The molecular genetics of *ABCA4*-related disease and allied retinal disorders

Benjamin Mc Clinton

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

The University of Leeds
School of Medicine

April 2023

The candidate confirms that the work submitted is their 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 3 of the thesis has appeared in publication as follows:

McClinton, B., Watson, C. M., Crinnion, L. A., McKibbin, M., Ali, M., Inglehearn, C. F., & Toomes, C. (2023). Haplotyping Using Long-Range PCR and Nanopore Sequencing to Phase Variants; Lessons Learned From the ABCA4 Locus. *Laboratory investigation; a journal of technical methods and pathology*, 100160. Advance online publication. https://doi.org/10.1016/j.labinv.2023.100160

The author performed long-range primer design and optimization, performed long-range PCR and nanopore sequencing of all samples, analysis of resulting data and drafting of the manuscript. The contribution of the other authors was to provide phenotype details of cases, aid analysis, and supervision and co-authoring of the paper.

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

McClinton, B., Crinnion, L. A., McKibbin, M., Mukherjee, R., Poulter, J. A., Smith, C. E. L., Ali, M., Watson, C. M., Inglehearn, C. F., & Toomes, C. (2023). Targeted nanopore sequencing enables complete characterisation of structural deletions initially identified using exon-based short-read sequencing strategies. *Molecular genetics & genomic medicine*, e2164. Advance online publication. https://doi.org/10.1002/mgg3.2164

The author performed long-range primer design and optimization, performed long-range PCR and nanopore sequencing of all samples, analysis of resulting data and drafting of the manuscript. The contribution of the other authors was to provide phenotype details of cases, initially identify the putative deletions, contributed to analysis, and supervision and co-authoring of the paper.

The work in Chapter 5 of the thesis has appeared in publication as follows:

Mc Clinton, B., Corradi, Z., McKibbin, M., Panneman, D. M., Roosing, S., Boonen, E. G. M., Ali, M., Watson, C. M., Steel, D. H., Cremers, F. P. M., Inglehearn, C. F., Hitti-Malin, R. J., & Toomes, C. (2023). Effective smMIPs-Based Sequencing of Maculopathy-Associated Genes in Stargardt Disease Cases and Allied Maculopathies from the UK. Genes, 14(1), 191. https://doi.org/10.3390/genes14010191

The author performed case selection and preparation, variant analysis and interpretation, interpretation of genotype-phenotype relationship and drafting the manuscript. The other authors performed library preparation, sequencing, and bioinformatic analysis. The other authors also provided variant interpretation guidelines which were followed in this experiment. The other authors provided phenotypic details and interpretation of genotype-phenotype relationship as well as supervision and co-authoring of the paper.

Acknowledgements

Firstly I would like to thank my lead supervisor, Dr Carmel Toomes for her invaluable supervision, support and advice throughout my PhD. I would also like to thank my supervisors Prof Chris Inglehearn, Dr Manir Ali and Dr Christopher Watson for their constant guidance, advice and expertise. I am grateful to my supervisors for the opportunity to work on such an interesting topic under their expert guidance. Additionally, I would like to thank the StarT network for funding this research and the members and fellows of the StarT network for making this such an enjoyable experience.

I would also like to thank my colleagues at St James' for their help, friendship and eternal patience, particularly George, Moe, Claire, Rowan, Kasia, Magda and Ummey.

I was fortunate to be able to go on two secondments during my project. I would like to thank Prof Frans Cremers, Dr Rebekkah Hitti-Malin and Zelia Corradi for facilitating a very productive and enjoyable secondment with Radboud MC. I would also like to thank Dr Gavin Arno for hosting me in UCL and for providing his encouragement and expertise during all CATCH experiments.

Finally I would like to thank my family and friends for their endless support and encouragement, during my PhD. In particular, I want to thank my partner Francesca for putting up with me this long, keep it up. Can't wait for what comes next.

Abstract

Inherited retinal dystrophies (IRDs) are a leading cause of blindness in people of working age, but many patients do not receive an informative result following genetic testing. An accurate genetic diagnosis is important for patients to inform genetic counselling and family planning, and a genetic diagnosis is a prerequisite for patients to be included in a growing number of clinical trials of therapies for IRDs.

Due to the read lengths obtained during next generation sequencing, it is difficult to establish the phase of detected variants. As a result, compound heterozygous variants are commonly assumed to be in *trans* and it is challenging to study interactions between variants such as complex alleles and *cis* acting modifiers. To address this, a method to phase contiguous segments of the *ABCA4* locus by long-range PCR and nanopore sequencing was developed. *ABCA4* was selected for study as pathogenic variants in *ABCA4* are the most common cause of IRD and the study of *ABCA4* has shed light on disease caused by mutations in other genes. This project was undertaken as part of the StarT consortium, which was established to investigate *ABCA4*-associated disease in detail. Using this method, it was shown that accurate phasing of large sections of *ABCA4* was possible, albeit practically challenging. An additional method to phase *ABCA4* using ultra-long reads obtained through amplification-free enrichment was then employed to phase variants of interest in two cases with *ABCA4*-related disease.

Deletions in genes related to IRDs are challenging to characterise following targeted sequencing. A workflow to characterise deletions by long-range PCR and nanopore sequencing was developed and used to characterise clinically relevant deletions in *CNGA1, CNGB1, EYS* and *PRPF31* to a nucleotide level resolution.

A single molecule Molecular Inversion Probe (smMIP) based panel had recently been developed to target 107 genes and loci associated with macular disease. This panel was used to screen a cohort of 57 UK cases with little or no prior screening. This screen resulted in the discovery of likely causative mutations in 63.2% of the examined cases, a much higher solve rate than in other cohorts, and the identification of four novel variants. The phenotypes of

cases with *ABCA4*-associated disease were compared to a previously proposed genotype-phenotype model. This comparison found the model to be largely accurate and clinically useful.

It is hoped that these findings will directly impact individuals affected by IRDs by improving diagnostic rates following genetic testing and allowing for more precise results. Full characterisation of causative mutations has the potential to benefit patients by facilitating genetic counselling and family planning and for inclusion in a growing number of gene therapy clinical trials.

