitting	Holoprosencephaly
OTX2	Eye field specification and splitting	Anophthalmia Retinal dystrophy
PTX3	Lens vesicle development	Congenital cataract Leucoma Peter's anomaly
CHD7	Neuroectoderm, lens vesicle	CHARGE syndrome
MAF	Lens placode, lens vesicle, primary lens fibres	Defects in lens, cornea, and iris (coloboma) Peter's anomaly
FOXE3	Lens placode	Peter's anomaly Cataract
PITX2 FOXC1	Periocular mesenchyme	Aniridia (Axenfeld-Rieger syndrome) Juvenile glaucoma (50%)

CRYA, CRYB, CRYG	Lens	Cataract
SOX1, SOX2, SOX3	Sensory placodes *SOX2* in lens placode	Anophthalmia

Table 1.1: A summary of major genes involved in ocular development. (Sources: (Forrester et al., 2020) and (Zagozewski et al., 2014)).

1.1.3 Basic anatomy of the human eye

The eye is composed of three layers namely: the outer layer, the middle layer, and the inner layer (Figure 1.2). The outer protective layer, or fibrous tunic is composed of the cornea and sclera, which maintain the shape of the eye globe and protect the inner structures against infections and structural damage. The cornea is a transparent avascular layer through which the light passes to the lens and retina. It is composed of five basic layers: epithelium, Bowman's membrane, lamellar stroma, Descemet's membrane, and the endothelium (Farjo et al., 2009). Each layer has a crucial role in maintaining corneal function. The corneal epithelium acts as an envelope that shields the cornea against foreign bodies and provides it with the essential nutrients and oxygen from tears.

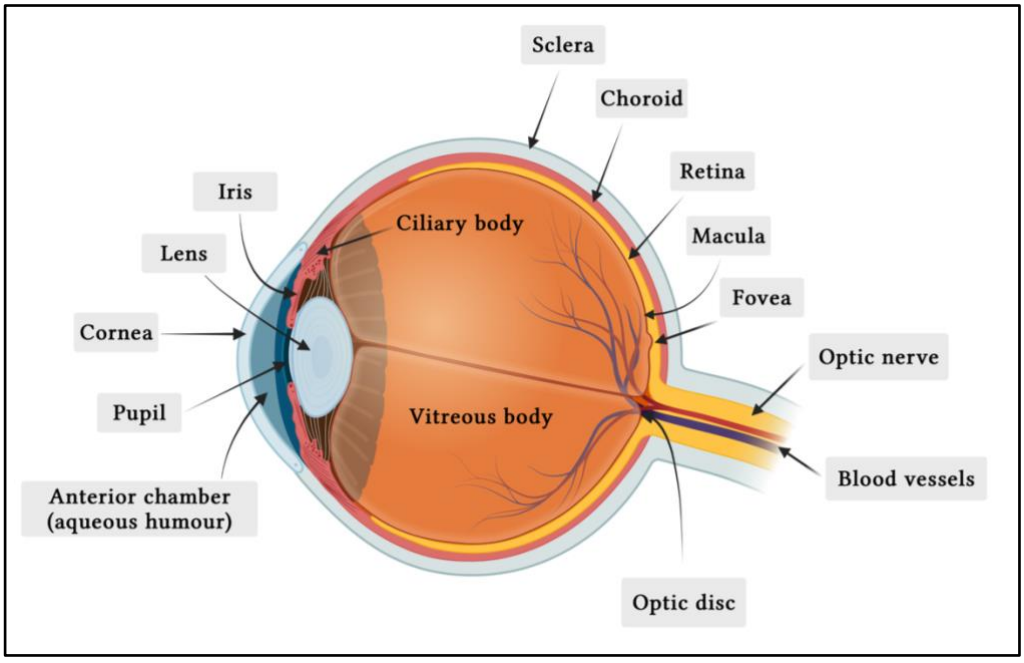


Figure 1.2: Schematic illustration of a human eye anatomy (cross section view). A cross section of the human eye showing the three major layers of the eye, including the outer layer (cornea and sclera), middle layer (iris, choroid, and ciliary body) and the innermost layer (retina). (Created with BioRender.com, Accessed on 22/09/2022).

The transparency of the cornea is significant for the light refraction, and it is preserved by strong layered collagen fibres (Forrester et al., 2020). In collaboration with the lens,

the cornea focuses the light on the photoreceptor cells in the retina and reduces its scattering. The sclera is a dense white connective tissue coat extending from the cornea all the way to the optic nerve. It attaches anteriorly to the limbus of the cornea and fuses posteriorly with the optic nerve sheath. It consists of proteoglycans, elastin, glycoprotein, and collagen fibrils that contribute to maintaining the eyeball shape and provide a resistance against trauma and damage to the vital eye components. The middle layer is made up of iris and ciliary body anteriorly and connects to the choroid posteriorly. The iris is the coloured muscular part of the eye that is located between the cornea and the lens. It controls the pupil size and thus regulates the amount of light reaching the retina. It is connected to the ciliary body which produces the aqueous humour and controls the lens power and shape. The choroid is a vascular layer that supplies the innermost layer of the retina with oxygen and nutrients. These ocular layers surround three transparent structures, the lens, the aqueous and the vitreous (Willoughby et al., 2010, Forrester et al., 2020).

1.2 Retina

The retina is a compound, light sensitive neural tissue lining the back of the eye on the inside. It is a part of the central nervous system and around 80% of all human sensory information is thought to be of retinal origin (Sharma and Ehinger, 2003). It consists of two main layers: the pigment layer (RPE) and the neural layer (outer nuclear layer (photoreceptor cells), inner nuclear layer (bipolar, horizontal, amacrine cells), and ganglion cell layer (ganglion cells and displaced amacrine cells (Lee et al., 2016)). It is like the film in a camera, in that it captures an image and allows our eyes to communicate that to the brain. In simple terms, the light rays enter the eye and are focussed by the cornea and lens onto light-sensitive cells called photoreceptors. Subsequently, these cells transform the light rays into electrical impulses which travel down the optic nerve to be processed by the brain.

The major layers and component cell types of the retina are shown in Figure 1.3.

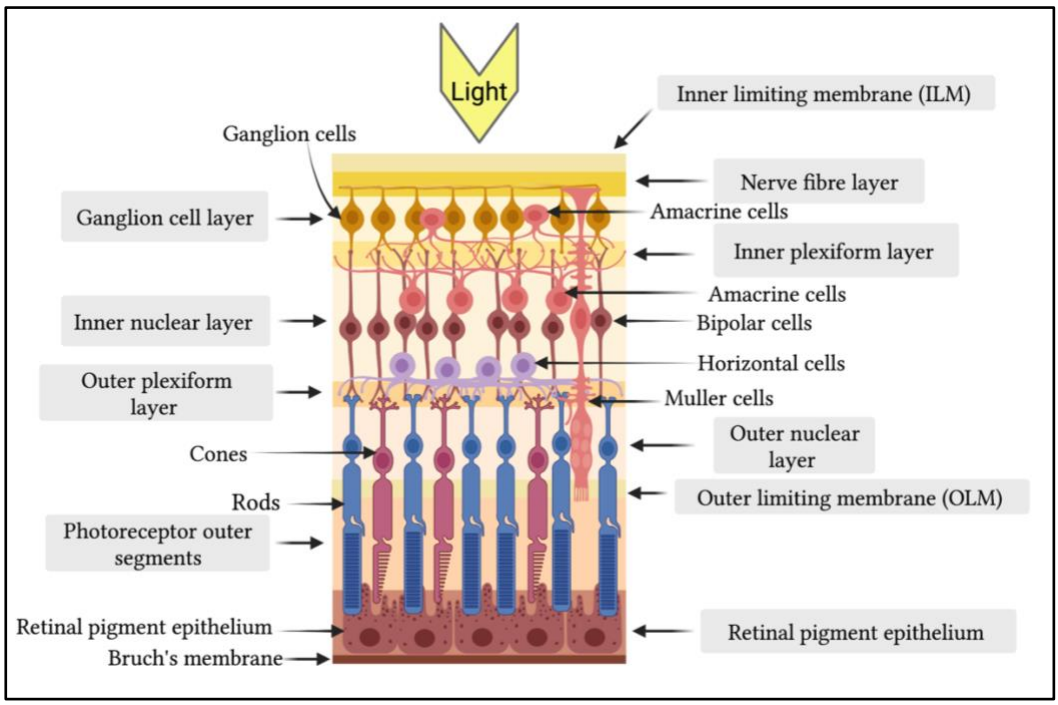


Figure 1.3: Schematic illustration of a cross section of the retina showing the cells and layers of the retina. In this cross section, the retinal cells are organised into two main layers: the retinal pigment epithelium (RPE) and the neural layer containing outer nuclear layer (rod and cone photoreceptors), inner nuclear layer (bipolar, horizontal, and amacrine cells), and ganglion cell layer (ganglion cells and displaced amacrine cells). (Created with BioRender.com, Accessed on 30/04/2023).

1.2.1 Retinal Pigment Epithelium (RPE)

The retinal pigment epithelium is of neuroectodermal embryonic origin (Section 1.1.1). It is formed of a single layer of polygonal pigmented cells located at the outermost layer of the retina (Yang et al., 2021b) between the photoreceptors and the choroid vasculature. Its basolateral membrane connects to Bruch's membrane, which separates the RPE from the choriocapillaris, while the inner apical membrane of the RPE harbours microvillous structures that extend between the light sensitive photoreceptor outer segments (POS) (Strauss, 2005, Yang et al., 2021b). RPE cells serve a critical function in preserving retinal homeostasis. The RPE is a part of the blood-retina barrier (BRB), which regulates the flow of various substances through the RPE cell-cell tight junctions. RPE cells transport water, ions, and metabolic end products from the subretinal region to the bloodstream, as well as supplying vital nutrients, including fatty acids, glucose, and retinol, from the bloodstream to the photoreceptors (Miller and Edelman, 1990, de la Cour, 1993, Strauss, 2005). The presence of melanin in RPE cells aids in the protection

of the retina and internal nerves from damage caused by ultra-violet light (Yang et al., 2021b). Importantly, the RPE maintains photoreceptor function by re-isomerizing all-*trans*-retinal back into 11-*cis*-retinal (Steinberg, 1985, Strauss, 2005) and therefore restoring excitability during the visual cycle (Section 1.4.2). It also supports photoreceptor function through its ability to phagocytose the detached outer segment discs shed by photoreceptor cells. Outer segments are maintained and renewed by forming new membrane discs at their proximal ends (Bok, 1993). This renewal, as well as the RPE's ability to generate a number of growth factors, is critical for photoreceptor function and choriocapillaris endothelium structural integrity (Strauss, 2005). Since the RPE performs these complex various functions, its shape and function are critical for normal vision, and any changes can result in retinopathies such as retinitis pigmentosa (RP), cone rod dystrophy (CORD), age related macular degeneration AMD and Stargardt disease, and ultimately to blindness.

1.2.2 Neural Retina

The neural retina is a thin, transparent layer that transforms light stimuli into neural impulses, which then undergo additional processing by ganglion cell axons before being transmitted to the brain (Forrester et al., 2020). It is made up of five main types of neurons stratified in three layers of neural cell bodies. Photoreceptors (rod and cone cells) are found in the outer nuclear layer (ONL), bipolar cells, horizontal cells and amacrine cells are found in the inner nuclear layer (INL), while the ganglion cell layer (GCL) contains ganglion cells and displaced amacrine cells (Willoughby et al., 2010, Lee et al., 2016). There are two layers composed mainly of synapses, the outer plexiform layer (OPL) and inner plexiform layer (IPL). Another type of retinal cells are the Müllerian glial cells, which are non-neuronal and extend throughout the retina, serving as an organisational backbone of the neural retina (Purnyn, 2013, Yao et al., 2018).

1.2.2.1 Outer nuclear layer (photoreceptors)

Light is focused by the cornea and lens onto the retina, a light-sensitive tissue lining the back of the eye. Vision begins when photoreceptor cells in the retina perceive light and convert the photons into a nerve signal in a process called phototransduction (Section 1.4). There are two types of photoreceptor cells that are distributed unequally across the retina: rods and cones. In the human retina, there is estimated to be 92 million rods (77-107 million) and 4.6 million cones (4-5 million) (Curcio et al., 1990). While only cones can be found in the foveola, rods predominate throughout the rest of the fovea and the entire peripheral retina (Hildebrand and Fielder, 2011). The photoreceptor cells can be

distinguished based on their morphologies and spectral sensitivities (Hussey et al., 2022). As a result of its high sensitivity and ability to detect the absorption of a single photon, rods are responsible for low light (night) vision (scotopic vision) and motion detection (Baylor et al., 1979). Cones are faster but less light-sensitive photoreceptors and are primarily involved in high-acuity daytime and colour vision (photopic vision) (Nathans et al., 1986). Cones do not saturate under high level illumination, whereas the response of rods does in bright light conditions (Mustafi et al., 2009). The sensitivity of photoreceptors is determined by which light-sensitive opsin protein they express, as different opsin proteins are sensitive to different wavelengths of light. Hence, cones can be categorised into three subtypes based on the expression of opsin proteins: S-opsin cones (short wavelength/blue cones), M-opsin cones (mid wavelength/green cones), or L-opsin cones (long wavelength/red cones) (Hussey et al., 2022). In contrast, all rods express one opsin, rhodopsin (RHO).

Both rods and cones consist of outer segment (photopigment), inner segment (mitochondria, endoplasmic reticulum, and Golgi apparatus), cell body (nucleus and axons) and the synaptic terminal (Figure 1.4). The outer segment is a modified cilium that stores the light capturing photopigments and interacts with the extended processes of the RPE. A non-motile cilium connects it to the inner segment, which is rich in mitochondria and functions as a cellular apparatus required for the formation of photopigment membrane discs. The shed discs in the outer segment are phagocytosed by the RPE (Bok, 1993). The outer segments in rods are long, cylindrical, and loaded with parallel membranous discs independent from the plasma membrane. In cones, they are short and conical with pointed ends and their membranous discs are continuous with the plasma membrane (Hildebrand and Fielder, 2011).

Rods and cones have diverse transduction methods that are largely influenced by how they react to various ranges of light intensity. For instance, a single photon of light is sufficient for a rod to provide a reliable response, whereas a cone requires more than 100 photons to produce a comparable reaction. Many rods communicate with a single rod bipolar cell, and numerous rod bipolar cells connect with a single amacrine cell, amplifying the response to even a small stimulus. For cones, a one-to-one interaction between cones, bipolar, and ganglion cells is necessary to maximise visual acuity. Only one cone contacts one bipolar cell, which in turn sends an input to a single retinal ganglion cell (Mustafi et al., 2009, Hildebrand and Fielder, 2011).

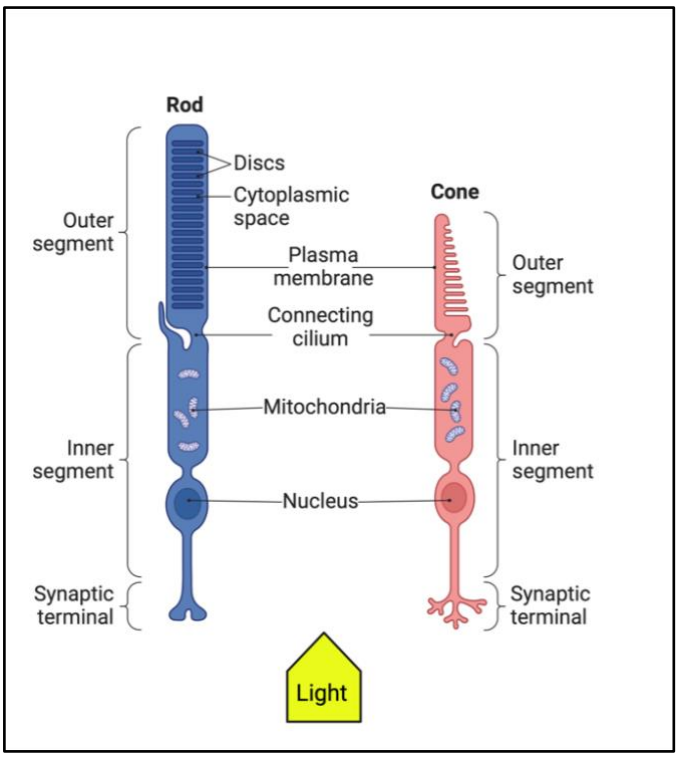


Figure 1.4: Morphology of photoreceptors (rods and cones). The structure of rods and cones showing the outer segment, the inner segment, the cell body, and the synaptic terminal. (Created with BioRender.com, Accessed on 29/09/2022).

1.2.2.2 Inner nuclear layer (interneurons)

The inner nuclear layer (interneurons) connects the photoreceptor layer with the ganglion cell layer. It consists of bipolar, horizontal, amacrine, and interplexiform cells. These interneurons act as a mediator between photoreceptors and ganglion cells, processing and transmitting signals.

The bipolar cells link the outer and inner retina and transmit visual data from photoreceptors (rods and cones) to amacrine and ganglion cells. The inner nuclear layer contains the cell bodies of bipolar cells, from which an axon and a primary dendrite project into the inner plexiform layer and the outer plexiform layer, respectively. There is one type of rod bipolar cell and are more than ten types of cone bipolar cells in the mammalian retina. The dendrites of cone bipolar cells only interact with cones and are engaged in photopic vision, whereas those of rod bipolar cells are in contact with rods and are involved in scotopic vision (Grünert, 2009). Based on their response to the neurotransmitter glutamate, bipolar cells can be categorised as OFF bipolar cells (cones) and ON bipolar cells (rods and cones). In response to the light, the cone OFF bipolar cells become hyperpolarised (inhibited) and the ON bipolar cells (rods and cones) become depolarised (excited) (Hildebrand and Fielder, 2011). Bipolar cells are the most prevalent

interneurons in the retina, accounting for about 40% of all retinal neurons (Masland, 2001).

Both direct and indirect signalling pathways from photoreceptors to ganglion cells incorporate bipolar cells. Conversely, horizontal cells are only implicated in indirect pathways. They are situated in the distal region of the inner nuclear layer, immediately after the photoreceptors, where they connect the rods, cones, and bipolar cells. They prevent the extensive spreading of signals from the axons and dendrites of other retinal cells by inducing a lateral inhibition in the surrounding regions. This process is essential for enhancing perception of contrast and colour in visual images (Purnyn, 2013). Amacrine cells are also involved in the indirect signalling pathways. Amacrine cells are categorised on the basis of size, horizontal dendritic fields, and sublaminar levels within the IPL. According to the dendritic field, there are two types: narrow-field and wide-field. In addition, there are uni-, bi-, and multi- stratified types. Amacrine cells transmit signals to ganglion cells or other bipolar cells after receiving input signals from bipolar cells (Hildebrand and Fielder, 2011, Purnyn, 2013).

Recently, the Campana cells, an unusual form of interneurons, have been identified in the retina. These cells are fundamentally similar to both amacrine and bipolar cells, yet they also differ from both in significant aspects. The Campana cells expand dendrites in OPL and axonal terminal in the IPL. These cells were found to play a role in visual signal processing by transmitting the visual signals from both rods and cones to retinal ganglion cells (RGCs) (Young et al., 2021).

1.2.2.3 Ganglion cell layer

Retinal ganglion cells (RGCs) are typical nerve cells that transmit visual information from the retina to the brain and provide feedback signals to the inner retina by forming gap junctions with other RGCs and amacrine cells (Vlasiuk and Asari, 2021). The ganglion cell body is found in the ganglion cell layer, and the IPL is where the ganglion cell dendrites interact with amacrine and bipolar cells. There are up to 20 different ganglion cell subtypes, with the midget and parasol types accounting for 80% of the total ganglion cell population (Hildebrand and Fielder, 2011). The intrinsically photosensitive retinal ganglion cells (ipRGCs) are a distinct class of ganglion cells that project widely throughout the brain. They are melanopsin-expressing cells that can react to light even in the absence of input from rods and cones. Additionally, ON and OFF subtypes of ipRGCs were shown to affect human visual perception and the circadian rhythm (Aranda and Schmidt, 2021).

1.2.3 Glial cells

In human retina, glial cell types include Müller cells, astrocytes, microglia, and infrequently oligodendrocytes. Müller cells have a cylindrical, fibre-like shape and are the most prevalent type in the retina, interacting with nearly all neurons. Their cell bodies are found in the inner nuclear layer, and their cell processes extend through the whole retina, holding its neuronal components together to maintain its structure and homeostasis (Bringmann et al., 2009a). The distal extensions of Müller cells extend to produce the outer limiting membrane, while the proximal extensions expand to form the *endfeet*, whose basal lamina creates the inner limiting membrane (Hildebrand and Fielder, 2011) (Figure 1.3). They are essential for the visual process to continue functioning optimally by controlling the uptake of neurotransmitters, clearing debris, regulating potassium levels, storing glycogen, electrically insulating receptors, and providing mechanical support for the neural retina (Bringmann et al., 2009b). They serve as a light collectors, directing light toward the cone and rod photoreceptors (Franze et al., 2007). Additionally, they are implicated in synaptogenesis and have a neuroprotection function (Purnyn, 2013). It should be emphasised that Müller cells contribute to the OS assembly, phagocytose the outer segments of cone cells, and take part in the cone visual cycle (Section 1.4.2.2) (Wang et al., 2004, Wang and Kefalov, 2011, Goldman, 2014).

1.3 Metabolism in retina

The RPE maintains both its metabolic balance and that of the neural retina by filtering metabolites in both directions through the interphotoreceptor matrix. RPE metabolism affects photoreceptors, and photoreceptors degenerate as RPE cells die. The transport of glucose across the RPE through the GLUT-1 glucose transporters is essential for photoreceptor growth and when GLUT-1 gene expression was blocked, photoreceptors were stunted (Swarup et al., 2019). Furthermore, RPE mitochondrial metabolism is crucial for the survival of photoreceptors. The loss of mitochondria changed the metabolism of the RPE cells, which caused the photoreceptors to degenerate (Zhao et al., 2011). The RPE can also be impacted by altered photoreceptors metabolism. Blocking AMPK, an intracellular energy sensor, expression in the mouse retina but not in the RPE led to fragmentation of mitochondria in the photoreceptors, which in turn disrupted the RPE cells (Xu et al., 2020). Metabolic dysregulation in the retina will cause retinal dystrophies such as age-related macular degeneration (AMD) and retinitis pigmentosa (RP) (Hurley, 2021).

1.4 Visual phototransduction and the visual cycle

1.4.1 Visual phototransduction

Phototransduction is a sensory transduction mechanism in the visual system and is one of the most well-studied G-protein signalling systems (Fu, 1995). It is the process of converting a photon of light into a neuronal signal (Lamb and Pugh, 2006). Visual pigments in the outer segment (OS) of retinal rod and cone photoreceptor cells capture light photons and generate an electrical response (Arshavsky et al., 2002). The plasma membrane of photoreceptor cells is then hyperpolarized as a result of a series of biochemical and electrophysiological actions (Ebrey and Koutalos, 2001). A visual pigment (photopigment) consists of opsin that is covalently linked to a chromophore (11-*cis*-retinal, 11cRAL, A1). Opsins are seven helix G-protein-coupled-receptors (GPCRs) that are involved in the conversion of light of various intensities and wavelengths into neural stimuli. Only one type of opsin is expressed by each photoreceptor. Rods have rhodopsin, which transduces dim light, whereas cones have photopsins, which provide colour vision. The chromophore is a light-absorbing retinaldehyde (11-*cis*-retinal) that functions as a strong antagonist, keeping the rhodopsin inactive (Lamb and Pugh, 2006, Arshavsky et al., 2002). In addition to rhodopsin, the heterotrimeric G protein transducin (G), the cyclic nucleotide phosphodiesterase (PDE6), and the cyclic nucleotide gated channels (CNGCs) in the plasma membrane are all required for the phototransduction cascade activation (Lamb and Pugh, 2006).

Activation of phototransduction: Upon the absorption of light in the outer segment, chromophore (11-*cis*-retinal) undergoes isomerization to all-*trans*-retinal (atRAL), inducing a conformational shift in the rhodopsin (RHO) to its activated form meta-rhodopsin II (RHO*). Rhodopsin achieves the conformational state (RHO*) in just one millisecond through a series of intramolecular changes (Arshavsky et al., 2002). RHO* then activates a G protein called transducin by interacting with its GDP-bound form that has alpha, beta, and gamma subunits. Through this interaction, transducin separates from GDP and binds GTP, causing the beta and gamma subunits to be expelled. Phosphodiesterase (PDE6) is activated by the alpha subunit-GTP transducin (Gα-GTP, G*), which binds to the inhibitory gamma subunits of PDE6 while also activating the alpha and beta subunits. Activated PDE6 (PDE6*) hydrolyses the cGMP to 5'GMP which in turn causes closure of cyclic nucleotide gated channels (cation channels). Sodium channel closure causes hyperpolarization, which in turn causes the voltage-gated calcium channel to close. This decrease in calcium levels, in turn, reduces the amount of the

neurotransmitter glutamate, which causes depolarization of ON bipolar cells (in rods and cones) and hyperpolarization of cone OFF bipolar cells (Lamb and Pugh, 2006, Fu and Yau, 2007, Wang and Kefalov, 2011) (Figure 1.5A).

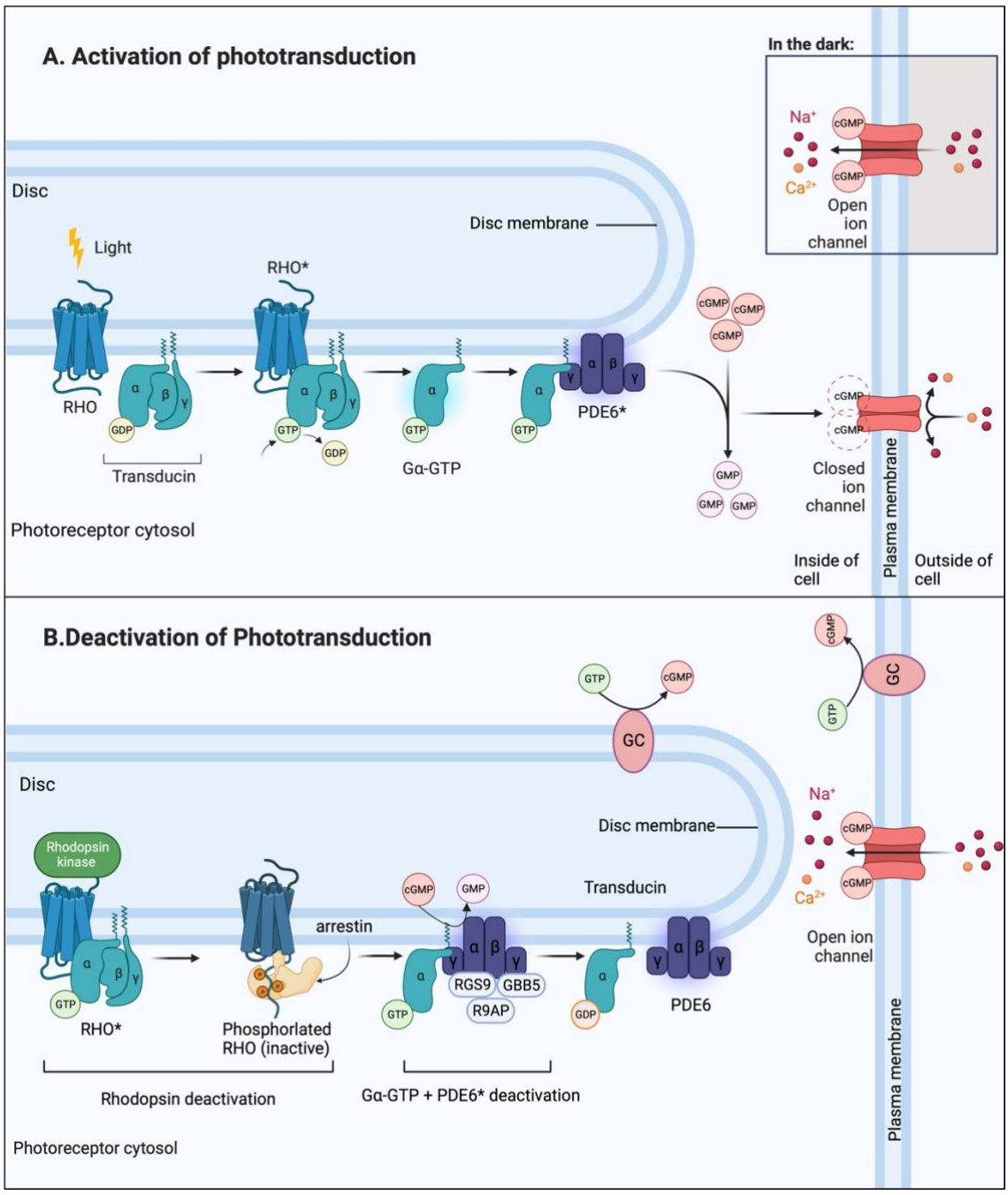


Figure 1.5: Representation of phototransduction cascade molecular steps. **A.** activation of phototransduction that results in closing cation channels (CNGCs) preventing further influx of Na+ and Ca+ ions. **B.** Deactivation of phototransduction cascade and recovery of dark state by reopening cation channels. RHO= rhodopsin (inactive); RHO*= rhodopsin (active); Gα-GTP= transducin (active); PDE6= phosphodiesterase 6 (inactive); PDE6*= phosphodiesterase 6 (active); RGS9= regulator of G protein signalling isoform 9; GBB5= complex of G protein beta 5; R9AP= RGS9 anchor protein; GC= guanylate cyclase. (Created with BioRender.com, Accessed on 29/09/2022).

Deactivation of phototransduction: Terminating the light response occurs when the Ca^{2+} level decreases, making the intracellular Ca-recoverin-RK complex dissociate into Ca^{2+}, recoverin, and rhodopsin kinase (RK), a G-protein coupled receptor kinase (GRK). Meta-rhodopsin II (RHO*) is phosphorylated by rhodopsin kinase to inactivate it and decrease its affinity for transducin. Arrestin then binds to the phosphorylated RHO to deactivate it completely. Furthermore, transducin is inactivated by stimulating the intrinsic GTPase activity, which hydrolyses GTP into GDP and subsequently converts Gα-GTP back to its inactive form Gα-GDP. This process is accelerated by a complex formed by the G-protein beta 5 (Gβ5), regulator of G protein signalling isoform 9 (RGS9), and RGS9 anchor protein (R9AP), which binds to the Gα-GTP-phosphodiesterase complex, resulting in dissociation of phosphodiesterase from Gα-GTP. Phosphodiesterase inactivation results in a cessation to PDE-mediated cGMP hydrolysis, with cGMP levels then being restored by the continued activity of guanylate cyclase (GC), leading to the reopening of cGMP gated channels (Arshavsky et al., 2002, Fu and Yau, 2007, Arshavsky and Wensel, 2013) (Figure 1.5B). Genes and proteins involved in phototransduction are detailed in Table 1.2.

Protein description	Role in phototransduction	Gene
Rod components		
Rhodopsin	Activation	*RHO*
Transducin α subunit	Activation	*GNAT1*
G protein β subunit 1	Activation	*GNB1*
Transducin γ subunit	Activation	*GNGT1*
PDE6 catalytic subunit	Activation	*PDE6A* *PDE6B*
PDE6 inhibitory subunit	Activation	*PDE6G*
Cyclic nucleotide-gated channel α	Activation and Ca+ feedback	*CNGA1*
Cyclic nucleotide-gated channel β	Activation and Ca+ feedback	*CNGB1*
G protein receptor kinase	Recovery	*GRK1*
Arrestin	Recovery	*SAG*
Recoverin Visinin	Ca+ feedback	*RCVRN*
Na+/Ca+, K+ ion exchanger	Ca+ feedback	*SLC24A1*
Cone components		
Long-wave sensitive opsin	Activation	*OPN1LW* *OPN1MW*

Short-wave sensitive opsin	Activation	*OPN1SW*
Transducin α subunit	Activation	*GNAT2*
G protein β subunit 3	Activation	*GNB3*
Transducin γ subunit	Activation	*GNGT2*
PDE6 catalytic subunit	Activation	*PDE6C*
PDE6 inhibitory subunit	Activation	*PDE6H*
Cyclic nucleotide-gated channel α	Activation and Ca+ feedback	*CNGA3*
Cyclic nucleotide-gated channel β	Activation and Ca+ feedback	*CNGB3*
G protein receptor kinase	Recovery	*GRK7*
Arrestin	Recovery	*ARR3*
Guanylyl cyclase activating protein	Ca+ feedback	*GUCA1C*
Na+/Ca+, K+ ion exchanger	Ca+ feedback	*SLC24A2*
Shared components		
Regulator of G-protein signalling 9	Recovery	*RGS9*
G-protein β subunit 5	Recovery	*GNB5*
RGS9 anchor protein	Recovery	*RGS9BP*
Guanylyl cyclase	Recovery and Ca+ feedback	*GUCY2D*
Guanylyl cyclase	Recovery and Ca+ feedback	*GUCY2F*
Guanylyl cyclase activating protein	Ca+ feedback	*GUCA1A*
Guanylyl cyclase activating protein	Ca+ feedback	*GUCA1B*

Table 1.2: Visual phototransduction genes and proteins. Proteins, genes, and their role in phototransduction. (Sources: RetNet: https://sph.uth.edu/retnet/sum-dis.htm#A-genes; (Lamb, 2020)).

1.4.2 The Visual cycle

All-*trans* retinol (vitamin A), the chromophore in the human visual system, is taken up by RPE from the choroid blood vessels and converted to 11-*cis* retinal then introduced into photoreceptors. The 11-*cis* retinal and opsin (GPCR) combine to produce the visual pigment, and after the pigment is photoactivated, visual phototransduction begins in the outer segments of photoreceptors (Wang and Kefalov, 2011).

1.4.2.1 Classic rod/cone visual cycle (retinoid cycle)

Upon photon absorption, 11-*cis* retinal undergoes photoisomerization to all-*trans* retinal, which causes changes in the conformation of opsin, starting the phototransduction cascade (Section 1.4.1). All-*trans* retinal must be converted back to 11-*cis* retinal through a sequence of enzymatic events that take place in photoreceptors and RPE, in order to render it ready to be used again, thereby preserving the retina in an active state, ready to undergo normal visual phototransduction.

11-*cis* retinal is isomerized to all-*trans* retinal by light, and all-*trans* retinal detaches from the active opsin (Scheerer et al., 2008). It is then released into the cytoplasm, where it is reduced to all-*trans* retinol either immediately by all-*trans* retinol dehydrogenase 8 (RDH8) or alternatively by forming a complex with phosphatidylethanolamine (PE) inside the disc membrane lumen known as N-retinylidene-PE, which is then transported to the cytoplasm by ABCA4 protein, and finally reduced by RDH8 (Rattner et al., 2000, Beharry et al., 2004). Interphotoreceptor retinoid binding protein (IRBP) binds to all-*trans* retinol to transport it from the outer segment of photoreceptors into the RPE (Gonzalez-Fernandez and Ghosh, 2008). In the RPE, lecithin retinol acyl transferase (LRAT) converts all-*trans* retinol to retinyl esters (REs) (Saari and Bredberg, 1989, Ruiz et al., 1999). These all-*trans* retinyl esters are then hydrolysed and isomerised to 11-*cis* retinol by the isomer-hydrolase retinal pigment epithelium 65 (RPE65) (Moiseyev et al., 2005, Redmond et al., 2005). Finally, the 11-*cis* retinol is subsequently oxidised to 11-*cis* retinal by the enzyme 11-*cis* retinol dehydrogenase 5 (RDH5) in the presence of NAD+ as a cofactor (Simon et al., 1995), and it is then coupled to IRBP to diffuse from the RPE to the photoreceptors where it interacts with the opsin and regenerates the visual pigment (Choi et al., 2021). Figure 1.6 shows the enzymatic reactions involved in visual cycle.

1.4.2.2 Cone visual cycle

The rate of visual chromophore reproduction through the classic visual cycle is insufficient to meet the need for 11-*cis* retinal necessary to sustain light sensitivity in a daytime setting. Cones and their photopigments control colour perception and daytime vision, and hence maintaining cone sensitivity to light requires an additional chromophore reproduction mechanism via Müller cells (Wang and Kefalov, 2011). The intraretinal cone visual cycle begins after the 11-*cis* retinal being photo-isomerized to 11-*trans* retinal, which is subsequently reduced to 11-*trans* retinol and transferred from the cone outer segment to Müller cells. In Müller cells, 11-*trans* retinol is isomerized to 11-*cis* retinol and then esterified to retinyl esters. When 11-*cis* retinol is returned to cone

photoreceptors, it undergoes oxidation to become 11-*cis* retinal, where it combines with cone opsins to produce cone pigments. Isomerase II, dihydroceramide desaturase 1, and multifunctional *O*-acyltransferase (MFAT) are believed to be the three enzymatic reactions that support the cone visual cycle (Muniz et al., 2009, Roosing et al., 2014, Tsin et al., 2018).

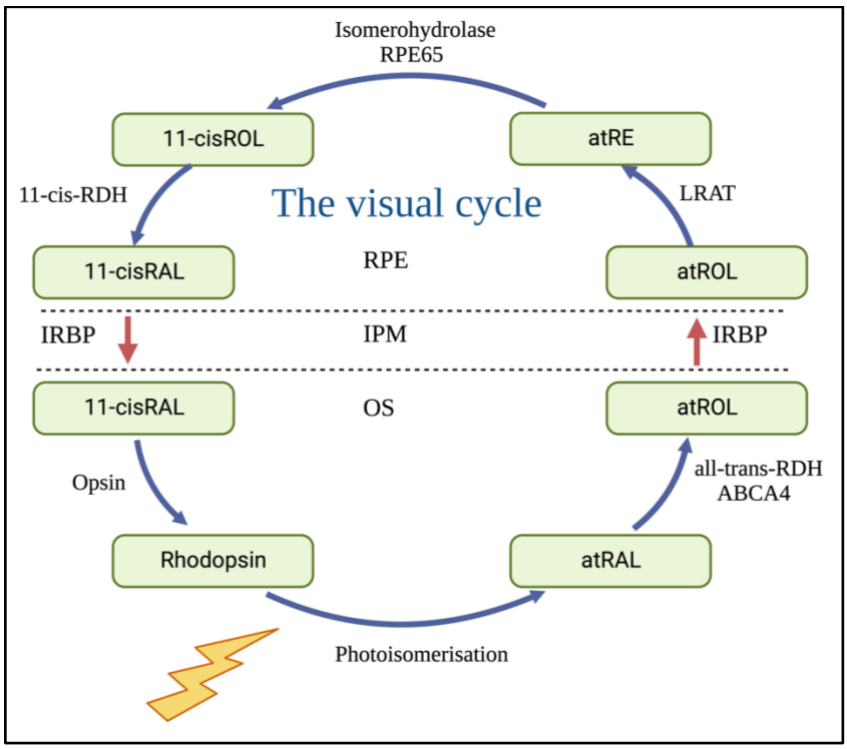


Figure 1.6: The visual cycle between photoreceptor outer segments and the RPE, including enzymatic reactions to recycle 11-*cis* retinal. 11-*cis*ROL= 11-*cis* retinol; 11-*cis*RAL= 11-*cis* retinal; atRAL= all-*trans* retinal; atROL= all-*trans* retinol; atRE= all-*trans* retinyl esters; 11-*cis*-RDH= 11-*cis* retinol dehydrogenase; all-*trans*-RDH= all-*trans* retinol dehydrogenase; LRAT= lecithin retinol acyltransferase; RPE= retinal pigment epithelium; IPM= interphotoreceptor matrix; OS= photoreceptor outer segment; IRBP= interstitial retinol-binding protein. (Created with BioRender.com, Accessed on 10/10/2022).

All of the elements involved in visual phototransduction (Table 1.2), as well as all of the enzymes and binding proteins involved in the visual cycle (Figure 1.6) significantly regulate human vision. Mutations in genes encoding proteins involved in these pathways are responsible for different forms of inherited retinal dystrophies (IRDs) (Section 1.5).

1.5 Inherited retinal diseases (IRDs)

Inherited retinal diseases (IRDs) are amongst the most common cause of irreversible reduced vision and blindness in children and young adults (Bunce et al., 2010). One in 3000 (Sahel et al., 2015)- 4000 (Ayuso and Millan, 2010) people worldwide are affected by these monogenic diseases, which are clinically and genetically heterogeneous. It is challenging to diagnose IRDs due to the fact that different IRDs have overlapping clinical symptoms (Hartong et al., 2006, Hamel, 2007, Chung and Traboulsi, 2009), while multiple phenotypes can result from the same gene (Hull et al., 2014). Various IRD phenotypes have been identified (Chen et al., 2021), and they are categorised according to the age at which symptoms first appear, the mode of inheritance, the location of retinal dysfunction (rod dominant, cone dominant, generalised retinal degenerations, and vitreoretinopathies), the rate of disease progression, and associations with syndromes (Holtan et al., 2020). The majority of IRD cases are non-syndromic and the disease is limited to the eye. However, more than 80 syndromic IRD forms have been described (Tatour and Ben-Yosef, 2020). Different types of IRDs have been linked to mutations in more than 280 genes (https://sph.uth.edu/retnet/), and our understanding of the genetic origins of IRDs is rapidly growing. Additional IRD loci have also been localised to chromosomal regions, though the causative genes remain unknown. In addition to encoding a wide range of proteins, such as structural and transmembrane proteins, visual cycle proteins, and proteins involved in phototransduction (Wright et al., 2010), it has been shown that these genes exhibit various inheritance patterns. IRDs can present clinically in a variety of ways, from moderate dysfunction to severe congenital diseases with early onset. The major non-syndromic and syndromic IRDs that form the foundation for this study are briefly reviewed below.

1.6 Non-syndromic retinal diseases

1.6.1 Retinitis pigmentosa (RP)

Retinitis pigmentosa (RP, OMIM 268000) is the most common inherited retinal dystrophy (IRD) and is a major cause of visual disability and blindness (Ferrari et al., 2011, Zhang, 2016, Verbakel et al., 2018). Depending on the geographic location, the prevalence of retinitis pigmentosa can range from 1:9000 (Na et al., 2017) to as high as 1:750 (Nangia et al., 2012). However, it has been estimated that the frequency is roughly 1:4000 globally (Pagon, 1988). It is characterised by the gradual loss of photoreceptors and the retinal pigment epithelium, which causes pigment deposits to form primarily in

the peripheral retina and then progress toward the central retina (Verbakel et al., 2018). The pigment released from the degenerating cells of RPE is what causes these pigmentary changes (Phelan and Bok, 2000). Rod photoreceptors degenerate first in the majority of RP cases, then cone photoreceptors start to disappear. This pattern of the photoreceptor involvement explains why patients initially present with decreased night vision and loss of peripheral vision, which is followed later in life by visual impairment in daylight settings. Up until later stages of the disease, the macula's functionality remains relatively intact, and in some cases is never lost. However, less common forms could involve cone rod degeneration and significantly affect the macula. For instance, macular atrophy can be evident from mid-stage of the disease caused by *RDS* and *CRX* mutations (Hamel, 2006, Verbakel et al., 2018).

The clinical manifestations of RP most commonly first appear in adolescence, as night blindness (nyctalopia) and trouble adjusting to the dark. These symptoms are followed by loss of peripheral vision or “tunnel vision”, then in some cases a decline in visual acuity caused by central cone degeneration, and eventually total blindness in adulthood (Phelan and Bok, 2000, Hartong et al., 2006, Hamel, 2006, Ferrari et al., 2011, Verbakel et al., 2018). Additionally, some RP patients may develop photophobia in the later stages (Pinckers et al., 1993). The fundus appearance (Figure 1.7) changes as the disease progresses. At an early stage it may appear normal, with few if any bone spicule shaped pigment deposits, and the optic disc appears normal. In the following stage, there is a progressive loss of rod photoreceptors, which results in the appearance of bone spicule deposits in the mid periphery. At this stage the retinal arteries become more narrowed, and the optic disc becomes moderately pale. In the end stage, examination would be expected to reveal a diffuse pigment deposit reaching the macular area, thin vessels, and waxy pallor of the optic disc (Hamel, 2006). Due to children's ability to compensate for peripheral vision loss and the fact that many patients may not notice the difficulty adjusting to the dark at night because of artificial lighting, it can be challenging to determine the exact age of onset. However, early onset RP subtypes frequently progress more rapidly, and the severity of the disease is correlated with the inheritance pattern (Verbakel et al., 2018). Generally, patients with autosomal dominant RP have the best long-term prognosis since they are able to maintain their central vision, while patients with X-linked RP typically experience a more severe course of the disease compared to those with autosomal recessive RP (Hamel, 2006, Verbakel et al., 2018).

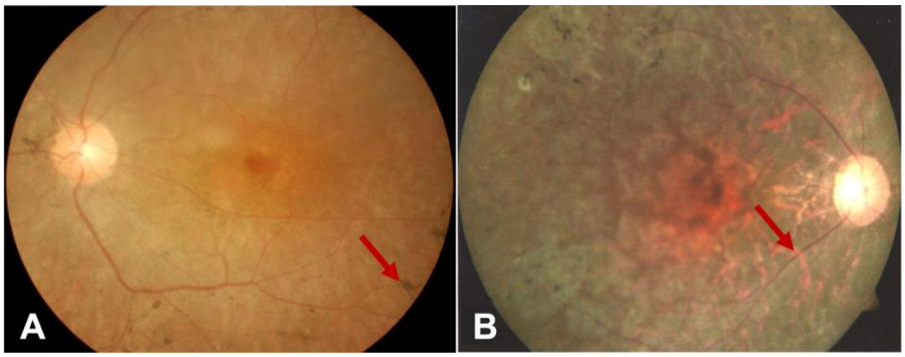


Figure 1.7: Colour images of fundus appearance in RP patients. A. Fundus examination of an RP patient showing bone spicules in the peripheral retina (red arrow), retinal arteriolar attenuation and RPE degeneration. **B.** Fundus examination of a different RP patient showing the spread of bone spicule pigmentation to the mid-peripheral retina, bilateral arteriolar attenuation (red arrow), RPE degeneration and pale optic disc. RP, retinitis pigmentosa; RPE, retinal pigment epithelium. (Adapted from (Yang et al., 2013) with permission from BMJ Copyright Clearance Centre, Licence number: 5412591037473).

Genetically, RP is a very heterogenous disease and can be classified according to the age of onset into early and late onset RP. The diagnosis of early onset RP can be mistaken with the diagnosis of LCA (Hamel, 2006), which is a severe visual impairment that presents from birth (Section 1.6.3). Late onset RP is diagnosed when symptoms appear during or after midlife. RP can also be classified based on the fundus appearance and the localisation of the lesions, with, for example, regional or sectorial, pericentral, parapapillary, and paraveinous disease-types(Hamel, 2006). Additionally, it can be categorized into syndromic (20-30%) and non-syndromic (70-80%) types. Patients with non-syndromic RP have only eye-specific symptoms and signs with no extraocular abnormalities.

The most popular classification of non-syndromic RP is based on the inheritance pattern, categorising it into autosomal dominant (30-40%), autosomal recessive (50-60%) and X-linked RP (5-15%) (Verbakel et al., 2018). Rarely, digenic (due to mutations in two different genes), mitochondrial, and X-linked dominant forms also occur (Ferrari et al., 2011). Up to date, mutations in more than 84 genes (https://sph.uth.edu/retnet/sum-dis.htm#A-genes) were found to be associated with non-syndromic RP and most of them are responsible for only one form of inheritance pattern, except for mutations in: *BEST1*, *NR2E3*, *NRL*, *RHO*, *RP1*, *RPE65* and *SAG*, which can cause both autosomal dominant and recessive forms. With the exception of *RHO*, where mutations account for roughly

25% of dominant RP, *USH2A*, where mutations represent about 20% of recessive diseases, and *RPGR*, where mutations account for approximately 70% of X-linked RP, mutations in the majority of RP genes cause only a small number of cases. Altogether, mutations in *RHO*, *USH2A*, and *RPGR* can account for 30% of all RP cases (Hartong et al., 2006).

Syndromic forms of RP are often caused by mutations in genes involved in ciliary function (Verbakel et al., 2018) and are usually inherited in a recessive manner (Ferrari et al., 2011). This is because the outer segments of photoreceptors are highly modified cilia which have evolved to capture a single external stimulus, light (Section 1.2.2.1). The most common syndromic form is Usher syndrome (USH), involving RP and deafness, a ciliopathy that accounts for about 14% of all RP cases (Boughman et al., 1983). Another well-known syndromic form of RP is Bardet-Biedl syndrome (BBS) (Mockel et al., 2011), where RP typically co-occurs with obesity, polydactyly, learning disability and kidney disease. BBS is less frequent than Usher syndrome (5%) (Ferrari et al., 2011).

1.6.1.1 Autosomal recessive retinitis pigmentosa (ARRP)

Apart from in the very rare circumstance where a de novo mutation has arisen, both parents of a person with autosomal recessive RP must be heterozygotes, which means that each parent must have one pathogenic mutation associated with RP. Each parent has a 50% chance of passing on the mutation to their offspring, and upon conception, there is therefore a 25% risk that their offspring will be affected (carrying two mutant alleles) and a 50% chance that they will be asymptomatic carriers. Carriers of a pathogenic variant often exhibits no symptoms (Fahim et al., 2020a). It is commonly recognised that children of consanguineous unions are more likely to develop recessive disorders because they inherit the autosomal recessive gene mutations from a common ancestor (Hamamy, 2012). Autosomal recessive RP is the most common inherited form of RP accounting for 50-60% of all RP cases (Hartong et al., 2006). The first decade is when ARRP symptoms often appear (Hamel, 2006), and it may overlap with other autosomal recessive retinopathies including Leber congenital amaurosis (LCA) (Fahim et al., 2020a). Mutations in more than 66 genes (https://sph.uth.edu/retnet/sum-dis.htm#A-genes) have been implicated in ARRP (Table 1.3). The majority of them only account for 1% or less of ARRP cases, however some of them, such *USH2A*, *EYS*, *CERKL*, *CRB1*, *PDE6A*, *PDE6B*, *RPE65*, *ABCA4*, *CNGA1*, and *SAG*, can be more prevalent (accounting for >1%) (Fahim, 2018). A wide range of retinal biochemical pathways, including the phototransduction cascade, vitamin A metabolism (visual cycle), cell-cell

communication, signalling, maintenance of photoreceptor cells and cilia, and phagocytosis, are regulated by proteins that are encoded by ARRP-related genes (Hartong et al., 2006).

The most prevalent gene involved in isolated ARRP, and recessive syndromic RP is the Usherin gene (*USH2A*; OMIM 276901). It spans 800 kb with 73 exons (exon 71 is cochlea-specific) at 1q41, and produces the Usherin protein (Bhardwaj et al., 2022). It is expressed in two isoforms as a result of alternative splicing: isoform a (short with 21 exons) and isoform b (long with 72 exons). The short isoform a is expressed in both the inner ear and the retina (Pearsall et al., 2002) whereas the long isoform b is strongly expressed in retinal photoreceptors and translated into the Usherin protein of 5202 amino acids that is attached to the cell membrane (Van Wijk et al., 2004, Reiners et al., 2005). Usherin protein is expressed in the connecting cilia in mammalian retina and involved in the cargo delivery from the inner segment (IS) to the outer segment (OS) of the photoreceptor cells (Van Wijk et al., 2004, Reiners et al., 2006). Usher syndrome type 2a (Section 1.7.1) and non-syndromic RP have both been linked to *USH2A* mutations (Chen et al., 2020, Reurink et al., 2021, Meng et al., 2021). The *USH2A* gene is fully penetrant in recessive conditions and there is no evidence of association with dominant diseases. This gene has more than 1100 mutations, including missense, nonsense, deletions, insertions, splicing, indels, and large rearrangements (Bhardwaj et al., 2022). Missense mutations in *USH2A* have recently been linked to the development of keratoconus in a Pakistani family (Ahmed et al., 2021).

The ATP-Binding Cassette, subfamily A, member 4 (*ABCA4*; OMIM 601691) gene was initially discovered in 1997 (Allikmets et al., 1997) and is the most frequent cause of Stargardt disease 1 (STGD1) (Section 1.6.2) (Sung et al., 2020). It is localised on chromosome 1 (1p22.1) and expressed in RPE and the outer segment of photoreceptor cells (Lenis et al., 2018). Its protein is essential for removing the intermediate metabolites of the visual cycle (Section 1.4.2) and its malfunction will eventually cause cytotoxicity in the RPE which subsequently affects the function of the photoreceptors (Bhardwaj et al., 2022). Mutations in *ABCA4* are known to cause in addition to Stargardt disease 1 (STGD1), retinitis pigmentosa 19 (RP19), cone -rod dystrophy 3 (CORD3), age related macular degeneration 2, fundus flavimaculatus (FFM), and early onset severe retinopathy (Huang et al., 2018). It is responsible for about 5-6% of all ARRP (Hartong et al., 2006) and up to 70% of STGD1/AR-MD/AR-CORD cases (Perea-Romero et al., 2021).

The phosphodiesterase 6 gene complex (PDE6-complex) is a family of catalytic genes that controls the amount of cGMP in the cytoplasm of photoreceptors, which is crucial for the phototransduction cascade (Section 1.4.1) (Gopalakrishna et al., 2017). It is a heterotetrametric complex with two catalytic subunits and two inhibitory subunits (Bhardwaj et al., 2022). Catalytic subunits in the rod photoreceptors are alpha (encoded by *PDE6A*) and beta (encoded by *PDE6B*), whereas the inhibitory subunits are two identical gamma subunits, which are encoded by *PDE6G*. Two identical alpha subunits (encoded by *PDE6C*) make up the cone photoreceptors catalytic core, which is inhibited by two identical cone specific gamma subunits (encoded by *PDE6H*) (Gopalakrishna et al., 2017, Kuehlewein et al., 2021, Deng et al., 2018a). Pathogenic variants in the rod-specific *PDE6* genes cause ARRP, whereas pathogenic variants in the cone-specific *PDE6* genes result in achromatopsia (Bhardwaj et al., 2022). Up to 4% of all ARRP cases in North America have been linked to the *PDE6A*, whereas *PDE6A*-related RP cases are rare in Spain, Japan, and the UK (Hayashi et al., 2021). In this gene, 40 pathogenic variants have been identified, the majority of which were missense and nonsense variants (Khan et al., 2021). On the other hand, *PDE6B* gene mutations induce a severe form of RP with an early onset disease and account for up to 8% of all ARRP cases (Kim et al., 2020).

The retinitis pigmentosa 25 (*RP25*) gene is also known as the *EYS* gene as it encodes a protein with homology to a Drosophila protein called Eyes Shut, which is a regulator of eye morphology in insects (Sengillo et al., 2018). In the human eye it is essential for the development and structural integrity of photoreceptors (Seko et al., 2018). The majority of cases of RP in Japan are caused by mutations in this gene, which accounts for 51% of all RP patients (Numa et al., 2020). The prevalence of mutations in this gene varies based on the geographic location, which can range from 5% in the Netherlands and Canada to up to 16% in Europe (Garcia-Delgado et al., 2021). Approximately 449 variants, including point variants, deletions, insertions, splicing changes, and complex rearrangements, have been found to date in this gene (Bhardwaj et al., 2022). The *EYS* gene variations are linked to relatively severe and early onset ARRP.

Mutations in the *RPE65* (the human retinal pigment epithelium 65) gene are responsible for 0.6-6% of RP (mostly ARRP) cases and 3-16% of LCA cases. It encodes a protein that is involved in visual cycle and phototransduction (formation of light sensitive pigments). More than 300 mutations were found in this gene and most of them were single base mutations (Aoun et al., 2021). Recent improvements have been made in the

treatment of LCA, which is caused by biallelic null or loss of function mutations in the *RPE65* gene. Only cases with biallelic *RPE65* mutations are eligible for this treatment, which replaces the missing gene through subretinal injection of an adeno-associated viral (AAV) vector carrying a human *RPE65* gene. There are lots of other gene therapies in clinical trials for IRDs and some will be discussed later in the discussion chapter (Section 6.3), but to date only this gene therapy for retinal degeneration has received approval from the Food and Drug Administration (Voretigene neparvovec-rzyl, Luxturna™, SparkTherapeutics) (Russell et al., 2017, Food and Administration, 2017) and the European Medicines Agency (EMA) in 2018. In order to determine a patient's eligibility for gene therapy in *RPE65*-related IRDs, it is essential to understand the pathogenicity of rare novel *RPE65* variants, particularly the variants of uncertain significance. For instance, when a patient has a monoallelic pathogenic variant in *RPE65* and a clear clinical picture of RP or LCA, sequencing the entire gene is required to find (or exclude) variants in the second allele. The pathogenicity of such variants is worthy of further investigation by *in silico* pathogenicity prediction tools and *in vitro* functional studies to evaluate the effects of these variants on protein abundance, localisation, and function. Additionally, individuals with potential compound heterozygous variants must undergo haplotype phasing and segregation studies prior to treatment, since *RPE65* gene therapy will be ineffective if the two variants are in *cis* rather than in *trans* (Sodi et al., 2021, Aoun et al., 2021).

Gene symbol (OMIM)	Proportion of cases (%)	Function	Gene symbol (OMIM)	Proportion of cases (%)	Function
***ABCA4* (601691)**	5-6%	Transport molecules into/out of photoreceptors	***MAK* (154235)**	≤1%	Regulation of retinal cilium and spermatogenesis
***AGBL5* (615900)**	Rare	Posttranslational modification of tubulin	***MERTK* (604705)**	≤1%	Regulation of phagocytosis in photoreceptor OS
***AHR* (600253)**	Rare	A transcription factor involved in biological response to xeno- and phytobiotics	***MVK* (251170)**	≤1%	Isoprenoid pathway
***ARHGEF18* (616432)**	≤1%	Regulation of wide spectrum of cellular functions	***NEK2* (604047)**	≤1%	Ciliary protein involved in cell division
***ARL6* (608845)**	≤1%	Ciliary function	***NEUROD1* (601724)**	≤1%	Maintenance of photoreceptor
ARL2BP	≤1%	Trafficking of ciliary proteins and factors	***NR2E3* (604485)**	≤1%	Ligand-dependent transcription factor
***BBS1* (209900)**	≤1%	Ciliary trafficking	***NRL* (162080)**	≤1%	Transcription factor/ required for rod photoreceptor development
***BBS2* (606151)**	≤1%	Cilia formation and function	***PCARE* (*C2orf71*/613425)**	≤1%	Regulation of OS disks development
***BEST1* (607854)**	≤1%	Transmembrane oligomeric chloride channel	***PDE6A* (180071)**	3-4%	Phototransduction cascade
***C8orf37* (614477)**	≤1%	Maintenance of physiological levels of OS membrane protiens	***PDE6B* (180072)**	5-8%	Phototransduction cascade
***CERKL* (608381)**	>1% More prevalent in Spain	Regulation of mitochondrial biology and metabolism	***PDE6G* (180073)**	Unknown	Phototransduction cascade
***CLCC1* (617539)**	≤1%	Maintenance of retinal structure and function	***POMGNT1* (606822)**	≤1%	Glycosylation pathway
***CLRN1* (606397)**	≤1%	Photoreceptor synapses	***PRCD* (610598)**	≤1%	Unknown
***CNGA1* (123825)**	2-5% 7.6% in Chinese 5.1% in Japanese	Phototransduction and formation of rod cells OS	***PROM1* (604365)**	≤1%	Regulation of OS disk morphogenesis

***CNGB1* (600724)**	≤1%	Phototransduction and regulation of ion flow into rod cells OS	***PROS1* (176880)**	≤1%	Anti-coagulant action
***CRB1* (604210)**	1% of ARRP 17% of LCA cases in Spain	Retinal structure, cell-cell interaction, and cell polarity	***RBP3* (180209)**	≤1%	Visual cycle
***CWC27* (617170)**	≤1%	Unknown	***REEP6* (609346)**	≤1%	Endoplasmic reticulum function and transport
***CYP4V2* (608614)**	≤1%	Fatty acid and steroid metabolism	***RGR* (600342)**	≤1%	Visual cycle
***DHDDS* (608172)**	≤1%	Photoreceptor structure	***RHO* (180380)**	≤1%	Phototransduction/ scotopic vision
***DHX38* (605584)**	≤1%	Pre-RNA splicing	***RLBP1* (180090)**	≤1%	Visual cycle
***EMC1* (616846)**	≤1%	Unknown	***RP1* (180100)**	≤1%	Regulates the protein transport between IS and OS of photoreceptor cells, maintenance of cilia structure
***ENSA* (603061)**	≤1%	Regulation of cell cycle progression	***RP1L1* (608581)**	≤1%	Interact with photoreceptor connecting cilia
***EYS* (*RP25*/602772)**	0-15.9% 15.9% in Spain 51% of Japanese RP patients	Maintenance of photoreceptor cells integrity and facilitate ciliary trafficking	***RPE65* (180069)**	0.6-6%	Visual cycle
***FAM161A* (613596)**	≤1%	Unknown	***SAG* (181031)**	>1%	Prevent light-dependent degeneration of photoreceptor cells
***GPR125* (612303)**	Rare	Unknown	***SAMD11* (616765)**	≤1%	Signal transduction and regulation of transcription
***HGSNAT* (610453)**	≤1%	Lysosomal membrane enzyme	***SLC7A14* (615720)**	≤1%	Cationic transporter protein
***IDH3B* (604526)**	≤1%	Citric acid cycle (krebs cycle)	***SPATA7* (609868)**	≤1%	Maintenance of protein localization at the photoreceptor connecting cilium (CC)

***IFT140* (614620)**	≤1%	Development and maintenance of OS of photoreceptor cells	***TRNT1* (612907)**	≤1%	tRNA function and protein synthesis
***IFT172* (607386**	≤1%	Intra-flagellar transport	***TTC8* (608132)**	≤1%	Ciliary function
***IMPG2* (607056)**	≤1%	Maturation and maintenance of photoreceptor OS	***TULP1* (602280)**	≤1%	Protein transport in photoreceptor cells
***KIAA1549* (6133344)**	≤1%	Photoreceptor function	***USH2A* (276901)**	9-17%	Maintenance of periciliary membrane complex/ regulate intracellular protein transport
***KIZ* (615757)**	≤1%	Centrosomes stability	***ZNF408* (616454)**	≤1%	Unknown
***LRAT* (604863)**	≤1%	Vitamin A metabolism/ visual cycle	***ZNF513* (613598)**	≤1%	Unknown

Table 1.3: Summery of genes involved in autosomal recessive (ARRP) pathogenesis. Genes, OMIM ID, estimated proportion (%) of all ARRP cases, and function of encoded protein, for genes implicated in ARRP. (Sources: RetNet: https://sph.uth.edu/retnet/sum-dis.htm#A-genes; OMIM: https://omim.org; GeneCards: https://www.genecards.org; (Ferrari et al., 2011, Hartong et al., 2006, Fahim, 2018, Bhardwaj et al., 2022)).

1.6.1.2 Autosomal dominant retinitis pigmentosa (ADRP)

Autosomal dominant forms of RP are the second most common inherited forms of RP (30-40%) (Bhardwaj et al., 2022). Generally, ADRP is milder than X-linked or recessive RP and some cases were reported to have relatively mild symptoms starting after the age of 50 (Hamel, 2006), but there are exceptions to this, with severe disease in some patients was reported (Hollingsworth and Gross, 2013). The majority of people with ADRP have a parent who is affected, which indicates that there is a 50% chance that their offspring will inherit the mutation. Variation in penetrance is common in ADRP, particularly when *PRPF8, PRPF31*, *HK1*, and *RP1* mutations are implicated (Hamel, 2006, Maubaret et al., 2011, Audo et al., 2012, Villanueva et al., 2014, Verbakel et al., 2018). Phenotypic heterogeneity in ADRP has been reported even within the same family members (intrafamilial variability) (Maubaret et al., 2011). To date, mutations in 31 different genes (Table 1.4) are associated with ADRP (RetNet; https://sph.uth.edu/retnet/sum-dis.htm#A-genes) and the most prevalent gene is rhodopsin (*RHO*) accounting for 30-40% of all ADRP cases. In addition, mutations in genes *PRPF31*, *PRPH2*, and *RP1* were reported as major genes underlying this form of RP (Ferrari et al., 2011, Audo et al., 2012). These genes encode proteins that play vital roles in maintaining photoreceptor outer segment structure and function, cilia structure, phototransduction, and splicing.

The rhodopsin gene (*RHO*), also known as *RP4,* was the first gene in which mutations were found to cause RP (McWilliam et al., 1989, Dryja et al., 1990a, Dryja et al., 1990b). It is located on chromosome 3q22.1 spanning 6.7 kb of DNA and consists of five exons (Meng et al., 2020). It encodes the rhodopsin protein, which accounts for more than 90% of the protein composition of the rod outer segment discs (Bhardwaj et al., 2022). In photoreceptors, the rhodopsin protein serves as the main molecule mediating vision and is one of the most studied G-protein coupled receptor proteins (GPCR). Once it is activated by photon of light, visual phototransduction is initiated (Section 1.4.1). Rhodopsin provides scotopic (low light) vision (Lenahan et al., 2020). Mutations in the *RHO* gene are known to cause 30-40% of all ADRP cases (Ferrari et al., 2011, Sudharsan and Beltran, 2019). There are currently over 150 distinct mutations in the *RHO* gene known to cause both dominant and recessive types of RP, with the vast majority being dominant (Sudharsan and Beltran, 2019, Meng et al., 2020, Bhardwaj et al., 2022). The genetic penetrance of the *RHO* gene is nearly 100%. Patients with *RHO*-mediated ADRP have revealed variable phenotypic expression between families with different mutations (Jacobson et al., 1991). Moreover, patients with the Pro23His mutation had better visual

acuity (Meng et al., 2020), whereas those with mutations in the cytoplasmic domain of the gene (e.g., Ter349Glu) had a severe form of ADRP (Sullivan et al., 2006a) . Of note, because of a founder effect, the mutation Pro23His is the most common *RHO* mutation in North America accounting for about 10% of ADRP cases (Sudharsan and Beltran, 2019, Meng et al., 2020), but is absent from European patients (Farrar et al., 1990).

Additionally, *PRPF31* mutations are a frequent cause of ADRP (RP11), accounting for 5–10% of all ADRP cases (Rose and Bhattacharya, 2016). The gene pre-mRNA processing factor 31 (*PRPF31*) is the most common splicing factor gene implicated in ADRP. There are four other pre-mRNA splicing factors associated with ADRP: *PRPF3* found in 1.5% of cases, *PRPF8* found in 2-3% of cases, and *PRPF4* and *PRPF6*, both of which are rare. These genes encode proteins that are essential for splicing in all cell types. However, pathogenic mutations in these genes are more likely to have a deleterious impact on the survival and function of retinal cells (Yang et al., 2021a). It has been suggested that this is because of the impact of splicing factors on ciliary function, and several groups have established a link between splicing defects and ciliopathy phenotypes at the cellular level (Roosing et al., 2015a, Wheway et al., 2015, Buskin et al., 2018). Because splicing factors are crucial for retinal development and function, *PRPF31* is abundantly expressed in the retina. The genetic variations in *PRPF31* were initially described in 2001 (Vithana et al., 2001). More than 100 variants have been discovered in *PRPF31* and most of them are located in exons 6-10. Due to haploinsufficiency, ADRP forms caused by *PRPF31* mutations show incomplete penetrance (Kiser et al., 2019). The alternative splicing of several genes, including *RHO*, *ROM1*, *FSCN2*, and *GNAT1*, has been seen to be affected by PRPF31 protein reduction (Azizzadeh Pormehr et al., 2020). It has been reported that it has bimodal phenotypic expressivity, in which individuals with *PRPF31* mutations were either asymptomatic (non-penetrant carriers) with normal fundus appearance and minimal or no psychophysical or ERG abnormalities, or symptomatic with adolescent onset nyctalopia, restricted visual fields, and undetectable ERG responses by 30 years of age (Evans et al., 1995). This may be attributed to a second allelic genetic influence, in which the non-penetrant carriers inherited a wild-type allele that was distinct from the one inherited by their symptomatic siblings. This suggests that there may be wild-type alleles that are differentially expressed and can potentially influence the penetrance of disease symptoms (Vithana et al., 2003, Waseem et al., 2007). Two possible modifier loci, the *CNOT3* gene (a master regulator of global gene expression) and an *MSR1* repeat element in the *PRPF31* promoter, could affect the

disease penetrance in *PRPF31*-related retinopathy. In retinal organoids and RPE derived from both non-penetrant carriers and RP11 patients, *PRPF31* haploinsufficiency has caused *CNOT3* mis-splicing, which in turn lowered its expression and other splicing machinery. On the other hand, inheritance of a 4-copy *MSR1* repeat element has only been identified in non-penetrant carriers; as a result, it has been suggested to be associated with disease non-penetrance (McLenachan et al., 2021).

Mutations in *PRPH2* (peripherin 2) also account for around 5-10% of ADRP cases (Rose and Bhattacharya, 2016). The *PRPH2* gene, also called retinal degeneration slow (*RDS*) after the phenotype observed in mice, encodes a glycoprotein that is co-localised in the outer segment of rods and cones with its non-glycosylated homologue, rod outer segment membrane protein 1 (ROM1), where both regulate the formation and function of the discs (Conley et al., 2019). RP causing mutations in this gene were first observed in 1991 (Farrar et al., 1991). More than 175 mutations in this gene were found to cause different types of retinal dystrophies including, in addition to ADRP: autosomal dominant macular dystrophy (MD), dominant adult vitelliform macular dystrophy (AVMD), dominant cone rod dystrophy (CORD), recessive LCA and digenic RP caused by heterozygous mutations in both *PRPH2* and *ROM1* genes (Coco-Martin et al., 2020). Although mutations in the *PRPH2* gene are usually fully penetrant, phenotypic variability and incomplete penetrance have also been reported (Soucy et al., 2023).

Approximately 5–10% of ADRP cases (or 8–10% in the UK population) are caused by *RP1* mutations. The retinitis pigmentosa 1 gene (*RP1*) encodes a photoreceptor-specific microtubule-associated protein that is located on the axoneme and the connecting cilium of the photoreceptors (Riazuddin et al., 2005). Variable penetrance and expressivity were observed in patients with mutation in the *RP1* gene (Audo et al., 2012). About 185 mutations have been found (Bhardwaj et al., 2022), the majority of which are truncation mutations that have a dominant-negative effect and cause ADRP (Verbakel et al., 2019). The majority of pathogenic *RP1* mutations that cause ADRP cluster in exon 4 between amino acids 500 and 1053 (Siemiatkowska et al., 2012). In *RP1*-mediated ADRP cases, it should be noted that the p.Arg677* and p.Leu762Tyrfs*17 mutations were described as being the most frequently reported variants (Audo et al., 2012).

Gene symbol (OMIM)	Proportion of cases (%)	Protein function	Gene symbol (OMIM)	Proportion of cases (%)	Protein function
ADIPOR1 **(601691)**	Rare	Uptake and retention of docosahexaenoic acid (DHA)/ promotes photoreceptor cell survival	***PRPF6*** **(613979)**	Rare	Pre-mRNA splicing
ARL3 **(604695)**	Rare	Regulate trafficking of prenylated proteins and ciliogenesis in the rod outer segment	***PRPF8*** **(607300)**	2-3%	Pre-mRNA splicing
BEST1 **(607854)**	Rare	Transmembrane oligomeric chloride channel	***PRPF31*** **(606419)**	5-10%	Pre-mRNA splicing
CA4 **(114760)**	Rare	Transport of carbon dioxide and bicarbonate	***PRPH2*** **(179605)**	5-10%	Photoreceptor outer segment structure (discs) and function
CRX **(602225)**	1%	Transcription factor/ differentiation and maintenance of photoreceptor cells	***RDH12*** **(608830)**	Unknown	Visual cycle
FSCN2 **(607643)**	Rare (3.3% in Japan)	Photoreceptor-specific paralog of fascin which crosslinks and bundles f-actin	***RHO*** **(180380)**	30-40%	Phototransduction/ Scotopic vision
GUCA1B **(602275)**	Rare (5% in Japan)	Phototransduction recovery	***ROM1*** **(1807221)**	Rare	Cellular structure
HK1 **(142600)**	Rare	Catalyses phosphorylation of glucose to glucose-6-phosphate	***RP1*** **(180100)**	5-10% (8-10% in the UK population)	Regulates the protein transport between IS and OS of photoreceptor cells, maintenance of cilia structure
IMPDH1 **(146690)**	2-3% (5-10% in US and Europe)	Cell growth regulation	***RP9*** **(180104)**	Rare	Splicing
IMPG1 **(602870)**	Rare	A component of photoreceptor extracellular matrix	***RPE65*** **(180069)**	Rare	Visual cycle
KIF3B **(603754)**	Rare	Chromosome movement and microtubule activity	***SAG*** **(181031)**	Unknown	Prevent light-dependent degeneration of photoreceptor cells
KLHL7 **(611119)**	1-2%	Ubiquitin-proteosome pathway/ Protein degradation	***SEMA4A*** **(607292)**	3-4% in Pakistan	Transmembrane semaphorin/ enhance T-cell activation

***NR2E3* (604485)**	1-2%	Ligand-dependent transcription factor	***SNRNP200* (601664)**	1-2% (5.8% in Chinese)	Splicing
***NRL* (162080)**	1%	Interacts with *CRX*/ Promotes transcription of rhodopsin and other retinal genes/ rod photoreceptor development	***SPP2* (602637)**	Rare	Unknown
***PRPF3* (607301)**	1.5%	Pre-mRNA splicing	***TOPORS* (609507)**	1-2%	Localised in basal body of connecting cilia in photoreceptors
***PRPF4* (607795)**	Rare	Pre-mRNA splicing			

Table 1.4: Summery of genes involved in autosomal dominant (ADRP) pathogenesis. Genes, OMIM ID, estimated proportion (%) in all ADRP cases, and protein function of genes implicated in ADRP. (Sources: RetNet: https://sph.uth.edu/retnet/sum-dis.htm#A-genes; OMIM: https://omim.org; GeneCards: https://www.genecards.org; (Ferrari et al., 2011, Hartong et al., 2006, Fahim, 2018, Bhardwaj et al., 2022, Wada et al., 2001, Rice et al., 2015, Daiger et al., 2015, Fahim et al., 2020a)).

1.6.1.3 X-linked retinitis pigmentosa (XLRP)

X-linked retinitis pigmentosa, a rare inherited retinal degeneration that manifests as an aggressive form of retinitis pigmentosa, starts early, progresses rapidly, and eventually leads to registered blindness in most cases. XLRP accounts for up to 20% of all RP cases (Breuer et al., 2002, Churchill et al., 2013, Tuupanen et al., 2022) and since it is caused by mutations in genes on the X chromosome, it mainly affects males, and is never passed by male-to-male transmission from one generation to the next. Female carriers have a 50% chance of passing the disease-causing variant to their sons, while their daughters will have 50% chance of being carriers. In affected males, symptoms first appear in childhood and include nyctalopia, a constriction of the peripheral visual field, central vision loss, and total blindness by the fourth decade of life (Nguyen et al., 2020). However, due to the preferential inactivation of the normal X chromosome (Fahim and Daiger, 2016), female carriers may also be affected, though their symptoms will be noticeably milder than those of male carriers (Chivers et al., 2021). XLRP can be mistakenly diagnosed as an autosomal dominant RP since a relatively severe phenotype can also occur in female carriers as mentioned above due to skewed X chromosome inactivation (Nanda et al., 2018). This demonstrates the importance of screening for X-linked genes in families with provisional autosomal dominant inheritance with no male-to-male transmission (Churchill et al., 2013). The majority of XLRP cases are caused by mutations in the retinitis pigmentosa GTPase regulator gene (*RPGR*; OMIM 312610) accounting for over 70% of XLRP-affected families (Li et al., 2016a, Nanda et al., 2018, Martinez-Fernandez De La Camara et al., 2018, Nguyen et al., 2020) , and the retinitis pigmentosa 2 gene (7-18%) (*RP2*; OMIM 312600) (Lyraki et al., 2018). Mutations in the oral-facial-digital 1 gene (*OFD1*; OMIM 311200) were identified as a rare cause of XLRP (Webb et al., 2012).

About 70% of all XLRP cases and 12-15% of simplex RP cases are caused by pathogenic variants in *RPGR* (Breuer et al., 2002, Churchill et al., 2013, Moreno-Leon et al., 2020). Additionally, CORD, MD, and extraocular disorders including sperm abnormalities, hearing loss, and primary cilia dyskinesia have been linked to *RPGR* mutations in some patients (Ayyagari et al., 2002, Iannaccone et al., 2003, Bukowy-Bieryłło et al., 2013). The *RPGR* gene was first cloned in 1996 (Meindl et al., 1996). It is located in chromosomal region Xp11.4, spanning 172 kb (Patnaik et al., 2015). Due to alternative splicing or post-translational modifications, it generates multiple transcripts, and two major transcripts are well distinguished: $RPGR^{ex1\text{-}19}$ and $RPGR^{ORF15}$ (Megaw et al., 2015).

The N-terminus (exon 2-11) of both transcripts encodes a structure similar to regulator of chromosomal condensation 1 (RCC1-like domain). The constitutive transcript *RPGR*$^{ex1-19}$ is widely expressed in many organs and can be found at the transition zone of primary and motile cilia (Vervoort et al., 2000, Hong et al., 2003, Megaw et al., 2015). *RPGR*$^{ex1-19}$ consists of 19 exons and encodes a protein of 815 amino acids. The *RPGR*^{ORF15} transcript is predominantly expressed in the retina and consists of exons 1 to 14 of *RPGR*$^{ex1-19}$ followed by a unique 3' terminus called open reading frame 15 (ORF15) (Vervoort et al., 2000, Moreno-Leon et al., 2021). It encodes an 1152 amino acid protein that localises within the photoreceptor connecting cilium (CC), a microtubule-based organelle analogous to the transition zone of primary (sensory) cilia that connects between the inner biosynthetic segment and the outer light-sensing segment (Hong et al., 2003). The ORF15 (c.1754-3459) includes exon 15 and a portion of intron 15 and encodes 567 amino acids with a highly repetitive purine-rich C terminal domain. It contains a ~ 1kb core region that is prone to error and challenging to sequence (c.2184-3162) and represents a mutational hotspot (Chiang et al., 2018, Tuupanen et al., 2022). The ORF15 region is responsible for two-thirds of all XLRP disease-causing mutations, the majority of which are out-of-frame deletions that result in shortened proteins (Sharon et al., 2003, Megaw et al., 2015). To date, over 500 variants have been identified in *RPGR*, most of them are frameshift and nonsense while 10% of them are splice site variants (Bhardwaj et al., 2022).

Another gene implicated in XLRP is retinitis pigmentosa 2 (*RP2*; OMIM 300757). It encodes a 350 amino acid protein (Schwahn et al., 1998) that has two domains: the tubulin folding cofactor C-like domain (TBCC domain) and the nucleoside diphosphate kinase-like domain (NDPK) (Breuer et al., 2002, Hurd et al., 2010). It is expressed at the basal body of the photoreceptor connecting cilium (CC) where it regulates protein transport from the photoreceptor inner segment to the outer segment (Evans et al., 2010). Biochemically, it also has a potential role in maintaining Golgi cohesion and directing proteins to the plasma membrane (Patil et al., 2011). Approximately 130 disease-causing mutations, including missense, splice site variations, nonsense, and insertion/deletion variants, have been described (Bhardwaj et al., 2022). Mutations in the *RP2* gene result in misfolding of the photoreceptor tubulin isoform, and disrupted protein trafficking machinery leading to photoreceptor degeneration and blindness (Patil et al., 2011). Patients with RP2 disease were shown to have severe symptoms, early disease onset, rapid disease progression, and early macular degeneration (Sharon et al., 2003, Jayasundera et al., 2010).

1.6.2 Cone and macular related disorders

There are two categories of inherited disorders leading to cone dysfunction: stationary and progressive. The stationary cone disorders, also known as the cone dysfunction syndromes, are congenital and exclusively affect cone photoreceptors. These syndromes include complete and incomplete achromatopsia (ACHM), blue cone monochromatism, oligocone trichromacy, bradyopsia, and Bornholm eye disease (Aboshiha et al., 2016). The progressive disorders, cone dystrophy (COD) and cone rod dystrophy (CORD), on the other hand, manifest as either a macular disease or a diffuse retinopathy with a predominance of macular involvement and later onset (in childhood or early adulthood) (Michaelides et al., 2006, Hamel, 2007). Macular dystrophies (MD) are diagnosed based on the fundus findings of bilateral, moderately symmetrical abnormalities involving the macula (Michaelides et al., 2003b). Symptoms of impaired central vision function are observed in patients with cone and macular related disorders. These conditions show genetic and clinical heterogeneity, and autosomal dominant, autosomal recessive, X-linked, and mitochondrial inheritance patterns have been described. They can also present as non-syndromic and syndromic forms.

Complete Achromatopsia (ACHM) is a rare autosomal recessive disorder that is characterised by loss of cone function and affects about 1 in 30,000 people. Patients with ACHM present with pendular nystagmus, poor visual acuity, lack of colour vision, and photophobia at birth or in the early stages of infancy (Thiadens et al., 2009, Michalakis et al., 2021). Patients may also demonstrate the so-called Flynn phenomenon, in which they experience a paradoxical constriction of their pupils when transitioning from light to darkness (Simon et al., 2004). Electroretinogram (ERG) shows absent cone responses with normal rod responses and fundus appearance is often normal. However, certain macular alterations, ranging from RPE abnormalities to atrophy, might be seen. A small proportion of ACHM patients have an incomplete version of the disease that is associated with supposedly improved colour vision and higher visual acuity, as well as some residual cone function on the ERG (Aboshiha et al., 2016). There are six genes associated with ACHM, including *CNGA3*, *CNGB3*, *ATF6*, *GNAT2*, *PDE6C*, and *PDE6H*. Mutations in genes *CNGA3* (OMIM 600053) and *CNGB3* (OMIM 605080) account for up to 90% of cases of ACHM (Michalakis et al., 2021). All ACHM genes are involved in phototransduction except for *ATF6*, which encodes a transmembrane transcription factor in the endoplasmic reticulum which plays a role in ER homeostasis (Yamamoto et al., 2007).

COD and CORD manifest as a diffuse retinopathy with predominant macular involvement. They are the most common hereditary progressive cone disorders with an incidence of 1:30,000 to 1:40,000 globally (Berger et al., 2010). Cone photoreceptor loss is a common feature of both diseases, which are clinically and genetically heterogeneous. In COD, the functional abnormalities are limited to the photopic system, whereas in CORD, rod dysfunction becomes widely involved as the disease progresses (Thiadens et al., 2012).

The mean onset of COD is at adolescence and patients with COD present with central vision loss, photophobia, and colour vision disturbance. Since the cone function is normal initially, nystagmus is frequently absent (Roosing et al., 2014). The course of the disease may vary, and some patients may become legally blind before 50 years of age (Thiadens et al., 2012). All three axes of colour vision are generally affected, and the hallmark for COD diagnosis is the reduced cone response in ERG with maintained rod response. Early nyctalopia caused by early rod involvement or simultaneous cone and rod loss on ERG distinguishes CORD from COD (Gill et al., 2019). Other clinical symptoms are similar to those of COD, but they start earlier and progress more rapidly, typically leading to legal blindness before the age of 40 (Thiadens et al., 2012). Macular abnormalities on fundoscopy for both COD and CORD can range from no abnormalities to bull's-eye maculopathy or RPE atrophy, and the optic nerve may exhibit variable degrees of temporal pallor. However, patients with CORD may also have peripheral pigment deposits and attenuated retinal vessels (Figure 1.8) (Roosing et al., 2014).

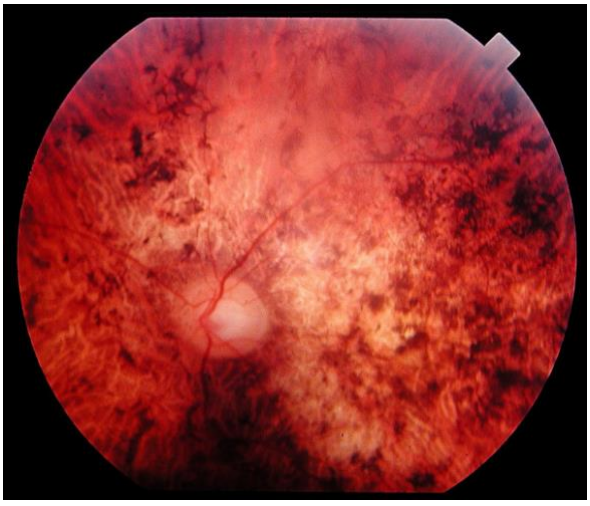


Figure 1.8: Colour image of fundus appearance in a CORD patient. A fundus examination of a patient with CORD reveals central retinal atrophy and irregularly shaped pigment deposits in the posterior pole. (Reproduced from (Hamel, 2007) with free license from BioMed Central).

Autosomal recessive, autosomal dominant, and X-linked inheritance patterns are all possible for COD and CORD. Currently, COD/CORD is known to be caused by mutations in 37 genes (RetNet; https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes). The majority of cases are recessively inherited, of which about 60% are caused by variants in *ABCA4* (Section 1.6.1.1). Ten genes have variants that are linked to AD COD/CORD, with *GUCY2D, PRPH2*, *CRX* (Chapter 4), and *GUCA1A* accounting for more than 75% of these variants. Variants in four genes are linked to XL COD/CORD, with *RPGR* (Section 1.6.1.3) accounting for 73% of cases. The proteins encoded by these genes are involved in phototransduction, photoreceptor outer segment morphogenesis, intraflagellar transport, and neurotransmitter release (Gill et al., 2019).

Macular dystrophies (MDs) are a subset of inherited retinal disorders result in progressive macular atrophy and severe visual impairment. Although the age of onset varies, the majority of cases appear in the first two decades of life. Clinical presentations include reduced central vision function with colour vision abnormalities and reading difficulties. Stargardt disease (STGD), Best disease (BD), X-linked retinoschisis (XLRS), autosomal dominant drusen (ADD), adult vitelliform dystrophy (AVMD), autosomal dominant "bull's-eye" macular dystrophy, and pattern dystrophy (PD) are the most prevalent subtypes of MDs. Notably, genetics also plays a role in the aetiology of age-related macular degeneration (ARMD) and about 20% of the patients have positive family history (Michaelides et al., 2003b, Rahman et al., 2020). To date, 21 genes have been linked to MD (RetNet: https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes).

Stargardt disease (STGD; OMIM 248200) is the most prevalent inherited juvenile MD, affecting 1:8000 to 1:10,000 individuals globally (Roosing et al., 2014). Patients frequently retain peripheral vision but exhibit impaired central vision, with decreased visual acuity, aberrant colour vision, photophobia, and sluggish dark adaption. Retinal flecks made up of lipofuscin pigment occur in the macula and macular atrophy and can be seen on fundus autofluorescence (FAF) imaging (Michaelides et al., 2003b). Stargardt disease is mainly an autosomal recessive disease, though rare dominant forms do exist. It exhibits a varied phenotype as well as age at onset and severity (Cremers et al., 2020).

In the predominant autosomal recessive disorder (STGD1), disease-causing mutations are found in the gene encoding the ATP Binding Cassette Subfamily A Member 4 (*ABCA4*) (Section 1.6.1.1). ABCA4 acts as a retinoid transporter, clearing potentially toxic all-trans retinal (ATR) from photoreceptors after photoexcitation. Defective retinoid transport caused by *ABCA4* mutations results in a build-up of all-*trans* retinal, which in turn leads

to formation of lipofuscin and its toxic by-products (A2E) in RPE and photoreceptors, ultimately causing cell dysfunction and death (Dean, 2005). It has been proposed that the severity of the disease depends on its onset; adult-onset (foveal sparing STGD) is linked to missense changes, whereas childhood-onset is linked to more deleterious variants.

Two rare autosomal dominant MDs, STGD3 and STGD4, caused by *ELOVL4* and *PROM1*, respectively, have symptoms that overlap with some of the STGD1 phenotypical manifestations (including bull's eye maculopathy) (Rahman et al., 2020).

1.6.3 Leber congenital amaurosis (LCA)

Leber congenital amaurosis (LCA; OMIM 204000), is the earliest and most severe type of inherited retinal disease and is characterised by blindness and severe visual impairment from birth. Dr. Theodore Leber first identified it in 1869, describing it as visual impairment in infants with nystagmus and a defective pupillary light reflex (Leber, 1869, Perrault et al., 1999). It accounts for at least 5% of all retinal dystrophies and about 20% of childhood blindness, with a population frequency of between 1/30,000 (Koenekoop, 2004) and 1/81,000 (Stone, 2007).

As early as six weeks of age, patients with LCA start showing symptoms. LCA is characterised by early and severe vision loss and nystagmus. Visual acuity ranges from 20/200 to light perception only or even non-light perception, and patients with LCA rarely have visual acuity better than 20/400 (Cremers et al., 2002). Patients may present with refractive errors as hyperopia in *GUCY2D* mutations. Other ocular features such as photophobia, nyctalopia, keratoconus, and juvenile cataracts may also present in LCA patients. Franceschetti's oculodigital sign can be a very important feature of LCA. Fundus appearance can vary from apparently normal to maculopathy, with or without bull's eye pattern, macular coloboma, or macular atrophy. The fundus of some LCA patients has also been reported to show typical RP-like abnormalities, including bone spicule peripheral pigmentation, vascular attenuation, pseudopapillary oedema, or coats-like vasculopathy. Amaurotic pupils and a severely reduced to non-detectable electrical signal on ERG are also typical features for LCA diagnosis (den Hollander et al., 2008, Huang et al., 2021).

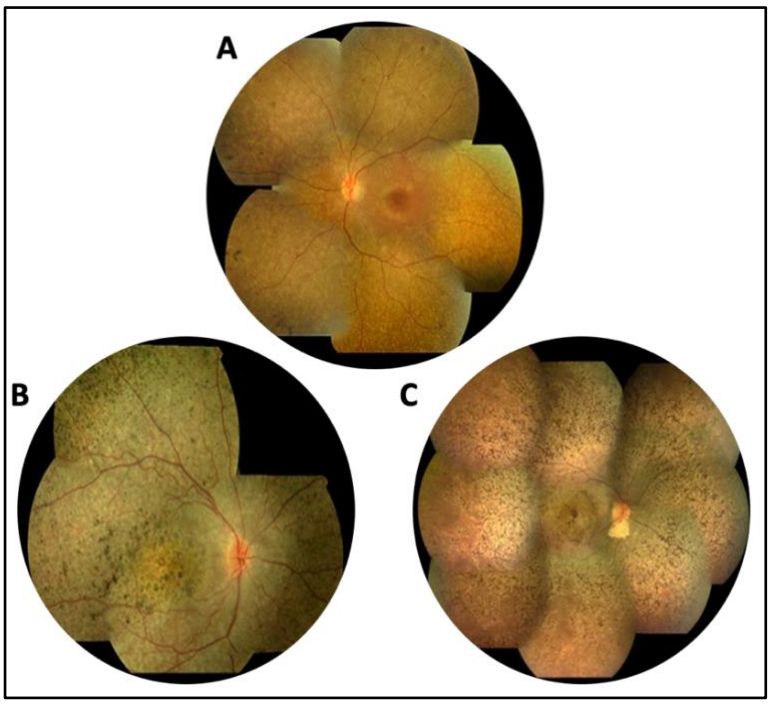


Figure 1.9: Colour images of fundus appearance in an LCA patient. A. Blonde fundus with peripheral white flecked lesions and a normal macula in *RPE65* - associated LCA. **B.** *CRB1*-associated LCA is characterised by nummular pigmentation, periarteriolar sparing of RPE and macular atrophy. **C.** Dense intraretinal pigmentation can be seen along with macular atrophy in *RDH12* - associated LCA. (Adapted from (Kumaran et al., 2017) with free licence from BMJ publishing group).

LCA is a heterogenous disease and is caused by mutations in 25 different genes (RetNet; https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes). The majority of cases are inherited in an autosomal recessive pattern. However, a small number of dominant cases caused by *CRX*, *IMPDH1*, and *OTX1* mutations have been documented (Kondkar and Abu-Amero, 2019). Some of these genotypes are linked to specific phenotypic features (Figure 1.9). The most prevalent causes of LCA are mutations in *GUCY2D* (6-21%), *RPE65* (4-16%), *CRB1* (9-17%), *CEP290* (15-30%), and *RDH12* (3-10%). LCA genes encode proteins that are essential for phototransduction, the visual cycle, ciliary transportation, photoreceptor structure and differentiation, and guanine synthesis (Huang et al., 2021).

1.6.4 Familial exudative vitreoretinopathy (FEVR)

Familial exudative vitreoretinopathy (FEVR) is a rare inherited angiogenesis disorder that affects the development of retinal blood vessels. It was first described in 1969 (Criswick and Schepens, 1969) and is characterised by incomplete peripheral retinal vascularization and poor vascular differentiation. Clinical and genetic variability in FEVR causes wide

variations in the disease manifestation and severity even within the same family. An avascular peripheral retina, which might be asymptomatic in the early stages, is the key feature of FEVR. Patients may experience secondary neovascularization in advanced stages, which can result in fibrosis, vitreoretinal tension, and retinal detachment (RD), leading to registered blindness. However, in some cases the disease never progresses beyond the avascular peripheral retina stage, and therefore carriers remain asymptomatic throughout their lives, but can still pass on the condition to their children (Gilmour, 2015).

Autosomal dominant, autosomal recessive, and X-linked inheritance patterns have been reported (Gilmour, 2015, Tauqeer and Yonekawa, 2018). The majority of cases of FEVR are caused by mutations in the genes encoding components of the Norrin-β-catenin signalling pathway; *FZD4*, *NDP*, *LRP5*, *TSPAN12* and *CTNNB1* (Chen et al., 1993, Toomes et al., 2004, Poulter et al., 2010, Collin et al., 2013, Drenser, 2016, Panagiotou et al., 2017). These genes produce proteins that play a role in eye organogenesis and angiogenesis by participating in both the canonical Wnt signalling pathway (β-catenin) and the Norrin signalling pathway (Tauqeer and Yonekawa, 2018). Other genes have also been implicated in rare cases. For example, mutations in *KIF11* were found to cause a FEVR-like syndrome called microcephaly with or without chorioretinopathy, lymphedema, or mental retardation (MCLMR) (Li et al., 2016b, Robitaille et al., 2014), while variants in *ZNF408* (Collin et al., 2013), *ATOH7* (Khan et al., 2012), *RCBTB1* (Wu et al., 2016), *JAG1* (Zhang et al., 2020), and *ILK* (Park et al., 2019) have also been shown to cause FEVR.

1.6.5 Optic atrophy (OA)

Optic atrophy (OA) is a condition that results in degeneration of retinal ganglion cells and their axons. Inherited optic neuropathies can be isolated, like Leber's hereditary optic neuropathy (LHON) and autosomal dominant optic atrophy (ADOA), or syndromic, like Behr's syndrome. The most frequent non-syndromic hereditary optic neuropathy is ADOA, with an estimated prevalence of 1:12,000 to 1:50,000 (Lenaers et al., 2012). Clinically, it manifests as progressive loss of visual fields and colour vision, loss of retinal ganglion cells and temporal pallor of the optic nerves. It is inherited in an autosomal dominant pattern and most commonly caused by mutations in the *OPA1* gene (Yu-Wai-Man and Chinnery, 2013). 20% of patients who have ADOA may also have additional systemic manifestations, which is known as ADOA+ (Yu-Wai-Man et al., 2010). Autosomal dominant OA has also been linked to mutations in the *AFG3L2*, *MFN2*, *MIEF1*, and *NR2F1* genes. Additionally, X-linked (*TIMM8A*), and autosomal recessive

inheritance patterns (*ACO2*, *NBAS*, *RTN4IP1*, and *TMEM126A*) are both possible for optic atrophy.

The *OPA1* gene (OMIM 605290) is localised on 3q28-q29 and encodes a mitochondrial dynamin-like GTPase family protein which is localised in the mitochondrial intermembrane space (IMS) (Han et al., 2022). It regulates the stability of mitochondrial network, mitochondrial bioenergetic output, and the accumulation of proapoptotic cytochrome c oxidase within the mitochondrial cristae spaces (Yu-Wai-Man et al., 2010). It is widely expressed, with varied levels of expression in various tissues. It contains five domains, three of which are highly conserved: the GTPase domain (exons 8-15), a dynamin central region (exons 16-24), and a C-terminal coiled coil region (GTPase effector domain (GED), exons 27-28) (Del Dotto et al., 2018). More than 500 disease causing variants in *OPA1* were reported in HGMD (accessed on August 2022) (Stenson et al., 2017), while the Leiden Open Variation Database (LOVD) for *OPA1* (https://www.lovd.nl/OPA1) listed 691 unique public variants. The mutation spectrum of the *OPA1* gene includes nonsense, missense, frameshift, and splice variants, small deletions and insertions and larger structural variants (Delettre et al., 2000, Delettre et al., 2001, Toomes et al., 2001a, Fuhrmann et al., 2009, Bonifert et al., 2014, Almind et al., 2011, Weisschuh et al., 2020a). Recessive mutations in *OPA1* result in Behr syndrome (BEHRS) (Schaaf et al., 2011, Bonneau et al., 2014) and mitochondrial DNA depletion syndrome type 14 (MTDPS-14) (Spiegel et al., 2016), while heterozygous mutations are linked to ADOA (57–89%) and ADOA+ (Delettre et al., 2000, Toomes et al., 2001a, Shimizu et al., 2003). Patients with Behr syndrome have bi-allelic variants in this gene, causing a rare, severe phenotype (Yu-Wai-Man et al., 2010, Schaaf et al., 2011, Bonneau et al., 2014, Carelli et al., 2015). There are no reports of patients with biallelic null alleles, and animal studies have indicated that such a genotype is fatal to embryos (Davies et al., 2007). The *OPA1* gene polymorphisms have also been associated with normal tension glaucoma (NTG) (Aung et al., 2002).