Table of Contents

List of Tables

List of Figures

Abbreviations

A2E	N-retinylidene-N-retinylethanolamine
A2PE	Phosphatidylpyridinium bisretinoid
AAV	Adeno-associated virus
ABC	ATP-binding cassette
ABCA4	ATP-binding cassette, subfamily A, member 4
ACMG	American College Medical Genetics
AMD	Age related macular degeneration
AONs	Antisense oligonucleotides
ATP	Adenosine triphosphate
AWAT	Acyl-CoA wax alcohol acyltransferase 2
BACs	Bacterial artificial chromosomes
BBS1	Bardet-Biedl syndrome 1
bp	Base pair
Cas	CRISPR-associated (Cas) endonuclease
CATCH	Cas9 assisted targeting of chromosome segments
CFTR	Cystic fibrosis transmembrane conductance regulator
CNGA1	Cyclic nucleotide-gated channel, alpha-1
CNGB1	Cyclic nucleotide-gated channel, beta-1
CNV	Copy number variation
CRD	Cone-rod dystrophy
CRISPR	Clustered regularly interspaced short palindromic repeats
CsgG	Curlin sigma S-dependent growth subunit G
ddNTPs	Dideoxynucleotides
DECIPHER	Database of chromosomal imbalance and phenotype in humans using Ensembl resources
DES1	Dihydroceramide desaturase-1
DMSO	Dimethyl sulfoxide

dNTP	Deoxyribonucleotide
EDTA	Ethylenediaminetetraacetic acid
ExoSAP	Exonuclease I, shrimp alkaline phosphatase
EYS	Eyes shut homolog
FEVR	Familial exudative vitreoretinopathy
GATK	Genome analysis toolkit
GEMs	Gel bead-in-emulsions
GIAB	Genome in a Bottle
gnomAD	Genome Aggregation Database
HiFi	High fidelity
HMW DNA	High molecular weight DNA
IGV	Integrative Genomics Viewer
IMPG2	Interphotoreceptor matrix proteoglycan 2
IPRB	Interphotoreceptor retinoid binding
IRD	Inherited retinal dystrophy
LB	Luria-Bertani
LORD	Late onset retinal dystrophy
LOVD	Leiden Open Variation Database
LRAT	Lecithin retinol acyl transferase
MD	Macular dystrophy
MIPs	Molecular inversion probes
N-ret-PE	N-Retinylidene-phosphatidylethanolamine
ND4	Nicotine adenine dinucleotide dehydrogenase, subunit 4
NDP	Norrin
NEB	New England Biolabs
NGS	Next generation sequencing
NPB	Nuclei preparation buffer
nt	Nucleotide
ONT	Oxford Nanopore Technologies

ORF15	Open reading frame 15
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PROM1	Prominin 1
PRPH2	Peripherin 2
qPCR	Quantitative PCR
RDH5	11-*cis*-retinol dehydrogenase
RDH8	Retinol dehydrogenase 8
RP	Retinitis pigmentosa
RPE	Retinal pigment epithelium
RPE65	Retinoid isomerohydrolase
RPGR	Retinitis pigmentosa GTPase regulator
smMIPs	Single-molecule molecular inversion probes
SMRT	Single-molecule real time
SNP	Single nucleotide polymorphism
STGD	Stargardt
TAE	Tris-acetate-EDTA
TCF4	Transcription factor 4
TE	Tris-ethylenediaminetetraacetic acid (EDTA)
UCL	University College London
UCSC	University of California, Santa Cruz
UMI	Unique molecule identifiers
USH2A	Usher syndrome type 2A
VUS	Variant of uncertain significance
WBC	White blood cells
WES	Whole exome sequencing
WGS	Whole genome sequencing
ZMW	Zero-mode waveguide

1 Introduction

1.1 General overview

Inherited retinal diseases (IRDs) represent a leading cause of blindness in people of working age in developed countries (Liew et al., 2014; Heath Jeffery et al., 2021), affecting approximately 1 in 2000-4000 people (Stone et al., 2017; Pontikos et al., 2020). Up to 36% of the human population are unaffected carriers for at least one recessive IRD causing mutation (Hanany et al., 2020). Over 300 genes and loci have been found to cause a wide variety of IRDs with overlapping features (https://web.sph.uth.edu/RetNet/sum-dis.htm#D-graph, accessed 21/2/2023), which although individually rare, collectively represent a significant disease burden. Of these, mutations in ATP-binding cassette, sub-family A, member 4 (*ABCA4*) represent the single most common cause of IRDs (Cremers et al., 2020). Pathogenic mutations in *ABCA4* cause autosomal recessive macular degeneration which commonly presents as early onset, severe disease. The frequency of *ABCA4*-related disease is still debated, however over 2000 pathogenic mutations in *ABCA4* have been identified, some of which are observed at higher frequencies (Cornelis et al., 2022). Due to its significance as a common cause of disease, much effort has been made to comprehensively characterise *ABCA4* and improve genetic testing. While advances in sequencing techniques have led to an increased understanding of *ABCA4* disease, significant areas remain underexplored.

This project was undertaken as part of a wider consortium, the StarT Network, created to characterise the function of *ABCA4*, search for sources of missing heritability and to develop potential novel treatments and models. As part of that, this project was established to search for missing heritability by using emerging long-read sequencing technologies to sequence the entirety of *ABCA4* and phase pathogenic variants. In parallel, these technologies were used to characterise copy number variants in IRD genes, a potential source of missing heritability in ABCA4 disease.

1.2 General structure of the eye

The human eye is a highly complex structure, consisting of a central vitreous fluid enclosed by three outer layers (Figure 1.1). The outermost of these is composed of the sclera and cornea. The sclera is the structural tissue which maintains the shape of the eye (Willoughby et al., 2010). The cornea is situated in the front of the eye. Both the cornea and the sclera are protective tissue which are highly similar, however the cornea is transparent and focuses light entering the eye (Müller et al., 2003).

The next layer of the eye is composed of the ciliary body, iris, and choroid. The ciliary body and iris work in tandem to control light entering the eye. The ciliary body influences the shape of the lens and holds it in place, while the iris controls the size of the pupil (Willoughby et al., 2010). The choroid is a highly vascular tissue and provides the basis of the vasculature which serves the outer retinal layers (Gao et al., 2020).
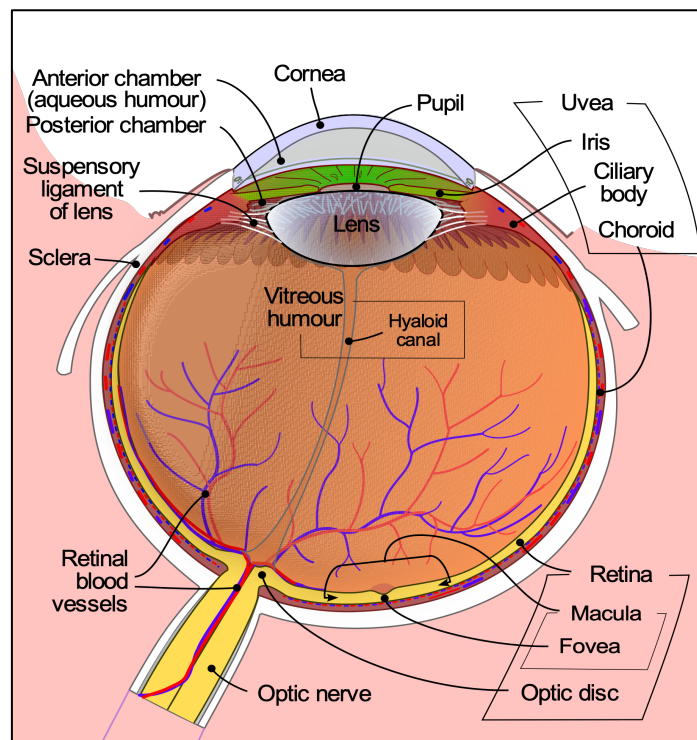
Figure 1.1 Diagram of human eye

Diagram of the human eye with major structures labelled. Adapted from "Schematic diagram of the human eye en" by Rh Castilhos and J March, licensed under CC BY-SA 3.0.

The innermost layer of the eye is the retina, the tissue responsible for light detection. Finally, the centre of the eye is the vitreous humour. This has several functions, including maintaining ocular pressure and homeostasis (Bhartiya et al., 2022).

1.2.1 The retina

1.2.1.1 Structure of the retina

The retina is an outgrowth of the brain and is a highly complex structure largely conserved across vertebrates (Hoon et al., 2014). In humans, the retinal progenitor cells form seven neuronal classes of cells (Luo et al., 2008). In order from the outermost layer of the retina to the innermost (i.e. the back of the eye moving towards the centre), this is organised into the retinal pigment epithelium (RPE), photoreceptors (cones and rods), horizontal cells, bipolar cells, amacrine cells and retinal ganglion cells, while Müller glia cells are arranged in a vertical orientation through the retina (Figure 1.2). Light must pass through the entire retina before it is detected by the photoreceptor. This has been theorized to induce scattering and reduce visual acuity, however the cells of the retina have been shown to be highly adapted to prevent this (Agte et al., 2011).

Additionally, the retina is supported by a network of tissues. The RPE is the primary supportive tissue of the retina (Strauss, 2005). It is a monolayer of cells which supports the retina by protecting photoreceptors from external stresses, forming part of the blood brain barrier and maintaining homeostasis by transporting nutrients to the photoreceptors (Simó et al., 2010).
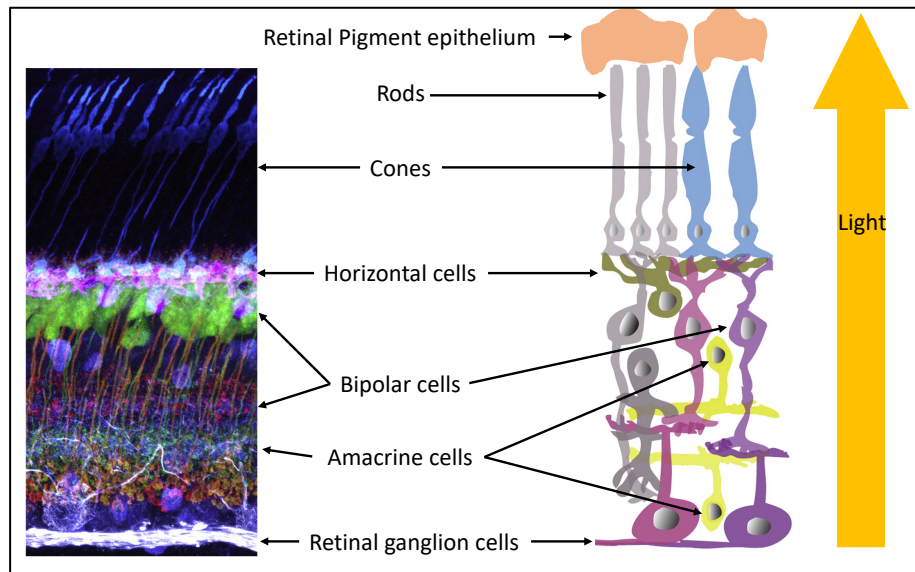
Figure 1.2 Diagram of retina

Transverse diagram of a healthy retina with supporting structures and neural cells labelled. The left side shows immunostaining of a mouse retina with cones (blue), horizontal cells (pink), bipolar cells (terminals red, cell bodies green), amacrine cells (purple), and ganglion cells (white). The right shows a schematic with the corresponding structures labelled. Reproduced with permission and adapted from (Hoon et al., 2014).

1.2.1.2 Light perception in the retina

The photoreceptors are the light sensitive cells in the retina. Photoreceptors exhibit distinct morphology, and are composed of the outer segment, the inner segment, the cell soma and the synaptic terminal. The outer segment is the site of light detection and is a specialised primary, non-motile, sensory cilia (Cohen, 1961; Röhlich, 1975; Besharse and Horst, 1990). The outer segment of the photoreceptor is composed of hundreds of stacked discs which are continuously renewed (Sjostrand, 1953). The outer segment discs are stacked vertically in the retina to capture maximum photons travelling along their axis (Goldberg et al., 2016). The outer segment discs are continuously shed and renewed to dispose of toxic biproducts while maintaining photoreceptor length (Kocaoglu et al., 2016). One of the ways in which the RPE supports photoreceptor function is by phagocytosis of shed outer segment discs (Sjostrand, 1953). Each RPE cell phagocytoses approximately 20 outer segment discs a day (Hartzell et al., 2005).