Solute carrier family 25, member 46 (*SLC25A46*; OMIM 610826) gene is located on chromosome 5 and composed of eight exons. It encodes a 418-amino acid mitochondrial solute carrier protein family member localised to the mitochondrial outer membrane (MOM). It is thought to have a transporter-like role across the inner mitochondrial membrane. It was found to be a reciprocal match to Ugo1 (Abrams et al., 2015), which is a modified mitochondrial solute carrier in the mitochondrial outer membrane in *Saccharomyces cerevisiae*, that functions as a mitochondrial fusion factor (Hoppins et al.,

2009). *SLC25A46* interacts with *OPA1*, *MFN2*, and mitochondrial contact site and cristae organising system (MICOS) complex to maintain mitochondrial morphology (Janer et al., 2016). Mutations in the *SLC25A46* gene were reported to cause recessive optic atrophy, axonal Charcot-Marie-Tooth disease (CMT), and cerebellar atrophy in four families including missense and indels mutations (Abrams et al., 2015). A homozygous splice site mutation was also reported in a patient with optic nerve and cerebellar atrophy (Nguyen et al., 2017). Furthermore, a homozygous missense mutation in *SLC25A46* was detected in a patient with Leigh syndrome, indicating that *SLC25A46* plays a role in the mitochondrial/endoplasmic reticulum pathway and is linked to neurodegenerative diseases with early onset (Janer et al., 2016).

Clearly many genes have been implicated to date in causing different forms of inherited retinal diseases providing important insights into disease causation, however, additional genes still remain to be elucidated.

1.7 Syndromic retinal diseases

The majority of IRD cases are non-syndromic, with the disease only affecting the eyes. However, over 80 syndromic IRD forms have been identified, with about 200 genes believed to be involved. Syndromic IRDs can be divided into two main disease categories, ciliopathies, and inborn errors of metabolism (IEMs). The foremost IEMs forms include congenital disorders of glycosylation, neuronal ceroid lipofuscinoses, mucopolysaccharidoses, and peroxisomal disorders, which frequently involve a retinal phenotype as part of a complex syndromic disorder. Ciliopathies are a group of genetic diseases resulting from mutations in genes involved in the primary cilia structure and function. The major forms that fall into this group are Usher syndrome (USH), Bardet-Biedl syndrome (BBS), Joubert syndrome (JBTS), Senior-Løken syndrome (SLN), and Alstrom syndrome (ALMS) (Tatour and Ben-Yosef, 2020). In the current study, BBS and USH syndromes are briefly overviewed.

1.7.1 Usher syndrome (USH)

Usher syndrome (USH; OMIM 276900) is clinically and genetically heterogenous and is the most common cause of syndromic hearing loss after Pendred syndrome (Castiglione et al., 2013). It accounts for 50% of individuals under 65 who are both deaf and blind, with a prevalence of 3.5-16.6 in 100,000 (Blanco-Kelly et al., 2015, Feenstra et al., 2022). It is a hereditary disorder characterised by sensorineural hearing loss, retinopathy (retinitis pigmentosa), and vestibular areflexia (Castiglione and Möller, 2022). Patients

with USH experience night blindness, then narrowing of the visual field or "tunnel vision" as a result of the degeneration of the rod cells. This can eventually lead to total blindness in some cases when the disease progresses to the cone cells (Reiners et al., 2005). USH is almost always inherited in an autosomal recessive pattern, and variants in genes involved in USH have also been reported to cause non-syndromic autosomal recessive RP (NS-ARRP) (McGee et al., 2010, Lenassi et al., 2015). It was revealed that USH genes such as *CDH23*, *USH2A*, *MYO7A*, *ADGRV1*, *PCDH15*, and *USH1C* can also cause dominant and recessive non-syndromic hearing loss without RP (Cesca et al., 2020). Clinically, it is categorised into three subtypes: Usher syndrome type 1 (USH1), Usher syndrome type 2 (USH2), and Usher syndrome type 3 (USH3) (Table 1.5).

Clinical subtype	Clinical features
Usher syndrome type 1 (USH1)	• Most severe form • Congenital, Severe to profound sensory hearing loss • Bilateral vestibular areflexia • Early onset of RP
Usher syndrome type 2 (USH2)	• Congenital, moderate-severe but relatively stable sensory hearing loss • Normal vestibular function • RP usually discovered in the second decade of life
Usher syndrome type 3 (USH3)	• Congenital, progressive sensory hearing loss • Progressive vestibular loss • Onset and severity of RP is highly variable

Table 1.5: Usher syndrome clinical subtypes. (Sources: (Fuster-García et al., 2021, Delmaghani and El-Amraoui, 2022, Castiglione and Möller, 2022))

USH can be categorized as a ciliopathy (Fuster-García et al., 2021) since it is caused by mutations in genes that encode proteins involved in different ciliary cell functions. It is one of several ciliopathies that impair non-motile or sensory cilia, including some forms of Bardet-Biedl syndrome, Alström syndrome, Meckel syndrome, Joubert syndrome, and Leber congenital amaurosis (LCA). USH has so far been linked to mutations in 17 different genes (Table 1.6), some of which encode proteins found in the retinal photoreceptor connecting cilium and hair cells in the inner ear (Castiglione and Möller, 2022).

Seven genes were identified for USH1, three for USH2, and two for USH3. Three atypical USH genes: *ABHD12*, *ARSG*, and *CEP78* and two modifier genes: *PDZD7* and *CEP250* were identified in association with USH. However, mutations in the *USH2A* gene are responsible for up to 80% of USH2 cases and over 50% of all USH cases (Toualbi et al., 2020, Feenstra et al., 2022). The penetrance in Usher syndrome is near 100% for all types (Castiglione and Möller, 2022).

Gene (OMIM)	Usher syndrome category	Protein function	Frequency
Usher type 1 (35-40%)			
***CDH23* (605516)**	USH1D and DFNB12	Structural protein: adhesion protein	10-20%
***CIB2* (605564)**	USH1J	Intracellular calcium homeostasis	Unknown
***ESPN* (606351)**	USH1M? and DFNB36	Actin-bundling protein	Unknown
***MYO7A* (276903)**	USH1B	Myosin/ structural component of cilia and microvilli	50-70%
***PCDH15* (605514)**	USH1F	Mediate calcium dependent cell-cell adhesion	5-10%
***USH1C* (605242)**	USH1C	Scaffold protein	6-20%
***USH1G* (607696)**	USH1G	Scaffold protein	0-5%
Usher type 2 (60-65%)			
***ADGVR1/GPR98* (602851)**	USH2C	G-protein coupled transmembrane receptor	5-20%
***WHRN* (607928)**	USH2D and DFNB31	PDZ scaffold protein	0-10%
***USH2A* (608400)**	USH2A	Maintenance of periciliary membrane complex/ regulate intracellular protein transport	50-80%
Usher type 3 (0-5%)			
***CLRN1* (606397)**	USH3A	Transmembrane protein	90-95%
***HARS1* (142810)**	USH3B	Protein coding gene	5-10%
Atypical USH			
***ABHD12* (613599)**	Atypical USH	Functional protein: enzyme involved in neurotransmission	Unknown
***ARSG* (618144)**	Atypical USH	Functional protein: enzyme (sulfatase)	Unknown
***CEP78* (617236)**	Atypical USH	Centriole-centriole cohesion during interphase of cell cycle	Unknown
Modifier genes			

***PDZD7* (612971)**	USH2?	Modifier gene	Unknown
***CEP250* (609689)**	Atypical USH?	Centriole-centriole cohesion during interphase of cell cycle	Unknown

Table 1.6: Summary of genes implicated in Usher syndrome. Gene involved in each USH subtype, protein function, and the estimated frequency. (Sources: RetNet: https://sph.uth.edu/retnet/sum-dis.htm#A-genes, (Fuster-García et al., 2021, Delmaghani and El-Amraoui, 2022, Castiglione and Möller, 2022))

1.7.2 Bardet-Biedl syndrome (BBS)

Bardet-Biedl syndrome (BBS, OMIM 209900) is a rare multisystem non-motile ciliopathy characterized by high genetic and clinical heterogeneity (Florea et al., 2021). Its estimated prevalence is 1:160,000 in Northern European populations (Forsythe and Beales, 2013). However, inbred, consanguineous, and isolated communities will often have a higher prevalence of BBS, such as Kuwait-Bedouin populations (1:17,000) and the Faroe Islands (1:3700) (Farag and Teebi, 1989, Hjortshøj et al., 2009, Ajmal et al., 2013, Khan et al., 2016).

It is a pleotropic disorder, with a number of features that include rod cone dystrophy, central obesity, postaxial polydactyly, cognitive impairment, hypogonadotropic hypogonadism and/or genitourinary abnormalities and kidney disease. Patients may also present with developmental delay, speech difficulties, brachydactyly or syndactyly, dental defects, ataxia, olfactory deficits, diabetes mellitus and congenital heart disease (Forsythe and Beales, 2013, Novas et al., 2015, Florea et al., 2021, Gupta et al., 2022). Before the genes for BBS were discovered, the disease was diagnosed based on its phenotypic manifestations (Forsythe and Beales, 2013, Beales et al., 1999). One of the most significant features of BBS is rod cone dystrophy, which affects 94–100% of affected individuals. Patients can experience night blindness, loss of peripheral vision, a reduction in colour perception, and a decrease in visual acuity.

Variable expressivity and pleiotropy of BBS are caused by a number of factors, including modifying variants, mutation types, interactions between BBS proteins, and interactions between BBS proteins and other non-disease-causing proteins (Florea et al., 2021). There are 26 genes in which mutations contribute to BBS (Table 1.7), each of which encodes a protein involved in the maintenance of primary cilia function. The BBSome multi-subunit complex is made up of eight BBS proteins: *BBS1*, *BBS2*, *BBS4*, *BBS5*, *BBS7*, *TTC8*, *BBS9*, and *BBIPI* (Gupta et al., 2022). The most frequently occurring mutations are in the *BBS1* and *BBS10* genes, accounting for 23% and 20% of BBS respectively. The formation

of BBSome complex is facilitated by three genes: *BBS6* (*MKKS*), *BBS10*, and *BBS12*, which produce chaperon-like BBS proteins. The BBS induced by mutations in these genes has more severe manifestations with earlier onset, higher prevalence of major diagnostic signs, and some signs of other ciliopathies, such as McKusick-Kaufman syndrome (MKKS). It should be noted that mutations in chaperonin genes are responsible for 30% of all BBS cases (Florea et al., 2021, Gupta et al., 2022).

Gene	Gene group	Gene	Gene group
ADIPOR1	Adiponectin receptor	***CEP19***	Centrosomal protein 19
BBS1	BBSome	***MKS1***	B9 domain containing MKS complex
BBS2	BBSome	***CEP290***	MKS complex
ARL6	ARF GTPase family	***WDPCP***	Ciliogenesis and planar polarity effector complex
BBS4	BBSome	***SDCCAG8***	microRNA protein coding host genes
BBS5	BBSome	***LZTFL1***	BBSome
BBS6/MKKS	Chaperonin-like protein	***BBIP1***	IFT-B1 complex RAB, member RAS oncogene GTPases
BBS7	BBSome	***IFT27***	IFT-B1 complex
TTC8	BBSome	***IFT74***	IFT-B1 complex
BBS9	BBSome	***C8orf37***	
BBS10	Chaperonin-like protein	***SCLT1***	
TRIM32	Tripartite motif containing ring finger proteins	***NPHP1***	NPHP complex
BBS12	Chaperonin-like protein	***SCAPER***	Zinc fingers C2H2-type

Table 1.7: Bardet-Biedl syndrome genes. (Sources: RetNet: https://sph.uth.edu/retnet/sum-dis.htm#A-genes, (Forsythe and Beales, 2013, Novas et al., 2015, Florea et al., 2021, Gupta et al., 2022).

1.8 Pathogenic variants identification through family studies

Many genes with disease-causing variants have been implicated in IRDs either as functional candidates or through family studies using candidate gene analysis, linkage analysis or autozygosity mapping.

1.8.1 Candidate gene analysis

In order to find pathogenic or risk variants linked to a particular disease, candidate gene studies have been at the forefront of genetic studies. Studies of candidate genes concentrate on those that have a known function that links them directly to a disease. In brief, the approach begins with the identification of a potential candidate gene that may be involved in the pathogenesis of the disease. This is followed by screening patients with the disease to identify variants in or close to those genes that may have a functional consequence, either by affecting gene regulation or altering the protein product. If variants are found in multiple cases and not in controls, this alone may constitute proof of their involvement in a Mendelian disease. Further proof may be obtained if variants are seen to segregate with the disease in a family. In multifactorial disease, implicating variants as risk factors in a disease would involve confirming the frequency of variants in the population, to determining whether there is a correlation between those variants and the phenotype using statistical methods (Tabor et al., 2002, Patnala et al., 2013). Such studies can be extended to interrogate genes agnostic to known functions using for example genome wide association studies (GWAS), a methodology frequently used for studies of multifactorial diseases including AMD, among other ocular disorders.

1.8.2 Linkage analysis

Genetic linkage is the tendency of DNA sequences that are adjacent to each other on a chromosome to be inherited together during the meiosis stage of sexual reproduction. The nearer two genes are on a chromosome, the lower the chance of recombination between them, and as a result, the two genetic markers are said to be linked and therefore more likely to be inherited together (Pulst, 1999). Sturtevant (Sturtevant, 1913) was the first to define a linkage map, which shows the linear positions of genes or genetic markers based on the frequencies of recombination between them during crossover of homologous chromosomes. Two genetic markers are thought to be farther apart the more frequently they recombine.

Human genetics researchers can use linkage maps to find new genes or genetic markers by testing for genetic linkage of the already identified markers to a disease or trait in a family. Genetic linkage analysis was formerly a widely used method for examining the segregation of traits in pedigrees, by determining whether specific phenotypes tended to be co-inherited with each other or with certain genetic markers, suggesting that the genes causing these traits are close to each other in the genome. Linkage analysis was frequently used to identify the chromosomal region most likely to contain the gene responsible for

a specific trait by evaluating the status of polymorphic markers. Polymorphic markers can be single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), or variable number tandem repeats (VNTRs). Linkage analysis may be either parametric or non-parametric, with parametric analysis being the type most commonly used for Mendelian trait analysis. Parametric linkage analysis relies on what we know about the phenotypes in the families and their segregation through a number of meiosis that are large enough to achieve statistically significant results. It makes assumptions about how traits are inherited and calculates a likelihood of linkage based on these assumptions. A statistical test called the LOD score (logarithm (base 10) of odds) can be used to investigate this. It measures the likelihood that the inheritance pattern in a particular pedigree in which the disease and the genetic marker are co-segregating is caused by linkage or by chance. Non-parametric linkage analysis (allele sharing), by comparison, comes to a result without assumptions about how the trait will behave in the family, and is more appropriate when, for example, there is variable penetrance or expression of a disease in a family.

Linkage analysis was used for the identification of loci in various types of inherited retinal diseases (Blanton et al., 1991, Shastry et al., 1995, Acland et al., 1998, Capeans et al., 1998). Linkage analysis requires the genotyping of a few markers to identify the specific chromosomal region shared between the affected individuals, but it cannot find the gene alone. This is followed by a more expensive targeted sequencing of each gene in the region. This required what was called positional cloning, before the availability of the human reference sequence (Pulst, 1999, Teare and Barrett, 2005, Teare and Santibañez Koref, 2014).

1.8.3 Autozygosity mapping

Autozygosity mapping, also called homozygosity mapping, is a method used to detect runs of homozygosity (ROH) in the DNA of individuals in whom recessively inherited disease have resulted from inheritance of a chromosomal region by descent from a common ancestor. In such patients the disease-causing mutation is frequently found in the ROHs in patients, especially if they come from consanguineous families (Quinodoz et al., 2021). Consanguinity is a social factor that increases the incidence of autosomal recessive disorders. In such families, both parents are likely to carry the same disease gene variant that was passed down from a common ancestor, meaning that the children are autozygous (homozygous inherited by descent) at the disease locus (Kannabiran et al., 2022). The closer the biological relationship between parents, the higher is the chance that their children will inherit one or both copies of the recessive gene (Hamamy, 2012).

1.9 DNA sequencing

Now that next-generation sequencing is available and widely affordable, it is increasingly possible to analyse a person's genome to find the disease-causing variants without performing a thorough family investigation. Large cohort studies like the UK 100,000 genome project have been made possible by this method (Turnbull et al., 2018). Nevertheless, performing segregation studies still helps in interpretation when it is available.

Exome or genome sequencing is now an important step toward the correlation of genotypes with phenotypic characters. Since Frederick Sanger first developed dideoxy sequencing in 1977 (Sanger et al., 1977b), DNA sequencing technologies have rapidly revolutionised molecular biology, medicine, genomics, and forensic sciences. DNA sequencing platforms have been developed over time, making sequencing a potent and practical research tool that can be employed easily in small labs. The advancement of sequencing can be divided into two stages. First generation sequencing includes Maxam-Gilbert and Sanger sequencing. More recently, to overcome some of the drawbacks of Sanger sequencing, several next generation sequencing (NGS) techniques have been developed. NGS methods can be further subdivided into second- and third-generation technologies.

1.9.1 First generation sequencing

Until about 25 years after the three-dimensional structure of DNA was discovered (Watson and Crick, 1953), early attempts to sequence DNA relied on analytical chemistry techniques that could only determine the composition of nucleotides, not their order (Holley et al., 1961). However, by combining these techniques with selective ribonuclease treatments, the first whole sequence of ribonucleic acid from *Saccharomyces cerevisiae* was produced (Holley et al., 1965). In the mid-1970s, the Sanger “plus and minus” system and the Maxam-Gilbert chemical cleavage approach were established. These two protocols were based on a single separation by polynucleotide length using electrophoresis (Sanger and Coulson, 1975, Maxam and Gilbert, 1977). Using the “plus and minus” method, the first DNA genome was sequenced from a bacteriophage (PhiX) (Sanger et al., 1977a).

The Maxam-Gilbert method involved cleaving a radiolabelled DNA chain at specific bases using chemical reagents and determining the length and location of the nucleotides by running the cleaved fragments on a polyacrylamide gel. Since it was the first technique

to gain wider implementation, it might be considered the actual beginning of first-generation DNA sequencing. However, this technique required the extensive use of hazardous chemicals, and as a result rapidly fell out of favour.

The development of Sanger's chain termination or dideoxy sequencing in 1977 (Sanger et al., 1977b), in contrast, represented a major advance that led to a step-change in the progress of DNA sequencing. This method uses chain-terminating dideoxy nucleotides (ddNTPs), which are chemical analogues of deoxynucleotides (dNTPs). The ddNTPs are unable to make bonds with the 5' phosphate of the next dNTP because they lack the 3' hydroxyl group that is necessary for the DNA chain extension. In the original protocol, the double stranded DNA was denatured into two single stranded DNA molecules (the templates), which were then coupled to a complementary primer to begin replicating in four separate polymerase reactions, each containing all four dNTPs and a small amount of one type of radiolabelled ddNTP. The products were then run in four lanes on a polyacrylamide gel and the positions of radioactive bands corresponded to the nucleotide sequence in the original template. Sanger sequencing underwent a number of improvements the following years, including the replacement of radiolabelling with fluorescence labelling, allowing researchers to perform sequencing in one reaction instead of four (Smith et al., 1985) and detection improvement by capillary-based electrophoresis (Swerdlow and Gesteland, 1990). Moreover, the innovation of the polymerase chain reaction (PCR) has also made it much easier to produce pure DNA at the high concentrations needed for sequencing. This facilitated the development of sequencing machines that became increasingly automated; the first one was established in 1991 (Hunkapiller et al., 1991). Sanger sequencing became the most popular technique for sequencing both small and large genomes, from bacteria and phages to humans. With a 99.99% base accuracy, it is the gold standard for confirming DNA sequences and can sequence a fragment of DNA commonly less than 1000 bp in length. It was crucial in helping researchers understanding the human genome landscape and enabled them to finish the first human genome sequence in 2004 (Consortium, 2004). However, the human genome project took over a decade to finish and cost roughly three billion dollars since it required enormous amounts of time and resources, demonstrating the obvious need for faster, higher throughput, and cheaper technology (NGS).

1.9.2 Next-generation sequencing (NGS)

Next-generation sequencing technology first appeared in 2004, offering a quick and affordable solution. In contrast to Sanger sequencing, NGS libraries are prepared in a cell free system. Instead of sequencing DNA molecules one by one, NGS facilitates millions of simultaneous parallel reactions, and the output is evaluated without the use of electrophoresis (Van Dijk et al., 2014). These improvements have enabled NGS technologies to generate a significant amount of data in less time and at a lower cost. The ability to generate high throughput data and sequence single-molecule DNA are the core elements of NGS (Hu et al., 2021). NGS can be further classified into second and third generation sequencing technologies.

1.9.2.1 Second generation sequencing (SGS)

The hallmark of second-generation "short-read" sequencing methods is the massive parallel sequencing of short (250–800 bp), clonally amplified DNA molecules (Tucker et al., 2009). Sequencing and clonal amplification are the two fundamental steps in short read sequencing techniques. A solid phase is necessary to amplify DNA fragments. This solid phase can be beads, as used in 454 pyrosequencing (Roche 454), ABI SOLiD technology, and Ion Torrent sequencing, or flow cell surfaces, as used in the Illumina sequencing by synthesis approach. Depending on the sequencing platform, either emulsion PCR, as in Roche 454 and Ion Torrent, or bridging PCR, as in Illumina, is used to amplify the anchored DNA fragments. The sequencing principle can be based on a "sequencing by synthesis" (SBS) method, as used by Illumina, Ion Torrent, and Roche 454, or on a "ligation mediated synthesis" method, as used by ABI SOLiD (Mardis, 2013, Van Dijk et al., 2014, Levy and Myers, 2016).

The first SGS technology was based on pyrosequencing (Nyrén and Lundin, 1985) and released by 454 Life Sciences (now Roche) (Margulies et al., 2005). In short, it involves generating light from phosphates by using an enzymatic cascade to yield read length of 400 bases and more than one million reads per run (Verma et al., 2017). It should be noted that Roche 454 was discontinued in mid-2016 (Levy and Myers, 2016). Following the success of Roche 454, a number of parallel sequencing methods emerged, with Illumina emerging as the most significant (Heather and Chain, 2016).

In this project, whole exome sequencing (WES) was carried out using Illumina NGS technology (Chapter 3). This technique is based on the "sequencing by synthesis" (SBS) method, which involves adding nucleotides to the extended DNA chain using DNA polymerase. After the genomic DNA has been fragmented, adapters are ligated at both

ends. The adapter tagged DNA molecules are then loaded onto a flow cell containing a lawn of complementary oligonucleotides to start solid-phase PCR. The other free end of the bound ssDNA bends over to prime from the next nearby complementary oligonucleotide in a process called ‘bridge’ amplification or bridging PCR. In each flow cell channel, repeated solid-phase PCR cycles create millions of dense ‘clusters’ of dsDNA. These clusters, each containing a single DNA molecule, are then sequenced by cycle reversible termination using fluorescently labelled nucleotides and reversible terminators. The amplified fragments are denatured followed by hybridizing a sequencing primer to a universal sequence in the amplicons. In each cycle, synthesis starts by adding four fluorescently labelled nucleotides together with DNA polymerase to the flow cell channel. The 3’ hydroxyl (OH) end is occupied by a fluorophore to block further addition. The resultant fluorescent signal can be imaged to determine the incorporated nucleotide (Figure 1.10). Deblocking the 3' OH end by enzymatic removal of the fluorescent moieties enables the addition of the next labelled nucleotide. This process repeats multiple times, and the sequence of each cluster is visualised as the wavelength (colour) of the fluorescence after each cycle (Zhou et al., 2010, Myllykangas et al., 2012, Heather and Chain, 2016, Verma et al., 2017, Hu et al., 2021).

Due to their capacity to generate highly accurate data at a reasonable price, short read sequencing technologies have come to dominate the market. Applying whole exome sequencing (WES), or targeted sequencing of gene or exon panels, has facilitated this achievement. Gene panels are custom-made panels utilizing SGS to sequence target genes and regions associated with various diseases. They are a quick and relatively inexpensive screening procedure but may prove costly if the gene involved is not in the panel, demanding the creation and testing of multiple gene panels. WES is an efficient and cost-effective “one size fits all” method that allow us to sequence only the coding regions, which are only about 1% of the human genome. It can be used to analyze the protein encoding regions (exons) where 85% of disease-causing mutations are believed to occur (Efthymiou et al., 2016, Schwarze et al., 2018, Barbitoff et al., 2020). The detection rate of causative mutations in IRDs using WES and targeted sequencing is low compared to WGS. Since WES only screens the exons of the genes, it is possible that disease-causing mutations are in non-coding regions, such as deep intronic variants, variants in regulatory regions, and variants in genes encoding non-coding RNAs, are missed. Technical issues with WES, such as biased read depth in high GC regions (Kieleczawa, 2006) and alignment challenges with highly repetitive regions (Ebbert et al., 2019, Watson et al., 2021), may also hinder the identification of the underlying genetic causes. Additionally,

mutations in novel genes, copy number variations (CNVs), and structural variants cannot be detected using targeted sequencing. Furthermore, some IRD cases may in fact be non-mendelian in their origin, perhaps because of digenic action, modifier genes, environmental, or other risk factors. For example, the sex of a patient is known to influence severity and age at onset for certain variants in the *ABCA4* gene in Stargardt disease (Runhart et al., 2020). Some of these issues can be overcome using WGS, which covers the entire genome, including both coding and noncoding regions and provides a greater detection of structural variations, non-coding variants, and variants in GC rich regions (PCR free WGS) compared to WES and targeted sequencing. Although it has more consistent and reliable coverage, WGS is generally not applied in everyday routine diagnostics currently, as it is still relatively expensive and more computationally demanding to store and analyze (Efthymiou et al., 2016). However, this is changing as the NHS plans to undertake 500,000 WGS for routine diagnostic care in 2023/2024. WGS can be used in a cost-effective manner by applying whole gene sequencing (GS), which enables capturing exonic, intronic, and regulatory regions for the target gene. GS was used to investigate 1054 unsolved Stargardt disease cases by sequencing the *ABCA4* gene using single-molecule molecular inversion probes (smMIPs). In 25% of the cases, this method allowed for the detection of pathogenic structural variations and deep intronic variants (Khan et al., 2020). Recently, NGS technologies have been globally used to increase the detection rate of IRD causing mutations. Regardless of the clinical diagnosis, up to two-thirds of cases with suspected IRDs were diagnosed using targeted sequencing (Watson et al., 2014, Consugar et al., 2015, Ellingford et al., 2016, Dockery et al., 2017, Hitti-Malin et al., 2022, Mc Clinton et al., 2023). When targeted sequencing fails to identify the underlying causal mutations, a broader approach employing WES or WGS is used (Riera et al., 2017, Carss et al., 2017, Wang et al., 2018, Zhang et al., 2018, Nash et al., 2022). Further discussion can be found in Section 6.2.1

However, as read length increases, sequencing accuracy decreases, which restricts these technologies to short sequence reads (Zhou et al., 2010, Myllykangas et al., 2012). Bias can arise in the amplification steps of short read sequencing, leading to uneven coverage. De novo assembly in the absence of a reference sequence, and phasing of haplotypes, are rarely possible using short-read sequencing. Assembling repetitive sequences can also be challenging, and some classes of structural variants are more difficult to detect due to the limited read length (Wenger et al., 2019, Hu et al., 2021). These limitations have led to the emergence of third generation sequencing technologies.

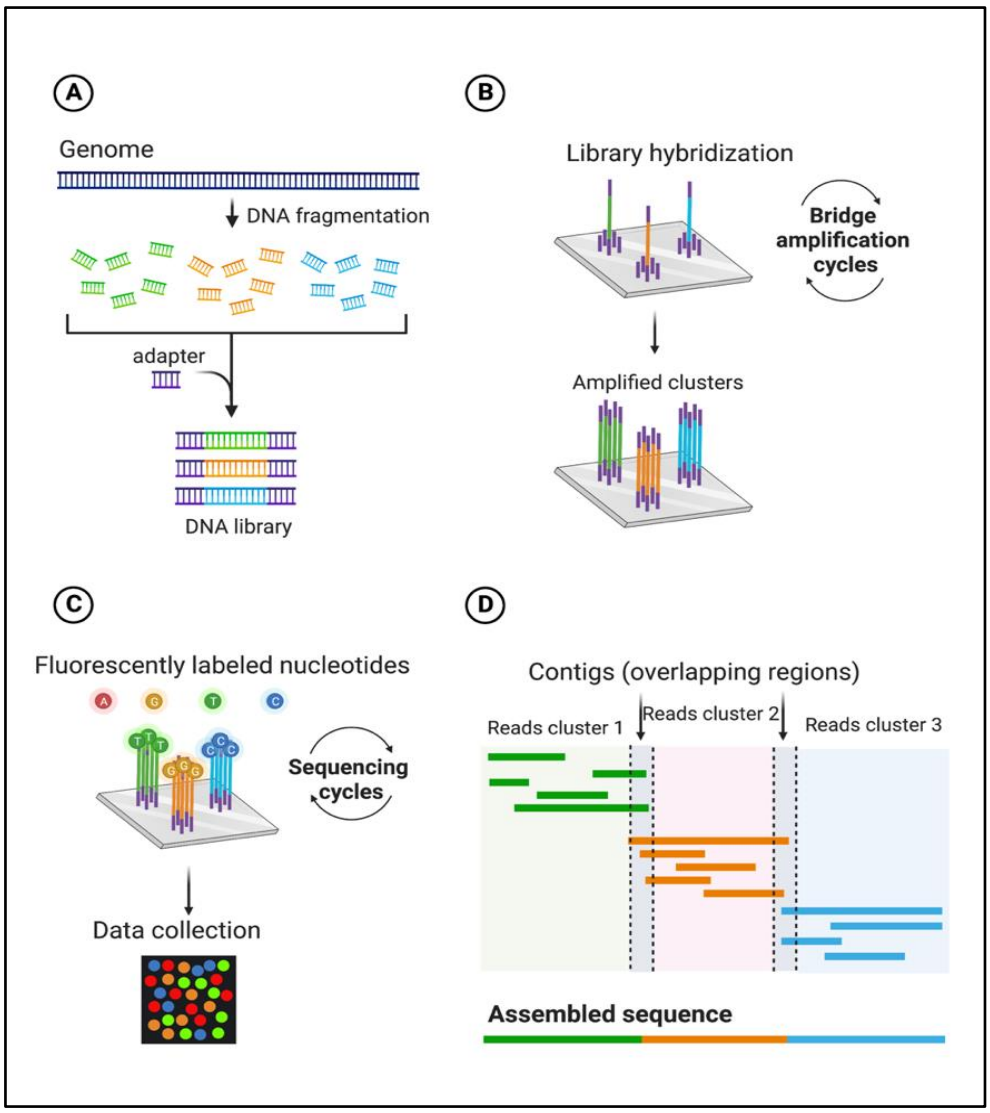


Figure 1.10: Second generation sequencing (the Illumina Sequencing by Synthesis method). **A.** Library preparation including DNA fragmentation and adapters ligation. **B.** DNA library bridge amplification. Multiple solid phase PCR cycles generate millions of amplified clusters. **C.** Cycle reversible termination sequencing uses fluorescently labelled nucleotides and reversible terminators, followed by data collection. **D.** Alignment and data analysis. (Created with BioRender.com, Accessed on 19/10/2022).

1.9.2.2 Third generation sequencing

Third generation sequencing, also called long read sequencing technology, was born out of the need for an approach that could generate long continuous reads (ranging from kilobases to >1Mb in length) directly from native DNA (Logsdon et al., 2020). It is a single molecular, amplification free (though the target molecules themselves may have been amplified), and real time sequencing approach. Two long-read technologies, from the companies Pacific Bioscience (PacBio) and Oxford Nanopore Technology (ONT), both of which were used in this study, are discussed briefly below.

The PacBio long-read platform utilizes single molecule real time technology (SMRT). PCR is not required for the DNA library preparation since a closed and circular single stranded DNA (ssDNA) template can be replicated automatically. This template is called SMRTbell template and is produced by ligation of hairpin adaptors to both ends of the dsDNA template. The sequencing reaction occurs in a SMRT cell chip that contains several tiny pores called zero mode waveguides (ZMW) (Figure 1.11A). In the SMRT cell, the DNA polymerase along with the SMRTbell template are immobilised at the bottom of ZMW to start replication. The fluorescently labelled nucleotide is incorporated into the target template by DNA polymerase during replication, and the light signal is recorded (Figure 1.11B). Only this nucleotide will produce detectable fluorescence before the dye is cleaved away and the signal is eliminated for that position. The light signals can be interpreted in real time, as nucleotide sequence, and the sequence produced via each ZMW is called a continuous long read (CLR) (Van Dijk et al., 2014, Heather and Chain, 2016, Xiao and Zhou, 2020). The PacBio sequencer can generate a read length of around 40 Kb but at 85% accuracy (Ambardar et al., 2016). The accuracy is improved by generating multiple subreads from a CLR, to give a circular consensus sequence (CCS) or HiFi. This improvement to the technology was achieved with the Sequel II system, generating highly accurate (up to 99.8% accuracy) long reads. HiFi sequencing generates shorter reads than traditional PacBio sequencing but reads are more accurate (Wenger et al., 2019, Logsdon et al., 2020).

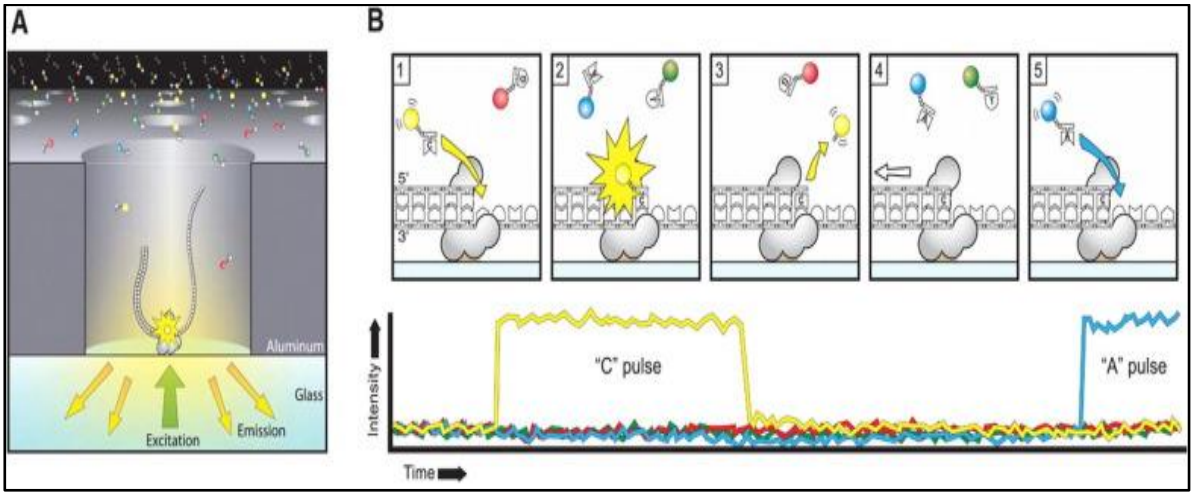


Figure 1.11: The long-read sequencing approach used by PacBio. A. Schematic representation of SMRT sequencing. The adaptor binds an immobilised polymerase at the bottom of a ZMW, which replicates an SMRTbell template, with the sequence captured as light excitation and emission. **B.** The SMRT sequencing process. When the fluorescently labelled nucleotide is incorporated into the target template by DNA polymerase, a light pulse is produced. (Reproduced from (Rhoads and Au, 2015) with permission from Elsevier).

Oxford Nanopore Technology (ONT) sequencing is another single molecule real time sequencing method based on a completely different technology. In this method, there is no need for PCR amplification or chemical labelling steps (Ambardar et al., 2016). The basis of ONT sequencing is the passage of a linear single stranded DNA molecule through a nanopore composed of staphylococcal α-hemolysin (αHL) protein (Clarke et al., 2009). The flow cell has a polymer membrane with hundreds to thousands of nanopores embedded in it. In order to facilitate strand capture and the loading of a processive enzyme at the 5'end of one strand, adapters are ligated to both ends of dsDNA prior to sequencing. This enzyme is utilised to guarantee millisecond-scale unidirectional single nucleotide displacement along the strand (Jain et al., 2016). The libraries are loaded along with an applied ion current onto a flow cell, and as each nucleotide passes through the pore, the sensor detects the disruptions in the ion current (Figure 1.12A). These disruptions are computationally interpreted using graphical models (Figure 1.12B) to give a sequencing readout (Jain et al., 2016, Logsdon et al., 2020, Hu et al., 2021).

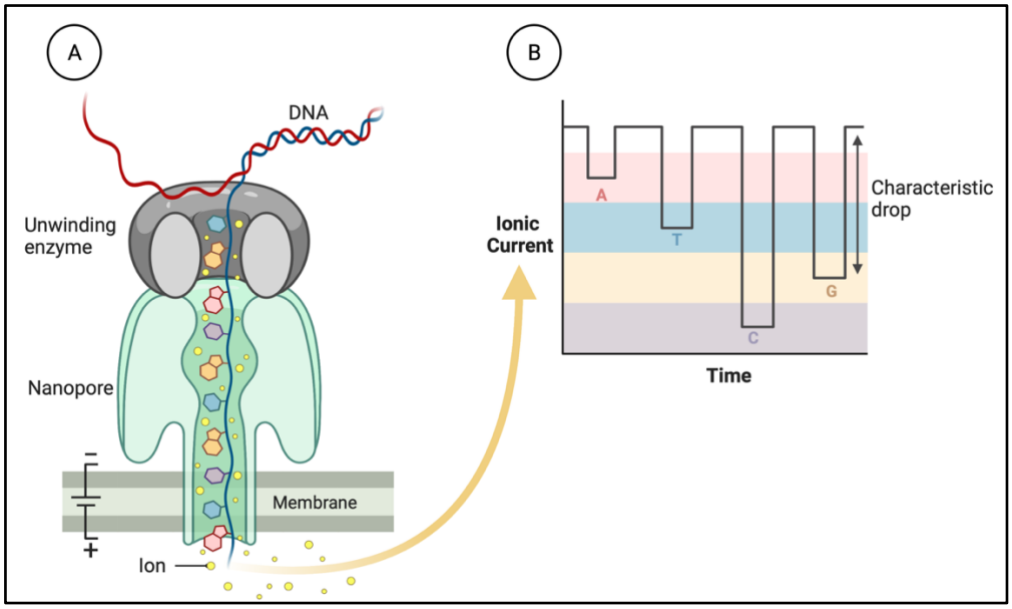


Figure 1.12: The sequencing principle used by ONT. A. A single stranded DNA molecule moves through the embedded nanopore under the pull of an ionic current. The obstruction of the nanopore by the ssDNA causes a change in the current that is measured by an electronic chip in the flow cell. **B.** The current variations are converted back into the original DNA sequence by a base-calling algorithm. (Created with BioRender.com, Accessed on 19/10/2022).

Long reads can be sequenced without interruption using ONT sequencing technology. However, it is limited by the need to prepare high molecular weight DNA (Hu et al., 2021). With reported accuracy of 87-98%, it can provide both long (10–100 Kb) and ultra-long reads (>100 Kb) (Logsdon et al., 2020). ONT platforms vary in terms of flow

cell capacity, output, and sequencing run time. Flongle flow cells are compatible with the standard MinION machine or the larger GridION and can be utilised for low throughput because they only have 126 nanopores and enable rapid sequencing at low cost. MinION is a portable, USB-sized device with one flow cell containing 512 channels. Each channel has four nanopores, giving a total of 2048 nanopores per flow cell. Up to five MinION-style flow cells are present in GridION. PromethION can accommodate up to 48 flow cells with 3000 channels and a total of 12,000 nanopores. While PromethION can produce 50–100 Gb of long read data, Flongle, MinION, and GridION can only produce 2–20 Gb of long read data per flow cell (Ambardar et al., 2016, Jain et al., 2016, Van Dijk et al., 2018, Logsdon et al., 2020, Hu et al., 2021). In this way, ONT technology allows users greater flexibility to undertake small, medium scale or much larger long-read sequencing projects and gives access to long-read technology on a relatively low budget.

1.10 Aims

The main aim of this project is to identify new variants involved in inherited retinal dystrophies, with a particular focus on those which have proved difficult to detect using current widely available technologies. Three major studies were carried out to accomplish this aim:

1) Analysing whole exome sequencing data from 24 previously unscreened individuals affected with different IRDs, to identify new pathogenic variants and potential causative genes.
2) Identifying new cases of a heterozygous *CRX* deletion causing late onset macular disease, characterising existing and new cases, and determining the mutational and pathobiological mechanisms underlying the disease phenotype in patients.
3) Developing a new approach to identifying new disease-causing variants in the *RPGR-ORF15* mutational hotspot using third generation sequencing (PacBio and ONT technologies).

Chapter 2
Materials and Methods

All primers were ordered from Sigma-Aldrich (Gillingham, Dorset, UK) and other reagents used in this study were purchased from Sigma-Aldrich or Thermofisher (Waltham, MA, USA), unless recommended otherwise.

The Room Temperature (RT) was in the range of 18-24°C.

All the methods described in this chapter were performed by the author except for:

1- DNA extraction from peripheral blood (Section 2.2.1), which was done by Leeds Genetics Laboratory, Leeds Teaching Hospitals NHS Trust, St James's University Hospital, Leeds, UK.
2- The WES library preparation (Section 2.10.1) was carried out in the Next Generation Sequencing facility (St James's University Hospital, Leeds, UK).
3- PacBio sequencing library preparation (Section 2.14.1.1) was performed in the Next Generation Sequencing facility (St James's University Hospital, Leeds, UK).
4- PacBio data analysis (Section 2.14.1.2) was undertaken post run by Dr Ian Carr (Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK).
5- Sanger sequencing for RPGR-ORF15 (Section 2.15) was performed at Manchester reference laboratory.

2.1 Patients

The affected individuals and their relatives were recruited prospectively at Ophthalmology clinics in the Eye Department, St James's University Hospital, Leeds, and in other UK and international centres collaborating with the Leeds Vision Research Group. Informed consent was obtained using a protocol which adhered to the tenets of the Declaration of Helsinki and was approved locally by the Leeds East Research Ethics committee (Project number 03/362) and by equivalent committees in other centres. The consent form obtained for historic samples referred only to screening of IRD genes. Samples recruited after 2017 signed a consent form which included a tick box for whether to feedback secondary/incidental findings. All participants were examined and investigated by experienced ophthalmologists or clinical geneticists to confirm their

clinical diagnosis, and detailed family history was taken. DNA Samples were extracted from either blood (Section 2.2.1) or saliva (Section 2.2.2) and samples from additional family members were taken when needed.

2.2 Genomic DNA extraction

Samples from many patients and relatives studied here had already been collected but, in some cases, patients and relatives were contacted for more samples to do further analysis.

2.2.1 DNA extraction from peripheral blood

DNA from peripheral blood was extracted by Leeds Genetics Laboratory, Leeds Teaching Hospitals NHS Trust, St James's University Hospital, Leeds, UK, using either a standard salt precipitation method or an automated DNA extractor such as the chemagic 360 or chemagic 360i, according to the manufacturer's protocols.

To extract DNA from fresh non-frozen blood using the salt precipitation method, 9ml of red blood cell lysis solution [155mM Ammonium Chloride (NH_4Cl), 10mM Potassium Bicarbonate ($KHCO_3$), and 1mM Ethylenediaminetetraacetic acid (EDTA) (pH 8.0)] was added to 3ml of whole blood in a polypropylene tube. The solution was mixed by shaking the tube for 10 minutes at room temperature (RT) and then centrifuged at 2000 *x g* for 10 minutes. The supernatant was removed and the white cell pellet containing DNA was resuspended in 3ml white cell lysis solution [25 mM EDTA (pH 8.0), 2% (v/v) Sodium dodecyl sulfate SDS] by pipetting. 1ml of 10M Ammonium Acetate (a protein precipitation solution) was added, and the mixture was vortexed for 20 seconds then centrifuged for 10 minutes at 2000 *x g*. The supernatant was transferred into a new tube and 3ml of isopropanol was added to precipitate the DNA. A centrifugation for 10 minutes at 2000 *x g* was also performed. The pellet was washed in 70% ethanol two times then left to air dry. 1x TE buffer (10 mM Tris HCl (pH 8.0) and 1 mM EDTA (pH 8.0)) was added to dissolve the precipitated DNA.

2.2.2 DNA extraction from saliva

Saliva samples were collected using the Oragene® DNA sample collection kit (DNA Genotek Inc., Kanata, Canada) and DNA extraction was done following the manufacturer's protocol. In short, the tubes were incubated at 50°C for a minimum of three hours. 500µl of the saliva/sampling mix was transferred to a 1.5ml Eppendorf tube, 20µl of PT-L2P was added and this was incubated on ice for 10 minutes. The tube was

then centrifuged at RT for 10 minutes at maximum speed (3500 *x g*). In a fresh tube, 500µl of supernatant was transferred and 600µl of 100% ethanol was added. The mixture was inverted 10 times and incubated for 10 minutes at RT. Another centrifugation was done at RT for 10 minutes at maximum speed. Supernatant was discarded and the pellet was washed in 500µl of 70% ethanol for one minute at RT. The tube was centrifuged for 5 minutes at maximum speed at RT. Ethanol was removed and the tube left open at RT for 30 minutes to ensure complete drying. After drying, 50µl of TE buffer was added to re-dissolve the pellet and stored at -20°C.

2.3 DNA quantification

The concentration of the extracted DNA was quantified using either a Qubit® Fluorometer (Invitrogen, Carlsbad, CA, USA) or Spectrophotometer (Nano Drop® 2000, ThermoFisher Scientific, Waltham, MA, USA) or both, according to the manufacturer's instructions.

2.3.1 DNA quantification using Qubit® Fluorometer

To measure the purified nucleic acid concentration using the Qubit® Fluorometer, an appropriate dsDNA broad range (BR) assay kit was used by adding 1-10µl of the sample to the Qubit dsDNA dilution buffer to make a final volume of 200µl. The samples were then vortexed and incubated at RT for 2 minutes. The samples were processed after calibrating the fluorometer using the standards provided by the assay kit.

2.3.2 DNA quantification using Nano Drop® 2000

DNA optical densities were measured using a spectrophotometer and 1µl of DNA samples. The spectrophotometer can be used to quantify and assess the purity of DNA, RNA, and proteins. It uses a photodetector and a wide spectral range of light stimulation from ultraviolet to visible light (190-840nm) to measure the sample's absorbance. The amount of the absorbed light can be used to identify the concentration of DNA or RNA in the solution by applying the Beer-Lambert law. This law draws a direct correlation between absorbance and concentration. DNA has a peak absorbance at 260nm while protein has a peak at 280nm. Purity ratio (A260/A280) was used to assess the quality of the DNA, where a ratio below 1.6 indicates high protein contamination while a value of 1.8 indicates low amounts of contamination.

2.4 Polymerase chain reaction (PCR)

Polymerase chain reaction is a molecular biology technique widely used to perform an *in vitro* cloning of DNA fragments. It is a rapid and sensitive method that can amplify even tiny amounts of DNA. To investigate a particular region of DNA, PCR was performed using specifically designed primers to amplify the region of interest including the sequences surrounding it. Initially, a complementary DNA strand is synthesised using these primers in 5'→3' direction and each newly synthesised strand will be used as a template for additional reactions to produce millions to billions of copies of DNA.

2.4.1 Primer design

To amplify products of various sizes, oligonucleotide primer pairs were designed using Primer3 program (http://bioinfo.ut.ee/primer3-0.4.0/). The genomic DNA (gDNA) sequence for the region of interest (ROI) was obtained from the NCBI (https://www.ncbi.nlm.mih.gov) and the UCSC genome browser, in FASTA format with exons in upper case and everything else in lower case. The primer pairs were designed to amplify the coding regions in addition to a minimum of 70 bp encompassing the 5' and 3' exonic regions. For optimal PCR, primer pairs were selected based on the length (range 18-24 bp), the melting temperature (range 57-62°C), and the GC content (<60%) according to the ROI. The annealing temperature was calculated using the following equation: melting temperature (T_m) = 2(A+T) + 4(G+C), where (A+T) denotes the total adenine and thymine residues in the primer sequence, and (G+C) denotes the total number of guanine and cytosine residues. The annealing temperature used was usually 5°C lower than the output of the above equation. The primer sequences were checked by running them through the BLAT tool in the UCSC genome browser (*in-silico* PCR) to ensure their specificity and to confirm that they will only anneal the ROI without binding to any other alternative areas in the genome. The dry primers were re-dissolved in sterile distilled water at 100μM and stored at -20°C. For working purposes, a 200μl solution of diluted primers (10μM) was made and stored at 4°C.

2.4.2 Routine PCRs: standard PCR

A 10µl reaction was prepared to amplify target DNA from 20-50ng gDNA by adding the following reagents at optimum concentration: 1µl of 10x PCR reaction buffer (200mM Tris-HCl (pH 8.4) and 500mM potassium chloride (KCl)) (final concentration 1x) (Invitrogen, Carlsbad, CA, USA), 0.5µl of 5mM dNTP (final concentration 0.25mM) (Invitrogen), 0.5µl of 10 µM forward and reverse primers each, 0.5µl (1 unit/ 1µl) *Taq* DNA polymerase (Invitrogen), and 0.3µl of 50mM $MgCl_2$ with a final concentration of 1.5mM. A negative control with no gDNA was included in each experiment. The thermal cycles were started with an initial denaturation of 95°C for two minutes followed by 30 cycles of denaturing at 95°C for 30 seconds, annealing at 55°C for 30 seconds, and extending at 72°C for 30 seconds. The final extension was carried out at 72°C for 10 minutes.

2.4.3 Modifications required to amplify purine rich/highly repetitive regions

To amplify purine rich regions, two types of DNA polymerases were used to ensure robust PCR performance. Including: Phusion® High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA; NEB) and LongAmp *Taq* DNA polymerase (NEB). Dimethyl sulphoxide (DMSO) was also used at 3% concentration in some instances.

2.4.3.1 Phusion high fidelity PCR for PacBio sequencing

Around 30-50ng gDNA was amplified using Phusion® High-Fidelity DNA polymerase kit (NEB). The reaction was performed by mixing 4µl of GC buffer (final concentration 1X) (NEB), 0.8µl of 5mM dNTPs (final concentration 200 µM) (Invitrogen), 1µl of each primer (final concentration 0.5µM each), 0.2µl Phusion High-Fidelity DNA polymerase (final concentration 0.4 units/20µl PCR reaction) (NEB), and dH_2O up to 20µl total volume. No additives were added. A positive control and a negative control with no gDNA were included for each pair of barcoded primers. The PCR reactions then were carried out by applying a touchdown PCR thermal cycle (Table 2.1).

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	98°C	1 minutes
4 cycles	Denaturation	98°C	10 seconds
	Annealing	67°C (Decrease by 0.5°C/ cycle)	30 seconds
	Extension	72°C	1 minute
26 cycles	Denaturation	98°C	10 seconds
	Annealing	65 °C	30 seconds
	Extension	72°C	1 minute
1	Final Extension	72°C	5 minutes
1	Hold	4°C	∞

Table 2.1:Thermocycling conditions applied in PacBio sequencing PCR.

2.4.3.2 Phusion high fidelity PCR for MinION sequencing (Pre-indexing)

For Nanopore MinION sequencing, amplicons were generated using the Phusion® High-Fidelity DNA Polymerase kit (NEB). A final volume of 20µl reaction mix was prepared by adding 20-50 ng of gDNA to a mixture containing 0.8µl of 5mM dNTPs (final concentration 200 µM) (Invitrogen), 0.2µl Phusion High-Fidelity DNA polymerase (final concentration 0.4 units/20µl PCR reaction) (NEB), 4µl of GC buffer (NEB) (final concentration 1X), 0.6µl DMSO (final concentration 3%), 1µl of each primer (final concentration 0.5µM each), and 11.4µl of dH_2O. A negative control with no gDNA was included in each experiment. This mixture was then placed in the thermal cycler under the following thermocycling conditions in Table 2.2.

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	98 °C	1 minute
30 cycles	Denaturation	98 °C	10 seconds
	Annealing	65 °C	30 seconds
	Extension	72°C	1 minute.
1	Final extension	72°C	5 minutes.
1	Hold	4 °C	∞

Table 2.2: Thermocycling conditions applied in MinION sequencing pre-indexing PCR.

2.4.3.3 Long range PCR: Long Amp™ *Taq* 2X Master Mix (Indexing)

100-200 fmol from each pre-indexed DNA sample (amplification product of Section 2.4.3.2) was transferred to a 96-well plate and its volume was adjusted to 24µl by adding nuclease free water. It was then mixed with 25µl Long Amp *Taq* 2X master mix (NEB) and 1µl of Oxford Nanopore Barcodes from PCR Barcoding Expansion 1-96 kit (EXP-PBC096) (Oxford nanopore technologies, Oxford, UK; ONT), one of BC1 to BC96 at 10µM (final concentration 0.2µM). Each sample has its unique barcode sequence to make a total volume of 50µl in each well. The mixture was spun down and amplified in the thermal cycler following the conditions in Table 2.3:

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	95 °C	3 minutes
30 cycles	Denaturation	95 °C	15 seconds
	Annealing	62 °C	15 seconds
	Extension	65°C	2 minutes.
1	Final extension	65°C	2 minutes.
1	Hold	4 °C	∞

Table 2.3: Thermocycling conditions applied in MinION sequencing indexing PCR.

2.4.4 Hot-shot master mix PCR

PCR reactions were performed by preparing a 10µl reaction volume to amplify 30-50ng of gDNA. Reaction mixes contained 50% hot shot diamond PCR master mix (Clent Life Science, Stourbridge, UK), 10 picomoles of each primer and dH_2O up to 10µl total volume. No other additives were added. A negative control with no gDNA was included in each experiment. The PCR thermal settings were then applied based on the amplicon size through the settings detailed in Table 2.4 and Table 2.5:

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	95 °C	2 minutes
40 cycles	Denaturation	95 °C	30 seconds
	Annealing	57°C -63 °C	30 seconds
	Extension	72°C	30 seconds
1	Final extension	72°C	3 minutes.
1	Hold	12 °C	∞

Table 2.4: Thermocycling conditions to confirm identified variants in WES samples (Chapter 3).

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	94 °C	10 minutes
40 cycles	Denaturation	94 °C	30 seconds
	Annealing	67 °C	30 seconds
	Extension	72°C	1 minute & 20 seconds
1	Final extension	72°C	4 minutes.
1	Hold	12 °C	∞

Table 2.5: Thermocycling conditions for the *CRX* deletion samples (Chapter 4).

2.4.5 Microsatellite marker genotyping

Microsatellite marker genotyping was carried out using hotshot diamond master mix PCR (Section 2.4.4), with the addition of a 5' -FAM (blue) fluorescent dye-labelled forward primer. Markers were identified and located using the UCSC genome browser (https://genome.ucsc.edu/). PCR was carried out using the thermocycling conditions in Table 2.6:

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	98°C	10 minutes
40 cycles	Denaturation	94°C	30 seconds
	Annealing	56°C	30 seconds
	Extension	72°C	30 seconds
1	Final extension	72°C	1 minute.
1	Hold	4°C	∞

Table 2.6: Thermocycling conditions applied in microsatellite genotyping.

Amplification then was confirmed by running 5µl of the reaction on a 2% agarose gel. Based on the strength of the band (band brightness on the gel compared to the marker), some products were used neat, and some were diluted. 1µl of either neat or diluted product was then mixed with 0.5µl ROX-500 size standard (Applied Biosystems, Waltham, Massachusetts, USA) and 8.5µl Hi-Di formamide (Applied Biosystems). The PCR fragments were size fractionated on a capillary DNA sequencing machine (ABI3130xI Genetic Analyzer; Applied Biosystem) and then the data were analysed using GeneMapper v.4 software (Applied Biosystems)

2.5 Agarose gel electrophoresis

DNA size fractionation was performed using agarose gel electrophoresis. Gels were made by dissolving between 1 and 3g (according to the band size) of molecular grade agarose powder (Bioline, London, UK) in 100ml 1 x Tris-acetate-EDTA (TAE) buffer (40 mM Tris HCl (pH 8.0), 1 mM EDTA (pH 8.0), and 19.4 mM Glacial acetic acid). The solution was heated in the microwave until it became clear, then cooled to (60-70˚C) and mixed with 3µl DNA advance stain Midori green (Nippon Genetics, Germany). This mixture was poured into a gel tray with gel combs inserted. Before electrophoresis, 1µl of 5x

DNA loading dye was added to 5µl of the PCR product. When the gel had set, samples were loaded inside the wells alongside either the molecular size standards EasyLadder I (Bioline, London, UK) or GeneRuler 1kb plus DNA Ladder (ThermoFisher Scientific, Waltham, MA, USA), depending on the product's size. The gel was then run at 110 V until obtaining the desired separation of the fragments. The visualization of DNA bands was carried out by a Bio-Rad molecular imager Gel Doc under the UV light. The resulting images were captured and analysed using ImageLab (v.4.0) software (Bio-Rad, Hemel Hempstead, UK).

2.6 DNA extraction from agarose gel

According to manufacturer's instructions, DNA bands were excised from agarose gels and purified using a MinElute gel extraction kit (Qiagen, Hilden, Germany). PCR products were run on a 2% gel and the bands were visualised using blue/green LED transilluminator (Nippon Genetics, Germany) and excised using a sterile scalpel. In a 1.5ml Eppendorf tube, three volumes (300µl) of buffer QG (supplied by manufacturer) were added to one volume of the excised gel containing the DNA fragment (100mg gel slice). The incubation was performed at 50°C for 10 minutes until the gel was completely dissolved in the buffer QG. After the gel slice had dissolved entirely, isopropanol was added in a proportion equal to gel volume (100µl of isopropanol for 100mg gel slice) and mixed properly. The mixture was placed in a MinElute column in a 2ml collection tube (provided in the kit) and then centrifuged for 1 minute at room temperature. The flow-through was discarded and the MinElute column was placed back in the collection tube. 500µl of buffer QG were added to the spin column which then centrifuged for 1 min and the flow-through was discarded once more. Multiple steps of purification were performed by adding 750µl of buffer PE (supplied by manufacturer) to the MinElute column and centrifuging for 1 min. The flow-through was discarded and the MinElute column was centrifuged for an additional 1 min at ≥ 10,000 *x g*. After removal of impurities, the MinElute column was placed into a clean 1.5ml microcentrifuge tube. To elute DNA, 10µl of buffer EB (10mM Tris.CL, pH 8.5) was added to the centre of the membrane and the tube was left to stand for 1 min before being centrifuged for an additional 1 min to collect the DNA. The collected pure DNA was then stored at -20°C until utilized.

2.7 Molecular cloning

2.7.1 TOPO TA DNA cloning

50µl PCR reactions for each sample were produced using a HotShot diamond master mix (Clent Life Science, Stourbridge, UK) following the steps in section 2.4.4. After extracting the correct band size from the gel (Section 2.6), the TOPO cloning reaction was performed using a pCR 2.1 TOPO vector (Figure 2.1) (Invitrogen, Carlsbad, CA, USA). For each sample, 2µl of the PCR product was added to 1µl salt solution (1.2 M NaCl and 0.06 M $MgCl_2$), 1 µl TOPO vector (supplied by manufacturer as linearized with single 3' thymidine (T) overhangs for TA cloning and topoisomerase I covalently bound to the vector), and 2µl nuclease free water. The reaction was then mixed gently and incubated at RT for 30 minutes.

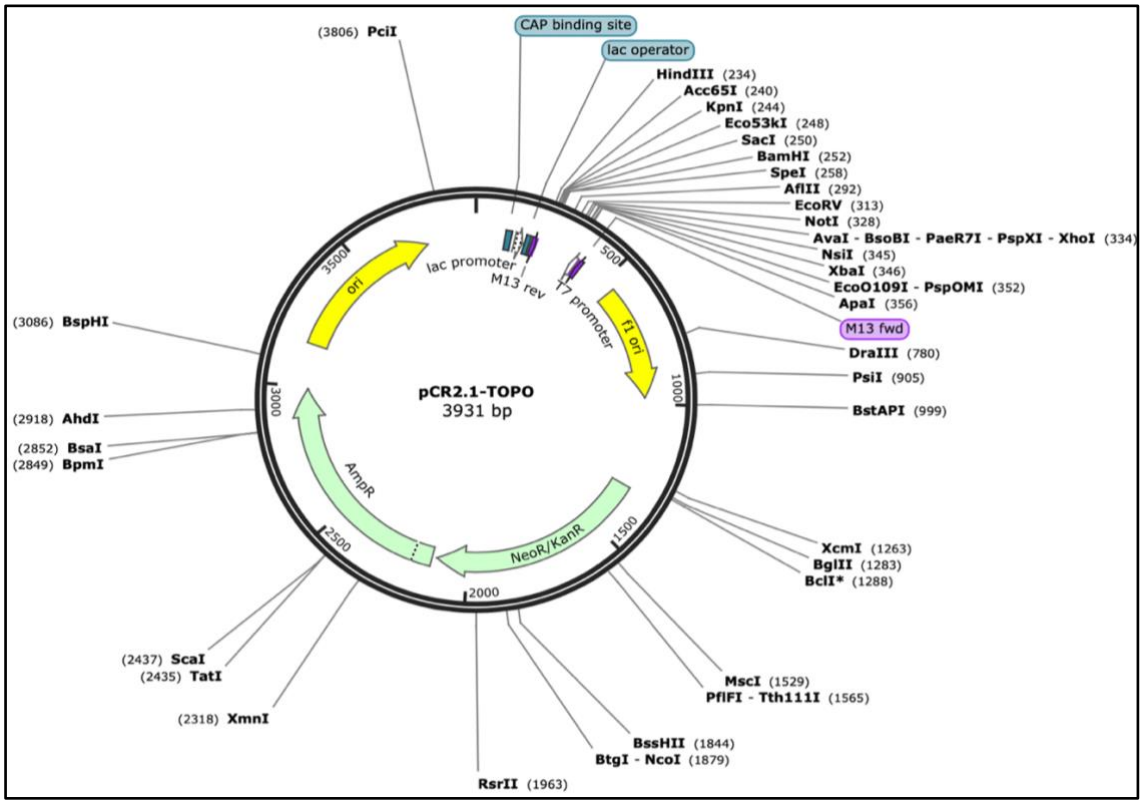


Figure 2.1:Map of the pCR 2.1 TOPO vector and restriction enzymes. (Created by SnapGene viewer; v.6.1.1).

2.7.2 Bacterial transformation and culture

The transformation was done by adding 2µl of TOPO cloning reaction (2.7.1) into a vial of One Shot® chemically competent *E. coli* (DH5α™ -T1^{R}) (Invitrogen, Carlsbad, CA, USA). This mix was incubated on ice for 30 minutes and then heat-shocked for 30 seconds at 42°C. To culture the transformed cells, it was mixed with 250µl of Super Optimal broth with Catabolite repression (S.O.C.) medium (2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, and 20mM glucose) and incubated for one hour at 37°C with shaking at 200 rpm. The cells were then spread on Luria-Bertani (LB) agar plates supplemented with ampicillin (25µg/ml) and incubated overnight at 37°C. The isolated colonies were picked using a sterile 1µl loop and added into a tube containing 5ml of LB broth (10g/L tryptone, 5g/L NaCl, 5g/L yeast extract) and then incubated at 37°C overnight in the shaker (250 rpm).

2.7.3 Plasmid DNA purification

The purification of plasmid DNA was performed using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. 1.5ml of cultured bacterial cells was transferred into a 1.5ml microcentrifuge tube and then centrifuged at 4000 *x g*. The supernatant was discarded, and the pelleted bacterial cells were resuspended in 250µl buffer containing RNase (buffer P1) (Qiagen) then lysed in 250µl NaOH/SDS buffer (buffer P2) (Qiagen). The lysate was neutralised by adding 350µl neutralisation buffer N3 (Qiagen), and centrifuged for 10 minutes at 17,000 *x g*. The cellular debris, denatured proteins, chromosomal DNA, and SDS were discarded. The remaining supernatant containing plasmid DNA was transferred into a QIAprep 2.0 spin column and centrifuged for 1 min. The flow-through was discarded and the spin column was washed by adding 500µl buffer PB (Qiagen) then centrifuged for 1 min and the flow-through was discarded. The spin column was washed again by adding 750µl buffer PE (Qiagen). The flow-through containing salts and endonucleases was discarded. Finally, the spin column was placed in a clean 1.5ml microcentrifuge tube and the plasmid DNA was eluted in 50µl buffer EB (10 mM Tris Cl, pH 8.5) (Qiagen).

2.7.4 Restriction enzyme digestion

To check insert sizes, a restriction enzyme digestion (Promega, Madison, Wisconsin, USA)) reaction was carried out in a total volume of 20µl by assembling the following components: 2µl restriction enzyme 10X buffer (Promega), 0.2µl acetylated bovine serum

albumin (BSA) (10 µg/ µl), 0.5µl restriction enzyme (*EcoR1*), 5µl DNA, and 12.3µl nuclease free water. The mixture was then incubated at 37°C for three hours and analysed by gel electrophoresis (1%) with a GeneRuler 1kb plus ladder as a size marker.

2.8 Bead-based purification

Purification of PCR amplicons was performed using paramagnetic beads (AxyPrep Mag PCR Clean-up kit, Axygen, union city, CA, USA). The beads were mixed well before use by vortex for 1 minute and then added to the PCR reaction in 1:1.8 ratio (reaction volume: beads ratio). This was mixed by pipetting up and down until it looked homogenous and then incubated in a rotator mixer for 15 minutes at room temperature. The mixture was then spun down and placed on a magnetic rack until the solution was clear. Keeping the tube in the magnetic rack, the supernatant was pipetted off carefully without disturbing the pellet. The beads pellet was washed twice with 200µl of freshly prepared 80% ethanol for 30 seconds each time. To remove any residual ethanol, the tube was spun down, placed back on the magnetic rack, and allowed to dry for ~ 1-2 minutes at room temperature with the lid off. The tube was removed from the magnetic rack to resuspend the bead pellet in the desired volume of nuclease free water and incubated for 2 minutes at RT. It was then placed back on the magnetic rack until the solution was clear and colourless then the supernatant was removed and placed in a clean 1.5ml Eppendorf tube.

2.9 Sanger sequencing

2.9.1 Enzymatic PCR products clean-up

Before Sanger sequencing, PCR products were treated with Exonuclease I and Shrimp Alkaline Phosphatase (ExoSAP-IT) (Affymetrix, Santa Clara, CA, USA) to remove the excess nucleotides and primers. The reaction was carried out by adding 1µl of ExoSAP-IT to 2.5µl of PCR product and then incubation at 37℃ for 15 minutes, followed by an inactivation at 80℃ for 15 minutes.

2.9.2 The sequencing reaction

The sequencing reaction was prepared according to manufacturer's instructions using the BigDye Terminator v3.1 kit (Applied Biosystems, Waltham, Massachusetts, USA). The reaction included 1.5µl 5x BigDye Terminator sequencing buffer (Applied Biosystems), 1µl of the diluted sequencing primer (1.6µM), 0.5µl BigDye Terminator v3.1 (Applied

Biosystems), 1µl ExoSAP-IT treated PCR product, and nuclease free water up to 10µl total volume. This reaction was performed for each primer and for each sample.

Sequencing reaction settings used were as follows: Initial denaturation was at 96℃ for 1 minute, followed by 25 cycles of denaturing at 96℃ for 10 seconds, annealing at 50℃ for 5 seconds, and extension at 60℃ for 4 minutes. All temperatures were ramped at 1℃/ second.

At the final step, the DNA was precipitated with 60µl 100% Ethanol and 5µl 125mM EDTA (final concentration: 8mM EDTA). The mixture was centrifuged at 3900 rpm (3000 *x g*) at 20℃ for 30 minutes. The supernatant was removed by upending the sequencing plate on tissue multiple times, then 60µl of 70% freshly prepared ethanol were added and the sequencing reactions were centrifuged at 2000 rpm (800 *x g*) at 4℃ for 15 minutes. The plate was upended again on a tissue and then left to air dry at 37℃ for 1 minute to remove the ethanol completely. The pellets were re-dissolved in 10µl Hi-Di formamide loading buffer (Applied Biosystem). An ABI3130xI Genetic analyser (Applied Biosystem) was used to perform the sequencing and the produced sequences were analysed on SeqScape (v2.5, Applied Biosystems) and SnapGene Viewer 6.0.4. Figure 2.2 shows the basic steps of Sanger sequencing.

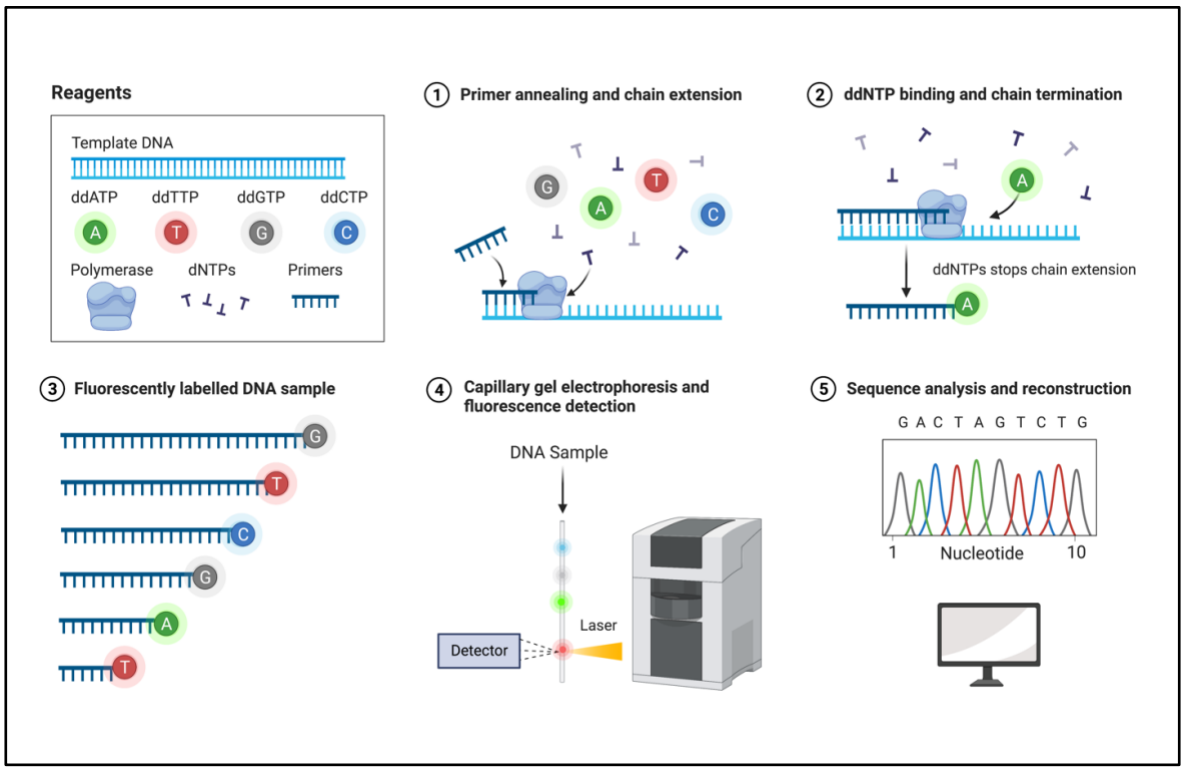


Figure 2.2: Basic steps of Sanger sequencing. (Created with BioRender.com, Accessed on 10/06/2022).

2.10 Whole Exome Sequencing (WES)

2.10.1 WES library preparation

A Qubit® dsDNA (Broad Range) assay kit (Invitrogen, Carlsbad, CA, USA) was used to quantify the genomic DNA (gDNA) in each sample (Section 2.3.1). The samples were then aliquoted at a concentration of ~ 70ng/µl as required by the NGS facility.

All WES reactions were carried out in the Next Generation Sequencing facility (St James's University Hospital, Leeds, UK) using the Agilent Technologies SureSelect^{XT} Human All Exon V6 kit (Agilent Technologies, Santa Clara, CA, USA), following the manufacturer's protocol (https://www.agilent.com/cs/library/usermanuals/Public/G7530-90000.pdf) for 200ng DNA samples. The reagents used in the library preparation were included in the SureSelect^{XT} Library Prep kit ILM, SureSelect^{XT} Target Enrichment and Herculase II Fusion DNA Polymerase Box-1, and SureSelect^{XT} Target Enrichment Kit ILM Indexing Hyp Module Box-2.

In a 1.5ml LoBind tube, 200ng of high quality gDNA was diluted in 1X low TE buffer (10mM Tris HCl (pH 8.0) and 1mM EDTA (pH 8.0)) to a total volume of 50µl followed by shearing the DNA to form fragments with size range from 150 to 200bp using a Covaris E220 ultrasonicator. The quality assessment of the samples was done using an Agilent 2200 TapeStation and D1000 ScreenTape (Agilent Technologies). The generated electropherogram should show a DNA fragment size peak between 150-200bp.

Next, DNA fragment ends were blunted and 5'phosphorylated using the SureSelectXT Library Prep Kit ILM (Agilent Technologies) to prepare an end repair mixture. For one reaction, 35.2µl of nuclease free water was added to 10µl of 10X end repair buffer, 1.6µl dNTP mix, 1µl T4 DNA polymerase, 2µl Klenow DNA polymerase, and 2.2µl of T4 polynucleotide kinase. In each PCR plate well, 52µl of master mix was added to 48µl of each sheared DNA sample and then incubated in a thermal cycler at 20℃ for 30 minutes with heated lid off. After repairing the ends, bead-based clean-ups were performed using 180µl Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics) (for each 100µl end repaired DNA sample), followed by two steps wash with 200µl freshly prepared 70% ethanol for each wash, and 32µl nuclease free water to elute the DNA. Adenylation of the 3' end of the fragments was performed by preparing a 20µl dA-tailing master mix by adding 5µl of 10 X Klenow polymerase buffer, 1µl dATP, and 3µl Exo (-

) Klenow to 11µl of nuclease free water. This mixture was then added to the end repaired, purified DNA samples, and the mix was incubated in the thermal cycler at 37℃ for 30 minutes. A purification step was performed again using 90µl Agencourt AMPure XP magnetic beads (for each 50µl dA-tailed DNA sample), followed by two steps wash with 200µl 70% ethanol for each wash, and finally eluted in 15µl nuclease free water. The paired-end adaptor was ligated by mixing 13µl of the dA-tailed, purified DNA sample with a mixture containing 15.5µl nuclease free water, 10µl 5 X T4 DNA ligase buffer, 10µl diluted SureSelect adaptor Oligo mix, and 1.5µl 5 X T4 DNA ligase. The plate was incubated for 15 minutes at 20℃ and the samples were purified with 90µl AMPure XP beads (for each 50µl adaptor-ligated DNA sample), two steps wash with 200µl freshly prepared 70% ethanol per wash, and 32µl nuclease free water for the final elution.

A pre-capture PCR amplification step was performed to amplify the adaptor-ligated libraries by preparing a pre-capture PCR reaction mix consisting of 6µl nuclease free water, 1.25µl SureSelect primer, 1.25µl SureSelect ILM Indexing Pre-Capture PCR Reverse primer, 10µl 5× Herculase II Reaction buffer, 0.5µl 100 mM dNTP mix, and 1µl Herculase II Fusion DNA Polymerase. In a PCR plate, 20µl of the PCR reaction mixture and 30µl of each purified DNA library sample was added to each well and mixed by pipetting. Using a thermal cycler, the PCR program was run as shown in Table 2.7.

Segment	No. of Cycles	Temperature	Time
1	1	98℃	2 minutes
2	10	98℃	30 seconds
		65℃	30 seconds
		72℃	1 minute
3	1	72℃	10 minutes
4	1	4℃	Hold

Table 2.7: Pre-Capture PCR amplification Program.

A final purification step was done as described above (using 90µl AMPure XP beads, 200µl 70% ethanol wash, and 30µl nuclease free water for DNA resuspension). The quality and quantity of the amplified, purified adaptor tagged libraries were assessed using an Agilent 2200 TapeStation with a high sensitivity D1000 ScreenTape. The

generated electropherogram should show a distribution with a DNA fragment size peak of 225-275bp (Figure 2.3).

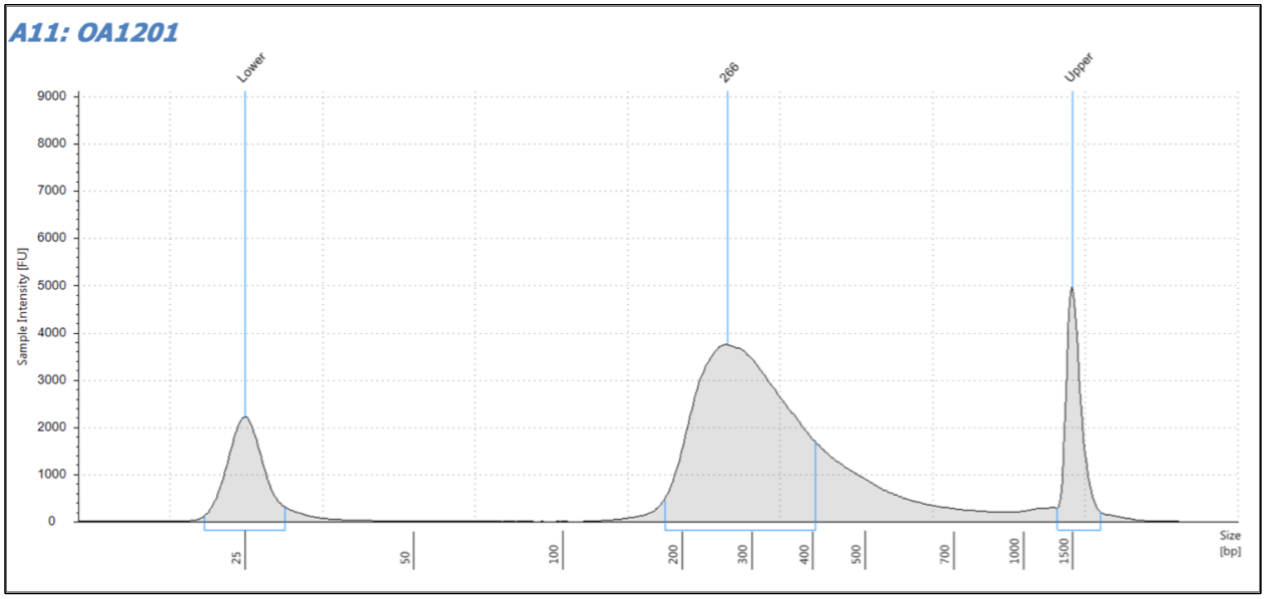


Figure 2.3: Pre-Capture quality and quantity assessment of amplified library of Sample OA1201 using 2200 TapeStation with a high sensitivity D1000 ScreenTape.

2.10.1.1 Hybridization and capture

One hybridization and capture were performed for each sample library and the initial DNA concentration required for the hybridization reaction should be 221ng/µl. To prepare the hybridization reaction, a hybridization buffer was made by mixing the components in Table 2.8 at room temperature (RT).

Reagents	Volume for 1 reaction
SureSelect Hyb 1	6.63 µl
SureSelect Hyb 2	0.27 µl
SureSelect Hyb 3	2.65 µl
SureSelect Hyb 4	3.45 µl
Total	13 µl

Table 2.8: Components used to prepare Hybridization Buffer.

Next, a SureSelect Block Mix was prepared by mixing 2.5µl SureSelect indexing block 1, 2.5µl SureSelect block 2, and 0.6µl SureSelect IML indexing block 3. In each PCR plate well, 3.4µl of gDNA library sample (221ng/µl) was mixed with 5.6µl of SureSelect Block Mix then transferred to a thermal cycler and the program was run as follows: 95℃

for 5 minutes then held for at least 5 minutes at 65℃. The appropriate dilution of SureSelect RNase block and the capture library hybridization mix were prepared based on the size of the capture library. The capture library hybridization mix was produced by mixing 13µl of hybridization buffer mixture (Table 2.8), 5µl of 10% RNase block solution, and 2µl of capture library (size < 3Mb) (2µl of 25% of RNase block solution and 5µl of capture library (size ≥ 3Mb)). 20µl of the capture library hybridization mix was added to the gDNA library + block mix and then incubated in a thermal cycler at 65℃ for 24 hours. The hybridized DNA library was then captured using previously prepared and washed streptavidin coated magnetic beads (Dynabeads MyOne Streptavidin T, Thermofisher Scientific). The hybridization mixture was transferred into a plate with 200µl of streptavidin beads and mixed vigorously for 30 minutes at RT then placed in a magnetic separator to collect the beads and until the solution was clear. Supernatant was removed and the beads were resuspended in 200µl SureSelect wash buffer 1, incubated at RT for 15 minutes, and again separated using the magnetic separator. Using SureSelect wash buffer 2, the beads were washed three times, recovered with the magnetic separator, and finally the captured DNA library was dissolved in 30µl of nuclease free water.

2.10.1.2 Indexing and sample processing

The captured DNA libraries were PCR-amplified with the appropriate indexing primer for each sample. To prepare the post-capture PCR reaction, 18.5µl nuclease free water was mixed with 10µl 5 X Herculase II Reaction Buffer, 1µl Herculase II Fusion DNA Polymerase, 0.5µl 100 mM dNTP mix, and 1µl SureSelect ILM Indexing Post-Capture Forward PCR Primer. 14µl of each sample library was added along with 5µl of the appropriate indexing primer to 31µl of the post-capture PCR reaction mixture in each well. The PCR amplification program was then run as shown in Table 2.9.

Segment	No. of Cycles	Temperature	Time
1	1	98℃	2 minutes
2	12	98℃	30 seconds
		57℃	30 seconds
		72℃	1 minute
3	1	72℃	10 minutes
4	1	4℃	Hold

Table 2.9: Post-Capture PCR amplification program.

AMPure XP beads were used to purify the indexed libraries and the library quality was assessed on an Agilent 2200 TapeStation with a high sensitivity D1000 ScreenTape. The electropherogram was expected to show a DNA fragment size peak localised between 250 and 350bp (Figure 2.4).

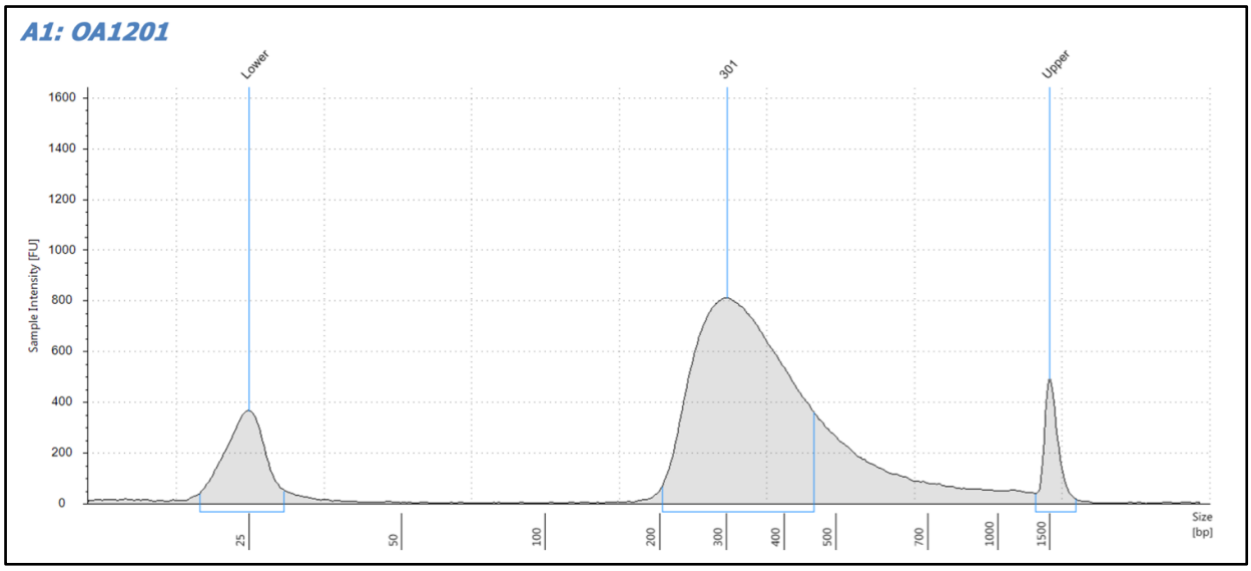


Figure 2.4: Post-Capture quality and quantity assessment of amplified library of Sample OA1201 using 2200 TapeStation with a high sensitivity D1000 ScreenTape.

Finally, 13 samples were pooled per lane with a final concentration of 10nM in a final volume of 50µl. The cluster amplification was carried out at the Leeds Next Generation Sequencing (NGS) facility using the appropriate Illumina Paired End Cluster generation kit, with read lengths between 100-150bp, on a single lane of an Illumina HiSeq 3000 Sequencer (Figure 2.5).

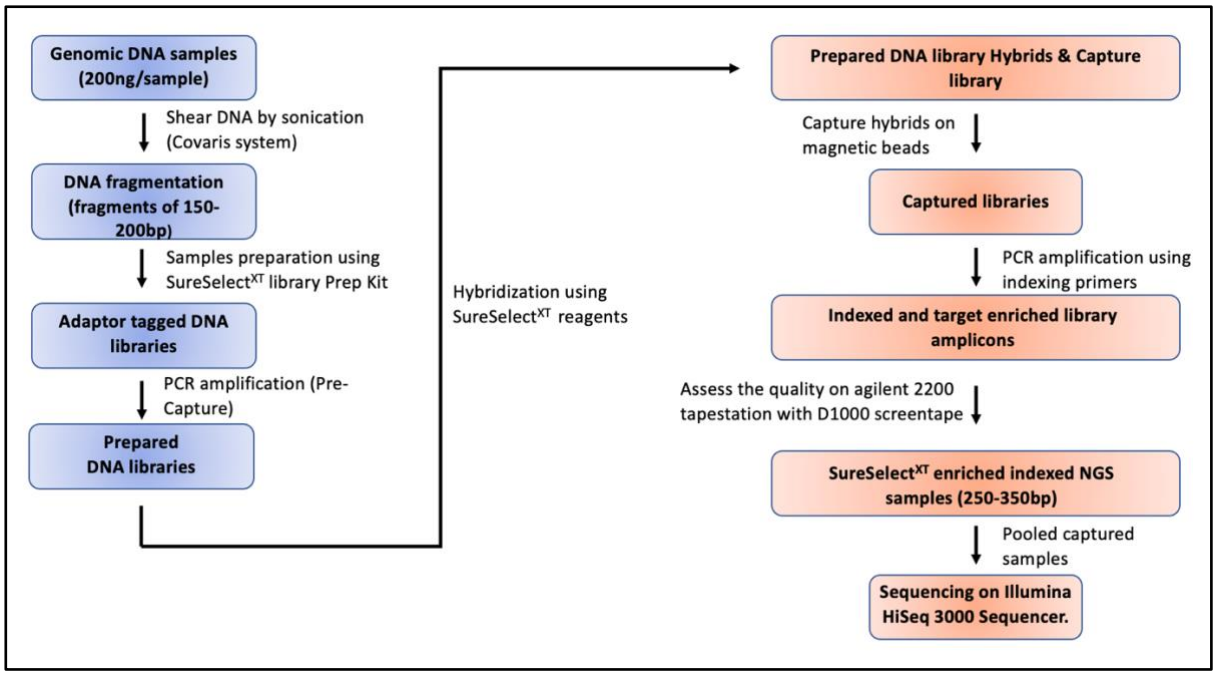


Figure 2.5: SureSelect^{XT} sample preparation workflow for Illumina paired-end sequencing. (Created by PowerPoint).

2.10.2 WES data analysis and bioinformatics

The output files from the WES analysis were analysed using in-house PerlScript (UNIX commands can be found in Appendix A.1). The quality control of the raw sequence data, the base quality scores, GC content and duplications were checked using the java based FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and reads with quality less than 20 were eliminated. The quality control was assessed throughout the NGS analysis in three main stages: raw data, alignment, and variant calling. To avoid adaptor contamination, Trim Galore software (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to remove the sequence adaptors (the first 13 bp by default). Trimmed sequences were then aligned against the reference genome (hg19/GRCh37) using the Burrows-Wheeler Aligner BWA (v0.7.12-r1.39) (Li and Durbin, 2009). It should be emphasised that all the analysis pipelines were built up for hg19/37 when the work started, thus it has been used throughout the thesis work even though, if the work had to start over, hg38 would probably be chosen. The generated SAM files were then converted to BAM files using SAMtools and then sorted by Picard (sort SAM). The Picard tools (v2.5.0) (https://broadinstitute.github.io/picard/) were used to mark and remove the PCR duplicates. Sorted BAM files were recalibrated and realigned locally around the indels using the Genome Analysis Tool Kit GATK (https://gatk.broadinstitute.org/hc/en-us)

(v3.5) (McKenna et al., 2010, DePristo et al., 2011). The aligned data files were visualised and explored using the Integrative Genomic Viewer (IGV) (https://www.broadinstitute.org/igv/). Figure 2.6 illustrates the major steps in WES pipeline.

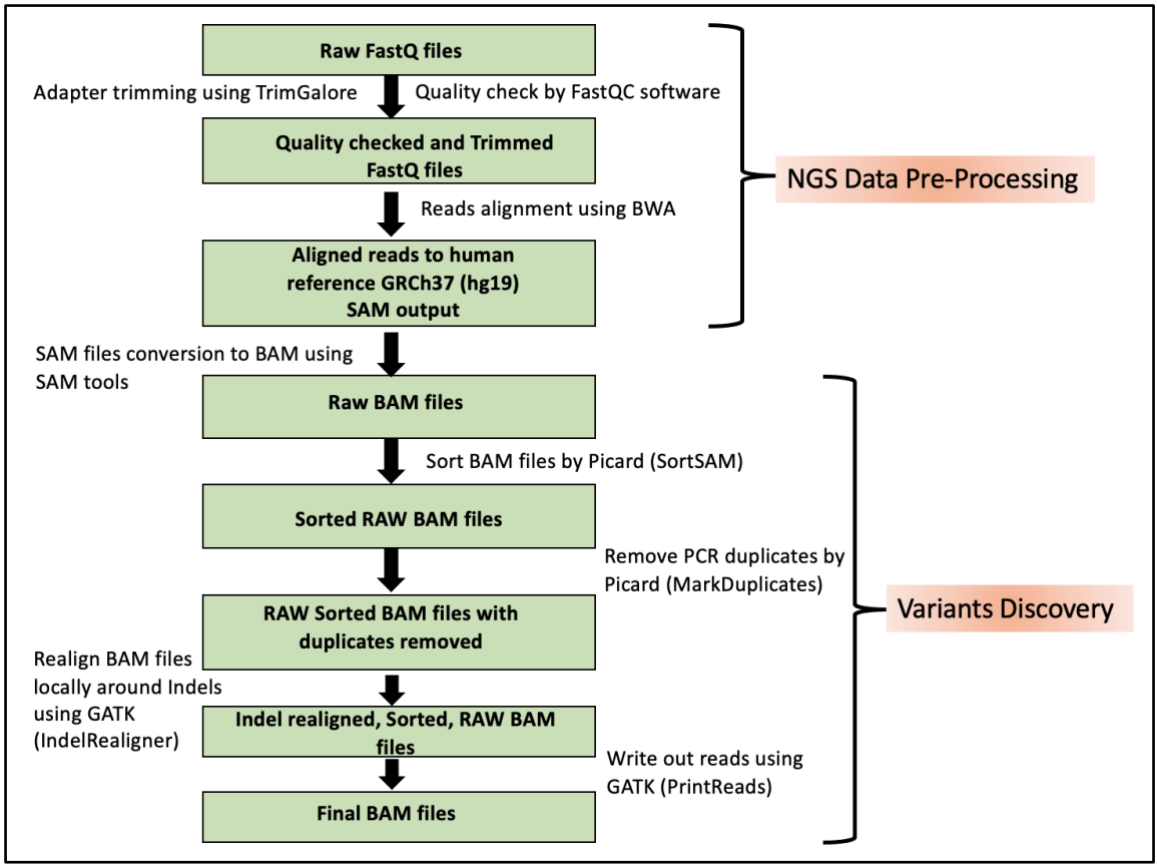


Figure 2.6: Bioinformatic Pipeline for Whole exome sequencing (data alignment). (Created by PowerPoint).

The HaplotypeCaller function in GATK was used to call the small insertion/deletions (indels) and single nucleotide variants (SNVs) in genomic variant call format (g.VCF). The genomic VCF files were converted to raw variant call format (raw.VCF). Hard filtering and recalibrating for SNVs and indels were performed separately before combining them to generate a final combined VCF file (Danecek et al., 2011).VCF files were filtered and annotated with the help of the VCF hacks package available at (https://github.com/david-a-parry/vcfhacks.git). Filtration of all VCF files was done using the following databases of known variants:

- The National Centre for Biotechnology Information's (NCBI) database of genetic variation (dbSNP 146) (https://www.ncbi.nlm.nih.gov/snp/) (Sherry et al., 2001).

- The Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org)

For each individual, the variant list was annotated using Variant Effect Predictor (VEP) software (McLaren et al., 2010). The filtration of the annotated variants was done by applying the following criteria: removing variants with minor allele frequency (MAF) of greater than or equal 1% in gnomAD and dbSNP 146 (in dominant and sporadic cases, variants with MAF ≥ 0.01% were filtered out), eliminating synonymous changes, selecting only coding DNA variants and variants in splice donor and acceptor sites (± 2 bp), and selecting variants with depth of coverage greater than 10 reads. The pathogenicity prediction score for each variant was calculated using Combined Annotation Dependent Depletion (CADD) (https://cadd.gs.washington.edu) (Kircher et al., 2014) and only variants with CADD score ≥ 15 were preserved and assessed additionally by other pathogenicity prediction tools such as: Sorting Intolerant from Tolerant (SIFT), MutationTaster, PolyPhen-2, and BLOSUM62 (Section 2.13.4). The variants with CADD score of 15 or above and predicted to be pathogenic in at least one of the other tools were retained and prioritised in the final lists. The final variant lists were initially compared to the existing IRD associated genes in the RetNet database (https://sph.uth.edu/retnet/) and checked further by investigating manually in gnomAD, dbSNP 146, Franklin databases and doing literature searches to look for the gene expression, function, and links with human inherited diseases. The variant detection pipeline and data filtration criteria is outlined in Figure 2.7.

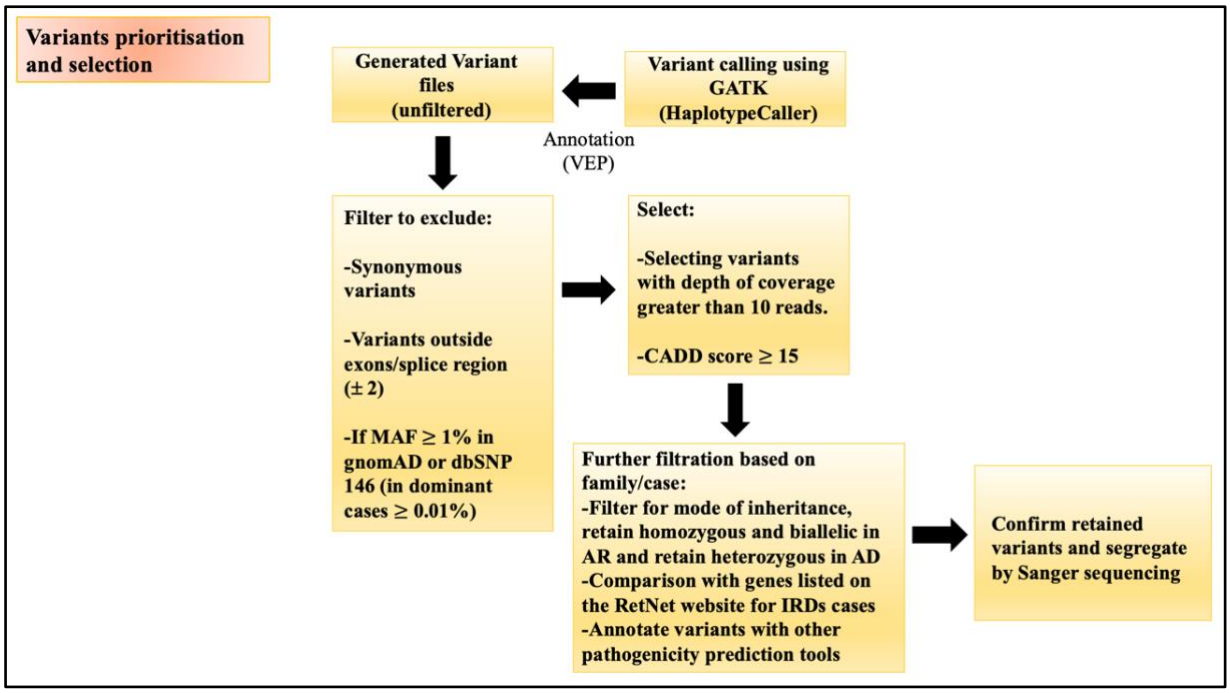


Figure 2.7: Schematic representation of variant detection and prioritisation pipeline.