Photoreceptors can be divided into rods and cones based on structure and function (Figure 1.3). Rods are generally used for achromatic light detection

and movement perception, particularly in low light environments (Lamb, 2015), however at the expense of visual acuity (Mustafi et al., 2009). Rods have a characteristic longer outer segment, with outer segment discs which are not connected to the plasma membrane (Mustafi et al., 2009). In contrast cones are used for higher resolution vision and colour detection (Lamb, 2015). Cones have shorter outer segments, with discs continuously connected to the plasma membrane, greatly increasing the surface area of the discs (Arikawa et al., 1992). In humans, rods are more abundant than cones, with cones comprising approximately 5% of the total photoreceptors (Busck and Osterber, 1935; Lamb, 2015).

Visual detection in photoreceptors is mediated by opsins, each of which is sensitive to different wavelengths of light (Shichida and Imai, 1998; Terakita, 2005; Goldberg et al., 2016). The opsins are transmembrane proteins which form a key part of the disc membrane (Humphries et al., 1997). The visual opsins are continuously renewed by incorporation of newly synthesised opsins into the membranes of the renewed discs (Lamb and Curtin, 2022). In humans, rhodopsin is the primary opsin found in rods, it detects green-blue light at very high sensitivity (Wald, 1951; Wald and Brown, 1956; Nathans and Hogness, 1984). In contrast, cones have three opsins which are used to detect light in different wavelengths, corresponding to red, green and blue light (Nathans et al., 1986). Trichromatic vision in humans is mediated through allelic diversity of the long (red) and medium (green) wave opsins (Bowmaker and Hunt, 2006). Light detection by the opsins is performed via the visual chromophore, 11-*cis*-retinal (Wang et al., 2014).This binds to the opsin in the outer segment discs of the photoreceptor, creating a photosensitive visual pigment.
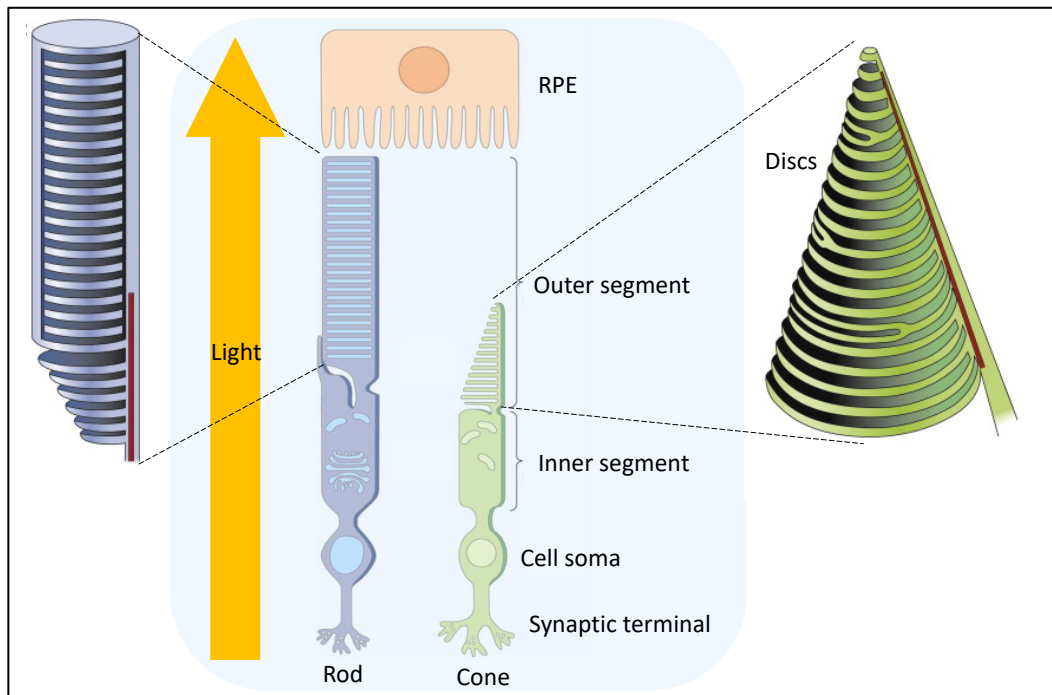
Figure 1.3 Photoreceptor structure

> In the centre the structure of the photoreceptors are shown with the direction of light indicated. Flanking this, the structure of the outer segments of rods (left) and cones (right) are shown. Reproduced with permission from (Mustafi et al., 2009; Kiser et al., 2012).

1.2.1.3 The phototransduction cascade

11-*cis*-retinal binds the opsin in the intradiscal space (Hubbard and Wald, 1952). When the photon is absorbed by 11-*cis*-retinal it prompts isomerisation to all-*trans*-retinal (Kiser and Palczewski, 2016). Light induced isomerisation of 11-*cis*-retinal to all-*trans*-retinal results in activation of the phototransduction cascade, leading to light detection (Pugh and Lamb, 1993; Luo et al., 2008; Molday et al., 2022).

The initial steps in the phototransduction cascade are similar between rods and cones. In rods, isomerisation of 11-*cis*-retinal leads to activation of rhodopsin to an active configuration, metarhodopsin II (Leskov et al., 2000) (R*, Figure 1.4). In cones the equivalent cone opsin is activated (Lamb and Curtin, 2022). This then leads to repeated recruitment of, and contact with, a G protein transducin (Fain et al., 2010). This contact results in the release of GDP and binding of GTP by the transducin, which causes the alpha subunit to disassociate. The disassociated alpha subunit remains bound to the recruited GTP (G*, Figure 1.4) (Fain et al., 2010). G* then binds to one of two

γ-subunits of phosphodiesterase (PDE, Figure 1.4) (Jeon et al., 2005). When both γ-subunits are bound by G*s, PDE becomes activated (PDE**, Figure 1.4). Activated PDE** then hydrolyses cyclic GMP (cGMP to GMP), reducing available cytoplasmic cyclic GMP (Barret et al., 2022). Reduced cGMP levels cause cyclic nucleotide gated channels (CNGC) to go from an open conformation to a shut conformation (Barret et al., 2022). The closed CNGCs result in hyperpolarisation of the cell, which then causes voltage gated calcium channels to close. Closure of the voltage gated calcium channels results in reduced glutamate secretion by the photoreceptor, leading to transmission of the signal (Hoon et al., 2014; Goldberg et al., 2016).

The fundamental principle of the phototransduction cascade is similar in rods and cones, however in cones the proteins responsible for the phototransduction cascade reside in the plasma membrane, whereas in rods the majority of the proteins reside in the disc membranes (Lamb and Curtin, 2022).
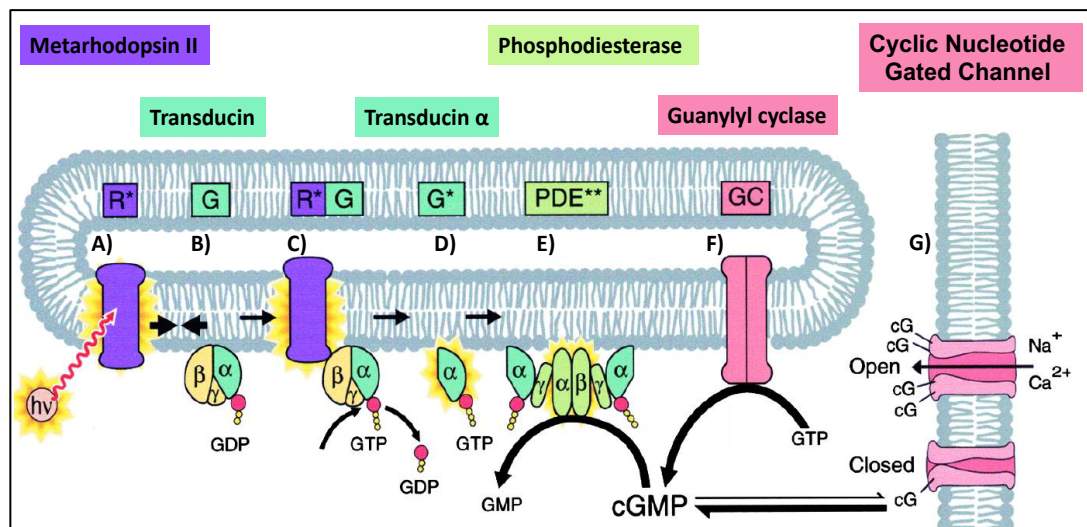
Figure 1.4 The phototransduction cascade

    A) A photon strikes rhodopsin, activating it to metarhodopsin II. **B)** Activated metarhodopsin II recruits transducin. **C)** Transducin is activated, replacing GDP with GTP. **D)** Transducin subunit α disassociates from transducin. **E)** The inhibitory subunits of PDE are bound by activated transducin α. Activated PDE then hydrolyses cGMP at an increased rate. **F)** Guanylyl cyclase continues to synthesise cGMP, but cGMP levels fall due to increased PDE activity **G)** Reduced cytoplasmic cGMP levels result in the closing of channels, hyperpolarising the photoreceptor. Adapted with permission from (Leskov et al., 2000)

Activation of the phototransduction cascade conveys the perception of the photon from the photoreceptor, through the retina to the brain for visual processing. Rods and cones utilise different but overlapping pathways to convey this signal, which minimizes signalling noise while retaining maximum sensitivity (Pahlberg and Sampath, 2011). The signalling pathways in the retina allow for initial segregation and processing of visual stimuli prior to signalling to the brain (Ichinose and Habib, 2022). Very broadly, rods and cones are connected to bipolar cells, which connect to ganglion cells which then pass the signal to the brain (Euler et al., 2014). In parallel, horizontal cells and amacrine cells modulate the signal pathways (Ichinose and Habib, 2022).

While rods and cells connect via separate but overlapping pathways, there is significant interaction between the processes at all levels of the signalling cascade (Wässle, 2004). Rods and cones connect directly to each other allowing direct integration between the pathways (Euler et al., 2014). Further, both rods and cones connect to cone bipolar cells, further integrating this signal (Völgyi et al., 2004) and the pathway primarily used by rods integrates into the cone bipolar pathways through A II amacrine cell (Wässle et al., 1991; Fain and Sampath, 2018) Horizontal cells also connect rods and cones and serve to provide another layer of interaction between the pathways (Szikra et al., 2014).

1.2.1.4 The visual cycle

Following activation of the phototransduction cascade, the reactive all-*trans*-retinal must then be cleared from the photoreceptor disc membranes for recycling back to 11-*cis*-retinal (Tsybovsky and Palczewski, 2014; Molday et al., 2022). A highly regulated system, known as the visual cycle, effectively clears all-*trans*-retinal from the photoreceptors (Figure 1.5). Briefly, the first step of the visual cycle involves all-*trans*-retinol dehydrogenase (RDH8), which is located in the outer segment cytoplasm. RDH8 reduces all-*trans*-retinal to all-*trans*-retinol (Palczewski et al., 1994; Rattner et al., 2000). All-*trans*-retinol can then be transported from the outer segment disc to the RPE by interphotoreceptor retinoid binding (IRPB) (Parker and Crouch, 2010). All-*trans*-retinol is then converted to retinyl esters, by lecithin retinol acyl transferase (LRAT) (Hu and Bok, 2010). At this point the retinyl esters can be stored as a pool of available substrate in the RPE and additional substrate,

derived from vitamin A, can be delivered to the RPE from blood circulation (Wang et al., 2014). Retinoid isomerohydrolase (RPE65) then converts this to 11-*cis*-retinol (Moiseyev et al., 2005), which is then in turn converted to 11-*cis*-retinal by 11-*cis*-retinol dehydrogenase (RDH5) (Saari, 2012). The regenerated 11-*cis*-retinal can then be delivered back to the outer segment, where it can bind to the opsin.

However, a fraction of dissociated all-*trans*-retinal reacts with phosphatidylethanolamine to form N-retinylidene-phosphatidylethanolamine (N-ret-PE) which remains trapped in the disc membrane (Anderson and Maude, 1970). This is unavailable for processing by RDH8. ABCA4 is a flippase which functions to make sequestered all-*trans*-retinal available for processing in the visual cycle. ABCA4 flips N-ret-PE with the retinylidene in the luminal space to its retinylidene facing into the cytoplasm (Quazi et al., 2012; Quazi and Molday, 2014; Molday et al., 2022). This allows RDH8 to act on the dissociated all-*trans*-retinal, beginning the visual cycle and allowing effective clearing of all-*trans*-retinal from the photoreceptor to the RPE where it is further processed (Maeda et al., 2008). Although alternate pathways of 11-*cis*-retinal regeneration have been found (Wang and Kefalov, 2009; Kaylor et al., 2017), dysfunction in the visual cycle results in pathogenic damage to the retina, indicating the importance of the visual cycle.