2.11 ExomeDepth and CNVs calling

The sequencing batches consisted of unrelated samples, which meant that ExomeDepth (Plagnol et al., 2012) could be used to identify copy number variations (CNVs) that were too large to be detected by the variant calling protocol described above (Section 2.10.2). The FASTQ files of all samples were analysed using the pipeline described in section 2.10.2 up to the BAM files indexing. The indexed BAM files were then analysed using the R package ExomeDepth (https://cran.r-project.org/web/packages/ExomeDepth/vignettes/ExomeDepth-vignette.pdf) (Plagnol et al., 2012) (R commands are listed in Appendix A.2). ExomeDepth analysis compares the read depths of all captured exons of the test sample with the read depths of a reference set of 5-10 samples. The reference samples and the test sample were all sequenced and processed in the same sequencing run to avoid confounding batch effects. The output csv. file was filtered by filtering out the common CNVs (Conrad et al., 2010b) and prioritising the calls with highest Bayes factor (the log10 of the likelihood ratio of data for the CNV call divided by that of the normal copy number call) (Plagnol et al., 2012). The zygosity was described based on the read ratio (RR) as follows:

RR= 0 → homozygous deletion

RR= 0.5 → heterozygous deletion

RR=1.5 → heterozygous duplication

RR= 2 → homozygous duplication

2.12 Homozygosity mapping

In consanguineous families, homozygosity mapping (Section 1.8.3) is a method used for mapping recessive traits. Its goal is to find and score consecutive homozygous genotypes, also known as "runs of homozygosity" (ROH), in a person's genome (Quinodoz et al., 2021). Children of consanguineous unions have ROHs for DNA regions that were co-inherited from ancestors who are common to both parents (McQuillan et al., 2008). Because recessive disease mutations are frequently discovered in ROHs (Wakeling et al., 2019), homozygosity mapping has been utilised in medical genetics to identify regions of the genome that should be prioritised for targeted mutational screening. Detection of homozygous regions from WES data was carried out using AutoMap (Autozygosity Mapper) (https://automap.iob.ch) (Quinodoz et al., 2021). The homozygosity mapping was done directly on VCF (Variant Call Format) calls from WES data using this tool, which is both web-based and downloadable. The output file contains both a PDF file with a graphical representation of autozygous regions along the autosomes and a text file with numerical data for the same information, such as the positions of the detected ROHs, their size, number of variants, and homozygosity percentage.

2.13 Bioinformatics

Bioinformatics is an interdisciplinary field that combines biology, computer science, mathematics, and statistics. It includes methods and software tools to analyse biological data, and, in this study, it was used to investigate the genes/proteins' structure and function, design primers, carry out large-scale analysis of WES, PacBio, and MinION sequencing data, and predict the mutation pathogenicity scores.

2.13.1 Online databases for literature searches

Candidate genes, phenotypes, pathogenic mutations, polymorphisms, protocols, techniques, and bioinformatics were investigated using wide range of online databases including: PubMed (https://pubmed.ncbi.nlm.nih.gov), Google Scholar (https://scholar.google.com), RetNet (https://sph.uth.edu/retnet/), GeneCards (https://www.genecards.org), the Ensembl Genome Browser (https://www.ensembl.org/index.html) and NCBI website (https://www.ncbi.nlm.nih.gov). Online Mendelian Inheritance in Man (OMIM) website

(https://omim.org) is a comprehensive online reference used to look over human genes and genetic phenotypes.

2.13.2 UCSC Genome Browser

The UCSC genome browser (Kent et al., 2002) is a web-based tool that acts as a multi-power microscope, allowing researchers to investigate all 23 chromosomes of the human genome from a full chromosome to a single nucleotide (https://genome.ucsc.edu/index.html). In an interactive and graphical display, the browser brings together the work of countless scientists throughout the world, including work done at UCSC. It also gives users access to the genomes of over a hundred different organisms. It has custom tracks with lots of items including databases of human genes and phenotypes such as Online Mendelian Inheritance in Man (OMIM available at https://omim.org), gene expression and variation such as ExAC, gnomAD, and dbSNP. Genomic sequence, protein sequence, polymorphism sites and frequencies, and intron-exon architectures were all explored in depth using this tool. The UCSC genome browser also offers tools such as BLAT and *in silico* PCR that were used to determine the specificity of primer sequences.

2.13.3 Bioinformatics comparative tools

BLAST alignment search tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi

EMBOSS Matcher, https://www.ebi.ac.uk/Tools/psa/emboss_matcher

DotMatcher, http://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher

EMBOSS stretcher, https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/

2.13.4 Mutation pathogenicity prediction software

2.13.4.1 Combined Annotation Dependent Depletion (CADD) score

The CADD score (Kircher et al., 2014) integrates various genomic annotations including gene model annotations, sequence context, evolutionary constraint, epigenetic measurements, and functional predictions into a single score using a machine learning model. It also scores the deleteriousness of any human single nucleotide variant (SNV) along with insertion/deletion events in the human genome. Most scoring tools and variant annotations tend to utilise a single information type e.g., conservation, and/or are confined in one scope e.g., missense changes. CADD therefore is a broadly applicable metric that

compares the annotations of fixed or nearly fixed human derived alleles that survived natural selections with simulated variants. It can prioritize deleterious, and disease causative variants within the individual genome sequences providing a genome-wide, data rich, and functional generic estimate of variant effect. The scores are generated in two forms. Raw scores are the immediate output from the model, giving negative values for observed variants and positive values for simulated variants. Since the raw scores have no definite unit, Phred-like scores (scaled C- scores) have been used to simplify the interpretation. The scaled C- scores range from 1-99. Variants with scores of 10 or higher are predicted to be the top 10% most deleterious variants in human genome, a score of 20 or more indicates the top 1% most deleterious and 30 or greater indicates the top 0.1% and so on (Rentzsch et al., 2019).

2.13.4.2 SIFT

Sorting Intolerant from Tolerant (https://sift.bii.a-star.edu.sg) is an algorithm that predicts the potential effect of single amino acid substitutions on protein function based on the sequence homology of amino acids and their physical properties. It can be used to analyse both naturally occurring single nucleotide variants (SNVs) and laboratory induced missense variants. The score ranges from 0.0 to 1.0, where an amino acid substitution with a score ≤ 0.05 is predicted to be damaging and a variant with score > 0.05 is predicted to be tolerated (Ng and Henikoff, 2001, Kumar et al., 2009).

2.13.4.3 PolyPhen-2

To annotate coding nonsynonymous single nucleotide polymorphisms (nsSNPs), Polymorphism Phenotyping v2 tool was used (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al., 2010). By applying the HumVar trained PolyPhen-2 (for Mendelian diseases), the damaging impact of missense mutations on the function and structure of the resulting protein was predicted. The score represents the probability that a mutation is damaging and ranges from 0.0 (tolerated) to 1.0 (deleterious). It can be interpreted as follows:

< 0.4 Benign

0.4-0.8 Possibly damaging.

> 0.8 Probably damaging.

2.13.4.4MutationTaster2

MutationTaster2 is a web-based software package used to evaluate the pathogenic impact of amino acid substitutions as well as intronic and synonymous changes, short insertions and/or deletions (indels) and splice site variants. It can predict variant pathogenicity based on all known single nucleotide polymorphisms (SNPs) and indels from 1000 Genomes project (1000G) (Consortium, 2012), ClinVar (Landrum et al., 2014), HapMap (Gibbs et al., 2003), and HGMD public (http://www.hgmd.cf.ac.uk/ac/index.php) (Stenson et al., 2014). Variants found in homozygous state more than four times in 1000G or in HapMap are considered neutral. Known pathogenic mutations in ClinVar are automatically predicted to be disease-causing (Schwarz et al., 2014). It ranks an alteration as one of four possible types:

Disease causing → probably deleterious.

Disease causing automatic → known to be deleterious.

Polymorphism → probably harmless

Polymorphism automatic → known to be harmless.

2.13.4.5BLOSUM62

BLOcks SUbstitution Matrix 62 was developed by analysing the frequencies of amino acid substitutions in conserved clusters of proteins (https://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt). Investigators determined the spectrum of normal substitution frequencies based on aligned proteins from about 2000 aligned sequence segments characterising more than 500 groups of related proteins (Henikoff and Henikoff, 1992). The BLOSUM62 scores are log-odds scores, and they measure the biological probability that a substitution could occur within the spectrum of normal variation relative to the probability of the substitution is disease causing. It ranges from +3 to -4, where substitutions with positive scores are more likely to be benign and those with negative scores are likely to be damaging.

2.13.4.6 Franklin database

It is the world's first open genomic community (https://franklin.genoox.com/clinical-db/home), founded by Genoox and named after Rosalind Franklin, whose research helped us understand DNA structure. It is a web-based tool that allows genomic experts to spend less time looking into variants and answering any genetic questions. It collects data from

several genetic databases to estimate variant pathogenicity and delivers a full assessment. The input can be a single nucleotide variant (SNV), copy number variation (CNV), or runs of homozygosity (ROH). For each variant, it provides information about its pathogenicity predictions, population frequencies, clinical evidence, references, and much more.

2.13.5 Integrative Genomics Viewer (IGV)

IGV (v.2.11.2) (https://software.broadinstitute.org/software/igv/) (Robinson et al., 2011) is a high-performance visualisation tool that enables real time exploration of a wide range of genomic data generated by next generation sequencing, array-based sequencing, and long read sequencing. It supports loading genomic files in different formats including nonindexed formats (e.g., GFF, BED, and WIG), aligned and indexed format (e.g., BAM & Goby), and multiresolution format (e.g., TDF, bigwig, and bigBed). Reads can be investigated for mutations, copy number variations, depth of coverage, gene expression, and methylation.

2.13.6 Protein alignment

Homologene (https://www.ncbi.nlm.nih.gov/homologene/) is an automated system for detecting homologs among annotated proteins in various sequenced eukaryotic genomes. It was used to examine the level of conservation for the amino acid of interest across several species by aligning the protein sequences using blasp. It also gives information about paralogues, orthologues, phenotype, and conserved domains.

2.14 Long read sequencing

2.14.1 PacBio® sequencing

Barcoded amplicons were generated using 48 pairs of 16 bp barcodes that are custom designed for PacBio system (Appendix D.4.2) by applying the steps in section 2.4.3.1. The barcoded amplicons were then pooled into a 1.5ml Eppendorf tube to be prepared for sequencing.

2.14.1.1 Library preparation

All PacBio reactions were carried out at the NGS facility (St James's University Hospital, Leeds, UK), and the steps were performed following the Pacific Biosciences Procedure and Checklist – Amplicon Template Preparation and Sequencing protocol, version 04

(June 2018) (https://dnatech.genomecenter.ucdavis.edu/wp-content/uploads/2018/07/PacBio-Amplicon-Template-Preparation-and-Sequencing-02-2018.pdf) with some adjustments as outlined below. All PacBio SMRT sequencing library preparation reactions were prepared using Pacific Bioscience SMRT sequencing kits (PacBio®, California, USA) following the manufacturer's instructions. All incubations were done using the BioRad DNA engine thermal cycler. 1µl of the pooled amplicons was diluted 1:10 with Qiagen buffer EB, and run on the Tapestation 2200 HSD5000 tapes, Qubit HS dsDNA kit and Nanodrop 8000 to determine the size, concentration, and purity of the sample. As an essential step prior to SMRTbell library preparation, pooled samples were purified using AMPure PB beads. The bead concentration was determined based on the amplicon size (2.272 kb) according to manufacturer's instructions and was added at 0.6X relative to the PCR reaction volume. The bead/DNA solution was mixed thoroughly by shaking the tube in VWR® vortex mixer at 2000 rpm for 10 minutes at room temperature. The tube was spun down to collect the beads and then placed in a magnetic bead rack until the solution appeared clear. Keeping the tube in the rack, the clear supernatant was pipetted off and the beads were washed twice with 1.5ml of freshly prepared 70% ethanol. The residual ethanol was removed by allowing the bead pellet to air dry for 60 seconds. A suitable volume of Qiagen buffer EB was added to the beads and mixed for 2 minutes in Eppedorf MixMate at 2000 rpm until the solution became homogenous. The tube was then spun down and placed back to the magnetic bead rack to collect the eluted sample and discard the beads. 1µl of the cleaned pooled amplicons was diluted 1:10 with Qiagen buffer EB, and run on the Tapestation 2200 HSD5000 tapes, Qubit HS dsDNA kit and Nanodrop 8000 to determine the size, concentration, and purity of the sample.

To repair any DNA damage, in a 0.2ml PCR tube, the reaction was prepared by adding the following reagents to a 500ng of sample: 5µl DNA damage repair buffer (final concentration 1X), 0.5µl NAD^+ (nicotinamide adenine dinucleotide) (final concentration 1X), 5µl ATP high (final concentration 1mM), 0.5µl dNTP (final concentration 0.1mM), dsDNA up to 37µl, and nuclease free water to adjust to 50µl total volume. The reaction was mixed well by flicking the tube and incubated at 37ºC for 60 minutes then held at 4ºC for one minute. Subsequently, to repair the DNA ends, 2.5µl of end repair mix was added to the repaired DNA reaction and incubated for 5 minutes at 25ºC then at 4ºC. The end repaired reaction was then transferred to a 1.5ml Lo-Bind tube and directly purified

using AMPure PB beads as explained above, using a 0.6X AMPure PB bead ration and eluting into 33µl of Qiagen buffer EB.

A blunt ligation reaction was then prepared on ice in a 0.2ml PCR tube by mixing the following components: 31µl of end repaired DNA, 2µl blunt adaptor (final concentration 1µM), 4µl template prep buffer (final concentration 1X), 2µl ATP low (final concentration 0.05mM), 1µl ligase (final concentration 0.75 U/ µl), and nuclease free water to adjust to 40µl total volume. It was incubated at 25ºC for 15 minutes and then at 65ºC for 10 minutes to inactivate the ligase. Exonucleases were added as follows: 0.5µl ExoIII and 0.5µl ExoVII to the ligated DNA to remove the failed ligation products. After incubating the mixture for one hour at 37ºC, an AMPure PB beads purification step was performed as explained above to remove all adaptor dimers. Again, using a 0.6X AMPure PB bead ration, and eluting into 32µl of Qiagen buffer EB. At this stage, a 1:10 dilution with Qiagen buffer EB of the ligated SMRTbell template was performed, and run on the Tapestation 2200 HSD5000 tapes, Qubit HS dsDNA kit to determine the size and concentration.

To ensure that no concatemers were present within the final SMRTbell template, a BluePippin size selection step was performed using the BluePippin from Sage Science (Beverly, MA, USA), a 0.75% Agarose Gel, and the 0.75% DF Low Voltage 1-6kb Marker S1 protocol (https://www.n-genetics.com/products/1275/1023/13318.pdf). A Broad Range was set to start at 1500bp and 3000bp and the protocol supplied by Sage Science to check and prepare the 0.75% gels was followed. The external marker S1 was used for size selection, and 30µl of SMRTbell template was mixed with 10µl of loading buffer supplied with the BluePippin gels. After the elution was complete, the gel was left for 45 minutes before removing the 40µl of eluted SMRTbell template to a Lo-Bind 1.5ml Eppendorf tube, to aid the recovery of long DNA fragments. 40µl of supplied 0.1% Tween was added to the elution well and left for 1 minute before being recovered to the same tube as the eluted SMRTbell template. 1:10 dilution with Qiagen buffer EB was used to check that the correct size had been recovered on the Tapestation 2200 HSD5000 tapes. This was followed by a further AMPure PB clean up, using a 0.6X AMPure PB bead ration, and eluting into 37µl of Qiagen buffer EB, and an additional repair DNA damage step as outlined above, without the end repair step. A final two AMPure PB clean up steps were performed, using a 0.6X AMPure PB bead ration, and eluting into 50µl of Qiagen buffer EB for the first one, and 10-14µl for the final one. 1µl of the SMRTbell

template was used to check the size and concentration on the Tapestation 2200 HSD5000 tapes, Qubit HS dsDNA kit. Samples were processed into SMRTbell complexes and then loaded into the sequel platform. The size and concentration of the SMRTbell Template was entered into the SMRT Link software (https://downloads.pacbcloud.com/public/software/installers/smrtlink_10.2.1.143962.zip), and the Sequel platform was run for 10-hour movie times. Data analysis was undertaken post run by Dr Ian Carr (Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK) (Section 2.14.1.2).

2.14.1.2 Data analysis and bioinformatics

Sequences from each lane were used to create circular consensus sequences (CCS) and exported as FASTQ files. Using the bespoke software package 'PacBioDemultiplexer', the locations of sequences homologous to the primers were determined in each CCS and used to determine if the CCS contained single or multiple copies of the original PCR amplicon. The indexes linked to each amplicon sequence were determined, and sequences with index combinations matching known samples exported to sample specific fasta files. Where necessary the reverse complement of the CCS was determined such that all the sequences were saved in the same orientation. Next, using a second bespoke application 'PacBioHomology', the individual amplicon sequences for each sequence were aligned to the reference sequence. Sequences with less than 70% homology to the reference sequence were discounted, with the alignment of the rest ranked by the number of exact matches found between the reference and amplicon sequences. The best 250 alignments were retained and used to generate a multiple sequence consensus alignment. This consensus was then examined to identify common single base variants as well as insertions and deletions, which were reported along with the number of CCS supporting the variant in the multiple sequence alignment. The first 100 alignments in the multiple consensus alignment were also reported to allow the accuracy of each variant reported to be manually determined by observation of the multiple alignments. An aggregated dataset of all the variants identified within a set of samples was generated using the python script p_Combine_Variants_Ref.py, which when provided with the appropriate information is able to reannotate the variants with respect to specific genomic sequences such as GenBank accession sequence etc, rather than the amplicon's sequence.

2.14.2 Oxford nanopore sequencing

2.14.2.1 Target enrichment and library preparation for MinION sequencing

Two PCR reactions were performed to generate *ORF15* amplification products that were uniquely indexed on a per-sample basis. A first PCR was carried out using *ORF15* specific primers tailed with universal sequencing tags (Section 2.4.3.2). The PCR amplicons were then subject to a bead-based purification to remove proteins, salts, dNTPs, and primers using AxyPrep paramagnetic beads (Axygen, Union City, CA, USA), as described in section 2.8. After cleaning up the amplicons, they were quantified by Qubit fluorometer (Section 2.3.1) and the molarity was calculated by the NEBiocalculator (https://nebiocalculator.neb.com/#!/ssdnaamt). To sequence multiple samples in a single flow cell, a second PCR was performed to incorporate the Oxford nanopore unique indexing barcodes to 100-200 fmol of each pre-indexed amplicon (Section 2.4.3.3). The barcoded libraries were purified again using AxyPrep paramagnetic beads (Section 2.8) and quantified by Qubit Fluorometer applying the steps in section 2.3.1.

Barcoded amplification products were pooled in equimolar quantities to a total mass of 5µg in a 1.5ml DNA LoBind Eppendorf tube. From this solution, 1µg of DNA was aliquoted and end-repaired by combining 3.5µl FFPE DNA repair buffer (NEB), 2µl FFPE DNA repair mix (NEB), 3.5µl Ultra II end-prep reaction buffer (NEB), 3µl Ultra II end-prep enzyme mix (NEB), made up with nuclease-free H_2O in a total reaction volume of 60µl. Buffers were used as supplied by (NEB). The reaction was incubated at 20 ºC for 5 minutes then 65 ºC for 5 minutes. The mixture was purified by adding 60µl AMPure XP beads (Beckman Coulter, Indianapolis, IN), then washed twice with 200µl of freshly prepared 70% ethanol and redissolved in 61µl nuclease free water.

Next, the sequencing adaptors (supplied by the manufacturer) were ligated to the double stranded amplimers. The reaction comprised of 60µl of end-prepped PCR amplimers, 25µl ligation buffer (LNB) (ONT), 10µl NEBNext Quick T4 DNA ligase (NEB), and 5µl adaptor mix (AMX) (ONT). The reaction was incubated for 10 minutes at room temperature then cleaned up using 40µl AMPure XP Beads; the beads were washed twice with 250µl short fragment buffer (SFB) (ONT). The pellet was eluted in 15µl elution buffer (ONT) then quantified using a Qubit fluorometer (Section 2.3.1) to enable library molarity to be calculated.

2.14.2.2 Priming and loading MinION flow cell

All solutions were supplied by the manufacturer (ONT) and were used according to instructions. To prepare the MinION flow cell (R9.4.1 FLO-MIN106D) for loading, the priming mix was created by adding 30µl of flush tether (FT) to a tube of thawed and mixed flush buffer (FB). 800µl of the priming mix was loaded into the flow cell priming port with care not to introduce any air bubbles. For loading, 50 fmol of the eluted library was made up to 12µl using nuclease free water mixed with 37.5µl of sequencing buffer (SQB) and 25.5µl of loading beads (LB). This mixture was then loaded into the flow cell via the SpotON port in a dropwise fashion. The MinION sequencer was run for 72 hours using MinKNOW software (v.3.6.5; ONT).

2.14.2.3 Use of flow cell wash kit with MinION sequencing

When *ORF15* amplification products were initially sequenced, pores were observed to be rapidly blocked, resulting in the production of relatively few reads. In subsequent runs a flow cell wash kit (WSH003) (ONT) was used to reactivate pores and boost instrument yields. The sequencer was run as described but paused after 4 hours. The wash mix was prepared by combining 2µl of wash solution (WMX) (ONT) with 398µl diluent (DIL) (ONT). The waste liquid was withdrawn from the waste port and discarded before 400µl of the wash mix was loaded to the priming port and left for 60 minutes. This was then removed from the waste port, followed by loading further priming mix through the priming port and more library was loaded into the SpotON port, following the steps in section 2.14.2.2.

2.14.2.4 Flongle sequencing

Long read sequencing was carried out on two types of nanopore flow cells. To sequence on Flongle, a separate sequencing library was created for each sample, using half volumes of the above-described end-prep and ligation reactions (Section 2.14.2.1). A Flongle flow cell (R.9.4.1) was next prepared for sequencing by loading 120µl priming mix (3µl of Flush Tether (FT) (ONT) and 117µl of Flush Buffer (FB) (ONT). 3-20 fmol of the library was then combined with 15µl of sequencing buffer (SQB) (ONT) and 10µl of loading beads (LB) (ONT) prior to loading onto the flow cell. A 24-hour Flongle sequencing run was initiated using MinKNOW software (v.3.6.0; ONT).

2.14.2.5 Data analysis & bioinformatics

Basecalling and sample demultiplexing was performed to convert the raw data from fast5 to FASTQ format using Guppy (v.5.0.16; https://nanoporetech.com). Trimming the adaptor sequences from the resulting reads was performed using Porechop (v.0.2.4; https://github.com/rrwick/Porechop).

NanoFilt (v.2.8.0;https://github.com/wdecoster/nanofilt; (De Coster et al., 2018)) was used to remove low quality reads based on their length and quality (Q score). Filtered reads were then aligned to the human reference genome (GRCh37/hg19) using Minimap2 (v.2.18; https://github.com/lh3/minimap2; (Li, 2018)). NanoPolish (v.0.11.0; https://nanopolish.readthedocs.io) was used to generate a VCF file, which was then combined with WhatsHap (v.0.17; https://whatshap.readthedocs.io/; (Martin et al., 2016)) to group individual reads according to their variant-specified haplotype. SAM to BAM file conversion, read sorting, and BAM file indexing was performed using Samtools (v.1.12; (Li et al., 2009)). NanoStat (v.1.5; https://github.com/wdecoster/nanostat ; (De Coster et al., 2018)) was used to calculate read metrics and statistics. BAM files were visualised using the Integrative Genomics Viewer (IGV; v.2.11.2; Section 2.13.5). Commands used in MinION analysis are listed in Appendix A.3.

2.15 Manchester reference laboratory Sanger sequencing of ORF15 (Chapter 5)

Four primer pairs (RPGR_Ex15-1F/ RPGR_Ex15-1R, RPGR_Ex15-2F/ RPGR_Ex15-2R, RPGR_Ex15-3F/ RPGR_Ex15-3R, and RPGR_Ex15-4F/ RPGR_Ex15-4R) were used to sequence the *ORF15* region of *RPGR*. The sequence-specific primers are listed in Appendix A.4. All of the RPGR exon *ORF15* primers were tailed with N13 tags (forward: GTAGCGCGACGGCCAGT and reverse: CAGGGCGCAGCGATGAC). The PCR mix used for primer pairs RPGR_Ex15-1F/ RPGR_Ex15-1R and RPGR_Ex15-4F/ RPGR_Ex15-4R consisted of 10µl Go*Taq* master mix (Promega, Madison, Wisconsin, USA), 2µl primer mix (final concentration 500nM), 2µl of genomic DNA, and 6µl nuclease free water. The PCR mix used for primer pairs RPGR_Ex15-2F/ RPGR_Ex15-2R and RPGR_Ex15-3F/ RPGR_Ex15-3R comprised 2µl of 10X PCR buffer minus $MgCl_2$ (Invitrogen), 1µl of 20 mM dNTP mix, 0.5µl of 50 mM $MgCl_2$, 1.20µl primer mix (final concentration 300nM), 0.20µl Platinum *Taq* DNA polymerase (Invitrogen), 2µl of genomic DNA, and 13.10µl nuclease free water. Thermocycling conditions for these PCR reactions are recorded in Table 2.10. Sanger sequencing reactions are recorded in Table

2.11. The sequencing run was performed on an ABI 3730 Genetic analyser (Applied Biosystem) and the sequences produced were analysed on sequence scanner software (v2.0; Applied Biosystem).

No. of cycles	**PCR stage**	**Temperature**	**Time**
1	Initial denaturation	95 °C	5 minutes
35 cycles	Denaturation	95 °C	30 seconds
	Annealing	60 °C	30 seconds
	Extension	72°C	2 minutes.
1	Final extension	72°C	5 minutes.
1	Hold	4 °C	∞

Table 2.10: Thermocycling conditions applied in *RPGR-ORF15* PCR for Sanger sequencing.

Primer	*RPGR*_Ex15-1 F+R *RPGR*_Ex15-2 F *RPGR*_Ex15-3 F+R	*RPGR*_Ex15-2 R	*RPGR*_Ex15-4 F+R
Component	X1 (µl)	X1 (µl)	X1 (µl)
Big Dye v3.1	1	4	0.25
Diluent	3		3.75
Nuclease free H_2O	3		3
Primer	2	2	2
Ampure	1	4	1

Table 2.11: *RPGR ORF15* Sanger sequencing reactions.

Chapter 3

Using Whole Exome Sequencing (WES) to identify pathogenic variants in previously unscreened IRD individuals

3.1 Introduction

Since the discovery of DNA structure (Watson and Crick, 1953) and the means to sequence DNA (Sanger et al., 1977b), significant progress has been made in understanding the complexity and diversity of human genomes in health and disease. Although technological breakthroughs in instruments and reagents supported the commencement of the human genome project, the first sequence of the human genome took about 13 years to be released (Lander et al., 2001). Because of the high cost and low throughput, more improved technologies were needed to allow scientists to sequence in less time and at a cheaper cost. With the emergence of next (second) generation sequencing NGS in 2004 (Pettersson et al., 2009), large-scale sequencing studies can now be undertaken, the full extent of natural human DNA variations can be quantified and disease-causing variants can now be identified more quickly and efficiently (Warr et al., 2015).

To find these variants, researchers can use whole gene sequencing (GS), whole genome sequencing (WGS), whole exome sequencing (WES) or targeted panels (Custom panels), as described in Section 1.9.2.1 to sequence the whole gene, the entire genome or a portion of it (Schwarze et al., 2018). WGS can detect all nucleotides in human genomic DNA and find variations in any area of the genome, but it needs more sequencing reagents and generates a big dataset, increasing the time, cost of analysis and data storage. WES, on the other hand, is a commonly used NGS approach that uses fewer sequencing chemicals and requires less time for bioinformatic analysis. It is an option to consider as it allows us to sequence only the coding regions, which are only ~1% of the genome but around 85% of disease-causing mutations are believed to be found in these protein coding regions (exons) (Choi et al., 2009, Van Dijk et al., 2014). Alternatively, GS can be used to carry out detailed analysis of a single gene and the surrounding locus, as has been done for *ABCA4* (Khan et al., 2020).

This chapter describes the screening of a cohort of 24 inherited retinal disease (IRD) cases in an attempt to identify the pathogenic variants in each case. These individuals had been diagnosed with a variety of IRDs, and some of them had ocular phenotypes that were accompanied by symptoms in other organs or tissues (Table 3.1). No previous genetic

screening had been carried out in these patients. Variants identified as likely to be pathogenic in this study will be reported to the ClinVar database to make them available to the IRD community.

However, the analyses described here are considered a research based preliminary screen. NHS genetic screening in the UK is carried out by accredited diagnostic laboratories using strictly regulated standard operating procedures. As these analyses did not follow such procedures, they are given to the relevant clinicians as research results, with no guarantee and with advice that they should be confirmed by a diagnostic laboratory. Furthermore, these analyses are not considered comprehensive. Filtering strategies used are deliberately relatively strict, and based on assumptions about inheritance patterns and clinical information that cannot be considered 100% certain. Where possible, the cases unsolved herein will be looked at again as analysis pipelines improve, to examine, for example, synonymous variants to see if they influence splicing, and to look at all variants rather than focussing on known IRD genes. They may also be subject to WGS for more comprehensive analyses of intron, promotor, and structural variants. The aim of this study was to provide an initial screen to identify “low hanging fruit”, so that later analyses could be more targeted.

NO.	SAMPLE ID	GENDER	AGE AT SAMPLING	ETHNICITY (COUNTRY OF ORIGIN)	INHERITANCE PATTERN	PHENOTYPE
1.	F1415	Male	37 y	South European (Greece)	Recessive	IRD + Cataract (posterior subcapsular)
2.	3869	Male	25 y	North European (UK)	Dominant	RP
3.	3800	Male	51 y	North European (UK)	Dominant	RP
4.	4211	Female	71 y	North European (UK)	Dominant	RP
5.	4212	Male	39 y	North European (UK)	Dominant	RP
6.	F1427	Female	28 y	North European (UK)	Dominant	FEVR
7.	OA1201	Female	9 y	South Asian (UK)	Recessive	OA
8.	OA1203	Female	16 y	South Asian (UK)	Recessive	OA
9.	12565367	Female	22 months	Arab (Oman)	Recessive	Syndromic: minor squint, visual impairment, and developmental delay
10.	12642093	Female	3 months	Arab (Oman)	Recessive	LCA
11.	11335421	Male	6 months	Arab (Oman)	Recessive	CORD
12.	11867145	Female	2 y	Arab (Oman)	Recessive	BBS
13.	11795135	Female	10 months	Arab (Oman)	Unknown	Syndromic: Cardiac and vision problems
14.	11935460	Male	9 y	Arab (Oman)	Recessive	RP and sensorineural hearing loss

15.	12194201	Male	9 months	Arab (Oman)	Unknown	Microcephaly, vision problems, Developmental Delay
16.	8956112	Male	8 y	Arab (Oman)	Unknown	Microdeletion syndrome??
17.	5543	Male	60 y	North European (UK)	Unknown	CORD
18.	5544	Female	45 y	African (UK)	Unknown	CORD
19.	5545	Female	55 y	North European (UK)	Unknown	PED with subsequent macular GA.
20.	5546	Male	55 y	North European (UK)	Unknown	AVMD
21.	3558	Male	74 y	North European (UK)	Recessive	RP
22.	3727	Male	64 y	North European (UK)	Unknown	MD
23.	5582	Female	63 y	North European (UK)	Unknown	MD
24.	5583	Female	58 y	North European (UK)	X-linked	RP

Table 3.1: Summery of clinical data and sources of 24 cases analysed by WES in search of disease-causing variants. (RP, retinitis pigmentosa; CORD, cone rod dystrophy; MD, macular dystrophy; FEVR, familial exudative vitreoretinopathy; OA, optic atrophy; LCA, Leber congenital amaurosis; BBS, Bardet-Biedl syndrome; AVMD, adult vitelliform macular dystrophy; PED, pigment epithelial detachment; GA, geographical atrophy).

3.2 Results

3.2.1 Whole exome sequencing (WES)

Twenty-four, previously unscreened, IRD cases were analysed by WES to try and identify their disease-causing variants (Table 3.1). All samples were quantified using Qubit fluorometer (Section 2.3.1) before undergoing WES by the Leeds University NGS facility using the SureSelect^{XT} Human All Exon V6 library capture reagent and Illumina HiSeq 3000 Sequencer (Sections 2.10.1). Samples were pooled and run on two lanes of the sequencer; the 1st pool included 13 samples while the 2nd included 11 samples.

After receiving the sequencing data back from the NGS facility, the raw sequence data for each sample was checked for quality control, and all passed (Section 2.10.2). Figure 3.1 shows a representative FastQC analysis report from one sample.

The samples were then processed using the bioinformatics pipeline outlined in Section 2.10.2. Variant prioritisation was based on the pipeline outlined in Figure 3.2. Briefly, all variants with a minor allele frequency (MAF) of greater than or equal to 1%, or in dominant cases greater than or equal to 0.01%, together with all synonymous variants and variants outside exons and their splice site regions (within 2bp of a splicing junction) were excluded from the resulting variants list. Only variants in coding DNA and those in splice donor and acceptor sites (±2 bp) were chosen, and only those with depth of coverage greater than ten reads and a CADD score of 15 or above were prioritised. Further filtration was carried out on a family/case specific basis, based on zygosity and likely mode of inheritance. Several additional pathogenicity prediction tools were used to further assess the results (Section 2.13.4). In addition, the R package ExomeDepth was used to look for copy number variations (CNVs) in the samples (Section 2.11).

The final variant lists in each family/case were compared to information on known genes associated with retinal diseases in the RetNet database (Accessed on February 2022) (Appendix B.1) in cases with clear IRD diagnosis, and to literature on function of the encoded protein and phenotypes associated with variants in it via PubMed searches. Genes and variants were also investigated further in dbSNP and gnomAD databases to check the allele frequency, zygosity in controls and probability of loss-of-function (pLI) scores. The top candidate variants were finally confirmed by Sanger sequencing (primer sequences are listed in Appendix B.2).

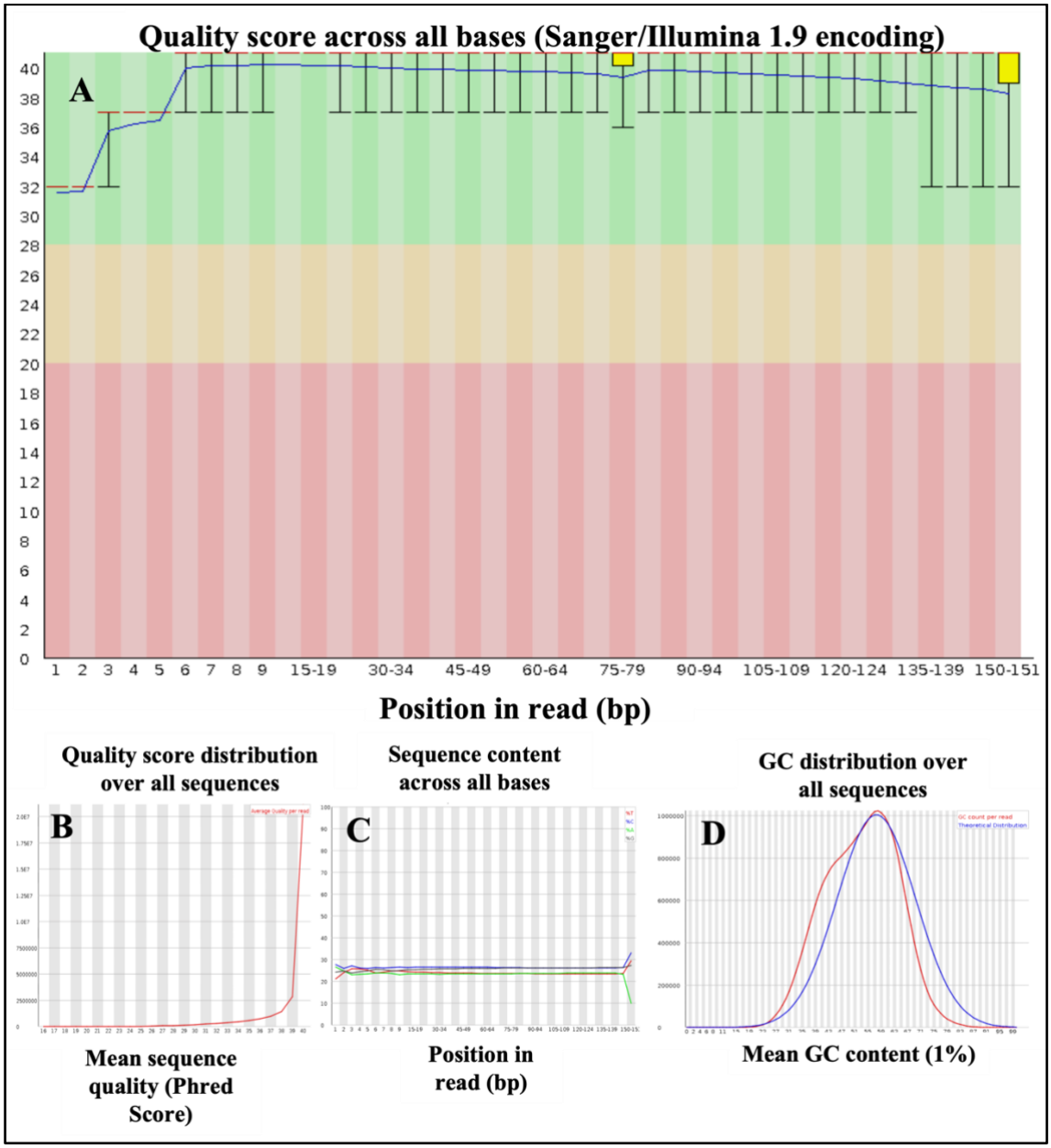


Figure 3.1: FastQC analysis report for sample 3800. A. Per base sequence quality. The central red line is the median value, the yellow box denotes the interquartile range (25-75%), the upper and lower whiskers represent the 10% and 90% points, and the blue line represents the mean quality. The Y-axis on the graph shows the quality scores (PHRED), a quality score of 20 means 99% base call accuracy, while a score of 30 means 99.9% accuracy. The background of the graph is divided into three colours: very good quality calls "green", reasonable quality calls "orange" and poor-quality calls "red". Sample 3800 had very good quality calls with zero poor quality calls from a total of 29,834,253 sequences. **B. Per sequence quality score.** Sample 3800 had 39 average quality per sequence. **C. Per base sequence content.** Sample 3800 had little to no difference between the different bases of a sequence run. **D. Per sequence GC content.** Sample 3800 had normal distribution of GC content over all sequences.

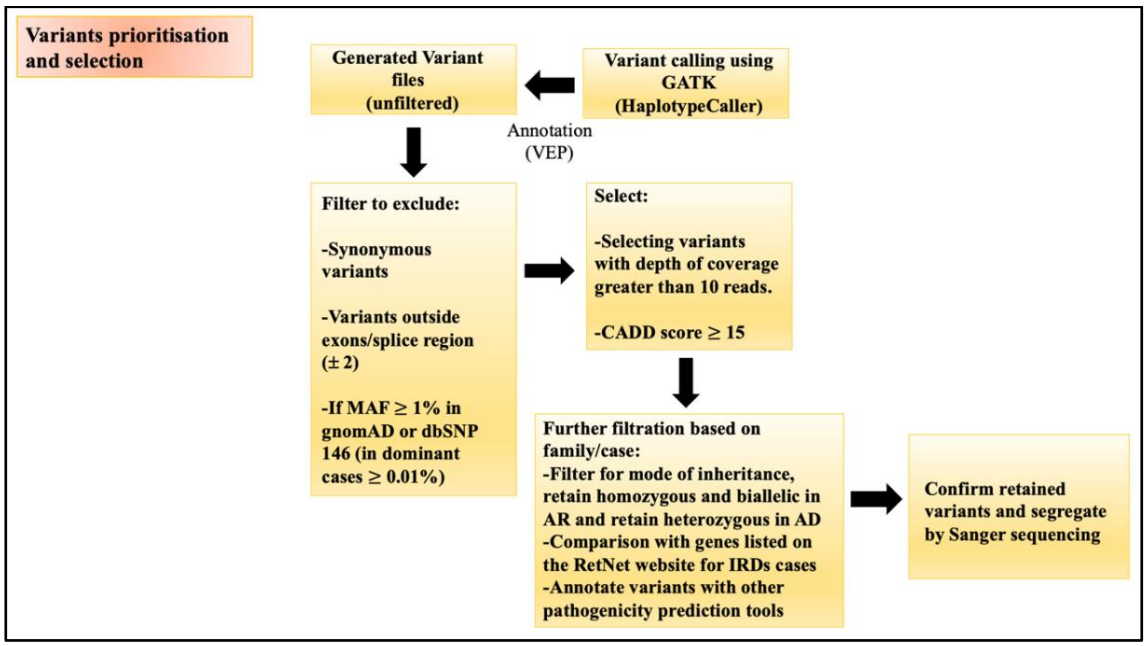


Figure 3.2: Schematic representation of variants detection and prioritisation pipeline.

3.2.2 Solved cases- variants identified in genes associated with IRDs

3.2.2.1 WES analysis of sample F1415 identifies a homozygous mutation in *CDHR1* causing non syndromic IRD

A male patient (F1415) and his sister (F1416) were both reported to have vision reduced to light perception only, with age at onset said to be school age. They were diagnosed with inherited retinal dystrophy (IRD) and posterior capsular cataract. Their parents' clinical data was not reported but both parents were said to be normal. They also have a female cousin (F1417) who had similar symptoms and whose two male children who were not checked but showed no symptoms (F1413 & F1414) (Figure 3.3).

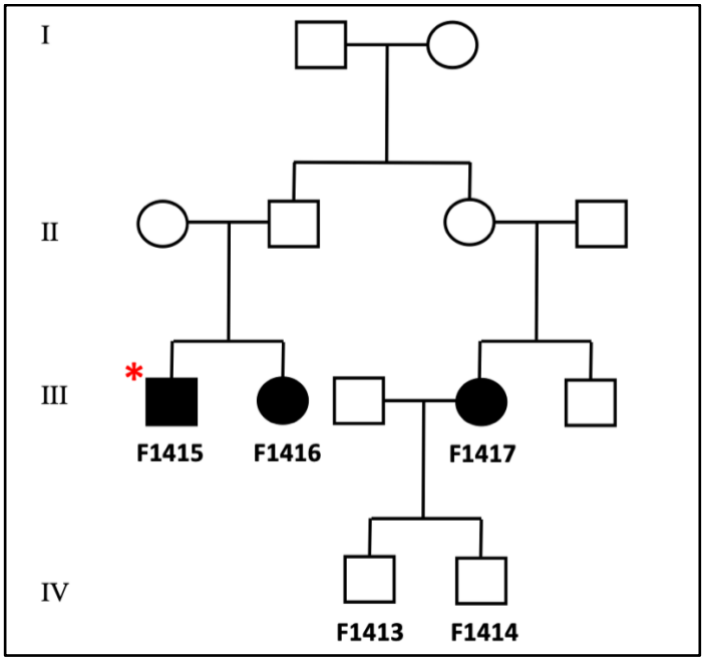


Figure 3.3: Pedigree of case F1415 family. The pedigree shows four generations of the family. The individuals from whom DNA was available are numbered with the affected individuals coloured black. (* = DNA was analysed by WES).

A recessive mode of inheritance was considered the most likely explanation in this family because the parents and affected cousin's offspring are unaffected (Figure 3.3) and members of this family live on a relatively small Greek island, suggesting possible endogamy. However, dominant inheritance with incomplete penetrance cannot be excluded. Homozygous or biallelic variants were prioritised in the VCF file as a first pass analysis. After screening for potential pathogenic variants in the known IRD genes listed in RetNet website (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (Accessed February 2022), it was noted that there is a homozygous nonsense (stop gained) mutation (GRCh37, Chr.10: 85971445, NM_033100, c.1527T>G, p.Y509*) in exon 14 of *CDHR1*

(Cadherin-related family member 1; OMIM 609502)). Mutations in *CDHR1* are known to cause autosomal recessive cone rod dystrophy (CORD) (Bolz et al., 2005, Henderson et al., 2010, Ostergaard et al., 2010). According to MutationTaster, the variant is predicted to be disease-causing and has a CADD score of 35, indicating that it is among the top 0.1 percent of the most deleterious mutations in the human genome. It was reported in dbSNP (rs1477733493) and found in gnomAD at a frequency of 0.0004% (1/251490).

To confirm the mutation and segregate it in the family, primer sequences were selected (given in Appendix B.2) to PCR across the variant and Sanger sequence five members of the family (F1413, F1414, F1415, F1416, and F1417). This showed that the *CDHR1* variant segregated with the disease phenotype in a manner consistent with recessive inheritance (Figure 3.3 and Figure 3.4) and revealed that both F1413 and F1414 are carriers. Furthermore, this nonsense mutation was reported previously in a homozygous status which strongly suggest that the IRD in this patient is indeed autosomal recessive (Arno et al., 2016, Carss et al., 2017) and that the *CDHR1* mutation is almost certainly the cause of this family's IRD phenotype.

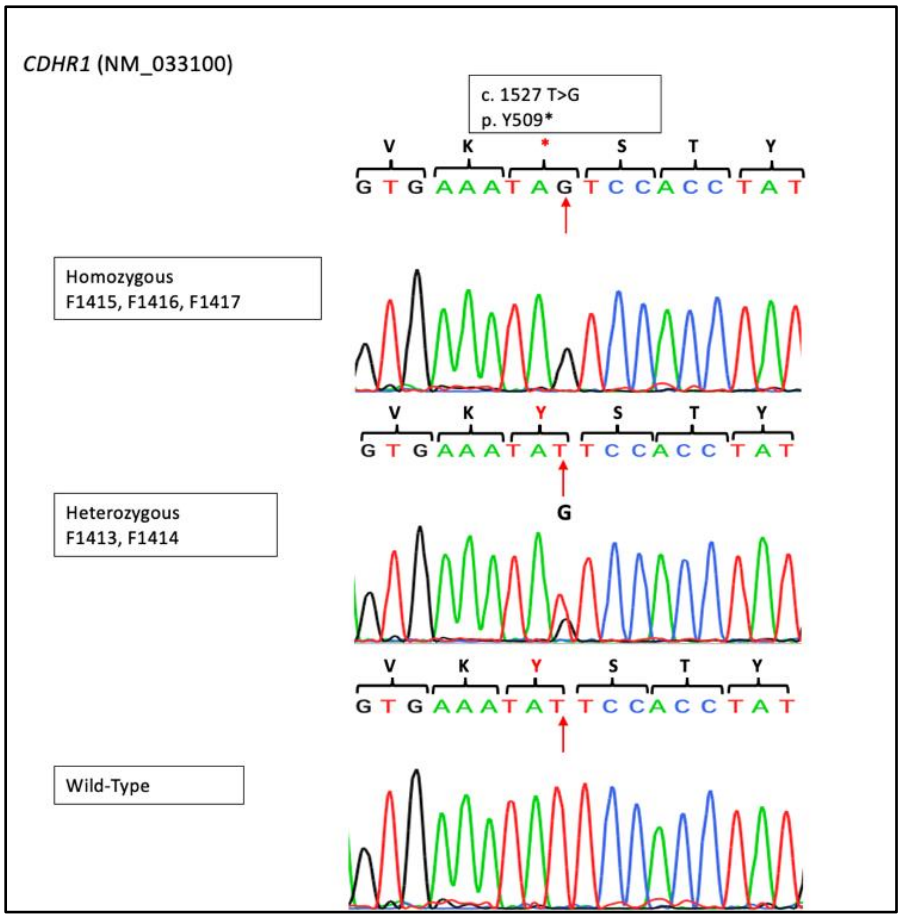


Figure 3.4: *CDHR1* mutation in sample F1415 and other family members. Sanger sequencing electropherograms for mutant (homozygous), carrier (heterozygous), and wild-type (healthy) sequences for the *CDHR1* variant (NM_033100, c. 1527T>G, p.Y509*).

3.2.2.2 WES analysis of sample 3869 identifies a heterozygous mutation in *RHO* causing autosomal dominant retinitis pigmentosa (ADRP)

A case of autosomal dominant retinitis pigmentosa (ADRP) (3869) was sent from Hull & East Yorkshire hospital to be investigated further to identify the causal variant. Heterozygous variants were prioritised and after screening the list obtained for genes known to be involved in IRDs, as listed in RetNet website (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (Accessed February 2022), two heterozygous variants were identified in two IRDs related genes (Table 3.2).

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
3	129247860	*RHO*	Het Missense (25.7)	NM_000539.3: c.284T>C, p.L95P
8	10467651	*RP1L1*	Het Frameshift (24.2)	NM_178857.6: c.3956_3957insAAGAAGAGGG, p.V1320Rfs*10

Table 3.2: List of candidate variants in case 3869 (male) after alignment, variant calling, filtering, and comparison with genes listed on the RetNet website. Chr. = chromosome, Het = heterozygous.

A heterozygous missense variant was identified in exon 1 of the *RHO* gene (GRCh37, Chr.3: 129247860, NM_000539, c.284T>C, p.L95P). Mutations in *RHO* (Rhodopsin; OMIM 180380) are a known cause of IRD and are reported to cause ~ 30% of autosomal dominant RP (McWilliam et al., 1989, Dryja et al., 1990a, Dryja et al., 1990b, Al-Maghtheh et al., 1993). The mutation is not present in the databases gnomAD (0/282650 (number of alleles for the closest variant)) and dbSNP. It has a CADD score of 25.7 and is predicted to be pathogenic using five additional pathogenicity prediction tools (Table 3.3). It has been published as disease-causing in a previous study (Roshandel et al., 2019) and is present in the *RHO* LOVD database (https://databases.lovd.nl/shared/genes/RHO).

Variant (Protein change)	**CADD score**	**SIFT**	**MutationTaster**	**PolyPhen2**	**BLOSUM62**	**Franklin**
c.284T>C (p.L95P)	25.7	Deleterious (0)	Disease causing	Probably damaging (0.942)	-3	Likely pathogenic

Table 3.3 : Pathogenic prediction scores for *RHO* mutation in sample 3869. The scores for *RHO* missense variant (NM_000539, c.284T>C, p.L95P) from six different pathogenicity prediction tools are shown.

To assess the evolutionary conservation of the affected amino acid, multiple protein alignments were made using HomoloGene (https://www.ncbi.nlm.nih.gov/homologene). The normal amino acid is conserved in all available species apart from chicken and frog (Figure 3.5).

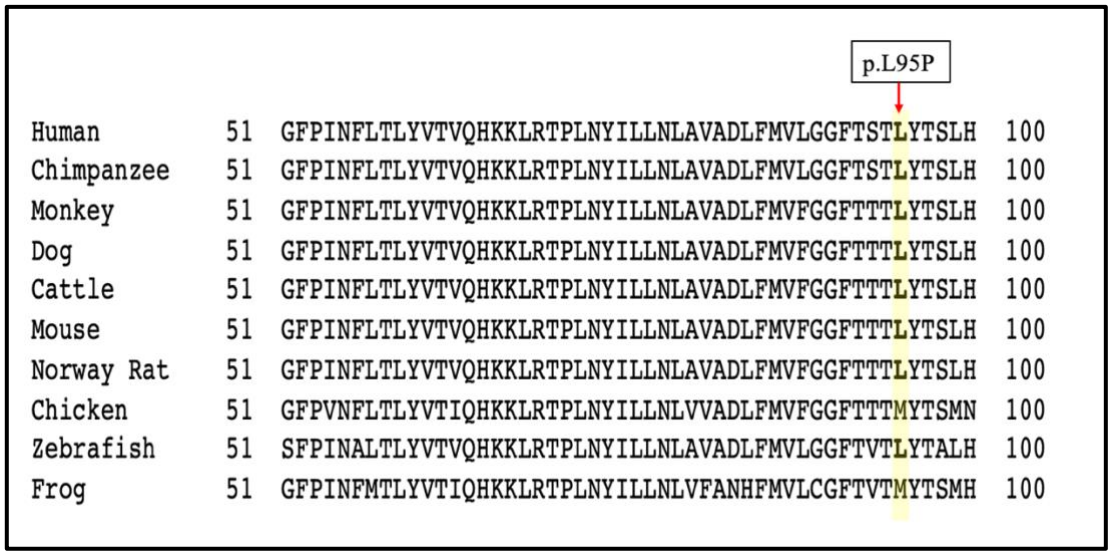


Figure 3.5: Amino acids sequence alignments of human *RHO* with homologues. Normal amino acid sequence of RHO in various species (HomoloGene, https://www.ncbi.nlm.nih.gov/homologene) showing the evolutionary conservation of leucine residue at position 95 in mammals.

Sanger sequencing validated the mutation in this case (Figure 3.6), but no segregation study was undertaken due to the lack of DNA from additional family members. Another heterozygous frameshift variant was identified in the *RP1L1* gene. Heterozygous variants in *RP1L1* are known to cause autosomal dominant macular degeneration (Akahori et al., 2010) while homozygous variants are associated with autosomal recessive RP (Davidson et al., 2013). However, since the patient here has autosomal dominant RP, it is considered highly likely that the ADRP in this case is due to the missense variant (NM_000539, c.284T>C, p.L95P) in the *RHO* gene.

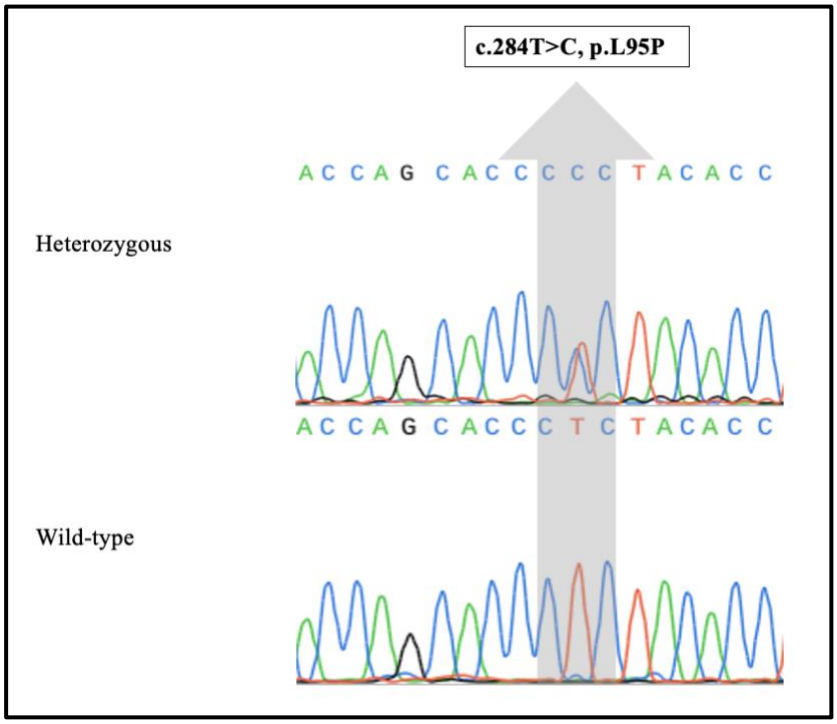


Figure 3.6: *RHO* mutation in sample 3869. Sanger sequencing electropherograms for mutant (heterozygous) and wild-type (healthy) sequences for the *RHO* missense variant (NM_000539, c.284T>C, p.L95P).

3.2.2.3 WES analysis of sample 4211 identifies a heterozygous mutation in *PRPF31* causing autosomal dominant retinitis pigmentosa (ADRP)

Another autosomal dominant retinitis pigmentosa (ADRP) case (4211) was sent from Hull and East Yorkshire hospital, and the genomic DNA was sequenced by WES to identify the disease-causing mutation. Heterozygous variants were prioritised and filtered against the genes listed in RetNet website (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (Accessed February 2022). Four candidate genes remained (Table 3.4) and based on the given inheritance pattern, *MYO7A*, *GRK1*, and *GRM6* genes were excluded because they are all known to be implicated in autosomal recessive diseases.

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
19	54627976	*PRPF31*	Het Frameshift (32)	NM_000539.3: c.797delC, p.S266*
11	76867967	*MYO7A*	Het Missense (27.4)	NM_000260.4: c.652G>A, p.D218N
13	114438091	*GRK1*	Het Missense (24.1)	NM_002929.3: c.1447G>T, p.D483Y
5	178413684	*GRM6*	Het Missense (22)	NM_000843.4: c.1571C>T, p.P524L

Table 3.4: List of candidate variants in case 4211 (female) after alignment, variant calling, filtering, and comparison with genes listed on the RetNet website. Chr. = chromosome, Het = heterozygous.

A heterozygous frameshift mutation (GRCh37, Chr.19: 54627976, NM_015629, c.797delC, p.S266*) was identified in exon 8 of *PRPF31* (Pre-mRNA-Processing Factor 31; OMIM 606419). *PRPF31* is a gene known to be mutated in patients with ADRP (Al-Maghtheh et al., 1996, Sullivan et al., 2006b, Venturini et al., 2012). This variant is absent from dbSNP and gnomAD databases (0/31404 (number of alleles for the closest variant)) and has a CADD score of 32 (top 0.1% of the most deleterious mutations) and predicted to be disease-causing on MutationTaster.

The presence of the variant was confirmed by Sanger sequencing, but no segregation analysis was done due to the lack of DNA from other family members (Figure 3.7). This variant has not been published previously and is absent from the *PRPF31* LOVD database. Further evidence could be obtained by testing for this variant in other family members to segregate this novel variant, which would also be of clinical benefit to the family, and it should then be reported to the IRD community. This was not completed due to lack of time or access to the necessary samples. Nevertheless, given that

heterozygous nonsense mutations in *PRPF31* are a well-known cause of dominant RP, it is considered highly likely that this variant is the cause of disease in this patient.

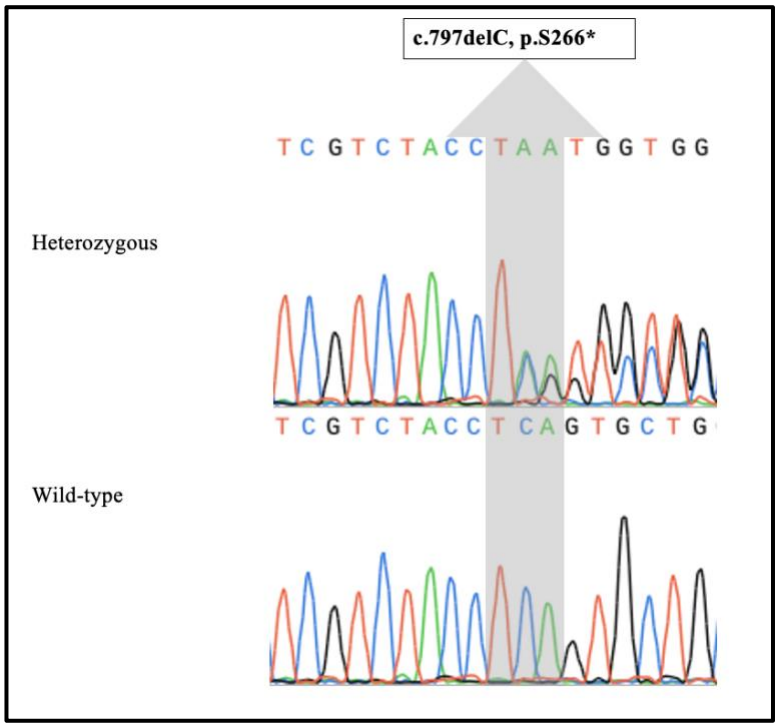


Figure 3.7: *PRPF31* mutation in sample 4211. Sanger sequencing electropherograms for mutant (heterozygous) and wild-type (healthy) sequences for the *PRPF31* frameshift variant (NM_015629, c.797delC, p.S266*).

3.2.2.4 WES analysis of sample 11335421 identifies a homozygous mutation in *CNGA3* causing cone rod dystrophy CORD

A male patient (11335421) was diagnosed with cone rod dystrophy (CORD) when his vision began to deteriorate at the age of four. In Oman, the diagnosis was made primarily on the FFERG results, which revealed drastically decreased rod and cone functions. This patient came from a consanguineous family and had two paternal cousins with similar symptoms. Other family members' DNA samples and clinical information were unavailable.

The mode of inheritance was assumed to be recessive based on the known consanguinity in the family, homozygous variants were prioritised and then filtered against the genes listed in RetNet website (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (Accessed February 2022). There was only one homozygous missense change, in exon 8 of *CNGA3* (GRCh37, Chr.2: 99012721, NM_001298, c.1088T>C, p.L363P). Recessive mutations in *CNGA3* (Cyclic Nucleotide Gated Channel Alpha 3; OMIM 600053) cause autosomal recessive achromatopsia (Arbour et al., 1997, Nishiguchi et al., 2005), and autosomal recessive cone rod dystrophy (CORD) (Nishiguchi et al., 2005, Li et al., 2014). The *CNGA3* missense variant identified is not detected in dbSNP and gnomAD databases (0/251156 (number of alleles for the closest variant)) and has been reported previously in homozygous state in patients with achromatopsia (Koeppen et al., 2010). It is predicted to be pathogenic according to different pathogenicity prediction tools (Table 3.5), with a CADD score of 25.1 (top 1% most deleterious mutations).

Variant (Protein change)	CADD score	SIFT	MutationTaster	PolyPhen2	BLOSUM62	Franklin
c.1088T>C (p.L363P)	25.1	Deleterious (0)	Disease causing	Probably damaging (0.998)	-3	Likely pathogenic

Table 3.5: Pathogenicity prediction scores for the *CNGA3* variant in sample 11335421. The Scores for *CNGA3* missense variant (NM_001298, c.1088T>C, p.L363P) from six different pathogenicity prediction tools are shown.

Sanger sequencing (Figure 3.8A) confirmed the mutation, and the multiple protein alignment website, HomoloGene (https://www.ncbi.nlm.nih.gov/homologene), was used to determine the evolutionary conservation, with the normal amino acid being entirely conserved across all available species (Figure 3.8B). This variant was therefore considered highly likely to be the cause of disease in this case.

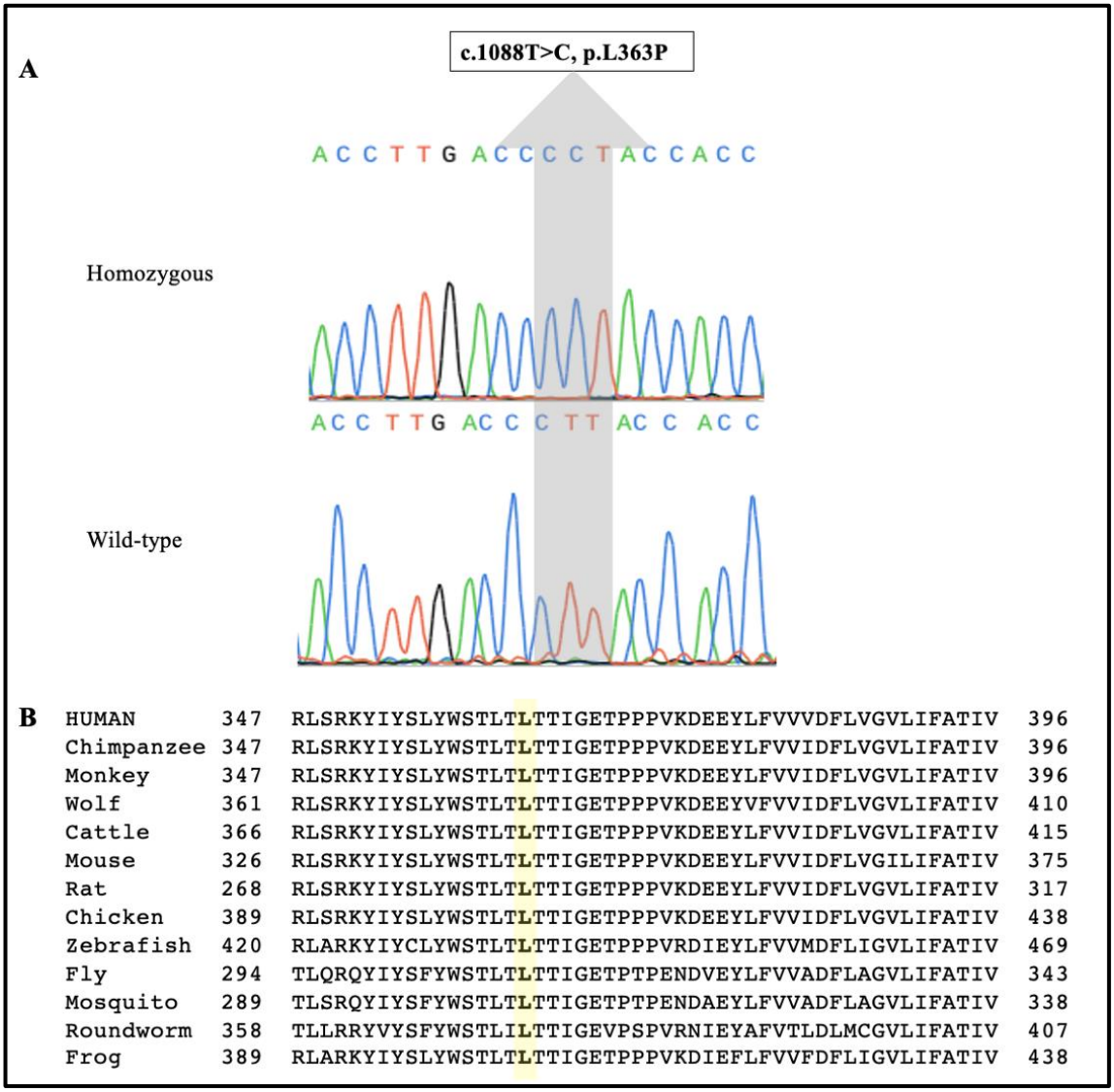


Figure 3.8: Analysis of *CNGA3* mutation in sample 11335421. A. Sanger sequencing electropherograms for mutant (homozygous) and wild-type (healthy) sequences for the *CNGA3* missense variant (NM_001298, c.1088T>C, p.L363P). **B.** Normal amino acid sequence alignments of CNGA3 (HomoloGene, https://www.ncbi.nlm.nih.gov/homologene) showing the evolutionary conservation of the leucine residue at position 363.

3.2.2.5 WES analysis of sample 11867145 identifies a homozygous mutation in *BBS10* causing Bardet-Biedl syndrome (BBS)

A female patient (11867145) from a consanguineous family of Omani origin was recruited from Oman for further investigation. She was one year and eight months old at the time of the initial assessment and was diagnosed with Bardet-Biedl syndrome. In addition to her vision problem, she also has dextrocardia, situs inversus, cardiac abnormalities, echogenic kidney, postaxial polydactyly in both hands and left foot, and slight developmental delay with no hearing problems. The patient's cousin (male) from her mother side had similar symptoms and was diagnosed with Bardet-Biedl syndrome.

The genomic DNA was only available for patient 11867145, and it was analysed using WES. The variant list was filtered as described in Section 3.2.1. Based on the family history and phenotype, homozygous variants were prioritised and compared against known IRDs genes (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (accessed February 2022). The patient was found to carry a homozygous frameshift mutation (GRCh37, Chr.12: 76741033, NM_024685, c.728-731delAAGA, p. K243Ifs*15) in exon 2 of gene *BBS10* (also known as *C12orf58* (Chromosome 12 Open Reading Frame 58); OMIM 610148). Mutations in *BBS10* (*C12orf58)* were reported to cause autosomal recessive Bardet-Biedl syndrome (Stoetzel et al., 2006, White et al., 2007). This mutation was reported in a homozygous state in 50 black South African families with Bardet-Biedl syndrome, indicating that BBS is more common in this ethnic group (Fieggen et al., 2016). The mutation is predicted to be disease-causing by MutationTaster and has a CADD score of 34 which means that it is in the top 0.1% of the most deleterious mutations in human genome. It is not present in gnomAD database (0/250484 (number of alleles for the closest variant)). However, it was found in dbSNP database (rs786204671). This mutation was confirmed by Sanger sequencing (Figure 3.9) and is considered highly likely to be the cause of disease in this patient.

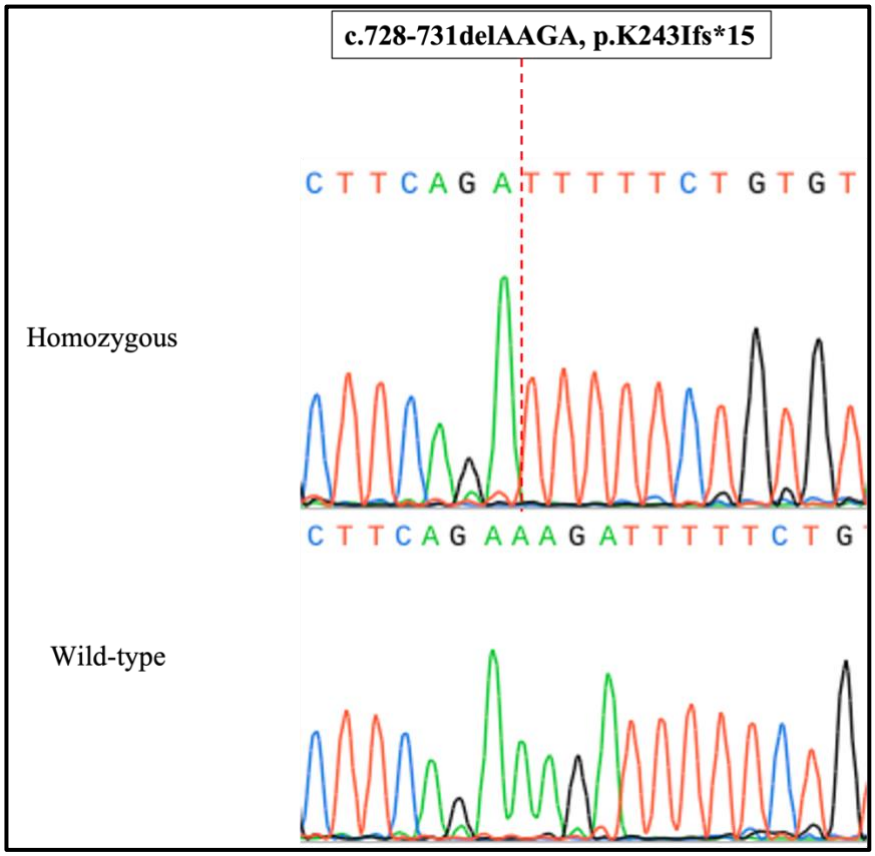


Figure 3.9: *BBS10* mutation in sample 11867145. Sanger sequencing electropherograms for mutant (homozygous) and wild-type (healthy) sequences for the *BBS10* frameshift variant (NM_024685, c.728-731delAAGA, p. K243Ifs*15).

3.2.2.6 WES analysis of sample 11935460 identifies a homozygous mutation in *USH2A* causing RP and sensorineural hearing loss

Sample 11935460 is from a male of Omani origin from a consanguineous family. He was initially assessed as a nine-year-old and diagnosed with retinitis pigmentosa (RP) at a local eye clinic. Apart from his vision problem, he was also diagnosed with sensorineural hearing loss and learning difficulties. No family history of similar symptoms was reported, and his siblings are healthy. Based on the consanguinity in the family, homozygous variants were prioritised. Two homozygous variants were found in two genes implicated in IRDs, *IFT172* and *USH2A* (Table 3.6).

A homozygous missense change (GRCh37, Chr.1: 216595427, NM_206933, c.252T>G, p.C84W) in exon 2 of gene *USH2A* (Usherin; OMIM 608400) was further investigated since the homozygous mutations in *USH2A* can cause autosomal recessive retinitis pigmentosa and recessive Usher syndrome type IIa (Seyedahmadi et al., 2004, Gao et al., 2021).

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
1	216595427	*USH2A*	Homo Missense (23.5)	NM_206933.4: c.252T>G, p.C84W
2	27682663	*IFT172*	Homo Missense (23.5)	NM_015662.3: c.2555A>T, p.E852V

Table 3.6: List of candidate variants in case 11935460 (male) after alignment, variant calling, filtering, and comparison against genes listed on the RetNet website. Chr. = chromosome, Homo = homozygous.

The identified missense mutation in *USH2A* is not found in dbSNP or gnomAD databases (0/250894). In all known species, the normal amino acid residue is totally conserved (Figure 3.10B), and the substitution was predicted to be disease-causing by five separate prediction tools (Table 3.7).

Variant (Protein change)	**CADD score**	**SIFT**	**MutationTaster**	**PolyPhen2**	**BLOSUM62**	**Franklin**
c.252T>G (p.C84W)	23.5	Deleterious (0)	Disease causing	Possibly damaging (0.972)	-2	VUS (ACMG; PM2, PM1)

Table 3.7: Pathogenic prediction scores for *USH2A* mutation in sample 11935460. The scores for *USH2A* missense variant **(**NM_206933**,** c.252T>G, p.C84W) from six different pathogenicity prediction tools are shown. VUS = variant of uncertain significance.

Primers (Appendix B.2) were designed to PCR across the candidate variant and the amplimer obtained was Sanger sequenced to confirm the presence of the mutation (Figure 3.10A).

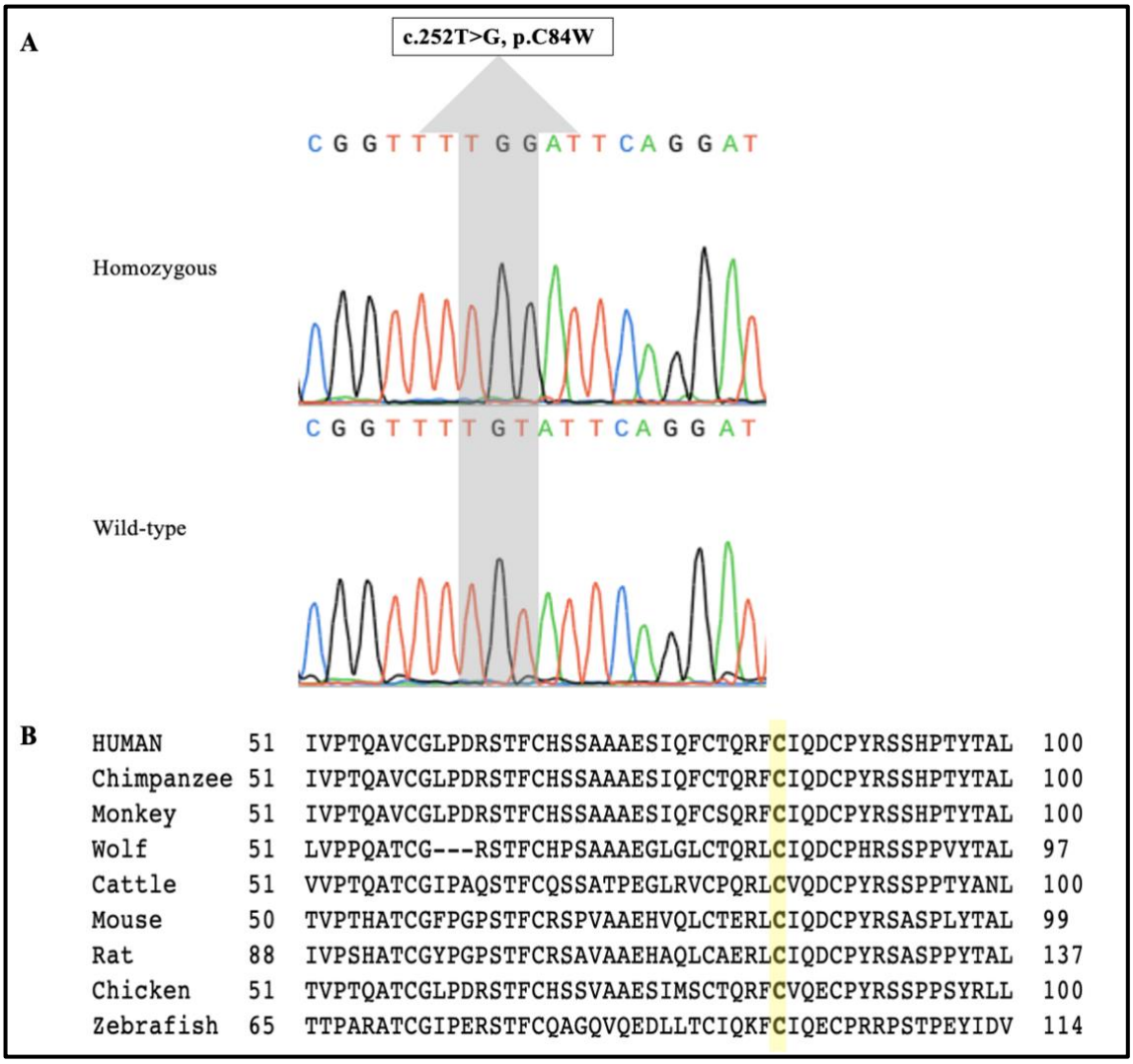


Figure 3.10: Analysis of *USH2A* mutation in sample 11935460. A. Sanger sequencing electropherograms for mutant (homozygous) and wild-type (healthy) sequences for the *USH2A* missense variant (NM_206933, c.252T>G, p.C84W). **B.** Normal amino acid sequence alignments of USH2A (HomoloGene, https://www.ncbi.nlm.nih.gov/homologene) showing the evolutionary conservation of the cysteine residue at position 84.

This patient has sensorineural hearing loss in addition to the RP meaning that he most probably has Usher syndrome. Based on these findings, the *USH2A* variant is considered the more likely of the two variants to be pathogenic in this case, causing RP and sensorineural hearing loss in this individual. However, since many IRD cases are not clear cut, and different variants in the same IRD genes can cause different phenotypes, all options should be considered. The second homozygous missense change (GRCh37, Chr.2: 27682663, NM_015662.3, c.2555A>T, p.E852V) in exon 24 of the *IFT172* gene (Intraflagellar transport 172; OMIM 607386) should therefore also be considered as a potential causative variant for the phenotype in this patient. It is not found in dbSNP or

gnomAD databases (0/281292) and was predicted to be pathogenic by five separate prediction tools (Table 3.8).

Variant (Protein change)	CADD score	SIFT	MutationTaster	PolyPhen2	BLOSUM62	Franklin
c.2555A>T (p.E852V)	23.5	Deleterious (0)	Disease causing	Probably damaging (0.974)	-2	VUS (ACMG; PM2)

Table 3.8: Pathogenic prediction scores for *IFT172* mutation in sample 11935460. The scores for *IFT172* missense variant (NM_015662.3, c.2555A>T, p.E852V) from six different pathogenicity prediction tools are shown. VUS = variant of uncertain significance.

Mutations in the *IFT172* gene are known to cause autosomal recessive RP and Bardet-Biedl syndrome (Bujakowska et al., 2015). Sanger sequencing should be performed to confirm this variant, and DNA samples from further family members should be screened to determine which of the two variants found in this study is the disease-causing variant in this family. However, this was not completed due to lack of time.

3.2.2.7 WES analysis of sample OA1201 identifies a homozygous mutation in *SLC25A46* causing autosomal recessive optic atrophy

WES analysis was performed on the genomic DNA from patient OA1201 with presumed autosomal recessive optic atrophy (as reported by the clinician). After the variants list was filtered as described in Section 3.2.1 and since the inheritance pattern was recessive, only homozygous or biallelic variants were kept. The final list was then compared to the known genes listed in RetNet (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (accessed February 2022). Three homozygous variants were identified (Table 3.9).

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
5	110092451	*SLC25A46*	Homo Missense (27.4)	NM_138773.4: c.670A>G, p.T224A
19	7440667	*ARHGEF18*	Homo Missense (22.3)	NM_001367823.1: c.337C>T, p.R113C
5	89923101	*ADGRV1*	Homo Missense (15.07)	NM_032119.4: c.746G>A, p.R249K

Table 3.9: List of candidate variants in case OA1201 (female) after alignment, variant calling, filtering, and comparison against genes listed on the RetNet website. Chr. = chromosome, Homo = homozygous.

A homozygous variant in *SLC25A46* (Solute Carrier Family 25, Member 46; OMIM 610826), a gene known to be mutated in recessive syndromic optic atrophy (optic atrophy with variable neurological findings including cerebellar ataxia, motor or sensory neuropathy, and pontocerebellar hypoplasia) was identified (Abrams et al., 2015, Janer et al., 2016).

The homozygous missense change (GRCh37, Chr.5: 110092451, NM_138773.4, c.670A>G, p.T224A) is in exon 7 of gene *SLC25A46*. It has been reported in dbSNP database (rs761276607) and found in South Asians at frequency of 0.000014 (3/210630) in gnomAD. It is predicted to be pathogenic according to the CADD score and another three different prediction tools (Table 3.10).

Gene	Variant (Protein change)	CADD score	SIFT	MutationTaster	PolyPhen2	BLOSUM62	Franklin
SLC25A46	**c.670A>G (p.T224A)**	27.4	Deleterious (0)	Disease causing	Possibly damaging (0.519)	0	VUS (ACMG; PP3, PM2)
ARHGEF18	**c.337C>T (p.R113C)**	22.3	Tolerated (0.07)	Polymorphism	Possibly damaging (0.543)	-3	Benign
ADGRV1	**c.746G>A (p.R249K)**	15.07	Tolerated (0.33)	Polymorphism	Benign (0.007)	2	Benign

Table 3.10: Pathogenic prediction scores for mutations in *SLC25A46*, *ARHGF18*, and *ADGRV1* in sample (OA1201). The scores for these mutations from six different pathogenicity prediction tools are shown. VUS = variant of uncertain significance.

The normal amino acid is fully conserved across all accessible species according to HomoloGene (https://www.ncbi.nlm.nih.gov/homologene) (Figure 3.11B) and the variant was confirmed by Sanger sequencing as segregating in the patient's unaffected parents (OA1200 and OA1202) in a manner consistent with recessive inheritance (Figure 3.11A).

Since the other two identified variants in genes *ARHGEF18* and *ADGRV1* were predicted to be benign by at least three pathogenicity prediction tools (Table 3.10), these findings suggest that the variant in *SLC25A46* is likely to be the pathogenic variant in this family.

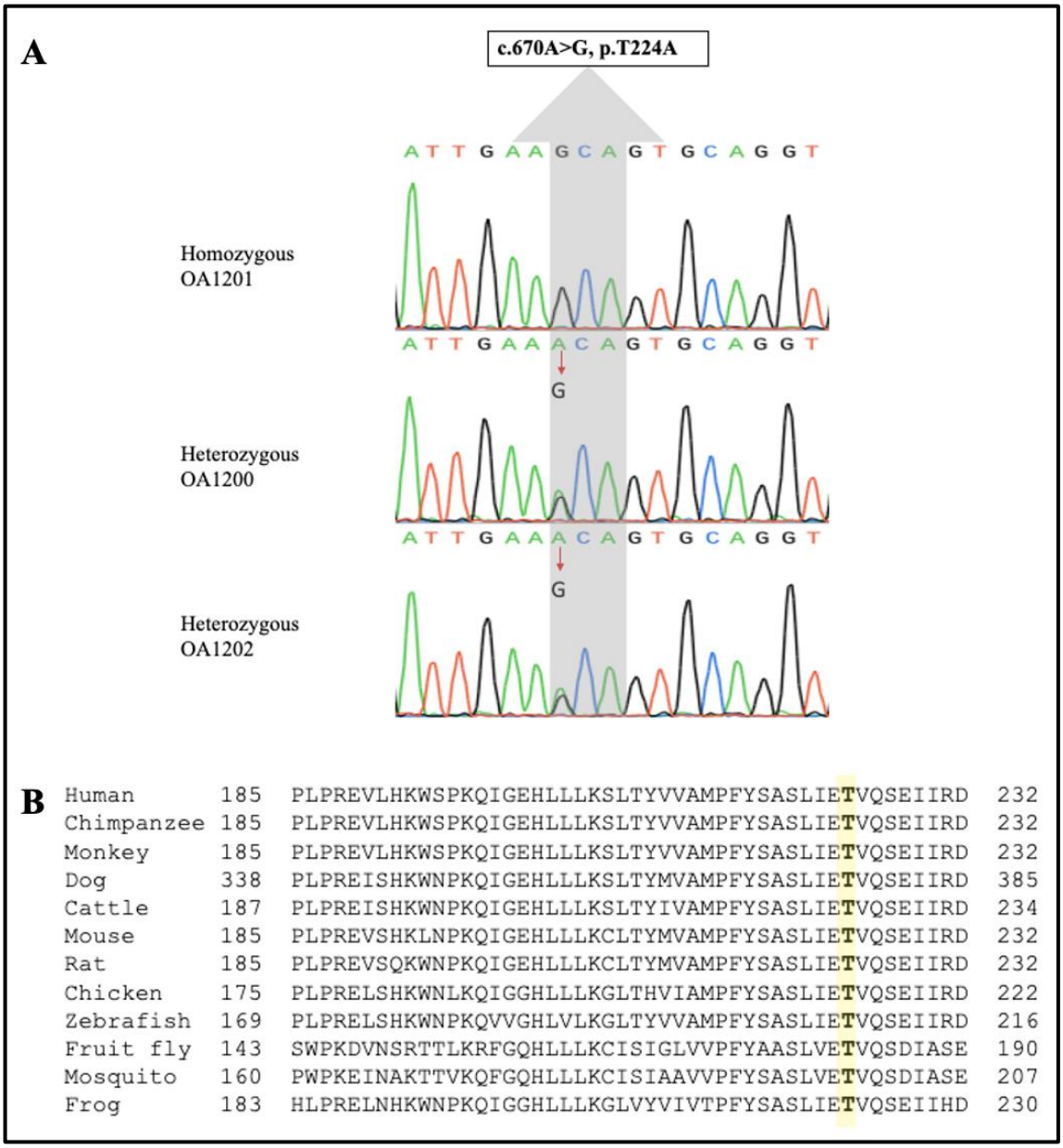


Figure 3.11: Analysis of *SLC25A46* mutation in sample OA1201. A. Sanger sequencing electropherograms for mutant (homozygous) sequence from affected individual OA1201 and heterozygous sequences from two carriers OA1200 and OA1202 for the *SLC25A46* missense variant (NM_138773, c.670A>G, p.T224A). **B.** Normal amino acid sequence alignments of SLC25A46 (HomoloGene, https://www.ncbi.nlm.nih.gov/homologene) showing the evolutionary conservation of threonine residue at position 224.

3.2.2.8 WES analysis of sample 5543 reveals compound heterozygous mutations in *ABCA4* causing CORD

Patient 5543 was diagnosed with late onset simplex CORD. The patient was asymptomatic when the condition was discovered by opticians at age 58 years old during a routine visit. His visual acuity was 6/7.5, with no sign of cystoid macular oedema (CMO) but upon examination there was foveal atrophy.

Both heterozygous and biallelic variants (homozygous and compound heterozygous) were retained for analysis since the patient's family history is unknown. When the resulting list was compared against the known IRD genes in RetNet database (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (accessed February 2022), only three heterozygous variants in two genes were highlighted (Table 3.11).

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
1	94471055	*ABCA4*	Het Missense (34)	NM_000350.3: c.6089G>A, p.R2030Q
1	94517254	*ABCA4*	Het Missense (26.7)	NM_000350.3: c.2588G>C, p.G863A
1	110151344	*GNAT2*	Het Missense (23.4)	NM_032119.4: c.370G>A, p.V124M

Table 3.11: List of candidate variants in case 5543 (male) after alignment, variant calling, filtering, and comparison against genes listed on the RetNet website. Chr. = chromosome, Het = heterozygous.

One of these genes was *ABCA4* (ATP-Binding Cassette, Subfamily A, Member 4; OMIM 601691), which was found to have presumed compound heterozygous variants, one in exon 17 (GRCh37, Chr.1:94517254, NM_000350.3, c.2588G>C, p.G863A) and another in exon 44 (GRCh37, Chr.1:94471055, NM_000350.3, c.6089G>A, p.R2030Q). Six pathogenicity prediction tools were used to assess the likely pathogenicity of these variants and five of them predicted that the missense variants were pathogenic (Table 3.12).

Variant (Protein change)	CADD score	SIFT	MutationTaster	PolyPhen2	BLOSUM62	Franklin
c.6089G>A (p.R2030Q)	34	Deleterious (0)	Disease causing	Probably damaging (1.000)	1	Pathogenic
c.2588G>C (p.G863A)	26.7	Deleterious (0)	Disease causing	Possibly damaging (0.864)	0	Pathogenic

Table 3.12: Pathogenic prediction scores for *ABCA4* mutation in sample (5543). The scores for two *ABCA4* missense variants (NM_000350.3, c.2588G>C, p.G863A and c.6089G>A, p.R2030Q) from six different pathogenicity prediction tools are shown.

The frequencies of both alleles were checked in publicly available databases. The variant c.2588G>C was reported in dbSNP (rs76157638) and found at frequency of 0.5% (601/118484) in ExAC and 0.4% (1214/282648) in gnomAD. Variant c.6089G>A was reported in dbSNP (rs61750641) and found at frequency of 0.03% (47/121404) in ExAC, and 0.03% (100/282784) in gnomAD. Both variants were reported previously (Birtel et al., 2018, Bauwens et al., 2019, Weisschuh et al., 2020b) and both are present in *ABCA4* LOVD database (https://databases.lovd.nl/shared/genes/ABCA4).

Sanger sequencing confirmed the mutations in the sample (Figure 3.12A), and the normal amino acid residues in each case were completely conserved across all available species (Figure 3.12B). However, lack of samples from additional family members meant it was not possible to test segregation, or to establish phase, meaning that these variants could be on the same allele. If further time were available, the haplotype phasing could be established using long read sequencing. Alternatively, it is possible to give a statistical estimation of the haplotype using the PHASE method, which utilises hidden Markov model (HMM) (Stephens et al., 2001).

The heterozygous variant in *GNAT2* was considered unlikely to cause the disease in this case because variants in this gene are associated with autosomal recessive achromatopsia (Kohl et al., 2002) and recessive CORD (Michaelides et al., 2003a), meaning a second variant is needed for it to be the disease-causing gene. However, the possibility of a second intronic or promotor variant cannot be excluded as this cannot be detected using WES.

Nevertheless, given the phenotype observed, the lack of any other obvious variants to account for the disease and the likely pathogenicity of these variants, these biallelic *ABCA4* mutations are considered the most likely cause of the CORD phenotype in this patient.

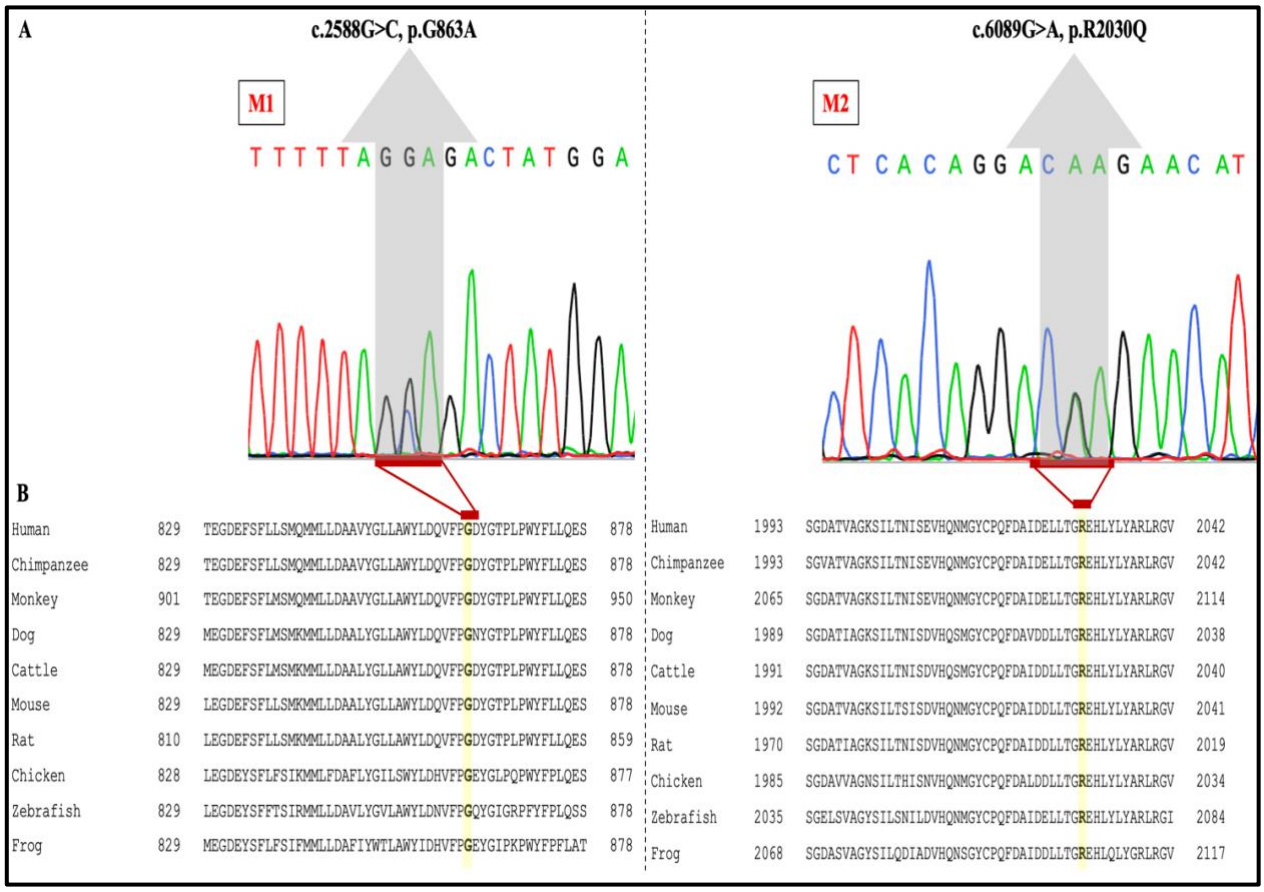


Figure 3.12: *ABCA4* mutations in sample 5543. A. *ABCA4* (NM 000350.3) Sanger sequencing electropherograms around the compound heterozygous variants (M1: c.2588G>C, p.G863A and M2: c.6089G>A, p.R2030Q) in sample 5543. **B.** Amino acid sequence alignments of ABCA4 (HomoloGene, https://www.ncbi.nlm.nih.gov/homologene) showing the evolutionary conservation of glycine at position 863 and arginine at position 2030.

3.2.3 Unsolved cases- Candidate variants

3.2.3.1 WES analysis of sample 5544 reveals a heterozygous mutation in *OPA1* as a possible cause of CORD

Patient 5544 was diagnosed with cone rod dystrophy (CORD) and has a sister and brother in Nigeria who may also be affected. However, the lack of further information means that inheritance pattern is not clear. The patient had a visual acuity (VA) of 6/60 in one eye and can recognise only hand movements (HM) in the other eye, central extensive macular pigmentation with peripheral bone spicules and peripheral avascularity. WES was performed on the proband's genomic DNA, and heterozygous and biallelic variants (homozygous and compound heterozygous) were retained. After filtering against the known IRD genes in the RetNet database (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (accessed February 2022), there were only seven heterozygous variants in total (Table 3.13) with a variant in *OPA1* being one of them (encoding the OPA1 mitochondrial dynamin like GTPase protein; OMIM 605290). The other six variants were in genes that are known to cause recessive diseases (*SAMD11*, *NBAS*, *MTTP*, *LCA5*, *CNGB3*, and *RBP3*), and since no second variants were found in these genes, they were considered unlikely to be involved in causing the IRD in this patient. However, second variants in these genes could occur in non-coding regions that WES, the method used in this study, is unable to identify.

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
4	100543853	*MTTP*	Het Missense (34)	NM_001386140.1: c.2533G>A, p.E845K
1	865625	*SAMD11*	Het Missense (26.6)	NM_001385641.1: c.700G>A, p.D234N
3	193355809	*OPA1*	Het Missense (25.1)	NM_130836: c.1050A>G, p.I350M
8	87641188	*CNGB3*	Het Missense (24.1)	NM_019098.5: c.1439G>A, p.R480Q
10	48390460	*RBP3*	Het Missense (23.1)	NM_002900.3: c.418G>A, p.G140S
6	80196789	*LCA5*	Het Missense (22.8)	NM_001122769.3: c.2026G>T, p.D676Y
2	15427245	*NBAS*	Het Missense (20.9)	NM_015909.4: c.5090G>T, p.R1697L

Table 3.13: List of candidate variants in case 5544 (female) after alignment, variant calling, filtering, and comparison against genes listed on the RetNet website. Chr. = chromosome, Het = heterozygous.

Heterozygous mutations in *OPA1* can cause autosomal dominant optic atrophy (Delettre et al., 2001, Toomes et al., 2001b). The heterozygous missense mutation c.1050A>G, p.I350M (GRCh37, Chr.3:193355809, NM_130836) is in exon 10 of the *OPA1* gene. It has been detected in heterozygous state in four individuals in gnomAD (African/African American) at a frequency of 0.000014 (4/282832). It is also found in dbSNP (rs771817617) at frequency of 0.000008 (1/120892) in ExAC. It has a CADD score of 25.1, the normal amino acid is completely conserved across all species (Figure 3.13B) and the substitution is predicted to cause disease by four out of six separate prediction tools (Table 3.14).

Gene	Variant (Protein change)	CADD score	SIFT	MutationTaster	PolyPhen2	BLOSUM62	Franklin
MTTP	**c.2533G>A (p.E845K)**	34	Deleterious (0.04)	Disease causing	Probably damaging (0.994)	1	VUS (ACMG; PM2)
SAMD11	**c.700G>A (p.D234N)**	26.6	Deleterious (0)	Disease causing	Possibly damaging (0.468)	1	Likely benign
OPA1	**c.1050A>G (p.I350M)**	25.1	Deleterious (0)	Disease causing	Possibly damaging (0.519)	1	VUS (ACMG; PM2, PM1, PP3)
CNGB3	**c.1439G>A (p.R480Q)**	24.1	Deleterious (0.02)	Disease causing	Possibly damaging (0.49)	1	Benign
RBP3	**c.418G>A (p.G140S)**	23.1	Tolerated (0.09)	Polymorphism	Probably damaging (0.998)	0	Benign
LCA5	**c.2026G>T (p.D676Y)**	22.8	Deleterious (0.01)	Polymorphism	Possibly damaging (0.563)	-3	Likely benign
NBAS	**c.5090G>T (p.R1697L)**	20.9	Tolerated (0.88)	Polymorphism	Benign (0)	-2	Likely benign

Table 3.14: Pathogenic prediction scores for the mutations identified in sample (5544). The scores from six different pathogenicity prediction tools are shown. VUS = variant of uncertain significance.

Due to lack of DNA samples from other family members, segregation studies could not be carried out. Sanger sequencing was performed on the sample, and the mutation was confirmed (Figure 3.13A).

The remaining ~ 400 exonic/splice variants listed with CADD scores of 15 and above were also investigated to look for a potential pathogenic variant in other genes. Bam files were further analysed on IGV to look for a second heterozygous variant in the above-mentioned recessive genes. Moreover, CNVs detection using ExomeDepth analysis as described in Section 2.11, did not reveal any significance CNVs. These results led to the conclusion that this patient's condition may have been caused by the *OPA1* mutation, though it is also important to consider variants in non-coding regions in the other genes with potentially pathogenic heterozygous variants, as mentioned above. However, the

fact that the patient was diagnosed with CORD rather than optic atrophy, together with the lack of segregation, leaves some doubt in this case, meaning that it is considered unsolved at this stage. The list of top five candidate variants based on CADD score can be found in Appendix B.4.11**.**

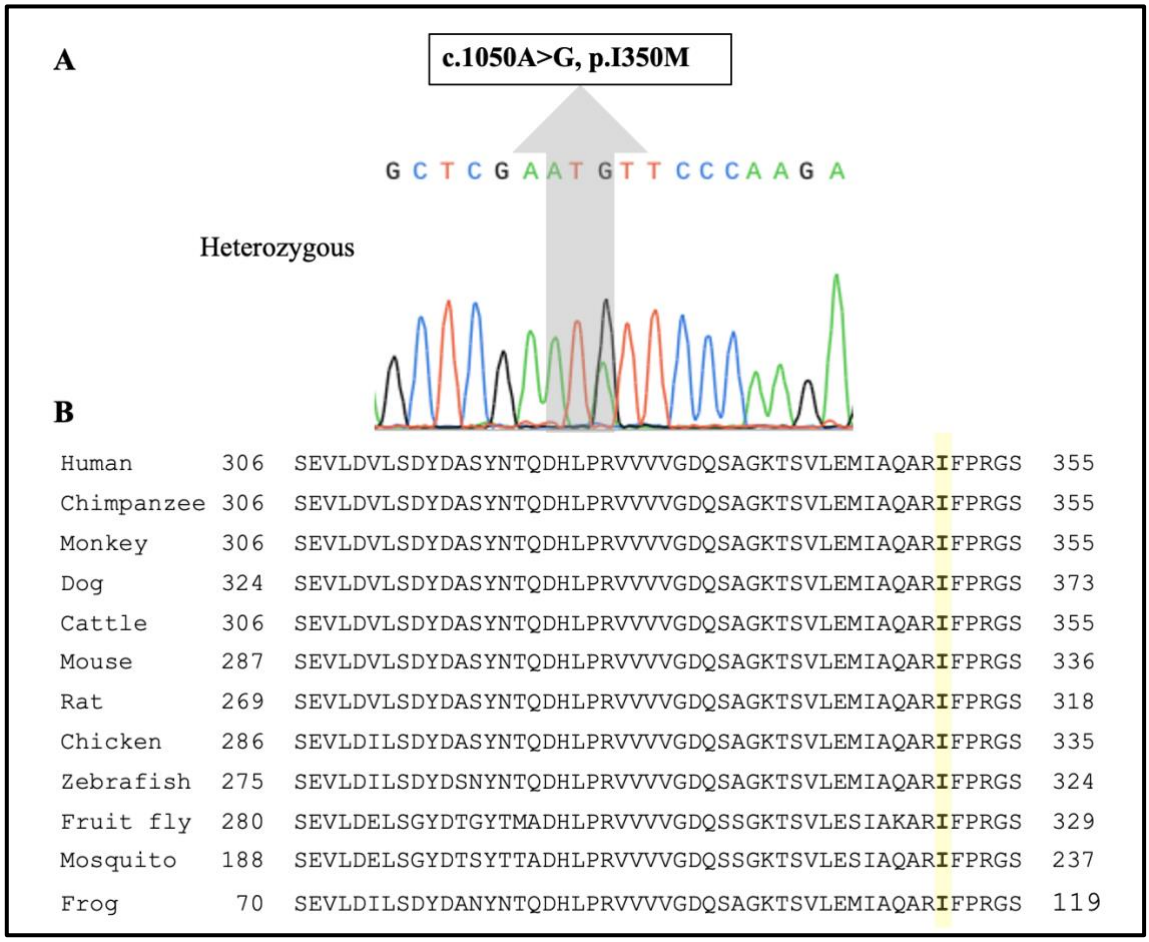


Figure 3.13: Analysis of *OPA1* mutation in sample 5544. **A.** Sanger sequencing electropherogram for mutant (heterozygous) sequence from affected individual 5544 for the *OPA1* missense variant c.1050A>G, p.I350M (NM_130837). **B.** Protein sequence alignments of OPA1 in various species (HomoloGene, https://www.ncbi.nlm.nih.gov/homologene) showing the evolutionary conservation of isoleucine residue at position 350.

3.2.3.2 WES analysis of sample 5546 reveals a heterozygous mutation in *C3* as a possible cause of AVMD

Patient (5546) was diagnosed with adult vitelliform macular dystrophy (AVMD). He presented at age 55 with poor near sight vision. On examination, he had a visual acuity of 6/18 in both eyes, a vitelliform large solid lesion in the right eye, subretinal fibrosis (SRF) and vitelliform pattern dystrophy in the left eye, with no basal laminar drusen or vitreomacular traction (VMT) and no central serous chorioretinopathy (CSC) on OCT scan. Other family members' clinical information and DNA samples were unavailable.

The proband's genomic DNA was subjected to WES. Due to the lack of a family history, heterozygous and biallelic variants (homozygous and compound heterozygous) with a CADD score of 15 or above were retained and compared to the known IRD genes in the RetNet database (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (accessed February 2022). Four variants remained after this analysis (Table 3.15), of which three did not match the reported phenotype.

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
10	102782113	*PDZD7*	Het Missense (24.6)	NM_001195263.2: c.572T>A, p.V191E
2	234959642	*SPP2*	Het Missense (23.9)	NM_006944.3: c.113C>T, p.S38F
19	6711203	*C3*	Het Missense (23.7)	NM_000064.4: c.1274G>T, p.R425L
X	49081230	*CACNA1F*	Hemi Missense (21.6)	NM_001256789.3: c.1870G>A, p.V624I

Table 3.15: List of candidate variants in case 5546 (male) after alignment, variant calling, filtering, and comparison against genes on the RetNet website. Chr. = chromosome, Het = heterozygous, and Hemi = hemizygous.

A heterozygous missense variant (GRCh37, Chr.19:6711203, NM_000064.4, c.1274G>T, p.R425L) was identified in exon 12 of the *C3* gene (Complement component 3; OMIM 120700). Variants in the *C3* gene were found to be associated with AMD (Maller et al., 2007, Yates et al., 2007). It was present in gnomAD at frequency of 0.00004 (11/278788) and dbSNP (rs201714568) at frequency of 0.00006 (2/32402). It has a CADD score of 23.7 and predicted to be pathogenic by four out of six pathogenicity prediction tools (Table 3.16).

Variant (Protein change)	CADD score	SIFT	MutationTaster	PolyPhen2	BLOSUM62	Franklin
c.1274G>T (p.R425L)	23.7	Deleterious (0)	Polymorphism	Probably damaging (0.987)	-2	VUS (ACMG; PM2, BP4)

Table 3.16: Pathogenic prediction scores for *C3* mutation in sample (5546). The scores for *C3* missense variant (NM_130837, c.1274G>T, p.R425L) from six different pathogenicity prediction tools are shown. VUS = variant of uncertain significance.

Sanger confirmation was not performed due to lack of time. Given that mutations in the same IRD gene can cause different IRD phenotypes, variants in the genes *PDZD7*, *SPP2*, and *CACNA1F* should be considered as potential disease-causing variants in this case. Because the majority of AVMD cases are not solved by variants in the known genes only, further investigation was carried out by looking at the remaining ~400 variants to look for potential new candidate genes. However, based on the phenotype, no significant variants were highlighted as the possible cause in this case. CNVs analysis was also performed by ExomeDepth (Section 2.11) and did not reveal any significant CNVs. Therefore, this case was considered unsolved. The list of top five candidate variants based on CADD score can be found in Appendix B.4.12.

3.2.3.3 WES analysis of sample 12565367 reveals a homozygous mutation in *MFSD9* as a possible cause of visual impairment and developmental delay

Sample 12565367 is an Omani female who was initially referred to the genetic clinic in Oman due to dysmorphic features. She was then seen at a local eye clinic and diagnosed with a minor squint and visual impairment at the age of 4 months. She started to show more distinct facial features (frontal bossing, hypertelorism, low set ears, blue sclera, down slanting eyes and micrognathia) around the age of 20 months, and her mother noticed that she was developmentally delayed compared to her siblings. Her parents are first cousins, and she has five healthy siblings with no family history of similar symptoms. Her genomic DNA was investigated by WES, and since the parents are first cousins, only homozygous variants were retained. The 26 homozygous variants identified were tested against the known IRD genes and only one potential gene, *MTTP*, was found (Table 3.17). *MTTP* (Microsomal Triglyceride Transfer Protein; OMIM 157147) can cause abetalipoproteinemia and retinal pigmentary degeneration (Narcisi et al., 1995), which does not fit the described phenotype perfectly. The variant was predicted to be tolerated on MutationTaster, SIFT, and was classed as a variant of unknown significance (VUS) in Franklin (Section 2.13.4.6).

To prioritise the remaining 25 homozygous variants, homozygous regions were identified by AutoMap (https://automap.iob.ch) (Quinodoz et al., 2021) (Figure 3.14) using the VCF file and only 8 variants were located in the identified regions of homozygosity (Table 3.17).

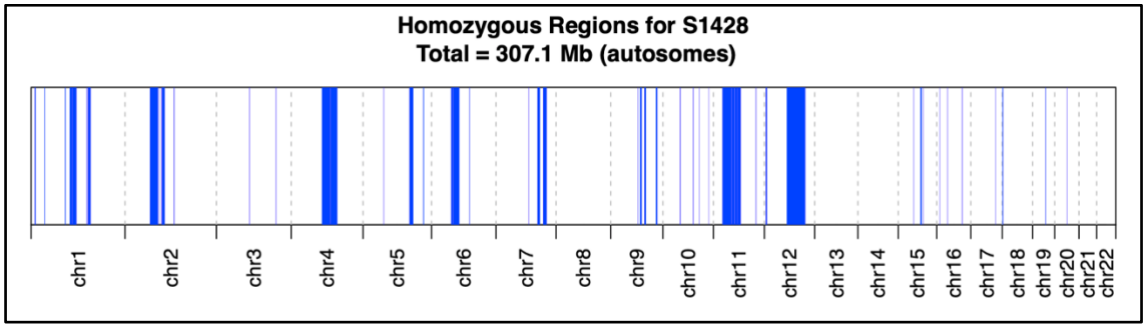


Figure 3.14: Homozygous regions for WES data from sample 12565367. Identified homozygous regions with a size of 307.1 Mb are shown in blue.

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation	Frequency (gnomAD)	Associated phenotypes/Gene function
2	103343425	*MFSD9*	Stop gained (39)	NM_032718.5:c.306G>A, p.W102*	352/250734 (0.14%)	Potential cause for developmental delay (Bartnik et al., 2014)
4	96762427	*PDHA2*	Missense (28.7)	NM_005390.5:c.1126C>A, p.R376S	0/251466	Spermatogenic failure (Yıldırım et al., 2018) /Pyruvate metabolism and estrogen receptor (ESR) mediated signalling
11	61908496	*INCENP*	Missense (24.9)	NM_001040694.2:c.1573C>T, p.P525S	4/249820 (0.0016%)	Nephronophthisis (Braun et al., 2016)/ Inner centromere protein
7	127222316	*GCC1*	Missense (23.4)	NM_024523.6:c.2080C>T, p.R694W	2/282762 (0.0007%)	Probably involved in maintaining Golgi structure
1	156526327	*IQGAP3*	Missense (23)	NM_178229.5:c.1288G>C, p.G430R	153/281736 (0.05%)	Predicted to be involved in actin regulation
4	84227441	*HPSE*	Missense (21.5)	NM_001098540.3:c.1121G>A, p.R374Q	39/282742 (0.01%)	Heparinase activity/ elevated in brain metastatic breast cancer (Zhang et al., 2011)
4	110222980	*COL25A1*	Missense (21.4)	NM_198721.4:c.196C>G, p.L66V	21/280942 (0.007%)	Encode a brain specific collagen/ congenital fibrosis of extraocular muscle (Shinwari et al., 2015)
4	100496109	*MTTP*	Missense (18.24)	NM_001386140.1:c.43T>C, p.Y15H	-/31398	Recessive abetalipoproteinemia (Narcisi et al., 1995)

Table 3.17: List of candidate homozygous variants in case 12565367 after alignment, variant calling and filtering against the homozygous regions identified by AutoMap using the VCF file of sample 12565367. Chr. = chromosome.

Based on the syndromic phenotype, which included developmental delay, a stop gained mutation in *MFSD9* (encoding the Major Facilitator Superfamily Domain containing protein 9) was identified as a variant of interest. Copy number variants in this gene have previously been reported as a potential cause for developmental delay and dysmorphic features (Bartnik et al., 2014). The homozygous stop gained mutation (GRCh37, Chr.2:103343425, NM_032718.5, c.306G>A, p.W102*) has a CADD score of 39 and was predicted to be pathogenic in MutationTaster and SIFT. It is reported in dbSNP (rs34960597) with allele frequency of 0.07% (201/264690) in TOPMED and 0.14% (352/250734) in gnomAD. The variant is located in a homozygous region spanning 7.26 Mb in chromosome 2. This mutation was confirmed on Sanger sequencing (Figure 3.15), but no segregation studies were done due to lack of additional family members' DNA samples. This result has been added to GeneMatcher on 21 September 2021 (https://genematcher.org) to see if other researchers have identified variants in this gene, but as yet none have been reported.

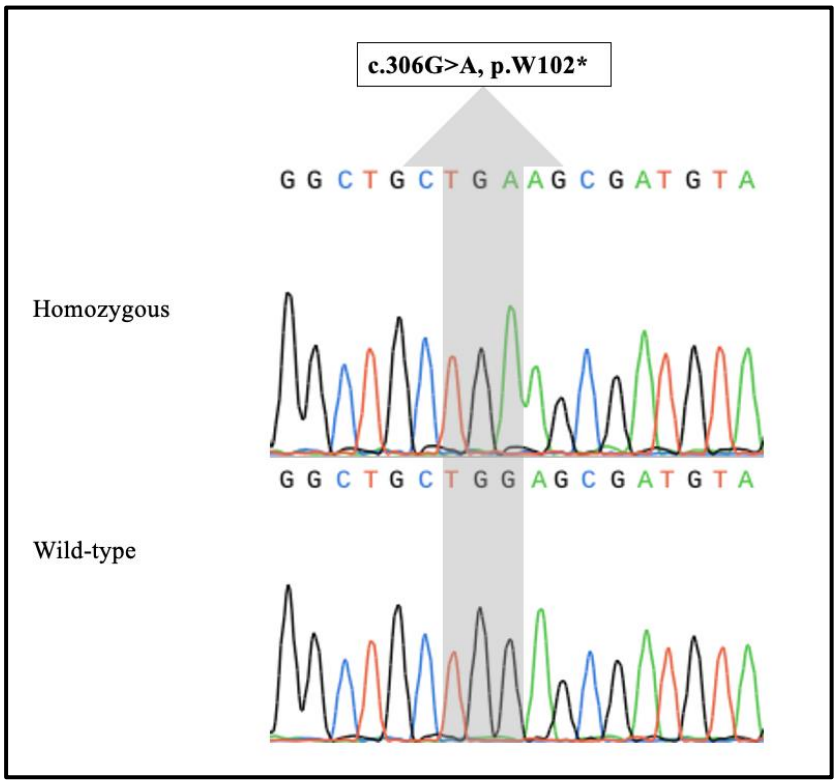


Figure 3.15: *MFSD9* mutation in sample 12565367. Sanger sequencing electropherograms for mutant (homozygous) and wild-type (healthy) sequences for the *MFSD9* stop gained variant (NM_032718.5, c.306G>A, p.W102*).

Variants in other candidates were further investigated by looking at their frequency in public databases, associated phenotypes, and gene function in the literature, but nothing significant was identified. CNV analysis was also performed for this sample using ExomeDepth (Section 2.11) revealing no significant homozygous CNVs over the homozygous regions identified by AutoMap. This sample was considered unsolved due to the ambiguous relevance of the variants identified and the lack of data demonstrating its relationship to the described phenotype. More DNA samples from other family members are necessary to confirm and segregate the *MFSD9* and other variants utilising Sanger sequencing.

3.2.3.4 WES analysis of sample 11795135 diagnosed with syndromic cardiac and vision problems

Genomic DNA from an Omani female (11795135) with a cardiac disorder, dysmorphic features, poorly defined vision problems and developmental delay was analysed by WES. According to the medical report, three of her relatives have congenital visual impairments, and her parents are unrelated. Both heterozygous and biallelic variants (homozygous and compound heterozygous) were examined in light of this unclear family history and diagnosis. Filtering against known IRD genes in the RetNet database (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (accessed February 2022) identified nine potential variants, and a heterozygous missense variant (GRCh37, Chr.16:16286757, NM_001171.6, c.1361C>T, p.T454I) was highlighted in the *ABCC6* gene. *ABCC6* variants have previously been implicated in an overlapping albeit not exactly matching phenotype. However, the variant was also classified as a VUS in the Franklin database, and no further samples from family members were available for segregation analysis. The other eight variants were found in genes that did not match the described phenotype. Since this is not a clear IRD, looking at genes that are only implicated in IRDs is not valid. The variant list was long (over 500 variants), even after filtration. These variants were investigated further by looking at the variant frequency in public database, associated phenotype, and gene function in the literature. The top five candidates based on CADD score are listed in Table 3.18. Moreover, CNVs analysis using ExomeDepth was also performed for this sample as described in Section 2.11 and did not reveal any significant CNVs. This case was therefore considered as unsolved.

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation	Frequency (gnomAD)	Associated phenotypes/Gene function
17	26851779	*FOXN1*	Het Missense (28.5)	NM_001369369.1:c.382C>T, p.R128W	179/279690 (0.06%)	T-cell immunodeficiency (Bosticardo et al., 2019)
9	139908461	*ABCA2*	Het Missense (27.8)	NM_001606.5: c.4270G>A, p.G1424S	10/275590 (0.004%)	Highly expressed in brain tissue, play a role in neural development
6	10529796	*GCNT2*	Het Missense (27.7)	NM_145649.5:c.652G>A, p.G218R	3/250520 (0.0011%)	Formation of blood group I antigen
18	44104721	*LOXHD1*	Het Missense (27.6)	NM_001384474.1: c.1357C>T, p.L1564F	5/157924 (0.003%)	Non-syndromic hearing loss / essential for normal hair cell function (Grillet et al., 2009)
16	16286757	*ABCC6*	Het Missense (22.3)	NM_001171.6: c.1361C>T, p.T454I	99/282628 (0.03%)	Pseudoxanthoma elasticum (Bergen et al., 2000)

Table 3.18: List of top five candidate variants via CADD score in case 11795135 after alignment, variant calling, and filtering. Chr. = chromosome, Het = heterozygous.

3.2.3.5 WES analysis of sample F1427 reveals a heterozygous deletion in *VCAN* as a possible cause of autosomal dominant FEVR

The genomic DNA from a dominant FEVR case (F1427) was analysed by WES screening. After filtering, only heterozygous variants were retained given the pattern of inheritance. Comparison with RetNet (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (accessed February 2022) revealed three heterozygous variants in known IRD genes (*IFT27*, *MERTK*, and *NBAS*), none of which appeared to exactly fit the given phenotype and the inheritance pattern.

However, CNVs analysis using ExomeDepth (Section 2.11) revealed a heterozygous deletion (GRCh37, chr5:82832827-82838087) of exon 8 in the *VCAN* gene (encoding the protein Versican, also called Chondroitin Sulfate Proteoglycan 2 or CSPG2; OMIM 118661). ExomeDepth estimated that 1241 reads were expected covering exon 8 based on the control samples, but only 823 reads were observed in sample F1427. The reads ratio value (RR) was 0.64, suggesting a heterozygous deletion. Heterozygous mutations in *VCAN* have been reported to cause autosomal dominant Wagner syndrome (OMIM 143200) (Miyamoto et al., 2005, Rothschild et al., 2013, Burin-des-Roziers et al., 2017). This syndrome is mainly characterised by an empty vitreous cavity with avascular strands or veils, chorioretinal atrophy, cataract by the age of 45, and retinal detachment (Graemiger et al., 1995). In one family reported in the literature, a *VCAN* variant was found to segregate with exudative retinal detachment in the context of a severe uncharacterized familial vitreoretinopathy (Brézin et al., 2011). A heterozygous splice-site mutation (c.9265+2T>A) in the *VCAN* gene was also identified in a family with FEVR (Panagiotou, 2018). Moreover, deletions overlapping exon 8 were reported as pathogenic (Burin-des-Roziers et al., 2017). This heterozygous deletion was therefore considered a possible pathogenic cause in this case and attempts were made to amplify the breakpoints with three primer sets (Appendix B.3), using long range PCR. Several technical variations were tried to optimise the PCR conditions, but no PCR product was observed. The precise breakpoints of this deletion need to be confirmed either by designing another set of primers or using WGS. Additional DNA samples from other family members should also be sequenced to determine whether this variant segregates with the phenotype. At this stage, the case was considered unsolved, but with a potentially significant variant requiring further investigation.

3.2.3.6 WES analysis of sample 3727 diagnosed with macular degeneration

Affected individual 3727 was diagnosed with macular degeneration. After WES analysis of genomic DNA, due to a lack of information about the patient's family history, both heterozygous and biallelic variants (homozygous and compound heterozygous) were evaluated. As putative candidate causal mutation of the disease, a heterozygous missense variant in *TLR3* was identified. The heterozygous mutation in gene *TLR3* (Toll-Like Receptor 3; OMIM 603029) (GRCh37, Chr.4:187003691, NM_003265.3, c.851A>T, p.N284I) was detected as a possible causative variant in this case based on RetNet database (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes). It has a CADD score of 23.4 and predicted to be pathogenic in four additional pathogenicity prediction tools (Table 3.19). It is present in gnomAD at frequency of 0.00004 (11/282848) and dbSNP (rs5743316) at frequency of 0.00006 (17/264690) in TOPMED. Verification by Sanger sequencing and segregation analysis was not performed due to lack of time for this variant to be checked.

The remaining ~ 100 exonic/splice variants list with CADD score of 15 and above were also checked to look for a potential pathogenic variant in other genes and the top five candidate variants based on CADD score are listed in Appendix B.4.13. Moreover, CNV analysis using ExomeDepth as described in Section 2.11, did not reveal any significant CNVs. This case was therefore considered unsolved.

Variant (Protein change)	CADD score	SIFT	MutationTaster	PolyPhen2	BLOSUM62	Franklin
c.851A>T (p.N284I)	23.4	Deleterious (0)	Disease causing	Probably damaging (1)	-3	VUS (ACMG; PM2)

Table 3.19: Pathogenic prediction scores for *TLR3* mutation in sample 3727. The scores for *TLR3* missense variant (NM_003265.3, c.851A>T, p.N284I) from six different pathogenicity prediction tools are shown.

3.2.3.7 Other unsolved cases

None of the remaining cases (3558, 3800, 4212, 5545, 5582, 5583, OA1203, 12194201, 12642093, and 8956112**)** were solved in this study. Variant lists were prepared for each sample as described in Figure 3.2, then for autosomal dominant cases, heterozygous variants were kept. For recessive cases only biallelic (homozygous or compound heterozygous) variants were retained and in cases with one heterozygous variant, eyeballing on IGV was done to look for a second heterozygous variant. The possibility of second variants in non-coding regions, which are difficult to identify by WES, should be acknowledged.

For cases with unknown family histories, both heterozygous and biallelic variants (homozygous or compound heterozygous) were checked. These lists were then manually compared against the RetNet database (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes), and in cases where the phenotype was poorly defined, genes containing candidate variants were also scanned against the published literature to look for their protein function and associated phenotypes. In addition, ExomeDepth was used, as described in Section 2.11, to identify rare CNVs as further candidate pathogenic variants.

However, in the cases detailed above, no point mutations or CNVs were detected in the coding regions of the known IRD genes. The most significant variants, namely rare variants with CADD scores of 15 and above which were found in genes not in RetNet, were investigated further using bioinformatics (e.g., other pathogenicity prediction tools). OMIM and PubMed were used to search for potential candidate variants in novel genes that were informed by the described phenotype and the inheritance pattern. In syndromic retinal disease cases such as cases 12194201 (a male patient diagnosed with microcephaly, vision problem, and developmental delay) and 8956112 (a male patient with a possible diagnosis of microdeletion syndrome), every gene in that list was examined. Samples 3558 and 5583 (Section 5.2.1.3) were also included in the PacBio long read sequencing analysis in order to search for a potential pathogenic variant in the challenging ORF15 region of the *RPGR* gene because 3558 was a male with recessive RP and 5583 was a female with X-linked RP, as reported by the clinician. The list of top five candidate variants for the remaining unsolved cases can be found in Appendix B.4.

3.3 Discussion

This chapter describes the use of WES in a cohort of 24 previously unscreened cases from patients with IRD/low vision. 17 of the individuals investigated had a variety of inherited retinal disorders (IRDs), two had recessive optic atrophy, one sample was from a patient with syndromic IRD, and four were from patients with poorly defined syndromes involving visual impairment. Each sample underwent WES analysis and variant filtering to identify the pathogenic variants in known IRD genes or in potential novel genes related to the reported phenotype. The R package, ExomeDepth, was used to detect CNVs in order to look for any exonic deletions/duplications that WES might have missed. Additionally, in some cases with recessive inheritance pattern, WES was combined with homozygosity mapping.

3.3.1 Mutations in genes associated with IRDs

Applying this approach has revealed nine likely pathogenic variants in genes known to be involved in IRDs (*CDHR1*, *RHO*, *PRPF31*, *CNGA3*, *BBS10*, *USH2A*, *SLC25A46*, and *ABCA4*) in 8 of the 24 cases (33.3%) (Table 3.20). These variants were confirmed by Sanger sequencing and were predicted to be pathogenic by several pathogenic prediction tools. The *SLC25A46* gene will be further discussed because case OA1201 had non-syndromic OA and the gene is known to be associated with syndromic OA.

3.3.1.1 *SLC25A46*

The homozygous missense change found in patient OA1201 (GRCh37, Chr.5: 110092451, NM_138773, c.670A>G, p.T224A) is in the *SLC25A46* gene. Homozygous mutations in *SLC25A46* gene have been linked to various neurological diseases with recessive inheritance, including Leigh syndrome (Janer et al., 2016, Li et al., 2021), syndromic optic atrophy with axonal Charcot-Marie-Tooth disease (CMT) and cerebellar atrophy (Abrams et al., 2015) and optic atrophy with progressive myoclonic ataxia (Charlesworth et al., 2016). It is well documented that mutations in *SLC25A46* are associated with a wide range of clinical features, with optic atrophy and axonal neuropathy being the most prevalent among all patients thus far. The severity of the disease, nevertheless, varies and can occasionally be fatal (Abrams et al., 2015, Janer et al., 2016). Patients with destabilizing mutations, or complete loss of function mutations, are at the severe end of the spectrum. Conversely, patients who have mutations that produce intermediate levels of the protein in the body may show mild disease symptoms.

The mildest form of the disease, with optic atrophy in the first decade of life, was observed in a patient with the p.G249D variant and a second mutation in *trans*. This patient developed spasticity and axonal neuropathy only in the late 40s (Abrams et al., 2015). Furthermore, the identification of the stable homozygous p.R257Q mutation has shed light on the clinical diversity of this disease and shown a direct proportion between the severity and protein levels, with higher levels being associated with a milder phenotype (Abrams et al., 2018). The age at sampling of the patient reported here was 9 years, therefore the mutation identified in this study may have resulted in intermediate levels of the protein and the neurological symptoms may appear later. Moreover, no neurological symptoms had been reported for this patient, so the homozygous *SLC25A46* change may cause a non-syndromic optic atrophy as it is not uncommon for a syndromic disease gene to cause a non-syndromic retinal disease (Estrada-Cuzcano et al., 2012). Functional analysis of the mutant protein and a thorough study of the phenotype in this patient is needed to clarify the contribution of this mutation to the disease. For instance, such studies might involve by introducing the mutation by gene editing into an induced pluripotent cell line, differentiating this into a neuronal cell type then examining mitochondrial morphology (Schuettpelz et al., 2023). Follow up in the clinic is also recommended to predict the prognosis of the disease in this patient.

Sample ID	Diagnosis	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
F1415	Recessive IRD + Cataract	Chr.10:85971445	*CDHR1*	Homo Stop gained (35)	NM_033100: c.1527T>G, p.Y509*
3869	Dominant RP	Chr.3:129247860	*RHO*	Het Missense (25.7)	NM_000539: c.284T>C, p.L95P
4211	Dominant RP	Chr.19:54627976	*PRPF31*	Het Frameshift (32)	NM_015629: c.797delC, p.S266*
11335421	Recessive RP	Chr.2:99012721	*CNGA3*	Homo Missense (25.1)	NM_001298: c.1088T>C, p.L363P
11867145	Recessive BBS	Chr.12:76741033	*BBS10*	Homo Frameshift (34)	NM_024685: c.728-731delAAGA, p. K243Ifs*15
11935460	Recessive RP	Chr.1:216595427	*USH2A*	Homo Missense (23.5)	NM_206933: c.252T>G, p.C84W
5543	CORD	Chr.1:94517254	*ABCA4*	C- het Missense (26.7 and 34)	NM_000350.3: c.2588G>C, p.G863A
		Chr.1:94471055			NM_000350.3: c.6089G>A, p.R2030Q
OA1201	Recessive OA	Chr.5:110092451	*SLC25A46*	Homo Missense (27.4)	NM_138773: c.670A>G, p.T224A

Table 3.20: A list of the likely pathogenic variants identified in this study. IRD = inherited retinal dystrophy, RP = retinitis pigmentosa, BBS = Bardet-Biedl syndrome, CORD = cone rod dystrophy, OA = optic atrophy, Chr = chromosome, Het = heterozygous, Homo = homozygous, C- het = compound heterozygous.

In order to make these findings available to the IRD community they will be made publicly available via the ClinVar database. In addition, these findings either have been or will be passed back to the relevant clinician for informed discussion with the families involved. However, given the tentative nature of these results, and as they have not yet been verified in an accredited diagnostic setting, they will be given as unconfirmed provisional findings. Where possible, further work will be undertaken to test additional family members, again ideally by an accredited NHS genetics laboratory.

3.3.2 Unsolved cases- Candidate variants

Six cases (25%) had candidate variants in genes linked to disorders other than those initially described (Table 3.21). This means that the initial diagnosis may change after molecular testing. In retinal diseases, it has been observed that some patients may carry pathogenic mutations in genes not particularly associated with the initial described phenotype (Liu et al., 2015, Wang et al., 2016b). Sanger sequencing confirmed the identified variants in only two cases, due to lack of time. However, further analysis is needed for all of them to confirm the genotype-phenotype correlations.

Sample ID	Diagnosis	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation	Associated phenotypes
5544	CORD	Chr.3:193355809	*OPA1*	Het Missense (25.1)	NM_130836: c.1050A>G, p.I350M	ADOA, ADOA+, Behr syndrome, PHARC, SCA5, SPG18, OA1
5546	AVMD	Chr.19:6711203	*C3*	Het Missense (23.7)	NM_000064.4: c.1274G>T, p.R425L	AMD
12565367	Recessive Syndromic minor squint, visual impairment, and developmental delay	Chr.2:103343425	*MFSD9*	Homo Stop gained (39)	NM_032718.5: c.306G>A, p.W102*	Developmental delay/intellectual disability
11795135	Syndromic cardiac and vision problem	Chr.16:16286757	*ABCC6*	Het Missense (22.3)	NM_001171.6: c.1361C>T, p.T454I	Pseudoxanthoma elasticum
F1427	FEVR	Chr.5:82832827-82838087	*VCAN*	Het exon 8 deletion		Wagner's disease
3727	MD	Chr.4:187003691	*TLR3*	Het Missense (23.4)	NM_003265.3: c.851A>T, p.N284I	AMD

Table 3.21: A list of candidate variants identified in genes inconsistent with the described phenotype. CORD = cone rod dystrophy, AVMD = adult vitelliform macular dystrophy, FEVR = familial exudative vitreoretinopathy, MD = macular dystrophy, AMD = age related macular dystrophy, ADOA = autosomal dominant optic atrophy, PHARC = polyneuropathy, hearing loss, retinitis pigmentosa, and cataract, SCA5 = spinocerebellar ataxia 5, SPG18 = spastic paraplegia 18, OA1 = ocular albinism type 1, Chr = chromosome, Het = heterozygous, Homo = homozygous.

3.3.2.1 *OPA1*

A heterozygous missense mutation (GRCh37, Chr.3:193355809, NM_130836, c.1050A>G, p.I350M) in the *OPA1* gene (Section 1.6.5) was identified as a potential cause of the CORD phenotype in patient 5544. It is widely expressed, with varied levels of expression in various tissues and encodes a protein that regulates mitochondrial stability and energy output (Yu-Wai-Man et al., 2010). Recessive mutations in *OPA1* result in Behr syndrome (BEHRS) (Schaaf et al., 2011, Bonneau et al., 2014) and mitochondrial DNA depletion syndrome type 14 (MTDPS-14) (Spiegel et al., 2016), while heterozygous mutations are linked to ADOA (57–89%) and ADOA+ (Delettre et al., 2000, Toomes et al., 2001b, Shimizu et al., 2003).

In the work described here, the heterozygous missense mutation identified in this patient was in exon 10 of the *OPA1* gene, which means it is located in the GTPase domain (exons 8-15). Isoleucine at this position is highly conserved across all species. According to the *OPA1* LOVD database, the majority of variants in exon 10 of *OPA1* are missense variants. It was observed that 50% of pathogenic *OPA1* variants are located in GTPase domain and the GTPase effector (Han et al., 2022). Missense variants are more likely than null alleles to result in a more severe phenotype (Ham et al., 2019), which may be related to the dominant negative effect of the partially inactive *OPA1* alleles. The GTPase activity is thought to be impaired by this dominant negative effect, which reduces the stability of the mitochondrial inner membrane structure and causes possible proton leakage (Olichon et al., 2003, Olichon et al., 2007, Weisschuh et al., 2021). Notably, the above-described patient was from Nigeria and this mutation was observed in gnomAD in the heterozygous state in four people worldwide, all of whom were African or African American.

The patient reported here was diagnosed with CORD, therefore a plausible explanation for these findings is that the ADOA could be mistaken for CORD, particularly if the diagnosis was given in the early stages of the disease as the proper diagnosis is frequently discovered in the second decade (Votruba et al., 1998). However, this is not the case here as the patient was first diagnosed in her 40s by an experienced ophthalmologist. The *OPA1* gene is a large gene with many variants, including large rearrangements and deletions, which the WES approach may overlook (Toomes et al., 2001b). It is also known that the penetrance of the *OPA1* gene varies from 43 to 88%, demonstrating the existence of unknown modifying factors that contribute to the variety of phenotypic expression

(Weisschuh et al., 2021). Additionally, the existence of asymptomatic carriers in ADOA (Cohn et al., 2007, Fuhrmann et al., 2010) may suggest that this patient is an asymptomatic carrier and could have both conditions, with the CORD masking the OA. Given that it is a missense variant, the variant described here may be a benign rare variant, and the disease-causing mutation in this case may be explained by undetected variants in *ABCA4* or another known CORD gene (Al-Khuzaei et al., 2021). These variants may have occurred in non-coding regions as deep intronic variants (Khan et al., 2020), or were classified as benign, frequent variants (hypomorphic alleles) that may be pathogenic under certain conditions (Zernant et al., 2017, Runhart et al., 2018). Structural variants such as a duplication or a deletion could also occur in non-coding regions (Bauwens et al., 2019). Monoallelic variants in the other six genes identified in this patient should also be considered for further analysis, either by looking for a second variant in non-coding region or segregating these variants by screening additional family members. Alternatively, this patient could have an unidentified mutation in a new CORD gene.

Mutations in the *OPA1* gene were identified in two further cases with RP from the Leeds cohort (Prof. Chris Inglehearn, personal communication). Additionally, four patients with phenotypes not listed as associated with the *OPA1* gene were reported in the *OPA1* database (https://databases.lovd.nl/shared/genes/*OPA1*), including ocular albinism type I (OA1; OMIM 300500); polyneuropathy, hearing loss, ataxia, retinitis pigmentosa and cataract (PHARC; OMIM 612674); spinocerebellar ataxia-5 (SCA5; OMIM 600224); and autosomal recessive spastic paraplegia-18 (SPG18, OMIM 611225) (Le Roux et al., 2019). Although all of these *OPA1* variants could be rare benign variants, these findings potentially suggest phenotypic expansion in which *OPA1* variations might sporadically present as pigmentary retinopathies. Hence more research is required to establish the relationship between mutations in the *OPA1* gene and pigmentary retinopathies. Further analysis for this patient will be carried out by the research team if possible, by segregating the identified *OPA1* variant in additional family members, investigating the other variants in the other six genes, and performing WGS on the patient's DNA and the DNA of other family members to look for large structural variants and variants in non-coding regions.

3.3.2.2 *VCAN*

In a case with a diagnosis of FEVR (F1427), copy number variant analysis identified a heterozygous deletion in exon 8 of the *VCAN* gene. Heterozygous mutations in this gene are known to cause an autosomal dominant vitreoretinopathy called Wagner syndrome (OMIM 143200) (Graemiger et al., 1995, Miyamoto et al., 2005, Rothschild et al., 2013, Burin-des-Roziers et al., 2017). An optically 'empty' vitreous with avascular strands and veils is the hallmark of this condition. Along with these symptoms, patients with Wagner syndrome may also exhibit moderate myopia, cataract, an ectopic fovea, chorioretinal degeneration with atrophy and retinal detachment (Graemiger et al., 1995).

The majority of the mutations that have been reported were found to affect either the conserved acceptor splice site of intron 7 or the donor splice site of intron 8 (Miyamoto et al., 2005, Mukhopadhyay et al., 2006, Kloeckener-Gruissem et al., 2006, Ronan et al., 2009). These splice site mutations cause exon 8 to be skipped, which results in an imbalanced quantitative ratio of versican transcripts, with isoforms V2 and V3 (which lack exon 8) upregulated and isoforms V0 and V1 downregulated. This demonstrates the significance of exon 8 in controlling the quantity of chondroitin sulphate chains and, consequently, the interactions between versican and other extracellular components to maintain the physiology of the vitreous gel. Based on these data, it was assumed that imbalanced versican isoforms were the possible molecular mechanism underlying Wagner syndrome (Brézin et al., 2011, Li et al., 2020). In two families, Wagner disease was also found to be caused by heterozygous deletions involving *VCAN* exon 8 (Burin-des-Roziers et al., 2017), which resulted in an imbalance of mRNA isoforms similar to that caused by exon 8 splice mutations (Mukhopadhyay et al., 2006, Kloeckener-Gruissem et al., 2013).

According to a report (Brézin et al., 2011), a French family had been initially diagnosed with FEVR but it was later discovered that they carried a pathogenic mutation in the splice acceptor site of intron 7 of the *VCAN* gene. Another *VCAN* heterozygous mutation (c.9265+2T>A) was identified in a family with FEVR (Panagiotou, 2018) indicating that *VCAN* mutations can also cause FEVR, and Wagner syndrome is a vitreoretinopathy that needs to be considered in the differential diagnosis of FEVR. However, it is necessary to investigate the phenotype in case F1427 in further depth, and to confirm the detected heterozygous deletion by designing another set of primers to identify its exact breakpoints and confirm its segregation by sequencing other family members using WGS. It also might be possible to investigate the expression profile of the *VCAN* gene by doing a

reverse transcription PCR on total RNA which can be extracted from various human tissues including blood cells. These possibilities are currently being investigated by other members of the Leeds Vision Research Group.

3.3.2.3 *MFSD9*

In case 12565367, the patient was described as having a developmental delay with minor squint and visual impairment, making the *MFSD9* gene a plausible candidate gene. The *MFSD9* gene encodes a protein that belongs to the major facilitator superfamily domain containing proteins family (MFSDs), which comprise the largest group of the solute carriers (SLCs). It is predicted to enable transmembrane transporter (SLCs) activity to translocate nutrients, waste, and drugs (Hediger et al., 2004). MFSD proteins are present in mouse neurons, the CNS vascular system and the brain (Ben-Zvi et al., 2014, Perland et al., 2016). Plasma and intracellular membranes in humans have both been shown to contain MFSD proteins (Perland et al., 2017a, Lekholm et al., 2017). The expression levels of *MFSD9* were altered after food deprivation and high fat diet in mouse brain areas (Perland et al., 2017b). *MFSD9* was found to be expressed in both central and peripheral nervous system in rat (Sreedharan et al., 2011). It is evolutionary conserved and was identified in several vertebrates including human (Perland et al., 2017b).

In a GWAS study for type 2 diabetes, a SNP (rs2540317) in *MFSD9* was found to be associated with decreased risk of type 2 diabetes in Amish, nominal association in Pima Indian and Mexican populations (Rampersaud et al., 2007). A whole genome exon-targeted oligonucleotide array comparative genomic hybridization (array CGH) was used to study 256 patients with developmental delay and intellectual disability. It has found 15 potential pathogenic CNVs in various genes, one of them was a deletion in *MFSD9* (Bartnik et al., 2014). It was not mentioned whether these CNVs are homozygous or heterozygous. No visual impairment was reported in these patients but since the visual impairment in the case studied here was poorly defined, it may be mild and may therefore have been missed in the case (or cases) described by Bartink and colleagues. Alternatively, cases in the Bartink study may have not undergone any visual testing due to their intellectual disability. It should be noted that the non-syndromic retinal dysfunction (Khan et al., 2017) and non-syndromic macular dystrophies (Roosing et al., 2015b, Xiang et al., 2021, Priluck and Breazzano, 2022) have both been linked to *MFSD8*, a member of the same major facilitator superfamily domain containing proteins family (MFSDs). To determine whether further patients existed with a genetic condition due to biallelic variants in this gene, this finding was reported in GeneMatcher on 21 September

2021, but no positive response has yet been reported. Additional genetic and experimental testing is required to determine whether variants in this gene are a novel cause of syndromic visual and/or other defects.

3.3.3 Unsolved cases

In the cohort of the current study, another 10 of the 24 cases have not yet been solved. Since most of the cases in this study were single cases with no known family history, both heterozygous and biallelic variations (compound heterozygous and homozygous) were retained in filtering, and homozygous variants were only chosen in cases where reported consanguinity was present. The variants were prioritized based on CADD score (≥15) instead of excluding it based on the pathogenicity prediction scores to avoid excluding pathogenic variants with low scores. While other scoring tools are limited to a single information type (e.g., conservation) or are limited in scope (e.g., missense change), CADD score is based on a range of genomic features and incorporates several annotations into one score (Section 2.13.4.1) (Kircher et al., 2014, Rentzsch et al., 2019). Additionally, variants with frequencies of less than 1% for recessive cases and less than 0.01% for dominant cases were retained. It is possible to apply a harder threshold but doing so may lead to the exclusion of some putative pathogenic variants that may be more common than expected in the population. It was more challenging to infer the mode of inheritance or determine whether candidate mutations segregated in a manner that was consistent with causing the disease phenotype in most of these individuals because they had no known family history and, for the most part, no DNA samples from extended family members were available. For some cases, the diagnosis was ambiguous and uncertain, while for others it was syndromic, meaning that genes not listed on the RetNet website could be involved, and leaving open the possibility of new disease genes, large deletions, or structural variants as an alternative possible cause. Analysing additional genes other than the ones associated with IRDs and using WGS to look for variants in the entire genome, including introns and upstream promotor regions, will help in identifying genes not yet known to be involved in IRDs, as well as finding large deletions, structural variants, and variants in non-coding regions of known IRD genes. Giving a clear clinical diagnosis helps in identifying the pathogenic variants at the molecular level by allowing comparison of previous phenotypes associated with variants in this gene to the findings in the patient and the pattern of inheritance (Best et al., 2022). Sequencing additional family members' DNA will also help to narrow the list of variants, confirm the causative mutations, and segregate them with the specified phenotype.

3.3.4 Lessons learned and future directions

Since the emergence of next-generation sequencing, the IRD diagnosis rate has increased to 50 -70% (García Bohórquez et al., 2021). For several years, a variety of NGS-based approaches, including targeted sequencing (TS), WES and WGS, were used to accurately genotype the patients with different forms of IRDs (Neveling et al., 2012, Shanks et al., 2013, Zhao et al., 2015, Wang et al., 2018, Zhang et al., 2018, Ng et al., 2019, Martín-Sánchez et al., 2020, Nash et al., 2022). Targeted sequencing (TS) is an economical method that enables the analysis of known IRD genes regardless of clinical diagnosis. It allows multiplexing many samples in a single run since it only sequences small genomic regions with adequate coverage. Other cost-saving advantages of TS include smaller file sizes, which enable cheaper storage and quicker processing. Moreover, employing TS strategies reduce the risk of identifying secondary or incidental findings. Therefore, it was an excellent option for the genetic diagnosis of IRD (Ellingford et al., 2016, Dockery et al., 2017, Hitti-Malin et al., 2022). However, if the newly identified IRD genes are not in the panel, a panel redesign is required to include them.

On the other hand, WES can capture the protein coding exons, which make up around 1% of the human genome and comprise about 85% of the disease coding variants. It allows the detection of disease-causing variants in new IRD genes without the requirement to update the panel. New genotype-phenotype correlations may be possible as a result, leading to the resolution of a previously unsolved diagnosis. Since most of the cases included in this study were single cases and some of them had syndromic disease or an ambiguous diagnosis, WES was chosen to identify the pathogenic variants in these individuals. The genetic diagnosis was found in 8 cases with a detection rate of 33.3% which is less than other studies (50-70%) (Riera et al., 2017, Wang et al., 2018, Zhang et al., 2018). In the current study, WES was unable to solve several cases (66.7%) despite accurate diagnoses and complete family histories. This may be because the causative mutations may affect rare novel genes with unidentified roles that require large cohorts for validation (Tiwari et al., 2016). The mutations may also be located in non-coding regions, such as splice variants that lie deep in the introns and create pseudoexons and intronic retention. By incorporating the deep intronic sequences in the TS panel, TS approaches can be used to identify deep intronic variants (Webb et al., 2012, Khan et al., 2019b). WES can accurately detect SNPs and small indels but larger structural variations (SVs) and variants in regulatory regions including the changes in regions that affect the expression levels of the gene such as topologically associated domains (TADs) (de Bruijn et al., 2020), the 5’ or 3’ UTRs, promotors and enhancers, are not usually captured so are

challenging to detect or characterise. Regions with high GC content (e.g., some first exons) are difficult to sequence at the chemistry level resulting in few or no mappable reads (Kieleczawa, 2006). Highly repetitive regions such as ORF15 in the *RPGR* gene are not sufficiently covered by WES and ambiguous alignments to the reference sequence can occur, concealing the real pathogenic event (Ebbert et al., 2019, Watson et al., 2021).

Moreover, the identification of copy number variations (CNVs) from WES data has aided in determining the pathogenesis of the disease in a number of IRD types (Bujakowska et al., 2017, Ellingford et al., 2018, Zampaglione et al., 2020). A range of tools using depth of coverage of WES data were developed (Kadalayil et al., 2015) to detect loss or gain in copy number genes/exons that WES variant calling missed. Additionally, a number of splicing prediction tools, such as SpliceAI, are available and can be used to predict splice altering variants but not intronic variants (Jaganathan et al., 2019). The tool used in this work, ExomeDepth, is a CNV calling method that uses read depth information from WES data; it was found to perform well compared to other similar tools (Tan et al., 2014). Although, it has resulted in the detection of pathogenic CNVs in this study and other IRDs (Ellingford et al., 2018, Zampaglione et al., 2020), it also has a relatively high false positive rate (Tan et al., 2014). CNV analysis using WES is challenging due to both short read lengths and GC content bias (Teo et al., 2012). The precise breakpoints may be challenging to locate if they are outside of the genome's coding regions due to the discontinuous nature of WES data.

These impediments can be overcome using WGS, which covers the entire genome, including both coding and noncoding regions and provides a greater detection of structural variations, non-coding variants, and variants in GC rich regions (PCR free WGS) compared to WES and targeted sequencing. It was successful in finding missed variants by WES in IRDs (Nishiguchi et al., 2013, Carss et al., 2017, Nash et al., 2022). Using WGS can also be of value in identifying modifier loci that may cause variable expressivity and penetrance of the variants in the major genes. These modifiers can be hypomorphic alleles such as the ones identified in *RPGR*-associated RP (Cehajic-Kapetanovic et al., 2020a) and *RDH12*-associated CORD (Wang et al., 2022). However, WGS is expensive and generates a large volume of data which must then be stored for future analyses. Alternatively, whole gene sequencing (GS), which enables capturing exonic, intronic, and regulatory regions for the target gene, is particularly effective in conditions where phenotype is a strong predictor of genotype and could also be used in the genes with monoallelic variants identified in this study. GS was used to investigate

1054 unsolved Stargardt disease cases by sequencing the *ABCA4* gene using single-molecule molecular inversion probes (smMIPs). In 25% of the cases, this method allowed for the detection of pathogenic structural variations and deep intronic variants (Khan et al., 2020).

Another possible explanation for the unsolved cases reported here is that they may in fact not be Mendelian inherited retinal dystrophies but could rather be the result of largely polygenic and multifactorial disease such as age-related macular dystrophy. It is evident that a number of IRDs can mimic the AMD phenotype, which, for example, has in the past led to debate about whether cases with *ABCA4* variants in AMD cohorts were proof that *ABCA4* variants contributed to AMD, or rather they were misdiagnosed cases of late onset Stargardts disease (Stone et al., 1998). Alternatively, there have been reports of digenic inheritance in IRDs, and it is possible that even more genetically complex forms may exist in a few rare cases. Finally, it remains possible that the IRD seen in these cases is in fact not genetic in origin, but is, for example, the result of long-term usage of drugs (Yusuf et al., 2022) or an immunological disorder (DeMaio et al., 2022).

For the unsolved cases reported here, employing WGS or whole gene sequencing may increase the detection rate (Tiwari et al., 2016). However, WGS may not be able to read across regions, such as ORF15, and due to its high cost and computational demands, it cannot be used in everyday routine diagnostics. This shows that a trustworthy, scalable, and affordable method to sequence across these difficult to sequence regions is required. In the following chapters, a series of cases with a large *CRX* deletion that was overlooked in early NGS studies (Chapter 4) and the use of long read sequencing techniques to read across a mutational hotspot (ORF15) in the *RPGR* gene, a region that is both difficult to sequence and difficult to align (Chapter 5), will be discussed.

Chapter 4

Late-onset autosomal dominant macular degeneration caused by deletion of *CRX* gene

4.1 Introduction

The Cone-Rod homeoboX-containing gene (*CRX*; OMIM 602225) encodes a transcription factor with a homeodomain most homologous to those of the orthodenticle homeobox 1 and 2 (OTX1 and OTX2) transcription factors. It is highly expressed in rod and cone photoreceptors of the retina and pinealocytes of the pineal glands and is crucial for the development and survival of photoreceptor cells (Freund et al., 1997, Furukawa et al., 1997, Furukawa et al., 1999) acting through synergistic interaction with other transcription factors, notably neural retina-specific leucine zipper protein (NRL), retinal homeobox protein (RAX) and nuclear receptor subfamily 2 group E member 3 (NR2E3) (Chen et al., 1997, Kimura et al., 2000, Peng et al., 2005). The *CRX* gene is located on chromosome 19q13.33 and comprises four exons, with exon one being non-coding. The remaining exons encode a 299-amino acid protein, with exon 3 (34-84 aa) covering most of the DNA-binding homeobox domain (homeodomain) and exon 4 (85-299 aa) encompassing the remaining >70% of the protein sequence, including a WSP (tryptophan-serine-proline) motif and OTX tail (Hull et al., 2014).

The *CRX* locus was first identified in 1994 as a site associated with autosomal dominant cone-rod dystrophy (CORD) in a single large family (Evans et al., 1994). Subsequently, *CRX* mutations have been associated with autosomal dominant CORD (Freund et al., 1997, Swain et al., 1997), autosomal dominant Retinitis pigmentosa (RP) (Sohocki et al., 1998), and both dominant and recessive forms of Leber congenital amaurosis (LCA) (Sohocki et al., 1998). More recently, *CRX* variants have also been shown to cause an autosomal dominant form of adult-onset macular dystrophy said to simulate benign concentric annular macular dystrophy (BCAMD), with a bull's eye macular lesion and fairly well-preserved visual acuity (VA) (Hull et al., 2014, Griffith et al., 2018, Nasser et al., 2019, Khan et al., 2019a, Ng et al., 2020).

Pathogenic variants in *CRX* are rare and most are found in a heterozygous state causing dominant disease. In total, over 100 disease-causing variants have been reported (www.lovd.nl/*CRX*), including missense, premature stop codon and frameshift changes and multiple exon deletions (Yi et al., 2019). Hull and colleagues reviewed 43 pathogenic variants (41 heterozygous and two homozygous) and noted that all but 2 consisted of

either missense mutations in exon 3, or frameshift and nonsense mutations occurring beyond the homeodomain in the large terminal exon 4 (Hull et al., 2014). A similar pattern was also noted in more recent studies (Yi et al., 2019, Nishiguchi et al., 2020). Normally, premature stop codon or frameshift variants other than those in or close to the last exon of a gene are generally thought to lead to nonsense-mediated decay (NMD) of the affected transcript, meaning they are effectively null alleles that produce no protein (Chang et al., 2007). However, premature stop codons caused by variants in or close to the last exon are protected from NMD by the same mechanism which ensures the normal stop codon does not elicit NMD. In *CRX*, since almost all the pathogenic frameshift and premature stop variants occur in the last exon, they are likely to escape NMD (Kurosaki and Maquat, 2016), meaning that a truncated protein would be produced.

One previous report described homozygosity for a putative null variant, which creates a premature stop codon in the first coding exon (exon 2), c.25insG (p.Pro9fs*9) (GenBank accession number AF024711), as the cause of LCA in a family (Silva et al., 2000). The authors proposed since the heterozygous carriers were unaffected that haploinsufficiency of *CRX* was not sufficient to cause a retinal disorder. Two more recent reports have also described putative null variants that appear to be tolerated and cause little discernible pathology in heterozygous state. The first describes a family with autosomal dominant CORD carrying a heterozygous deletion of *CRX* exons 3 and 4 (breakpoints undefined), that would be predicted to lead to NMD of the mutant *CRX* transcript. However, interpretation in this family was complex due to a *PRPH2* variant also segregating. One individual was shown to carry only the *CRX* deletion and to have a CORD phenotype, while another with the same genotype did not, suggesting that haploinsufficiency (the situation when one copy of a gene is inactivated or deleted and the other remaining functional copy of the gene is insufficient to produce the required gene product to maintain normal function) may not be sufficient to cause disease in all cases (Jones et al., 2017). The second reported a consanguineous family with LCA in which there is a homozygous 56 kb deletion of *CRX* and the two flanking genes (Ibrahim et al., 2018). The heterozygous carrier parents were also examined, and two of the four were found to have subclinical macular abnormalities but no evidence of clinical disease.

The work in this chapter was initiated by Dr Claire Smith (Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK), who identified a heterozygous deletion encompassing *CRX* and neighbouring genes *TPRX1* (Tetrapeptide repeat homeobox 1; OMIM 611166) and *SULT2A1* (Sulfotransferase

family 2A, dehydroepiandrosterone-preferring member 1; OMIM 125263) in a patient with late-onset macular dystrophy. This finding led her to re-examine and verify data generated by Dr James Poulter (Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK), from which a similar finding had been discounted on the assumption that *CRX* deletion carriers were asymptomatic so this alone could not account for disease. She then developed a PCR assay for the deletion and used it to screen further IRD patients, and also a cohort of patients with Age-related macular dystrophy (AMD). This led to the identification of a further case with late onset MD and a case with AMD. At this point Dr Smith moved on to another project and the author continued the research on this project.

Through collaboration within the UK inherited retinal disease consortium (UK IRDC), an additional four samples with late-onset MD and *CRX* deletions were identified. The presence of a similar deletion in these additional cases was confirmed using the PCR assay developed by Dr Smith (Section 2.4.4). To determine whether all cases had the exact same deletion, and to determine the exact location and likely underlying mechanism of the deletion, PCR products were cloned (Section 2.7) and Sanger sequenced (Section 2.9) in some cases, while in others, Sanger sequencing alone was used to characterise the breakpoints. Haplotype analysis was performed both by microsatellite genotyping (Section 2.4.5) and using SNPs in the sequence generated immediately adjacent to the breakpoint, to discern whether the deletion was due to separate mutation events at a mutation hotspot or to a founder effect.

The results of these findings were published in the journal Ophthalmology, with the author as joint first author with Dr Smith (Yahya et al., 2023).

4.2 Results

All the genomic coordinates given refer to the GRCh37/hg19 human genome assembly.

4.2.1 Variant identification

The genomic DNA from a female patient who received a diagnosis of CORD with central macular atrophy in her 50s (patient 4776, Table 4.1, Figure 4.1A) was subjected to WES and the initial variants list was compared to the genes, in which variants cause IRDs, listed on RetNet website (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes). The family history in this patient is unknown. She is one of an affected sibling pair, and due to the late onset of the disease, dominant inheritance was possible but other modes of inheritance could not be excluded. Only three heterozygous variants with CADD score of 15 or above in genes *CDH23*, *MERTK*, and *CDH3* were identified. It appeared unlikely that these variants were causing the disease because all three genes are known to be involved in diseases with autosomal recessive inheritance. There is no published evidence that pathogenic variants in these genes can cause dominant diseases, or that they act in a digenic fashion to cause disease together, though it was not possible to exclude these hypotheses. ExomeDepth analysis was performed by comparing the read depth profile of this sample with nine other control samples sequenced in the same run lane. The read depth analysis identified a minimum heterozygous deletion of chr19:48,305,033-48,389,514 (Figure 4.2A), which included the entire coding regions of *TPRX1*, *CRX*, and *SULT2A1*. In a second female patient who was diagnosed with bull's-eye maculopathy at age 78 (patient 3888, Table 4.1, Fig 4.1B), read depth analysis in data generated by sequencing 108 inherited retinal disease associated genes (Weisschuh et al., 2018) revealed a heterozygous deletion minimally spanning chr19:48,337,702-48,343,224 (Figure 4.2B), which covered all *CRX* exons. It should be emphasised that in neither case was there a significant second variant found in the *CRX* gene. A third male individual who was diagnosed with late-onset MD at age 20 (patient GC16591, Table 4.1, Figure 4.1D), was found to have a heterozygous deletion of the *TPRX1, CRX*, and *SULT2A1* genes in genome sequence generated by the UK 100,000 Genomes Project (Figure 4.2C). The breakpoints were defined, and the centromeric breakpoint was located at approximately chr19:48,296,800. The telomeric breakpoint was difficult to place accurately due the presence of a block of 10 tandem repeats (~ 5.3 kb) containing small NF-90 associated RNAs (SNARs), spanning approximately chr19:48,407,000-

48,465,000. Investigation of all coding and non-coding variants in the *CRX* region and beyond has been made possible by whole genome sequencing. Functional analysis through Hidden Markov Models (FATHMM; version 2.3) (Shihab et al., 2013) was used to predict the consequences of all variants in the region chr19:47,010,944- 49,499,828, but again no significant second *CRX* variant was identified. Furthermore, no variants in the genes indicated above (*CDH23*, *MERTK*, and *CDH3*) were found in these other cases therefore, it seems far more likely that the *CRX* deletion alone explains the phenotype in these cases.

Lab ID/sex	Age at onset	Diagnosis	Clinical notes	Deletion detected by	Breakpoint confirmed in
4776/F	50s	Bull's eye maculopathy	Nyctalopia 2nd decade, reduced central visual acuity (6/60 OD, 6/24 OS) 6th decade. Posterior pole staphyloma. Sister affected, same diagnosis	Exome sequencing	Sequenced PCR product
3888/F	78	Bull's eye maculopathy		MIPs screening	Cloned amplimer
GC16591/M	20	Bull's eye maculopathy	Mild photo aversion, normal peripheral fields, no family history. Acuities 6/60 bilaterally, symmetrical area of degeneration at posterior pole. Electrophysiology has shown minimal progression in 10 years, generalised retinal dysfunction affecting rods and cones.	Genome sequencing	Sequenced PCR product
5587/F	36	Macular dystrophy	Photophobic, poor accommodation, normal colour vision. MfERG - poor central responses. pERG Extinguished R/L. Snellen visual acuity 6/12 OD, 6/18 OS.	Targeted sequencing	ND
5588/F	50	Macular dystrophy	Vision never good, sat at front of class at school. Noticed distortion at 50. Pendular nystagmus, not photophobic, no nyctalopia. LogMar visual acuity 0.46 R, 0.64 L. Vitreomacular traction on OCT. MfERG - low amplitude bilaterally, pERG non recordable. Mother of 5587	Targeted sequencing	Sequenced PCR product
5581/F	Late 40s	Macular dystrophy	Father developed rapidly progressive macular degeneration at 52, died at 73 severely sight impaired. No other family members symptomatic	Targeted sequencing	PCR
5598/M	53	Cone rod dystrophy	Visual acuity 20/300 both eyes, vision loss began at 53 yrs. AF with few hyperfluorescent lesions (not pisciform). Low photopic flicker on FFERG, MfERG with low foveal waveforms. OCT with foveal atrophy	Exome sequencing	Sequenced PCR product
3258/M	Late 60s	Wet AMD	Presented with poor vision, scar in left eye. Asymptomatic in right eye despite early wet AMD. No previous visual problems. Visual acuity 6/18 OD, 3/60 OS	PCR	Cloned amplimer

Table 4.1: Phenotypes and screening method used to detect and characterise the deletion for each patient. (Reused from (Yahya et al., 2023), no permission needed). AMD = age related macular degeneration, OD = right eye, OS = left eye, MfERG = multifocal electroretinogram, pERG = pattern electroretinogram, R = right, L = left, AF = autofluorescence imaging, LogMar = logarithm of the minimum angle of resolution, OCT = optical coherence tomography scan, FFERG = full field electroretinogram, MIPs = molecular inversion probes, PCR = polymerase chain reaction, ND = not done.

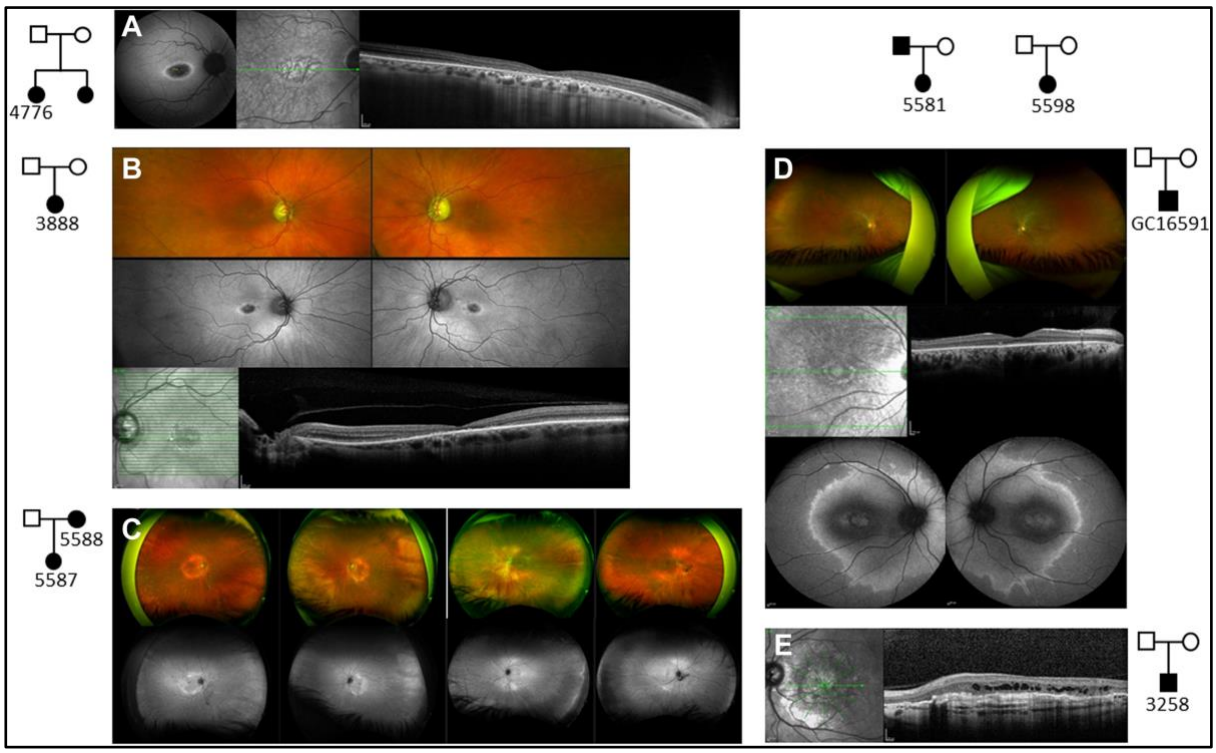


Figure 4.1: Clinical and family information in individuals with *CRX* deletion. A. Family tree, fundus autofluorescence image, and OCT for individual 4776. The fundus shows a central area of reduced autofluorescence with a surrounding ring of increased autofluorescence, consistent with the bull's-eye maculopathy seen on clinical examination. On the OCT image, there is central loss of the outer retinal layers with choroidal hypertransmission. **B.** Family tree, pseudo-color fundus photographs, and autofluorescence images and OCT for individual 3888. There is reduced fundus autofluorescence centrally in both eyes, with a surrounding ring of increased autofluorescence, consistent with a diagnosis of bull's-eye maculopathy. The structural OCT image shows loss of outer retinal lamination centrally. **C.** Family tree and pseudo-color and autofluorescence fundus images for individuals 5588 (mother, left images) and 5587 (daughter, right images). A ring of increased autofluorescence is visible in the mother, whereas in the daughter patches of increased autofluorescence are also visible in the macular region. **D.** Pseudo-color fundus, OCT, and autofluorescence images and family tree for individual GC16591. Bull's-eye maculopathy is evident, but there are also changes nasal to the optic disc and a ring of increased autofluorescence around the central retina and posterior pole. **E.** Structural OCT images and family tree for individual 3258. Subretinal hyperreflective material and intraretinal cysts are evident on the OCT image, indicating a diagnosis of wet age-related macular disease. (Reused from (Yahya et al., 2023), no permission needed).

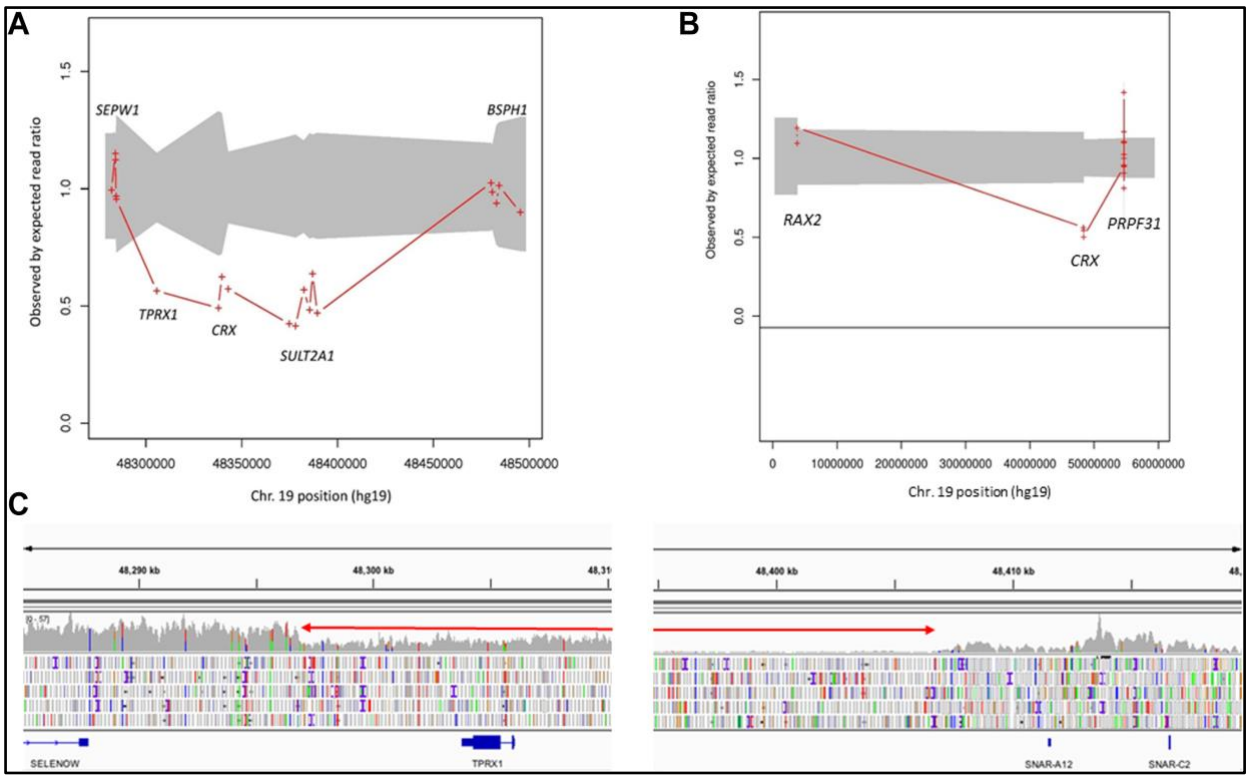


Figure 4.2: Sequence analysis in individuals with *CRX* deletion. A. ExomeDepth plot of whole-exome sequencing data across the deleted region in individual 4776, showing the drop in read depth over *TPRX1*, *CRX*, and *SULT2A*. **B.** ExomeDepth plot of molecular inversion probes–generated sequence data from individual 3888, showing a drop in read depth for all exons of the *CRX* gene. **C.** Integrative Genomics Viewer (IGV) image of genomic sequence at each end of the deletion in individual GC16591. The left-hand image shows a clear drop in read depth across the deleted region at the centromeric end of the deletion. However, the 5.3 kb tandem repeat cluster at the telomeric end means sequences align multiple times, giving the false impression of a sudden increase in read depth at the point where these repeats start, masking the precise end point of the deletion in the genomic sequence. An increase in read depth at this point is also visible in normal genomic sequence (not shown). (Reused from (Yahya et al., 2023), no permission needed).

4.2.2 PCR assay

Primers were designed to confirm the presence of a deletion in the above cases and to identify the exact breakpoints in each case. Given the repetitive nature of the sequence adjacent to the deleted region, the common ethnic origin of the subjects, and the fact that panel and exome sequencing do not define breakpoints, it was hypothesized that the deletion might be the same or similar in all three cases. Multiple primers were designed using whole genome sequencing data. Due to the repetitive nature of the region, it was impossible to design unique PCR primers.

Various primer combinations were tried by Dr Smith and primers CRXhetdel3_F (CCTCCCATCTCAGCCTCCTA) and CRXhetdel3_R (CAAGGAGGAATGTGCAGTGG) were the optimum primer combination. The PCR conditions were then further optimised by the author by testing different annealing temperatures (61°C, 63°C, 65°C, and 67°C) and the optimum conditions used were as described in Section 2.4.4. With these conditions in place, these primers should not amplify any product in normal human genomic DNA; it amplified a unique product of approximately 1150 bp only when the deletion was present (Figure 4.3).

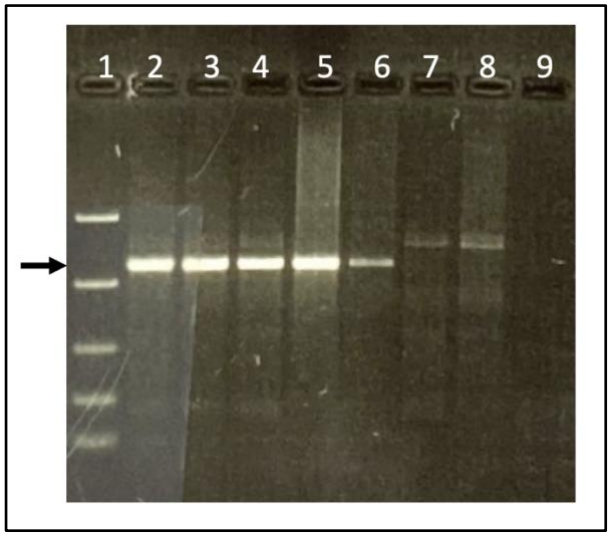


Figure 4.3: PCR analysis in individuals with *CRX* deletion. Agarose gel electrophoresis (1.5%) of the PCR amplification product (~1150 bp; marked by arrow) of primers CRXhetdel3_F and CRXhetdel3_R. Lane 1 = Bioline EasyLadder1 size marker (100, 250, 500, 1000 and 2000 bp), Lanes 2-6 show PCR amplification products in five individuals with *CRX* deletion. Lanes 7 and 8 show non-specific amplification products from control individuals. Lane 9 is a no DNA control.

These primers were then used to screen additional cases and found to amplify the same-sized band in each case, implying the deletion breakpoints were similar in each case (Section 4.2.3).

4.2.3 Additional screening

A screen for mutations in 105 genes associated with inherited retinal diseases (Ellingford et al., 2016), identified a mother (5588) and daughter (5587) who were diagnosed with MD at ages 50 and 36, respectively, as well as a female patient (5581) who was diagnosed with MD in her 40s and has a father who developed rapidly progressing MD at age 52. Read depth analysis identified a heterozygous deletion of chr19:48337702-48389514 encompassing all *CRX* exons in each of these cases. Patient 5598, a male from the United States who was diagnosed with CORD at age 53, was found to have a deletion spanning the *CRX* gene in a whole exome sequencing-based panel screening for individuals with inherited eye diseases (Guidugli et al., 2019). PCR amplification with the above primers (CRXhetdel3_F and CRXhetdel3_R) was performed to confirm the presence of the deletion in these four individuals. In each case, a band of approximately 1150 bp was amplified, confirming the presence of the deletion with similar breakpoints to those described above (Figure 4.4). Screening data and clinical information for these patients can be found in Table 4.1 and Figure 4.1.

Primers CRXhetdel3_F and CRXhetdel3_R were also utilised to test for the presence of the same deletion in a cohort of 405 white British patients with exudative (wet) AMD in order to evaluate the hypothesis that the late-onset macular disease in these individuals may overlap with AMD. In patient 3258 (Table 4.1, Figure 4.1E), who was diagnosed with wet AMD in his late 60s, the amplification yielded a band of approximately 1150 bp (Figure 4.4), confirming the presence of the deletion with similar breakpoints. With these primers, no product was amplified in the remaining cases.

Moreover, primers CRXhetdel3_F and CRXhetdel3_R were used by Dr Smith to test whether this deletion is common in the normal population by amplifying 382 white British control DNA samples, all of which proved negative. These primers were also employed by Dr Smith to establish that the deletion reported by Ibrahim et al (Ibrahim et al., 2018) (data not shown), which deleted the same three genes and was claimed to have caused subclinical foveal abnormalities in two of the four carriers examined, was not the same deletion as the one described here. The heterozygous deletion carriers identified by Ibrahim et al. were younger than the individuals studied here (Rob Koenekoop, personal communication, June 29, 2018) and may thus be at risk of developing MD in later life.

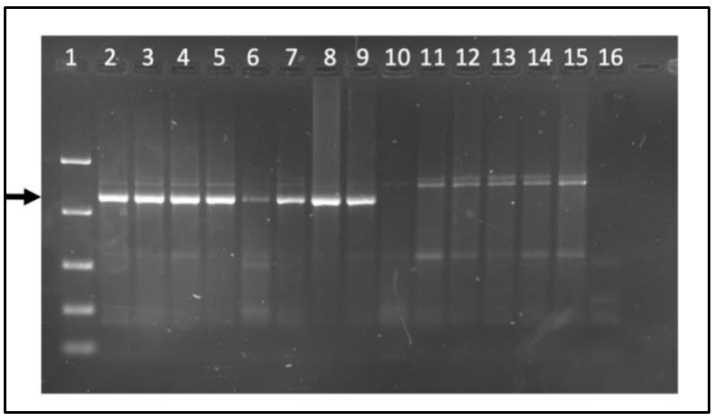


Figure 4.4: PCR analysis in additional individuals with *CRX* deletion. Agarose gel electrophoresis (1.5%) of the PCR amplification (~1150 bp; marked by arrow) of primers CRXhetdel3_F and CRXhetdel3_R in additional cases. Lane 1 = Bioline EasyLadder1 size marker (100, 250, 500, 1000 and 2000 bp), lane 2 = 3258, lane 3 = 3888, lanes 4 and 5, clinician consent withdrawn due to reasons outside of the author's control, lane 6 = 4776, lane 7 = 5588, lane 8 = GC16591, lane 9 = 5598. Lanes 10-13 show amplification products in other samples with late-onset macular dystrophy which proved negative on testing, lanes 14 and 15 show non-specific PCR products from control individuals and lane 16 is a no DNA control. (Reused from (Yahya et al., 2023), no permission needed).

4.2.4 Sanger sequencing and breakpoint characterisation

To determine the precise breakpoints in each case, the breakpoint PCR products from patients 3888 and 3258 were cloned and Sanger sequenced. Due to the presence of nonspecific bands in all cases, the band of interest was gel extracted using MinElute gel extraction kit following the steps detailed in Section 2.6 and then purified to get a cleaner product. Following that, the product was cloned using a pCR 2.1 TOPO vector (the vector map can be found in Appendix C.1) applying the TA cloning method described in Section 2.7.1. The cloning reaction was then transformed into chemically competent *E. coli* cells and plated on LB-agar plates and incubated overnight. To generate eight clones for each case, the colonies from each sample were picked from the plates and transferred to eight tubes per sample containing LB-broth to grow the cultures (Section 2.7.2). The plasmid DNA was then purified using QIAprep spin miniprep kit (Section 2.7.3) and the clones were digested with *EcoR1* restriction enzyme (Section 2.7.4). The insert in the eight clones was checked by agarose gel electrophoresis (1%) (Figure 4.5).

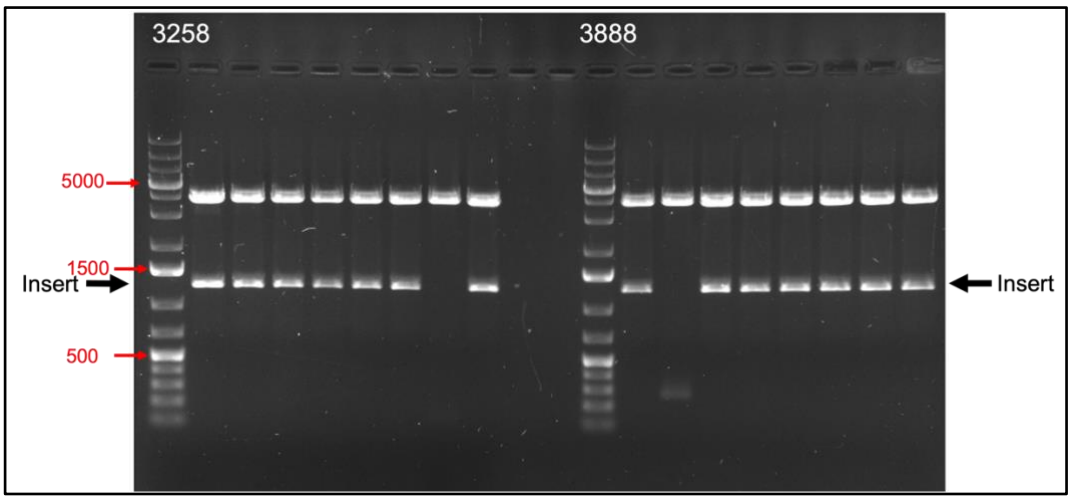


Figure 4.5: A test digest of eight colonies obtained by TA cloning. Agarose gel electrophoresis (1%) showing the insert of eight clones of the *CRX* deletion in patients 3258 and 3888 digested out with *EcoR1* restriction enzyme. The insert size was ~1200 bp as shown by the arrow. The first lane is for the GeneRuler 1kb plus DNA ladder (75, 200, 300, 400, 500, 700, 1000, 1500, 2000, 3000, 4000, 5000, 7000, 10000, and 20000 bp). Lanes 2-9 show the insert bands in eight clones from each case.

Sanger sequencing was then performed on two clones for each case, using six different internal primers (Appendix C.2). The breakpoints in other four deletion carriers (4776, 5588, GC16591 and 5598) were determined by Sanger sequencing alone using the same sequencing primers. Utilizing a software called DNA Baser (https://www.dnabaser.com/), the resulting contigs were assembled and manually examined to avoid potential mapping errors over repetitive regions. 53 bp at the start of cloned PCR products were removed using the VecScreen BLAST tool (https://www.ncbi.nlm.nih.gov/tools/vecscreen/) because they were vector sequences. Figure 4.6 shows the breakpoint sequence generated by Sanger sequencing and Figure 4.7 shows the Sanger sequencing chromatogram in addition to a graphic representation of the genomic locus and deletion.

CCTCCCATCTCAGCCTCCTAAGTAGCTGGAATTACAAGTGTGTACCATCATGCTCAGTTAATTTTAATATTTTTATAGAGATGGGGTCTCATTATA
TTGCCCAGACTGGTCTCAAGCAGTCCTCCTGCCTCAGCCTCCCAAAGTTTTTTTTTTGAGACGGAGTCTTGCTCTGTCACCAGGCTGGAGTGCAGT
GGCATGATCTCGGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCATGCACCAC
TACACCCAGCTAATTTTTGTATTTTTAGTAGAGATGAGGTTTCACCATGTTGGCTAGGATGGTCTCGATCTCTTGACCTTGTGATCCACTCGCCTC
GGCCTCCCAAAGTGCTGGGATTTCAGGCGTGAGCCACTGCATGTGGTCCATTTTAGTTTTAATATGCATTTCCCTGATGATTTTTGATGTCATGGG
CTTACTGGCCTTTTTGCATATCTTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGGAGGCTGGCTCTGTCACCCAGGCTGGAGTCAGTGGCACGATC
TTGGTTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCCCCTGCCTCAGCCTCCTGAACAGCTGGGATTACAGGCACCCGTCCCCATGCCCGG
GTAATTTTTGTATATTTAGTAGAGATGGGGTTTTCAGCATGTTGGCCAGGCTGGTCTCGAACTCCTTACCTCAAGTGATCCATCCGCCTCAGCCTC
CCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCCGGCTTCTTGAATTAGTTTCTAGGAGAAGCCAAGAACCCCCCCGGGCTAAGCCTCAATT
TTGGGGCTCGCCTGTGCTGCATCAGCTTCACACCTAGAGAAGAGAGTGAAAGAGAAGGGAGAGCCGTCCTTTTGGGCTGCCTTGAAATAAAGGTTG
ACCCATGGGTTCATGTATTGTCATTCATTCTCATTTTCTCTTCCTCTCTTTCTCTCTCCCCTGCCCCTTAAGTTAGCTTTTCTAAACCCTACCT
GGATAAGGATAAGAAATAGAAGGAGGGGACACTTTAGGATGCTACAAAATAAAATACAACAACAACAACAATAAAAGCAGCAGCAACAACAAC

Figure 4.6: Breakpoint sequence. DNA sequence from patient 4776 spanning the breakpoint PCR (~ 1150bp), with breakpoint assigned by BLAT in bold, underlined and marked in red (CC). On the centromeric side, the deletion allele was in cis with the C allele of SNP rs11673598 (bold, underlined and marked in orange), found in 17.4% of Europeans in gnomAD, in all deletion carriers sequenced. On the telomeric side, the deletion was in cis, again in all deletion carriers tested, with the A allele of SNP rs879322303 (bold, underlined and marked in orange), which accounts for 2.5% of Europeans in gnomAD (Table 4.2).

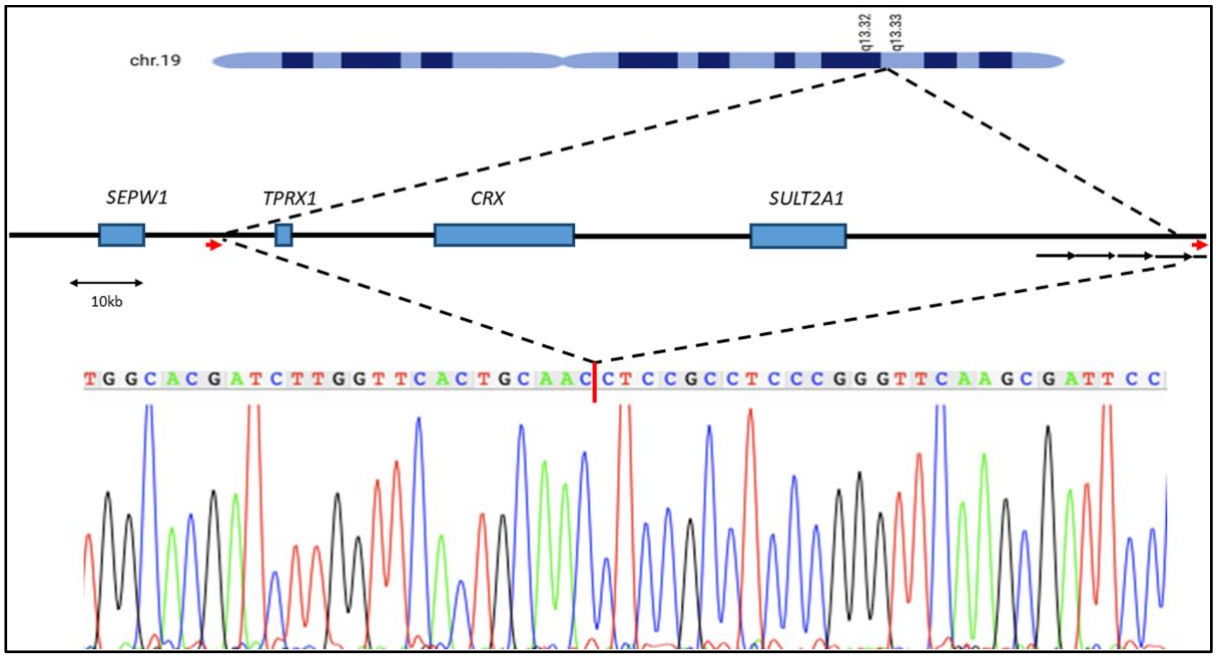


Figure 4.7: Graphical representation and Sanger sequencing chromatogram of the *CRX* deletion. 19q13.33 genomic region spanning *TPRX1*, *CRX*, *SULT2A1*, and adjacent genes, with cut-away showing the deleted region and below that a sequence trace of the breakpoint. Red arrows indicate the AluSx1 short, interspersed repeats within which the breakpoints occur, and the black arrows at the distal end of the locus denote small NF-90 associated RNA–containing tandem 5.3 kb repeats. (Reused from (Yahya et al., 2023), no permission needed).

Using the assembled sequence from Sanger sequencing, breakpoint locations were then identified using the BLAT function of the UCSC genome browser and the deletion was the same in each case. However, due to local sequence homology, the precise location of the breakpoint was ambiguous. Therefore, it was assigned as the most 3-prime base that

could apply, according to Human Genome Variation Society (HGVS) guidelines. The centromeric breakpoint was assigned as chr19:48,296,898. The telomeric breakpoint gave 100% alignment within the fourth 5.3 kb small NF-90 associated RNA containing repeat (SNARs), giving a location of chr19:48,422,892, making this a deletion of 125,995 bp (g. 48,296,898_48,422,892del). However, because of the 5.3 kb repeats, the telomeric sequence also aligned at 10 additional locations between chr19:48,406,792 and chr19:48,460,162, with sequence identities ranging from 99.5% to 96.5%.

The sequences adjacent to the breakpoints, obtained either via cloning and sequencing or from direct sequencing of the PCR product in the six cases listed above, were analysed using Emboss Matcher (EMBL-EBI, https://www.ebi.ac.uk/Tools/psa/emboss_matcher/). This revealed 294 bp of sequences with 78% homology immediately adjacent to the deletion breakpoints at each end. Furthermore, the repeat masker function of the UCSC genome browser revealed that these sequences are AluSx1 sequences.

4.2.5 Haplotype and SNP analysis

To detect whether the deletion is a result of a common ancestral founder or a mutational hotspot, genotyping was carried out with microsatellite markers spanning the ~3 Mb (chr19:47,010,944-49,858,428) or approximately 6.5 cM of chromosome 19 between the microsatellite markers D19S412 and D19S550, containing the *TPRX1*, *CRX* and *SULT2A1* genes (Kong et al., 2002). The genotyping was performed as described in Section 2.4.5, using five markers spanning the *CRX* locus, one intragenic (D19S902), two distal and two proximal to the *CRX* gene. As marker D19S902 is within the deleted region, the presence of only one allele in every case provides further confirmation that one of the alleles was deleted.

Haplotype analyses in these cases were restricted because it was not possible to obtain additional family members, which would have allowed the establishment of allele phase in each case. Nevertheless, for example, DNA from case 3258 has a genotype of 174/174 for marker D19S606, while individual 5581 has genotype of 180/184 (Figure 4.8) and individual GC16591 has genotype 176/182. It is therefore obvious that none of the 6 alleles carried by these three individuals carries the same genotype for this marker, so they cannot share a common haplotype at this marker. Similarly, for each other microsatellite marker tested it was not possible to identify a single allele shared by all affected individuals, which would suggest a founder haplotype (Table 4.2). From the table, it can be concluded that marker D19S606 genotypes show the deletion segregating

on at least 2 different haplotype backgrounds, whereas marker D19S879 genotypes imply segregation with at least three different haplotype backgrounds.

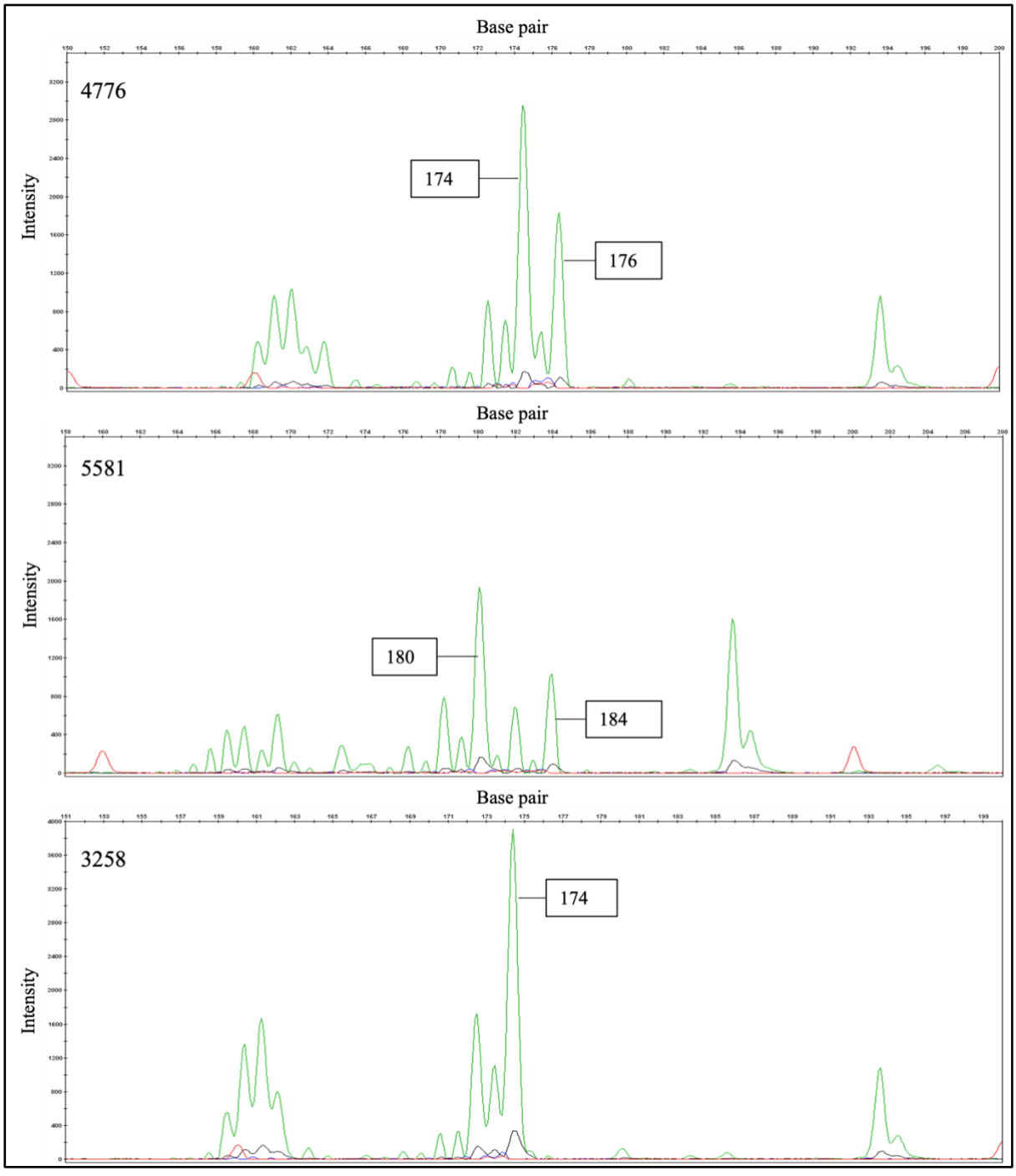


Figure 4.8: Representative electropherograms of marker D19S606 for cases 4776, 5581, and 3258. Case 4776 has a genotype of 174/176 for this marker while case 5581 has genotype of 180/184 for the same marker. It is obvious that these two individuals carry four different alleles indicating that they do not share a common haplotype at this marker. Cases 4776 and 3258 share a haplotype of allele 174 and could potentially be related.