A complementary system exists in cone cells to clear all-*trans*-retinal (Muniz et al., 2009). This system, known as the intra-retinal cone visual cycle, recycles all-*trans*-retinal in the cone cell and the Müller cell following release of all-*trans*-retinal from the cone pigment (Tsin et al., 2018). As in the classical visual cycle, this is converted to all-*trans*-retinol. In this cycle, all-*trans*-retinol is instead transported to the Müller cells where it is converted to 11-*cis*-retinol by dihydroceramide desaturase-1 (DES1) (Tsin et al., 2018). Then it is converted to 11-*cis*-retinyl esters by acyl-CoA wax alcohol acyltransferase 2 (AWAT) (Tsin et al., 2018). Following this, the retinyl esters are converted to 11-*cis*-retinol by retinyl ester hydrolase (Schreiber et al., 2012). At this point, the 11-*cis*-retinol is transported back to the cone where it is converted to available 11-*cis*-retinal.
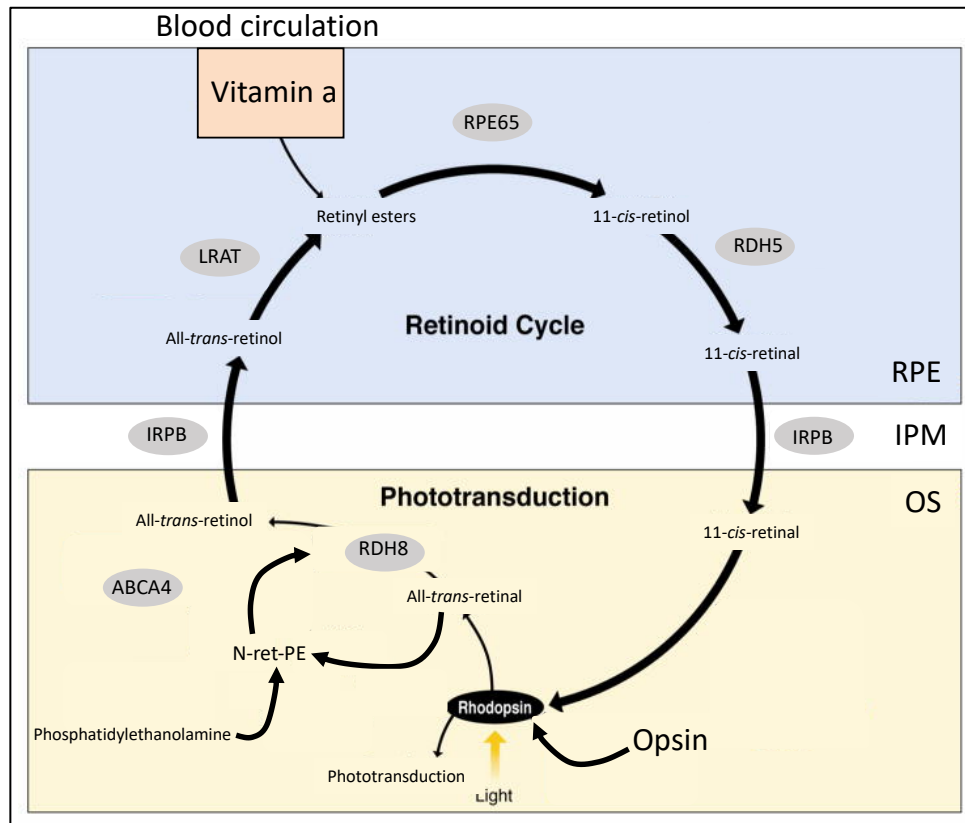
Figure 1.5 The visual cycle

Schematic showing the visual cycle in rods with key enzymes labelled. IPM refers to the interphotoreceptor matrix. Adapted with permission from (Kiser et al., 2012).

1.2.1.5 Layout of the retina

While the structure of the retina is largely conserved in function across vertebrates, the distribution of rods and cones shows significant divergence across species (Curcio et al., 1990). In humans, rods form the bulk of photoreceptors (approximately 91 million rods compared to 4.5 million cones) and these are generally distributed throughout the periphery of the retina (Purves et al., 2001). Rods are highly adapted to detection of light in low light conditions. As a result of this, they are less well adapted to light detection during the day, with slower response kinetics and longer recovery times compared to cones (Lamb and Pugh, 2004; Lamb et al., 2015).

Rods are extremely effective at photon detection in low light conditions compared to cones. While rods are only slightly more sensitive to photons than cones, rods are far more effective at amplifying the signal obtained from a single photon. Photoreceptors can spontaneously signal in the absence of

a photon (Ala-Laurila et al., 2011). Further, as each rod bipolar cell is connected to multiple rods, each rod bipolar cell will receive a signal from a rod cell approximately every 4 seconds, even in total darkness (Lamb, 2015). In darkness rod bipolar cells effectively filter this noise, allowing for highly sensitive detection of single photons (Berntson et al., 2016). In contrast in cones, increased noise levels reduce sensitivity to single photons (Lamb, 2015). The rod bipolar cells join the cone signalling pathway, where they are highly effective at conveying single photon signals to the retinal ganglion cells, without introducing noise during light conditions when the rods are saturated (Ala-Laurila and Rieke, 2014).