Marker	Position (hg19)	4776	3888	3258	5581	5587	5588	GC16591	5598	Control
D19S412	47010944-47011334	92/102	90/102	90/92	90/90	90/92	ND	99/109	ND	94/101
D19S606	47973563-47973917	174/176	180/184	174/174	180/184	174/180	ND	176/182	ND	176/178
rs11673598	48296505	C	C	C	ND	ND	C	C	C	ND
D19S902	48332028-48332279	198/ -	206/ -	208/ -	210/ -	198/ -	ND	198/-	ND	210/212
rs879322303	48423428	A	A	A	ND	ND	A	A	A	ND
D19S879	49515918-49516189	251/253	261/263	251/251	251/251	255/259	ND	255/257	ND	259/263
D19S550	49858037-49858428	143/156	143/156	143/156	143/156	143/156	ND	145/161	ND	143/143

Table 4.2: Microsatellite genotypes and phased SNP alleles in cis with the deletion in *CRX* deletion carriers. D19S412, D19S606 (proximal to the deletion), D19S902 (in *CRX* intron 1), D19S879 and D19S550 (telomeric to the deletion) were genotyped in patients 4776, 3888, 3258, 5581 and 5587. These microsatellite markers lie within a 3Mb region centred on the *CRX* gene. D19S902 is in intron 1 of the *CRX* gene. The phased genotypes for SNPs immediately outside the breakpoints on each side, shown in Figure 4.6, were derived from sequencing of either cloned or PCR amplified breakpoint containing amplimers. ND = not done. (Reused from (Yahya et al., 2023), no permission needed).

In addition, the breakpoint PCR product sequences were compared to the hg19 reference genome using the BLAT function of the UCSC genome browser. This identified two non-reference variants relatively rare in the European population which were present in all breakpoint sequences. The first variant (T>C; rs11673598) is a known SNP and the patients in this study all carry the rare allele (C) with an allele frequency of 17.4% in genomAD. The second variant (C>A; rs879322303) is also a known SNP with allele frequency of 2.5% in genomAD. Again, all of the deletion carriers studied here had the rare A allele.

Given that these variants are within the 500 bp immediately adjacent to the breakpoint, this finding contrasts with the results obtained by the microsatellite markers and may in fact suggest that this deletion only occurred once and that these alleles form a founder haplotype. It should be noted that there was a common region of variation between all the clones, the poly (T) run in the middle of the sequence, with size ranging from 25 to 32. However, the sizes of these variants were different even between clones from the same patient, and since only one allele was amplified, this suggests that they were artefacts.

Even when using proofreading *Taq* DNA polymerases, poly (T) runs are a hotspot for PCR-induced errors (Clarke et al., 2001). Although this poly (T) was noted in dbSNP (rs59456686), with a size of 25 appearing to be the most common length, there is always a chance that PCR-induced errors have been uploaded into the databases.

4.3 Discussion

In this study, a heterozygous deletion of the entire coding region of *CRX* and both flanking genes has been shown to cause a late-onset dominant macular dystrophy in six apparently unrelated families, and in one patient with AMD. The deletion was successfully detected in MIPs sequence, WES and WGS, in three different individuals. Determining the specific breakpoint position in the NGS data was difficult due to the presence of a block of 10 tandem repeats that included SNAR repeats. Subsequently though, it was possible to identify the exact breakpoints by combining the PCR test with cloning and Sanger sequencing, and this revealed that all eight individuals carry the same heterozygous ~ 126 kb deletion encompassing the *TPRX1*, *CRX*, and *SULT2A1* locus.

It seems unlikely that loss of *SULT2A1* and *TPRX1* contributes to the phenotype observed in this patient series. *TPRX1* (Tetrapeptide repeat homeobox 1; OMIM 611166) is a member of the paired-like homeobox gene family of transcription factors, which encodes a homeodomain protein like CRX and a DNA-binding domain. It is hypothesised to be involved in the regulation of preimplantation embryo development and embryo genome activation since it is expressed in early human embryos (Madissoon et al., 2016). Variants in this gene are not associated with any condition or trait in any species; it is expressed at very low levels in all tissues except retina and testis (Protein Atlas, version 19.3, https://www.proteinatlas.org/ENSG00000178928-TPRX1/tissue). The *SULT2A1* gene (Sulfotransferase family 2A, dehydroepiandrosterone-preferring member 1; OMIM 125263) encodes dehydroepiandrosterone sulfotransferase (DHEAST), which is an enzyme that catalyses a wide variety of steroids in human liver and adrenal tissue and is responsible for sulfation of bile acids in the human liver (Comer et al., 1993). It is highly expressed in human liver and adrenal tissues, and variants in it contribute to prostate cancer risk (Kwon et al., 2012) and promote metastasis in hepatocellular carcinoma (He et al., 2022).

In contrast, variants in *CRX* are well documented in patients with macular dystrophies. Most of the variants reported are heterozygous missense/nonsense, small deletions, duplications, or insertions. Indeed, patients with heterozygous *CRX* variants have been reported to have the same phenotype identified in this study (Hull et al., 2014, Griffith et

al., 2018, Nasser et al., 2019, Khan et al., 2019a, Ng et al., 2020). We therefore infer that it is almost certainly the complete loss of one copy of the *CRX* gene that is the cause of late onset macular disease in this patient series.

It has been suggested by other researchers that many *CRX* variants might not be true null alleles because premature stop or frameshift variants in or close to the large fourth exon may escape NMD, leading to the production of a truncated protein (Silva et al., 2000, Hull et al., 2014). In those reports, the authors hypothesised that heterozygosity was tolerated, indicating that *CRX* haploinsufficiency is not sufficient to trigger a retinal disease since the heterozygous carriers were unaffected. Additionally, two recent reports have described putative null variants that, in the heterozygous condition, seem to be tolerated and do not appear to have a significant pathological effect (Jones et al., 2017, Ibrahim et al., 2018). In a consanguineous family with a deletion similar to that reported here, described by Ibrahim et al (Ibrahim et al., 2018), two of the four adult heterozygous carriers (aged 38 and 41 years) show subclinical macular abnormalities. This may imply that these individuals will ultimately progress to the phenotype reported here, or could alternatively suggest other variants, either in the remaining copy of *CRX* or in another gene or genes, or environmental factors, may modify the phenotype such that non-penetrance is possible.

The variant reported in this study deletes *CRX* and is an unambiguous *CRX* null allele that will almost certainly result in insufficient CRX protein. The finding of a causal link between this variant and IRD demonstrates that the lack of CRX protein due to haploinsufficiency, or total loss of function of one allele of the gene, is not asymptomatic but rather causes a late-onset macular phenotype, with variation in age at onset (20-78 years) and fundus appearance. Cases 4776 (Figure 4.1A) and 3888 (Figure 4.1B) have a classic bull's-eye maculopathy with an incomplete ring of central macular atrophy. Cases 5588 and 5587 (Figure 4.1C), in contrast, have a much larger ring of atrophy located closer to the temporal arcades. Case GC16591 (Figure 4.1D), like 4476 and 3888, has a bull's-eye maculopathy, but there are also changes nasal to the optic disc and a ring of increased autofluorescence around the posterior pole. For cases 5581 and 5598, no clinical images were available, but diagnoses of CORD and MD and information available from notes are consistent with the range of phenotypes described above.

The same deletion was also found in case 3258 (Figure 4.1E), who was diagnosed with wet age-related macular degeneration (AMD) in the late 60s. The existence of the deletion in this case may indicate that the patient may have an underlying *CRX* phenotype that

progressed to AMD. It could also imply that those who carry this deletion may be more likely to acquire AMD in the future. However, this patient could have other alleles that increase susceptibility to AMD. The finding of apparently Mendelian alleles in AMD patients is not without precedent. A study using whole exome sequencing on a cohort of 218 patients with AMD discovered a pathogenic heterozygous variant in *PRPH2* causal for autosomal central areolar choroidal dystrophy (CACD), as well as enrichment of heterozygous *ABCA4* variants. Additionally, it can be challenging to differentiate between AMD and AMD-mimicking dystrophies based only on fundus examination. These observations imply a blurring of the boundaries between the largely Mendelian inherited retinal dystrophies and the largely polygenic and multifactorial age-related macular dystrophy. It is evident that a number of IRDs can mimic the AMD phenotype, which, for example, has in the past led to debate about whether cases with *ABCA4* variants in AMD cohorts were proof that *ABCA4* variants contributed to AMD, or rather they were misdiagnosed cases of late onset Stargardts disease (Stone et al., 1998). Therefore, genetic screening for AMD patients for variants in a number of IRD genes may be useful to exclude AMD-mimicking dystrophies (Kersten et al., 2018), and these data presented in this chapter imply that *CRX* should be included in such screening.

It was difficult to identify the exact positions of breakpoints in the available NGS data due to the presence of the repetitive elements on the flanking sequences surrounding the deletion. Structural variations account for more variations between individuals than SNVs and are the main cause underlying chromosomal rearrangements (Conrad et al., 2010a, Sudmant et al., 2015, Sanchis-Juan et al., 2018). Mechanisms giving rise to chromosomal rearrangements such as the one described here include recombination-based processes such as nonallelic homologous recombination (NAHR) (Lupski, 1998) and replication-based processes such as microhomology mediated break-induced replication (MMBIR) or fork stalling and template switching (FoSTes) (Lee et al., 2007, Zhang et al., 2009). The repetitive nature of the region around the breakpoint potentially facilitates events leading to structural variations and makes this region more likely to be a mutational hotspot. The presence of the block of 5.3 kb SNARs, which contain the Alu repeats immediately adjacent to the deletion breakpoints, provides a substrate for homologous recombination events which could lead to either duplication or deletion secondary to an unequal crossover of genomic sequences (Lupski, 1998). In light of that, large stretches of homology (294 bp) were identified in the sequences surrounding the breakpoint junctions, indicating that NAHR is the most likely mechanism by which this deletion arose.

Microsatellite genotyping revealed that the deletion was segregating on multiple haplotype backgrounds, but analysis of SNPs immediately adjacent to the breakpoint suggests it has arisen only once on a single founder allele. The contrast between the single haplotype detected by single-nucleotide polymorphisms within 500 bp of the breakpoints and the multiple haplotypes detected by markers 300 kb proximal (D19S606) and 1 Mb distal (D19S879) suggests that the deletion arose only once but has existed in the population for many generations. This is plausible because this adult-onset condition would not be subject to significant negative selection. The late age at onset may also explain the apparent lack of a family history in most of these individuals because in most cases parents were deceased.

Previous versions of next-generation sequencing did not look for deletions, and when coverage is low, more recent screens may fail to detect heterozygous deletions. The identification of this variant in seven apparently unrelated families with inherited retinal disease, the theory that it may have first appeared many generations ago, and the existence of reports indicating that such an allele might be asymptomatic all together imply that this may turn out to be a relatively common cause of MD that has been overlooked in current screening. The PCR test described here will facilitate quick, easy, and inexpensive screening of large patient cohorts and may lead to the identification of further cases.

Screening population controls did not reveal any further deletion carriers. In 13,096 clinical genetics patients with various inherited nonretinal conditions screened by array comparative genomic hybridisation (CGH) (Leeds Teaching Hospitals NHS Trust, Leeds, UK), only 1 case was identified with a deletion encompassing *CRX*. This deletion spanned 3.4 Mb, and the individual also carried an inversion of chromosome 19. At birth this individual was found to have exomphalos, urogenital sinus, congenital heart defect, annular pancreas, ductal aneurysm, and bowel atresia. The UCSC genome browser seems to show 14 copy number variants deleting the *CRX* gene. Of these, 2 encompassing a similar region, nsv1066782 and nssv3573812, were identified in a study of 29,084 patients with developmental disorders (Coe et al., 2014), but precise breakpoints are unclear. The remaining 12 were identified by array CGH in a comparative screen of 30 normal humans and 30 chimpanzees (Perry et al., 2008), but the same study noted that copy number variant boundaries were likely to be overestimated with the platform used, and these findings remain unverified.

In a published response to a paper on these results, Mustafi and Chao noted that the hypothesis of haploinsufficiency being the most likely explanation for the late-onset

macular phenotype, is contrary to earlier reports implying that heterozygous loss of function *CRX* alleles were asymptomatic carriers (Mustafi and Chao, 2022). They reviewed a number of published cases where hypomorphic alleles in other IRD genes were found to be involved in late-onset phenotypes in diseases such as *RPGR*-associated RP (Cehajic-Kapetanovic et al., 2020a), *RDH12*-associated CORD (Wang et al., 2022) and *ABCA4*-associated macular degenerative disorders (Zernant et al., 2017). In light of this, they suggested that the *CRX* results presented might be better explained if the deleted allele was paired with a functional hypomorphic allele (biallelic variation) which might be in non-coding regions, including deep-intronic variants, variants in regulatory regions (e.g., promotors, enhancers, 5' untranslated region (UTR), 3' UTR, and topologically associated domains (TADs)) that had been missed during analyses, making this condition inherited as a recessive disorder.

The group responded (Inglehearn et al., 2022) by noting that the terms hypomorphic, modifier or variable penetrance allele are often used more or less interchangeably to describe the behaviour of alleles with a wide range of frequencies and effect sizes. If, by a hypomorphic allele, Mustafi and Chao meant a rare allele with an almost Mendelian effect size, such as the *RPGR* and *RDH12* genes they cited, then this hypothesis can be theoretically possible. However, the proposed mechanism of haploinsufficiency is far more likely. This is because no second potential disease-causing allele was identified in seven cases, including one case for which genome sequence was available (the eighth AMD case was screened only by breakpoint PCR so an additional variant could not be ruled out in this case). To be more certain that these variants had not been excluded through filtering, exomes were re-examined in four cases using a more conservative, non-stringent filtering strategy. No second variant with a MAF below 1% and CADD score of > 10 was observed, and neither was any single variant found in multiple cases as in the *RDH12* allele. Furthermore, no two cases shared the same haplotype in the trans allele. However, this more conservative analysis was not carried out on the cases for which genome sequence was available due to the need to respond quickly to the correspondence piece, so intronic, regulatory or other none-coding variants meeting these lower filtering thresholds cannot be excluded.

On the other hand, if Mustafi and Chao were suggesting the action of a relatively common hypomorphic allele, such as the *ABCA4* (p.Asn1868Ile) variant they also cited, which is found in about 7% of people, then this was entirely possible and could not be excluded. However, such alleles are probably present in many human genes as part of the pool of

variation that makes each human individual unique. The action of such alleles is only possible to detect when studying very large cohorts, and it would be impossible to demonstrate an impact of such an allele in a small patient cohort such as the one available in this study.

Given that only exome or targeted sequencing data were available for all but one of the cases described here, it was not possible to rule out the occurrence of either of these variant types outside the coding region. Expression/splicing quantitative trait loci (eQTLs and sQTLs) are one class of non-coding variants that have been shown to modify the penetrance of coding variants in rare disorders (Castel et al., 2018). Therefore, the GTEx portal (GTEx locus browser; https://gtexportal.org/home/locusBrowserPage/ENSG00000105392.15) was examined for this type of variant at the *CRX* gene locus. However, no significant eQTLs/ sQTLs have been characterised for *CRX* to date.

Mustafi and Chao also recommended using long read sequencing to identify the non-coding variants (Miller et al., 2021a). The variants most commonly missed by short-read sequencing are in fact low complexity regions and complex rearrangements, but the hypomorphic variant examples mentioned by Mustafi and Chao are single nucleotide polymorphisms, which are easily detected by conventional short-read sequencing. Therefore, the difficulty in characterising such variants is not in identifying them but rather in demonstrating their relevance. Moreover, whole genome sequencing can be of value in identifying variants in non-coding regions. As well as the UK 100,000 Genome case (patient GC16591), the sister of patient 4776 was subject to WGS at Edinburgh Genomics and analysed by Dr Claire Smith. This was analysed in the same way as the WES, with a similar software pipeline, which admittedly was not designed to detect non-coding variants. Nevertheless, no candidate second variant was identified, in this or the UK 100,000 Genome case. However, this data was not included in the paper or shown in this thesis because the ophthalmologist involved (but not the patient) withdrew his consent for us to work on this patient. In the response, the group therefore concluded that the strategy employed in this study is valid, and even though Mustafi and Chao's theory was considered as a potential alternative explanation, haploinsufficiency remains the most likely cause of the disease in these patients.

In conclusion, this project has identified a heterozygous deletion of *CRX* as the cause of retinal disease in six apparently unrelated families with MD and an individual with wet AMD, confirming that haploinsufficiency for *CRX* is pathogenic. The deletion seems

likely to be a relatively old founder mutation caused by non-allelic homologous recombination due to Alu sequences at each breakpoint. A PCR assay that will facilitate the rapid screening of large patient cohorts was also described. Publishing these findings will help others also examine further cohorts of central retinal dystrophy cases to determine whether this mutation has been overlooked in previous screens and may therefore prove to be a relatively common cause of late-onset retinal disease. Further examination of WGS to determine whether a second predisposing allele exists in these patients would also be beneficial, but there are no plans to do that at this stage.

Chapter 5

Screening *RPGR-ORF15* on long read sequencing

5.1 Introduction

DNA sequencing approaches have advanced significantly since the landmark report of Sanger sequencing (Sanger et al., 1977b). Next (Second) generation sequencing has become the dominant technology for determining a molecular diagnosis in patients with rare genetic diseases (Lu et al., 2016, Lipworth et al., 2020) due to its capability to deliver large volumes of highly accurate sequence data at relatively low cost. However, short-read DNA sequencing has several reported limitations (Van Dijk et al., 2018). Generic enrichment PCR conditions can lead to non-uniform or absent coverage; *de-novo* assembly and haplotype phasing is rarely possible; structural variants can prove difficult to detect; and the alignment of repetitive regions in short-read sequence data remains challenging. It is likely that these issues underlie many of the approximately half of cases with a suspected Mendelian disease that remain undiagnosed following whole-exome sequencing (Frésard and Montgomery, 2018, Wenger et al., 2019, Miller et al., 2021a)

More recently, third generation single molecule sequencing (TGS) platforms such as the Sequel and Revio instruments (Pacific Biosciences), in addition to the nanopore range of devices (Oxford Nanopore Technologies; ONT) were born out of the need for a technology that could sequence larger reads (>10kb) at a faster rate (Lu et al., 2016, Wenger et al., 2019). This is facilitating the investigation of so-called dark and camouflaged genomic loci, which have remained refractory to short-read analyses either due to informatic difficulties (*e.g.,* an inability to determine an unambiguous mapping position) or wet-laboratory processes that relate to their underlying genomic architecture (*e.g.,* the high-GC content of some first exons). These studies are increasing our understanding of the frequency and complexity of structural variants and enabling improved analysis of challenging genomic regions (Kieleczawa, 2006, Huddleston and Eichler, 2016, Ebbert et al., 2019, Watson et al., 2021).

Most X-linked RP (XLRP) cases are caused by mutations in the Retinitis Pigmentosa GTPase Regulator (*RPGR*; OMIM 312610) gene (Section 1.6.1.3), which accounts for more than 70% of XLRP-affected families (Vervoort et al., 2000, Breuer et al., 2002). Mutations in another two genes, Retinitis Pigmentosa 2 (*RP2*) and Oral-Facial-Digital 1 (*OFD1*), were also found to cause XLRP (Webb et al., 2012, Lyraki et al., 2018), and three additional loci, *RP6*, *RP24*, and *RP34*, have been proposed to be associated with

XLRP (https://web.sph.uth.edu/RetNet/sum-dis.htm#A-genes) (Accessed on January 2023). RPGR localizes within the connecting cilium (CC) of photoreceptors and is thought to maintain photoreceptors by regulating cilia function. It has been suggested that RPGR protein plays a role in sorting and trafficking rhodopsin to the outer segment (Pawlyk et al., 2016, Lyraki et al., 2016, Megaw et al., 2017).

There are several isoforms of the *RPGR* gene, but *RPGR-ORF15* (isoform C; NM 001034853.2) is predominantly expressed in the retina and frequently mutated in RP. Over 60% of all *RPGR* mutations are clustered in the unique *ORF15* exon at the 3' terminus of the predominant retinal transcript, which is purine-rich and highly repetitive and encodes a 567-amino acid protein sequence rich in glutamic acid and glycine residues (Chiang et al., 2018). It contains a ~900 bp core region that has been identified as a hotspot for disease-causing variations (Vervoort et al., 2000), the most prevalent of which are frameshift mutations due to small out of frame deletions/duplications (Neidhardt et al., 2008, Li et al., 2016a, Tuupanen et al., 2022). The occurrence of these mutations is thought to be due to the slippage of DNA polymerase in the highly repetitive, purine rich region (Ball et al., 2005). Furthermore, it has been suggested that super helical tension caused by the repeats in this region of ORF15 leads to the formation of hairpins and other complex structures that cause instability and polymerase arrest (Holder et al., 2015, De Bustos et al., 2016).

The susceptibility of the *ORF15* repetitive sequence to instability and polymerase slippage or arrest are thought to underlie the difficulty researchers experience in sequencing the *ORF15* region of *RPGR* by Sanger sequencing. Attempts to overcome this issue have used a range of different mutation detection approaches, including direct Sanger sequencing (Breuer et al., 2002), cloning the PCR product and then Sanger sequencing (Zhang et al., 2002, Ebenezer et al., 2005), and direct sequencing of the repetitive part of *RPGR-ORF15* with nested sequencing primers (Bader et al., 2003). However, these approaches remain technically challenging.

It is also well documented that standard short-read NGS captures the outer edges of the *ORF15* exon but is unable to read through the repetitive ~900 bp central region (Wang et al., 2014, Huang et al., 2015, Li et al., 2016a, Chiang et al., 2018). An NGS-based approach has been developed using a *de novo* assembly pipeline, which reportedly overcomes the limitations of the traditional pipeline but requires considerable optimisation to reduce the number of false positive calls (Maggi et al., 2020). WES requires capture of exons by probe hybridisation, and GC-rich probes may form

thermostable hairpin loops or other secondary structures which would fail to denature completely and therefore capture targets less effectively than single copy DNA. PCR amplification of such sequences, a necessary step during library preparation, may also create bias. Using WGS, the library preparation for which is PCR-free (though cluster amplification on the flowcell still requires a bridging PCR step), gives improved coverage in most GC-rich regions (Pipis et al., 2019) and has been found to perform better than WES (Nash et al., 2022),. However, this is not the case for *ORF15* (Figure 5.1). This may be due to the repetitive nature of the region, together with the presence of common polymorphic indels, making it difficult to align the sequence.

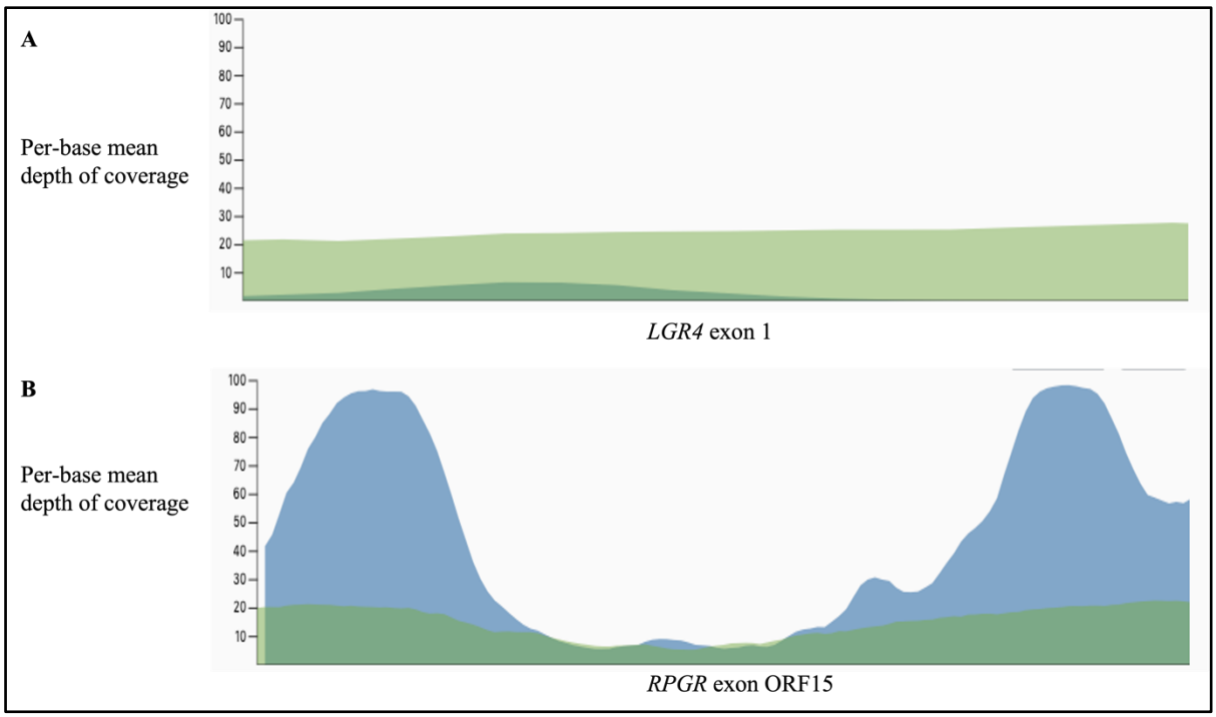


Figure 5.1: Read depth plots generated using data from gnomAD, comparing the mean depth of coverage between WGS (green) and WES (blue) for two GC-rich regions. A. WGS gives good depth of coverage for exon 1 of the *LGR4* gene, while WES gives at best poor coverage and at worst no reads at all over much of the region. **B.** Poor depth of coverage is generated over the *ORF15* exon of the *RPGR* gene by both WGS and WES.

A scalable, high-throughput, reliable approach is therefore required to screen this exon. The primary aim of the work described in this chapter was to determine whether either of the two long-read sequencing methods available in the NGS facility (St. James's University Hospital, Leeds, UK) was more effective than conventional Illumina short-read sequencing for mutation screening in *RPGR-ORF15*. As a secondary aim, we also sought to compare two platforms to see which was the more reliable and cost-effective

approach. Nanopore sequencing (Flongle and MinION; Oxford Nanopore Technology, ONT) and PacBio single molecule real time (SMRT) sequencing were assessed as a screening strategy for the identification of pathogenic mutations in *RPGR-ORF15* in PCR-amplified *ORF15* DNA samples from 218 individuals with various retinal diseases.

The 218 DNA samples screened were from affected males, as well as carrier females, from families with clinical diagnoses of IRD, the majority of whom had RP with a probable or possible X-linked inheritance pattern. Carrier females were tested in this investigation because it has been documented that female carriers of X-linked RP exhibit variable disease expression, ranging from moderate (Chivers et al., 2021) to severe ocular symptoms (Nanda et al., 2018). Cases with central retinal dystrophies were also included because it is well documented that variants in *RPGR-ORF15* can also cause cone or cone rod dystrophy, and macular degenerative disorders in addition to RP (Ayyagari et al., 2002, Ebenezer et al., 2005, Wang et al., 2021, Hadalin et al., 2021). Some of these cases had been tested previously using different screening methods such as MIPs, whole exome or clinical exome sequencing but remained unsolved as screening began. Others had no prior testing. A third group were known carriers of *ORF15* pathogenic variants identified in prior screening, and these were included as positive control samples to validate this method. These included two males with c.3334C>T; p.(Glu1112Ter), three females and one male with c.2405_2406delAG; p.(Glu802Glyfs*32), two males with c.2426_2427delAG; p.(Glu809Glyfs*25), and one male with c.2236_2237delGA; p.(Glu746Argfs*23). Finally, 27 samples from unaffected male controls were also examined to assess population variation. All sequence variants are numbered based on transcript NM_001034853.2. A breakdown of the categories of all individuals screened is shown in Table 5.1, and clinical details for all the individuals included in this study can be found in Appendix D.1.

Category	Gender	
	Male	Female
Unsolved peripheral retinopathy	96	40
Unsolved central retinopathy	37	36
Known variant carriers	6	3
Controls	27	

Table 5.1: A breakdown of the categories of all individuals screened for *ORF15* pathogenic variants. The four categories of 218 cases and 27 controls screened on either PacBio or nanopore sequencing or both platforms, for *ORF15* pathogenic variants are shown. The number of the samples in each category is displayed.

For PacBio sequencing, one-step high fidelity PCR using barcoded primers was performed to produce *ORF15* amplimers (Section 2.4.3.1). The library preparation was performed at the University of Leeds in-house sequencing facility based at St. James's University Hospital, following the steps detailed in Section 2.14.1.1, and the libraries were then sequenced on a PacBio Sequel platform.

For nanopore sequencing, two PCR reactions were performed to generate *ORF15* amplification products. A first pre-indexing PCR was performed as detailed in Section 2.4.3.2, and a second indexing PCR was carried out to incorporate the unique indexes on a per-sample basis (Section 2.4.3.3). Based on the flow cell used, ligation-based library preparation was carried out by the author, and the samples were sequenced on either MinION (2.14.2.1) or Flongle (Section 2.14.2.4) flow cells.

5.2 Results

5.2.1 Screening *RPGR-ORF15* by SMRT PacBio sequencing

In this section, SMRT PacBio sequencing was used to screen 216 of the cases described in Table 5.1, of which 207 unsolved cases, nine known variant carriers, in addition to all 27 of the controls. These are broken down to their various categories in Table 5.2.

Category	**Gender**	
	Male	Female
Unsolved peripheral retinopathy	94	40
Unsolved central retinopathy	37	36
Known variant carriers	6	3
Controls	27	

Table 5.2: A breakdown of the categories of individuals screened on PacBio sequencing for *ORF15* pathogenic variants. The four categories of 216 cases and 27 controls are shown and the number of the samples in each category is displayed.

5.2.1.1 PCR assay and barcoding

To sequence the entire *ORF15* region, a specific primer pair (MA_ORF15_F2 and MA_ORF15_R) (Appendix D.2) were designed by Dr James Poulter (Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK) in the adjacent DNA sequence using primer 3. These primers amplified under the PCR conditions described in Appendix D.2.1, using the Phusion high fidelity DNA polymerase kit, but proved unreliable when tested on control DNA samples. Several technical variations were tried to optimise the PCR conditions, including trying different annealing temperatures (e.g., touchdown 70→65, 67→60, 67→63, 67ºC and 65ºC), increasing thermal cycle repeats, and using additives such as DMSO (3%) and $MgCl_2$. PCR conditions were optimised for these primers to generate a ~1.8 kb amplicon spanning the entire 1.7 kb of *ORF15* (Figure 5.2A).

One run on a PacBio sequencer, performed by the University of Leeds in-house sequencing facility, cost £1160, but was expected to give a read depth far in excess of that required for a single sample. Unlike the nanopore sequencing technology, no smaller device was available giving fewer reads at reduced cost. It was therefore necessary to design a new set of primers utilising the above primer pair with a number of barcodes attached to allow demultiplexing of pooled samples, in order to make PacBio analysis more cost-effective. Amplification using these barcoded primers was performed in a two-stage PCR. First, *ORF15* was amplified using the non-barcoded primer pair (MA_ORF15_F2 and MA_ORF15_R) as described above, then the PCR product was re-amplified for a further 35 cycles under the same conditions (Appendix D.2.1) using different combinations of barcoded primers (Appendix D.2.2). However, this led to the amplification of multiple bands of unknown origin (Figure 5.2B).

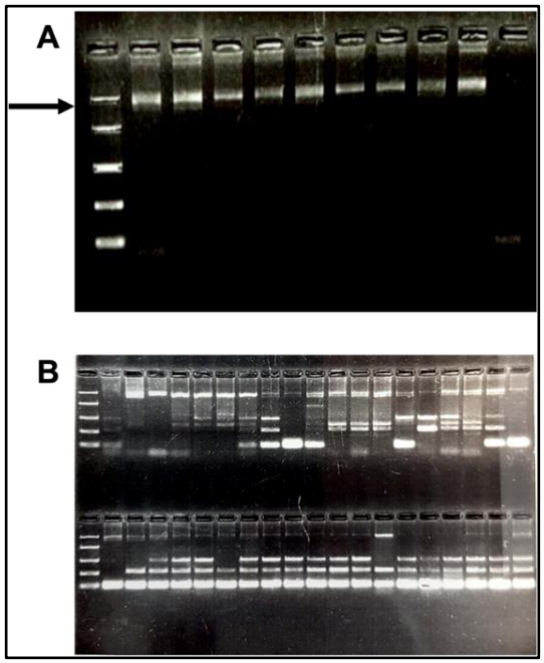


Figure 5.2: PCR analysis of samples screened for *RPGR-ORF15* pathogenic variants. A. Agarose gel electrophoresis (1.5%) of the PCR amplification product for *ORF15* (~1.8 kb; marked by arrow) of primers MA_ORF15_F2 and MA_ORF15_R. First lane shows Bioline EasyLadder1 size marker (100, 250, 500, 1000, and 2000) bp. Each well displays *ORF15* PCR products from different DNA samples except for the last well which contains a negative control (No DNA). **B.** Agarose gel electrophoresis (1.5%) of the PCR products for *ORF15* using different barcoded primer combinations of primer pair MA_ORF15_F2 and MA_ORF15_R. The product size is ~1.8 kb. The first lane shows Bioline EasyLadder1 size marker (100, 250, 500, 1000, and

2000) bp. Each well is loaded with PCR products from a different DNA sample amplified with a unique primer combination. Multiple bands of incorrect size and unknown origin can be seen.

To eliminate these bands, a troubleshooting process was applied including, carrying out the PCR using the barcoded primer directly on the genomic DNA; diluting the PCR product of the first PCR at 1:10 and at 1:20, then doing the second PCR with barcoded primers on the diluted products; ExoSAP-IT was used on the first PCR product then re-PCR with the barcoded primer; finally, an AMPure beads clean-up of the second PCR product with the barcoded primers was done. However, none of these attempts eliminated these additional bands of incorrect sizes.

The set of barcoded primers had been designed with additional 5 bp sequences on them known as padding sequences, added to address concern about unequal ligation efficiencies. It was thought that these padding sequences may be causing these multiple bands because they may have the potential to allow the ends to hybridise and form a circular DNA molecule, or to link up and concatemerize into double or treble circles. These circles would then run on the gel as open or closed circles rather than linear DNA molecules, giving the different band sizes seen.

Since the first primer pair (MA_ORF15_F2 and MA_ORF15_R) had worked only intermittently, another set of primer pairs without barcodes were designed and ordered (MA_ORF15_F3, MA_ORF15_R3 and MA_ORF15_F4, MA_ORF15_R4, see Appendix D.3). Different primer combinations (F3/R3, F3/R4, F4/R4, and F4/R3) were optimised by trying to use different buffers (GC or HF buffers supplied by the Phusion high fidelity polymerase kit manufacturers), adding DMSO at 3%, and applying different thermal conditions (touchdown 63°C→61°C, 65°C→63°C, and 67°C→65°C) and cycles (25 and 30 cycles) (Appendix D.3.1). The most reliable primer pair proved to be MA_ORF15_F3 and MA_ORF15_R4. Further optimisation steps were applied until the most accurate and reliable conditions were achieved (described in Section 2.4.3.1). The validity of these conditions was assessed on several samples using the non-barcoded primer pair MA_ORF15_F3 and MA_ORF15_R4 to confirm that this produces a 2.272 kb amplicon encompassing the whole ORF15 region in all samples.

A total of 48 PacBio barcodes (16 bp each) were added to the primer set MA_ORF15_F3 and MA_ORF15_R4 and used to generate different combinations of barcoded samples. This time the 5 bp padding sequences were not included. This experiment was then

repeated in the same way as described previously, by attempting to re-PCR from the previous PCR product in order to add barcodes on, using the same PCR conditions (Section 2.4.3.1). Again, multiple bands of incorrect size were observed. It was then decided to try PCR amplifying with the barcoded primers MA_ORF15_F3 and MA_ORF15_R4 directly from genomic DNA to generate amplicons of barcoded samples. The optimum option was to use the barcoded primer pair MA_ORF15_F3 and MA_ORF15_R4 applying the conditions in Section 2.4.3.1. Different combinations of barcoded primers were tested by PCR amplifying directly from one control genomic DNA to ensure this approach gave the correct band size and to decide which combinations work best. The successful combinations of barcoded primers were then used to amplify the *ORF15* region in all samples (Figure 5.3). The sequences of barcoded primers are listed in Appendix D.3.2.

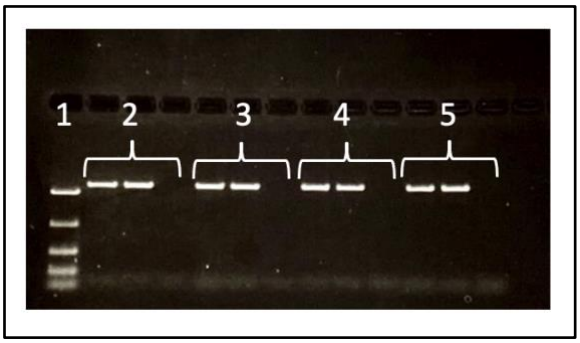


Figure 5.3: PCR analysis of samples screened for *RPGR-ORF15* pathogenic variants. Agarose gel electrophoresis (1.5%) of the PCR products for *ORF15* using four different barcoded primer combinations of primer pair MA_ORF15_F3 and MA_ORF15_R4. The product size is 2.272 kb. Lane 1 = Bioline EasyLadder1 size marker (100, 250, 500, 1000, and 2000) bp. Brackets 2, 3, 4, and 5 represent four different primer pair combinations. For each primer pair combination, the first well shows the PCR product from the patient's DNA sample, the second well shows the PCR product from a positive control and the last well shows a no DNA control.

5.2.1.2 Testing of 46 cases and 27 controls

To test the effectiveness of this approach, 46 DNA samples from patients with peripheral retinopathy and central retinopathy, selected from the larger set described in Section 5.2.1, were screened for disease causative variants in *RPGR-ORF15*. Nine of these were the set of controls with previously known *ORF15* pathogenic variants, included as positive controls. The remaining 37 were unsolved at the time of screening. Finally, the 27 unaffected individuals were included as controls to show the range of polymorphic variation in *ORF15* in the population.

A 2.272 kb amplicon was generated from each of these samples using a unique combination of barcode attached to primer pair MA_ORF15_F3 and MA_ORF15_R4, as described in Section 2.4.3.1, then samples were pooled and sent to the University of Leeds in-house sequencing facility based at St. James's University Hospital. The library was prepared by Mrs. Morag Raynor in the sequencing facility, as detailed in Section 2.14.1.1, and the samples were processed into SMRTbell complexes which were then loaded into the PacBio Sequel platform. The Sequel platform was run for 10 hours using the SMRT Link software (https://downloads.pacbcloud.com/public/software/installers/smrtlink_10.2.1.143962.zip). The output files were analysed by Dr Ian Carr in the Leeds Institute of Medical Research, University of Leeds, as outlined in Section 2.14.1.2. Text files containing 100 alignments randomly selected from a larger multiple sequence alignment, composed of sequences with a quality score of greater than Q20 for 99.5% of positions, were visually inspected. This analysis confirmed that the PacBio sequencing platform is capable of reading across the *ORF15* region.

The sequence output confirmed the existence of all previously identified variants in the nine samples with RP included as positive controls. In addition, two previously reported pathogenic variants were identified in two new cases with RP, c.2323_2324delAG; p.(Arg775Glufs*59) (Carss et al., 2017, Beigi et al., 2021) (ClinVar accession number: VCV000438144.18) in case 3219 (Figure 5.4) and c.2041_2042delAA; p.(Lys681Glyfs*2) (ClinVar accession number: VCV000865836.2) in case 3685.The presence of these variants was confirmed on nanopore long read sequencing using MinION flow cell. 27 unaffected male controls were also tested to determine the degree of the normal variations in *RPGR-ORF15* in the normal population. No pathogenic mutations were identified in these control samples or in the eleven obligate negative cases solved for variants in other IRD genes during this study; however, benign single nucleotide variants and in-frame deletions and duplications were detected. Twelve of these control samples with benign variants were also rerun on a MinION to confirm the variants observed.

Full details of all cases and controls screened and variants found are given in Appendix D.1.

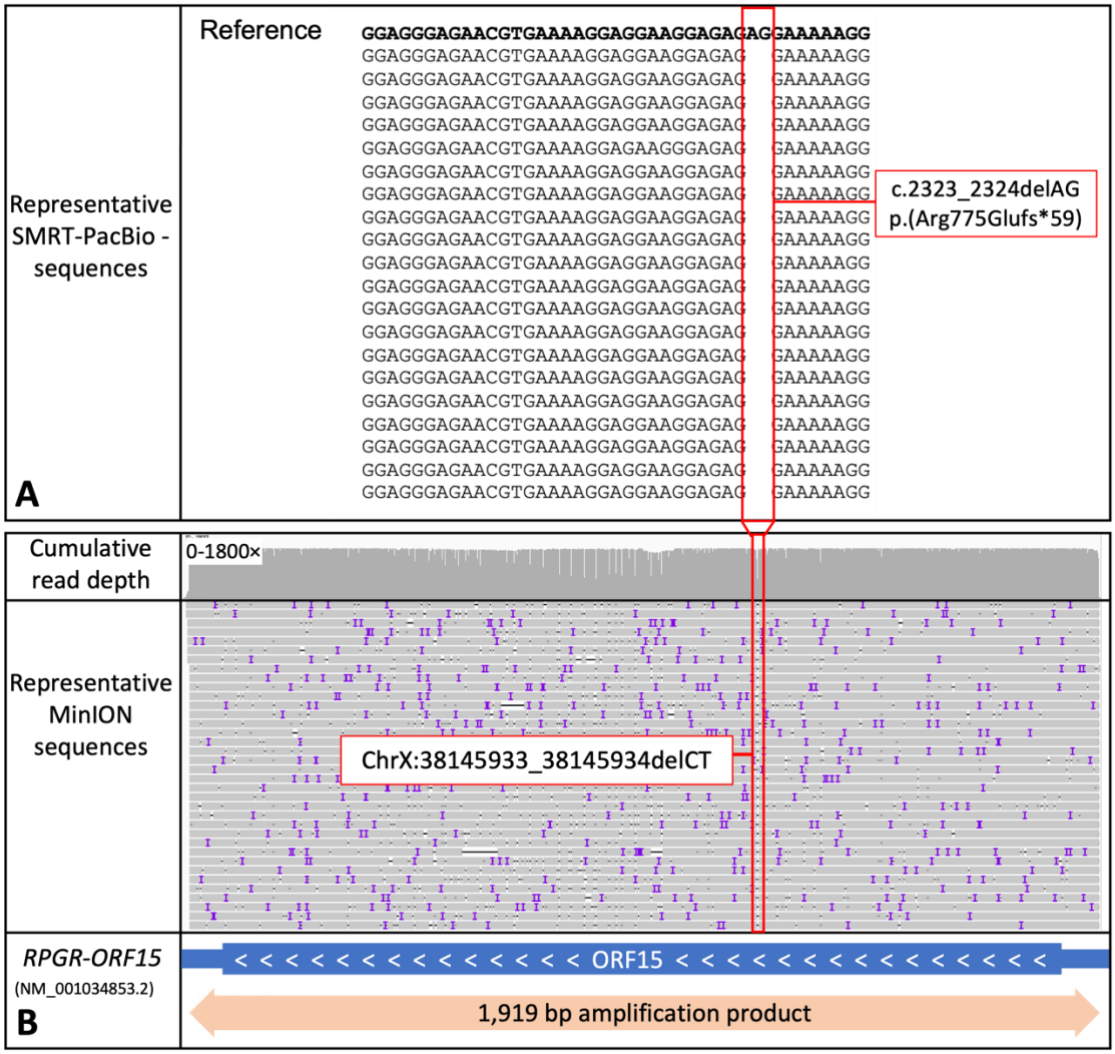


Figure 5.4: Sequencing *RPGR-ORF15* on two different long-read sequencing platforms. A. Long-read sequencing alignment at the *RPGR-ORF15* locus produced from a 2.272 kb amplimer using PacBio Sequel instrument, showing the first 20 reads out of 100 alignments randomly selected from a larger multiple sequence alignment, composed of sequences with a quality score of greater than Q20 for 99.5% of positions. The sequence shows a hemizygous two base-pair deletion (c.2323_2324delAG (NM_001034853.2), p.(Arg775Glufs*59)) in a male patient with RP (3219, identified in this study). **B.** Long-read sequencing alignment at the *RPGR-ORF15* locus generated from a 1.919 kb amplimer using a nanopore MinION sequencer, confirming the absence of the two base-pair sequence at the same location (ChrX:38145933_38145934delCT (hg19)) in patient 3219.

5.2.1.3 Scale up and screening of an additional 170 cases

Using this same approach, an additional 170 cases were examined to screen for putative disease-causing variants in *RPGR-ORF15*, bringing the total number screened by PacBio to 216, as detailed in Table 5.2. These included 64 males and 38 females with peripheral retinopathy and 32 males and 36 females with central retinopathy.

Among 102 DNA samples from patients with peripheral retinopathy who were negative in previous testing or had not previously been tested, three potential frameshift pathogenic variants were identified in one male (3558) and three females (3606, 4686, and 5228); c.2488delG, p.(Glu830Argfs*259);; c.2608_2609insG, p.(Glu870Glyfs*209), and c.2764delG, p.(Glu922Argfs*167). These variants were all located in the centre of the highly repetitive region of *ORF15,* are predicted to be likely pathogenic by the Franklin (Section 2.13.4.6) database and are not present in the databases dbSNP, gnomAD, and ClinVar. However, it is good practise in genetic screening, even in a research setting, to verify potential pathogenic variants by a second means, to exclude false positives and reduce the likelihood of sample mix up. Screening these samples by nanopore sequencing (Flongle) is therefore needed to confirm the authenticity of the detected variants and to evaluate this method's sensitivity and specificity. In the remaining 68 patients with central dystrophies, two males (5225 and 5251) were found to carry the variant c.2608_2609insG, p.(Glu870Glyfs*209). Two further frameshift variants, c.2971delG, p.(Glu991Lysfs*98) (in case 5867) and c.3317delA, p.(Lys1106Serfs*25) (in cases 4659 and 5242), were detected in patients with peripheral retinopathy. However, the read percentage of the mutant allele was much lower than the predicted 50% (in females) and 100% (in males), therefore they were considered to be less likely to be pathogenic and less likely to be disease-causing variants (Table 5.3). This also needs to be confirmed by nanopore sequencing.

PATIENT ID	GENDER	CLINICAL DIAGNOSIS	NUCLEOTIDE CHANGE (NM_001034853.2)	PROTEIN CHANGE	CATEGORY	VALIDATED ON MINION?
135	M	X-linked RP	c.2426_2427delAG	p.(Glu809Glyfs*25)	Known control	Yes
749	F	X-linked RP	c.2405_2406delAG	p.(Glu802Glyfs*32)	Known control	Yes
752	F	X-linked RP	c.2405_2406delAG	p.(Glu802Glyfs*32)	Known control	Yes
753	M	X-linked RP	c.2405_2406delAG	p.(Glu802Glyfs*32)	Known control	Yes
1475	M	RP	c.2236_2237delGA	p.(Glu746Argfs*23)	Known control	No
1487	F	X-linked RP	c.2405_2406delAG	p.(Glu802Glyfs*32)	Known control	Yes
3219	M	X-linked RP	c.2323_2324delAG	p.(Arg775Glufs*59)	Unsolved case	Yes
3558	M	AR-RP	c.2488delG	p.(Glu830Argfs*259)	Unsolved case	No
3606	F	AR-RP	c.2608_2609insG	p.(Glu870Glyfs*209)	Unsolved case	No
3685	M	X-linked RP	c.2041_2042delAA	p. (Lys681Glyfs*2)	Unsolved case	Yes
4659	F	AD-RP	c.3317delA	p.(Lys1106Serfs*25)	Unsolved case	No
4686	F	Multiplex RP	c.2608_2609insG	p.(Glu870Glyfs*209)	Unsolved case	No
5156	M	X-linked MD	c.3334 C>T	p.(Gln1112*)	Known control	Yes
5225	M	AR bull’s eye maculopathy	c.2608_2609insG	p.(Glu870Glyfs*209)	Unsolved case	No
5228	F	Multiplex RP	c.2764delG	p.(Glu922Argfs*167)	Unsolved case	No
5242	F	AR? Severe RP	c.3317delA	p.(Lys1106Serfs*25)	Unsolved case	No
5251	M	Cone dystrophy	c.2608_2609insG	p.(Glu870Glyfs*209)	Unsolved case	No
5257	M	X-linked RP	c.3334 C>T	p.(Gln1112*)	Known control	Yes
5586	M	RP	c.2426_2427delAG	p.(Glu809Glyfs*25)	Known control	Yes

5867	M	Unusual peripheral retinal degeneration, LD, obesity	c.2971delG	p.(Glu991Lysfs*98)	Unsolved case	No

Table 5.3: Summary of pathogenic variants identified in *RPGR-ORF15* using PacBio sequencer. This table details all the pathogenic variants identified across all four runs on PacBio sequencer. M = male, F = female, RP= Retinitis pigmentosa, MD = Macular dystrophy, AR= Autosomal recessive, AD= Autosomal dominant.

5.2.1.4 SNVs and in frame deletions and duplications

In both patients and controls, a large number of benign single nucleotide variants (SNVs) and in-frame deletion/duplication variants (indels) have been found. The number of indels varies from zero to nine in each individual, and some of these indels were also found in healthy controls. These in-frame variants were classified as benign, likely benign, or variants of uncertain significance (VUS). In this cohort, the frequency of in-frame deletions and duplications was 48% (104/216) in the patient group whereas in the control group it was 33% (9/27). In this small study, there is therefore no significant difference in the frequency of in-frame indels between patient and control cohorts. However, in a previous study the number of in-frame indels was approximately two times higher in patients than in controls (Li et al., 2016a). Some of the identified benign SNVs and in-frame deletions and duplications noted in this study are the same benign variants reported previously (Shu et al., 2008, Li et al., 2016a). They were also observed in both cases and controls on MinION sequencing. The identified SNPs and indels in both cases and controls are listed in Appendix D.4.

5.2.2 Screening RPGR-ORF15 by Nanopore sequencing

In this section, Nanopore sequencing was used to screen 54 samples, of which 35 were unsolved cases, seven known variant carriers, and 12 controls. All but two are drawn from samples screened by PacBio (Table 5.2). These are broken down to their various categories in Table 5.4.

Category	**Gender**	
	Male	Female
Unsolved peripheral retinopathy	30	1
Unsolved central retinopathy	4	
Known variant carriers	5	2
Controls	12	

Table 5.4: A breakdown of the categories of individuals screened on nanopore sequencing, for *ORF15* pathogenic variants. The four categories of 42 cases and 12 controls are shown and the number of the samples in each category is displayed.

5.2.2.1 PCR assay and barcoding

For nanopore long-read sequencing on both Flongle and MinION flow cells, *ORF15* was amplified using the PacBio non-barcoded primers MA_ORF15_F3 and MA_ORF15_R4 that worked previously. To pool the samples, primers were tailed with nanopore universal sequencing tags to facilitate attachment of nanopore indexes for demultiplexing. These primers were amplified using the Phusion high fidelity DNA polymerase kit under the same PCR conditions used in the PacBio method (outlined in Section 2.4.3.1). However, they proved to be unreliable when tested on control DNA samples.

To optimize the PCR conditions, a range of technical changes were made. Various annealing temperatures were tried (at 70°C, touchdown 70°C →67°C, and 70°C →65°C), as well as use of the additive DMSO (3%) with either GC or HF buffer (supplied by the Phusion high fidelity polymerase kit manufacturers). None of these attempts worked, so another set of primers (MinORF15_F1, MinORF15_R1, MinORF15_F2, MinORF15_R2, MinORF15_F3, MinORF15_R3, MinORF15_F4, MinORF15_R4, MinORF15_F5, MinORF15_R5, MinORF15_F6, MinORF15_R6, MinORF15_F7, MinORF15_R7, MinORF15_F8, MinORF15_R8, MinORF15_F9, MinORF15_R9, MinORF15_F10, MinORF15_F11) tailed with the universal sequencing tags were designed and ordered. Different primer combinations were optimised by trying different buffers (GC or HF buffer), and different annealing temperatures (65°C and 63°C) with and without adding DMSO at 3%. All primers sequences are listed in Appendix D.5.

MinORF15_F9 and MinORF15_R5 were found to be the most reliable primer pair, and the most accurate PCR conditions are described in Section 2.4.3.2 (Pre-indexing PCR protocol). Using the primer pair MinORF15_F9 and MinORF15_R5, the validity of these conditions was examined in several samples, and it was determined that all samples generated a 1.919 kb amplicon containing the whole *ORF15* region (Figure 5.5).

To sequence multiple samples in a single flow cell, an indexing PCR using Long Amp *Taq* 2X master mix (Section 2.4.3.3) was performed. The pre-indexed PCR amplification products were cleaned up using AxyPrep paramagnetic beads (Section 2.8) then quantified by Qubit fluorometer (Section 2.3.1). Molarity was calculated using the NEBiocalculator and the unique indexing barcodes (EXP-PBC096; ONT) were added to the samples before pooling them in equimolar quantities for library preparation.

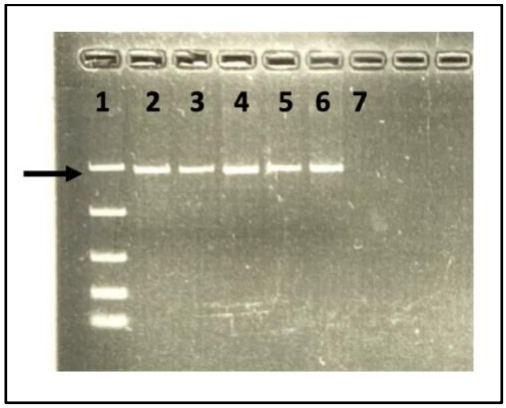


Figure 5.5: PCR analysis of samples screened for *RPGR-ORF15* pathogenic variants. Agarose gel electrophoresis (1.5%) of the pre-indexed PCR amplification products for *ORF15* (~1.9 kb; marked by arrow) of primers MinORF15_F9 and MinORF15_R5. First lane shows Bioline EasyLadder1 size marker (100, 250, 500, 1000, and 2000) bp. Lanes 2, 3, 4, 5, and 6 display *ORF15* PCR products from five different DNA samples. Lane 7 is a negative control (No DNA).

5.2.2.2 Testing the method by screening 5 samples on Flongle

Applying the pre-indexing PCR protocol (Section 2.4.3.2), the genomic DNA from five males (laboratory sample numbers 3217, 3239, 3439, 5231, 5586) with RP was amplified using Phusion high fidelity DNA polymerase and the primer pair MinORF15_F9 and MinORF15_R5 tailed with nanopore universal sequencing tags to generate a 1.919 kb fragment containing *ORF15*.

One male patient with RP (5586) was known to be hemizygous for the *RPGR-ORF15* pathogenic variant c.2426_2427delAG, p.(Glu809Glyfs*25) and was included as a positive control. The remaining four were males with RP who remained unsolved following previous analysis of targeted or exome enriched NGS datasets. As stated above (Section 5.1), conventional short-read NGS approaches perform poorly when they are used to sequence this exon (Wang et al., 2014, Huang et al., 2015, Li et al., 2016a, Chiang et al., 2018). Figure 5.6A shows an example of a short-read alignment at the *ORF15* locus that was produced using a HiSeq3000 (Illumina, Inc.). While there is sufficient read depth and coverage at the extremities of the *ORF15* exon, there are no alignments spanning the central ~900 bp repetitive region.

Ligation-based library preparation was performed on *ORF15* amplification products, and the flow cell was prepared as described in Section 2.14.2.4. 3-20 fmol of each library was loaded on individual Flongle flow cells and the sequencing was run for 24 hours. The

output data was analysed as detailed in Section 2.14.2.5 and the resultant bam files were visualised using the IGV (Section 2.13.5).

Visual examination of the aligned sequence reads confirmed that the nanopore workflow is capable of generating full-length *RPGR-ORF15* sequences. Unique sequences flanking the repeat enabled the long reads to be anchored to the target locus, generating sufficient read coverage across the highly repetitive *ORF15* sequence to enable mutation detection. Identification of the previously reported *ORF15* pathogenic variant c.2426_2427delAG, p.(Glu809Glyfs*25) is demonstrated in Figure 5.6B. No pathogenic or likely pathogenic variants were observed in the remaining four individuals. Run yields obtained were between 9 and 56 Mb, corresponding to read counts of between 6.41K and 34.04K, as detailed in Table 5.5. The c.2426_2427delAG variant was confirmed using Sanger sequencing by the Manchester reference laboratory (Northwest Genomic Laboratory Hub, Manchester centre for genomic medicine, St Mary's Hospital, Manchester, UK) (Section 2.15) (Figure 5.6C).

Sample ID	Run number	Loading concentration	Run yield	Read counts
3217	1	11.50 fmol	55.65 Mb	34.04 K
5231	2	20.00 fmol	29.21 Mb	23.02 K
3239	3	10.45 fmol	17.57 Mb	17.14 K
3439	4	20.00 fmol	8.85 Mb	6.41 K
5586	5	22.27 fmol	36.50 Mb	31.96 K

Table 5.5: Flongle Run metrics for *RPGR-ORF15* amplicon in five samples.

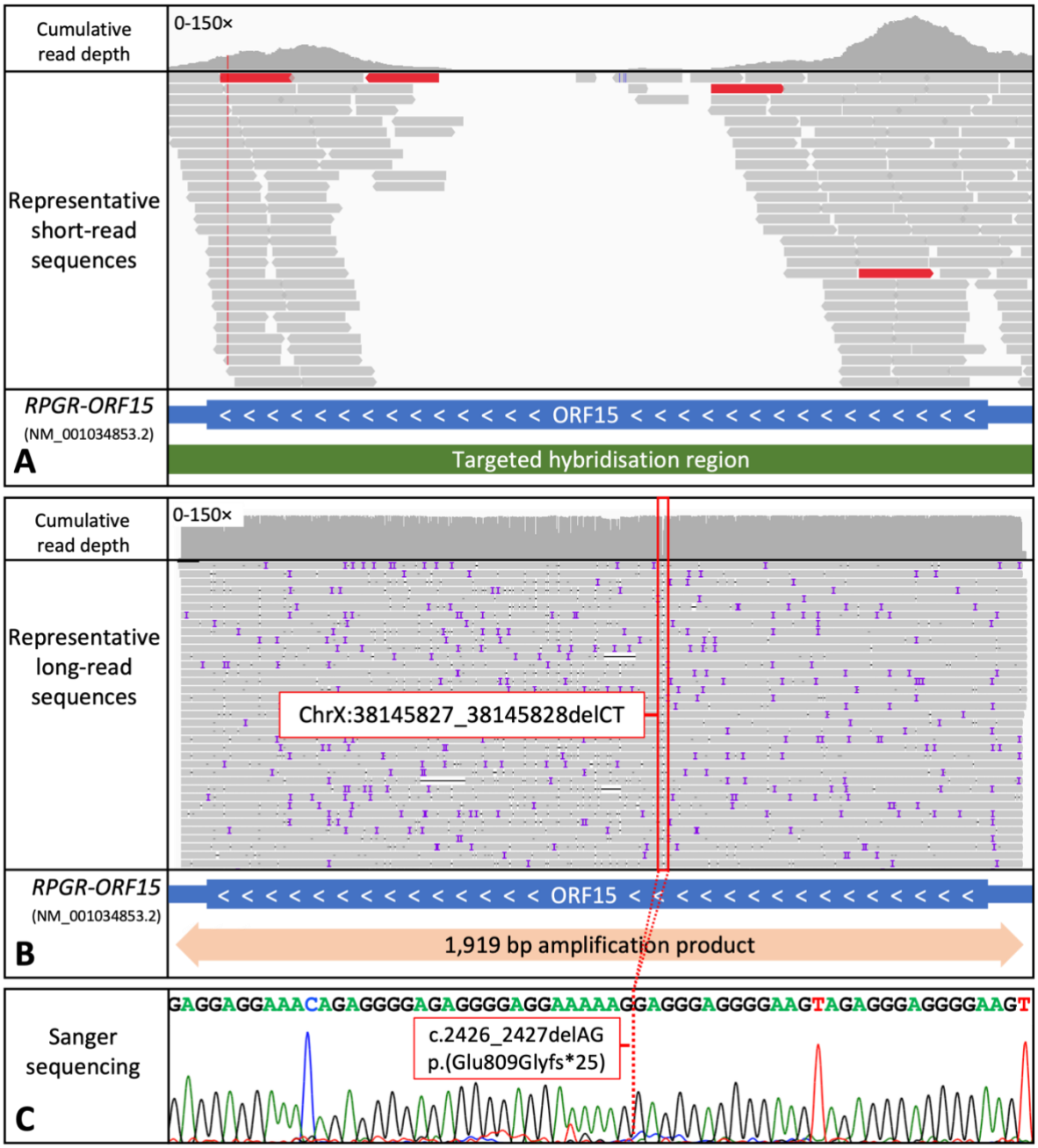


Figure 5.6: Sequencing *RPGR-ORF15*. A. Short-read next generation sequencing at the *RPGR-ORF15* locus. Hybridisation capture enrichment was performed prior to sequencing on an Illumina HiSeq 3000. Aligned sequence reads were viewed using the Integrative Genomic Viewer (IGV). There is an absence of mapped reads across the central region. **B.** Long-read sequencing alignment at the *RPGR-ORF15* locus generated from a 1.919 kb amplimer using a nanopore Flongle sequencer. Read depth is consistent throughout the *ORF15* exon and male RP patient 5586 (a known control) is seen to be hemizygous for the two base-pair deletion (NM_001034853.2) c.2426_2427delAG, p.(Glu809Glyfs*25) (ChrX: 38145827_38145828delCT (hg19). **C.** Sanger sequencing electropherogram generated by the Manchester Reference Laboratory confirms the deletion of a two base-pair sequence at the dashed

vertical line. This analysis was carried out as described in Section 2.15 by NHS diagnostic laboratory staff.

The hemizygous mutation (c.2426 2427delAG, p.(Glu809Glyfs*25)) should account for 100% of reads, as this is a male patient, but appears heterozygous when zoomed in on IGV in Figure 5.6B. However, close examination of the cumulative read count plot reveals that this is actually a function of the basecaller, which designates the deleted bases as different bases in different reads even though the variant is the same in all, rather than the aligner. This gives the appearance of a variable deletion rather than a clean 2 bp deletion (Figure 5.7).

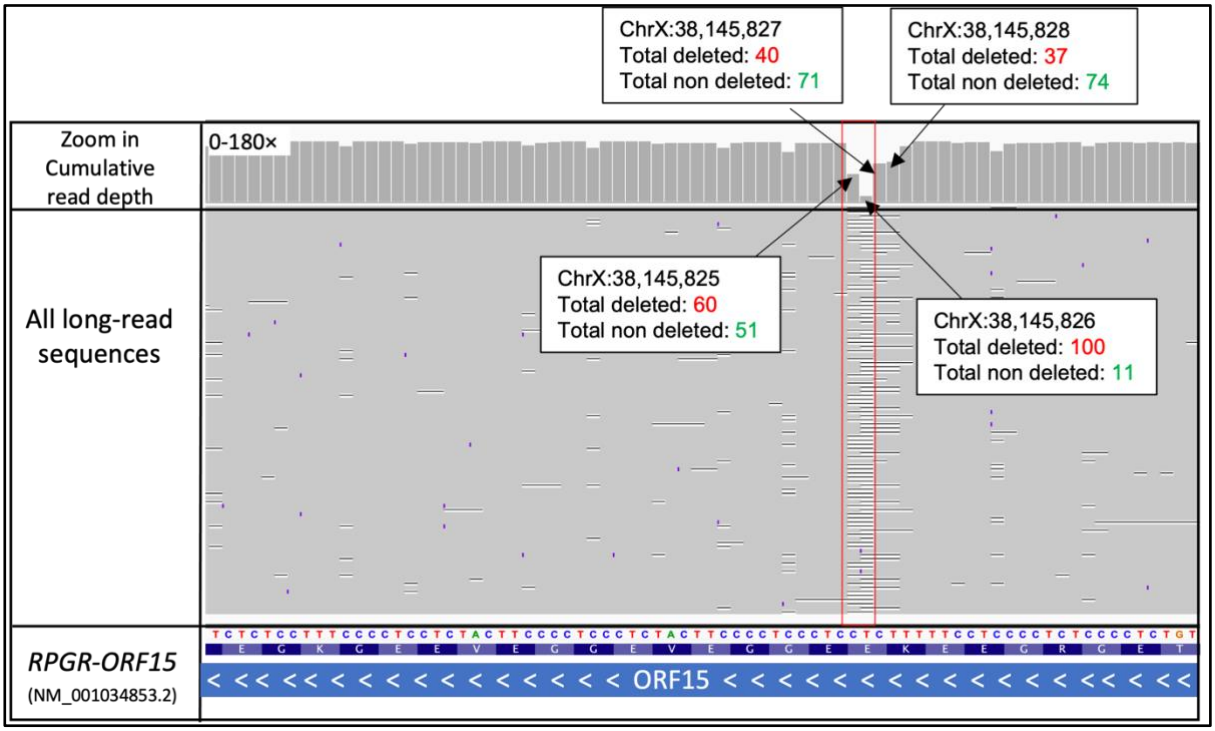


Figure 5.7: Zoom in IGV view in patient 5586 (Figure 5.6B). Zoom in IGV view of the hemizygous 2 bp deletion in patient 5586 (Figure 5.6B), (NM_001034853.2) c.2426_2427delAG, p.(Glu809Glyfs*25) (ChrX: 38145827_38145828delCT (hg19), showing the total deleted and non-deleted reads.

While Flongle sequencing of *ORF15* was successful, yields were between 1 and 5% of the conservatively anticipated 1 Gb sequencer-output. Furthermore, it was evident from cumulative read traces that, by contrast with a normal Flongle sequencing run, throughput from the flow cell after loading the *ORF15* amplimer slowed rapidly, within minutes. Figure 5.8A shows the cumulative read trace of a sequencing run targeting a non-

repetitive region spanning 11.5 kb of the *ALMS1* gene, with no known difficulty for conventional DNA sequence analysis. A proportion of pores remained open and able to generate significant numbers of new reads after 24 hours (Figure 5.8B). In contrast, when the *ORF15* amplimer was sequenced, the cumulative read count plateaued within the first hour (Figure 5.8C) and pore availability declined rapidly within 35 minutes of loading (Figure 5.8D).

Several post-amplification clean-up protocols were investigated to determine whether contaminants were blocking the pores. These included post PCR clean-ups using AMPure XP beads, manual gel extraction of the PCR product and automated separation using the Pippin Prep System (Sage Science, Beverly, MA, USA). These protocols in addition to the re-sequencing were carried out by Dr Christopher Watson, Northeast and Yorkshire Genomic Laboratory Hub, Central Lab, St. James's University Hospital, Leeds, UK. However, no improvement was obtained (data not shown). We therefore hypothesised that the reduced throughput of the *ORF15* amplimer may be a consequence of the formation of secondary structures within the repetitive *ORF15* sequence such that these structures then progressively blocked the flow cell pores until no further reads could be generated.

5.2.2.3 Use of a nuclease wash kit to improve yield

To explore the hypothesis that the formation of secondary structures within the repetitive *ORF15* sequence may be the cause of the reduced throughput, a flow cell wash kit (WSH003, ONT) containing DNase I was used. Wash kits are intended to facilitate reuse of MinION flow cells by digesting, and therefore removing, any residual DNA from the flow cell pores before a different library is loaded. We hypothesised that application of a nuclease wash treatment would clear the pores and allow reloading of a further aliquot of the same *ORF15* library; thereby increasing yield and cumulative read count at the target locus. However, the nuclease wash can only be used on the MinION flow cell, not the lower-throughput Flongle, because opening and resecuring the Flongle flow cell cover is not a procedure supported by ONT. At this point, experiments were therefore switched to use MinION flow cells.

Ligation-based library preparation was performed on *ORF15* amplification products from the same five samples mentioned above as per Section 2.14.2.1 to test the method, and the MinION flow cell was prepared as described in Section 2.14.2.2. 50 fmol of the library was loaded on a MinION flow cell and the sequencing was run for 72 hours. The output data was analysed as described in Section 2.14.2.5. Reads with a length of 1700-2000 bp

and a quality score of 18 were retained, and the resulting bam files were visualised using the IGV (Section 2.13.5).

An expected rapid decline in the cumulative read count (Figure 5.8E) and pore availability (Figure 5.8F), over a period of 2-3 hours was observed. Use of the DNase I wash kit (Section 2.14.2.3) led to an immediate rebound in cumulative read output and pore availability, but this declined within a similar timeframe, requiring a further DNase I nuclease treatment. The resultant output, though still well below the manufacturer's expected yield, was considerably increased as a result of washing and reloading.

To establish the likely limit of the rewashing protocol, multiple aliquots of an *ORF15* amplimer library from twelve indexed samples were sequenced on a single MinION flow cell over a period of three days, washing and reloading nine times. A cumulative read count trace and pore availability plot for this run are shown in Figures 5.8G and H respectively. Pore availability continued to rebound after each wash but declined over the course of the run until little benefit was gained from further reloading. Although long-range PCR target enrichment has been used in other studies (Chiang et al., 2018), combining it with the wash kit and long-read nanopore sequencing is a novel approach.

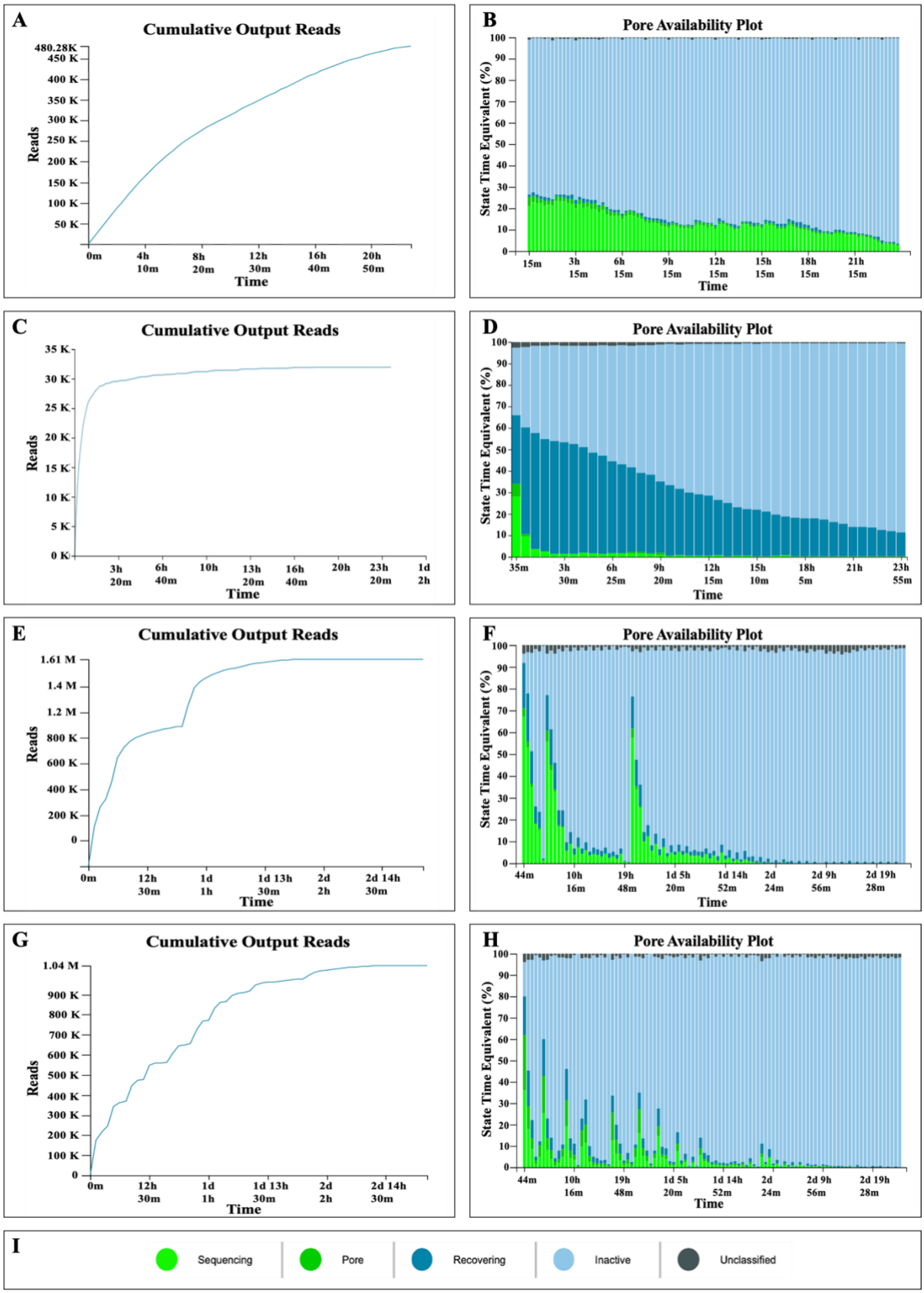


Figure 5.8: Cumulative read counts and pore availability plots for Flongle and MinION long-read sequencing of *RPGR-ORF15*, with and without the use of a flow cell wash kit. A. Cumulative read count plot for Flongle sequencing of an 11.5 kb PCR amplimer containing the *ALMS1* gene. **B.** Pore availability over time for the Flongle run plotted in A. The plot shows that

pores were available over a 24-hour period, with a slow decline over that time. **C.** Cumulative read count plot for Flongle sequencing of a 1.9 kb PCR amplimer containing *RPGR-ORF15*. Pore availability dropped rapidly within the first hour and reads produced declined to almost none within three hours. **D.** Pore availability for the Flongle run shown in C, demonstrating that pores rapidly became "unavailable" over the first hour of sequencing, resulting in a dramatic decrease in data acquisition. **E.** Cumulative read count plot for MinION sequencing of the same *RPGR-ORF15*-containing amplimer in five pooled, tagged samples, with application of flow cell wash buffer after 3 and 24 hours. After each wash, the rate at which reads were acquired recovered to near the original starting rate, then rapidly declined again over the first hour. As a result, throughput was more than doubled from the point of the first wash. **F.** Pore availability for the MinION run plotted in E, with two wash treatments. After washing, pores recovered from "unavailable" to the "single pore" state, increasing the rate of data acquisition, though they then rapidly dropped over the next hour. **G.** Cumulative read count plot for MinION sequencing of the *RPGR-ORF15* amplicon in 12 pooled samples, with nine washing steps over a two-day period. Throughput rebounds after every washing step but this effect declines progressively over the course of the run. **H.** Pore availability for the MinION sequencing run of exon ORF15 plotted in G, which included nine treatments to reactivate the pores. After every wash, the pores recovered from the "unavailable" state to the "single pore" state, increasing the rate of data acquisition. **I.** Colour key showing the pore status during sequencing in the nanopore runs shown in B, D, F and H.

5.2.2.4 Screening the remaining 49 samples listed in Table 5.4

This customised *ORF15* workflow was applied to a further 49 individuals, and the initial five samples were reanalyzed. A breakdown of the 54 screened individuals is shown in Table 5.4 and full clinical details of each sample can be found in Appendix D.1. All these individuals also underwent PacBio screening, except for two who underwent MinION screening alone. These included 30 males and one female with unsolved peripheral retinopathy and four males with unsolved central retinopathy. In addition, four males and two females with RP and one male with macular dystrophy, each carrying known *ORF15* pathogenic variants (two with c.3334C>T, p.(Gln1112Ter) (Figure 5.11B); two with c.2426_2427delAG, p.(Glu809Glyfs*25) (Figure 5.6B); and three with c.2405_2406delAG, p.(Glu802Glyfs*32)), were included. 12 unaffected individuals were tested as controls to assess population variation. Figure 5.9 illustrates the variant, c.2405 2406delAG, p.(Glu802Glyfs*32) in a male (hemizygous, A and B) and Figure 5.10 shows the same variant in a female (heterozygous, A and B).

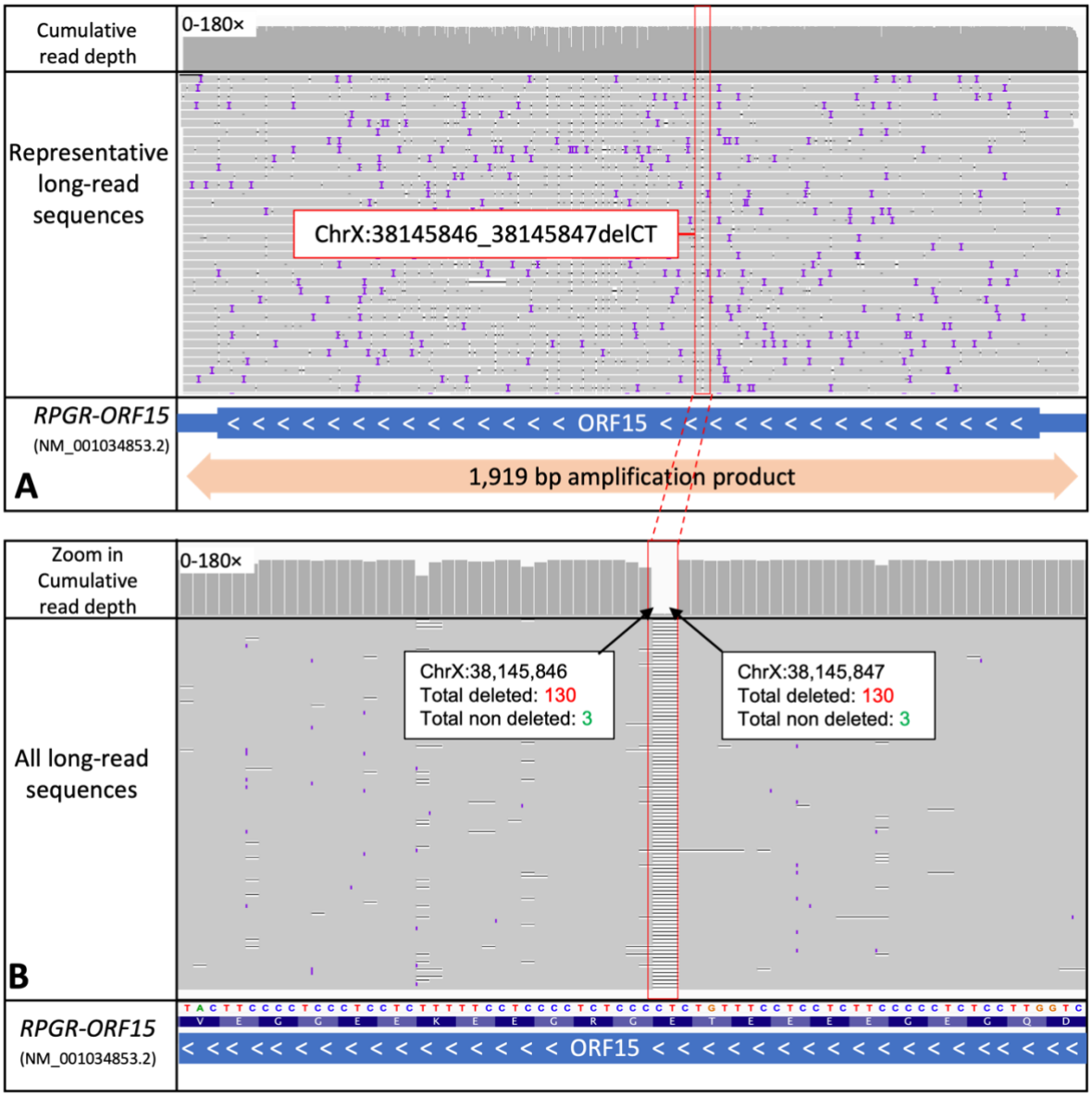


Figure 5.9: Long-read sequencing alignment at the *RPGR-ORF15* locus generated using a nanopore MinION sequencer. A. The male RP patient (753, a known control) is hemizygous for the two base-pair deletion (NM_001034853.2) c.2405 2406delAG, p.(Glu802Glyfs*32) (ChrX: 38145846_38145847delCT (hg19). **B.** Zoom in view of the same 2 bp deletion showing the total deleted and non-deleted reads.

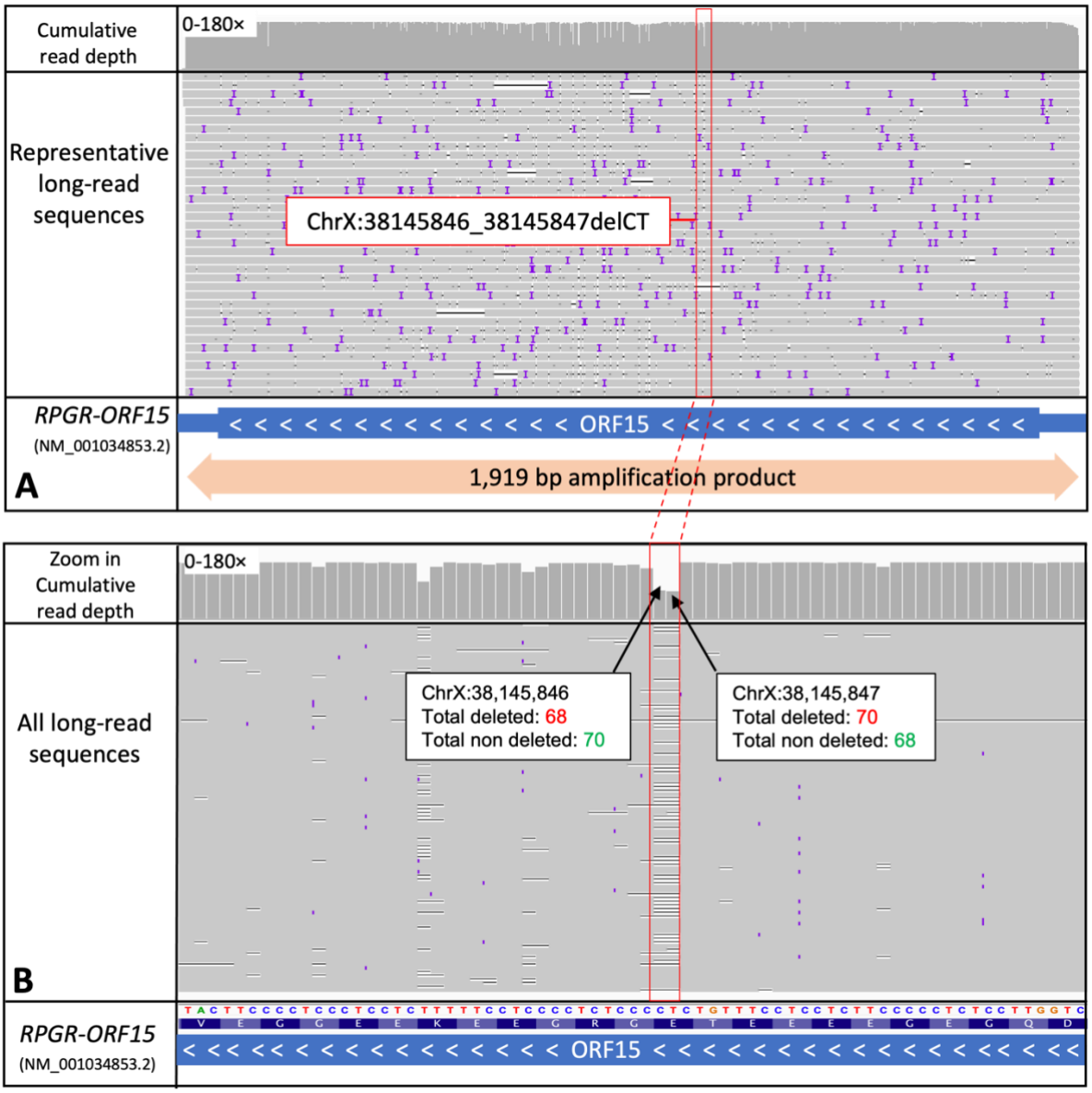


Figure 5.10: Long-read sequencing alignment at the *RPGR-ORF15* locus generated using a nanopore MinION sequencer. A. The female RP patient (749, a known control) is heterozygous for the two base-pair deletion (NM_001034853.2) c.2405 2406delAG, p.(Glu802Glyfs*32) (ChrX: 38145846_38145847delCT (hg19). **B.** Zoom in view of the same 2 bp deletion showing the total deleted and non-deleted reads.

Libraries from each DNA sample were indexed as described above and combined in pools of up to 24 cases, then sequenced on a MinION flow cell, with multiple nuclease washes performed. All previously known variants were observed and two new cases of RP were identified to be caused by the *ORF15* (NM_001034853.2) variants c.2041_2042delAA, p.(Lys681Glyfs*2), ChrX:38146210_38146211del (hg19) (Case 3685; Figure 5.11A), and c.2323_2324delAG, p.(Arg775Glufs*59), ChrX:38145933_3814594del (hg19) (Carss et al., 2017, Beigi et al., 2021) (Case 3219; Figure 5.4). These variants were also

identified via PacBio sequencing (Section 5.2.1.2), and both have been previously reported on the ClinVar database as either likely pathogenic (c.2041_2042delAA; accession number: VCV000865836.2) or pathogenic (c.2323_2324delAG; accession number: VCV000438144.18) respectively. In addition, several benign single nucleotide variants (SNVs) and in frame deletions and duplications were observed, in both cases and controls. These are listed in Appendix D.4.

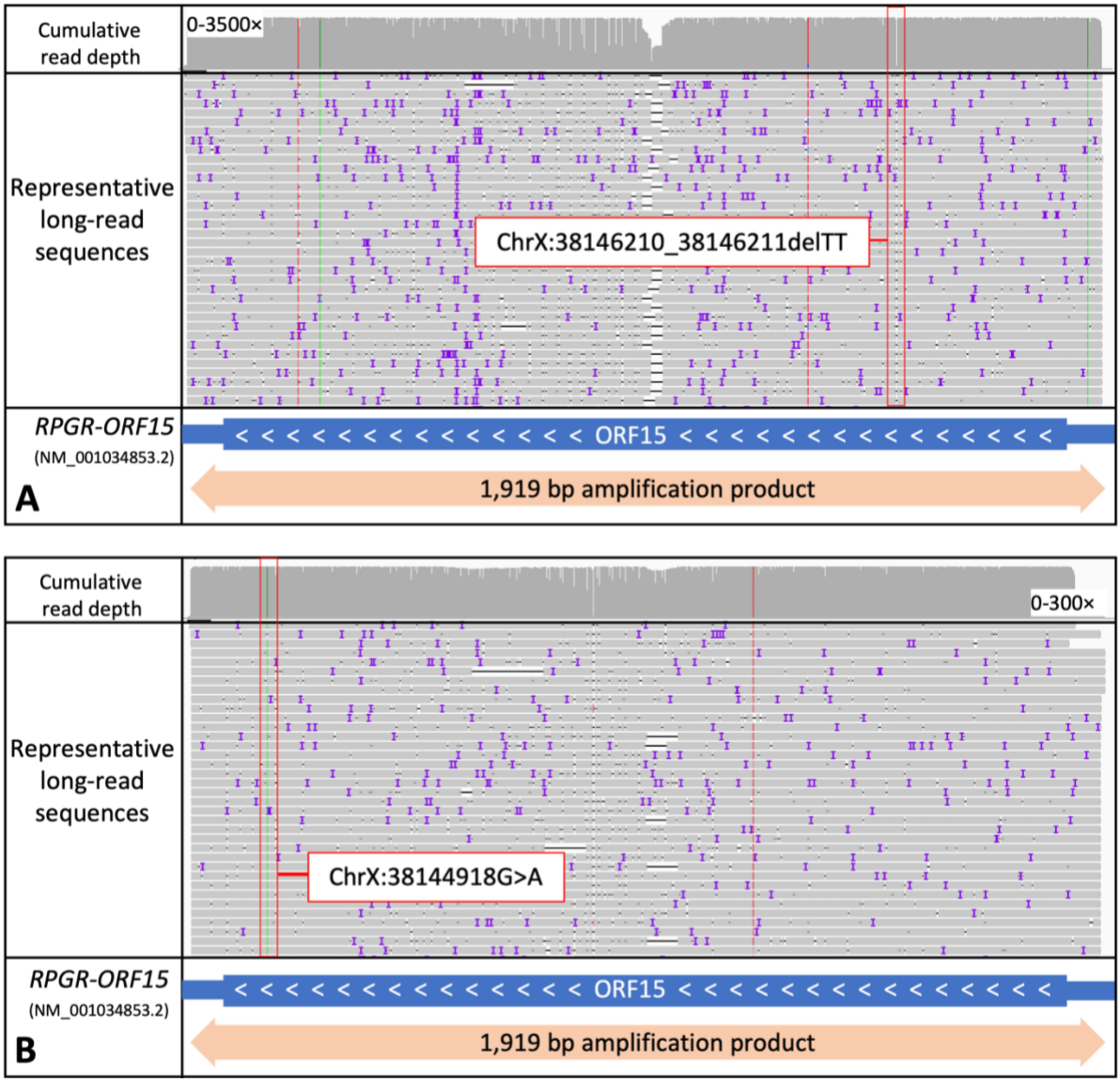


Figure 5.11: Long-read sequencing alignment at the *RPGR-ORF15* locus generated using a nanopore MinION sequencer. **A.** The male RP patient (3685) is hemizygous for a new two base-pair deletion (NM_001034853.2) c.2041_2042delAA, p.(Lys681Glyfs*2) (ChrX: 38146210_38146211delTT (hg19). **B.** The male RP patient (5156, a known control) is hemizygous for a published nonsense variant (NM_001034853.2) c.3334C>T, p.(Gln1112Ter) (ChrX:38144918G>A (hg19).

5.2.3 Discussion

PCR and Sanger sequencing to amplify the *RPGR-ORF15* locus presents a technical challenge. The sequencing protocol used by the Manchester Reference laboratory and described in Section 2.15 is able to generate high quality sequence, as shown in Figure 5.6C, but is complex and involves multiple PCRs, each with unique conditions. Other approaches have worked, but these are difficult and are often unreliable. Sequencing *RPGR-ORF15* using NGS-based methods is also prone to failure (Wang et al., 2014, Huang et al., 2015, Li et al., 2016a, Chiang et al., 2018, Maggi et al., 2020). Short-read WES, using sequencing by synthesis chemistry, results in poor depth of coverage over many highly repetitive and/or GC-rich regions (Wang et al., 2014). This is usually overcome using WGS, which has no amplification step in the library preparation, but even this approach fails with *ORF15*.

In this study, two long read sequencing technologies, PacBio and nanopore sequencing, were used to determine whether they could generate *ORF15* sequence, and which was the most effective of the two for screening this notoriously hard-to-sequence mutation hotspot. A total of 218 individuals with peripheral or central retinopathy were screened either by Pacbio or Minion, with many screened by both methods (Table 5.1). In particular, where there were significant findings, one platform was used to verify the results of the other. However, over the course of this project a number of cases from these groups were solved through other screening going on in parallel within the laboratory. These are therefore considered to be obligate negative cases in the subsequent analysis described here (Figure 5.12). Full clinical details for all 218 IRD cases tested across both platforms in this study are detailed in Appendix D.1.

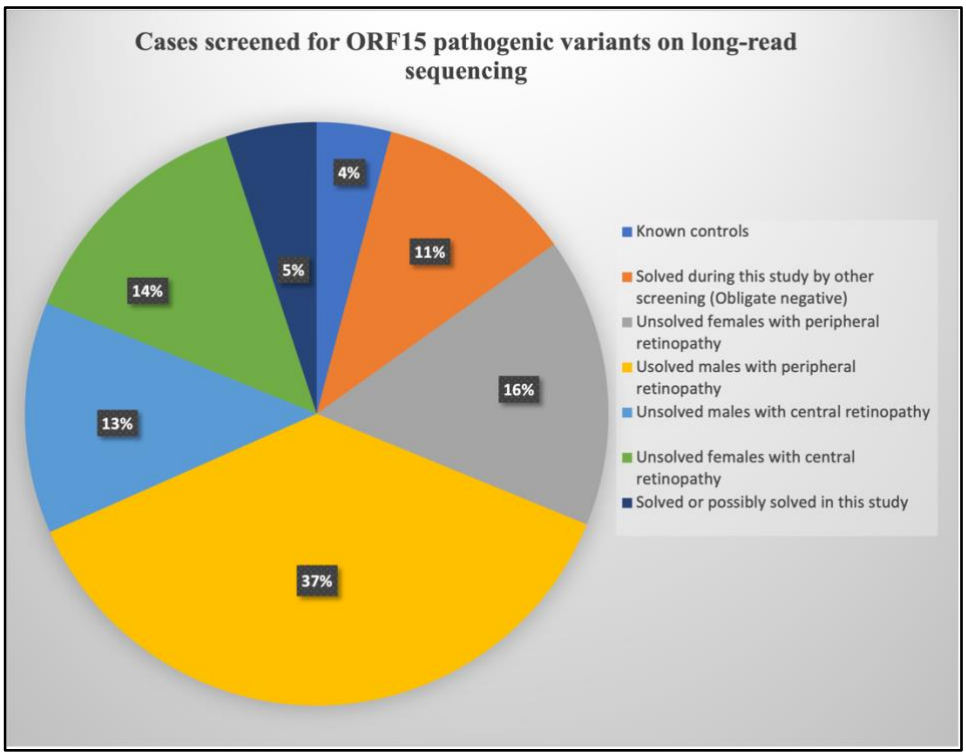


Figure 5.12: A pie chart illustrating the clinical data of 218 cases screened on either PacBio or nanopore sequencing or both platforms, for *ORF15* pathogenic variants. The cases are categorised into seven categories (legend) and the percentage of the cases in each category is displayed.

A single step of high-fidelity PCR followed by PacBio sequencing was applied to a cohort of 216 individuals with unsolved peripheral and central retinal dystrophies, as detailed in Table 5.2. In order to evaluate normal variations in *ORF15* in the population, 27 unaffected controls were tested. To amplify this region and pool the samples for cost-effectiveness, different combinations of barcoded primers were utilized. Despite the difficulties in determining the optimum PCR conditions, amplifying of *ORF15* was successful and sequencing of pools of up to 130 samples on PacBio was effective.

40 of these same cases screened by PacBio, plus an additional two cases, underwent further screening using long-range PCR target enrichment together with long read nanopore sequencing, to validate the PacBio sequencing and compare the performance of the platforms. In addition, 12 of the same controls screened by PacBio were included in nanopore sequence analysis, again to evaluate normal variations in *ORF15* in the population on this platform. Thus, a total of 54 samples were screened by nanopore sequencing, as detailed in Table 5.4. The sequencing of *ORF15* was first tried on Flongle,

which was effective and showed a sufficient coverage over the *ORF15* region. However, during the run, sequencing pores become "unavailable", possibly due to secondary structures formed by the *ORF15* repetitive sequence. This reduced the number of reads generated to around 5% of those expected. Therefore, the experiments were switched to use the MinION flow cells, which allowed repeated use of a flow cell wash kit containing DNase I. The target locus was amplified using a two-step PCR which incorporated per-sample barcodes prior to sequencing on a MinION flow cell. The DNase I wash kit was originally designed to allow flow cells to be reused to reduce the per-experiment cost by digesting any remaining DNA fragments and unclog the pores (Lipworth et al., 2020). After applying the wash kit, a further aliquot of the same library, or a freshly prepared library was reloaded. It has been established that DNase I treatment restores pores to an "active" state, resulting in higher per-run yields and cumulative read depth at the target *ORF15* locus.

Both approaches were shown to be reliable and able to read across the *ORF15* region. PacBio sequencing was superior in terms of cost and scalability. While sequencing 24 pooled indexed samples on MinION would cost approximately £40 per sample (MinION flow cell ~ £720/24, library prep/run ~ £80, 24 indexes ~ £80, and wash kit ~ £80), sequencing 130 barcoded pooled samples on PacBio would cost about £8.92 per sample (£1160 per lane/130 including library prep and sequencing). Furthermore, PacBio sequencing required fewer steps and less time to prepare the samples for sequencing compared to MinION sequencing, which required a two-step PCR and several clean up steps. However, the Minion has the advantage that it can be used at both small and large scale, with a Flongle costing around £100, while the PacBio platform as used in this study required upfront cost for a full lane totaling at least £1160.

In comparison to NGS, both methods allowed screening the entire length of the *ORF15* exon and resulted in a depth-of-coverage that allowed detection of the previously verified variants, and further pathogenic variants in previously unsolved cases. The PacBio method identified seven probable pathogenic variants in unsolved cases and two of them (c.2323_2324delAG; p.(Arg775Glufs*59) and c.2041_2042delAA; p.(Lys681Glyfs*2)) were confirmed on MinION nanopore sequencing. It is planned to use Flongle sequencing to confirm the remaining five newly identified novel pathogenic variants. In comparison to the healthy control group, the patient cohort had a higher percentage of in-frame deletions or duplications. This approach offers a useful tool for further research into the clinical significance of in-frame deletions or duplications.