In contrast, cones are highly adapted for light perception in good light conditions at high visual acuity. Cones are largely restricted to the centre of the retina, forming the macula (Figure 1.6) (Provis et al., 2005). The macula is responsible for central vision in humans in light conditions and constitutes less than 4% of the area of the retina (Provis et al., 2005). Contained within the macula is the fovea, the region of the eye responsible for highest acuity vision. The fovea is highly adapted for capturing high acuity vision. The fovea is served by an increased number of ganglion cells compared to the peripheral retina and has a reduced number of rods (Curcio and Allen, 1990). Further, the fovea is not served by the retinal vasculature, instead relying on the choriocapillaris. The absence of the vasculature seen in other parts of the retina is theorised to reduce light scattering (Provis et al., 2005). The centre of the fovea, the foveola, is responsible for highest acuity vision and does not contain any rods (Packer et al., 1989; Hendrickson, 1994). In the foveola, the overlying layers of the retina are diverted to minimize light scattering, the cones in this region are distally innervated and this region lacks overlying retinal blood vessels (Purves et al., 2001; Lujan et al., 2011). As only cones are present in the foveola, incoming light is minimally scattered and very high visual acuity is obtained (Provis et al., 2005).
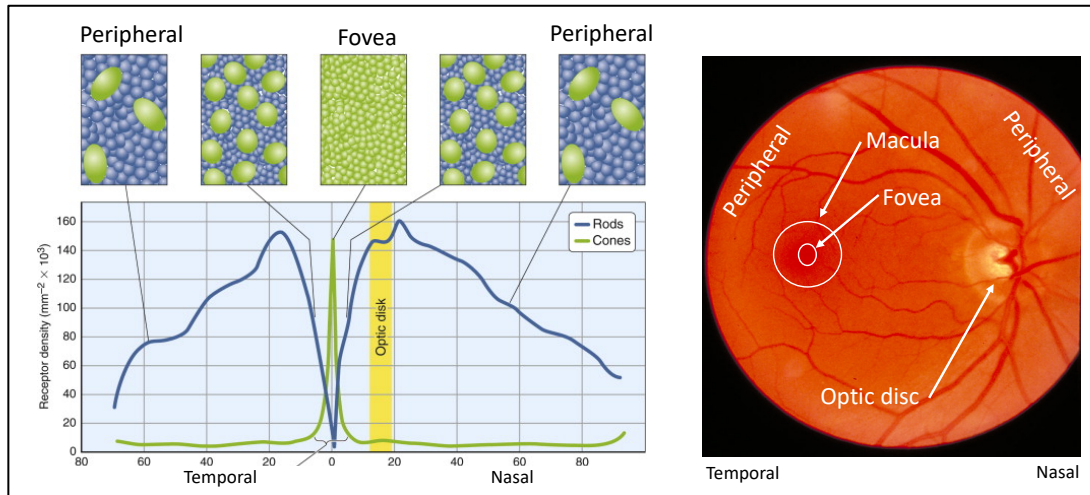
Figure 1.6 Fundus image of healthy retina

> **Left** Distribution of rods and cones throughout the macula. Adapted with permission from (Mustafi et al., 2009)**. Right** Image of a retina with the optic disc containing the optic nerve, macula and fovea labelled. Adapted from "Ophthalmoscopy" by Community Eye Health, licensed under CC BY-NC 2.0.

1.3   Inherited retinal dystrophies

The retina is a highly complex system and reflecting this, mutations in over 300 genes have been found to result in pathological disruption (https://web.sph.uth.edu/RetNet/sum-dis.htm#D-graph, accessed 21/2/2023). Mutations in genes with a broad range of functions have been associated with IRDs, with a variety of pathways affected (Ortega and Jastrzebska, 2022). For example mutations in genes associated with cilia development and function (Adams et al., 2007; Mockel et al., 2011), the phototransduction cascade (Kiser and Palczewski, 2016; Ortega and Jastrzebska, 2022), disc morphogenesis (Yang et al., 2008a) and the visual cycle (Tsin et al., 2018) among others have been described.

IRDs represent a spectrum of conditions and are generally classified based on clinical manifestation with significant overlap between conditions. A number of IRDs have been found to be caused by mutations in more than one gene and a number of genes have been found associated with more than one IRD (Figure 1.7) (Sheck et al., 2021).
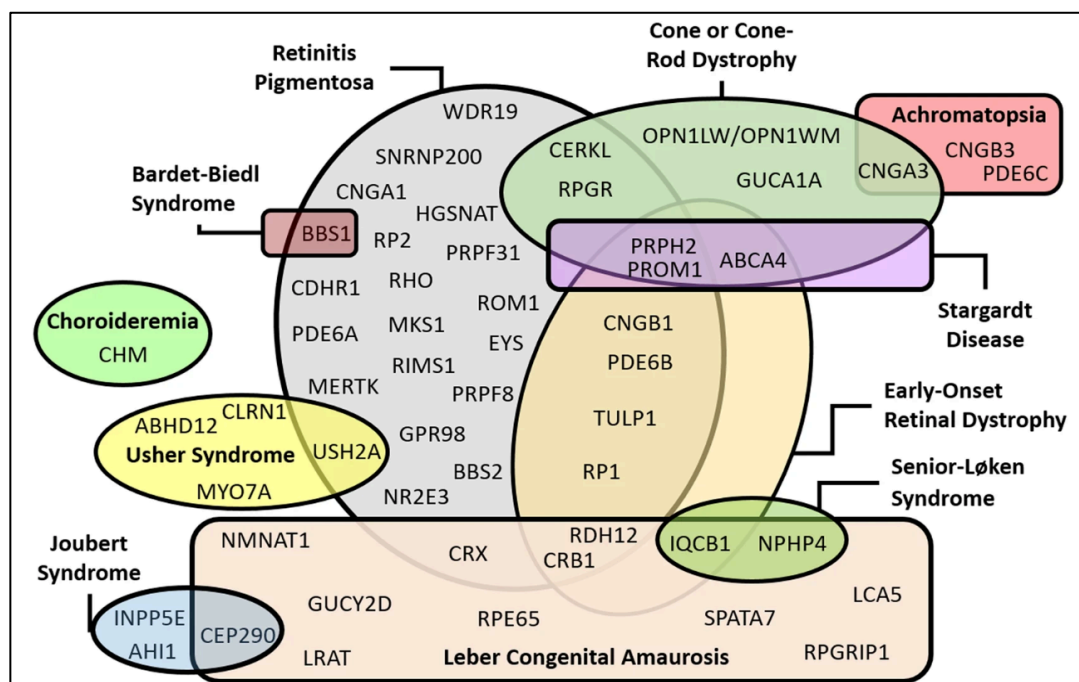
Figure 1.7 Genes associated with IRDs

> IRDs are a genetically heterogenous group of conditions with overlapping clinical features. While some IRDs are caused by mutations in only one gene, others are caused by mutations in a range of genes. Similarly, some mutations in some genes result in a single IRD, whereas mutations in other genes can result in a range of IRDs. Adapted from (Motta et al., 2018) under CC BY 4.0.