Based on previous studies (Li et al., 2016a, Maggi et al., 2020, Tuupanen et al., 2022), we anticipated that male RP cases would have the highest rate of *ORF15* variants identified. Four unsolved cases out of 96 males with peripheral retinopathy (4.1%) and two unsolved cases of 37 males with central retinopathy (5.4%) were solved or probably solved by this approach. In addition, five out of 40 females with unsolved peripheral retinopathy were solved or probably solved in this study (12.5%). These figures, and particularly the proportion of males with RP solved, is lower than in other studies. This might be a consequence of the high level of consanguinity in the Bradford Pakistani community, which has been linked to an increase in recessive genetic disease inheritance in Yorkshire over other types of inheritance (Arciero et al., 2021). This in turn may mean that, in a Yorkshire RP cohort, cases of recessive RP will tend to be overrepresented, which will in turn reduce the frequency of XLRP cases. Alternatively, this may be a consequence of Yorkshire ophthalmologists taking up genetic testing for *ORF15* at the Manchester Reference laboratory, and only sending unsolved cases to the research laboratory after some *ORF15* cases have been solved by NHS testing. Further testing and segregation studies are needed to confirm the significance of the discovered variants.

One recognised limitation of the workflow required for both the PacBio and nanopore platforms is the requirement to perform PCR-based target-enrichment. This can lead to polymerase slippage across low-complexity repeats, and biased amplification of parental alleles (although this concern is mitigated when analysing hemizygous male cases). Two additional long-read target enrichment strategies are being developed which avoid PCR amplification. CRISPR/Cas9 workflows enable specific cleavage sites to be generated in bulk genomic DNA, prior to the ligation of instrument-specific sequencing adapters (Gilpatrick et al., 2020). This process involves dephosphorylating the DNA, introducing new cuts with a Cas9-gRNA complex, and binding nanopore sequencing adapters to the cuts encompassing the region of interest, before loading the sample onto the nanopore sequencer. ReadUntil sequencing (also known as adaptive sampling) using ReadFish software package, allows nanopore devices to selectively reject off-target sequences from the pore, in real-time, by reversing the voltage across individual nanopores (Miller et al., 2021a, Payne et al., 2021). This can be done by applying a signal comparison algorithm (dynamic time warping (DTW), which compares the signal from the pore with previously calculated signals for sequences of interest. However, both workflows suffer from relatively low yield and on-target read depth, which may further exacerbate the already much reduced run yields obtained from nanopore sequencing of *ORF15*.

Although little is known about the mechanism of action of *RPGR-ORF15*, it has been observed in animal models that RPGR protein deficiency or abnormal RPGR protein synthesis results in opsin mis-localization, suggesting that RPGR is crucial for protein trafficking in the photoreceptor connecting cilia (Shu et al., 2005, Pawlyk et al., 2016). Through its interaction with other ciliary proteins, it regulates ciliogenesis and maintains cilia function (Khanna et al., 2005, Patnaik et al., 2015). The *ORF15* exon contains a long, purine-rich, highly repetitive sequence which was found to be essential for correct subcellular localisation and full function of RPGR. This was demonstrated by delivering an AAV vector expressing either a long form of the human *RPGR-ORF15* gene, which had 1/3 of the *ORF15* repetitive region (126 codons) deleted, or a short form, which had the majority of *ORF15* (314 codons) deleted, to *Rpgr* null mice. The long form preserved the RPGR function and provided stability, while the shorter form has failed to rescue the mouse phenotype (Pawlyk et al., 2016). This shows that the RPGR function was restored with the benefit of enhanced stability using a moderately shortened *RPGR-ORF15* replacement gene, but not a more substantially shortened one.

As well as the benefit to patients from obtaining a confirmed molecular diagnosis, it is particularly crucial to develop a reliable, scalable, and affordable screening approach for *RPGR-ORF15* in order to offer patients the opportunity to participate in the ongoing clinical trials. This not only benefits patients but also provides direct support for clinical trials themselves, giving a larger patient pool in which to test the effectiveness of the therapy. Two major clinical trials are currently underway regarding gene therapy for X-linked RP due to variants in RPGR (https://clinicaltrials.gov/). These are led by Oxford-based biotechnology company MeiraGTx and by US company Applied Genetic Technologies Corp. The *ORF15* repetitive region has proved important in the development of these gene therapies for this form of RP. An *RPGR* vector containing an in-frame deletion of the *ORF15* repeat similar to the longer version used by Pawlyk and colleagues, described above (total deletion 378 base pairs or 126 codons) is now being tested in human clinical trials by MeiraGTx (NCT03252847). Those experiments prove that this version does restore some function in the *Rpgr* knockout mouse. However, some argue that short length *RPGR* may not fully restore RPGR function and may lead to diminished glutamylation, which is required for cone function (Martinez-Fernandez de la Camara et al., 2022). Meanwhile, Applied Genetic Technologies Corp. (AGTC) are testing full length RPGR (NCT03316560). They previously showed that when administering full-length *RPGR* to *RPGR*-mutant dogs versus a shorter variant (45 bp deletion within *ORF15*), the full-length form was favoured (Song et al., 2020). It is

therefore essential to monitor and compare the outcomes of clinical trials using the shortened RPGR in order to assess their impact on human.

In conclusion, these novel findings show that long-read sequencing by both PacBio and nanopore platforms can read through the region of *RPGR-ORF15* that is refractory to analysis by short read NGS. In nanopore sequencing, use of a flow cell wash kit containing DNase I, unblocks the pores, allowing researchers to increase yield by reloading further aliquots of the library over a 72-hour period. Both approaches described here allows the sequencing of indexed pooled libraries. The PacBio sequencer is more scalable compared to MinION sequencer since up to 130 barcoded samples can be pooled and sequenced in a single run at lower price, providing a rapid, cost-effective screening protocol for this notoriously hard-to-sequence mutation hotspot. However, the Minion provides the flexibility to do small scale experiments to optimise use of the platform. Both approaches therefore have some benefits to the user.

Chapter 6

General discussion and conclusion

6.1 Overview of main findings

The work described in this thesis was mainly focused on using NGS technologies, including second and third generation platforms, to detect novel variants implicated in inherited retinal diseases, with an emphasis on those that have proven challenging to detect.

Chapter three describes the use of WES to analyse 24 previously unscreened individuals affected with different types of IRDs, including syndromic and non-syndromic types. The genomic DNA from each sample was sequenced using WES, and the data was then subject to a first pass analysis to identify the most likely disease-causing variant(s) associated with the described phenotype. This approach was combined with CNVs analysis using ExomeDepth tool. Pathogenic or likely pathogenic variants were identified and then confirmed by Sanger sequencing in eight individuals, including variants in known IRDs genes *CDHR1*, *RHO*, *CNGA3*, *BBS10*, *ABCA4*, *USH2A*, *PRPF31*, and *SLC25A46*. Furthermore, it was identified that three genes, *OPA1* in a patient with CORD, *MFSD9* in a patient with syndromic IRD, and *VCAN* in a patient with FEVR, all harbour potential pathogenic variants that are inconsistent with the described phenotypes. However, about 60% of the cohort studied here remained unsolved, suggesting that there could be novel genes or variants in non-coding regions that are not detected by the methods used in this study. This low detection rate might be attributed to the fact that the filtering strategy used in this study excluded synonymous changes and changes beyond the immediate 2 bp splice donor and acceptor sites for each exon, which may contribute to the pathogenesis of IRDs by altering splicing. In addition, the analysis involved working on single cases, some of whom lacked family history while others had an ambiguous diagnosis, further complicating the variant filtering process (further discussed below).

In chapter four of this thesis, genetic analysis was performed on six unrelated families (Seven patients) with sporadic or multiplex late onset macular degeneration and one patient with AMD. The seven affected individuals presented with late onset bull's eye macular dystrophy, with some also reporting nyctalopia since their teens. These eight affected individuals were found to carry a heterozygous ~ 126 kb deletion encompassing *CRX* and neighbouring genes *TPRX1* and *SULT2A1*. Though initially overlooked, the

deletion was detected in MIPs, WES, and WGS, in three different individuals and it was subsequently detected and confirmed in the other individuals by the PCR assay designed in this study. Combining the PCR assay with cloning and Sanger sequencing enabled the precise breakpoints in all eight individuals to be identified. Microsatellite genotyping and SNPs analysis around the breakpoints suggested that this deletion seems likely to be a relatively old founder mutation caused by non-allelic homologous recombination (NAHR) due to *Alu* sequences at each breakpoint. Previous studies suggest that carriers for a single defective *CRX* allele had normal ocular phenotype. The findings in this chapter revealed a novel mutation mechanism and hotspot due to repeat structure at this locus and confirmed that *CRX* haploinsufficiency causes a late onset macular dystrophy phenotype (Yahya et al., 2023). Furthermore, the PCR assay described will expedite the rapid screening of large patient cohorts. Publishing these findings will help others to screen further cohorts of central retinal dystrophy cases for this deletion. This publication (Yahya et al., 2023) led to correspondence with others (Mustafi and Chao, 2022) who suggested that hypomorphic second alleles missed in our screening could be contributing to disease in these cases, though we countered (Inglehearn et al., 2022) that alleles of major effect could be excluded. Further examination of WGS to exclude the possibility that a second predisposing allele exists in these patients would be beneficial, but there are no plans to do that at this stage (discussed further in Section 4.3).

Chapter five describes the use of third generation sequencing technologies to identify disease-causing variants in the hard to sequence mutation hotspot *RPGR-ORF15*. Both long read nanopore sequencing (Flongle and MinION) and SMRT PacBio sequencing were employed to screen a cohort of 218 affected individuals with either peripheral or central retinopathy, including nine cases with four known mutations, to test the validity of the methods. Unaffected controls were also included to assess population variation. A single step of high-fidelity PCR followed by PacBio sequencing was successful in reading through the *ORF15* region. Nanopore sequencing was first tested using a Flongle flow cell which was successful in reading through a 2kb PCR-amplified fragment containing *ORF15*. However, the yield was less than expected because this G-rich, highly repetitive region blocks the available sequencing pores. The experiments were then switched to use MinION flow cell to allow using a flow cell wash kit containing DNase I. The use of the DNase I treatment allowed repeated re-loading of the flow cell and increased the number of reads that were mapped to *ORF15* locus. Both methods allowed the sequencing of indexed pooled samples and thereby the detection of previously verified mutations and additional pathogenic variants in previously unsolved cases. The detection rate using this

approach was lower than in other studies using different approaches, and this may be explained by the high level of consanguinity in the Bradford Pakistani community, which has been linked to an increase in recessive genetic disease inheritance in Yorkshire over other types of inheritance. Although long range PCR target enrichment has been used in other studies, combining it with long read sequencing is a novel approach. PacBio sequencer was easier to perform and superior in terms of cost and scalability, providing a rapid cost-effective protocol. However, the Minion has the advantage that it can be used at both small and large scale (further discussion can be found in Section 5.3).

6.2 IRDs genetic testing, challenges, and future prospects

Understanding the genetic basis of IRDs is essential for early and accurate diagnosis to improve the quality of life for patients. Genetic testing can aid in identifying the disease-causing variant(s), allowing for establishing a clear genetic diagnosis for the affected individual. Understand the inheritance pattern of the condition allows clinical geneticists and genetic counsellors to provide patients with a better understanding of the risk and help them make well-informed decisions about their family planning. Genetic testing is also a fundamental step for affected individuals to determine their eligibility for clinical trials and emerging treatments.

6.2.1 NGS (SGS)

Mutation detection in IRDs can be achieved by applying traditional methods of genetic sequencing such as Sanger sequencing but it is expensive and time consuming to test every gene associated with retinal diseases. Using the arrayed primer extension reaction for genotyping on microarrays are also available; however, they can only detect the known mutations (Jaakson et al., 2003). Recently, NGS technologies have been globally used to increase the detection rate of IRD causing mutations. Targeted sequencing (TS) of all known and candidate IRD genes has been developed and widely used in diagnostic settings (Consugar et al., 2015, Ellingford et al., 2016, Dockery et al., 2017, Hitti-Malin et al., 2022, Watson et al., 2014). It can yield high diagnostic rates and for up to two-thirds of cases with IRDs, a genetic explanation can be given (Khan et al., 2020, Stephenson et al., 2021, Mc Clinton et al., 2023). Regardless of the clinical diagnosis, TS is a good option to screen known IRDs genes since it enables the sequencing of smaller genomic regions, including non-coding regions, which improves the coverage of clinically significant genes (García Bohórquez et al., 2021). It is an economical method since multiplexing several samples allows for faster processing, and reduced data storage.

However, TS approaches cannot be used to detect mutations in novel genes; as a result, a more comprehensive technique, such as WES or WGS, can be used.

Since the cost of the TS and WES is now nearly equal, WES was chosen to identify the pathogenic variants in cases included in this thesis (Chapter 3) because they were single cases and some of them had syndromic disease or an ambiguous diagnosis. Without having to update the panel, WES allows the discovery of possible new genes by identifying disease-causing variants in the protein-coding regions of any gene. The detection rate using WES in diagnosing IRDs ranges from 50-70% (Riera et al., 2017, Wang et al., 2018, Zhang et al., 2018). However, the detection rate in this study was less than that (~ 30%). This may be because some of these cases have no family history. The likelihood of identifying a new pathogenic variant or gene in a single case using WES alone is low, but a number of strategies can aid in identifying novel genes in WES data. These include a linkage-based strategy, homozygosity mapping, a double-hit strategy, an overlap strategy, a *de novo* strategy and a candidate strategy (Gilissen et al., 2012). Some of these strategies typically require either sequencing data from multiple family members (e.g., trio WES) (Wang et al., 2016a) or a large control group (Dewey et al., 2016). Other possible explanations for missing pathogenic variants include incorrect diagnosis, which might have negatively impacted the filtering method employed, or clinical heterogeneity, which might render causative variants in any given gene very rare. Furthermore, multiple IRD phenotypes can result from mutations in the same IRD gene (Hull et al., 2014, Georgiou et al., 2021), meaning that a filtering strategy based on phenotype data could miss unexpected novel phenotype associations. Some cases may present with phenocopies of retinal degenerations that resemble the clinical signs of inherited retinal dystrophies. In these cases, the phenocopies may be caused by bacterial or viral inflammation in addition to immunological causes (Kellner et al., 1996). For example, a number of conditions may resemble the clinical signs of RP and are categorised as pseudo-retinitis pigmentosa. It is vital to differentiate them from hereditary RP since they can be treated and do not have an underlying genetic component (Verbakel et al., 2018).

Nonetheless, in some cases, despite the clear diagnosis and full family history, the pathogenic mutations were not detected. This could be explained by the fact that the causative variants may be located in non-coding regions including, deep intronic variants, variants in regulatory regions, as well as variants in genes encoding non-coding RNAs (e.g., micro RNAs (miRNAs), small nuclear RNAs (snRNAs), and long non-coding RNAs). Deep intronic variations can result either in skipping an exon or creating a

pseudoexon to induce aberrant splicing. They have been proposed as disease causing variants in IRDs (Webb et al., 2012, Khan et al., 2019b, Weisschuh et al., 2020c, Qian et al., 2021). Variations in regulatory regions have an impact on the regulatory elements (e.g., promotors, enhancers, 5' untranslated region (UTR), 3' UTR, and topologically associated domains (TADs)) that regulate the expression of genes and proteins. Non-coding variations are increasingly being shown to have a significant impact in rare diseases (Tomaselli et al., 2017, Bauwens et al., 2019, de Bruijn et al., 2020). Moreover, the filtering process used in this study excluded synonymous variants, therefore including synonymous changes in variant interpretation may improve the detection rate. Synonymous changes may contribute to the pathogenesis of human disorders through various mechanisms such as splicing alterations (Bali and Bebok, 2015, Zhu et al., 2022). The causative mutations can be complex alleles which have been proposed as disease-causing variants in Stargardt disease (Salles et al., 2017). Common variants that are classified as benign variants (hypomorphic alleles) may be pathogenic under certain conditions, which also have been reported in *ABCA4*-related diseases (Zernant et al., 2017, Runhart et al., 2018). All of these possibilities are compounded by the risk in WES that one allele might be detected but discounted because the second allele is missed in a gene associated with recessive disease. Additionally, mutations in more than one gene might be the underlying cause (García Bohórquez et al., 2021). Modifier alleles in other genes can cause variable expressivity and penetrance in patients carrying the same mutation. The phenotype variability can be in age of onset, severity, or degeneration pattern. For instance, two modifier loci, the *CNOT3* gene (a master regulator of global gene expression) and an *MSR1* repeat element in the *PRPF31* promoter were found to affect the disease penetrance in *PRPF31*-related retinopathy (Section 1.6.1.2) (McLenachan et al., 2021). It has also been shown that the clinical severity of the disease and age of onset are correlated with X-inactivation ratios in female carriers of *RPGR* pathogenic variants (Fahim et al., 2020b). While analysing a patient's genome, variants of uncertain significance (VUS) should therefore also be considered. It is challenging to classify these variants due to the nature of the variant itself (e.g., missense, synonymous change, in frame deletion/duplication, and non-coding variants), and its association with the phenotype is unclear (Federici and Soddu, 2020). Reclassification of VUS can be achieved by segregating the variant within the family, observing the phenotype changes through functional testing of the mutation effect *in vitro*, or collecting more cases with the same VUS and phenotype (Iancu et al., 2021). Technical issues with WES, such as biased read depth in high GC regions (Kieleczawa, 2006) and alignment challenges with

highly repetitive regions (Ebbert et al., 2019, Watson et al., 2021), may also hinder the identification of the underlying genetic causes.

These issues might be mitigated to some extent using WGS, which covers the whole genome and provides a better detection of structural variations, non-coding variants, and variants in GC rich regions compared to WES. With PCR-free WGS, libraries may be created without being amplified by PCR, which prevents DNA polymerase errors, improves coverage of GC-rich regions, and allows for the detection of inversions and balanced translocations (Pipis et al., 2019). WGS was successful in identifying missed variants in IRDs (Nishiguchi et al., 2013, Carss et al., 2017, Nash et al., 2022). It should be noted that some of the unsolved cases described here had ambiguous syndromes that might be caused by a CNV encompassing multiple adjacent genes, so employing WGS may improve the detection rate. WGS is more effective for finding CNVs since it has a greater coverage (98% coverage of the genome) than WES (95% coverage of targeted exons) does (Lelieveld et al., 2015), and CNVs are known to be involved in IRDs (Ellingford et al., 2018, Zampaglione et al., 2020). It is challenging to discover synonymous or non-coding variations using WGS without a candidate gene or region, since it generates huge volumes of variant data. Hence, it would be beneficial to begin by working on synonymous or non-coding variations in genes in which variants cause IRDs, particularly in recessive cases where one allele has been found. Also, sequencing large cohorts and more family members may be useful in reducing the variant list. When resources and time are limited, it can be beneficial to prioritise patients based on the severity of the phenotype and the availability of the DNA samples from other family members. Furthermore, applying whole gene sequencing (GS), which enables capturing exonic, intronic, and regulatory regions for the target gene, can be an economical alternative for WGS. GS was used to investigate 1054 unsolved Stargardt disease cases by sequencing the *ABCA4* gene using single-molecule molecular inversion probes (smMIPs). This method allowed for the detection of pathogenic structural variations and deep intronic variants in 25% of the cases (Khan et al., 2020). It would be also helpful to combine WES or WGS with RNA sequencing to investigate how the identified variant affects the expression or function of the protein although this will only be effective if the genes are expressed in the blood or fibroblasts of the patient, which is often not the case for IRD genes. In 35% of previously undiagnosed probands with a rare muscle condition, RNA sequencing helped finding the disease-causing splicing variants (Cummings et al., 2017). The identification of non-coding variants will boost the detection rate and catalyse the discovery of novel disease mechanisms. This can be achieved following the

recommendations suggested by Ellingford et al for adaptation of the ACMG guidelines to improve the clinical interpretation of these variants (Ellingford et al., 2022). However, WGS is costly due to the high sequencing costs as well as the processing requirements for computing power, labour hours, and storage.

Of note, ophthalmological imaging has undergone significant advancements to refine the phenotype in order to make a more accurate diagnosis and prognosis. Machine learning has been utilised to evaluate fundus and OCT images in AMD patients (Schmidt-Erfurth et al., 2018, Russakoff et al., 2019) and fundus autofluorescence imaging in patients with Stargardt disease (Charng et al., 2020). Accurate phenotyping will be made possible by machine learning, which will advance NGS screening and improve the detection rate (Dockery et al., 2021). Researchers from Moorfield Eye Hospital have employed an artificial intelligence system in partnership with DeepMind Health to assess OCT images and other eye testing and suggest the best option for referral (De Fauw et al., 2018).

It should be noted that the long lists of variants produced by WES or WGS may contain incidental or secondary findings (SFs) unrelated to the initial sequencing purpose. For the patient and the clinician, these findings could be useful or of medical value. According to the American College of Medical Genetics and Genomics (ACMG), 73 genes (SF v3.0) should be investigated, and the results should be reported to the ordering clinician (Miller et al., 2021c, Miller et al., 2021b). Of note, the *RPE65* gene was included in this list since there is now an FDA-approved gene therapy for it. As a result of being notified of a potentially pathogenic variant in one of these 73 genes, the patient will have a better chance of receiving an early diagnosis and the benefits of therapeutic intervention. The ACMG also proposed that patients should have the choice to opt out of SF testing. This can be done by signing a thorough consent that states what information will and will not be disclosed, along with receiving pre- and post-test counselling (Dockery et al., 2021). Applying TS methods reduces the probability of identifying incidental or secondary findings. In some IRD studies, to reduce the chance of such findings, the whole genome is sequenced, but only the variants in the genes involved in IRDs are investigated using a variant filtering with a synthetic gene panel (Carss et al., 2017).

6.2.2 TGS

Long read sequencing platforms can produce extremely long reads in real time (ranging from kilobases to > 1Mb in length) directly from native DNA (Logsdon et al., 2020), enabling the identification of structural variations and *de novo* assembly, as well as haplotype phasing (Shi et al., 2016, Wang et al., 2020, Sakamoto et al., 2021). Since it

does not require PCR amplification, PCR-related difficulties are avoided. It provides a method for the investigation of so-called "dark and camouflaged" genomic loci that have remained resistant to short read NGS analysis because of informatic challenges (e.g., the inability to determine an unambiguous mapping position) or wet-laboratory procedures related to their underlying genomic architecture (e.g., the high GC content of some first exons) (Huddleston and Eichler, 2016, Ebbert et al., 2019, Watson et al., 2021). Using PacBio HiFi and Oxford nanopore ultra-long read sequencing, the truly complete sequence of human genome (telomere to telomere (T2T CHM13)) was published (Nurk et al., 2022). The PacBio method is a useful option for identifying DNA modifications and novel isoforms as it can be used to sequence both RNA and DNA (Chen et al., 2019, Hu et al., 2021). However, in addition to the large size of these devices, PacBio needs extensive start-up funding. In contrast, ONT instruments are portable, compact, and affordable, making them a good choice for low-resource settings and field applications. These methods have a significant degree of uncertainty, between 2-15% for ONT devices (Hu et al., 2021) and over 10% for PacBio devices (Quail et al., 2012). To improve the error rate, several adjustments have been made for PacBio such as combining short read and PacBio approaches (Berbers et al., 2020) and using HiFi sequencing (up to 99.8% accuracy). ONT techniques' error rates have decreased by modifying the nanopore protein to process nucleotides more slowly and utilising better base calling models (McCombie et al., 2019).

The first- and second-generation sequencing methods were employed in the research described in this thesis. However, using these techniques has revealed certain technological difficulties. Therefore, long read sequencing technologies were employed to find probable disease-causing variants in regions that were difficult to sequence using traditional methods. *ORF15* exon in the *RPGR* gene, a G-rich, highly repetitive region that is a mutation hotspot for XLRP (Vervoort et al., 2000), was screened using both SMRT PacBio sequencing (the Sequel) and nanopore sequencing (Flongle and MinION; ONT). Both approaches were effective in reading through the region of the *RPGR ORF15* that is refractory to short read NGS analysis. A workflow is described that enables the sequencing of indexed pooled libraries using a MinION sequencer (up to 24 individuals in a single flow cell) and PacBio sequencer (up to 130 individuals per run). This suggests that PacBio is superior in terms of pricing and scalability. Researchers looking to sequence other similarly interactable genomic regions will likely find this protocol useful. This protocol, however, requires PCR-based target enrichment, which might result in polymerase slippage over low complexity repeats and biased parental allele amplification.

To avoid PCR amplification, CRISPR/Cas9 enrichment long read sequencing (Gilpatrick et al., 2020) can be used. CAs9 Targeting of CHromosomal segment (CATCH)-nanopore sequencing was used to find complex structural rearrangements in *ABCA4* that could result in cryptic splicing. Also, it made it possible to sequence the NGS-interactable *OPN1LW*/*OPN1MW* gene array (Arno et al., 2022). Nanopore devices may use ReadUntil sequencing, also known as adaptive sampling, to selectively reject off-target sequences from the pore in real-time by reversing the voltage across individual nanopores (Miller et al., 2021a, Payne et al., 2021).

Our understanding of the genetic basis of IRDs will improve as a result of TGS effectiveness in locating novel disease-causing variants that were refractory to NGS analysis. In addition, applying new and improved in-silico tools to interpret novel variants identified by sequencing in the lab, together with relevant functional analyses, are essential in increasing detection rates and determining the impact of these novel variants on protein function. Human iPSC-derived retinal organoids can be used as *in vitro* 3D models because there is no requirement for transient expression of the mutation and because it provides a more retina-like environment (further discussed in Section 6.3). As more IRDs genes are found, the rapid enhancements in DNA sequencing technologies, in-silico tools, and functional analyses will enable further understanding of how the retina functions. Despite such enhancements, it is likely that the underlying genetic causes in a small percentage of screened patients will remain unidentified, either because they are the result of complex multifactorial disease or because they are non-genetic forms of disease. Understanding in these cases will ultimately come from other areas of retinal research.

Nevertheless, the continuous improvement in genetic screening methods will increase the likelihood of providing an absolute molecular diagnosis, aid in the recruitment of patients into gene therapy clinical trials and improve counselling for them so they may make knowledgeable choices about their family planning, education, and lifestyle.

6.3 Implications for patients, healthcare professionals, and overview of gene therapy

Because IRDs are heterogeneous, it is essential that the patient receives a highly specific clinical and molecular diagnosis along with an exact inheritance pattern through molecular genetic testing. This will help healthcare professionals provide the patient and their family with the proper management and counselling. Genetic testing will also give

prognostic information, which is crucial for identifying infants and young children who are at risk of developing systemic problems like renal disease to benefit from early intervention and therapy. Precise genetic diagnosis will permit preimplantation and prenatal diagnosis in families who decided to undergo genetic testing. In this thesis, additional genetic testing methods were developed that will assist healthcare professionals in providing accurate diagnoses quickly and affordably. Among these are a PCR assay to check a large cohort of patients with macular degeneration for a heterozygous deletion encompassing *CRX* (Chapter 4), and a method to screen exon *ORF15* of the *RPGR* gene using MinION and PacBio sequencers (Chapter 5). However, patients are becoming better informed of clinical trials and gene therapy, which is the main reason they request genetic testing.

The eye is a suitable target for gene therapies due to the blood retina barrier (BRB), which ensures that the introduction of a foreign antigen like a viral vector would be tolerated without inducing a severe inflammatory reaction (Kaur et al., 2008). The barrier inhibits the locally injected vector's undesirable systemic effects, considering that only a small amount is required to provide a therapeutic response. The eye is easily accessible by surgery and the disease progression can be monitored using different non-invasive approaches such as fundus autofluorescence, spectral domain OCT, visual acuity, and bio-microscopy tests (Liang et al., 2001, Sahel et al., 2015). Several strategies are being tested to find effective therapies for IRDs, including molecular therapies, stem cell-based therapies, and retinal prostheses.

Gene augmentation (also known as gene replacement therapy) is a DNA-based molecular therapy in which a therapeutic vector is used to deliver a normal copy of the mutant gene into the host cell. Therapeutic vectors can be either viral vectors, such as adeno-associated viruses (AAVs), lentiviral vectors (LVs), adenoviruses, or helper dependent adenoviruses, or non-viral vectors such as nanoparticles (NPs), naked DNA, or liposomes. AAVs are frequently used in IRDs gene therapies because of their high affinity for certain retinal cells and minimal immunogenicity. The most advanced molecular therapeutic strategy is gene augmentation, which was used in developing a gene therapy for biallelic *RPE65* pathogenic variants in Leber congenital amaurosis (LCA) in humans. The therapy was approved by the FDA (Voretigene neparvovec, Luxturna™, SparkTherapeutics) (Russell et al., 2017, Food and Administration, 2017) and the European Medicines Agency (EMA) in 2018. After subretinal injection of an AAV2 carrying the *PRPH2* transgene into the *Prph2* knockout Retinal Degeneration Slow

mouse model, the structural integrity of the photoreceptors was restored (Ali et al., 2000). The *MERTK* gene was delivered to the retinas of *Mertk* knockout RCS rats via AAV (Smith et al., 2003) and LVs (Tschernutter et al., 2005), and this has improved ERG performance and saved more photoreceptors. This led to a phase I clinical trial in humans in Saudi Arabia (NCT01482195), which was initiated using subretinal administration of rAAV2-VMD2-hMERTK (LaVail et al., 2016). In an *Abca4*-knockout mouse model of Stargardt disease, recombinant AAV with an AAV5 capsid (rAAV2/5 serotype) was used to package the large *Abca4* gene (8.9 kb) as the AAV2 cargo capacity is restricted to 4.7 kb. The subretinal injection of rAAV2/5 vector containing *Abca4* resulted in decreased lipofuscin content and improved retinal morphology (Allocca et al., 2008). Additionally, LV based gene therapy decreased lipofuscin accumulation in mouse *Abca4* knockout models, paving the way to testing the safety of administering the lentivirus-mediated *ABCA4* gene (SAR422459) in a phase I/II clinical trial in humans (NCT01367444) (Davis, 2018). The first gene therapy clinical trial for *RPGR*-associated RP using AAV8.GRK1.coRPGR-ORF15 vector was initiated in 2017 (NCT03116113; Biogen Inc.), and the initial results of a dose-escalation interventional study was published in 2020 (Cehajic-Kapetanovic et al., 2020b). A shortened form of the human *RPGR-ORF15* gene using the AAV2/5.GRK1.RPGR-ORF15 vector is being tested in patients with *RPGR-ORF15* mutations (NCT04671433; MeiraGTx). Also Applied Genetic Technologies Corp. is evaluating the safety and efficacy of vector AAV2tYF.GRK1.coRPGR-ORF15 (NCT04850118) (Martinez-Fernandez de la Camara et al., 2022). Other gene therapy clinical trials using AAVs, or LVs are being tested in patients with Usher syndrome (*MYO7A* gene; NCT02065011, NCT01505062), choroideremia (*CHM* gene; NCT01461213, NCT02341807, and NCT03496012), and achromatopsia (*CNGB3* gene; NCT02599922 and NCT03001310).

Gene replacement therapy is costly, time-consuming, and ineffective for autosomal dominant forms of IRDs. Genome editing is another DNA-based molecular therapy that offers an alternative promising approach. It can be done *in vivo* or in patient derived cells followed by cell transplantation (*ex vivo*) (Sanjurjo-Soriano and Kalatzis, 2018). The most advanced system, clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9, is frequently used in genome editing (Hung et al., 2016, Tabebordbar et al., 2016). It is based on single guide RNA (sgRNA), which guides the endonuclease Cas9 to target a genomic position where it creates double stranded DNA breaks (DSBs) upstream of the protospacer adjacent motif (PAM) site. These breaks can be repaired then by either nonhomologous end-joining (NHEJ) or homology-directed repair (HDR) (Ran

et al., 2013). NHEJ is an error-prone genome editing mechanism that introduces insertions or deletions (indels) of various lengths. Therefore, this generates frameshifts and/or premature stop codons within open reading frames of target genes, leading to creation of non-functional proteins (Liu et al., 2017). HDR, on the other hand, is a high-fidelity pathway which requires a homology containing donor DNA template to repair the cut. HDR mediated genome modification allows for precise introduction of any required genetic engineering (Ziccardi et al., 2019). Since RPE and photoreceptors lack a HDR repair mechanism, alternative strategies using the NHEJ mechanism have been developed (Suzuki et al., 2016). A new Cas9/RecA system was designed based on HDR and allowed for correcting of the *Pde6b* mutation in *rd1* mice (Cai et al., 2019). A CRISPR/Cas9 based approach (EDIT-101) was developed to target the splice mutation c.2991+1655A>G in *CEP290* that causes LCA10 and showed promising results in mice and nonhuman primate models (Maeder et al., 2019). This approach reached a phase I/II trial (NCT03872479; Editas Medicine/Allergan), making it the first *in vivo* use of CRISPR/Cas9 technology in human (Ledford, 2020). However, the company were forced to shelve the trial after finding that the response obtained was not broad enough. They are now seeking a new partner and identifying the appropriate patient population, since the individuals who showed improvement in corrected vision were those who were homozygous for the mutation. Genome editing can also be achieved using zinc-finger nucleases (ZFNs) and transcription activator like effector nucleases (TALENs). These molecules create DSBs, which consequently trigger one of the DNA repair mechanisms including microhomology-mediated end joining, NHEJ, or HDR. To create the DSBs, a DNA binding protein domain must guide the nuclease to the target locus. These approaches are time-consuming and challenging since new guiding proteins need to be created for each target (Schierling et al., 2012, Gaj et al., 2013). Nonetheless, the development of these approaches and attempts to enhance their safety and efficiency will contribute to fulfilling the promise in the treatment of IRDs.

Another molecular therapeutic approach depends on the use of antisense oligonucleotides (AONs) to modify aberrant splicing. AONs consist of small DNA or RNA molecules that can be delivered to their target pre-mRNA either as naked oligonucleotides or expressed by viral vectors such as AAV (Ziccardi et al., 2019). It is not applicable to all mutations, but they are effective in restoring normal splicing by eliminating abnormal splicing events caused by activation of cryptic splicing sites. In patients with compromised immune systems, Formivirsen was the first AON-based medication authorised to treat cytomegalovirus retinitis in patients with AIDS (Group, 2002). About 15% of all IRD-

causing mutations affect splicing and among these is the recurrent deep intronic mutation in *CEP290* (c.2991+1665A>G) that causes LCA (den Hollander et al., 2006). In both *in vitro* and *in vivo* models, AON-based treatment for *CEP290*-associated LCA (sepofarsen) has demonstrated promising outcomes (Collin et al., 2012, Garanto et al., 2016, Duijkers et al., 2018), leading to a phase I/II clinical trial (Cideciyan et al., 2019). Unfortunately, it was revealed in April 2022 that phase II/III will not be pursued because it failed to improve vision in patients with LCA. AON based therapies have also been developed to correct deep intronic mutations in *OPA1* (c.610+364G>A) that causes optic neuropathies (Bonifert et al., 2016), *CHM* (c.315-4587T>A) that causes choroideremia (Garanto et al., 2018), *USH2A* (c.7595-2144A>G) that causes Usher syndrome (Slijkerman et al., 2016), and *ABCA4* (c.4539+2001G>A and c.4539+2028C>T) that causes Stargardt disease (Albert et al., 2018). However, there are still some difficulties with the usage of AONs, such as delivery, durability, and off-target effects.

Other promising approaches in treating IRDs include stem cell-based therapy, which is applicable to patients with IRDs who are at an advanced stage of degeneration. It aims to incorporate the externally injected cells to replace either RPE or photoreceptors and therefore restore visual function. The major forms of retinal stem cell therapy are human embryonic stem cells (hESCs) and the induced pluripotent stem cells (iPSCs). Patient derived somatic cells such as fibroblasts and lymphoblasts can be reprogramed into iPSCs, which in turn differentiate into retinal precursor cells (Garcia et al., 2015). These cells can be delivered to the eye by either intravitreal or subretinal injection. The genome editing tools can be used to correct the mutations in the patient derived stem cells, to be eventually transplanted back to the patient (Takahashi et al., 2007). Genetic modification would not be necessary for the use of hESCs. However, compared to iPSCs generated from the patient, employing hESC requires ethical consideration and has a larger risk of immunological rejection following transplant (Vázquez-Domínguez et al., 2019). Using hESC-derived RPE reported promising outcomes in patients with AMD and Stargardt disease (Schwartz et al., 2015, Song et al., 2015). Transplantation of iPSC-derived RPE cells to patients with AMD has also shown success in improving the visual acuity with no immune response induced (Mandai et al., 2017).

Human iPSC-derived retinal organoids provide a new method for disease modelling in the investigation of the pathophysiology of inherited retinal diseases. This technique has, for example, greatly facilitated research into the role of ciliary function in the retina. Mutations in the *CEP290* gene were investigated by looking at the cilia biogenesis and

function in iPSC-derived retinal organoids from *CEP290*-LCA and *CEP290*-Joubert syndrome patients (Shimada et al., 2017). For the purpose of studying the morphology and localization of the photoreceptors, another model of *RPGR* mutation-specific RP retinal organoids was developed. CRISPR-Cas9 was used in this study to repair the mutation, and the results revealed a significant improvement (Deng et al., 2018b). Additionally, iPSC-derived retinal organoids have been used as an *in vitro* 3D model for evaluating the effectiveness of novel therapeutic approaches, including assessing the protective effects of 4-hydroxytamoxifen and diethylstilbestrol on the health of photoreceptors (Ito et al., 2017). Retinal organoids can also be used in the investigation of gene therapies by, for example, evaluating the effectiveness of gene delivery in humans (Garita-Hernandez et al., 2020). Human iPSC-derived retinal organoids can also serve as a source of cells for transplantation in therapeutic settings. In June 2020, the first clinical trial using human iPSC-derived retinal organoids (conducted at Kobe City Eye Hospital in Japan) was approved (jRCTa050200027) (Sugita et al., 2016).

The therapeutic strategies described above focus primarily in saving existing vision and preventing disease progression. These approaches will not benefit patients with end-stage disease, whose retinas have degenerated beyond a point where such therapies will be effective. Patients with such advanced retinal diseases may benefit from retinal prosthetic implants. These implants can interact directly with inner retinal neurons, which are still present in advanced stages of retinal degeneration, bypassing the damaged photoreceptors. Retinal prosthetic implants contain an image capture device, an image processor, a stimulator chip, and an electrode array. The light emitted by visible objects is transformed into microelectrodes and then directed onto the inner and middle retinal neurons. Numerous devices have been developed, and they are divided into epiretinal, subretinal, and suprachoroidal prosthesis based on where they are placed anatomically (Yue et al., 2016, Vázquez-Domínguez et al., 2019). The most popular epiretinal prosthetic device is Argus II, which has been implanted in over 250 RP patients and has shown some signs of improving visual performance and function (Da Cruz et al., 2013, da Cruz et al., 2016, Dagnelie et al., 2017).

As well as the strategies described above, there are other therapeutic approaches being developed to treat IRDs, and patients are becoming aware of these therapies. It is important to note that therapeutic strategies such as gene replacement are gene specific, while strategies such as correcting splicing with an antisense oligonucleotide are mutation specific. Patients given a treatment that does not work for their particular form of IRD

will have no therapeutic benefit, and if such patients are enrolled inappropriately onto a clinical trial, they may prevent the trial from reaching a successful conclusion. This shows that it is essential for IRD patients and their clinicians to know and understand exactly which genetic form of IRD they have, not only for the purposes of informed genetic counselling but also to ensure that they receive the appropriate treatment.

6.4 Wider significance in ophthalmic genetics

Most individuals affected with IRDs will be initially diagnosed by a general ophthalmologist, paediatric ophthalmologist, or optician. Questions like, "What type of genetic testing is applicable? Who needs to be tested? What type of clinical care following testing is appropriate? Who will cover testing costs?" will emerge after the patient has been diagnosed.

Both understanding the genetic basis of the disease and availability of therapies will help healthcare professionals to offer a precision medicine programme for each patient. Precision medicine is a new method of treating and preventing disease that considers a person's unique genetic background, environmental factors, and lifestyle (König et al., 2017). For precision medicine to succeed, accurate molecular diagnosis should be established. The cost of genetic testing limits it from being available for every patient. Insurance companies often will not pay for genetic testing unless there is solid proof that the results will predict the clinical condition and influence clinical care (Capasso, 2014). Patients without insurance may engage in research studies or pay for their own testing. The expense of molecular investigations is a significant impediment to testing. Therefore, molecular testing will be more widely available as the price declines.

Even though genetic testing has now become more affordable, there is still a critical need for trained ophthalmologists to interpret the testing results and explain the findings to the patients and their families. Nowadays, establishing a large international database and close collaboration between ophthalmologists and clinical geneticists is necessary in order to provide patients with the proper clinical care and treatment. A number of such databases do exist, including the various gene specific LOVD databases, the HGMD database and ClinVar. However, having one comprehensive source dedicated to human IRDs, hosted perhaps by an organisation such as the European Inherited Disease Consortium, would greatly simplify variant interpretation. Such a database could also be informed by the recently published human Pangenome (Liao et al., 2023), which will improve understanding of normal genetic variation in people of different ethnicity.

6.5 Concluding remarks

With mutations in over 280 genes, including non-coding and modifying variations, IRDs are genetically complicated diseases. When modifier loci have been identified in some IRDs, the difference between Mendelian and multifactorial forms of IRDs has been blurred, further complicating molecular diagnoses. NGS-based molecular genetic testing has addressed several challenges, and the advent of long read sequencing has, to some extent, further improved findings. Many unresolved cases may be solved by using comprehensive long read whole genome sequencing to look at modifier loci and variations in non-coding regions. Long read sequencing using PacBio or nanopore sequencing needs further refinement in terms of bioinformatic analysis and result interpretation, particularly when used to screen challenging to sequence regions, and further improvements in the technology will help to bring costs down. The genetic detection rate in IRDs may also be increased by enhancing in-silico tools and utilising deep learning technology to deliver an accurate clinical phenotyping. Accurate molecular diagnosis and prognosis will increase the patients' benefit from the different therapeutic approaches once they are approved and available. Once the disease-causing variant(s) are identified and based on the disease prognosis, the appropriate therapeutic option can be chosen. However, the safety and efficiency of this treatment should be tested first on cellular/animal models. We can fight blindness and provide hope and vision for many visually impaired individuals worldwide only when clinicians, scientists, funding agencies, and industry collaborate nationally and internationally.

Bibliography

ABOSHIHA, J., DUBIS, A. M., CARROLL, J., HARDCASTLE, A. J. & MICHAELIDES, M. 2016. The cone dysfunction syndromes. British Journal of Ophthalmology, 100, 115-121.

ABRAMS, A. J., FONTANESI, F., TAN, N. B., BUGLO, E., CAMPEANU, I. J., REBELO, A. P., KORNBERG, A. J., PHELAN, D. G., STARK, Z. & ZUCHNER, S. 2018. Insights into the genotype-phenotype correlation and molecular function of SLC25A46. Human mutation, 39, 1995-2007.

ABRAMS, A. J., HUFNAGEL, R. B., REBELO, A., ZANNA, C., PATEL, N., GONZALEZ, M. A., CAMPEANU, I. J., GRIFFIN, L. B., GROENEWALD, S. & STRICKLAND, A. V. 2015. Mutations in SLC25A46, encoding a UGO1-like protein, cause an optic atrophy spectrum disorder. Nature genetics, 47, 926-932.

ACLAND, G. M., RAY, K., MELLERSH, C. S., GU, W., LANGSTON, A. A., RINE, J., OSTRANDER, E. A. & AGUIRRE, G. D. 1998. Linkage analysis and comparative mapping of canine progressive rod–cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. Proceedings of the National Academy of Sciences, 95, 3048-3053.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. Nature methods, 7, 248-249.

AHMED, A. N., TAHIR, R., KHAN, N., AHMAD, M., DAWOOD, M., BASIT, A., YASIN, M., NOWSHID, M., MARWAN, M. & SULTAN, K. 2021. USH2A gene variants cause Keratoconus and Usher syndrome phenotypes in Pakistani families. BMC ophthalmology, 21, 1-8.

AJMAL, M., KHAN, M. I., NEVELING, K., TAYYAB, A., JAFFAR, S., SADEQUE, A., AYUB, H., ABBASI, N. M., RIAZ, M. & MICHEAL, S. 2013. Exome sequencing identifies a novel and a recurrent BBS1 mutation in Pakistani families with Bardet-Biedl syndrome. Molecular vision, 19, 644.

AKAHORI, M., TSUNODA, K., MIYAKE, Y., FUKUDA, Y., ISHIURA, H., TSUJI, S., USUI, T., HATASE, T., NAKAMURA, M. & OHDE, H. 2010. Dominant mutations in RP1L1 are responsible for occult macular dystrophy. The American Journal of Human Genetics, 87, 424-429.

AL-KHUZAEI, S., BROADGATE, S., FOSTER, C. R., SHAH, M., YU, J., DOWNES, S. M. & HALFORD, S. 2021. An overview of the genetics of ABCA4 retinopathies, an evolving story. Genes, 12, 1241.

AL-MAGHTHEH, M., VITHANA, E., TARTTELIN, E., JAY, M., EVANS, K., MOORE, T., BHATTACHARYA, S. & INGLEHEARN, C. F. 1996. Evidence for a major retinitis pigmentosa locus on 19q13. 4 (RP11) and association with a unique bimodal expressivity phenotype. American journal of human genetics, 59, 864.

AL-MAGHTHEH, M., GREGORY, C., INGLEHEARN, C., HARDCASTLE, A. & BHATTACHARYA, S. 1993. Rhodopsin mutations in autosomal dominant retinitis pigmentosa. Human Mutation, 2, 249-255.

ALBERT, S., GARANTO, A., SANGERMANO, R., KHAN, M., BAX, N. M., HOYNG, C. B., ZERNANT, J., LEE, W., ALLIKMETS, R. & COLLIN, R. W. 2018. Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. The American Journal of Human Genetics, 102, 517-527.

ALI, R. R., SARRA, G.-M., STEPHENS, C., ALWIS, M. D., BAINBRIDGE, J. W., MUNRO, P. M., FAUSER, S., REICHEL, M. B., KINNON, C. & HUNT, D. M. 2000. Restoration of photoreceptor ultrastructure and function in retinal degeneration slow mice by gene therapy. Nature genetics, 25, 306-310.

ALI, R. R. & SOWDEN, J. C. 2011. DIY eye. Nature, 472, 42-43.

ALLIKMETS, R., SINGH, N., SUN, H., SHROYER, N. F., HUTCHINSON, A., CHIDAMBARAM, A., GERRARD, B., BAIRD, L., STAUFFER, D. & PEIFFER, A. 1997. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Starqardt macular dystrophy. Nature genetics, 15, 236-246.

ALLOCCA, M., DORIA, M., PETRILLO, M., COLELLA, P., GARCIA-HOYOS, M., GIBBS, D., KIM, S. R., MAGUIRE, A., REX, T. S. & DI VICINO, U. 2008. Serotype-dependent packaging of large genes in adeno-associated viral vectors results in effective gene delivery in mice. The Journal of clinical investigation, 118, 1955-1964.

ALMIND, G. J., GRØNSKOV, K., MILEA, D., LARSEN, M., BRØNDUM-NIELSEN, K. & EK, J. 2011. Genomic deletions in OPA1 in Danish patients with autosomal dominant optic atrophy. BMC medical genetics, 12, 1-5.

AMBARDAR, S., GUPTA, R., TRAKROO, D., LAL, R. & VAKHLU, J. 2016. High throughput sequencing: an overview of sequencing chemistry. Indian journal of microbiology, 56, 394-404.

AOUN, M., PASSERINI, I., CHIURAZZI, P., KARALI, M., DE RIENZO, I., SARTOR, G., MURRO, V., FILIMONOVA, N., SERI, M. & BANFI, S. 2021. Inherited retinal diseases due to RPE65 variants: from genetic diagnostic management to therapy. International Journal of Molecular Sciences, 22, 7207.

ARANDA, M. L. & SCHMIDT, T. M. 2021. Diversity of intrinsically photosensitive retinal ganglion cells: circuits and functions. Cellular and Molecular Life Sciences, 78, 889-907.

ARBOUR, N. C., ZLOTOGORA, J., KNOWLTON, R. G., MERIN, S., ROSENMANN, A., KANIS, A. B., ROKHLINA, T., STONE, E. M. & SHEFFIELD, V. C. 1997. Homozygosity mapping of achromatopsia to chromosome 2 using DNA pooling. Human molecular genetics, 6, 689-694.

ARCIERO, E., DOGRA, S. A., MALAWSKY, D. S., MEZZAVILLA, M., TSISMENTZOGLOU, T., HUANG, Q. Q., HUNT, K. A., MASON, D., SHARIF, S. M., VAN HEEL, D. A., SHERIDAN, E., WRIGHT, J., SMALL, N., CARMI, S., ILES, M. M. & MARTIN, H. C. 2021. Fine-scale population structure and demographic history of British Pakistanis. Nature Communications, 12, 7189.

ARNO, G., HULL, S., CARSS, K., DEV-BORMAN, A., CHAKAROVA, C., BUJAKOWSKA, K., VAN DEN BORN, L. I., ROBSON, A. G., HOLDER, G. E. & MICHAELIDES, M. 2016. Reevaluation of the retinal dystrophy due to recessive alleles of RGR with the discovery of a cis-acting mutation in CDHR1. Investigative ophthalmology & visual science, 57, 4806-4813.

ARNO, G., JURKUTE, N., BOLES, T. C., MICHAELIDES, M., MAHROO, O. A. R. & WEBSTER, A. R. 2022. Cas9 targeted long-read nanopore sequencing of ABCA4 and the OPN1LW/OPN1MW gene array for investigation of inherited retinal disease. Investigative Ophthalmology & Visual Science, 63, 516–A0093-516–A0093.

ARSHAVSKY, V. Y., LAMB, T. D. & PUGH, E. N., JR. 2002. G proteins and phototransduction. Annu Rev Physiol, 64, 153-87.

ARSHAVSKY, V. Y. & WENSEL, T. G. 2013. Timing is everything: GTPase regulation in phototransduction. Investigative ophthalmology & visual science, 54, 7725-7733.

AUDO, I., MOHAND-SAÏD, S., DHAENENS, C. M., GERMAIN, A., ORHAN, E., ANTONIO, A., HAMEL, C., SAHEL, J. A., BHATTACHARYA, S. S. & ZEITZ, C. 2012. RP1 and autosomal dominant rod–cone dystrophy: Novel mutations, a review of published variants, and genotype–phenotype correlation. Human mutation, 33, 73-80.

AUNG, T., OCAKA, L., EBENEZER, N. D., MORRIS, A. G., KRAWCZAK, M., THISELTON, D. L., ALEXANDER, C., VOTRUBA, M., BRICE, G. & CHILD, A. H. 2002. A major marker for normal tension glaucoma: association with polymorphisms in the OPA1 gene. Human genetics, 110, 52-56.

AYUSO, C. & MILLAN, J. M. 2010. Retinitis pigmentosa and allied conditions today: a paradigm of translational research. Genome Medicine, 2, 34.

AYYAGARI, R., DEMIRCI, F. Y., LIU, J., BINGHAM, E. L., STRINGHAM, H., KAKUK, L. E., BOEHNKE, M., GORIN, M. B., RICHARDS, J. E. & SIEVING, P. A. 2002. X-linked recessive atrophic macular degeneration from RPGR mutation. Genomics, 80, 166-171.

AZIZZADEH PORMEHR, L., AHMADIAN, S., DAFTARIAN, N., MOUSAVI, S. A. & SHAFIEZADEH, M. 2020. PRPF31 reduction causes mis-splicing of the phototransduction genes in human organotypic retinal culture. European Journal of Human Genetics, 28, 491-498.

BADER, I., BRANDAU, O., ACHATZ, H., APFELSTEDT-SYLLA, E., HERGERSBERG, M., LORENZ, B., WISSINGER, B., WITTWER, B. R., RUDOLPH, G. N. & MEINDL, A. 2003. X-linked retinitis pigmentosa: RPGR mutations in most families with definite X linkage and clustering of mutations in a short sequence stretch of exon ORF15. Investigative Ophthalmology & Visual Science, 44, 1458-1463.

BALI, V. & BEBOK, Z. 2015. Decoding mechanisms by which silent codon changes influence protein biogenesis and function. The international journal of biochemistry & cell biology, 64, 58-74.

BALL, E. V., STENSON, P. D., ABEYSINGHE, S. S., KRAWCZAK, M., COOPER, D. N. & CHUZHANOVA, N. A. 2005. Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. Human mutation, 26, 205-213.

BARBITOFF, Y. A., POLEV, D. E., GLOTOV, A. S., SEREBRYAKOVA, E. A., SHCHERBAKOVA, I. V., KISELEV, A. M., KOSTAREVA, A. A., GLOTOV, O. S. & PREDEUS, A. V. 2020. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. Scientific reports, 10, 1-13.

BARTNIK, M., NOWAKOWSKA, B., DERWIŃSKA, K., WIŚNIOWIECKA-KOWALNIK, B., KĘDZIOR, M., BERNACIAK, J., ZIEMKIEWICZ, K., GAMBIN, T., SYKULSKI, M. & BEZNIAKOW, N. 2014. Application of array comparative genomic hybridization in 256 patients with developmental delay or intellectual disability. Journal of applied genetics, 55, 125-144.

BAUWENS, M., GARANTO, A., SANGERMANO, R., NAESSENS, S., WEISSCHUH, N., DE ZAEYTIJD, J., KHAN, M., SADLER, F., BALIKOVA, I. & VAN CAUWENBERGH, C. 2019. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. Genetics in Medicine, 21, 1761-1771.

BAYLOR, D. A., LAMB, T. & YAU, K.-W. 1979. Responses of retinal rods to single photons. The Journal of physiology, 288, 613-634.

BEALES, P., ELCIOGLU, N., WOOLF, A., PARKER, D. & FLINTER, F. 1999. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. Journal of medical genetics, 36, 437-446.

BEHARRY, S., ZHONG, M. & MOLDAY, R. S. 2004. N-retinylidene-phosphatidylethanolamine is the preferred retinoid substrate for the photoreceptor-specific ABC transporter ABCA4 (ABCR). Journal of Biological Chemistry, 279, 53972-53979.

BEIGI, F., DEL POZO-VALERO, M., MARTIN-MERIDA, I., MANAVIAT, M. R., AYUSO, C. & GHASEMI, N. 2021. Homozygous females for a X-linked RPGR-ORF15 mutation in an Iranian family with retinitis pigmentosa. Experimental Eye Research, 211, 108714.

BEN-ZVI, A., LACOSTE, B., KUR, E., ANDREONE, B. J., MAYSHAR, Y., YAN, H. & GU, C. 2014. Mfsd2a is critical for the formation and function of the blood–brain barrier. Nature, 509, 507-511.

BERBERS, B., SALTYKOVA, A., GARCIA-GRAELLS, C., PHILIPP, P., ARELLA, F., MARCHAL, K., WINAND, R., VANNESTE, K., ROOSENS, N. H. & DE KEERSMAECKER, S. C. 2020. Combining short and long read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified Bacillus. Scientific reports, 10, 1-13.

BERGEN, A. A., PLOMP, A. S., SCHUURMAN, E. J., TERRY, S., BREUNING, M., DAUWERSE, H., SWART, J., KOOL, M., VAN SOEST, S. & BAAS, F. 2000. Mutations in ABCC6 cause pseudoxanthoma elasticum. Nature genetics, 25, 228-231.

BERGER, W., KLOECKENER-GRUISSEM, B. & NEIDHARDT, J. 2010. The molecular basis of human retinal and vitreoretinal diseases. Progress in retinal and eye research, 29, 335-375.

BEST, S., LORD, J., ROCHE, M., WATSON, C. M., POULTER, J. A., BEVERS, R. P., STUCKEY, A., SZYMANSKA, K., ELLINGFORD, J. M. & CARMICHAEL, J. 2022. Molecular diagnoses in the congenital malformations caused by ciliopathies cohort of the 100,000 Genomes Project. Journal of Medical Genetics, 59, 737-747.

BHARDWAJ, A., YADAV, A., YADAV, M. & TANWAR, M. 2022. Genetic dissection of non-syndromic retinitis pigmentosa. Indian Journal of Ophthalmology, 70, 2355-2385.

BIRTEL, J., EISENBERGER, T., GLIEM, M., MÜLLER, P. L., HERRMANN, P., BETZ, C., ZAHNLEITER, D., NEUHAUS, C., LENZNER, S. & HOLZ, F. G. 2018. Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. Scientific reports, 8, 1-11.

BLANCO-KELLY, F., JAIJO, T., ALLER, E., AVILA-FERNANDEZ, A., LÓPEZ-MOLINA, M. I., GIMÉNEZ, A., GARCÍA-SANDOVAL, B., MILLÁN, J. M. & AYUSO, C. 2015. Clinical aspects of Usher syndrome and the USH2A gene in a cohort of 433 patients. JAMA ophthalmology, 133, 157-164.

BLANTON, S. H., HECKENLIVELY, J. R., COTTINGHAM, A. W., FRIEDMAN, J., SADLER, L. A., WAGNER, M., FRIEDMAN, L. H. & DAIGER, S. P. 1991. Linkage mapping of autosomal dominant retinitis pigmentosa (RP1) to the pericentric region of human chromosome 8. Genomics, 11, 857-869.

BOK, D. 1993. The retinal pigment epithelium: a versatile partner in vision. Journal of cell science, 1993, 189-195.

BOLZ, H., EBERMANN, I. & GAL, A. 2005. Protocadherin-21 (PCDH21), a candidate gene for human retinal dystrophies. Mol Vis, 11, 929-33.

BONIFERT, T., KARLE, K. N., TONAGEL, F., BATRA, M., WILHELM, C., THEURER, Y., SCHOENFELD, C., KLUBA, T., KAMENISCH, Y. & CARELLI, V. 2014. Pure and syndromic optic atrophy explained by deep intronic OPA1 mutations and an intralocus modifier. Brain, 137, 2164-2177.

BONIFERT, T., MENENDEZ, I. G., BATTKE, F., THEURER, Y., SYNOFZIK, M., SCHÖLS, L. & WISSINGER, B. 2016. Antisense oligonucleotide mediated splice correction of a deep intronic mutation in OPA1. Molecular Therapy-Nucleic Acids, 5, e390.

BONNEAU, D., COLIN, E., OCA, F., FERRÉ, M., CHEVROLLIER, A., GUÉGUEN, N., DESQUIRET-DUMAS, V., N'GUYEN, S., BARTH, M. & ZANLONGHI, X. 2014. Early-onset Behr syndrome due to compound heterozygous mutations in OPA1. Brain, 137, e301-e301.

BOSTICARDO, M., YAMAZAKI, Y., COWAN, J., GIARDINO, G., CORSINO, C., SCALIA, G., PRENCIPE, R., RUFFNER, M., HILL, D. A. & SAKOVICH, I. 2019. Heterozygous FOXN1 variants cause low TRECs and severe T cell lymphopenia, revealing a crucial role of FOXN1 in supporting early thymopoiesis. The American Journal of Human Genetics, 105, 549-561.

BOUGHMAN, J., VERNON, M. & SHAVER, K. 1983. Usher syndrome: definition and estimate of prevalence from two high-risk populations. Journal of chronic diseases, 36, 595-603.

BRAUN, D. A., SCHUELER, M., HALBRITTER, J., GEE, H. Y., PORATH, J. D., LAWSON, J. A., AIRIK, R., SHRIL, S., ALLEN, S. J. & STEIN, D. 2016. Whole exome sequencing identifies causative mutations in the majority of consanguineous or familial cases with childhood-onset increased renal echogenicity. Kidney international, 89, 468-475.

BREUER, D. K., YASHAR, B. M., FILIPPOVA, E., HIRIYANNA, S., LYONS, R. H., MEARS, A. J., ASAYE, B., ACAR, C., VERVOORT, R. & WRIGHT, A. F. 2002. A comprehensive mutation analysis of RP2 and RPGR in a North American cohort of families with X-linked retinitis pigmentosa. The American Journal of Human Genetics, 70, 1545-1554.

BRÉZIN, A. P., NEDELEC, B., BARJOL, A., ROTHSCHILD, P.-R., DELPECH, M. & VALLEIX, S. 2011. A new VCAN/versican splice acceptor site mutation in a French Wagner family associated with vascular and inflammatory ocular features. Molecular Vision, 17, 1669.

BRINGMANN, A., IANDIEV, I., PANNICKE, T., WURM, A., HOLLBORN, M., WIEDEMANN, P., OSBORNE, N. N. & REICHENBACH, A. 2009a. Cellular signaling and factors involved in Müller cell gliosis: neuroprotective and detrimental effects. Progress in retinal and eye research, 28, 423-451.

BRINGMANN, A., PANNICKE, T., BIEDERMANN, B., FRANCKE, M., IANDIEV, I., GROSCHE, J., WIEDEMANN, P., ALBRECHT, J. & REICHENBACH, A. 2009b. Role of retinal glial cells in neurotransmitter uptake and metabolism. Neurochemistry international, 54, 143-160.

BUJAKOWSKA, K. M., FERNANDEZ-GODINO, R., PLACE, E., CONSUGAR, M., NAVARRO-GOMEZ, D., WHITE, J., BEDOUKIAN, E. C., ZHU, X., XIE, H. M. & GAI, X. 2017. Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. Genetics in Medicine, 19, 643-651.

BUJAKOWSKA, K. M., ZHANG, Q., SIEMIATKOWSKA, A. M., LIU, Q., PLACE, E., FALK, M. J., CONSUGAR, M., LANCELOT, M.-E., ANTONIO, A. & LONJOU, C. 2015. Mutations in IFT172 cause isolated retinal degeneration and Bardet–Biedl syndrome. Human molecular genetics, 24, 230-242.

BUKOWY-BIERYŁŁO, Z., ZIĘTKIEWICZ, E., LOGES, N. T., WITTMER, M., GEREMEK, M., OLBRICH, H., FLIEGAUF, M., VOELKEL, K., RUTKIEWICZ, E. & RUTLAND, J. 2013. RPGR mutations might cause reduced orientation of respiratory cilia. Pediatric pulmonology, 48, 352-363.

BUNCE, C., XING, W. & WORMALD, R. 2010. Causes of blind and partial sight certifications in England and Wales: April 2007–March 2008. Eye, 24, 1692-1699.

BURIN-DES-ROZIERS, C., ROTHSCHILD, P. R., LAYET, V., CHEN, J. M., GHIOTTI, T., LEROUX, C., CREMERS, F. P., BRÉZIN, A. P. & VALLEIX, S. 2017. Deletions overlapping VCAN exon 8 are new molecular defects for Wagner disease. Human Mutation, 38, 43-47.

BUSKIN, A., ZHU, L., CHICHAGOVA, V., BASU, B., MOZAFFARI-JOVIN, S., DOLAN, D., DROOP, A., COLLIN, J., BRONSTEIN, R. & MEHROTRA, S. 2018.

Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. Nature communications, 9, 1-19.

CAI, Y., CHENG, T., YAO, Y., LI, X., MA, Y., LI, L., ZHAO, H., BAO, J., ZHANG, M. & QIU, Z. 2019. In vivo genome editing rescues photoreceptor degeneration via a Cas9/RecA-mediated homology-directed repair pathway. Science advances, 5, eaav3335.

CAPASSO, J. E. 2014. The cost of genetic testing for ocular disease: who pays? Current Opinion in Ophthalmology, 25, 394-399.

CAPEANS, C., BLANCO, M., LAREU, M., BARROS, F., PIÑEIRO, A., SANCHEZ-SALORIO, M. & CARRACEDO, A. 1998. Linkage analysis in a large Spanish family with X-linked retinitis pigmentosa: phenotype—genotype correlation. Clinical genetics, 54, 26-32.

CARELLI, V., SABATELLI, M., CARROZZO, R., RIZZA, T., SCHIMPF, S., WISSINGER, B., ZANNA, C., RUGOLO, M., LA MORGIA, C. & CAPORALI, L. 2015. 'Behr syndrome'with OPA1 compound heterozygote mutations. Brain, 138, e321-e321.

CARSS, K. J., ARNO, G., ERWOOD, M., STEPHENS, J., SANCHIS-JUAN, A., HULL, S., MEGY, K., GROZEVA, D., DEWHURST, E. & MALKA, S. 2017. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. The American Journal of Human Genetics, 100, 75-90.

CASTEL, S. E., CERVERA, A., MOHAMMADI, P., AGUET, F., REVERTER, F., WOLMAN, A., GUIGO, R., IOSSIFOV, I., VASILEVA, A. & LAPPALAINEN, T. 2018. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. Nature genetics, 50, 1327-1334.

CASTIGLIONE, A., BUSI, M. & MARTINI, A. 2013. Syndromic hearing loss: an update. Hearing, Balance and Communication, 11, 146-159.

CASTIGLIONE, A. & MÖLLER, C. 2022. Usher syndrome. Audiology Research, 12, 42-65.

CEHAJIC-KAPETANOVIC, J., MCCLEMENTS, M. E., WHITFIELD, J., SHANKS, M., CLOUSTON, P. & MACLAREN, R. E. 2020a. Association of a novel intronic variant in RPGR with hypomorphic phenotype of X-linked retinitis pigmentosa. JAMA ophthalmology, 138, 1151-1158.

CEHAJIC-KAPETANOVIC, J., XUE, K., MARTINEZ-FERNANDEZ DE LA CAMARA, C., NANDA, A., DAVIES, A., WOOD, L. J., SALVETTI, A. P., FISCHER, M. D., AYLWARD, J. W. & BARNARD, A. R. 2020b. Initial results from a

first-in-human gene therapy trial on X-linked retinitis pigmentosa caused by mutations in RPGR. Nature medicine, 26, 354-359.

CESCA, F., BETTELLA, E., POLLI, R., LEONARDI, E., ASPROMONTE, M. C., SICILIAN, B., STANZIAL, F., BENEDICENTI, F., SENSI, A. & CIORBA, A. 2020. Frequency of Usher gene mutations in non-syndromic hearing loss: higher variability of the Usher phenotype. Journal of Human Genetics, 65, 855-864.

CHANG, Y.-F., IMAM, J. S. & WILKINSON, M. F. 2007. The nonsense-mediated decay RNA surveillance pathway. Annual review of biochemistry, 76, 51-74.

CHARLESWORTH, G., BALINT, B., MENCACCI, N. E., CARR, L., WOOD, N. W. & BHATIA, K. P. 2016. SLC25A46 mutations underlie progressive myoclonic ataxia with optic atrophy and neuropathy. Movement Disorders, 31, 1249-1251.

CHARNG, J., XIAO, D., MEHDIZADEH, M., ATTIA, M. S., ARUNACHALAM, S., LAMEY, T. M., THOMPSON, J. A., MCLAREN, T. L., DE ROACH, J. N. & MACKEY, D. A. 2020. Deep learning segmentation of hyperautofluorescent fleck lesions in Stargardt disease. Scientific Reports, 10, 16491.

CHEN, C., SUN, Q., GU, M., QIAN, T., LUO, D., LIU, K., XU, X. & YU, S. 2020. Multimodal imaging and genetic characteristics of Chinese patients with USH2A-associated nonsyndromic retinitis pigmentosa. Molecular genetics & genomic medicine, 8, e1479.

CHEN, H., GAO, F., HE, M., DING, X. F., WONG, A. M., SZE, S. C., YU, A. C., SUN, T., CHAN, A. W. H. & WANG, X. 2019. Long-read RNA sequencing identifies alternative splice variants in hepatocellular carcinoma and tumor-specific isoforms. Hepatology, 70, 1011-1025.

CHEN, S., WANG, Q.-L., NIE, Z., SUN, H., LENNON, G., COPELAND, N. G., GILBERT, D. J., JENKINS, N. A. & ZACK, D. J. 1997. Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. Neuron, 19, 1017-1030.

CHEN, T.-C., HUANG, D.-S., LIN, C.-W., YANG, C.-H., YANG, C.-M., WANG, V. Y., LIN, J.-W., LUO, A. C., HU, F.-R. & CHEN, P.-L. 2021. Genetic characteristics and epidemiology of inherited retinal degeneration in Taiwan. NPJ genomic medicine, 6, 1-8.

CHEN, Z., BATTINELLI, E., FIELDER, A., BUNDEY, S., SIMS, K., BREAKEFIELD, X. & CRAIG, I. 1993. A mutation in the Norrie disease gene (NDP) associated with X–linked familial exudative vitreoretinopathy. Nature genetics, 5, 180-183.

CHIANG, J. P., LAMEY, T. M., WANG, N. K., DUAN, J., ZHOU, W., MCLAREN, T. L., THOMPSON, J. A., RUDDLE, J. & DE ROACH, J. N. 2018. Development of

high-throughput clinical testing of RPGR ORF15 using a large inherited retinal dystrophy cohort. Investigative Ophthalmology & Visual Science, 59, 4434-4440.

CHIVERS, M., LI, N., PAN, F., WIEFFER, H., SLOWIK, R. & LEARTSAKULPANITCH, J. 2021. The Burden of X-Linked Retinitis Pigmentosa on Patients and Society: A Narrative Literature Review. ClinicoEconomics and Outcomes Research: CEOR, 13, 565.

CHOI, E. H., DARUWALLA, A., SUH, S., LEINONEN, H. & PALCZEWSKI, K. 2021. Retinoids in the visual cycle: role of the retinal G protein-coupled receptor. Journal of lipid research, 62.

CHOI, M., SCHOLL, U. I., JI, W., LIU, T., TIKHONOVA, I. R., ZUMBO, P., NAYIR, A., BAKKALOĞLU, A., ÖZEN, S. & SANJAD, S. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences, 106, 19096-19101.

CHUNG, D. C. & TRABOULSI, E. I. 2009. Leber congenital amaurosis: clinical correlations with genotypes, gene therapy trials update, and future directions. Journal of American Association for Pediatric Ophthalmology and Strabismus, 13, 587-592.

CHURCHILL, J. D., BOWNE, S. J., SULLIVAN, L. S., LEWIS, R. A., WHEATON, D. K., BIRCH, D. G., BRANHAM, K. E., HECKENLIVELY, J. R. & DAIGER, S. P. 2013. Mutations in the X-linked retinitis pigmentosa genes RPGR and RP2 found in 8.5% of families with a provisional diagnosis of autosomal dominant retinitis pigmentosa. Investigative ophthalmology & visual science, 54, 1411-1416.

CIDECIYAN, A. V., JACOBSON, S. G., DRACK, A. V., HO, A. C., CHARNG, J., GARAFALO, A. V., ROMAN, A. J., SUMAROKA, A., HAN, I. C. & HOCHSTEDLER, M. D. 2019. Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. Nature medicine, 25, 225-228.

CLARKE, J., WU, H.-C., JAYASINGHE, L., PATEL, A., REID, S. & BAYLEY, H. 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nature nanotechnology, 4, 265-270.

CLARKE, L., REBELO, C., GONCALVES, J., BOAVIDA, M. & JORDAN, P. 2001. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Molecular Pathology, 54, 351.

COCO-MARTIN, R. M., SANCHEZ-TOCINO, H. T., DESCO, C., USATEGUI-MARTÍN, R. & TELLERÍA, J. J. 2020. PRPH2-related retinal diseases: Broadening the clinical spectrum and describing a new mutation. Genes, 11, 773.

COE, B. P., WITHERSPOON, K., ROSENFELD, J. A., VAN BON, B. W., VULTO-VAN SILFHOUT, A. T., BOSCO, P., FRIEND, K. L., BAKER, C., BUONO, S. &

VISSERS, L. E. 2014. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nature genetics, 46, 1063-1071.

COHN, A. C., TOOMES, C., POTTER, C., TOWNS, K. V., HEWITT, A. W., INGLEHEARN, C. F., CRAIG, J. E. & MACKEY, D. A. 2007. Autosomal dominant optic atrophy: penetrance and expressivity in patients with OPA1 mutations. American journal of ophthalmology, 143, 656-662. e1.

COLLIN, R. W., DEN HOLLANDER, A. I., VAN DER VELDE-VISSER, S. D., BENNICELLI, J., BENNETT, J. & CREMERS, F. P. 2012. Antisense oligonucleotide (AON)-based therapy for Leber congenital amaurosis caused by a frequent mutation in CEP290. Molecular Therapy-Nucleic Acids, 1, e14.

COLLIN, R. W., NIKOPOULOS, K., DONA, M., GILISSEN, C., HOISCHEN, A., BOONSTRA, F. N., POULTER, J. A., KONDO, H., BERGER, W. & TOOMES, C. 2013. ZNF408 is mutated in familial exudative vitreoretinopathy and is crucial for the development of zebrafish retinal vasculature. Proceedings of the National Academy of Sciences, 110, 9856-9861.

COMER, K. A., FALANY, J. L. & FALANY, C. 1993. Cloning and expression of human liver dehydroepiandrosterone sulphotransferase. Biochemical Journal, 289, 233-240.

CONLEY, S. M., STUCK, M. W., WATSON, J. N., ZULLIGER, R., BURNETT, J. L. & NAASH, M. I. 2019. Prph2 initiates outer segment morphogenesis but maturation requires Prph2/Rom1 oligomerization. Human molecular genetics, 28, 459-475.

CONRAD, D. F., BIRD, C., BLACKBURNE, B., LINDSAY, S., MAMANOVA, L., LEE, C., TURNER, D. J. & HURLES, M. E. 2010a. Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nature genetics, 42, 385-391.

CONRAD, D. F., PINTO, D., REDON, R., FEUK, L., GOKCUMEN, O., ZHANG, Y., AERTS, J., ANDREWS, T. D., BARNES, C., CAMPBELL, P., FITZGERALD, T., HU, M., IHM, C. H., KRISTIANSSON, K., MACARTHUR, D. G., MACDONALD, J. R., ONYIAH, I., PANG, A. W., ROBSON, S., STIRRUPS, K., VALSESIA, A., WALTER, K., WEI, J., WELLCOME TRUST CASE CONTROL, C., TYLER-SMITH, C., CARTER, N. P., LEE, C., SCHERER, S. W. & HURLES, M. E. 2010b. Origins and functional impact of copy number variation in the human genome. Nature, 464, 704-12.

CONSORTIUM, G. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature, 491, 56-65.

CONSORTIUM, I. H. G. S. 2004. Finishing the euchromatic sequence of the human genome. Nature, 431, 931-945.

CONSUGAR, M. B., NAVARRO-GOMEZ, D., PLACE, E. M., BUJAKOWSKA, K. M., SOUSA, M. E., FONSECA-KELLY, Z. D., TAUB, D. G., JANESSIAN, M.,

WANG, D. Y. & AU, E. D. 2015. 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, 17, 253-261.

CREMERS, F. P., LEE, W., COLLIN, R. W. & ALLIKMETS, R. 2020. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. Progress in retinal and eye research, 79, 100861.

CREMERS, F. P., VAN DEN HURK, J. A. & DEN HOLLANDER, A. I. 2002. Molecular genetics of Leber congenital amaurosis. Human molecular genetics, 11, 1169-1176.

CRISWICK, V. & SCHEPENS, C. 1969. Familial exudative vitreoretinopathy. American journal of ophthalmology, 68, 578-594.

CUMMINGS, B. B., MARSHALL, J. L., TUKIAINEN, T., LEK, M., DONKERVOORT, S., FOLEY, A. R., BOLDUC, V., WADDELL, L. B., SANDARADURA, S. A. & O'GRADY, G. L. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Science translational medicine, 9, eaal5209.

CURCIO, C. A., SLOAN, K. R., KALINA, R. E. & HENDRICKSON, A. E. 1990. Human photoreceptor topography. Journal of comparative neurology, 292, 497-523.

DA CRUZ, L., COLEY, B. F., DORN, J., MERLINI, F., FILLEY, E., CHRISTOPHER, P., CHEN, F. K., WUYYURU, V., SAHEL, J. & STANGA, P. 2013. The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss. British Journal of Ophthalmology, 97, 632-636.

DA CRUZ, L., DORN, J. D., HUMAYUN, M. S., DAGNELIE, G., HANDA, J., BARALE, P.-O., SAHEL, J.-A., STANGA, P. E., HAFEZI, F. & SAFRAN, A. B. 2016. Five-year safety and performance results from the Argus II retinal prosthesis system clinical trial. Ophthalmology, 123, 2248-2254.

DAGNELIE, G., CHRISTOPHER, P., ARDITI, A., DA CRUZ, L., DUNCAN, J. L., HO, A. C., OLMOS DE KOO, L. C., SAHEL, J. A., STANGA, P. E. & THUMANN, G. 2017. Performance of real-world functional vision tasks by blind subjects improves after implantation with the Argus® II retinal prosthesis system. Clinical & experimental ophthalmology, 45, 152-159.

DAIGER, S. P., BOWNE, S. J. & SULLIVAN, L. S. 2015. Genes and mutations causing autosomal dominant retinitis pigmentosa. Cold Spring Harbor perspectives in medicine, 5, a017129.

DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C. A., BANKS, E., DEPRISTO, M. A., HANDSAKER, R. E., LUNTER, G., MARTH, G. T. & SHERRY, S. T. 2011. The variant call format and VCFtools. Bioinformatics, 27, 2156-2158.

DAVIDSON, A. E., SERGOUNIOTIS, P. I., MACKAY, D. S., WRIGHT, G. A., WASEEM, N. H., MICHAELIDES, M., HOLDER, G. E., ROBSON, A. G., MOORE, A. T. & PLAGNOL, V. 2013. RP1L1 variants are associated with a spectrum of inherited retinal diseases including retinitis pigmentosa and occult macular dystrophy. Human Mutation, 34, 506-514.

DAVIES, V. J., HOLLINS, A. J., PIECHOTA, M. J., YIP, W., DAVIES, J. R., WHITE, K. E., NICOLS, P. P., BOULTON, M. E. & VOTRUBA, M. 2007. Opa1 deficiency in a mouse model of autosomal dominant optic atrophy impairs mitochondrial morphology, optic nerve structure and visual function. Human molecular genetics, 16, 1307-1318.

DAVIS, J. L. 2018. The blunt end: surgical challenges of gene therapy for inherited retinal diseases. American journal of ophthalmology, 196, xxv-xxix.

DE BRUIJN, S. E., FIORENTINO, A., OTTAVIANI, D., FANUCCHI, S., MELO, U. S., CORRAL-SERRANO, J. C., MULDERS, T., GEORGIOU, M., RIVOLTA, C. & PONTIKOS, N. 2020. Structural variants create new topological-associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa. The American Journal of Human Genetics, 107, 802-814.

DE BUSTOS, A., CUADRADO, A. & JOUVE, N. 2016. Sequencing of long stretches of repetitive DNA. Scientific reports, 6, 1-7.

DE COSTER, W., D'HERT, S., SCHULTZ, D. T., CRUTS, M. & VAN BROECKHOVEN, C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics, 34, 2666-2669.

DE FAUW, J., LEDSAM, J. R., ROMERA-PAREDES, B., NIKOLOV, S., TOMASEV, N., BLACKWELL, S., ASKHAM, H., GLOROT, X., O'DONOGHUE, B., VISENTIN, D., VAN DEN DRIESSCHE, G., LAKSHMINARAYANAN, B., MEYER, C., MACKINDER, F., BOUTON, S., AYOUB, K., CHOPRA, R., KING, D., KARTHIKESALINGAM, A., HUGHES, C. O., RAINE, R., HUGHES, J., SIM, D. A., EGAN, C., TUFAIL, A., MONTGOMERY, H., HASSABIS, D., REES, G., BACK, T., KHAW, P. T., SULEYMAN, M., CORNEBISE, J., KEANE, P. A. & RONNEBERGER, O. 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine, 24, 1342-1350.

DE LA COUR, M. D. 1993. Ion transport in the retinal pigment epithelium. A study with double barrelled ion-selective microelectrodes. Acta ophthalmologica. Supplement, 1-32.

DEAN, M. 2005. The genetics of ATP-binding cassette transporters. Methods in enzymology, 400, 409-429.

DEL DOTTO, V., FOGAZZA, M., CARELLI, V., RUGOLO, M. & ZANNA, C. 2018. Eight human OPA1 isoforms, long and short: What are they for? Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1859, 263-269.

DELETTRE, C., GRIFFOIN, J.-M., KAPLAN, J., DOLLFUS, H., LORENZ, B., FAIVRE, L., LENAERS, G., BELENGUER, P. & HAMEL, C. P. 2001. Mutation spectrum and splicing variants in the OPA1 gene. Human genetics, 109, 584-591.

DELETTRE, C., LENAERS, G., GRIFFOIN, J.-M., GIGAREL, N., LORENZO, C., BELENGUER, P., PELLOQUIN, L., GROSGEORGE, J., TURC-CAREL, C. & PERRET, E. 2000. Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. Nature genetics, 26, 207-210.

DELMAGHANI, S. & EL-AMRAOUI, A. 2022. The genetic and phenotypic landscapes of Usher syndrome: from disease mechanisms to a new classification. Human Genetics, 1-27.

DEMAIO, A., MEHROTRA, S., SAMBAMURTI, K. & HUSAIN, S. 2022. The role of the adaptive immune system and T cell dysfunction in neurodegenerative diseases. J Neuroinflammation, 19, 251.

DEN HOLLANDER, A. I., KOENEKOOP, R. K., YZER, S., LOPEZ, I., ARENDS, M. L., VOESENEK, K. E., ZONNEVELD, M. N., STROM, T. M., MEITINGER, T. & BRUNNER, H. G. 2006. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. The American Journal of Human Genetics, 79, 556-561.

DEN HOLLANDER, A. I., ROEPMAN, R., KOENEKOOP, R. K. & CREMERS, F. P. 2008. Leber congenital amaurosis: genes, proteins and disease mechanisms. Progress in retinal and eye research, 27, 391-419.

DENG, W.-T., KOLANDAIVELU, S., DINCULESCU, A., LI, J., ZHU, P., CHIODO, V. A., RAMAMURTHY, V. & HAUSWIRTH, W. W. 2018a. Cone phosphodiesterase-6γ'subunit augments cone PDE6 holoenzyme assembly and stability in a mouse model lacking both rod and cone PDE6 catalytic subunits. Frontiers in molecular neuroscience, 11, 233.

DENG, W. L., GAO, M. L., LEI, X. L., LV, J. N., ZHAO, H., HE, K. W., XIA, X. X., LI, L. Y., CHEN, Y. C., LI, Y. P., PAN, D., XUE, T. & JIN, Z. B. 2018b. Gene Correction Reverses Ciliopathy and Photoreceptor Loss in iPSC-Derived Retinal Organoids from Retinitis Pigmentosa Patients. Stem Cell Reports, 10, 1267-1281.

DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPAKIS, A. A., DEL ANGEL, G., RIVAS, M. A. & HANNA, M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics, 43, 491-498.

DEWEY, F. E., MURRAY, M. F., OVERTON, J. D., HABEGGER, L., LEADER, J. B., FETTEROLF, S. N., O'DUSHLAINE, C., VAN HOUT, C. V., STAPLES, J. & GONZAGA-JAUREGUI, C. 2016. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science, 354, aaf6814.

DOCKERY, A., STEPHENSON, K., KEEGAN, D., WYNNE, N., SILVESTRI, G., HUMPHRIES, P., KENNA, P. F., CARRIGAN, M. & FARRAR, G. J. 2017. Target 5000: target capture sequencing for inherited retinal degenerations. Genes, 8, 304.

DOCKERY, A., WHELAN, L., HUMPHRIES, P. & FARRAR, G. J. 2021. Next-generation sequencing applications for inherited retinal diseases. International Journal of Molecular Sciences, 22, 5684.

DRENSER, K. A. 2016. Wnt signaling pathway in retinal vascularization. Eye and Brain, 8, 141.

DRYJA, T. P., MCGEE, T. L., HAHN, L. B., COWLEY, G. S., OLSSON, J. E., REICHEL, E., SANDBERG, M. A. & BERSON, E. L. 1990a. Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. New England Journal of Medicine, 323, 1302-1307.

DRYJA, T. P., MCGEE, T. L., REICHEL, E., HAHN, L. B., COWLEY, G. S., YANDELL, D. W., SANDBERG, M. A. & BERSON, E. L. 1990b. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. Nature, 343, 364-366.

DUIJKERS, L., VAN DEN BORN, L. I., NEIDHARDT, J., BAX, N. M., PIERRACHE, L. H., KLEVERING, B. J., COLLIN, R. W. & GARANTO, A. 2018. Antisense oligonucleotide-based splicing correction in individuals with leber congenital amaurosis due to compound heterozygosity for the c. 2991+ 1655A> G mutation in CEP290. International journal of molecular sciences, 19, 753.

EBBERT, M. T., JENSEN, T. D., JANSEN-WEST, K., SENS, J. P., REDDY, J. S., RIDGE, P. G., KAUWE, J. S., BELZIL, V., PREGENT, L. & CARRASQUILLO, M. M. 2019. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. Genome biology, 20, 1-23.

EBENEZER, N. D., MICHAELIDES, M., JENKINS, S. A., AUDO, I., WEBSTER, A. R., CHEETHAM, M. E., STOCKMAN, A., MAHER, E. R., AINSWORTH, J. R. & YATES, J. R. 2005. Identification of novel RPGR ORF15 mutations in X-linked progressive cone-rod dystrophy (XLCORD) families. Investigative Ophthalmology & Visual Science, 46, 1891-1898.

EBREY, T. & KOUTALOS, Y. 2001. Vertebrate photoreceptors. Prog Retin Eye Res, 20, 49-94.

EFTHYMIOU, S., MANOLE, A. & HOULDEN, H. 2016. Next generation sequencing in neuromuscular diseases. Current opinion in neurology, 29, 527.

ELLINGFORD, J. M., AHN, J. W., BAGNALL, R. D., BARALLE, D., BARTON, S., CAMPBELL, C., DOWNES, K., ELLARD, S., DUFF-FARRIER, C. & FITZPATRICK, D. R. 2022. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. Genome Medicine, 14, 73.

ELLINGFORD, J. M., BARTON, S., BHASKAR, S., O'SULLIVAN, J., WILLIAMS, S. G., LAMB, J. A., PANDA, B., SERGOUNIOTIS, P. I., GILLESPIE, R. L. & DAIGER, S. P. 2016. Molecular findings from 537 individuals with inherited retinal disease. Journal of medical genetics, 53, 761-767.

ELLINGFORD, J. M., HORN, B., CAMPBELL, C., ARNO, G., BARTON, S., TATE, C., BHASKAR, S., SERGOUNIOTIS, P. I., TAYLOR, R. L. & CARSS, K. J. 2018. 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.

ESTRADA-CUZCANO, A., KOENEKOOP, R. K., SENECHAL, A., DE BAERE, E. B., DE RAVEL, T., BANFI, S., KOHL, S., AYUSO, C., SHARON, D. & HOYNG, C. B. 2012. BBS1 mutations in a wide spectrum of phenotypes ranging from nonsyndromic retinitis pigmentosa to Bardet-Biedl syndrome. Archives of ophthalmology, 130, 1425-1432.

EVANS, K., AL-MAGHTHEH, M., FITZKE, F., MOORE, A., JAY, M., INGLEHEARN, C., ARDEN, G. & BIRD, A. 1995. Bimodal expressivity in dominant retinitis pigmentosa genetically linked to chromosome 19q. British journal of ophthalmology, 79, 841-846.

EVANS, K., FRYER, A., INGLEHEARN, C., DUVALL-YOUNG, J., WHITTAKER, J. L., GREGORY, C. Y., BUTLER, R., EBENEZER, N., HUNT, D. M. & BHATTACHARYA, S. 1994. Genetic linkage of cone–rod retinal dystrophy to chromosome 19q and evidence for segregation distortion. Nature genetics, 6, 210-213.

EVANS, R. J., SCHWARZ, N., NAGEL-WOLFRUM, K., WOLFRUM, U., HARDCASTLE, A. J. & CHEETHAM, M. E. 2010. The retinitis pigmentosa protein RP2 links pericentriolar vesicle transport between the Golgi and the primary cilium. Human molecular genetics, 19, 1358-1367.

FAHIM, A. 2018. Retinitis pigmentosa: recent advances and future directions in diagnosis and management. Current Opinion in Pediatrics, 30, 725-733.

FAHIM, A., DAIGER, S. & WELEBER, R. 2020a. Nonsyndromic Retinitis Pigmentosa Overview Clinical Characteristics of Nonsyndromic Retinitis Pigmentosa Clinical Manifestations of Retinitis Pigmentosa. GeneReviews.

FAHIM, A. T. & DAIGER, S. P. 2016. The role of X-chromosome inactivation in retinal development and disease. Retinal Degenerative Diseases. Springer.

FAHIM, A. T., SULLIVAN, L. S., BOWNE, S. J., JONES, K. D., WHEATON, D. K. H., KHAN, N. W., HECKENLIVELY, J. R., JAYASUNDERA, K. T., BRANHAM, K. H., ANDREWS, C. A., OTHMAN, M. I., KAROUKIS, A. J., BIRCH, D. G. & DAIGER, S. P. 2020b. X-Chromosome Inactivation Is a Biomarker of Clinical Severity in Female Carriers of RPGR-Associated X-Linked Retinitis Pigmentosa. Ophthalmol Retina, 4, 510-520.

FARAG, T. I. & TEEBI, A. S. 1989. High incidence of Bardet Biedl syndrome among the Bedouin. Clinical genetics, 36, 463-464.

FARJO, A., MCDERMOTT, M. & SOONG, H. K. 2009. Corneal anatomy, physiology, and wound healing. Ophthalmology, 44, 203-208.

FARRAR, G., KENNA, P., REDMOND, R., MCWILLIAM, P., BRADLEY, D., HUMPHRIES, M., SHARP, E., INGLEHEARN, C. F., BASHIR, R. & JAY, M. 1990. Autosomal dominant retinitis pigmentosa: absence of the rhodopsin proline→ histidine substitution (codon 23) in pedigrees from Europe. American journal of human genetics, 47, 941.

FARRAR, G. J., KENNA, P., JORDAN, S. A., KUMAR-SINGH, R., HUMPHRIES, M. M., SHARP, E. M., SHEILS, D. M. & HUMPHRIES, P. 1991. A three-base-pair deletion in the peripherin–RDS gene in one form of retinitis pigmentosa. Nature, 354, 478-480.

FEDERICI, G. & SODDU, S. 2020. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. Journal of Experimental & Clinical Cancer Research, 39, 46.

FEENSTRA, H. M., AL-KHUZAEI, S., SHAH, M., BROADGATE, S., SHANKS, M., KAMATH, A., YU, J., JOLLY, J. K., MACLAREN, R. E. & CLOUSTON, P. 2022. Phenotypic and Genetic Characteristics in a Cohort of Patients with Usher Genes. Genes, 13, 1423.

FERRARI, S., DI IORIO, E., BARBARO, V., PONZIN, D., S SORRENTINO, F. & PARMEGGIANI, F. 2011. Retinitis pigmentosa: genes and disease mechanisms. Current genomics, 12, 238-249.

FIEGGEN, K., MILLIGAN, C., ESTERHUIZEN, A. & HENDERSON, B. 2016. Bardet Biedl syndrome in South Africa: a single founder mutation: the new millennium. South African Medical Journal, 106, 72-74.

FLOREA, L., CABA, L. & GORDUZA, E. V. 2021. Bardet–Biedl Syndrome—Multiple Kaleidoscope Images: Insight into Mechanisms of Genotype–Phenotype Correlations. Genes, 12, 1353.

FOOD, U. & ADMINISTRATION, D. 2017. FDA approves novel gene therapy to treat patients with a rare form of inherited vision loss. US Dep Heal Hum Serv.

FORRESTER, J. V., DICK, A. D., MCMENAMIN, P. G., ROBERTS, F. & PEARLMAN, E. 2020. The eye e-book: basic sciences in practice, Elsevier Health Sciences.

FORSYTHE, E. & BEALES, P. L. 2013. Bardet–Biedl syndrome. European journal of human genetics, 21, 8-13.

FRANZE, K., GROSCHE, J., SKATCHKOV, S. N., SCHINKINGER, S., FOJA, C., SCHILD, D., UCKERMANN, O., TRAVIS, K., REICHENBACH, A. & GUCK, J. 2007. Müller cells are living optical fibers in the vertebrate retina. Proceedings of the National Academy of Sciences, 104, 8287-8292.

FRÉSARD, L. & MONTGOMERY, S. B. 2018. Diagnosing rare diseases after the exome. Molecular Case Studies, 4, a003392.

FREUND, C. L., GREGORY-EVANS, C. Y., FURUKAWA, T., PAPAIOANNOU, M., LOOSER, J., PLODER, L., BELLINGHAM, J., NG, D., HERBRICK, J.-A. S. & DUNCAN, A. 1997. Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. Cell, 91, 543-553.

FU, Y. 1995. Phototransduction in Rods and Cones. In: KOLB, H., FERNANDEZ, E. & NELSON, R. (eds.) Webvision: The Organization of the Retina and Visual System. Salt Lake City (UT).

FU, Y. & YAU, K.-W. 2007. Phototransduction in mouse rods and cones. Pflügers Archiv-European Journal of Physiology, 454, 805-819.

FUHRMANN, N., ALAVI, M., BITOUN, P., WOERNLE, S., AUBURGER, G., LEO-KOTTLER, B., YU-WAI-MAN, P., CHINNERY, P. & WISSINGER, B. 2009. Genomic rearrangements in OPA1 are frequent in patients with autosomal dominant optic atrophy. Journal of medical genetics, 46, 136-144.

FUHRMANN, N., SCHIMPF, S., KAMENISCH, Y., LEO-KOTTLER, B., ALEXANDER, C., AUBURGER, G., ZRENNER, E., WISSINGER, B. & ALAVI, M. V. 2010. Solving a 50 year mystery of a missing OPA1 mutation: more insights from the first family diagnosed with autosomal dominant optic atrophy. Molecular neurodegeneration, 5, 1-13.

FURUKAWA, T., MORROW, E. M. & CEPKO, C. L. 1997. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. Cell, 91, 531-541.

FURUKAWA, T., MORROW, E. M., LI, T., DAVIS, F. C. & CEPKO, C. L. 1999. Retinopathy and attenuated circadian entrainment in Crx-deficient mice. Nature genetics, 23, 466-470.

FUSTER-GARCÍA, C., GARCÍA-BOHÓRQUEZ, B., RODRÍGUEZ-MUÑOZ, A., ALLER, E., JAIJO, T., MILLÁN, J. M. & GARCÍA-GARCÍA, G. 2021. Usher syndrome: genetics of a human ciliopathy. International journal of molecular sciences, 22, 6723.

GAJ, T., GERSBACH, C. A. & BARBAS, C. F. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology, 31, 397-405.

GAO, F.-J., WANG, D.-D., CHEN, F., SUN, H.-X., HU, F.-Y., XU, P., LI, J., LIU, W., QI, Y.-H. & LI, W. 2021. Prevalence and genetic–phenotypic characteristics of patients with USH2A mutations in a large cohort of Chinese patients with inherited retinal disease. British Journal of Ophthalmology, 105, 87-92.

GARANTO, A., CHUNG, D. C., DUIJKERS, L., CORRAL-SERRANO, J. C., MESSCHAERT, M., XIAO, R., BENNETT, J., VANDENBERGHE, L. H. & COLLIN, R. W. 2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. Human molecular genetics, 25, 2552-2563.

GARANTO, A., VAN DER VELDE-VISSER, S. D., CREMERS, F. P. & COLLIN, R. W. Antisense oligonucleotide-based splice correction of a deep-intronic mutation in CHM underlying choroideremia. Retinal Degenerative Diseases: Mechanisms and Experimental Therapy, 2018. Springer, 83-89.

GARCÍA BOHÓRQUEZ, B., ALLER, E., RODRÍGUEZ MUÑOZ, A., JAIJO, T., GARCÍA GARCÍA, G. & MILLÁN, J. M. 2021. Updating the genetic landscape of inherited retinal dystrophies. Frontiers in cell and developmental biology, 1332.

GARCIA, J. M., MENDONÇA, L., BRANT, R., ABUD, M., REGATIERI, C. & DINIZ, B. 2015. Stem cell therapy for retinal diseases. World Journal of Stem Cells, 7, 160.

GARCIA-DELGADO, A. B., VALDES-SANCHEZ, L., MORILLO-SANCHEZ, M. J., PONTE-ZUÑIGA, B., DIAZ-CORRALES, F. J. & DE LA CERDA, B. 2021. Dissecting the role of EYS in retinal degeneration: clinical and molecular aspects and its implications for future therapy. Orphanet Journal of Rare Diseases, 16, 222.

GARITA-HERNANDEZ, M., ROUTET, F., GUIBBAL, L., KHABOU, H., TOUALBI, L., RIANCHO, L., REICHMAN, S., DUEBEL, J., SAHEL, J.-A. & GOUREAU, O. 2020. AAV-mediated gene delivery to 3D retinal organoids derived from human induced pluripotent stem cells. International journal of molecular sciences, 21, 994.

GEORGIOU, M., ALI, N., YANG, E., GREWAL, P. S., ROTSOS, T., PONTIKOS, N., ROBSON, A. G. & MICHAELIDES, M. 2021. Extending the phenotypic spectrum of PRPF8, PRPH2, RP1 and RPGR, and the genotypic spectrum of early-onset severe retinal dystrophy. Orphanet Journal of Rare Diseases, 16, 128.

GIBBS, R. A., BELMONT, J. W., HARDENBOL, P., WILLIS, T. D., YU, F., YANG, H., CH'ANG, L.-Y., HUANG, W., LIU, B., SHEN, Y., TAM, P. K.-H., TSUI, L.-C., WAYE, M. M. Y., WONG, J. T.-F., ZENG, C., ZHANG, Q., CHEE, M. S., GALVER, L. M., KRUGLYAK, S., MURRAY, S. S., OLIPHANT, A. R., MONTPETIT, A., HUDSON, T. J., CHAGNON, F., FERRETTI, V., LEBOEUF, M., PHILLIPS, M. S., VERNER, A., KWOK, P.-Y., DUAN, S., LIND, D. L., MILLER, R. D., RICE, J. P., SACCONE, N. L., TAILLON-MILLER, P., XIAO, M., NAKAMURA, Y., SEKINE, A., SORIMACHI, K., TANAKA, T., TANAKA, Y., TSUNODA, T., YOSHINO, E., BENTLEY, D. R., DELOUKAS, P., HUNT, S., POWELL, D., ALTSHULER, D., GABRIEL, S. B., ZHANG, H., ZENG, C., MATSUDA, I., FUKUSHIMA, Y., MACER, D. R., SUDA, E., ROTIMI, C. N., ADEBAMOWO, C. A., ANIAGWU, T., MARSHALL, P. A., MATTHEW, O., NKWODIMMAH, C., ROYAL, C. D. M., LEPPERT, M. F., DIXON, M., STEIN, L. D., CUNNINGHAM, F., KANANI, A., THORISSON, G. A., CHAKRAVARTI, A., CHEN, P. E., CUTLER, D. J., KASHUK, C. S., DONNELLY, P., MARCHINI, J., MCVEAN, G. A. T., MYERS, S. R., CARDON, L. R., ABECASIS, G. R., MORRIS, A., WEIR, B. S., MULLIKIN, J. C., SHERRY, S. T., FEOLO, M., ALTSHULER, D., DALY, M. J., SCHAFFNER, S. F., QIU, R., KENT, A., DUNSTON, G. M., KATO, K., NIIKAWA, N., KNOPPERS, B. M., FOSTER, M. W., CLAYTON, E. W., WANG, V. O., WATKIN, J., GIBBS, R. A., BELMONT, J. W., SODERGREN, E., WEINSTOCK, G. M., et al. 2003. The International HapMap Project. Nature, 426, 789-796.

GILISSEN, C., HOISCHEN, A., BRUNNER, H. G. & VELTMAN, J. A. 2012. Disease gene identification strategies for exome sequencing. Eur J Hum Genet, 20, 490-7.

GILL, J. S., GEORGIOU, M., KALITZEOS, A., MOORE, A. T. & MICHAELIDES, M. 2019. Progressive cone and cone-rod dystrophies: clinical features, molecular genetics and prospects for therapy. British Journal of Ophthalmology, 103, 711-720.

GILMOUR, D. 2015. Familial exudative vitreoretinopathy and related retinopathies. Eye, 29, 1-14.

GILPATRICK, T., LEE, I., GRAHAM, J. E., RAIMONDEAU, E., BOWEN, R., HERON, A., DOWNS, B., SUKUMAR, S., SEDLAZECK, F. J. & TIMP, W. 2020. Targeted nanopore sequencing with Cas9-guided adapter ligation. Nature biotechnology, 38, 433-438.

GOLDMAN, D. 2014. Müller glial cell reprogramming and retina regeneration. Nature Reviews Neuroscience, 15, 431-442.

GONZALEZ-FERNANDEZ, F. & GHOSH, D. 2008. Focus on molecules: interphotoreceptor retinoid-binding protein (IRBP). Experimental eye research, 86, 169-170.

GOPALAKRISHNA, K. N., BOYD, K. & ARTEMYEV, N. O. 2017. Mechanisms of mutant PDE6 proteins underlying retinal diseases. Cellular signalling, 37, 74-80.

GRAEMIGER, R. A., NIEMEYER, G., SCHNEEBERGER, S. A. & MESSMER, E. P. 1995. Wagner vitreoretinal degeneration: follow-up of the original pedigree. Ophthalmology, 102, 1830-1839.

GRIFFITH, J. F., DEBENEDICTIS, M. J. & TRABOULSI, E. I. 2018. A novel dominant CRX mutation causes adult-onset macular dystrophy. Ophthalmic genetics, 39, 120-124.

GRILLET, N., SCHWANDER, M., HILDEBRAND, M. S., SCZANIECKA, A., KOLATKAR, A., VELASCO, J., WEBSTER, J. A., KAHRIZI, K., NAJMABADI, H. & KIMBERLING, W. J. 2009. Mutations in LOXHD1, an evolutionarily conserved stereociliary protein, disrupt hair cell function in mice and cause progressive hearing loss in humans. The American Journal of Human Genetics, 85, 328-337.

GROUP, V. S. 2002. Safety of intravitreous fomivirsen for treatment of cytomegalovirus retinitis in patients with AIDS. American Journal of Ophthalmology, 133, 484-498.

GRÜNERT, U. 2009. Retinal Bipolar Cells. In: BINDER, M. D., HIROKAWA, N. & WINDHORST, U. (eds.) Encyclopedia of Neuroscience. Berlin, Heidelberg: Springer Berlin Heidelberg.

GUIDUGLI, L., MEHINE, M., TUUPANEN, S., KÄMPJÄRVI, K., WELLS, K., KÄNSÄKOSKI, J., VALORI, M., SAARINEN, I., MUONA, M. & SANKILA, E.-M. K. 2019. Characterization of copy number variants (CNVs) identified by genetic testing of inherited retinal disorders. Investigative Ophthalmology & Visual Science, 60, 2944-2944.

GUPTA, N., D'ACIERNO, M., ZONA, E., CAPASSO, G. & ZACCHIA, M. Bardet–Biedl syndrome: The pleiotropic role of the chaperonin-like BBS6, 10, and 12 proteins. American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 2022. Wiley Online Library, 9-19.

HADALIN, V., SUSTAR, M., VOLK, M., MAVER, A., SAJOVIC, J., JARC-VIDMAR, M., PETERLIN, B., HAWLINA, M. & FAKIN, A. 2021. Cone Dystrophy Associated with a Novel Variant in the Terminal Codon of the RPGR-ORF15. Genes (Basel), 12.

HAM, M., HAN, J., OSANN, K., SMITH, M. & KIMONIS, V. 2019. Meta-analysis of genotype-phenotype analysis of OPA1 mutations in autosomal dominant optic atrophy. Mitochondrion, 46, 262-269.

HAMAMY, H. 2012. Consanguineous marriages. Journal of community genetics, 3, 185-192.

HAMEL, C. 2006. Retinitis pigmentosa. Orphanet journal of rare diseases, 1, 1-12.

HAMEL, C. P. 2007. Cone rod dystrophies. Orphanet J Rare Dis, 2, 7.

HAN, J., LI, Y., YOU, Y., FAN, K. & LEI, B. 2022. Autosomal dominant optic atrophy caused by six novel pathogenic OPA1 variants and genotype–phenotype correlation analysis. BMC ophthalmology, 22, 1-9.

HARTONG, D. T., BERSON, E. L. & DRYJA, T. P. 2006. Retinitis pigmentosa. The Lancet, 368, 1795-1809.

HAYASHI, T., MIZOBUCHI, K., KAMEYA, S., YOSHITAKE, K., IWATA, T. & NAKANO, T. 2021. A new PDE6A missense variant p. Arg544Gln in rod–cone dystrophy. Documenta Ophthalmologica, 143, 107-114.

HE, T., TAO, B., YI, C., ZHANG, C., ZHANG, P., SHAO, W., LI, Y., CHEN, Z., LU, L. & JIA, H. 2022. 27-Hydroxycholesterol promotes metastasis by SULT2A1-dependent alteration in hepatocellular carcinoma. Cancer science, 113, 2575.

HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics, 107, 1-8.

HEAVNER, W. & PEVNY, L. 2012. Eye development and retinogenesis. Cold Spring Harb Perspect Biol, 4.

HEDIGER, M. A., ROMERO, M. F., PENG, J.-B., ROLFS, A., TAKANAGA, H. & BRUFORD, E. A. 2004. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. Pflügers Archiv, 447, 465-468.

HENDERSON, R. H., LI, Z., ABD EL AZIZ, M. M., MACKAY, D. S., ELJININI, M. A., ZEIDAN, M., MOORE, A. T., BHATTACHARYA, S. S. & WEBSTER, A. R. 2010. Biallelic mutation of protocadherin-21 (PCDH21) causes retinal degeneration in humans. Molecular vision, 16, 46.

HENIKOFF, S. & HENIKOFF, J. G. 1992. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences, 89, 10915-10919.

HILDEBRAND, G. D. & FIELDER, A. R. 2011. Anatomy and physiology of the retina. Pediatric retina. Springer.

HITTI-MALIN, R. J., DHAENENS, C.-M., PANNEMAN, D. M., CORRADI, Z., KHAN, M., DEN HOLLANDER, A. I., FARRAR, G. J., GILISSEN, C., HOISCHEN, A. & VAN DE VORST, M. 2022. Using single molecule Molecular Inversion Probes as a cost-effective, high-throughput sequencing approach to target all genes and loci associated with macular diseases. medRxiv.

HJORTSHØJ, T. D., GRØNSKOV, K., BRØNDUM-NIELSEN, K. & ROSENBERG, T. 2009. A novel founder BBS1 mutation explains a unique high prevalence of Bardet–Biedl syndrome in the Faroe Islands. British Journal of Ophthalmology, 93, 409-413.

HOLDER, I. T., WAGNER, S., XIONG, P., SINN, M., FRICKEY, T., MEYER, A. & HARTIG, J. S. 2015. Intrastrand triplex DNA repeats in bacteria: a source of genomic instability. Nucleic acids research, 43, 10126-10142.

HOLLEY, R. W., APGAR, J., EVERETT, G. A., MADISON, J. T., MARQUISEE, M., MERRILL, S. H., PENSWICK, J. R. & ZAMIR, A. 1965. Structure of a ribonucleic acid. Science, 147, 1462-1465.

HOLLEY, R. W., APGAR, J., MERRILL, S. H. & ZUBKOFF, P. L. 1961. Nucleotide and oligonucleotide compositions of the alanine-, valine-, and tyrosine-acceptor "soluble" ribonucleic acids of yeast. Journal of the American Chemical Society, 83, 4861-4862.

HOLLINGSWORTH, T. & GROSS, A. K. 2013. The severe autosomal dominant retinitis pigmentosa rhodopsin mutant Ter349Glu mislocalizes and induces rapid rod cell death. Journal of Biological Chemistry, 288, 29047-29055.

HOLTAN, J. P., SELMER, K. K., HEIMDAL, K. R. & BRAGADÓTTIR, R. 2020. Inherited retinal disease in Norway–a characterization of current clinical and genetic knowledge. Acta Ophthalmologica, 98, 286-295.

HONG, D.-H., PAWLYK, B., SOKOLOV, M., STRISSEL, K. J., YANG, J., TULLOCH, B., WRIGHT, A. F., ARSHAVSKY, V. Y. & LI, T. 2003. RPGR isoforms in photoreceptor connecting cilia and the transitional zone of motile cilia. Investigative Ophthalmology & Visual Science, 44, 2413-2421.

HOPPINS, S., HORNER, J., SONG, C., MCCAFFERY, J. M. & NUNNARI, J. 2009. Mitochondrial outer and inner membrane fusion requires a modified carrier protein. Journal of Cell Biology, 184, 569-581.

HU, T., CHITNIS, N., MONOS, D. & DINH, A. 2021. Next-generation sequencing technologies: An overview. Human Immunology, 82, 801-811.

HUANG, C.-H., YANG, C.-M., YANG, C.-H., HOU, Y.-C. & CHEN, T.-C. 2021. Leber's Congenital Amaurosis: Current Concepts of Genotype-Phenotype Correlations. Genes, 12, 1261.

HUANG, X., YUAN, L., XU, H., ZHENG, W., CAO, Y., YI, J., GUO, Y., YANG, Z., LI, Y. & DENG, H. 2018. Identification of a novel mutation in the ABCA4 gene in a Chinese family with retinitis pigmentosa using exome sequencing. Bioscience Reports, 38.

HUANG, X.-F., WU, J., LV, J.-N., ZHANG, X. & JIN, Z.-B. 2015. Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing. Genetics in Medicine, 17, 307-311.

HUDDLESTON, J. & EICHLER, E. E. 2016. An incomplete understanding of human genetic variation. Genetics, 202, 1251-1254.

HULL, S., ARNO, G., PLAGNOL, V., CHAMNEY, S., RUSSELL-EGGITT, I., THOMPSON, D., RAMSDEN, S. C., BLACK, G. C., ROBSON, A. & HOLDER, G. E. 2014. The phenotypic variability of retinal dystrophies associated with mutations in CRX, with report of a novel macular dystrophy phenotype. Investigative ophthalmology & visual science, 55, 6934-6944.

HUNG, S. S., CHRYSOSTOMOU, V., LI, F., LIM, J. K., WANG, J.-H., POWELL, J. E., TU, L., DANISZEWSKI, M., LO, C. & WONG, R. C. 2016. AAV-mediated CRISPR/Cas gene editing of retinal cells in vivo. Investigative ophthalmology & visual science, 57, 3470-3476.

HURD, T., ZHOU, W., JENKINS, P., LIU, C.-J., SWAROOP, A., KHANNA, H., MARTENS, J., HILDEBRANDT, F. & MARGOLIS, B. 2010. The retinitis pigmentosa protein RP2 interacts with polycystin 2 and regulates cilia-mediated vertebrate development. Human molecular genetics, 19, 4330-4344.

HURLEY, J. B. 2021. Retina Metabolism and Metabolism in the Pigmented Epithelium: A Busy Intersection. Annu Rev Vis Sci, 7, 665-692.

HUSSEY, K. A., HADYNIAK, S. E. & JOHNSTON JR, R. J. 2022. Patterning and Development of Photoreceptors in the Human Retina. Frontiers in Cell and Developmental Biology, 795.

IANCU, I.-F., AVILA-FERNANDEZ, A., ARTECHE, A., TRUJILLO-TIEBAS, M. J., RIVEIRO-ALVAREZ, R., ALMOGUERA, B., MARTIN-MERIDA, I., DEL POZO-VALERO, M., PEREA-ROMERO, I., CORTON, M., MINGUEZ, P. & AYUSO, C. 2021. Prioritizing variants of uncertain significance for reclassification using a rule-based algorithm in inherited retinal dystrophies. npj Genomic Medicine, 6, 18.

IANNACCONE, A., BREUER, D., WANG, X., KUO, S., NORMANDO, E., FILIPPOVA, E., BALDI, A., HIRIYANNA, S., MACDONALD, C. & BALDI, F. 2003. Clinical and immunohistochemical evidence for an X linked retinitis pigmentosa

syndrome with recurrent infections and hearing loss in association with an RPGR mutation. Journal of medical genetics, 40, e118-e118.

IBRAHIM, M., ALARCON-MARTINEZ, T., LOPEZ, I., FAJARDO, N., CHIANG, J. & KOENEKOOP, R. 2018. A complete, homozygous CRX deletion causing nullizygosity is a new genetic mechanism for Leber congenital amaurosis. Scientific reports, 8, 1-6.

INGLEHEARN, C. F., YAHYA, S., SMITH, C. E. L., POULTER, J. A., ALI, M., TOOMES, C., ELLINGFORD, J., BLACK, G. C., ARNO, G. & WEBSTER, A. R. 2022. Reply. Ophthalmology.

ITO, S. I., ONISHI, A. & TAKAHASHI, M. 2017. Chemically-induced photoreceptor degeneration and protection in mouse iPSC-derived three-dimensional retinal organoids. Stem Cell Res, 24, 94-101.

JAAKSON, K., ZERNANT, J., KÜLM, M., HUTCHINSON, A., TONISSON, N., GLAVAČ, D., RAVNIK-GLAVAČ, M., HAWLINA, M., MELTZER, M. & CARUSO, R. 2003. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Human mutation, 22, 395-403.

JACOBSON, S. G., KEMP, C. M., SUNG, C.-H. & NATHANS, J. 1991. Retinal function and rhodopsin levels in autosomal dominant retinitis pigmentosa with rhodopsin mutations. American journal of ophthalmology, 112, 256-271.

JAGANATHAN, K., PANAGIOTOPOULOU, S. K., MCRAE, J. F., DARBANDI, S. F., KNOWLES, D., LI, Y. I., KOSMICKI, J. A., ARBELAEZ, J., CUI, W. & SCHWARTZ, G. B. 2019. Predicting splicing from primary sequence with deep learning. Cell, 176, 535-548. e24.

JAIN, M., OLSEN, H. E., PATEN, B. & AKESON, M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome biology, 17, 1-11.

JANER, A., PRUDENT, J., PAUPE, V., FAHIMINIYA, S., MAJEWSKI, J., SGARIOTO, N., DES ROSIERS, C., FOREST, A., LIN, Z. Y. & GINGRAS, A. C. 2016. SLC 25A46 is required for mitochondrial lipid homeostasis and cristae maintenance and is responsible for Leigh syndrome. EMBO molecular medicine, 8, 1019-1038.

JAYASUNDERA, T., BRANHAM, K. E., OTHMAN, M., RHOADES, W. R., KAROUKIS, A. J., KHANNA, H., SWAROOP, A. & HECKENLIVELY, J. R. 2010. RP2 phenotype and pathogenetic correlations in X-linked retinitis pigmentosa. Archives of ophthalmology, 128, 915-923.

JONES, K. D., WHEATON, D. K., BOWNE, S. J., SULLIVAN, L. S., BIRCH, D. G., CHEN, R. & DAIGER, S. P. 2017. Next-generation sequencing to solve complex

inherited retinal dystrophy: A case series of multiple genes contributing to disease in extended families. Molecular vision, 23, 470.

KADALAYIL, L., RAFIQ, S., ROSE-ZERILLI, M. J., PENGELLY, R. J., PARKER, H., OSCIER, D., STREFFORD, J. C., TAPPER, W. J., GIBSON, J. & ENNIS, S. 2015. Exome sequence read depth methods for identifying copy number changes. Briefings in bioinformatics, 16, 380-392.

KANNABIRAN, C., PARAMESWARAPPA, D. & JALALI, S. 2022. Genetics of Inherited Retinal Diseases in Understudied Populations. Frontiers in Genetics, 321.

KAUR, C., FOULDS, W. & LING, E. 2008. Blood–retinal barrier in hypoxic ischaemic conditions: basic concepts, clinical features and management. Progress in retinal and eye research, 27, 622-647.

KELLNER, U., HELBIG, H. & FOERSTER, M. H. 1996. [Phenocopies of hereditary retinal degenerations]. Ophthalmologe, 93, 680-7.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. Genome research, 12, 996-1006.

KERSTEN, E., GEERLINGS, M. J., PAUPER, M., COROMINAS, J., BAKKER, B., ALTAY, L., FAUSER, S., DE JONG, E. K., HOYNG, C. B. & DEN HOLLANDER, A. I. 2018. Genetic screening for macular dystrophies in patients clinically diagnosed with dry age-related macular degeneration. Clinical genetics, 94, 569-574.

KHAN, A. A., WARYAH, Y. M., IQBAL, M., BAIG, H. M. A., RAFIQUE, M. & WARYAH, A. M. 2021. P. arg102ser is a common Pde6? mutation causing autosomal recessive retinitis pigmentosa in Pakistani families. Journal of the Pakistan Medical Association, 71, 816-821.

KHAN, A. O., NERI, P. & AL TENEIJI, A. M. 2019a. A recurrent arcuate retinopathy in familial cone-rod dystrophy secondary to heterozygous CRX deletion. Ophthalmic Genetics, 40, 493-499.

KHAN, K., LOGAN, C. V., MCKIBBIN, M., SHERIDAN, E., ELCIOGLU, N. H., YENICE, O., PARRY, D. A., FERNANDEZ-FUENTES, N., ABDELHAMED, Z. I. & AL-MASKARI, A. 2012. Next generation sequencing identifies mutations in Atonal homolog 7 (ATOH7) in families with global eye developmental defects. Human molecular genetics, 21, 776-783.

KHAN, K. N., EL-ASRAG, M. E., KU, C. A., HOLDER, G. E., MCKIBBIN, M., ARNO, G., POULTER, J. A., CARSS, K., BOMMIREDDY, T., BAGHERI, S., BAKALL, B., SCHOLL, H. P., RAYMOND, F. L., TOOMES, C., INGLEHEARN, C. F., PENNESI, M. E., MOORE, A. T., MICHAELIDES, M., WEBSTER, A. R., ALI, M., DISEASES, F. N. B.-R. & CONSORTIUM, U. I. R. D. 2017. Specific Alleles of

CLN7/MFSD8, a Protein That Localizes to Photoreceptor Synaptic Terminals, Cause a Spectrum of Nonsyndromic Retinal Dystrophy. Investigative Ophthalmology & Visual Science, 58, 2906-2914.

KHAN, M., CORNELIS, S. S., KHAN, M. I., ELMELIK, D., MANDERS, E., BAKKER, S., DERKS, R., NEVELING, K., VAN DE VORST, M. & GILISSEN, C. 2019b. Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. Human Mutation, 40, 1749-1759.

KHAN, M., CORNELIS, S. S., POZO-VALERO, M. D., WHELAN, L., RUNHART, E. H., MISHRA, K., BULTS, F., ALSWAITI, Y., ALTALBISHI, A. & DE BAERE, E. 2020. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. Genetics in Medicine, 22, 1235-1246.

KHAN, S., MUHAMMAD, N., KHAN, M., KAMAL, A., REHMAN, Z. & KHAN, S. 2016. Genetics of human Bardet–Biedl syndrome, an updates. Clinical genetics, 90, 3-15.

KHANNA, H., HURD, T. W., LILLO, C., SHU, X., PARAPURAM, S. K., HE, S., AKIMOTO, M., WRIGHT, A. F., MARGOLIS, B. & WILLIAMS, D. S. 2005. RPGR-ORF15, which is mutated in retinitis pigmentosa, associates with SMC1, SMC3, and microtubule transport proteins. Journal of Biological Chemistry, 280, 33580-33587.

KIELECZAWA, J. 2006. Fundamentals of sequencing of difficult templates—an overview. Journal of biomolecular techniques: JBT, 17, 207.

KIM, Y. N., SONG, J. S., OH, S. H., KIM, Y. J., YOON, Y. H., SEO, E.-J., SEOL, C. A., LEE, S.-M., CHOI, J.-M. & SEO, G. H. 2020. Clinical characteristics and disease progression of retinitis pigmentosa associated with PDE6B mutations in Korean patients. Scientific Reports, 10, 1-12.

KIMURA, A., SINGH, D., WAWROUSEK, E. F., KIKUCHI, M., NAKAMURA, M. & SHINOHARA, T. 2000. Both PCE-1/RX and OTX/CRX interactions are necessary for photoreceptor-specific gene expression. Journal of Biological Chemistry, 275, 1152-1160.

KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics, 46, 310-315.

KISER, K., WEBB-JONES, K. D., BOWNE, S. J., SULLIVAN, L. S., DAIGER, S. P. & BIRCH, D. G. 2019. Time course of disease progression of PRPF31-mediated retinitis pigmentosa. American journal of ophthalmology, 200, 76-84.

KLOECKENER-GRUISSEM, B., BARTHOLDI, D., ABDOU, M.-T., ZIMMERMANN, D. R. & BERGER, W. 2006. Identification of the genetic defect in the original Wagner syndrome family. Molecular Vision, 350-355.

KLOECKENER-GRUISSEM, B., NEIDHARDT, J., MAGYAR, I., PLAUCHU, H., ZECH, J.-C., MORLÉ, L., PALMER-SMITH, S. M., MACDONALD, M. J., NAS, V. & FRY, A. E. 2013. Novel VCAN mutations and evidence for unbalanced alternative splicing in the pathogenesis of Wagner syndrome. European Journal of Human Genetics, 21, 352-356.

KOENEKOOP, R. K. 2004. An overview of Leber congenital amaurosis: a model to understand human retinal development. Survey of ophthalmology, 49, 379-398.

KOEPPEN, K., REUTER, P., LADEWIG, T., KOHL, S., BAUMANN, B., JACOBSON, S. G., PLOMP, A. S., HAMEL, C. P., JANECKE, A. R. & WISSINGER, B. 2010. Dissecting the pathogenic mechanisms of mutations in the pore region of the human cone photoreceptor cyclic nucleotide-gated channel. Human mutation, 31, 830-839.

KOHL, S., BAUMANN, B., ROSENBERG, T., KELLNER, U., LORENZ, B., VADALA, M., JACOBSON, S. G. & WISSINGER, B. 2002. Mutations in the cone photoreceptor G-protein α-subunit gene GNAT2 in patients with achromatopsia. The American Journal of Human Genetics, 71, 422-425.

KONDKAR, A. A. & ABU-AMERO, K. K. 2019. Leber congenital amaurosis: Current genetic basis, scope for genetic testing and personalized medicine. Experimental Eye Research, 189, 107834.

KONG, A., GUDBJARTSSON, D. F., SAINZ, J., JONSDOTTIR, G. M., GUDJONSSON, S. A., RICHARDSSON, B., SIGURDARDOTTIR, S., BARNARD, J., HALLBECK, B. & MASSON, G. 2002. A high-resolution recombination map of the human genome. Nature genetics, 31, 241-247.

KÖNIG, I. R., FUCHS, O., HANSEN, G., VON MUTIUS, E. & KOPP, M. V. 2017. What is precision medicine? European respiratory journal, 50.

KUEHLEWEIN, L., ZOBOR, D., STINGL, K., KEMPF, M., NASSER, F., BERND, A., BISKUP, S., CREMERS, F. P., KHAN, M. I. & MAZZOLA, P. 2021. Clinical phenotype of pde6b-associated retinitis pigmentosa. International journal of molecular sciences, 22, 2374.

KUMAR, P., HENIKOFF, S. & NG, P. C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols, 4, 1073-1081.

KUMARAN, N., MOORE, A. T., WELEBER, R. G. & MICHAELIDES, M. 2017. Leber congenital amaurosis/early-onset severe retinal dystrophy: clinical features, molecular genetics and therapeutic interventions. British journal of ophthalmology, 101, 1147-1154.

KUROSAKI, T. & MAQUAT, L. E. 2016. Nonsense-mediated mRNA decay in humans at a glance. Journal of cell science, 129, 461-467.

KWON, E. M., HOLT, S. K., FU, R., KOLB, S., WILLIAMS, G., STANFORD, J. L. & OSTRANDER, E. A. 2012. Androgen metabolism and JAK/STAT pathway genes and prostate cancer risk. Cancer epidemiology, 36, 347-353.

LAMB, T. D. 2020. Evolution of the genes mediating phototransduction in rod and cone photoreceptors. Progress in retinal and eye research, 76, 100823.

LAMB, T. D. & PUGH, E. N., JR. 2006. Phototransduction, dark adaptation, and rhodopsin regeneration the proctor lecture. Invest Ophthalmol Vis Sci, 47, 5137-52.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J.-F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. Nature, 409, 860-921.

LANDRUM, M. J., LEE, J. M., RILEY, G. R., JANG, W., RUBINSTEIN, W. S., CHURCH, D. M. & MAGLOTT, D. R. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research, 42, D980-D985.

LAVAIL, M. M., YASUMURA, D., MATTHES, M. T., YANG, H., HAUSWIRTH, W. W., DENG, W.-T. & VOLLRATH, D. Gene therapy for MERTK-associated retinal degenerations. Retinal Degenerative Diseases: Mechanisms and Experimental Therapy, 2016. Springer, 487-493.

LE ROUX, B., LENAERS, G., ZANLONGHI, X., AMATI-BONNEAU, P., CHABRUN, F., FOULONNEAU, T., CAIGNARD, A., LERUEZ, S., GOHIER, P. & PROCACCIO, V. 2019. OPA1: 516 unique variants and 831 patients registered in an updated centralized Variome database. Orphanet journal of rare diseases, 14, 1-9.

LEBER, T. 1869. Ueber Retinitis pigmentosa und angeborene Amaurose. Archiv für Ophthalmologie, 15, 1-25.

LEDFORD, H. 2020. CRISPR treatment inserted directly into the body for first time. Nature, 579, 185-186.

LEE, J. A., CARVALHO, C. M. & LUPSKI, J. R. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. cell, 131, 1235-1247.

LEE, S. C. S., WELTZIEN, F., MADIGAN, M. C., MARTIN, P. R. & GRÜNERT, U. 2016. Identification of AII amacrine, displaced amacrine, and bistratified ganglion cell types in human retina with antibodies against calretinin. Journal of Comparative Neurology, 524, 39-53.

LEKHOLM, E., PERLAND, E., ERIKSSON, M. M., HELLSTEN, S. V., LINDBERG, F. A., ROSTAMI, J. & FREDRIKSSON, R. 2017. Putative membrane-bound transporters MFSD14A and MFSD14B are neuronal and affected by nutrient availability. Frontiers in Molecular Neuroscience, 10, 11.

LELIEVELD, S. H., SPIELMANN, M., MUNDLOS, S., VELTMAN, J. A. & GILISSEN, C. 2015. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. Human mutation, 36, 815-822.

LENAERS, G., HAMEL, C., DELETTRE, C., AMATI-BONNEAU, P., PROCACCIO, V., BONNEAU, D., REYNIER, P. & MILEA, D. 2012. Dominant optic atrophy. Orphanet journal of rare diseases, 7, 1-12.

LENAHAN, C., SANGHAVI, R., HUANG, L. & ZHANG, J. H. 2020. Rhodopsin: a potential biomarker for neurodegenerative diseases. Frontiers in Neuroscience, 14, 326.

LENASSI, E., VINCENT, A., LI, Z., SAIHAN, Z., COFFEY, A. J., STEELE-STALLARD, H. B., MOORE, A. T., STEEL, K. P., LUXON, L. M. & HÉON, E. 2015. A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants. European Journal of Human Genetics, 23, 1318-1327.

LENIS, T. L., HU, J., NG, S. Y., JIANG, Z., SARFARE, S., LLOYD, M. B., ESPOSITO, N. J., SAMUEL, W., JAWORSKI, C., BOK, D., FINNEMANN, S. C., RADEKE, M. J., REDMOND, T. M., TRAVIS, G. H. & RADU, R. A. 2018. Expression of ABCA4 in the retinal pigment epithelium and its implications for

Stargardt macular degeneration. Proceedings of the National Academy of Sciences, 115, E11120-E11127.

LEVY, S. E. & MYERS, R. M. 2016. Advancements in next-generation sequencing. Annu Rev Genomics Hum Genet, 17, 95-115.

LI, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 34, 3094-3100.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25, 1754-1760.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078-2079.

LI, J., TANG, J., FENG, Y., XU, M., CHEN, R., ZOU, X., SUI, R., CHANG, E. Y., LEWIS, R. A. & ZHANG, V. W. 2016a. Improved diagnosis of inherited retinal dystrophies by high-fidelity PCR of ORF15 followed by next-generation sequencing. The Journal of Molecular Diagnostics, 18, 817-824.

LI, J.-K., FEI, P., LI, Y., HUANG, Q.-J., ZHANG, Q., ZHANG, X., RAO, Y.-Q., LI, J. & ZHAO, P. 2016b. Identification of novel KIF11 mutations in patients with familial exudative vitreoretinopathy and a phenotypic analysis. Scientific Reports, 6, 1-7.

LI, Q., MADDEN, J. A., LIN, J., SHI, J., ROSEN, S. M., SCHMITZ-ABE, K. & AGRAWAL, P. B. 2021. Reanalysis of Exome Data Identifies Novel SLC25A46 Variants Associated with Leigh Syndrome. Journal of Personalized Medicine, 11, 1277.

LI, S., HUANG, L., XIAO, X., JIA, X., GUO, X. & ZHANG, Q. 2014. Identification of CNGA3 mutations in 46 families: common cause of achromatopsia and cone-rod dystrophies in Chinese patients. JAMA ophthalmology, 132, 1076-1083.

LI, S., LI, M., SUN, L., ZHAO, X., ZHANG, T., HUANG, L., HUANG, S., CHEN, C., WANG, Z. & DING, X. 2020. Identification of Novel Copy Number Variations of VCAN Gene in Three Chinese Families with Wagner Disease. Genes, 11, 992.

LIANG, F.-Q., ANAND, V., MAGUIRE, A. M. & BENNETT, J. 2001. Intraocular delivery of recombinant virus. Vision research protocols, 125-139.

LIAO, W.-W., ASRI, M., EBLER, J., DOERR, D., HAUKNESS, M., HICKEY, G., LU, S., LUCAS, J. K., MONLONG, J., ABEL, H. J., BUONAIUTO, S., CHANG, X. H., CHENG, H., CHU, J., COLONNA, V., EIZENGA, J. M., FENG, X., FISCHER, C., FULTON, R. S., GARG, S., GROZA, C., GUARRACINO, A., HARVEY, W. T., HEUMOS, S., HOWE, K., JAIN, M., LU, T.-Y., MARKELLO, C., MARTIN, F. J., MITCHELL, M. W., MUNSON, K. M., MWANIKI, M. N., NOVAK, A. M., OLSEN, H. E., PESOUT, T., PORUBSKY, D., PRINS, P., SIBBESEN, J. A., SIRÉN, J.,

TOMLINSON, C., VILLANI, F., VOLLGER, M. R., ANTONACCI-FULTON, L. L., BAID, G., BAKER, C. A., BELYAEVA, A., BILLIS, K., CARROLL, A., CHANG, P.-C., CODY, S., COOK, D. E., COOK-DEEGAN, R. M., CORNEJO, O. E., DIEKHANS, M., EBERT, P., FAIRLEY, S., FEDRIGO, O., FELSENFELD, A. L., FORMENTI, G., FRANKISH, A., GAO, Y., GARRISON, N. A., GIRON, C. G., GREEN, R. E., HAGGERTY, L., HOEKZEMA, K., HOURLIER, T., JI, H. P., KENNY, E. E., KOENIG, B. A., KOLESNIKOV, A., KORBEL, J. O., KORDOSKY, J., KOREN, S., LEE, H., LEWIS, A. P., MAGALHÃES, H., MARCO-SOLA, S., MARIJON, P., MCCARTNEY, A., MCDANIEL, J., MOUNTCASTLE, J., NATTESTAD, M., NURK, S., OLSON, N. D., POPEJOY, A. B., PUIU, D., RAUTIAINEN, M., REGIER, A. A., RHIE, A., SACCO, S., SANDERS, A. D., SCHNEIDER, V. A., SCHULTZ, B. I., SHAFIN, K., SMITH, M. W., SOFIA, H. J., ABOU TAYOUN, A. N., THIBAUD-NISSEN, F., TRICOMI, F. F., et al. 2023. A draft human pangenome reference. Nature, 617, 312-324.

LIPWORTH, S., PICKFORD, H., SANDERSON, N., CHAU, K. K., KAVANAGH, J., BARKER, L., VAUGHAN, A., SWANN, J., ANDERSSON, M. & JEFFERY, K. 2020. Optimized use of Oxford Nanopore flowcells for hybrid assemblies. Microbial genomics, 6.

LIU, C., ZHANG, L., LIU, H. & CHENG, K. 2017. Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications. Journal of controlled release, 266, 17-26.

LIU, X., XIAO, J., HUANG, H., GUAN, L., ZHAO, K., XU, Q., ZHANG, X., PAN, X., GU, S. & CHEN, Y. 2015. Molecular genetic testing in clinical diagnostic assessments that demonstrate correlations in patients with autosomal recessive inherited retinal dystrophy. JAMA ophthalmology, 133, 427-436.

LOGSDON, G. A., VOLLGER, M. R. & EICHLER, E. E. 2020. Long-read human genome sequencing and its applications. Nature Reviews Genetics, 21, 597-614.

LU, H., GIORDANO, F. & NING, Z. 2016. Oxford Nanopore MinION sequencing and genome assembly. Genomics, proteomics & bioinformatics, 14, 265-279.

LUPSKI, J. R. 1998. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends in genetics, 14, 417-422.

LYRAKI, R., LOKAJ, M., SOARES, D. C., LITTLE, A., VERMEREN, M., MARSH, J. A., WITTINGHOFER, A. & HURD, T. 2018. Characterization of a novel RP2–OSTF1 interaction and its implication for actin remodelling. Journal of cell science, 131, jcs211748.

LYRAKI, R., MEGAW, R. & HURD, T. 2016. Disease mechanisms of X-linked retinitis pigmentosa due to RP2 and RPGR mutations. Biochemical Society Transactions, 44, 1235-1244.

MADISSOON, E., JOUHILAHTI, E.-M., VESTERLUND, L., TÖHÖNEN, V., KRJUTŠKOV, K., PETROPOULOS, S., EINARSDOTTIR, E., LINNARSSON, S., LANNER, F. & MÅNSSON, R. 2016. Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. Scientific Reports, 6, 1-14.

MAEDER, M. L., STEFANIDAKIS, M., WILSON, C. J., BARAL, R., BARRERA, L. A., BOUNOUTAS, G. S., BUMCROT, D., CHAO, H., CIULLA, D. M. & DASILVA, J. A. 2019. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. Nature medicine, 25, 229-233.

MAGGI, J., ROBERTS, L., KOLLER, S., REBELLO, G., BERGER, W. & RAMESAR, R. 2020. De Novo Assembly-Based Analysis of RPGR Exon ORF15 in an Indigenous African Cohort Overcomes Limitations of a Standard Next-Generation Sequencing (NGS) Data Analysis Pipeline. Genes (Basel), 11.

MALLER, J. B., FAGERNESS, J. A., REYNOLDS, R. C., NEALE, B. M., DALY, M. J. & SEDDON, J. M. 2007. Variation in complement factor 3 is associated with risk of age-related macular degeneration. Nature genetics, 39, 1200-1201.

MANDAI, M., WATANABE, A., KURIMOTO, Y., HIRAMI, Y., MORINAGA, C., DAIMON, T., FUJIHARA, M., AKIMARU, H., SAKAI, N. & SHIBATA, Y. 2017. Autologous induced stem-cell–derived retinal cells for macular degeneration. New England Journal of Medicine, 376, 1038-1046.

MARDIS, E. R. 2013. Next-generation sequencing platforms. Annu Rev Anal Chem, 6, 287-303.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y.-J. & CHEN, Z. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437, 376-380.

MARTIN, M., PATTERSON, M., GARG, S., FISCHER, S. O., PISANTI, N., KLAU, G. W., SCHÖENHUTH, A. & MARSCHALL, T. 2016. WhatsHap: fast and accurate read-based phasing. BioRxiv, 085050.

MARTÍN-SÁNCHEZ, M., BRAVO-GIL, N., GONZÁLEZ-DEL POZO, M., MÉNDEZ-VIDAL, C., FERNÁNDEZ-SUÁREZ, E., RODRÍGUEZ-DE LA RÚA, E., BORREGO, S. & ANTIÑOLO, G. 2020. A multi-strategy sequencing workflow in inherited retinal dystrophies: routine diagnosis, addressing unsolved cases and candidate genes identification. International journal of molecular sciences, 21, 9355.

MARTINEZ-FERNANDEZ DE LA CAMARA, C., CEHAJIC-KAPETANOVIC, J. & MACLAREN, R. E. 2022. Emerging gene therapy products for RPGR-associated X-linked retinitis pigmentosa. Expert Opinion on Emerging Drugs, 27, 431-443.

MARTINEZ-FERNANDEZ DE LA CAMARA, C., NANDA, A., SALVETTI, A. P., FISCHER, M. D. & MACLAREN, R. E. 2018. Gene therapy for the treatment of X-linked retinitis pigmentosa. Expert opinion on orphan drugs, 6, 167-177.

MASLAND, R. H. 2001. The fundamental plan of the retina. Nature neuroscience, 4, 877-886.

MAUBARET, C. G., VACLAVIK, V., MUKHOPADHYAY, R., WASEEM, N. H., CHURCHILL, A., HOLDER, G. E., MOORE, A. T., BHATTACHARYA, S. S. & WEBSTER, A. R. 2011. Autosomal dominant retinitis pigmentosa with intrafamilial variability and incomplete penetrance in two families carrying mutations in PRPF8. Investigative ophthalmology & visual science, 52, 9304-9309.

MAXAM, A. M. & GILBERT, W. 1977. A new method for sequencing DNA. Proceedings of the National Academy of Sciences, 74, 560-564.

MC CLINTON, B., CORRADI, Z., MCKIBBIN, M., PANNEMAN, D. M., ROOSING, S., BOONEN, E. G., ALI, M., WATSON, C. M., STEEL, D. H. & CREMERS, F. P. 2023. Effective smMIPs-Based Sequencing of Maculopathy-Associated Genes in Stargardt Disease Cases and Allied Maculopathies from the UK. Genes, 14, 191.

MCCOMBIE, W. R., MCPHERSON, J. D. & MARDIS, E. R. 2019. Next-generation sequencing technologies. Cold Spring Harbor perspectives in medicine, 9, a036798.

MCGEE, T. L., SEYEDAHMADI, B. J., SWEENEY, M. O., DRYJA, T. P. & BERSON, E. L. 2010. Novel mutations in the long isoform of the USH2A gene in patients with Usher syndrome type II or non-syndromic retinitis pigmentosa. Journal of medical genetics, 47, 499-506.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S. & DALY, M. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research, 20, 1297-1303.

MCLAREN, W., PRITCHARD, B., RIOS, D., CHEN, Y., FLICEK, P. & CUNNINGHAM, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics, 26, 2069-2070.

MCLENACHAN, S., ZHANG, D., GRAINOK, J., ZHANG, X., HUANG, Z., CHEN, S. C., ZAW, K., LIMA, A., JENNINGS, L., ROSHANDEL, D., MOON, S. Y., HEATH JEFFERY, R. C., ATTIA, M. S., THOMPSON, J. A., LAMEY, T. M., MCLAREN, T. L., DE ROACH, J., FLETCHER, S. & CHEN, F. K. 2021. Determinants of Disease Penetrance in PRPF31-Associated Retinopathy. Genes (Basel), 12.

MCQUILLAN, R., LEUTENEGGER, A.-L., ABDEL-RAHMAN, R., FRANKLIN, C. S., PERICIC, M., BARAC-LAUC, L., SMOLEJ-NARANCIC, N., JANICIJEVIC, B., POLASEK, O. & TENESA, A. 2008. Runs of homozygosity in European populations. The American Journal of Human Genetics, 83, 359-372.

MCWILLIAM, P., FARRAR, G. J., KENNA, P., BRADLEY, D. G., HUMPHRIES, M. M., SHARP, E. M., MCCONNELL, D. J., LAWLER, M., SHEILS, D. & RYAN, C. 1989. Autosomal dominant retinitis pigmentosa (ADRP): localization of an ADRP gene to the long arm of chromosome 3. Genomics, 5, 619-622.

MEGAW, R., ABU-ARAFEH, H., JUNGNICKEL, M., MELLOUGH, C., GURNIAK, C., WITKE, W., ZHANG, W., KHANNA, H., MILL, P. & DHILLON, B. 2017. Gelsolin dysfunction causes photoreceptor loss in induced pluripotent cell and animal retinitis pigmentosa models. Nature Communications, 8, 271.

MEGAW, R. D., SOARES, D. C. & WRIGHT, A. F. 2015. RPGR: Its role in photoreceptor physiology, human disease, and future therapies. Experimental eye research, 138, 32-41.

MEINDL, A., DRY, K., HERRMANN, K., MANSON, E., CICCODICOLA, A., EDGAR, A., CARVALHO, M., ACHATZ, H., HELLEBRAND, H. & LENNON, A. 1996. A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X–linked retinitis pigmentosa (RP3). Nature genetics, 13, 35-42.

MENG, D., RAGI, S. D. & TSANG, S. H. 2020. Therapy in rhodopsin-mediated autosomal dominant retinitis pigmentosa. Molecular Therapy, 28, 2139-2149.

MENG, X., LIU, X., LI, Y., GUO, T. & YANG, L. 2021. Correlation between genotype and phenotype in 69 Chinese patients with USH2A mutations: A comparative study of the patients with Usher syndrome and nonsyndromic retinitis pigmentosa. Acta Ophthalmologica, 99, e447-e460.

MICHAELIDES, M., ALIGIANIS, I., HOLDER, G., SIMUNOVIC, M., MOLLON, J., MAHER, E., HUNT, D. & MOORE, A. 2003a. Cone dystrophy phenotype associated with a frameshift mutation (M280fsX291) in the α-subunit of cone specific transducin (GNAT2). British journal of ophthalmology, 87, 1317-1320.

MICHAELIDES, M., HARDCASTLE, A. J., HUNT, D. M. & MOORE, A. T. 2006. Progressive cone and cone-rod dystrophies: phenotypes and underlying molecular genetic basis. Survey of ophthalmology, 51, 232-258.

MICHAELIDES, M., HUNT, D. & MOORE, A. 2003b. The genetics of inherited macular dystrophies. Journal of medical genetics, 40, 641-650.

MICHALAKIS, S., GERHARDT, M., RUDOLPH, G., PRIGLINGER, S. & PRIGLINGER, C. 2021. Achromatopsia: Genetics and Gene Therapy. Molecular diagnosis & therapy, 1-9.

MILLER, D. E., SULOVARI, A., WANG, T., LOUCKS, H., HOEKZEMA, K., MUNSON, K. M., LEWIS, A. P., FUERTE, E. P. A., PASCHAL, C. R. & WALSH, T. 2021a. Targeted long-read sequencing identifies missing disease-causing variation. The American Journal of Human Genetics, 108, 1436-1449.

MILLER, D. T., LEE, K., CHUNG, W. K., GORDON, A. S., HERMAN, G. E., KLEIN, T. E., STEWART, D. R., AMENDOLA, L. M., ADELMAN, K. & BALE, S. J. 2021b. ACMG SF v3. 0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). Genetics in medicine, 23, 1381-1390.

MILLER, D. T., LEE, K., GORDON, A. S., AMENDOLA, L. M., ADELMAN, K., BALE, S. J., CHUNG, W. K., GOLLOB, M. H., HARRISON, S. M. & HERMAN, G. E. 2021c. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: a policy statement of the American College of Medical Genetics and Genomics (ACMG). Genetics in Medicine, 23, 1391-1398.

MILLER, S. S. & EDELMAN, J. L. 1990. Active ion transport pathways in the bovine retinal pigment epithelium. The Journal of physiology, 424, 283-300.

MIYAMOTO, T., INOUE, H., SAKAMOTO, Y., KUDO, E., NAITO, T., MIKAWA, T., MIKAWA, Y., ISASHIKI, Y., OSABE, D. & SHINOHARA, S. 2005. Identification of a novel splice site mutation of the CSPG2 gene in a Japanese family with Wagner syndrome. Investigative ophthalmology & visual science, 46, 2726-2735.

MOCKEL, A., PERDOMO, Y., STUTZMANN, F., LETSCH, J., MARION, V. & DOLLFUS, H. 2011. Retinal dystrophy in Bardet–Biedl syndrome and related syndromic ciliopathies. Progress in retinal and eye research, 30, 258-274.

MOISEYEV, G., CHEN, Y., TAKAHASHI, Y., WU, B. X. & MA, J.-X. 2005. RPE65 is the isomerohydrolase in the retinoid visual cycle. Proceedings of the National Academy of Sciences, 102, 12413-12418.

MORENO-LEON, L., WEST, E. L., O'HARA-WRIGHT, M., LI, L., NAIR, R., HE, J., ANAND, M., SAHU, B., CHAVALI, V. R. M., SMITH, A. J., ALI, R. R., JACOBSON, S. G., CIDECIYAN, A. V. & KHANNA, H. 2021. RPGR isoform imbalance causes ciliary defects due to exon ORF15 mutations in X-linked retinitis pigmentosa (XLRP). Hum Mol Genet, 29, 3706-3716.

MORENO-LEON, L., WEST, E. L., O'HARA-WRIGHT, M., LI, L., NAIR, R., HE, J., ANAND, M., SAHU, B., CHAVALI, V. R. M. & SMITH, A. J. 2020. RPGR isoform imbalance causes ciliary defects due to exon ORF15 mutations in X-linked retinitis pigmentosa (XLRP). Human molecular genetics, 29, 3706-3716.

MUKHOPADHYAY, A., NIKOPOULOS, K., MAUGERI, A., DE BROUWER, A. P., VAN NOUHUYS, C. E., BOON, C. J., PERVEEN, R., ZEGERS, H. A., WITTEBOL-

POST, D. & VAN DEN BIESEN, P. R. 2006. Erosive vitreoretinopathy and wagner disease are caused by intronic mutations in CSPG2/Versican that result in an imbalance of splice variants. Investigative ophthalmology & visual science, 47, 3565-3572.

MUNIZ, A., BETTS, B. S., TREVINO, A. R., BUDDAVARAPU, K., ROMAN, R., MA, J.-X. & TSIN, A. T. 2009. Evidence for two retinoid cycles in the cone-dominated chicken eye. Biochemistry, 48, 6854-6863.

MUSTAFI, D. & CHAO, J. R. 2022. Re: Yahya et al.: Late-onset autosomal dominant macular degeneration caused by deletion of the CRX gene (Ophthalmology. 2022 Aug 5; S0161-6420 (22) 00565-6. Ophthalmology.

MUSTAFI, D., ENGEL, A. H. & PALCZEWSKI, K. 2009. Structure of cone photoreceptors. Progress in retinal and eye research, 28, 289-302.

MYLLYKANGAS, S., BUENROSTRO, J. & JI, H. P. 2012. Overview of sequencing technology platforms. Bioinformatics for high throughput sequencing. Springer.

NA, K.-H., KIM, H. J., KIM, K. H., HAN, S., KIM, P., HANN, H. J. & AHN, H. S. 2017. Prevalence, age at diagnosis, mortality, and cause of death in Retinitis Pigmentosa in Korea—a nationwide population-based study. American journal of ophthalmology, 176, 157-165.

NANDA, A., SALVETTI, A. P., CLOUSTON, P., DOWNES, S. M. & MACLAREN, R. E. 2018. Exploring the variable phenotypes of RPGR carrier females in assessing their potential for retinal gene therapy. Genes, 9, 643.

NANGIA, V., JONAS, J. B., KHARE, A. & SINHA, A. 2012. Prevalence of retinitis pigmentosa in India: the Central India eye and medical study. Acta ophthalmologica, 90, e649-e650.

NARCISI, T. M., SHOULDERS, C. C., CHESTER, S. A., READ, J., BRETT, D. J., HARRISON, G. B., GRANTHAM, T. T., FOX, M. F., POVEY, S. & DE BRUIN, T. W. 1995. Mutations of the microsomal triglyceride-transfer–protein gene in abetalipoproteinemia. American journal of human genetics, 57, 1298.

NASH, B. M., MA, A., HO, G., FARNSWORTH, E., MINOCHE, A. E., COWLEY, M. J., BARNETT, C., SMITH, J. M., LOI, T. H. & WONG, K. 2022. Whole Genome Sequencing, Focused Assays and Functional Studies Increasing Understanding in Cryptic Inherited Retinal Dystrophies. International journal of molecular sciences, 23, 3905.

NASSER, F., KURTENBACH, A., KOHL, S., OBERMAIER, C., STINGL, K. & ZRENNER, E. 2019. Retinal dystrophies with bull's-eye maculopathy along with negative ERGs. Documenta Ophthalmologica, 139, 45-57.

NATHANS, J., THOMAS, D. & HOGNESS, D. S. 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science, 232, 193-202.

NEIDHARDT, J., GLAUS, E., LORENZ, B., NETZER, C., LI, Y., SCHAMBECK, M., WITTMER, M., FEIL, S., KIRSCHNER-SCHWABE, R. & ROSENBERG, T. 2008. Identification of novel mutations in X-linked retinitis pigmentosa families and implications for diagnostic testing. Molecular vision, 14, 1081.

NEVELING, K., COLLIN, R. W., GILISSEN, C., VAN HUET, R. A., VISSER, L., KWINT, M. P., GIJSEN, S. J., ZONNEVELD, M. N., WIESKAMP, N. & DE LIGT, J. 2012. Next-generation genetic testing for retinitis pigmentosa. Human mutation, 33, 963-972.

NG, C. C., CARRERA, W. M., MCDONALD, H. R. & AGARWAL, A. 2020. Heterozygous CRX R90W mutation-associated adult-onset macular dystrophy with phenotype analogous to benign concentric annular macular dystrophy. Ophthalmic Genetics, 41, 485-490.

NG, P. C. & HENIKOFF, S. 2001. Predicting deleterious amino acid substitutions. Genome research, 11, 863-874.

NG, T. K., TANG, W., CAO, Y., CHEN, S., ZHENG, Y., XIAO, X. & CHEN, H. 2019. Whole exome sequencing identifies novel USH2A mutations and confirms Usher syndrome 2 diagnosis in Chinese retinitis pigmentosa patients. Scientific reports, 9, 1-11.

NGUYEN, M., BOESTEN, I., HELLEBREKERS, D., MULDER-DEN HARTOG, N., DE COO, I., SMEETS, H. & GERARDS, M. 2017. Novel pathogenic SLC25A46 splice-site mutation causes an optic atrophy spectrum disorder. Clinical genetics, 91, 121-125.

NGUYEN, X.-T.-A., TALIB, M., VAN SCHOONEVELD, M. J., BRINKS, J., TEN BRINK, J., FLORIJN, R. J., WIJNHOLDS, J., VERDIJK, R. M., BERGEN, A. A. & BOON, C. J. 2020. RPGR-associated dystrophies: clinical, genetic, and histopathological features. International journal of molecular sciences, 21, 835.

NGUYEN-BA-CHARVET, K. T. & CHÉDOTAL, A. 2014. Development of retinal layers. Comptes Rendus Biologies, 337, 153-159.

NISHIGUCHI, K. M., KUNIKATA, H., FUJITA, K., HASHIMOTO, K., KOYANAGI, Y., AKIYAMA, M., IKEDA, Y., MOMOZAWA, Y., SONODA, K. H. & MURAKAMI, A. 2020. Association of CRX genotypes and retinal phenotypes confounded by variable expressivity and electronegative electroretinogram. Clinical & Experimental Ophthalmology, 48, 644-657.

NISHIGUCHI, K. M., SANDBERG, M. A., GORJI, N., BERSON, E. L. & DRYJA, T. P. 2005. Cone cGMP-gated channel mutations and clinical findings in patients with

achromatopsia, macular degeneration, and other hereditary cone diseases. Human mutation, 25, 248-258.

NISHIGUCHI, K. M., TEARLE, R. G., LIU, Y. P., OH, E. C., MIYAKE, N., BENAGLIO, P., HARPER, S., KOSKINIEMI-KUENDIG, H., VENTURINI, G. & SHARON, D. 2013. Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene. Proceedings of the National Academy of Sciences, 110, 16139-16144.

NOVAS, R., CARDENAS-RODRIGUEZ, M., IRIGOÍN, F. & BADANO, J. L. 2015. Bardet–Biedl syndrome: is it only cilia dysfunction? FEBS letters, 589, 3479-3491.

NUMA, S., OISHI, A., HIGASA, K., OISHI, M., MIYATA, M., HASEGAWA, T., IKEDA, H. O., OTSUKA, Y., MATSUDA, F. & TSUJIKAWA, A. 2020. EYS is a major gene involved in retinitis pigmentosa in Japan: genetic landscapes revealed by stepwise genetic screening. Scientific reports, 10, 1-9.

NURK, S., KOREN, S., RHIE, A., RAUTIAINEN, M., BZIKADZE, A. V., MIKHEENKO, A., VOLLGER, M. R., ALTEMOSE, N., URALSKY, L. & GERSHMAN, A. 2022. The complete sequence of a human genome. Science, 376, 44-53.

NYRÉN, P. & LUNDIN, A. 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Analytical biochemistry, 151, 504-509.

OLICHON, A., BARICAULT, L., GAS, N., GUILLOU, E., VALETTE, A., BELENGUER, P. & LENAERS, G. 2003. Loss of OPA1 perturbates the mitochondrial inner membrane structure and integrity, leading to cytochrome c release and apoptosis. Journal of Biological Chemistry, 278, 7743-7746.

OLICHON, A., LANDES, T., ARNAUNÉ-PELLOQUIN, L., EMORINE, L. J., MILS, V., GUICHET, A., DELETTRE, C., HAMEL, C., AMATI-BONNEAU, P. & BONNEAU, D. 2007. Effects of OPA1 mutations on mitochondrial morphology and apoptosis: relevance to ADOA pathogenesis. Journal of cellular physiology, 211, 423-430.

OSTERGAARD, E., BATBAYLI, M., DUNO, M., VILHELMSEN, K. & ROSENBERG, T. 2010. Mutations in PCDH21 cause autosomal recessive cone-rod dystrophy. Journal of medical genetics, 47, 665-669.

PAGON, R. A. 1988. Retinitis pigmentosa. Survey of ophthalmology, 33, 137-177.

PANAGIOTOU, E. 2018. Using NGS to identify new genes and modifiers underlying Familial Exudative Vitreoretinopathy (FEVR). Thesis submitted for the degree of PhD. University of Leeds.

PANAGIOTOU, E. S., SORIANO, C. S., POULTER, J. A., LORD, E. C., DZULOVA, D., KONDO, H., HIYOSHI, A., CHUNG, B. H.-Y., CHU, Y. W.-Y. & LAI, C. H. 2017. Defects in the cell signaling mediator β-catenin cause the retinal vascular condition FEVR. The American Journal of Human Genetics, 100, 960-968.

PARK, H., YAMAMOTO, H., MOHN, L., AMBÜHL, L., KANAI, K., SCHMIDT, I., KIM, K.-P., FRACCAROLI, A., FEIL, S. & JUNGE, H. J. 2019. Integrin-linked kinase controls retinal angiogenesis and is linked to Wnt signaling and exudative vitreoretinopathy. Nature communications, 10, 1-14.

PATIL, S. B., HURD, T. W., GHOSH, A. K., MURGA-ZAMALLOA, C. A. & KHANNA, H. 2011. Functional analysis of retinitis pigmentosa 2 (RP2) protein reveals variable pathogenic potential of disease-associated missense variants. PLoS One, 6, e21379.

PATNAIK, S. R., RAGHUPATHY, R. K., ZHANG, X., MANSFIELD, D. & SHU, X. 2015. The role of RPGR and its interacting proteins in ciliopathies. Journal of ophthalmology, 2015.

PATNALA, R., CLEMENTS, J. & BATRA, J. 2013. Candidate gene association studies: a comprehensive guide to useful in silicotools. BMC genetics, 14, 1-11.

PAWLYK, B. S., BULGAKOV, O., SUN, X., ADAMIAN, M., SHU, X., SMITH, A. J., BERSON, E. L., ALI, R. R., KHANI, S. & WRIGHT, A. F. 2016. Photoreceptor rescue by an abbreviated human RPGR gene in a murine model of X-linked retinitis pigmentosa. Gene therapy, 23, 196-204.

PAYNE, A., HOLMES, N., CLARKE, T., MUNRO, R., DEBEBE, B. J. & LOOSE, M. 2021. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. Nature biotechnology, 39, 442-450.

PEARSALL, N., BHATTACHARYA, G., WISECARVER, J., ADAMS, J., COSGROVE, D. & KIMBERLING, W. 2002. Usherin expression is highly conserved in mouse and human tissues. Hearing research, 174, 55-63.

PENG, G.-H., AHMAD, O., AHMAD, F., LIU, J. & CHEN, S. 2005. The photoreceptor-specific nuclear receptor Nr2e3 interacts with Crx and exerts opposing effects on the transcription of rod versus cone genes. Human molecular genetics, 14, 747-764.

PEREA-ROMERO, I., GORDO, G., IANCU, I. F., POZO-VALERO, D., ALMOGUERA, B., BLANCO-KELLY, F., CARREÑO, E., JIMENEZ-ROLANDO, B., LOPEZ-RODRIGUEZ, R. & LORDA-SANCHEZ, I. 2021. Genetic landscape of 6089 inherited retinal dystrophies affected cases in Spain and their therapeutic and extended epidemiological implications. Scientific reports, 11, 1-13.

PERLAND, E., HELLSTEN, S. V., LEKHOLM, E., ERIKSSON, M. M., ARAPI, V. & FREDRIKSSON, R. 2017a. The novel membrane-bound proteins MFSD1 and MFSD3 are putative SLC transporters affected by altered nutrient intake. Journal of Molecular Neuroscience, 61, 199-214.

PERLAND, E., HELLSTEN, S. V., SCHWEIZER, N., ARAPI, V., REZAYEE, F., BUSHRA, M. & FREDRIKSSON, R. 2017b. Structural prediction of two novel human atypical SLC transporters, MFSD4A and MFSD9, and their neuroanatomical distribution in mice. PLoS One, 12, e0186325.

PERLAND, E., LEKHOLM, E., ERIKSSON, M. M., BAGCHI, S., ARAPI, V. & FREDRIKSSON, R. 2016. The putative SLC transporters Mfsd5 and Mfsd11 are abundantly expressed in the mouse brain and have a potential role in energy homeostasis. PloS one, 11, e0156912.

PERRAULT, I., ROZET, J.-M., GERBER, S., GHAZI, I., LEOWSKI, C., DUCROQ, D., SOUIED, E., DUFIER, J.-L., MUNNICH, A. & KAPLAN, J. 1999. Leber congenital amaurosis. Molecular genetics and metabolism, 68, 200-208.

PERRY, G. H., YANG, F., MARQUES-BONET, T., MURPHY, C., FITZGERALD, T., LEE, A. S., HYLAND, C., STONE, A. C., HURLES, M. E. & TYLER-SMITH, C. 2008. Copy number variation and evolution in humans and chimpanzees. Genome research, 18, 1698-1710.

PETTERSSON, E., LUNDEBERG, J. & AHMADIAN, A. 2009. Generations of sequencing technologies. Genomics, 93, 105-111.

PHELAN, J. K. & BOK, D. 2000. A brief review of retinitis pigmentosa and the identified retinitis pigmentosa genes. Mol Vis, 6, 24.

PINCKERS, A., VAN AAREM, A. & KEUNEN, J. 1993. Colour vision in retinitis pigmentosa. International Ophthalmology, 17, 143-146.

PIPIS, M., ROSSOR, A. M., LAURA, M. & REILLY, M. M. 2019. Next-generation sequencing in Charcot–Marie–Tooth disease: opportunities and challenges. Nature Reviews Neurology, 15, 644-656.

PLAGNOL, V., CURTIS, J., EPSTEIN, M., MOK, K. Y., STEBBINGS, E., GRIGORIADOU, S., WOOD, N. W., HAMBLETON, S., BURNS, S. O., THRASHER, A. J., KUMARARATNE, D., DOFFINGER, R. & NEJENTSEV, S. 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics, 28, 2747-54.

POULTER, J. A., ALI, M., GILMOUR, D. F., RICE, A., KONDO, H., HAYASHI, K., MACKEY, D. A., KEARNS, L. S., RUDDLE, J. B. & CRAIG, J. E. 2010. Mutations in TSPAN12 cause autosomal-dominant familial exudative vitreoretinopathy. The American Journal of Human Genetics, 86, 248-253.

PRILUCK, A. Z. & BREAZZANO, M. P. 2022. Novel MFSD8 mutation causing non-syndromic asymmetric adult-onset macular dystrophy. Ophthalmic genetics, 1-5.

PULST, S. M. 1999. Genetic linkage analysis. Archives of neurology, 56, 667-672.

PURNYN, H. 2013. The mammalian retina: structure and blood supply. Neurophysiology, 45, 266-276.

QIAN, X., WANG, J., WANG, M., IGELMAN, A. D., JONES, K. D., LI, Y., WANG, K., GOETZ, K. E., BIRCH, D. G. & YANG, P. 2021. Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. Frontiers in Genetics, 12, 647400.

QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics, 13, 341.

QUINODOZ, M., PETER, V. G., BEDONI, N., ROYER BERTRAND, B., CISAROVA, K., SALMANINEJAD, A., SEPAHI, N., RODRIGUES, R., PIRAN, M. & MOJARRAD, M. 2021. AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. Nature communications, 12, 1-7.

RAHMAN, N., GEORGIOU, M., KHAN, K. N. & MICHAELIDES, M. 2020. Macular dystrophies: clinical and imaging features, molecular genetics and therapeutic options. British Journal of Ophthalmology, 104, 451-460.

RAMPERSAUD, E., DAMCOTT, C. M., FU, M., SHEN, H., MCARDLE, P., SHI, X., SHELTON, J., YIN, J., CHANG, Y.-P. C. & OTT, S. H. 2007. Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. Diabetes, 56, 3053-3062.

RAN, F. A., HSU, P. D., LIN, C.-Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S. & ZHANG, Y. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell, 154, 1380-1389.

RATTNER, A., SMALLWOOD, P. M. & NATHANS, J. 2000. Identification and characterization of all-trans-retinol dehydrogenase from photoreceptor outer segments, the visual cycle enzyme that reduces all-trans-retinal to all-trans-retinol. Journal of Biological Chemistry, 275, 11034-11043.

REDMOND, T. M., POLIAKOV, E., YU, S., TSAI, J.-Y., LU, Z. & GENTLEMAN, S. 2005. Mutation of key residues of RPE65 abolishes its enzymatic role as isomerohydrolase in the visual cycle. Proceedings of the National Academy of Sciences, 102, 13658-13663.

REINERS, J., MARKER, T., JURGENS, K., REIDEL, B. & WOLFRUM, U. 2005. Photoreceptor expression of the Usher syndrome type 1 protein protocadherin 15 (USH1F) and its interaction with the scaffold protein harmonin (USH1C).

REINERS, J., NAGEL-WOLFRUM, K., JÜRGENS, K., MÄRKER, T. & WOLFRUM, U. 2006. Molecular basis of human Usher syndrome: deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease. Experimental eye research, 83, 97-119.

RENTZSCH, P., WITTEN, D., COOPER, G. M., SHENDURE, J. & KIRCHER, M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic acids research, 47, D886-D894.

REURINK, J., DOCKERY, A., OZIĘBŁO, D., FARRAR, G. J., OŁDAK, M., TEN BRINK, J. B., BERGEN, A. A., RINNE, T., YNTEMA, H. G. & PENNINGS, R. J. 2021. Molecular Inversion Probe-Based Sequencing of USH2A Exons and Splice Sites as a Cost-Effective Screening Tool in USH2 and arRP Cases. International journal of molecular sciences, 22, 6419.

RHOADS, A. & AU, K. F. 2015. PacBio sequencing and its applications. Genomics, proteomics & bioinformatics, 13, 278-289.

RIAZUDDIN, S. A., ZULFIQAR, F., ZHANG, Q., SERGEEV, Y. V., QAZI, Z. A., HUSNAIN, T., CARUSO, R., RIAZUDDIN, S., SIEVING, P. A. & HEJTMANCIK, J. F. 2005. Autosomal recessive retinitis pigmentosa is associated with mutations in RP1 in three consanguineous Pakistani families. Investigative ophthalmology & visual science, 46, 2264-2270.

RICE, D. S., CALANDRIA, J. M., GORDON, W. C., JUN, B., ZHOU, Y., GELFMAN, C. M., LI, S., JIN, M., KNOTT, E. J. & CHANG, B. 2015. Adiponectin receptor 1 conserves docosahexaenoic acid and promotes photoreceptor cell survival. Nature communications, 6, 1-15.

RIERA, M., NAVARRO, R., RUIZ-NOGALES, S., MÉNDEZ, P., BURÉS-JELSTRUP, A., CORCÓSTEGUI, B. & POMARES, E. 2017. Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. Scientific reports, 7, 1-13.

ROBINSON, J. T., THORVALDSDÓTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. Nature biotechnology, 29, 24-26.

ROBITAILLE, J. M., GILLETT, R. M., LEBLANC, M. A., GASTON, D., NIGHTINGALE, M., MACKLEY, M. P., PARKASH, S., HATHAWAY, J., THOMAS, A. & ELLS, A. 2014. Phenotypic overlap between familial exudative

vitreoretinopathy and microcephaly, lymphedema, and chorioretinal dysplasia caused by KIF11 mutations. JAMA ophthalmology, 132, 1393-1399.

RONAN, S. M., TRAN-VIET, K.-N., BURNER, E. L., METLAPALLY, R., TOTH, C. A. & YOUNG, T. L. 2009. Mutational hot spot potential of a novel base pair mutation of the CSPG2 gene in a family with Wagner syndrome. Archives of ophthalmology, 127, 1511-1519.

ROOSING, S., HOFREE, M., KIM, S., SCOTT, E., COPELAND, B., ROMANI, M., SILHAVY, J. L., ROSTI, R. O., SCHROTH, J. & MAZZA, T. 2015a. Functional genome-wide siRNA screen identifies KIAA0586 as mutated in Joubert syndrome. Elife, 4, e06602.

ROOSING, S., THIADENS, A. A., HOYNG, C. B., KLAVER, C. C., DEN HOLLANDER, A. I. & CREMERS, F. P. 2014. Causes and consequences of inherited cone disorders. Progress in retinal and eye research, 42, 1-26.

ROOSING, S., VAN DEN BORN, L. I., SANGERMANO, R., BANFI, S., KOENEKOOP, R. K., ZONNEVELD-VRIELING, M. N., KLAVER, C. C., VAN LITH-VERHOEVEN, J. J., CREMERS, F. P. & DEN HOLLANDER, A. I. 2015b. Mutations in MFSD8, encoding a lysosomal membrane protein, are associated with nonsyndromic autosomal recessive macular dystrophy. Ophthalmology, 122, 170-179.

ROSE, A. & BHATTACHARYA, S. 2016. Variant haploinsufficiency and phenotypic non-penetrance in PRPF31-associated retinitis pigmentosa. Clinical genetics, 90, 118-126.

ROSEN, D. & MAHABADI, N. 2021. Embryology, Optic Cup. StatPearls. Treasure Island (FL).

ROSHANDEL, D., RAFATI, M., KHORAMI, S., NOVIN BAHERAN, N., JALALI, S., TABATABAIE, R., REZAI, S., AHMADIEH, H. & GHAFFARI, S. R. 2019. Rhodopsin gene mutation analysis in Iranian patients with autosomal dominant retinitis pigmentosa. International Ophthalmology, 39, 2523-2531.

ROTHSCHILD, P.-R., AUDO, I., NEDELEC, B., GHIOTTI, T., BRÉZIN, A. P., MONIN, C. & VALLEIX, S. 2013. De novo splice mutation in the versican gene in a family with Wagner syndrome. JAMA ophthalmology, 131, 805-807.

RUIZ, A., WINSTON, A., LIM, Y.-H., GILBERT, B. A., RANDO, R. R. & BOK, D. 1999. Molecular and biochemical characterization of lecithin retinol acyltransferase. Journal of Biological Chemistry, 274, 3834-3841.

RUNHART, E. H., KHAN, M., CORNELIS, S. S., ROOSING, S., DEL POZO-VALERO, M., LAMEY, T. M., LISKOVA, P., ROBERTS, L., STÖHR, H., KLAVER, C. C. W., HOYNG, C. B., CREMERS, F. P. M. & DHAENENS, C. M. 2020.

Association of Sex With Frequent and Mild ABCA4 Alleles in Stargardt Disease. JAMA Ophthalmol, 138, 1035-1042.

RUNHART, E. H., SANGERMANO, R., CORNELIS, S. S., VERHEIJ, J. B., PLOMP, A. S., BOON, C. J., LUGTENBERG, D., ROOSING, S., BAX, N. M. & BLOKLAND, E. A. 2018. The common ABCA4 variant p. Asn1868Ile shows nonpenetrance and variable expression of Stargardt disease when present in trans with severe variants. Investigative ophthalmology & visual science, 59, 3220-3231.

RUSSAKOFF, D. B., LAMIN, A., OAKLEY, J. D., DUBIS, A. M. & SIVAPRASAD, S. 2019. Deep learning for prediction of AMD progression: a pilot study. Investigative ophthalmology & visual science, 60, 712-722.

RUSSELL, S., BENNETT, J., WELLMAN, J. A., CHUNG, D. C., YU, Z.-F., TILLMAN, A., WITTES, J., PAPPAS, J., ELCI, O. & MCCAGUE, S. 2017. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. The Lancet, 390, 849-860.

SAARI, J. & BREDBERG, D. L. 1989. Lecithin: retinol acyltransferase in retinal pigment epithelial microsomes. Journal of Biological Chemistry, 264, 8636-8640.

SAHEL, J.-A., MARAZOVA, K. & AUDO, I. 2015. Clinical characteristics and current therapies for inherited retinal degenerations. Cold Spring Harbor perspectives in medicine, 5, a017111.

SAKAMOTO, Y., ZAHA, S., SUZUKI, Y., SEKI, M. & SUZUKI, A. 2021. Application of long-read sequencing to the detection of structural variants in human cancer genomes. Computational and Structural Biotechnology Journal, 19, 4207-4216.

SALLES, M. V., MOTTA, F. L., DIAS DA SILVA, E., VARELA, P., COSTA, K. A., FILIPPELLI-SILVA, R., MARTIN, R. P., CHIANG, J., PESQUERO, J. B. & SALLUM, J. M. F. 2017. Novel Complex ABCA4 Alleles in Brazilian Patients With Stargardt Disease: Genotype–Phenotype Correlation. Investigative Ophthalmology & Visual Science, 58, 5723-5730.

SANCHIS-JUAN, A., STEPHENS, J., FRENCH, C. E., GLEADALL, N., MÉGY, K., PENKETT, C., SHAMARDINA, O., STIRRUPS, K., DELON, I. & DEWHURST, E. 2018. Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short-and long-read genome sequencing. Genome medicine, 10, 1-10.

SANGER, F., AIR, G. M., BARRELL, B. G., BROWN, N. L., COULSON, A. R., FIDDES, J. C., HUTCHISON, C., SLOCOMBE, P. M. & SMITH, M. 1977a. Nucleotide sequence of bacteriophage φX174 DNA. nature, 265, 687-695.

SANGER, F. & COULSON, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology, 94, 441-448.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977b. DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences, 74, 5463-5467.

SANJURJO-SORIANO, C. & KALATZIS, V. 2018. Guiding lights in genome editing for inherited retinal disorders: implications for gene and cell therapy. Neural Plasticity, 2018.

SCHAAF, C. P., BLAZO, M., LEWIS, R. A., TONINI, R. E., TAKEI, H., WANG, J., WONG, L.-J. & SCAGLIA, F. 2011. Early-onset severe neuromuscular phenotype associated with compound heterozygosity for OPA1 mutations. Molecular genetics and metabolism, 103, 383-387.

SCHEERER, P., PARK, J. H., HILDEBRAND, P. W., KIM, Y. J., KRAUS, N., CHOE, H.-W., HOFMANN, K. P. & ERNST, O. P. 2008. Crystal structure of opsin in its G-protein-interacting conformation. Nature, 455, 497-502.

SCHIERLING, B., DANNEMANN, N., GABSALILOW, L., WENDE, W., CATHOMEN, T. & PINGOUD, A. 2012. A novel zinc-finger nuclease platform with a sequence-specific cleavage module. Nucleic acids research, 40, 2623-2638.

SCHMIDT-ERFURTH, U., BOGUNOVIC, H., SADEGHIPOUR, A., SCHLEGL, T., LANGS, G., GERENDAS, B. S., OSBORNE, A. & WALDSTEIN, S. M. 2018. Machine learning to analyze the prognostic value of current imaging biomarkers in neovascular age-related macular degeneration. Ophthalmology Retina, 2, 24-30.

SCHUETTPELZ, J., JANER, A., ANTONICKA, H. & SHOUBRIDGE, E. A. 2023. The role of the mitochondrial outer membrane protein SLC25A46 in mitochondrial fission and fusion. Life Sci Alliance, 6.

SCHWAHN, U., LENZNER, S., DONG, J., FEIL, S., HINZMANN, B., VAN DUIJNHOVEN, G., KIRSCHNER, R., HEMBERGER, M., BERGEN, A. A. & ROSENBERG, T. 1998. Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nature genetics, 19, 327-332.

SCHWARTZ, S. D., REGILLO, C. D., LAM, B. L., ELIOTT, D., ROSENFELD, P. J., GREGORI, N. Z., HUBSCHMAN, J.-P., DAVIS, J. L., HEILWELL, G. & SPIRN, M. 2015. Human embryonic stem cell-derived retinal pigment epithelium in patients with age-related macular degeneration and Stargardt's macular dystrophy: follow-up of two open-label phase 1/2 studies. The Lancet, 385, 509-516.

SCHWARZ, J. M., COOPER, D. N., SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. Nature methods, 11, 361-362.

SCHWARZE, K., BUCHANAN, J., TAYLOR, J. C. & WORDSWORTH, S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genetics in Medicine, 20, 1122-1130.

SEKO, Y., IWANAMI, M., MIYAMOTO-MATSUI, K., TAKITA, S., AOI, N., UMEZAWA, A. & KATO, S. 2018. The manner of decay of genetically defective EYS gene transcripts in photoreceptor-directed fibroblasts derived from retinitis pigmentosa patients depends on the type of mutation. Stem cell research & therapy, 9, 1-11.

SENGILLO, J. D., LEE, W., NAGASAKI, T., SCHUERCH, K., YANNUZZI, L. A., FREUND, K. B., SPARROW, J. R., ALLIKMETS, R. & TSANG, S. H. 2018. A distinct phenotype of eyes shut homolog (EYS)-retinitis pigmentosa is associated with variants near the C-terminus. American journal of ophthalmology, 190, 99-112.

SEYEDAHMADI, B. J., RIVOLTA, C., KEENE, J. A., BERSON, E. L. & DRYJA, T. P. 2004. Comprehensive screening of the USH2A gene in Usher syndrome type II and non-syndromic recessive retinitis pigmentosa. Experimental eye research, 79, 167-173.

SHANKS, M. E., DOWNES, S. M., COPLEY, R. R., LISE, S., BROXHOLME, J., HUDSPITH, K. A., KWASNIEWSKA, A., DAVIES, W. I., HANKINS, M. W. & PACKHAM, E. R. 2013. Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. European journal of human genetics, 21, 274-280.

SHARMA, R. K. & EHINGER, B. E. 2003. Development and structure of the retina. Adler's Physiology of the Eye: Clinical Applications, 10, 319-347.

SHARON, D., SANDBERG, M. A., RABE, V. W., STILLBERGER, M., DRYJA, T. P. & BERSON, E. L. 2003. RP2 and RPGR mutations and clinical correlations in patients with X-linked retinitis pigmentosa. The American Journal of Human Genetics, 73, 1131-1146.

SHASTRY, B. S., HEJTMANCIK, J. F., PLAGER, D. A., HARTZER, M. K. & TRESE, M. T. 1995. Linkage and candidate gene analysis of X-linked familial exudative vitreoretinopathy. Genomics, 27, 341-344.

SHERRY, S. T., WARD, M.-H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic acids research, 29, 308-311.

SHI, L., GUO, Y., DONG, C., HUDDLESTON, J., YANG, H., HAN, X., FU, A., LI, Q., LI, N. & GONG, S. 2016. Long-read sequencing and de novo assembly of a Chinese genome. Nature communications, 7, 12065.

SHIHAB, H. A., GOUGH, J., COOPER, D. N., STENSON, P. D., BARKER, G. L., EDWARDS, K. J., DAY, I. N. & GAUNT, T. R. 2013. Predicting the functional,

molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Human mutation, 34, 57-65.

SHIMADA, H., LU, Q., INSINNA-KETTENHOFEN, C., NAGASHIMA, K., ENGLISH, M. A., SEMLER, E. M., MAHGEREFTEH, J., CIDECIYAN, A. V., LI, T., BROOKS, B. P., GUNAY-AYGUN, M., JACOBSON, S. G., COGLIATI, T., WESTLAKE, C. J. & SWAROOP, A. 2017. In Vitro Modeling Using Ciliopathy-Patient-Derived Cells Reveals Distinct Cilia Dysfunctions Caused by CEP290 Mutations. Cell Rep, 20, 384-396.

SHIMIZU, S., MORI, N., KISHI, M., SUGATA, H., TSUDA, A. & KUBOTA, N. 2003. A novel mutation in the OPA1 gene in a Japanese patient with optic atrophy. American journal of ophthalmology, 135, 256-257.

SHINWARI, J. M., KHAN, A., AWAD, S., SHINWARI, Z., ALAIYA, A., ALANAZI, M., TAHIR, A., POIZAT, C. & AL TASSAN, N. 2015. Recessive mutations in COL25A1 are a cause of congenital cranial dysinnervation disorder. The American Journal of Human Genetics, 96, 147-152.

SHU, X., FRY, A., TULLOCH, B., MANSON, F., CRABB, J., KHANNA, H., FARAGHER, A., LENNON, A., HE, S. & TROJAN, P. 2005. RPGR ORF15 isoform co-localizes with RPGRIP1 at centrioles and basal bodies and interacts with nucleophosmin. Human Molecular Genetics, 14, 1183-1197.

SHU, X., MCDOWALL, E., BROWN, A. F. & WRIGHT, A. F. 2008. The human retinitis pigmentosa GTPase regulator gene variant database. Hum Mutat, 29, 605-8.

SIEMIATKOWSKA, A. M., ASTUTI, G. D., ARIMADYO, K., DEN HOLLANDER, A. I., FARADZ, S. M., CREMERS, F. P. & COLLIN, R. W. 2012. Identification of a novel nonsense mutation in RP1 that causes autosomal recessive retinitis pigmentosa in an Indonesian family. Molecular vision, 18, 2411.

SILVA, E., YANG, J.-M., LI, Y., DHARMARAJ, S., SUNDIN, O. H. & MAUMENEE, I. H. 2000. A CRX null mutation is associated with both Leber congenital amaurosis and a normal ocular phenotype. Investigative ophthalmology & visual science, 41, 2076-2079.

SIMON, A., HELLMAN, U., WERNSTEDT, C. & ERIKSSON, U. 1995. The Retinal Pigment Epithelial-specific 11-cis Retinol Dehydrogenase Belongs to the Family of Short Chain Alcohol Dehydrogenases (∗). Journal of Biological Chemistry, 270, 1107-1112.

SIMON, G. B., ABRAHAM, F. & MELAMED, S. 2004. Pingelapese achromatopsia: correlation between paradoxical pupillary response and clinical features. British journal of ophthalmology, 88, 223-225.

SLIJKERMAN, R. W., VACHÉ, C., DONA, M., GARCÍA-GARCÍA, G., CLAUSTRES, M., HETTERSCHIJT, L., PETERS, T. A., HARTEL, B. P., PENNINGS, R. J. & MILLAN, J. M. 2016. Antisense oligonucleotide-based splice correction for USH2A-associated retinal degeneration caused by a frequent deep-intronic mutation. Molecular Therapy-Nucleic Acids, 5, e381.

SMITH, A. J., SCHLICHTENBREDE, F. C., TSCHERNUTTER, M., BAINBRIDGE, J. W., THRASHER, A. J. & ALI, R. R. 2003. AAV-Mediated gene transfer slows photoreceptor loss in the RCS rat model of retinitis pigmentosa. Molecular therapy, 8, 188-195.

SMITH, L. M., FUNG, S., HUNKAPILLER, M. W., HUNKAPILLER, T. J. & HOOD, L. E. 1985. The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. Nucleic acids research, 13, 2399-2412.

SODI, A., BANFI, S., TESTA, F., DELLA CORTE, M., PASSERINI, I., PELO, E., ROSSI, S., SIMONELLI, F. & ITALIAN, I. R. D. W. G. 2021. RPE65-associated inherited retinal diseases: consensus recommendations for eligibility to gene therapy. Orphanet Journal of Rare Diseases, 16, 257.

SOHOCKI, M. M., SULLIVAN, L. S., MINTZ-HITTNER, H. A., BIRCH, D., HECKENLIVELY, J. R., FREUND, C. L., MCINNES, R. R. & DAIGER, S. P. 1998. A range of clinical phenotypes associated with mutations in CRX, a photoreceptor transcription-factor gene. The American Journal of Human Genetics, 63, 1307-1315.

SONG, C., DUFOUR, V. L., CIDECIYAN, A. V., YE, G.-J., SWIDER, M., NEWMARK, J. A., TIMMERS, A. M., ROBINSON, P. M., KNOP, D. R. & CHULAY, J. D. 2020. Dose range finding studies with two RPGR transgenes in a canine model of X-linked retinitis pigmentosa treated with subretinal gene therapy. Human Gene Therapy, 31, 743-755.

SONG, W. K., PARK, K.-M., KIM, H.-J., LEE, J. H., CHOI, J., CHONG, S. Y., SHIM, S. H., DEL PRIORE, L. V. & LANZA, R. 2015. Treatment of macular degeneration using embryonic stem cell-derived retinal pigment epithelium: preliminary results in Asian patients. Stem cell reports, 4, 860-872.

SONTAKKE, Y. 2018. Textbook of human embryology. CBS Publishers & Distributors, London.

SOUCY, M., KOLESNIKOVA, M., KIM, A. H. & TSANG, S. H. 2023. Phenotypic variability in PRPH2 as demonstrated by a family with incomplete penetrance of autosomal dominant cone-rod dystrophy. Doc Ophthalmol.

SPIEGEL, R., SAADA, A., FLANNERY, P. J., BURTÉ, F., SOIFERMAN, D., KHAYAT, M., EISNER, V., VLADOVSKI, E., TAYLOR, R. W. & BINDOFF, L. A.

2016. Fatal infantile mitochondrial encephalomyopathy, hypertrophic cardiomyopathy and optic atrophy associated with a homozygous OPA1 mutation. Journal of medical genetics, 53, 127-131.

SREEDHARAN, S., STEPHANSSON, O., SCHIÖTH, H. B. & FREDRIKSSON, R. 2011. Long evolutionary conservation and considerable tissue specificity of several atypical solute carrier transporters. Gene, 478, 11-18.

STEINBERG, R. 1985. Interactions between the retinal pigment epithelium and the neural retina. Documenta Ophthalmologica, 60, 327-346.

STENSON, P. D., MORT, M., BALL, E. V., EVANS, K., HAYDEN, M., HEYWOOD, S., HUSSAIN, M., PHILLIPS, A. D. & COOPER, D. N. 2017. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Human genetics, 136, 665-677.

STENSON, P. D., MORT, M., BALL, E. V., SHAW, K., PHILLIPS, A. D. & COOPER, D. N. 2014. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Human genetics, 133, 1-9.

STEPHENS, M., SMITH, N. J. & DONNELLY, P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet, 68, 978-89.

STEPHENSON, K. A., ZHU, J., WYNNE, N., DOCKERY, A., CAIRNS, R. M., DUIGNAN, E., WHELAN, L., MALONE, C. P., DEMPSEY, H. & COLLINS, K. 2021. Target 5000: a standardized all-Ireland pathway for the diagnosis and management of inherited retinal degenerations. Orphanet journal of rare diseases, 16, 1-8.

STOETZEL, C., LAURIER, V., DAVIS, E. E., MULLER, J., RIX, S., BADANO, J. L., LEITCH, C. C., SALEM, N., CHOUERY, E. & CORBANI, S. 2006. BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. Nature genetics, 38, 521-524.

STONE, E. M. 2007. Leber congenital amaurosis–a model for efficient genetic testing of heterogeneous disorders: LXIV Edward Jackson Memorial Lecture. American journal of ophthalmology, 144, 791-811. e6.

STONE, E. M., WEBSTER, A. R., VANDENBURGH, K., STREB, L. M., HOCKEY, R. R., LOTERY, A. J. & SHEFFIELD, V. C. 1998. Allelic variation in ABCR associated with Stargardt disease but not age-related macular degeneration. Nature Genetics, 20, 328-329.

STRAUSS, O. 2005. The retinal pigment epithelium in visual function. Physiological reviews, 85, 845-881.

STURTEVANT, A. H. 1913. The linear arrangement of six sex? linked factors in Drosophila, as shown by their mode of association. Journal of experimental zoology, 14, 43-59.

SUDHARSAN, R. & BELTRAN, W. A. 2019. Progress in gene therapy for rhodopsin autosomal dominant retinitis pigmentosa. Retinal Degenerative Diseases, 113-118.

SUDMANT, P. H., RAUSCH, T., GARDNER, E. J., HANDSAKER, R. E., ABYZOV, A., HUDDLESTON, J., ZHANG, Y., YE, K., JUN, G. & HSI-YANG FRITZ, M. 2015. An integrated map of structural variation in 2,504 human genomes. Nature, 526, 75-81.

SUGITA, S., IWASAKI, Y., MAKABE, K., KAMAO, H., MANDAI, M., SHIINA, T., OGASAWARA, K., HIRAMI, Y., KURIMOTO, Y. & TAKAHASHI, M. 2016. Successful Transplantation of Retinal Pigment Epithelial Cells from MHC Homozygote iPSCs in MHC-Matched Models. Stem Cell Reports, 7, 635-648.

SULLIVAN, L. S., BOWNE, S. J., BIRCH, D. G., HUGHBANKS-WHEATON, D., HECKENLIVELY, J. R., LEWIS, R. A., GARCIA, C. A., RUIZ, R. S., BLANTON, S. H. & NORTHRUP, H. 2006a. Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. Investigative ophthalmology & visual science, 47, 3052-3064.

SULLIVAN, L. S., BOWNE, S. J., SEAMAN, C. R., BLANTON, S. H., LEWIS, R. A., HECKENLIVELY, J. R., BIRCH, D. G., HUGHBANKS-WHEATON, D. & DAIGER, S. P. 2006b. Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. Investigative ophthalmology & visual science, 47, 4579-4588.

SUNG, Y.-C., YANG, C.-H., YANG, C.-M., LIN, C.-W., HUANG, D.-S., HUANG, Y.-S., HU, F.-R., CHEN, P.-L. & CHEN, T.-C. 2020. Genotypes predispose phenotypes—clinical features and genetic spectrum of ABCA4-associated retinal dystrophies. Genes, 11, 1421.

SUZUKI, K., TSUNEKAWA, Y., HERNANDEZ-BENITEZ, R., WU, J., ZHU, J., KIM, E. J., HATANAKA, F., YAMAMOTO, M., ARAOKA, T. & LI, Z. 2016. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature, 540, 144-149.

SWAIN, P. K., CHEN, S., WANG, Q.-L., AFFATIGATO, L. M., COATS, C. L., BRADY, K. D., FISHMAN, G. A., JACOBSON, S. G., SWAROOP, A. & STONE, E. 1997. Mutations in the cone-rod homeobox gene are associated with the cone-rod dystrophy photoreceptor degeneration. Neuron, 19, 1329-1336.

SWARUP, A., SAMUELS, I. S., BELL, B. A., HAN, J. Y. S., DU, J., MASSENZIO, E., ABEL, E. D., BOESZE-BATTAGLIA, K., PEACHEY, N. S. & PHILP, N. J. 2019. Modulating GLUT1 expression in retinal pigment epithelium decreases glucose levels

in the retina: impact on photoreceptors and Müller glial cells. Am J Physiol Cell Physiol, 316, C121-c133.

SWERDLOW, H. & GESTELAND, R. 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic acids research, 18, 1415-1419.

TABEBORDBAR, M., ZHU, K., CHENG, J. K., CHEW, W. L., WIDRICK, J. J., YAN, W. X., MAESNER, C., WU, E. Y., XIAO, R. & RAN, F. A. 2016. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science, 351, 407-411.

TABOR, H. K., RISCH, N. J. & MYERS, R. M. 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genetics, 3, 391-397.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. cell, 131, 861-872.

TAN, R., WANG, Y., KLEINSTEIN, S. E., LIU, Y., ZHU, X., GUO, H., JIANG, Q., ALLEN, A. S. & ZHU, M. 2014. An evaluation of copy number variation detection tools from whole-exome sequencing data. Human mutation, 35, 899-907.

TATOUR, Y. & BEN-YOSEF, T. 2020. Syndromic inherited retinal diseases: genetic, clinical and diagnostic aspects. Diagnostics, 10, 779.

TAUQEER, Z. & YONEKAWA, Y. 2018. Familial exudative vitreoretinopathy: pathophysiology, diagnosis, and management. The Asia-Pacific Journal of Ophthalmology, 7, 176-182.

TEARE, M. D. & BARRETT, J. H. 2005. Genetic linkage studies. The Lancet, 366, 1036-1044.

TEARE, M. D. & SANTIBAÑEZ KOREF, M. F. 2014. Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing. Briefings in functional genomics, 13, 378-383.

TEO, S. M., PAWITAN, Y., KU, C. S., CHIA, K. S. & SALIM, A. 2012. Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics, 28, 2711-2718.

THIADENS, A. A., PHAN, T. M. L., ZEKVELD-VROON, R. C., LEROY, B. P., VAN DEN BORN, L. I., HOYNG, C. B., KLAVER, C. C., ROOSING, S., POTT, J.-W. R. & VAN SCHOONEVELD, M. J. 2012. Clinical course, genetic etiology, and visual outcome in cone and cone–rod dystrophy. Ophthalmology, 119, 819-826.

THIADENS, A. A., SLINGERLAND, N. W., ROOSING, S., VAN SCHOONEVELD, M. J., VAN LITH-VERHOEVEN, J. J., VAN MOLL-RAMIREZ, N., VAN DEN BORN, L. I., HOYNG, C. B., CREMERS, F. P. & KLAVER, C. C. 2009. Genetic

etiology and clinical consequences of complete and incomplete achromatopsia. Ophthalmology, 116, 1984-1989. e1.

TIWARI, A., LEMKE, J., ALTMUELLER, J., THIELE, H., GLAUS, E., FLEISCHHAUER, J., NÜRNBERG, P., NEIDHARDT, J. & BERGER, W. 2016. Identification of novel and recurrent disease-causing mutations in retinal dystrophies using whole exome sequencing (WES): benefits and limitations. PloS one, 11, e0158692.

TOMASELLI, P. J., ROSSOR, A. M., HORGA, A., JAUNMUKTANE, Z., CARR, A., SAVERI, P., PISCOSQUITO, G., PAREYSON, D., LAURA, M. & BLAKE, J. C. 2017. Mutations in noncoding regions of GJB1 are a major cause of X-linked CMT. Neurology, 88, 1445-1453.

TOOMES, C., BOTTOMLEY, H. M., JACKSON, R. M., TOWNS, K. V., SCOTT, S., MACKEY, D. A., CRAIG, J. E., JIANG, L., YANG, Z. & TREMBATH, R. 2004. Mutations in LRP5 or FZD4 underlie the common familial exudative vitreoretinopathy locus on chromosome 11q. The American Journal of Human Genetics, 74, 721-730.

TOOMES, C., MARCHBANK, N. J., MACKEY, D. A., CRAIG, J. E., NEWBURY-ECOB, R. A., BENNETT, C. P., VIZE, C. J., DESAI, S. P., BLACK, G. C. & PATEL, N. 2001a. Spectrum, frequency and penetrance of OPA1 mutations in dominant optic atrophy. Human Molecular Genetics, 10, 1369-1378.

TOOMES, C., MARCHBANK, N. J., MACKEY, D. A., CRAIG, J. E., NEWBURY-ECOB, R. A., BENNETT, C. P., VIZE, C. J., DESAI, S. P., BLACK, G. C. M., PATEL, N., TEIMORY, M., MARKHAM, A. F., INGLEHEARN, C. F. & CHURCHILL, A. J. 2001b. Spectrum, frequency and penetrance of OPA1 mutations in dominant optic atrophy. Human Molecular Genetics, 10, 1369-1378.

TOUALBI, L., TOMS, M. & MOOSAJEE, M. 2020. USH2A-retinopathy: From genetics to therapeutics. Experimental Eye Research, 201, 108330.

TSCHERNUTTER, M., SCHLICHTENBREDE, F., HOWE, S., BALAGGAN, K., MUNRO, P., BAINBRIDGE, J., THRASHER, A., SMITH, A. & ALI, R. 2005. Long-term preservation of retinal function in the RCS rat model of retinitis pigmentosa following lentivirus-mediated gene therapy. Gene therapy, 12, 694-701.

TSIN, A., BETTS-OBREGON, B. & GRIGSBY, J. 2018. Visual cycle proteins: structure, function, and roles in human retinal disease. Journal of Biological Chemistry, 293, 13016-13021.

TUCKER, T., MARRA, M. & FRIEDMAN, J. M. 2009. Massively parallel sequencing: the next big thing in genetic medicine. The American Journal of Human Genetics, 85, 142-154.

TURNBULL, C., SCOTT, R. H., THOMAS, E., JONES, L., MURUGAESU, N., PRETTY, F. B., HALAI, D., BAPLE, E., CRAIG, C. & HAMBLIN, A. 2018. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. Bmj, 361.

TUUPANEN, S., GALL, K., SISTONEN, J., SAARINEN, I., KÄMPJÄRVI, K., WELLS, K., MERKKINIEMI, K., VON NANDELSTADH, P., SARANTAUS, L. & KÄNSÄKOSKI, J. 2022. Prevalence of RPGR-Mediated Retinal Dystrophy in an Unselected Cohort of Over 5000 Patients. Translational vision science & technology, 11, 6-6.

VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y. & THERMES, C. 2014. Ten years of next-generation sequencing technology. Trends in genetics, 30, 418-426.

VAN DIJK, E. L., JASZCZYSZYN, Y., NAQUIN, D. & THERMES, C. 2018. The third revolution in sequencing technology. Trends in Genetics, 34, 666-681.

VAN WIJK, E., PENNINGS, R. J., TE BRINKE, H., CLAASSEN, A., YNTEMA, H. G., HOEFSLOOT, L. H., CREMERS, F. P., CREMERS, C. W. & KREMER, H. 2004. Identification of 51 novel exons of the Usher syndrome type 2A (USH2A) gene that encode multiple conserved functional domains and that are mutated in patients with Usher syndrome type II. The American Journal of Human Genetics, 74, 738-744.

VÁZQUEZ-DOMÍNGUEZ, I., GARANTO, A. & COLLIN, R. W. 2019. Molecular therapies for inherited retinal diseases—current standing, opportunities and challenges. Genes, 10, 654.

VENTURINI, G., ROSE, A. M., SHAH, A. Z., BHATTACHARYA, S. S. & RIVOLTA, C. 2012. CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. PLoS genetics, 8, e1003040.

VERBAKEL, S. K., VAN HUET, R. A., BOON, C. J., DEN HOLLANDER, A. I., COLLIN, R. W., KLAVER, C. C., HOYNG, C. B., ROEPMAN, R. & KLEVERING, B. J. 2018. Non-syndromic retinitis pigmentosa. Progress in retinal and eye research, 66, 157-186.

VERBAKEL, S. K., VAN HUET, R. A., DEN HOLLANDER, A. I., GEERLINGS, M. J., KERSTEN, E., KLEVERING, B. J., KLAVER, C. C., PLOMP, A. S., WESSELING, N. L. & BERGEN, A. A. 2019. Macular dystrophy and cone-rod dystrophy caused by mutations in the RP1 gene: extending the RP1 disease spectrum. Investigative Ophthalmology & Visual Science, 60, 1192-1203.

VERMA, M., KULSHRESTHA, S. & PURI, A. 2017. Genome sequencing. Bioinformatics, 3-33.

VERVOORT, R., LENNON, A., BIRD, A. C., TULLOCH, B., AXTON, R., MIANO, M. G., MEINDL, A., MEITINGER, T., CICCODICOLA, A. & WRIGHT, A. F. 2000.

Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nature genetics, 25, 462-466.

VILLANUEVA, A., WILLER, J. R., BRYOIS, J., DERMITZAKIS, E. T., KATSANIS, N. & DAVIS, E. E. 2014. Whole exome sequencing of a dominant retinitis pigmentosa family identifies a novel deletion in PRPF31. Investigative ophthalmology & visual science, 55, 2121-2129.

VITHANA, E. N., ABU-SAFIEH, L., ALLEN, M. J., CAREY, A., PAPAIOANNOU, M., CHAKAROVA, C., AL-MAGHTHEH, M., EBENEZER, N. D., WILLIS, C. & MOORE, A. T. 2001. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13. 4 (RP11). Molecular cell, 8, 375-381.

VITHANA, E. N., ABU-SAFIEH, L., PELOSINI, L., WINCHESTER, E., HORNAN, D., BIRD, A. C., HUNT, D. M., BUSTIN, S. A. & BHATTACHARYA, S. S. 2003. Expression of PRPF31 mRNA in patients with autosomal dominant retinitis pigmentosa: a molecular clue for incomplete penetrance? Investigative ophthalmology & visual science, 44, 4204-4209.

VLASIUK, A. & ASARI, H. 2021. Feedback from retinal ganglion cells to the inner retina. PLOS ONE, 16, e0254611.

VOTRUBA, M., MOORE, A. T. & BHATTACHARYA, S. S. 1998. Clinical features, molecular genetics, and pathophysiology of dominant optic atrophy. Journal of medical genetics, 35, 793-800.

WADA, Y., ABE, T., TAKESHITA, T., SATO, H., YANASHIMA, K. & TAMAI, M. 2001. Mutation of human retinal fascin gene (FSCN2) causes autosomal dominant retinitis pigmentosa. Investigative ophthalmology & visual science, 42, 2395-2400.

WAKELING, M. N., LAVER, T. W., WRIGHT, C. F., DE FRANCO, E., STALS, K. L., PATCH, A.-M., HATTERSLEY, A. T., FLANAGAN, S. E. & ELLARD, S. 2019. Homozygosity mapping provides supporting evidence of pathogenicity in recessive Mendelian disease. Genetics in Medicine, 21, 982-986.

WANG, B., TSENG, E., BAYBAYAN, P., ENG, K., REGULSKI, M., JIAO, Y., WANG, L., OLSON, A., CHOUGULE, K. & BUREN, P. V. 2020. Variant phasing and haplotypic expression from long-read sequencing in maize. Communications Biology, 3, 78.

WANG, C., SUN, L., ZHENG, H. & HU, Y.-Q. 2016a. Detecting multiple variants associated with disease based on sequencing data of case–parent trios. Journal of human genetics, 61, 851-860.

WANG, J., WANG, Y., LI, S., XIAO, X., YI, Z., JIANG, Y., LI, X., JIA, X., WANG, P. & JIN, C. 2022. Clinical and genetic analysis of RDH12-associated retinopathy in 27

Chinese families: a hypomorphic allele leads to cone-rod dystrophy. Investigative ophthalmology & visual science, 63, 24-24.

WANG, J., ZHANG, V. W., FENG, Y., TIAN, X., LI, F.-Y., TRUONG, C., WANG, G., CHIANG, P.-W., LEWIS, R. A. & WONG, L.-J. C. 2014. Dependable and efficient clinical utility of target capture-based deep sequencing in molecular diagnosis of retinitis pigmentosa. Investigative ophthalmology & visual science, 55, 6213-6223.

WANG, J.-S. & KEFALOV, V. J. 2011. The cone-specific visual cycle. Progress in retinal and eye research, 30, 115-128.

WANG, L., ZHANG, J., CHEN, N., WANG, L., ZHANG, F., MA, Z., LI, G. & YANG, L. 2018. 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, 9, 360.

WANG, X., FENG, Y., LI, J., ZHANG, W., WANG, J., LEWIS, R. A. & WONG, L.-J. 2016b. Retinal diseases caused by mutations in genes not specifically associated with the clinical diagnosis. Plos one, 11, e0165405.

WANG, X., IANNACCONE, A. & JABLONSKI, M. M. 2004. Contribution of Müller cells toward the regulation of photoreceptor outer segment assembly. Neuron glia biology, 1, 291-296.

WANG, Y., LIU, S., ZHAI, Y., LIU, Y., WAN, X., WANG, W., WANG, F. & SUN, X. 2021. Identification of a novel RPGR mutation associated with X-linked cone-rod dystrophy in a Chinese family. BMC ophthalmology, 21, 1-10.

WARR, A., ROBERT, C., HUME, D., ARCHIBALD, A., DEEB, N. & WATSON, M. 2015. Exome sequencing: current and future perspectives. G3: Genes, Genomes, Genetics, 5, 1543-1550.

WASEEM, N. H., VACLAVIK, V., WEBSTER, A., JENKINS, S. A., BIRD, A. C. & BHATTACHARYA, S. S. 2007. Mutations in the gene coding for the pre-mRNA splicing factor, PRPF31, in patients with autosomal dominant retinitis pigmentosa. Investigative ophthalmology & visual science, 48, 1330-1334.

WATSON, C. M., CRINNION, L. A., LINDSAY, H., MITCHELL, R., CAMM, N., ROBINSON, R., JOYCE, C., TANTELES, G. A., HALLORAN, D. J. & PENA, S. D. 2021. Assessing the utility of long-read nanopore sequencing for rapid and efficient characterization of mobile element insertions. Laboratory Investigation, 101, 442-449.

WATSON, C. M., EL-ASRAG, M., PARRY, D. A., MORGAN, J. E., LOGAN, C. V., CARR, I. M., SHERIDAN, E., CHARLTON, R., JOHNSON, C. A. & TAYLOR, G. 2014. Mutation screening of retinal dystrophy patients by targeted capture from tagged pooled DNAs and next generation sequencing. PloS one, 9, e104281.

WATSON, J. D. & CRICK, F. H. The structure of DNA. Cold Spring Harbor symposia on quantitative biology, 1953. Cold Spring Harbor Laboratory Press, 123-131.

WEBB, T. R., PARFITT, D. A., GARDNER, J. C., MARTINEZ, A., BEVILACQUA, D., DAVIDSON, A. E., ZITO, I., THISELTON, D. L., RESSA, J. H. & APERGI, M. 2012. Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). Human molecular genetics, 21, 3647-3654.

WEISSCHUH, N., FELDHAUS, B., KHAN, M. I., CREMERS, F. P., KOHL, S., WISSINGER, B. & ZOBOR, D. 2018. Molecular and clinical analysis of 27 German patients with Leber congenital amaurosis. PLoS One, 13, e0205380.

WEISSCHUH, N., MAZZOLA, P., HEINRICH, T., HAACK, T., WISSINGER, B., TONAGEL, F. & KELBSCH, C. 2020a. First submicroscopic inversion of the OPA1 gene identified in dominant optic atrophy–a case report. BMC Medical Genetics, 21, 1-6.

WEISSCHUH, N., OBERMAIER, C. D., BATTKE, F., BERND, A., KUEHLEWEIN, L., NASSER, F., ZOBOR, D., ZRENNER, E., WEBER, E. & WISSINGER, B. 2020b. Genetic architecture of inherited retinal degeneration in Germany: A large cohort study from a single diagnostic center over a 9-year period. Human Mutation, 41, 1514-1527.

WEISSCHUH, N., SCHIMPF-LINZENBOLD, S., MAZZOLA, P., KIENINGER, S., XIAO, T., KELLNER, U., NEUHANN, T., KELBSCH, C., TONAGEL, F. & WILHELM, H. 2021. Mutation spectrum of the OPA1 gene in a large cohort of patients with suspected dominant optic atrophy: Identification and classification of 48 novel variants. Plos one, 16, e0253987.

WEISSCHUH, N., STURM, M., BAUMANN, B., AUDO, I., AYUSO, C., BOCQUET, B., BRANHAM, K., BROOKS, B. P., CATALÁ-MORA, J. & GIORDA, R. 2020c. Deep-intronic variants in CNGB3 cause achromatopsia by pseudoexon activation. Human mutation, 41, 255-264.

WENGER, A. M., PELUSO, P., ROWELL, W. J., CHANG, P.-C., HALL, R. J., CONCEPCION, G. T., EBLER, J., FUNGTAMMASAN, A., KOLESNIKOV, A. & OLSON, N. D. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nature biotechnology, 37, 1155-1162.

WHEWAY, G., SCHMIDTS, M., MANS, D. A., SZYMANSKA, K., NGUYEN, T.-M. T., RACHER, H., PHELPS, I. G., TOEDT, G., KENNEDY, J. & WUNDERLICH, K. A. 2015. An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes. Nature cell biology, 17, 1074-1087.

WHITE, D. R., GANESH, A., NISHIMURA, D., RATTENBERRY, E., AHMED, S., SMITH, U. M., PASHA, S., RAEBURN, S., TREMBATH, R. C. & RAJAB, A. 2007. Autozygosity mapping of Bardet–Biedl syndrome to 12q21. 2 and confirmation of FLJ23560 as BBS10. European journal of human genetics, 15, 173-178.

WILLOUGHBY, C. E., PONZIN, D., FERRARI, S., LOBO, A., LANDAU, K. & OMIDI, Y. 2010. Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function–a review. Clinical & Experimental Ophthalmology, 38, 2-11.

WRIGHT, A. F., CHAKAROVA, C. F., EL-AZIZ, A., MAI, M. & BHATTACHARYA, S. S. 2010. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. Nature Reviews Genetics, 11, 273-284.

WU, J.-H., LIU, J.-H., KO, Y.-C., WANG, C.-T., CHUNG, Y.-C., CHU, K.-C., LIU, T.-T., CHAO, H.-M., JIANG, Y.-J. & CHEN, S.-J. 2016. Haploinsufficiency of RCBTB1 is associated with Coats disease and familial exudative vitreoretinopathy. Human molecular genetics, 25, 1637-1647.

XIANG, Q., CAO, Y., XU, H., YANG, Z., TANG, L., XIANG, J., LI, J., DENG, H. & YUAN, L. 2021. Novel MFSD8 Variants in a Chinese Family with Nonsyndromic Macular Dystrophy. Journal of Ophthalmology, 2021.

XIAO, T. & ZHOU, W. 2020. The third generation sequencing: the advanced approach to genetic diseases. Translational pediatrics, 9, 163.

XU, L., BROWN, E. E., KEUTHAN, C. J., GUBBI, H., GRELLIER, E.-K., ROGER, J., SWAROOP, A., DU, J. & ASH, J. D. 2020. AMP-activated-protein kinase (AMPK) is an essential sensor and metabolic regulator of retinal neurons and their integrated metabolism with RPE. bioRxiv, 2020.05.22.109165.

YAHYA, S., SMITH, C. E. L., POULTER, J. A., MCKIBBIN, M., ARNO, G., ELLINGFORD, J., KÄMPJÄRVI, K., KHAN, M. I., CREMERS, F. P. M., HARDCASTLE, A. J., CASTLE, B., STEEL, D. H. W., WEBSTER, A. R., BLACK, G. C., EL-ASRAG, M. E., ALI, M., TOOMES, C., INGLEHEARN, C. F., INGRAM, S., TAYLOR, R., MANSON, F., SERGOUNIOTIS, P., PONTIKOS, N., CHEETHAM, M., FIORENTINO, A., DOWNES, S., YU, J., HALFORD, S., BROADGATE, S., VAN HEYNINGEN, V., AMBROSE, J. C., ARUMUGAM, P., BEVERS, R., BLEDA, M., BOARDMAN-PRETTY, F., BOUSTRED, C. R., BRITTAIN, H., CAULFIELD, M. J., CHAN, G. C., ELGAR, G., FOWLER, T., GIESS, A., HAMBLIN, A., HENDERSON, S., HUBBARD, T. J. P., JACKSON, R., JONES, L. J., KASPERAVICIUTE, D., KAYIKCI, M., KOUSATHANAS, A., LAHNSTEIN, L., LEIGH, S. E. A., LEONG, I. U. S., LOPEZ, J. F., MALEADY-CROWE, F., MCENTAGART, M., MINNECI, F., MOUTSIANAS, L., MUELLER, M., MURUGAESU, N., NEED, A. C., O'DONOVAN, P., ODHAMS, C. A., PATCH, C.,

PEREIRA, M. B., PEREZ-GIL, D., PULLINGER, J., RAHIM, T., RENDON, A., ROGERS, T., SAVAGE, K., SAWANT, K., SCOTT, R. H., SIDDIQ, A., SIEGHART, A., SMITH, S. C., SOSINSKY, A., STUCKEY, A., TANGUY, M., TAYLOR TAVARES, A. L., THOMAS, E. R. A., THOMPSON, S. R., TUCCI, A., WELLAND, M. J., WILLIAMS, E., WITKOWSKA, K. & WOOD, S. M. 2023. Late-Onset Autosomal Dominant Macular Degeneration Caused by Deletion of the CRX Gene. Ophthalmology, 130, 68-76.

YAMAMOTO, K., SATO, T., MATSUI, T., SATO, M., OKADA, T., YOSHIDA, H., HARADA, A. & MORI, K. 2007. Transcriptional induction of mammalian ER quality control proteins is mediated by single or combined action of ATF6α and XBP1. Developmental cell, 13, 365-376.

YANG, C., GEORGIOU, M., ATKINSON, R., COLLIN, J., AL-AAMA, J., NAGARAJA-GRELLSCHEID, S., JOHNSON, C., ALI, R., ARMSTRONG, L. & MOZAFFARI-JOVIN, S. 2021a. Pre-mRNA Processing Factors and Retinitis Pigmentosa: RNA Splicing and Beyond. Frontiers in cell and developmental biology, 2062.

YANG, L., YIN, X., WU, L., CHEN, N., ZHANG, H., LI, G. & MA, Z. 2013. Targeted exome capture and sequencing identifies novel PRPF31 mutations in autosomal dominant retinitis pigmentosa in Chinese families. BMJ open, 3, e004030.

YANG, S., ZHOU, J. & LI, D. 2021b. Functions and diseases of the retinal pigment epithelium. Frontiers in Pharmacology, 1976.

YAO, X., SON, T., KIM, T.-H. & LU, Y. 2018. Functional optical coherence tomography of retinal photoreceptors. Experimental Biology and Medicine, 243, 1256-1264.

YATES, J. R., SEPP, T., MATHARU, B. K., KHAN, J. C., THURLBY, D. A., SHAHID, H., CLAYTON, D. G., HAYWARD, C., MORGAN, J. & WRIGHT, A. F. 2007. Complement C3 variant and the risk of age-related macular degeneration. New England Journal of Medicine, 357, 553-561.

YI, Z., XIAO, X., LI, S., SUN, W. & ZHANG, Q. 2019. Pathogenicity discrimination and genetic test reference for CRX variants based on genotype-phenotype analysis. Experimental Eye Research, 189, 107846.

YILDIRIM, Y., OURIACHI, T., WOEHLBIER, U., OUAHIOUNE, W., BALKAN, M., MALIK, S. & TOLUN, A. 2018. Linked homozygous BMPR1B and PDHA2 variants in a consanguineous family with complex digit malformation and male infertility. European Journal of Human Genetics, 26, 876-885.

YOUNG, B. K., RAMAKRISHNAN, C., GANJAWALA, T., WANG, P., DEISSEROTH, K. & TIAN, N. 2021. An uncommon neuronal class conveys visual

signals from rods and cones to retinal ganglion cells. Proceedings of the National Academy of Sciences, 118, e2104884118.

YU-WAI-MAN, P. & CHINNERY, P. F. 2013. Dominant optic atrophy: novel OPA1 mutations and revised prevalence estimates. Ophthalmology, 120, 1712-1712. e1.

YU-WAI-MAN, P., GRIFFITHS, P. G., GORMAN, G., LOURENCO, C., WRIGHT, A., AUER-GRUMBACH, M., TOSCANO, A., MUSUMECI, O., VALENTINO, M. & CAPORALI, L. 2010. Multi-system neurological disease is common in patients with OPA1 mutations. Brain, 133, 771-786.

YUE, L., WEILAND, J. D., ROSKA, B. & HUMAYUN, M. S. 2016. Retinal stimulation strategies to restore vision: Fundamentals and systems. Progress in retinal and eye research, 53, 21-47.

YUSUF, I. H., CHARBEL ISSA, P. & AHN, S. J. 2022. Novel imaging techniques for hydroxychloroquine retinopathy. Front Med (Lausanne), 9, 1026934.

ZAGOZEWSKI, J., ZHANG, Q. & EISENSTAT, D. 2014. Genetic regulation of vertebrate eye development. Clinical genetics, 86, 453-460.

ZAMPAGLIONE, E., KINDE, B., PLACE, E. M., NAVARRO-GOMEZ, D., MAHER, M., JAMSHIDI, F., NASSIRI, S., MAZZONE, J. A., FINN, C. & SCHLEGEL, D. 2020. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. Genetics in Medicine, 22, 1079-1087.

ZERNANT, J., LEE, W., COLLISON, F. T., FISHMAN, G. A., SERGEEV, Y. V., SCHUERCH, K., SPARROW, J. R., TSANG, S. H. & ALLIKMETS, R. 2017. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. Journal of medical genetics, 54, 404-412.

ZHANG, F., KHAJAVI, M., CONNOLLY, A. M., TOWNE, C. F., BATISH, S. D. & LUPSKI, J. R. 2009. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nature genetics, 41, 849-853.

ZHANG, L., SULLIVAN, P. S., GOODMAN, J. C., GUNARATNE, P. H. & MARCHETTI, D. 2011. MicroRNA-1258 Suppresses Breast Cancer Brain Metastasis by Targeting HeparanaseMicro RNA-1258 Suppresses Breast Cancer Brain Metastasis. Cancer research, 71, 645-654.

ZHANG, L., SUN, Z., ZHAO, P., HUANG, L., XU, M., YANG, Y., CHEN, X., LU, F., ZHANG, X. & WANG, H. 2018. Whole-exome sequencing revealed HKDC1 as a candidate gene associated with autosomal-recessive retinitis pigmentosa. Human molecular genetics, 27, 4157-4168.

ZHANG, L., ZHANG, X., XU, H., HUANG, L., ZHANG, S., LIU, W., YANG, Y., FEI, P., LI, S. & YANG, M. 2020. Exome sequencing revealed Notch ligand JAG1 as a novel candidate gene for familial exudative vitreoretinopathy. Genetics in Medicine, 22, 77-84.

ZHANG, Q. 2016. Retinitis pigmentosa: progress and perspective. The Asia-Pacific Journal of Ophthalmology, 5, 265-271.

ZHANG, Q., ACLAND, G. M., WU, W. X., JOHNSON, J. L., PEARCE-KELLING, S., TULLOCH, B., VERVOORT, R., WRIGHT, A. F. & AGUIRRE, G. D. 2002. Different RPGR exon ORF15 mutations in Canids provide insights into photoreceptor cell degeneration. Human molecular genetics, 11, 993-1003.

ZHAO, C., YASUMURA, D., LI, X., MATTHES, M., LLOYD, M., NIELSEN, G., AHERN, K., SNYDER, M., BOK, D., DUNAIEF, J. L., LAVAIL, M. M. & VOLLRATH, D. 2011. mTOR-mediated dedifferentiation of the retinal pigment epithelium initiates photoreceptor degeneration in mice. J Clin Invest, 121, 369-83.

ZHAO, L., WANG, F., WANG, H., LI, Y., ALEXANDER, S., WANG, K., WILLOUGHBY, C. E., ZANEVELD, J. E., JIANG, L. & SOENS, Z. T. 2015. Next-generation sequencing-based molecular diagnosis of 82 retinitis pigmentosa probands from Northern Ireland. Human genetics, 134, 217-230.

ZHOU, X., REN, L., LI, Y., ZHANG, M., YU, Y. & YU, J. 2010. The next-generation sequencing technology: a technology review and future perspective. Science China Life Sciences, 53, 44-57.

ZHU, T., SHEN, Y., SUN, Z., HAN, X., WEI, X., LI, W., LU, C., CHENG, T., ZOU, X. & LI, H. 2022. Clinical and molecular features of a Chinese cohort with syndromic and nonsyndromic retinal dystrophies related to the CEP290 gene. American Journal of Ophthalmology.

ZICCARDI, L., CORDEDDU, V., GADDINI, L., MATTEUCCI, A., PARRAVANO, M., MALCHIODI-ALBEDI, F. & VARANO, M. 2019. Gene therapy in retinal dystrophies. International journal of molecular sciences, 20, 5722.

Appendix A

A.1 List of UNIX commands used in WES analysis

- **Trim the adaptors and do quality control**

$ trim_galore -q 20 --fastqc_args "--outdir /data/bssy/path/to/outdir" --illumina --gzip -o /data/bssy/path/to/outdir --length 20 --paired /data/bssy/path/to/file/sample_R1_001.fastq.gz /data/bssy/path/to/file/sample_R2_001.fastq.gz

- **Align the sample to the human genome**

$ bwa mem -t 12 -M /home/ref/b37/human_g1k_v37.fasta /pathto/val_files_from_previous_step_R1.gz /pathto/val_files_from_previous_step_R2.gz -v 1 -R '@RG\tID:Add_sample_ID\tSM:Add_sample_ID\tPL:Illumina\tPU:HiSeq3000\tLB:$Samplename_exome\' -M | samtools view -Sb - > /path/sample_bwa.bam

- **Next sort the alignment**

$ java -Xmx4g -jar /home/picard/picard-tools-2.5.0/picard.jar SortSam I=/path/sample_bwa.bam O=/path/sample_bwa.sort.bam SO=coordinate CREATE_INDEX=TRUE

- **remove original bam to save space:**

$ rm -i /path/sample_bwa.bam

- **Mark PCR duplicates**

$ java -Xmx4g -jar /home/picard/picard-tools-2.5.0/picard.jar MarkDuplicates I=sample_bwa.sort.bam O=sample_bwa.sort.dedup.bam M=sample_bwa.sort.metrics CREATE_INDEX=TRUE

- **Delete pre-deduplicated bam to save space**

$ rm -i /path/sample_bwa.sort.bam

- **Create indel realigner targets**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T RealignerTargetCreator -R /home/ref/b37/human_g1k_v37.fasta -known /home/ref/b37/1000G_phase1.indels.b37.vcf -known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I sample_bwa.sort.dedup.bam -o sample_bwa.sort.dedup.intervals

- **Perform indel realignment**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T IndelRealigner -R /home/ref/b37/human_g1k_v37.fasta -known /home/ref/b37/1000G_phase1.indels.b37.vcf -known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I sample_bwa.sort.dedup.bam -targetIntervals sample_bwa.sort.dedup.intervals -o sample_bwa.sort.dedup.indelrealn.bam

- **Delete pre-indelrealn bam and gzip interval file to save space:**

$ rm -i sample_bwa.sort.dedup.bam

$ gzip sample_bwa.sort.dedup.intervals

- **Perform base quality recalibration**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T BaseRecalibrator -R /home/ref/b37/human_g1k_v37.fasta -knownSites /home/ref/b37/1000G_phase1.indels.b37.vcf -knownSites /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -knownSites /home/ref/b37/dbSnp146.b37.vcf.gz -I sample_bwa.sort.dedup.indelrealn.bam -o sample_bwa.sort.dedup.indelrealn.recal.grp -nct 6

- **print reads**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T PrintReads -R /home/ref/b37/human_g1k_v37.fasta -I sample_bwa.sort.dedup.indelrealn.bam -BQSR sample_bwa.sort.dedup.indelrealn.recal.grp -o sample_bwa.sort.dedup.indelrealn.recal.bam

- **Delete old bam (the non-recal file)**

$ rm -i sample_bwa.sort.dedup.indelrealn.bam

- **Generate g.vcf file for each sample using Haplotype Caller**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T HaplotypeCaller --emitRefConfidence GVCF --variant_index_type LINEAR --variant_index_parameter 128000 -R /home/ref/b37/human_g1k_v37.fasta -D /home/ref/b37/dbSnp146.b37.vcf.gz -stand_call_conf 30 -stand_emit_conf 10 -I /data/bssy/results/**sample**_bwa.sort.dedup.indelrealn.recal.bam -o /data/bssy/results/**sample**.g.vcf

- **Convert the raw.g.vcf to a raw.vcf (for autozygosity)**

$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T GenotypeGVCFs -R /home/ref/b37/human_g1k_v37.fasta -D /home/ref/b37/dbSnp146.b37.vcf.gz -stand_call_conf 30 -stand_emit_conf 10 -V

/data/bssy/results/sample.g.vcf -o /data/bssy/results/sample.combined.raw.vcf -nda --showFullBamList -nt 8

- **Split Variants SNP**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T SelectVariants -R /home/ref/b37/human_g1k_v37.fasta -selectType SNP --variant /data/bssy/path/to/file.vcf -o /data/bssy/path/to/file-snps.vcf

- **Split Variants INDEL**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T SelectVariants -R /home/ref/b37/human_g1k_v37.fasta --variant /data/bssy/path/to/file.vcf -selectType INDEL -selectType MNP -o /data/bssy/path/to/file-indels.vcf

- **Hard filtering SNP**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T VariantFiltration -R /home/ref/b37/human_g1k_v37.fasta -V /data/bssy/path/to/file-snps.vcf --filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MappingQualityRankSum < -12.5" --filterName "snp_hard_filter" -o /data/bssy/path/to/file.fltd-snps.vcf

- **Hard filtering INDEL**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T VariantFiltration -R /home/ref/b37/human_g1k_v37.fasta -V /data/bssy/path/to/file-indels.vcf --filterExpression "QD < 20 || FS > 200.0 || ReadPosRankSum < -20.0" --filterName "indel_hard_filter" -o /data/bssy/path/to/file.fltd-indels.vcf

- **Combine Filtered Variants**

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T CombineVariants -R /home/ref/b37/human_g1k_v37.fasta --variant /data/bssy/path/to/file.fltd-snps.vcf --variant /data/bssy/path/to/file.fltd-indels.vcf -o /data/bssy/path/to/file.fltd-combined.vcf --genotypemergeoption UNSORTED

- **Filter variants in gnomad >/= 1%**

$ perl /home/vcfhacks-v0.2.0/filterVcfOnVcf.pl -f /home/ref/ExAC/gnomad.exomes.r2.0.1.sites.vcf.gz -w -y 0.01 -b -i /data/bssy/path/to/file.fltd-combined.vcf -o /data/bssy/path/to/file.fltd-combined.gnomAD.vcf

- **Annotate variants with VEP**

$ perl /home/variant_effect_predictor/variant_effect_predictor.pl --offline --vcf --everything --dir_cache /home/variant_effect_predictor/cache_b37 --dir_plugins /home/variant_effect_predictor/cache_b37/Plugins --plugin SpliceConsensus --fasta /home/variant_effect_predictor/cache_b37/homo_sapiens/92_GRCh37/Homo_sapiens.GRCh37.75.dna.primary_assembly.fa.gz -i /data/bssy/path/to/file.fltd-combined.gnomAD.vcf -o /data/bssy/path/to/file.fltd-combined_gnomad.vep.vcf -fork 6

- **For recessive and X-linked cases**
- **Select Biallelic and X linked variants**

$ perl /home/vcfhacks-v0.2.0/findBiallelic.pl -i /data/bs16gn/path/to/file.fltd-combined_gnomad.vep.vcf --x_linked 2 -s sample_ID --consensus_splice_site -n 1 -o /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom

- **Rank on CADD Score**

$ perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl -c /data/shared/cadd/v1.3/*.gz -i /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom -o /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3 -n /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3_NOTFOUND.tsv --progress -d

- **GeneAnnotator_vcfhacks**

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl -d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb --i /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3 -o /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3.geneanno

- **AnnovcfToSimple_vcfhacks_xlsx with -f gives only the functional variants**

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3.geneanno --vep --gene_anno --functional -o /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3.geneanno.simple.xlsx

- **AnnovcfToSimple_vcfhacks_xlsx with -f gives only the functional variants canonical_only**

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3.geneanno --vep --gene_anno --canonical_only --functional -o /data/bssy/path/to/file.fltd-combined_AR_gnomad.vep.hom.cadd1.3.geneanno.simple.canonicalOnly.xlsx

- **For dominant cases**
- **Select dominant variants**

$ perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl -i /data/bssy/path/to/file.fltd-combined_gnomad.vep.vcf --consensus_splice_site -s sample_ID -o /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom

- **Rank on CADD Score**

$ perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl -c /data/shared/cadd/v1.3/*.gz -i /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom -o /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom.cadd1.3 -n /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom.cadd1.3_NOTFOUND.tsv --progress -d

- **GeneAnnotator_vcfhacks**

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl -d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb --i /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom.cadd1.3 -o /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom.cadd1.3.geneanno

- **AnnovcfToSimple_vcfhacks_xlsx with only the functional variants**

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom.cadd1.3.geneanno --vep --gene_anno --functional -o /data/bssy/path/to/file.fltd-combined_AD_gnomad.vep.hom.cadd1.3.geneanno.simple.xlsx

A.2 List of R commands used in ExomeDepth analysis

```
$ R

> library(“ExomeDepth”)

> data(exons.hg19)

> print(head(exons.hg19))

> Samar_bam_files <- c(“/path/to/File1.bam”, “/path/to/File2.bam”,”/path/to/File3.bam”, etc.)

> Samar.counts <- getBamCounts(bed.frame = exons.hg19, bam.files = my_bam_files,
include.chr = FALSE, referenceFasta = "/nobackup/medmela/ref/human_g1k_v37.fasta")

> ExomeCount.dafr <- as(Samar.counts[, colnames(Samar.counts)], 'data.frame')

> print(head(ExomeCount.dafr))

> Samar.test <- Samar.counts$File1.bam

> Samar.ref.samples <-c(“File2.bam”,”File3.bam”, etc.)

> Samar.reference.set<-as.matrix(ExomeCount.dafr[,Samar.ref.samples])

> Samar.choice<-select.reference.set(test.counts=Samar.test,reference.counts=
Samar.reference.set,bin.length=(ExomeCount.dafr$end -
ExomeCount.dafr$start)/1000,n.bins.reduced=10000)

> print(Samar.choice[[1]])

> Samar.matrix <-as.matrix( ExomeCount.dafr[, Samar.choice$reference.choice, drop =
FALSE])

> Samar.reference.selected<-apply(X= Samar.matrix,MAR=1,FUN=sum)

> all.exons <-new('ExomeDepth',test= Samar.test, reference=
Samar.reference.selected,formula='cbind(test, reference)~1')

> all.exons<-CallCNVs(x=all.exons,transition.probability=10^-
4,chromosome=ExomeCount.dafr$chromosome,start=ExomeCount.dafr$start,end=ExomeCoun
t.dafr$end,name=ExomeCount.dafr$exon)

> head(all.exons@CNV.calls)

> data(Conrad.hg19)

> head(Conrad.hg19.common.CNVs)

> all.exons<-AnnotateExtra(x=all.exons, reference.annotation=Conrad.hg19.common.CNVs,
min.overlap=0.5, column.name='Conrad.hg19')
```

```
> print(head(all.exons@CNV.calls))

> exons.hg19.GRanges <-
GenomicRanges::GRanges(seqnames=exons.hg19$chromosome,IRanges::IRanges(start=exons.
hg19$start,end=exons.hg19$end),names=exons.hg19$name)

> all.exons <- AnnotateExtra(x=all.exons, reference.annotation=exons.hg19.GRanges,
min.overlap=0.0001, column.name='exons.hg19')

> all.exons@CNV.calls[3:6,]

> output.file <- 'exome_calls_File1.csv'

> write.csv(file=output.file,x=all.exons@CNV.calls,row.names=FALSE)
```

A.3 List of commands used in MinION analysis

- **Demultiplex using Guppy v.5.0.16**

singularity run --nv --bind /nobackup:/nobackup /nobackup/containers/guppy-gpu-5.0.16.simg guppy_basecaller -x "cuda:0" -q 0 --compress_fastq --trim_barcodes -i /nobackup/medcmwa/20211005_Run97/Samar_Run2/20211005_1105_MN27859_FAR17792_4b5b04e0/fast5/ -s /nobackup/medcmwa/20211005_Run97/Samar_Run2/20211005_1105_MN27859_FAR17792_4b5b04e0/fastq_barcode/ --flowcell FLO-MIN106 --kit SQK-LSK109 --barcode_kits EXP-PBC096

- **Adaptor trimming using Porechop v.0.2.4**

porechop -i ${BARCODE_DIR}/${BARCODE_ID}.merged.fastq.gz -o ${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.fastq.gz >> ${BARCODE_DIR}/${BARCODE_ID}.porechop.log 2>> ${BARCODE_DIR}/${BARCODE_ID}.porechop.log

- **Filtering based on read length using NanoFilt v.2.8.0**

gunzip -c ${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.fastq.gz | NanoFilt -l ${LEN_MIN} --maxlength ${LEN_MAX} --logfile ${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.nanofilt.length.log 2> ${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.nanofilt.length.log | gzip > ${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.${LEN_MIN}_${LEN_MAX}.fastq.gz

- **Filtering based on read quality using NanoFilt v.2.8.0**

gunzip -c ${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.${LEN_MIN}_${LEN_MAX}.fastq.gz | NanoFilt -q ${QUAL_MIN} --logfile ${BARCODE_DIR}/${BARCODE_ID}.Q${QUAL_MIN}.nanofilt.qual.log 2> ${BARCODE_DIR}/${BARCODE_ID}.Q${QUAL_MIN}.nanofilt.qual.log | gzip > ${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.fastq.gz

- **Align filtered read to human refernce genome (GRCh37/hg19) using Minimap2 v.2.18**

```
minimap2 -ax map-ont -t 4 /nobackup/bssy/resources/hg19.minimap2.ont
${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.${LEN_MIN}_${LEN_MAX}.Q${QUA
L_MIN}.fastq.gz >
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.sa
m 2>>
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.mi
nimap2.log
```

- **Sort the SAM file and output to BAM**

```
samtools view -bh
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.sa
m | samtools sort -o
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.sor
t.bam
```

- **Index the BAM file**

```
samtools index
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.sor
t.bam
```

- **Run NanoStat v.1.5 on raw fastq**

```
NanoStat --fastq ${BARCODE_DIR}/fastq_run* >>
${BARCODE_DIR}/${BARCODE_ID}.nanostat.RawFq.txt
```

- **Run NanoStat v.1.5 on length filtered fastq**

```
NanoStat --fastq
${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.${LEN_MIN}_${LEN_MAX}.fastq.gz
>>
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.nanostat.LenFiltFq.txt
```

- **Run NanoStat v.1.5 on quality filtered fastq**

```
NanoStat --fastq
${BARCODE_DIR}/${BARCODE_ID}.adapt_trim.${LEN_MIN}_${LEN_MAX}.Q${QUA
L_MIN}.fastq.gz >>
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.na
nostat.LenFilt_QualFilt_Fq.txt
```

- **Run NanoStat v.1.5 on final BAM**

```
NanoStat --bam
${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.sor
t.bam >>
```

${BARCODE_DIR}/${BARCODE_ID}.${LEN_MIN}_${LEN_MAX}.Q${QUAL_MIN}.nanostat.BAM.txt

A.4 Sequence-specific primers used in *RPGR ORF15* Sanger sequencing

Primers	Primer sequence 5'-3'	Location	Fragment size (bp)
*RPGR*_Ex15-1F	N13-GGTATGGCAGGAAATTGATTG	c.1754-102_1754-82	807
*RPGR*_Ex15-1R	N13-CTACTTCCCCTCCCTCTACTTC	c.2437_2458	
*RPGR*_Ex15-2F	N13-GAAGAAGTGGAGGGAGAACGT	c.2284_2304	596
*RPGR*_Ex15-2R	N13-TCCTCTTCCCCCTCCCA	c.2863_2879	
*RPGR*_Ex15-3F	N13-GGGAAGAGGAGGAAGGAGAAT	c.2843_2863	456
*RPGR*_Ex15-3R	N13-GTTTGCCATATTTCACAGATCC	c.3277_3298	
*RPGR*_Ex15-4F	N13-AAGGCAGGATGGAGAGGAGT	c.3234_3253	290
*RPGR*_Ex15-4R	N13-GCCCGTTATATGCAAGGC	c.*47_*64	

Table A.1: PCR/Sequencing primers used by Manchester reference laboratory in RPGR-ORF15 Sanger sequencing.

Appendix B

B.1 List of RetNet genes used in WES data filtration

ABCA4	*BBS1*	*CDH3*	*CORD1*	*FHASD*	*IFT27*	*MAPKAPK3*
ABCC6	*BBS10*	*CDHR1*	*CORD17*	*FLVCR1*	*IFT81*	*MCDR3*
ABHD12	*BBS12*	*CEP164*	*CORD4*	*FSCN2*	*IMPDH1*	*MCDR4*
ACBD5	*BBS2*	*CEP19*	*CORD8*	*FZD4*	*IMPG1*	*MCDR5*
ACO2	*BBS4*	*CEP250*	*CRB1*	*GDF6*	*IMPG2*	*MDDC*
ADAM9	*BBS5*	*CEP290*	*CRX*	*GNAT1*	*INPP5E*	*MERTK*
ADAMTS18	*BBS7*	*CEP78*	*CSPP1*	*GNAT2*	*INVS*	*MFN2*
ADGRV1	*BBS9*	*CERKL*	*CTNNA1*	*GNB3*	*IQCB1*	*MFRP*
ADIPOR1	*BCAMD*	*CFB*	*CYP4V2*	*GNPTG*	*ITM2B*	*MFSD8*
AFG3L2	*BEST1*	*CFH*	*DFNB31*	*GPR125*	*JAG1*	*MIR204*
AGBL5	*C12orf65*	*CHM*	*DHDDS*	*GPR179*	*KCNJ13*	*MKKS*
AHI1	*C1QTNF5*	*CIB2*	*DHX38*	*GRK1*	*KCNV2*	*MKS1*
AHR	*C2*	*CLN3*	*DMD*	*GRM6*	*KIAA1549*	*MRST*
AIPL1	*C21orf2*	*CLRN1*	*DRAM2*	*GUCA1A*	*KIF11*	*MT-ATP6*
ALMS1	*C2orf71*	*CLUAP1*	*DTHD1*	*GUCA1B*	*KIZ*	*MT-TH*
ARHGEF18	*C3*	*CNGA1*	*EFEMP1*	*GUCY2D*	*KLHL7*	*MT-TL1*
ARL2BP	*C8orf37*	*CNGA3*	*ELOVL4*	*HARS*	*KSS*	*MT-TP*
ARL3	*CA4*	*CNGB1*	*EMC1*	*HGSNAT*	*LAMA1*	*MT-TS2*
ARL6	*CABP4*	*CNGB3*	*ERCC6*	*HK1*	*LCA5*	*MTTP*
ARMS2	*CACD*	*CNNM4*	*ESPN*	*HMCN1*	*LHON*	*MVK*
ARSG	*CACNA2D4*	*COD2*	*EVR3*	*HMX1*	*LRAT*	*MYO7A*
ASRGL1	*CAPN5*	*CODA1*	*EXOSC2*	*HTRA1*	*LRIT3*	*NBAS*
ATF6	*CC2D2A*	*COL11A1*	*EYS*	*IDH3B*	*LRP5*	*NDP*
ATXN7	*CCT2*	*COL2A1*	*FAM161A*	*IFT140*	*LZTFL1*	*NEK2*
BBIP1	*CDH23*	*COL9A1*	*FBLN5*	*IFT172*	*MAK*	*NEUROD1*

Cont. List of RetNet genes used in WES data filtration

NMNAT1	*PDE6A*	*PRPF6*	*RP1L1*	*SPP2*	*USH1G*
NPHP1	*PDE6B*	*PRPF8*	*RP2*	*TEAD1*	*USH1H*
NPHP3	*PDE6C*	*PRPH2*	*RP22*	*TIMM8A*	*USH1K*
NPHP4	*PDE6G*	*PRPS1*	*RP24*	*TIMP3*	*USH2A*
NR2E3	*PDE6H*	*RAB28*	*RP29*	*TLR3*	*VCAN*
NR2F1	*PDZD7*	*RAX2*	*RP32*	*TLR4*	*VRD1*
NRL	*PEX1*	*RB1*	*RP34*	*TMEM126A*	*WDPCP*
NYX	*PEX2*	*RBP3*	*RP6*	*TMEM216*	*WDR19*
OAT	*PEX7*	*RBP4*	*RP63*	*TMEM237*	*WFS1*
OFD1	*PGK1*	*RCBTB1*	*RP9*	*TOPORS*	*WFS2*
OPA1	*PHYH*	*RCD1*	*RPE65*	*TREX1*	*ZNF408*
OPA2	*PITPNM3*	*RD3*	*RPGR*	*TRIM32*	*ZNF423*
OPA3	*PLA2G5*	*RDH11*	*RPGRIP1*	*TRNT1*	*ZNF513*
OPA4	*PLK4*	*RDH12*	*RPGRIP1L*	*TRPM1*	
OPA5	*PNPLA6*	*RDH5*	*RS1*	*TSPAN12*	
OPA6	*POC1B*	*REEP6*	*RTN4IP1*	*TTC8*	
OPA8	*POC5*	*RGR*	*SAG*	*TTLL5*	
OPN1LW	*POMGNT1*	*RGS9*	*SAMD11*	*TTPA*	
OPN1MW	*PRCD*	*RGS9BP*	*SDCCAG8*	*TUB*	
OPN1SW	*PRD*	*RHO*	*SEMA4A*	*TUBGCP4*	
OTX2	*PRDM13*	*RIMS1*	*SLC24A1*	*TUBGCP6*	
PANK2	*PROM1*	*RLBP1*	*SLC25A46*	*TULP1*	
PAX2	*PRPF3*	*RNANC*	*SLC7A14*	*UNC119*	
PCDH15	*PRPF31*	*ROM1*	*SNRNP200*	*USH1C*	
PCYT1A	*PRPF4*	*RP1*	*SPATA7*	*USH1E*	

B.2 Sanger sequencing primers

Gene	Primers sequences	Product size
CDHR1 **(Exon 14)**	Forward: TCTCTCCACCAATTTCTCCCC	339 bp
	Reverse: GCCTCTCTCCACTTCTCTTCA	
RHO **(Exon 1)**	Forward: TCTACGTCACCGTCCAGCAC	248 bp
	Reverse: ACTCTCCCAGACCCCTCCAT	
PRPF31 **(Exon 8)**	Forward: CCCACCTCTCTGCTTTCTTCT	328 bp
	Reverse: GCTCCTGAGTGCTACCGTCA	
SLC25A46 **(Exon 7)**	Forward: GCATCCTACGTTACAGATTTCTGA	393 bp
	Reverse: GCAACGTGCGATATAACTGTG	
MFSD9 **(Exon 4)**	Forward: GGCATCATGGAACTGAGCATC	262 bp
	Reverse: GGACTCTAGCCAGGACAAACA	
CNGA3 **(Exon 8)**	Forward: TACTGAAGTTTTCCCGGCTCT	421 bp
	Reverse: TCAATCTTGGCCTGGAACTCT	
BBS10 **(Exon 2)**	Forward: ACTTTTGTGGAAGAGTGGGAAG	326 bp
	Reverse: TGTCTGAAACTGTGCTTCTGA	
ABCA4 **(c.2588G>C)**	Forward: ACTCACCTTCACCGCCAAG	225 bp
	Reverse: GGGAGCCTGAGAATAGCCAT	
ABCA4 **(c.6089G>A)**	Forward: ACTCCTATGTGGCCACAACA	228 bp
	Reverse: GGTCATCCCTCCACTCCTTG	
USH2A **(Exon 2)**	Forward: GTCACGAGGTCTTTTCCCAAG	376 bp
	Reverse: CTCAGGTTTCAGCCATACAGC	
OPA1 **(Exon 10)**	Forward: CTCAGAGCAGCATTACAAATAGGT	270 bp
	Reverse: TCAACACAATGAACAGGTCTCAC	
ABCA4	Forward: TCAAAGGGGAAGTGGAGAGAA	428 bp
	Reverse: TTGCTATGCTTGGGTGGGGA	
SNRNP200 **(Exon 21)**	Forward: GTCTTGTAGGATGCGGTGAAC	404 bp
	Reverse: GTGAAATGATGTCTGCACGGA	

Table B.1: PCR/Sequencing primers used to confirm variants identified by WES (Chapter 3)

B.3 Primers sequences to sequence across VCAN breakpoints

Primer name	Forward	Reverse	Product size
VCAN_het_del1	AACCACCTCTCATCGACAGG	ATTTTGCAGCGATCAGGTCC	23555bp
VCAN_het_del2	CACCTCTCATCGACAGGGAA	ACACTGGTCTCCGCTGTATC	23640bp
VCAN_het_del3	GCCAAGGAAACAGAAACCGA	AGGGAAAATGGTGGTGTTGC	20081bp

B.4 List of candidate genes for unsolved samples

B.4.1 Sample 3800

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
4	190876277	*FRG1*	Het Stop gained (39)	NM_004477.3: c.403A>T, p.R135*
2	33413889	*LTBP1*	Het Missense (35)	NM_206943.4: c.1672C>T, p.R558W
12	80616011	*OTOGL*	Het Missense (35)	NM_001378609.3: c.475C>T, p.R159W
12	57431758	*MYO1A*	Het Missense (34)	NM_005379.4: c.1856G>A, p.R619Q
2	241536279	*CAPN10*	Het Missense (33)	NM_023083.4: c.1663C>T, p.R555C

Table B.2: List of top five candidate variants based on CADD score in case 3800 after alignment, variant calling, and filtering. The patient is male with autosomal dominant retinitis pigmentosa, only heterozygous variants were retained. Het = heterozygous.

B.4.2 Sample 5545

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
9	33466142	*NOL6*	Het Missense (35)	NM_022917.5: c.2291G>A, p.R764H
11	130278465	*ADAMTS8*	Het Missense (35)	NM_007037.6: c.1993G>A, p.G665S
13	25671210	*PABPC3*	Het Stop gained (35)	NM_030979.3: c.874C>T, p.Q292*
16	74425787	*NPIPB15*	Het Stop gained (35)	NM_001306094.1: c.1141G>T, p.E381*
17	48628401	*SPATA20*	Het Missense (34)	NM_022827.4: c.1426C>T, p.R476W

Table B.3: List of top five candidate variants based on CADD score in case 5545 after alignment, variant calling, and filtering. The patient is female with pigment epithelial detachment with macular geographical atrophy (inheritance pattern is unknown). Heterozygous, biallelic, and homozygous variants were retained. Het = heterozygous.

B.4.3 Sample 5582

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
19	36297423	*PRODH2*	Het Stop gained (36)	NM_021232.2 c.910C>T, p.R304*
11	64323613	*SLC22A11*	Het Stop gained (35)	NM_018484.4: c.142C>T, p.R48*
15	90328681	*ANPEP*	Het Missense (35)	NM_001150.3: c.2803C>T, p.R935W
16	57938727	*CNGB1*	Het Frameshift (35)	NM_001297.5: c.2544dup, p.L849Afs*3
1	5934941	*NPHP4*	Het Missense (33)	NM_015102.5: c.3037G>A, p.E1013K

Table B.4: List of top five candidate variants based on CADD score in case 5582 after alignment, variant calling, and filtering. The patient is female with macular dystrophy (inheritance pattern is unknown). Heterozygous, biallelic, and homozygous variants were retained. The frameshift mutation in *CNGB1* is heterozygous in a gene that causes a recessive disease, hence ExomDepth was investigated for second alleles.

B.4.4 Sample OA1203

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
11	72539495	*ATG16L2*	Het Stop gained (44)	NM_033388.2 c.1564C>T, p.R522*
7	151962134	*KMT2C*	Het Stop gained (39)	NM_170606.3: c.1173C>A, p.C391*
13	25671369	*PABPC3*	Het Stop gained (38)	NM_030979.3: c.1033G>T, p.E345*
1	149931640	*OTUD7B*	Het Stop gained (37)	NM_020205.4: c.808C>T, p.R270*
6	152621820	*SYNE1*	Het Missense (35)	NM_182961.4 c.17638C>T, p.R5880C

Table B.5: List of top five candidate variants based on CADD score in case OA1203 after alignment, variant calling, and filtering. The patient is female with autosomal recessive optic atrophy. Both biallelic and homozygous variants were retained.

B.4.5 Sample 12194201

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
17	55193512	*AKAP1*	Het Stop gained (44)	NM_003488.4: c.2322G>A, p.W774*
4	103826685	*SLC9B1*	Het Stop gained (42)	NM_139173.4: c.1318A>T, p.K440*
2	32422819	*SLC30A6*	Het Stop gained (40)	NM_017964.5: c.709C>T, p.R197*
6	28269139	*PGBD1*	Het Stop gained (37)	NM_032507.4: c.1508C>A, p.S503*
9	133057290	*HMCN2*	Het Missense (37)	NM_001291815.2: c.769G>T, p.V257F

Table B.6: List of top five candidate variants based on CADD score in case 12194201 after alignment, variant calling, and filtering. The patient is male with microcephaly, vision problem, and developmental delay (inheritance pattern is unknown). Heterozygous, biallelic, and homozygous variants were retained.

B.4.6 Sample 12642093

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
16	30953855	*FBXL19*	Homo Missense (28.6)	NM_001382779.1: c.1625C>A, p.P542Q
3	133969435	*RYK*	Homo Frameshift (23.2)	NM_002958.4: c.57insG, p.R20Qfs*59
3	113376110	*USF3*	Homo in-frame deletion (17.6)	NM_001009899.4: c.4416_4418del, p.Q1473del

Table B.7: List of top candidate variants based on CADD score in case 12642093 after alignment, variant calling, and filtering. The patient is female with autosomal recessive Leber congenital amaurosis (with history of consanguinity). Only homozygous variants were retained.

B.4.7 Sample 8956112

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
12	30784844	*IPO8*	Het Stop gained (42)	NM_006390.4: c.3001C>T, p.R1001*
1	144921963	*PDE4DIP*	Het Stop gained (37)	NM_001350520.1: c.1555C>T, p.E519*
17	48646234	*CACNA1G*	Het Stop gained (37)	NM_018896.5: c.246G>A, p.W82*
11	17667125	*OTOG*	Het Missense (35)	NM_001292063.2: c.8476C>T, p.R2826C
16	74425787	*NPIPB15*	Het Stop gained (35)	NM_001306094.1: c.1141G>T, p.E381*

Table B.8: List of top five candidate variants based on CADD score in case 8956112 after alignment, variant calling, and filtering. The patient is male with a possible microdeletion syndrome (inheritance pattern is unknown). Heterozygous, biallelic, and homozygous variants were retained.

B.4.8 Sample 3558

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
1	86173649	*ZNHIT6*	Het Missense (28.8)	NM_017953.4: c.319G>A, p.V107M
X	2836184	*ARSD*	Hemi Missense (27.2)	NM_001669.4: c.524G>A, p.G175D
14	70039881	*CCDC177*	Homo Missense (23.8)	NM_001271507.2: c.459G>C, p.E153D
11	121458843	*SORL1*	Het Missense (22.9)	NM_003105.6 c. c.3929C>T, p.A1310V
1	16901021	*NBPF1*	Het Stop gained (22.8)	NM_017940.6: c.2359T>C, p.S787P

Table B.9: List of top five candidate variants based on CADD score in case 3558 after alignment, variant calling, and filtering. The patient is male with autosomal recessive RP. Both biallelic and homozygous variants were retained. No point mutations or structural variations were detected in known IRDs genes. This case was included in *RPGR-ORF15* screening (Chapter 5).

B.4.9 Sample 5583

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
14	64066363	*WDR89*	Het Stop gained (35)	NM_080666.4: c.298C>T, p.R100*
5	236649	*SDHA*	Het Missense (29.4)	NM_004168.4: c.1367C>T, p.S456L
3	141535879	*GRK7*	Het Missense (25.5)	NM_139209.3: c.1649G>A, p.C550Y
14	92547337	*ATXN3*	Het Missense (25.3)	NM_004993.6: c. c.844C>T, p.R282W
19	15918201	*OR10H1*	Het Stop gained (24.8)	NM_013940.4: c.647T>C>C, p.L216P

Table B.10: List of top five candidate variants based on CADD score in case 5583 after alignment, variant calling, and filtering. The patient is female with X-linked RP. Both biallelic and homozygous variants were retained. This case was included in *RPGR-ORF15* screening (Chapter 5).

B.4.10 Sample 4212

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
13	109610115	*MYO16*	Het Stop gained (44)	NM_001198950.3: c.2005C>T, p.R669*
8	86178918	*CA13*	Het Stop gained (41)	NM_198584.3: c.436G>T, p.G146*
4	190876277	*FRG1*	Het Stop gained (39)	NM_004477.3: c.403A>T, p.R135*
1	14059374	*PRDM2*	Het Stop gained (38)	NM_012231.4: c.228G>A, p.W76*
8	82439323	*FABP12*	Het Stop gained (38)	NM_001105281.3: c.280C>T, p.Q94*

Table B.11: List of top five candidate variants based on CADD score in case 4212 after alignment, variant calling, and filtering. The patient is male with autosomal dominant RP. Only heterozygous variants were retained.

B.4.11 Sample 5544

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
19	1058727	*ABCA7*	Het Stop gained (42)	NM_019112.4: c.5260C>T, p.R1754*
15	68619124	*ITGA11*	Het Stop gained (41)	NM_001004439.2: c.2079C>G, p.Y693*
6	39272336	*KCNK17*	Het Stop gained (38)	NM_031460.4: c.448C>T, p.R150*
11	65810704	*GAL3ST3*	Het Stop gained (37)	NM_033036.3: c.570C>A, p.Y190*
6	46679244	*PLA2G7*	Het Stop gained (36)	NM_005084.4: c.652C>T, p.R218*

Table B.12: List of top five candidate variants based on CADD score in case 5544 after alignment, variant calling, and filtering. The patient is female with CORD (inheritance pattern is not clear). Heterozygous, biallelic, and homozygous variants were retained.

B.4.12 Sample 5546

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
16	31418867	*ITGAD*	Het Stop gained (36)	NM_005353.3: c.736C>T, p.R246*
8	21768200	*DOK2*	Het Missense (35)	NM_003974.4: c.602G>A, p.R201H
9	33466142	*NOL6*	Het Missense (35)	NM_022917.5: c.2291G>A, p.R764H
19	58385479	*ZNF814*	Het Stop gained (35)	NM_001144989.2: c.1279C>T, p.Q427*
7	128491324	*FLNC*	Het Missense (34)	NM_001458.5: c.5578C>T, p.R1860C

Table B.13: List of top five candidate variants based on CADD score in case 5546 after alignment, variant calling, and filtering. The patient is male with AVMD (inheritance pattern is unknown). Heterozygous, biallelic, and homozygous variants were retained.

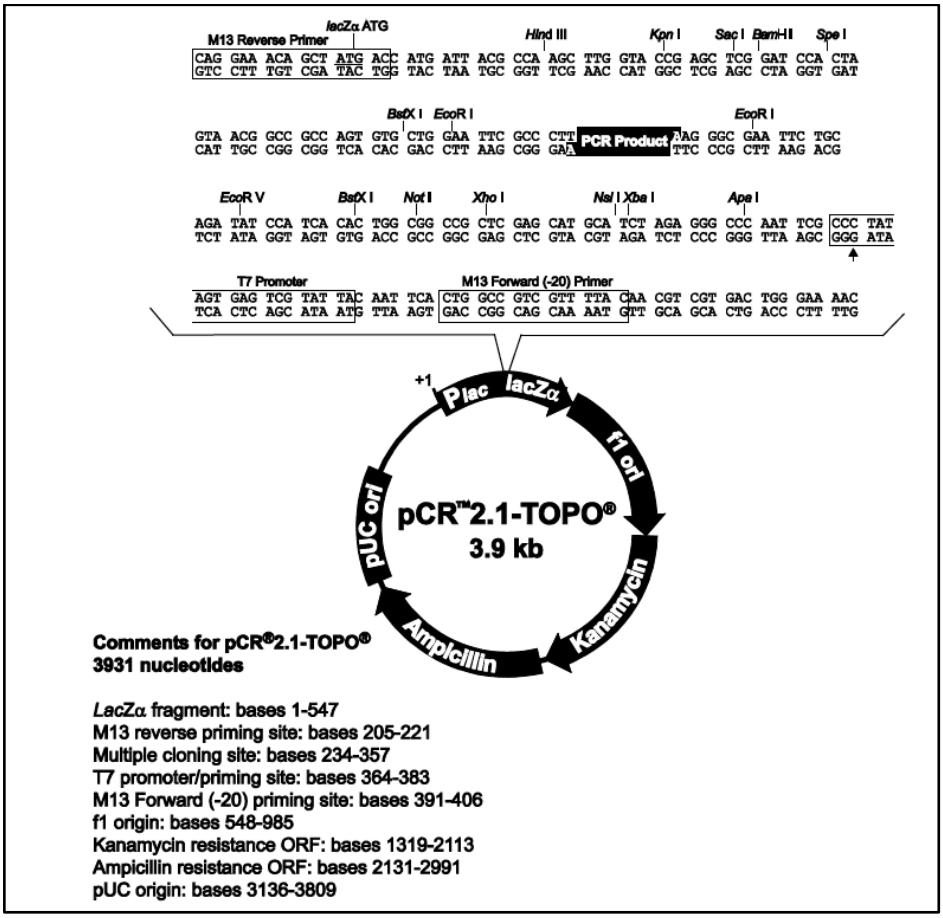
B.4.13 Sample 3727

Chr.	Position (GRCh37)	Gene	Mutation type (CADD score)	Mutation
3	108147680	*MYH15*	Het Stop gained (37)	NM_014981.2: c.3421C>T, p.R1141*
4	42545953	*ATP8A1*	Het Missense (35)	NM_006095.2: c.1703G>A, p.R568Q
3	197408001	*RUBCN*	Het Missense (34)	NM_014687.4: c.2429G>A, p.R810Q
4	148887925	*ARHGAP10*	Het Missense (34)	NM_024605.4.2: c.1651G>A, p.A551T
5	1879861	*IRX4*	Het Missense (33)	NM_016358.3: c.493C>T, p.R165C

Table B.14: List of top five candidate variants based on CADD score in case 3727 after alignment, variant calling, and filtering. The patient is male with MD (inheritance pattern is unknown). Heterozygous, biallelic, and homozygous variants were retained.

Appendix C

C.1 Map of pCR 2.1 TOPO vector used in TA cloning for *CRX* deletion cases



C.2 Internal primers used to sequence the breakpoint containing clone or amplimer in *CRX* deletion carriers

Primer Name	Primer sequence
CRX-del-1F	TACAGGCATGCACCACTACA
CRX-del-1R	CCCTGCCCCTTAAGTTAGCT
CRX-del-4F	TGATGTCATGGGCTTACTGG
CRX-del-4R	CCTCCTGAACAGCTGGGATT
CRX-del-5	GTGTGTACCATCATGCTCAGT
CRX-del-7	GAGGCTGGCTCTGTCACC

Appendix D

D.1 Clinical details and identified variants of 218 individuals included in *RPGR-ORF15* screening on both MinION and PacBio sequencers

Patient ID	Sex	Inheritance pattern	Clinical diagnosis	Subsequently solved	Category	Identified variants (NM_001034853.2)	Sequenced on minion	Sequenced on pacbio
119	F	Simplex	RP		Unsolved case			x
122	M	XL	RP		Unsolved case			x
135	M	XL	RP		Known control	c.2426_2427delAG p.(Glu809Glyfs*25)	x	x
410	F	AR	MD		Unsolved case			x
415	F		Sectoral RP		Unsolved case			x
427	M	XL/AR	RP		Unsolved case			x
460	F	AD RP	RP		Unsolved case			x
468	M	Sporadic	RP		Unsolved case			x
469	F	Sporadic	RP/RCD		Unsolved case			x
470	M	Sporadic	RP		Unsolved case			x
535	M		RP		Unsolved case			x

548	F	AD	RP		Unsolved case			x
651	M	AD	RP	NR2E3	Obligate negative		x	x
741	M	XL	RP		Unsolved case			x
744	M		RP		Unsolved case		x	x
749	F	XL	RP		Known control	c.2405_2406delAG p.(Glu802Glyfs*32)	x	x
752	F	XL	RP		Known control	c.2405_2406delAG p.(Glu802Glyfs*32)		x
753	M	XL	RP		Known control	c.2405_2406delAG p.(Glu802Glyfs*32)	x	x
760	M		Pigmentary retinopathy		Unsolved case			x
764	M	AD	RP		Unsolved case		x	x
771	F		RP		Unsolved case		x	x
918	M	XL?	RP		Unsolved case			x
999	F		RP		Unsolved case			x
1024	M	AR/XL	RP		Unsolved case			x

1045	M	AR/XL	RP		Unsolved case		x	x
1190	M		Pattern dystrophy		Unsolved case			x
1194	M		RP		Unsolved case			x
1284	M		Cone dystrophy		Unsolved case			x
1475	M		RP		Known control	c.2236_2237delGA p.(Glu746Argfs*23)		x
1486	M	Sporadic	RP		Unsolved case			x
1487	F	XL	RP		Known control	c.2405_2406delAG p.(Glu802Glyfs*32)	x	x
1599	F		CORD		Unsolved case			x
1600	F		Pattern dystrophy		Unsolved case			x
1607	F	AD	MD		Unsolved case			x
1670	M	AD	CACD/LORD	C1QTNF5	Obligate negative			x
1677	M	AR	RP		Unsolved case			x
1764	M	AD	Macular coloboma		Unsolved case		x	x
1765	M	AD	Macular coloboma		Unsolved case			x
1786	M	AD	RP		Unsolved case			x

1794	M	AD/XL	RP	Unsolved case		x
1798	M		RP	Unsolved case		x
2002	M	XL	RP	Unsolved case	x	x
2050	M	XL	RP	Unsolved case	x	
2077	F	XL	RP	Unsolved case		x
2092	M	Sporadic	RP	Unsolved case		x
2117	F	Sporadic	CORD	Unsolved case		x
2126	M	XL	RP	Unsolved case		x
2148	F		Generalised RD	Unsolved case		x
2185	M		Nyctalopia, photophobia	Unsolved case		x
2190	F	AD	MD	Unsolved case		x
2194	M	AD	RCD	Unsolved case	x	x
2196	F	AD	MD	Unsolved case		x
2197	M		RP	Unsolved case		x
2207	M	Simplex	RP	Unsolved case		x
2215	F		MD	Unsolved case		x

2216	M	AR	RP unaffected father		Unsolved case		x
2270	M	AD	MD		Unsolved case		x
2272	M		MD		Unsolved case		x
2379	F		RP		Unsolved case		x
2400	F		RP		Unsolved case		x
2403	F	AD	RP		Unsolved case		x
2444	M		RCD	RP1	Obligate negative	x	x
2457	F		RP		Unsolved case		x
2461	M		RCD		Unsolved case		x
2463	F	Simplex	Cone dystrophy		Unsolved case		x
2475	F	AR	RP inversa		Unsolved case		x
2477	M	Simplex	RP	RP2	Obligate negative	x	x
2479	M	Simplex	RP		Unsolved case		x
2482	M		Unaffected parent RP		Unsolved case		x
2484	M?		Parents 1st cousin		Unsolved case		x

			Retinal dystrophy, nyctalopia, photophobia					
2729	M	AD/Sporadic	CORD		Unsolved case			x
2762	M	Simplex/XL	RP	CACNA1F	Obligate negative		x	x
2773	F	AR	RP		Unsolved case			x
2843	M		RP inversa		Unsolved case			x
3216	M	XL?	RP	BBS1	Obligate negative		x	x
3217	M	XL?	RP	USH2A	Obligate negative		x	x
3218	M	XL?	RP		Unsolved case			x
3219	M	XL?	RP		Unsolved case	c.2323_2324delAG p.(Arg775Glufs*59)	x	x
3220	M	XL?	RP		Unsolved case			x
3233	M	XL	RP		Unsolved case			x
3236	M	XL?	RP		Unsolved case		x	x
3239	M	XL?	RP	CLN3	Obligate negative		x	x
3241	M	XL?	RP		Unsolved case		x	x

3244	M	XL?	RP	IDH3B	Obligate negative		x	x
3246	M		RP		Unsolved case			x
3351	M		RP		Unsolved case			x
3439	M	XL?	RP		Unsolved case		x	x
3528	F		MD (NCMD?)		Unsolved case			x
3536	F		MD	ABCA4	Obligate negative			x
3558	M	AR	RP		Unsolved case	c.2488delG p.(Glu830Argfs*259)		x
3606	F	AR	RP		Unsolved case	c.2608_2609insG p.(Glu870Glyfs*209)		x
3654	M	Simplex	MD (Best disease)	BEST1	Obligate negative			x
3656	F		MD	ABCA4	Obligate negative			x
3670	F		MD	BEST1	Obligate negative			x
3681	F	Simplex	MD		Unsolved case			x
3684	F		Pericentral RP		Unsolved case			x
3685	M	XL/AR	RP		Unsolved case	c.2041_2042delAA p. (Lys681Glyfs*2)	x	x
3691	F	Simplex	RCD		Unsolved case			x

3693	M	Multiplex	RP		Unsolved case		x
3698	F		Sectoral RP		Unsolved case		x
3719	M		MD		Unsolved case		x
3726	M	AR	RP		Unsolved case		x
3775	M	XL	RP		Unsolved case	x	x
3776	M	XL	RP		Unsolved case	x	
3787	M		RCD		Unsolved case		x
3798	F		Pattern dystrophy	PRPH2	Obligate negative		x
3877	F		RP		Unsolved case		x
3910	F	AR	RP		Unsolved case		x
3941	M		MD		Unsolved case		x
3952	F		MD		Unsolved case		x
3962	M		RCD		Unsolved case		x
3974	F	AD	RP		Unsolved case		x
3994	F	AR	RP		Unsolved case		x
3996	M		LORD	RP1L1	Obligate negative	x	x

4018	M		Reticular Pattern Dystrophy		Unsolved case			x
4084	F	AD	RP		Unsolved case			x
4153	M	AR	RP	USH2A	Obligate negative		x	x
4168	M		MD		Unsolved case			x
4169	F		MD		Unsolved case			x
4172	M		RCD		Unsolved case			x
4181	F		MD		Unsolved case			x
4187	F		MD		Unsolved case			x
4203	M		CACD		Unsolved case			x
4395	M	AD	RP	PRPF31	Obligate negative		x	x
4401	M	AR	RP	REEP6	Obligate negative		x	x
4656	M	AD	MD		Unsolved case			x
4659	F	AD	RP		Unsolved case	c.3317delA p.(Lys1106Serfs*25)		x
4662	M	AD Multiplex	RP		Unsolved case			x
4663	F		MD		Unsolved case			x
4666	F	Simplex	PPRCA		Unsolved case			x

4667	F	Simplex	Achromatopsia	Unsolved case			x
4668	M	AD	RP	Unsolved case	c.2608_2609insG p.(Glu870Glyfs*209)	x	x
4670	M	Simplex	Retinitis punctata albescence	Unsolved case			x
4671	F	Multiplex	CACD	Unsolved case			x
4676	M?			Unsolved case			x
4677	M	AD/AR	RP	Unsolved case			x
4680	F	AD	RP	Unsolved case			x
4681	M	AD	RP	Unsolved case			x
4684	M	AD	MD	Unsolved case			x
4686	F	Multiplex	RP	Unsolved case			x
4691	M		RP	Unsolved case		x	x
4692	M	Simplex	Cone or MD	Unsolved case			x
4697	M	AR	RP	Unsolved case			x
4699	M	Simplex/XL	RP	Unsolved case		x	x
4700	F	Multiplex	CRD	Unsolved case			x

4703	M	Simplex	CACD	Unsolved case		x
4706	?	Simplex	RP and bull's eye MD	Unsolved case		x
4709	M	Simplex	RP	Unsolved case	x	x
4710	F	Simplex	LORD	Unsolved case		x
4742	M		LORD	Unsolved case		x
4745	F	Multiplex	AVMD	Unsolved case		x
4751	unknown	Simplex	MD	Unsolved case		x
4753	F	Simplex	RCD	Unsolved case		x
4754	M	Multiplex	RP	Unsolved case		x
4755	M		RCD	Unsolved case		x
4756	M	Multiplex	MD	Unsolved case	x	x
4759	F	Simplex	Vitelliform MD	Unsolved case		x
4770	F	Simplex	CACD	Unsolved case		x
4774	M	Simplex	RP	Unsolved case		x
4780	M	Simplex	RCD	Unsolved case		x
4786	M	Simplex	RP/BCAD/	Unsolved case		x

			pericentric RP				
4788	M	Simplex	CRD and coloboma	Unsolved case			x
4789	M	Simplex	MD or NCMD	Unsolved case		x	x
4819	M		Pericentric RP	Unsolved case			x
4823	F	AD	RP	Unsolved case			x
5107	M	Simplex	RP	Unsolved case			x
5152	F	AR	MD	Unsolved case			x
5156	M	XL	MD	Known control	c.3334 C>T p.(Gln1112*)	x	x
5191	M	Simplex	RP and Asperger's	Unsolved case			x
5192	unknown	AR?	ABCA4 retinopathy?	Unsolved case			x
5193	M	Simplex	RP	Unsolved case			x
5194	F	Simplex	Atypical RCD	Unsolved case			x
5208	F		Early onset retinopathy	Unsolved case			x
5212	F	Simplex	RP and	Unsolved case			x

			short stature				
5213	F		Atypical bests/ pattern dystrophy	Unsolved case			x
5214	M	AD?	Pattern dystrophy/ CACD	Unsolved case			x
5225	unknown	AR?	Bull's eye maculopathy	Unsolved case	c.2608_2609insG p.(Glu870Glyfs*209)		x
5228	F	Multiplex	RP and high myopia	Unsolved case	c.2764delG p.(Glu922Argfs*167)		x
5231	M	AD?	Severe RP	Unsolved case		x	x
5232	M	Simplex	Late onset RCD	Unsolved case			x
5239	M		Mild RP	Unsolved case			x
5242	F	AR?	Severe RP	Unsolved case	c.3317delA p.(Lys1106Serfs*25)		x
5243	F		Asymptomatic pigmentary disturbance	Unsolved case			x
5251	M		Cone dystrophy	Unsolved case	c.2608_2609insG		x

			(?ABCA4?)		p.(Glu870Glyfs*209)		
5257	M	XL?	RP	Known control	c.3334 C>T p.(Gln1112*)	x	x
5258	F		CORD and hyperoxaluria	Unsolved case			x
5260	M			Unsolved case			x
5262	M		Serous retinal detachment and RD	Unsolved case			x
5264	M		Bull's eye maculopathy	Unsolved case			x
5266	F		RCD	Unsolved case			x
5267	F	Consanguinity/ simplex	CORD/CSNB	Unsolved case			x
5270	M	Simplex	CORD	Unsolved case			x
5272	M	Simplex/XL?	RP	Unsolved case			x
5274	M		Stargardts?	Unsolved case			x
5278	M		Early onset RCD	Unsolved case			x

5279	F	Multiplex	RP		Unsolved case			x
5347	M		Odd RD, myopia, pseudostaphyloma, pigment clumps		Unsolved case			x
5349	M	Simplex	Atypical CORD	ABCA4	Obligate negative			x
5550	F				Unsolved case		x	x
5586	M		RP		Known control	c.2426_2427delAG p.(Glu809Glyfs*25)	x	x
5610	F		MD or COD	PROM1	Obligate negative			x
5857	F		MD or COD	ABCA4	Obligate negative			x
5861	M		Stargardts?	ABCA4	Obligate negative			x
5864	M		Cone dystrophy	ABCA4	Obligate negative			x
5865	M		MD or COD	ABCA4	Obligate negative			x
5867	M		Unusual peripheral retinal degeneration, LD, obesity		Unsolved case	c.2971delG p.(Glu991Lysfs*98)		x
5868	F	AR	CORD		Unsolved case			x

5869	F	AR	CORD	Unsolved case	x
5870	F	AD variable penetrance or XL	RCD	Unsolved case	x
5871	M	AD	RCD	Unsolved case	x
5872	M		COD	Unsolved case	x
5873	F		MD	Unsolved case	x
5874	M		MD	Unsolved case	x
5875	M		RP	Unsolved case	x
5972	M		LCA/early RP	Unsolved case	x
5974	M		LCA/early RP	Unsolved case	x
5977	M		LCA/early RP	Unsolved case	x
5980	M		Sever early RP	Unsolved case	x

Table D.1: Clinical details and identified variants of all individuals included in *RPGR ORF15* screening using both MinION and PacBio sequencers. The diagnosis of the patients was either peripheral or central retinopathy. M = male, F = female, RP= Retinitis pigmentosa, MD = Macular dystrophy, RCD = Rod cone dystrophy, COD = Cone dystrophy, CORD = Cone rod dystrophy, LCA = Leber congenital amaurosis, CACD = Central areolar choroidal dystrophy, LORD = Late onset retinal degeneration, RD = Retinal detachment, NCMD = North Carolina macular dystrophy, PPRCA = Pigmented paravenous retinochoroidal atrophy, AVMD = Adult vitelliform macular dystrophy, BCAD = Benign concentric annular dystrophy, CSNB = Congenital stationary night blindness, LD= Learning disabilities, AR= Autosomal recessive, AD= Autosomal dominant, XL= X-linked.

D.2 List of non-barcoded PCR primers tested in PacBio sequencing study

Forward primer (MA_ORF15_F2)	Reverse primer (MA_ORF15_R)	Product size
AGAAAGAGGATGATGAAGTGGAAACT	TTTAAATTGTCTGACTGGCCATAATC	1,892 bp

D.2.1 Workflow of PCR optimisation for primers MA_ORF15_F2 and MA_ORF15_R

First, the optimisation started by trying the Biolab protocol as follows:

Component	20 µl Reaction	50 µl Reaction	Final Concentration
Nuclease-free water	to 20 µl	to 50 µl	
5X Phusion HF or GC Buffer	4 µl	10 µl	1X
10 mM dNTPs	0.4 µl	1 µl	200 µM
10 µM Forward Primer	1 µl	2.5 µl	0.5 µM
10 µM Reverse Primer	1 µl	2.5 µl	0.5 µM
Template DNA	1 µl	1 µl	< 250 ng
DMSO (optional)	(0.6 µl)	(1.5 µl)	3%
Phusion DNA Polymerase	0.2 µl	0.5 µl	1.0 units/50 µl PCR

Two set of experiment were attempted:

1. GC buffer was used in the first set.
2. HF buffer was used in the second.

And the initial thermocycling conditions were:

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	98 °C	1 minute
30 cycles	Denaturation	98 °C	10 seconds
	Annealing	70 °C	30 seconds
	Extension	72°C	1 minute.
1	Final extension	72°C	5 minutes.
1	Hold	4 °C	∞

Then, the variation used were as follows:

1- Touch down the annealing temperature 70°C→65°C, 67°C→60°C, and 67°C→63°C. When we started to see the correct bands at annealing temperature 67°C→63°C, we tried two separate annealing temperature 67°C and 65°C.
2- Additives as DMSO 3% and $MgCl_2$ were used.

We found that using the annealing temperature 65°C and using GC buffer with adding DMSO 3% gave us a correct band size and they were the best conditions to use.

D.2.2 List of barcoded primers of MA_ORF15_F2 and MA_ORF15_R used in PacBio sequencing study

Forward primer (MA_ORF15_F2)	
F1:	GGTAGTCAGACGATGCGTCATAGAAAGAGGATGATGAAGTGGAAACT
F2:	GGTAGCTATACATGACTCTGCAGAAAGAGGATGATGAAGTGGAAACT
F3:	GGTAGTACTAGAGTAGCACTCAGAAAGAGGATGATGAAGTGGAAACT
F4:	GGTAGTGTGTATCAGTACATGAGAAAGAGGATGATGAAGTGGAAACT
F5:	GGTAGACACGCATGACACACTAGAAAGAGGATGATGAAGTGGAAACT
F6:	GGTAGGATCTCTACTATATGCAGAAAGAGGATGATGAAGTGGAAACT
F7:	GGTAGACAGTCTATACTGCTGAGAAAGAGGATGATGAAGTGGAAACT
F8:	GGTAGATGATGTGCTACATCTAGAAAGAGGATGATGAAGTGGAAACT
Reverse primer (MA_ORF15_R)	
R1:	CCATCATGACGCATCGTCTGATTTAAATTGTCTGACTGGCCATAATC
R2:	CCATCGCAGAGTCATGTATAGTTTAAATTGTCTGACTGGCCATAATC
R3:	CCATCGAGTGCTACTCTAGTATTTAAATTGTCTGACTGGCCATAATC
R4:	CCATCCATGTACTGATACACATTTAAATTGTCTGACTGGCCATAATC
R5:	CCATCAGTGTGTCATGCGTGTTTAAATTGTCTGACTGGCCATAATC
R6:	CCATCGCATATAGTAGAGATCTTTAAATTGTCTGACTGGCCATAATC
R7:	CCATCCAGCAGTATAGACTGTTTAAATTGTCTGACTGGCCATAATC

Table D.2: Sequences of barcoded primer pair (MA_ORF15_F2 and MA_ORF15_R).

Red: padding sequence (5 bp), Green: barcode sequence, Black: primer sequence.

D.3 List of newly designed non barcoded primers (MA_ORF15_F3, MA_ORF15_R3 and MA_ORF15_F4, MA_ORF15_R4) tested in PacBio sequencing study

Primer name	Primer sequence
MA_ORF15_F3	CAGTAGAAAAGCCAGACAGTTACATG
MA_ORF15_R3	CTATTGTCTTTGGCTCCTTAACACAG
MA_ORF15_F4	CCTGAGGCAATAGAATTTAGTAGTGG
MA_ORF15_R4	GTATATTCCTGTTTCCTAAAGCTGCC

Table D.3: Sequences of different non barcoded primer combinations tested for PCR amplification of *ORF15* region for PacBio sequencing. F = Forward, R = Reverse.

Product size for different combinations:

- F3 – R3 → 2119bp
- F4 – R4 → 2198bp
- F4 – R3 → 2045bp
- F3 – R4 → 2272bp

D.3.1 Workflow of PCR optimisation for primers MA_ORF15_F3, MA_ORF15_R3 and MA_ORF15_F4, MA_ORF15_R4

Four primer combinations were tried (F3/R3, F4/R4, F3/R4, F4/R3) and for each combination the following steps were performed:

1- GC buffer +/- DMSO 3%
2- HF buffer +/- DMSO 3%

At first these experiments were performed by applying the following thermocycling conditions:

No. of cycles	PCR stage	Temperature	Time
1	Initial denaturation	98°C	1 minutes
4 cycles	Denaturation	98°C	10 seconds
	Annealing	63°C (Decrease by 0.5°C/ cycle)	30 seconds
	Extension	72°C	1 minute
21 cycles	Denaturation	98°C	10 seconds
	Annealing	61 °C	30 seconds
	Extension	72°C	1 minute
1	Final Extension	72°C	5 minutes
1	Hold	4°C	∞

The bands of the correct sizes have appeared only in one primer combination, F3/R4 and that was when using the GC/HF buffer with no addition of DMSO 3%, these bands were unclear so, we decided to try higher annealing temperature.

1- Touchdown 65°C→63°C
2- Touchdown 67°C→65°C

We found that applying the annealing temperature (Touchdown 67°C→65°C) and using the GC buffer with no DMSO added, gave us clear bands and it was the best conditions to use.

D.3.2 List of barcoded primers of MA_ORF15_F3 and MA_ORF15_R4 used in PacBio sequencing study

Barcode ID	Forward primer (MA_ORF15_F3)	Barcode ID	Reverse primer (MA_ORF15_R4
F1	GCGCTCTGTGTGCAGCCAGTAGAAAAGCCAGACAGTTACATG	R1	TCATATGTAGTACTCTGTATATTCCTGTTTCCTAAAGCTGCC
F2	TCATGAGTCGACACTACAGTAGAAAAGCCAGACAGTTACATG	R2	GCGATCTATGCACACGGTATATTCCTGTTTCCTAAAGCTGCC
F3	TATCTATCGTATACGCCAGTAGAAAAGCCAGACAGTTACATG	R3	TGCAGTCGAGATACATGTATATTCCTGTTTCCTAAAGCTGCC
F4	ATCACACTGCATCTGACAGTAGAAAAGCCAGACAGTTACATG	R4	GACTCTGCGTCGAGTCGTATATTCCTGTTTCCTAAAGCTGCC
F5	AGAGCATCTCTGTACTCAGTAGAAAAGCCAGACAGTTACATG	R5	TACAGCGACGTCATCGGTATATTCCTGTTTCCTAAAGCTGCC
F6	TGTGAGTCAGTACGCGCAGTAGAAAAGCCAGACAGTTACATG	R6	GCGCAGACTACGTGTGGTATATTCCTGTTTCCTAAAGCTGCC
F7	AGAGACACGATACTCACAGTAGAAAAGCCAGACAGTTACATG	R7	GTCTCTGCGATACAGCGTATATTCCTGTTTCCTAAAGCTGCC
F8	CTGCTAGAGTCTACAGCAGTAGAAAAGCCAGACAGTTACATG	R8	AGTATGAGATAGCTCGGTATATTCCTGTTTCCTAAAGCTGCC
F9	AGCACTCGCGTCAGTGCAGTAGAAAAGCCAGACAGTTACATG	R9	GCGACGAGTACTCATGGTATATTCCTGTTTCCTAAAGCTGCC
F10	TCATGCACGTCTCGCTCAGTAGAAAAGCCAGACAGTTACATG	R10	AGTATCACAGTCGCTGGTATATTCCTGTTTCCTAAAGCTGCC
F11	CGCATCGACTACGCTACAGTAGAAAAGCCAGACAGTTACATG	R11	AGACGTAGATCACAGCGTATATTCCTGTTTCCTAAAGCTGCC
F12	CGTAGCGTGCTATCACCAGTAGAAAAGCCAGACAGTTACATG	R12	CGTGTCATGCTACTCAGTATATTCCTGTTTCCTAAAGCTGCC
F13	ATGCTGATGACTGCGACAGTAGAAAAGCCAGACAGTTACATG	R13	TGTGAGACTGCATGTCGTATATTCCTGTTTCCTAAAGCTGCC
F14	TGCGTGAGCTGTACATCAGTAGAAAAGCCAGACAGTTACATG	R14	GCTCAGTGCGCTACTGGTATATTCCTGTTTCCTAAAGCTGCC
F15	CGATCATCTATAGACACAGTAGAAAAGCCAGACAGTTACATG	R15	ACTATCGCGCACGCAGGTATATTCCTGTTTCCTAAAGCTGCC
F16	CGACGTATCTGACAGTCAGTAGAAAAGCCAGACAGTTACATG	R16	TGACACTCTGCACGCGGTATATTCCTGTTTCCTAAAGCTGCC
F17	CACGTCACTAGAGCGACAGTAGAAAAGCCAGACAGTTACATG	R17	CAGACGTGACTGATATGTATATTCCTGTTTCCTAAAGCTGCC

F18	TGTCGCAGCTACTAGTCAGTAGAAAAGCCAGACAGTTACATG	R18	GCACTGTAGTGATCGTGTATATTCCTGTTTCCTAAAGCTGCC
F19	CATACGCTGTGTAGCACAGTAGAAAAGCCAGACAGTTACATG	R19	CAGTGCGAGACAGTAGGTATATTCCTGTTTCCTAAAGCTGCC
F20	AGTCGCATGACTGTGTCAGTAGAAAAGCCAGACAGTTACATG	R20	AGTAGTGCTACTCGACGTATATTCCTGTTTCCTAAAGCTGCC
F21	CAGTACTGCACGATCGCAGTAGAAAAGCCAGACAGTTACATG	R21	ATGCGAGATCTGCTCAGTATATTCCTGTTTCCTAAAGCTGCC
F22	GTGCTGAGCATCAGACCAGTAGAAAAGCCAGACAGTTACATG	R22	TGAGACATACTGAGTGGTATATTCCTGTTTCCTAAAGCTGCC
F23	CACTGATCGATATGCACAGTAGAAAAGCCAGACAGTTACATG	R23	ATGTGCACTAGTGTACGTATATTCCTGTTTCCTAAAGCTGCC
F24	TACAGTGTCTGCTGCGCAGTAGAAAAGCCAGACAGTTACATG	R24	TCAGCTGACGATGTGAGTATATTCCTGTTTCCTAAAGCTGCC
F25	TACAGATAGTGTAGCGCAGTAGAAAAGCCAGACAGTTACATG	R25	ACTGATGCGCACATGTGTATATTCCTGTTTCCTAAAGCTGCC
F26	TCGTAGAGCTCGAGACCAGTAGAAAAGCCAGACAGTTACATG	R26	CTACTCTCAGCAGTGAGTATATTCCTGTTTCCTAAAGCTGCC
F27	GAGCTGCGCACTCGATCAGTAGAAAAGCCAGACAGTTACATG	R27	ATCTACATCACGACTCGTATATTCCTGTTTCCTAAAGCTGCC
F28	GCGATGTCGCTATGTGCAGTAGAAAAGCCAGACAGTTACATG	R28	ATATAGTACAGCGTCTGTATATTCCTGTTTCCTAAAGCTGCC
F29	CGAGAGTCAGCGCATACAGTAGAAAAGCCAGACAGTTACATG	R29	GACACGACTAGATCGCGTATATTCCTGTTTCCTAAAGCTGCC
F30	TCACGATGAGCACGTACAGTAGAAAAGCCAGACAGTTACATG	R30	TACGAGTCTGTCATACGTATATTCCTGTTTCCTAAAGCTGCC
F31	GACTGAGATCATGATCCAGTAGAAAAGCCAGACAGTTACATG	R31	ACTCAGCTACATAGTGGTATATTCCTGTTTCCTAAAGCTGCC
F32	ACGACATGATACTGCTCAGTAGAAAAGCCAGACAGTTACATG	R32	ACGTATCATAGTGAGAGTATATTCCTGTTTCCTAAAGCTGCC
F33	ATACAGCACAGATGTGCAGTAGAAAAGCCAGACAGTTACATG	R33	GAGTCGTATCGCTCATGTATATTCCTGTTTCCTAAAGCTGCC
F34	ACAGTCGATATCTCTCCAGTAGAAAAGCCAGACAGTTACATG	R34	GCGATCACGAGTAGACGTATATTCCTGTTTCCTAAAGCTGCC

F35	GCTCGATCACATGACGCAGTAGAAAAGCCAGACAGTTACATG	R35	CTAGACGTACATGTCGGTATATTCCTGTTTCCTAAAGCTGCC
F36	GTCGTACACGTGCGACCAGTAGAAAAGCCAGACAGTTACATG	R36	TAGCAGTCACTGTGCGGTATATTCCTGTTTCCTAAAGCTGCC
F37	ACTCATATCTAGAGTGCAGTAGAAAAGCCAGACAGTTACATG	R37	CGTCATGCGATAGCTAGTATATTCCTGTTTCCTAAAGCTGCC
F38	ACTGATCTGTCGCGCTCAGTAGAAAAGCCAGACAGTTACATG	R38	GCGCAGTCGTCTGTATGTATATTCCTGTTTCCTAAAGCTGCC
F39	CACTAGCTCTGACTACCAGTAGAAAAGCCAGACAGTTACATG	R39	ATGAGCTACGTACAGAGTATATTCCTGTTTCCTAAAGCTGCC
F40	GCTGTCATGTACTAGCCAGTAGAAAAGCCAGACAGTTACATG	R40	GTCGCGAGTCTATCAGGTATATTCCTGTTTCCTAAAGCTGCC
F41	TATACATACACGCACTCAGTAGAAAAGCCAGACAGTTACATG	R41	ACATCGATCTGCACTAGTATATTCCTGTTTCCTAAAGCTGCC
F42	TGTGACGACGCGTCTCCAGTAGAAAAGCCAGACAGTTACATG	R42	AGTATAGCATAGACGCGTATATTCCTGTTTCCTAAAGCTGCC
F43	GACGTGAGCATGCACTCAGTAGAAAAGCCAGACAGTTACATG	R43	GTGAGAGCGTGACTCTGTATATTCCTGTTTCCTAAAGCTGCC
F44	CTCGATACGTGTAGCTCAGTAGAAAAGCCAGACAGTTACATG	R44	TGTCAGTAGATGACTCGTATATTCCTGTTTCCTAAAGCTGCC
F45	GTGTCTAGACAGCTGTCAGTAGAAAAGCCAGACAGTTACATG	R45	TCGTACGAGATCGACAGTATATTCCTGTTTCCTAAAGCTGCC
F46	GATGCATGCGTACGCACAGTAGAAAAGCCAGACAGTTACATG	R46	CTACATGTGACTCGAGGTATATTCCTGTTTCCTAAAGCTGCC
F47	TATCAGAGCAGCGATGCAGTAGAAAAGCCAGACAGTTACATG	R47	GCGCTATAGTGCTCGTGTATATTCCTGTTTCCTAAAGCTGCC
F48	TGTCTCGTGCTGAGACCAGTAGAAAAGCCAGACAGTTACATG	R48	CTGTGTAGAGAGCACAGTATATTCCTGTTTCCTAAAGCTGCC

Table D.4: Sequences of barcoded primer pair (MA_ORF15_F3 and MA_ORF15_R4) used in PacBio sequencing study.

Red: barcode sequence (16 bp), Black: primer sequence.

D.4 SNVs and in frame indels identified by both MinION and PacBio in cases and unaffected controls in *RPGR ORF15* screening study

No.	SNVs	No.	Indels
1.	rs12687163	**1.**	rs764268405
2.	rs12688514	**2.**	rs751710678
3.	rs147619484	**3.**	rs200824587
4.	rs5917557	**4.**	rs199663434
5.	rs761510942	**5.**	rs199896738
6.	rs756083902	**6.**	rs772859148
7.	rs62636730	**7.**	rs777850798
8.	rs5917557	**8.**	rs1164395323
9.	rs111787313	**9.**	rs201134185
10.	rs78736275	**10.**	
11.	rs151247357	**11.**	
12.	rs752979508	**12.**	
13.	rs770346168	**13.**	
14.	rs755138994	**14.**	
15.	rs867319999	**15.**	
16.	rs750364695	**16.**	
17.	rs776423695	**17.**	
18.	rs1328655753	**18.**	
19.	rs1250133030	**19.**	
20.	rs777091269	**20.**	
21.	rs201131185	**21.**	
22.	rs1343966503	**22.**	
23.	rs1279895966	**23.**	
24.	rs371230592	**24.**	
25.	rs1262060159	**25.**	

26.	rs1241379586	**26.**	
27.	rs1441737211	**27.**	
28.	rs868119890	**28.**	
29.	rs192410099	**29.**	
30.	rs187844918	**30.**	
31.	rs1412771722	**31.**	
32.	rs201655057	**32.**	
33.	rs766376194	**33.**	
34.	rs1312552676	**34.**	
35.	rs1064797363	**35.**	
36.	rs1341173289	**36.**	
37.	rs769179127	**37.**	

D.5 List of primer combinations tested in MinION sequencing study

Primer name	**Primer sequence**	**Product size**
MinORF15_F1	TTTCTGTTGGTGCTGATATTGCAGGGAAAGTTGCAAGCAAGA	2106 bp
MinORF15_R1	ACTTGCCTGTCGCTCTATCTTCAGTGCCCGTTATATGCAAGG	
MinORF15_F2	TTTCTGTTGGTGCTGATATTGCAAACCCATAATATCCAAATCCA	1875 bp
MinORF15_R1	ACTTGCCTGTCGCTCTATCTTCAGTGCCCGTTATATGCAAGG	
MinORF15_F3	TTTCTGTTGGTGCTGATATTGCCGGTATGGCAGGAAATTGAT	1920 bp
MinORF15_R1	ACTTGCCTGTCGCTCTATCTTCAGTGCCCGTTATATGCAAGG	
MinORF15_F4	TTTCTGTTGGTGCTGATATTGCACGCAGCCAGCTACGACTAT	2554 bp

MinORF15_R2	ACTTGCCTGTCGCTCTATCTTCCACAGCTGCATCAGTTGCTT	
MinORF15_F1	TTTCTGTTGGTGCTGATATTGCAGGGAAAGTTGCAAGCAAGA	2784 bp
MinORF15_R3	ACTTGCCTGTCGCTCTATCTTCGAAGAAACTGAGGCCCAATG	
MinORF15_F5	TTTCTGTTGGTGCTGATATTGCCATGGAAGGTGCAAGTGAGA	2114 bp
MinORF15_R2	ACTTGCCTGTCGCTCTATCTTCCACAGCTGCATCAGTTGCTT	
MinORF15_F6	TTTCTGTTGGTGCTGATATTGCCAGGCTGACACTGATGGAGA	2526 bp
MinORF15_R4	ACTTGCCTGTCGCTCTATCTTCAGGCCAAAATTTACCAGTGC	
MinORF15_F7	TTTCTGTTGGTGCTGATATTGCTGGAGGACACAGTCAGAAGG	2978 bp
MinORF15_R3	ACTTGCCTGTCGCTCTATCTTCGAAGAAACTGAGGCCCAATG	
MinORF15_F8	TTTCTGTTGGTGCTGATATTGCAGCAGCCTGAGGCAATAGAA	1991 bp
MinORF15_R2	ACTTGCCTGTCGCTCTATCTTCCACAGCTGCATCAGTTGCTT	
MinORF15_F9	TTTCTGTTGGTGCTGATATTGCTGATGAAGTGGAAACTGACCA	1919 bp
MinORF15_R5	ACTTGCCTGTCGCTCTATCTTCTGTCTGACTGGCCATAATCG	
MinORF15_F10	TTTCTGTTGGTGCTGATATTGCTGAGTCCTTTCTTTGGCAACT	2214 bp
MinORF15_R6	ACTTGCCTGTCGCTCTATCTTCTGACATAAAATCAATTTAATAACACG	
MinORF15_F11	TTTCTGTTGGTGCTGATATTGCTGGCAGAAATAGCAGGTATGAA	2392 bp
MinORF15_R7	ACTTGCCTGTCGCTCTATCTTCGGGGGAAATACACGAAAAT	

MinORF15_F1	TTTCTGTTGGTGCTGATATTGCAGGGAAAGTTGCAAGCAAGA	2419 bp
MinORF15_R8	ACTTGCCTGTCGCTCTATCTTCCCTTAAGCATCTGTGCCATTT	
MinORF15_F1	TTTCTGTTGGTGCTGATATTGCAGGGAAAGTTGCAAGCAAGA	2559 bp
MinORF15_R9	ACTTGCCTGTCGCTCTATCTTCTTCTTTAAAATGTGAATGCCTCAA	

Table D.5: Sequences of different primer combinations tested for pre-indexing PCR amplification of *ORF15* region for MinION sequencing. The universal sequencing tags are underlined. The primer combination worked is orange shaded. F= Forward primer, R= Reverse primer.