IRDs generally do not uniformly affect the retina. For example retinitis pigmentosa (RP) (OMIM: #268000) is the most common IRD, affecting 1 in 3,000 to 5,000 people (Haim, 2002). Mutations in 75 genes have been associated with non-syndromic RP (https://web.sph.uth.edu/RetNet/sum-dis.htm?csrt=17870442760574951690#B-diseases, accessed 13/4/2023), representing a broad spectrum of pathways. Broadly, RP initially affects rods and the peripheral retina, resulting in night blindness and loss of peripheral vision (Newton and Megaw, 2020). In contrast, macular dystrophies (MD) primarily affect the macula initially, generally resulting in initial loss of visual acuity and central vision. Mutations in 18 genes have been associated with non-syndromic MD (https://web.sph.uth.edu/RetNet/sum-dis.htm?csrt=17870442760574951690#B-diseases, accessed 13/4/2023). Other structures such as the retinal vasculature can also be affected in IRDs, for example familial exudative vitreoretinopathy (FEVR, OMIM: #133780) which results in abnormal retinal angiogenesis with phenotypic heterogeneity (Panagiotou et al., 2017).

Similarly, IRDs unevenly affect different cell types. Mutations in 28 genes result in cone-rod dystrophy (CRD) (https://web.sph.uth.edu/RetNet/sum-dis.htm?csrt=17870442760574951690#B-diseases, accessed 13/4/2023), which generally results in initial death of cone cells which may progress to damage to the peripheral retina (Huang et al., 2016). In contrast, optic atrophy (OMIM: #165500) is caused by mutations in 13 genes (https://web.sph.uth.edu/RetNet/sum-dis.htm?csrt=17870442760574951690#B-diseases, accessed 13/4/2023). These genes have a range of functions, for example mitochondrial function, and pathogenic mutations result in damage to the retinal ganglion cells (Yu-Wai-Man et al., 2009).

Many genes which when mutated cause IRD are expressed in the retina and supportive tissues, although mutations in genes with widespread expression and functions have been found. Disruption in these processes result in a spectrum of disorders which can affect the retina in isolation, or can be one part of a multi-system disorder (Werdich et al., 2014). Mutations in genes related to cilia development and function are a well-known cause of syndromic conditions, for example mutations in *BBS1* can cause Bardet Biedl syndrome (OMIM: #209900) which displays a range of symptoms such as obesity, intellectual disability and polydactyly in addition to RP (Scheidecker et al., 2014). Mutations in a number of genes have been associated with both syndromic and non-syndromic conditions (Leitch et al., 2008; Molin et al., 2013; Werdich et al., 2014; Chiang et al., 2015). For example mutations in *USH2A* have been associated with both syndromic (Usher Syndrome type 2A, #276901) and non-syndromic IRDs (Retinitis Pigmentosa-39, #613809) (Molina-Ramírez et al., 2020).

1.3.1 Stargardt disease

Given its importance in maintaining the health of the retina, it is no surprise that disruption of the visual cycle is a frequent cause of IRDs. Reflecting this, mutations in *ABCA4* are the single most common cause of IRDs, causing up to 30% of all IRD cases (Hanany et al., 2020). Pathogenic mutations in *ABCA4* cause MD with an autosomal recessive inheritance pattern, resulting in central vision loss. The frequency of ABCA4-related disease is difficult to estimate and differs between populations (Cremers et al., 2020). Descriptions of MD similar to *ABCA4*-related disease were made in the 19th century with definitive

clinical description of what became known as Stargardt disease at the beginning of the 20th century (Lang, 1885; Stargardt, 1909).

Classical Stargardt disease (STGD) typically presents in young people with diagnosis typical in adolescence or young adulthood. The age of onset can be difficult to determine exactly as many individuals are initially unaware of the changes in their vision and damage to the macula may precede vision loss (Cremers et al., 2020). This is particularly true in cases with initial sparing of the fovea (Nakao et al., 2012; Fujinami et al., 2013; van Huet et al., 2014). The initial symptoms observed by individuals with STGD are central vision loss, progressing to reduced dark adaption in a subset of patients (Fishman et al., 1999; Derwent et al., 2004; Salvatore et al., 2014). Individuals may also show impaired colour discrimination and photophobia (Anderson et al., 2003). Mutations in *ABCA4* result in a spectrum of conditions, including early onset, severe IRDs such as STGD, and cone-rod dystrophy to relatively late onset, mild disease (Cremers et al., 2020). The onset of symptoms is correlated with residual function of ABCA4 protein with severe mutations resulting in early onset cone-rod dystrophy and milder mutations resulting in late onset disease (Lambertus et al., 2016; Runhart et al., 2019). Further, variable penetrance has been observed with some mutations being identified as resulting in disease with reduced penetrance (Zernant et al., 2017; Runhart et al., 2018; Allikmets et al., 2018).

1.3.1.1   Clinical features of *ABCA4* disease

Cases with *ABCA4*-related disease typically show macular degeneration presenting as progressive damage to the central macula. While the rate and extent of this can vary, this typically results in a distinctive 'beaten copper' appearance upon fundus imaging (Figure 1.8). Typically, although this can vary greatly, the RPE is initially affected leading to choriocapillaris involvement (Duncker et al., 2014; H. Song et al., 2015; Tanaka et al., 2018). End stage disease can lead to deterioration of the choroid, resulting in visible sclera (Lee et al., 2018). Additionally, cases with *ABCA4*-related disease also present with characteristic lipofuscin accumulation which can appear as flecks when viewed, particularly under short wave autofluorescence. These flecks can vary in extend and morphology, however their presence is highly indicative of *ABCA4*-related disease (Cukras et al., 2012; Chen et al., 2019). Frequently the tissue surrounding the optic nerve will be spared from damage,

particularly in early stages of disease (Cideciyan et al., 2005). Taken together, presentation of these clinical manifestations allows for relatively confident diagnosis of *ABCA4*-related disease from clinical signs alone.
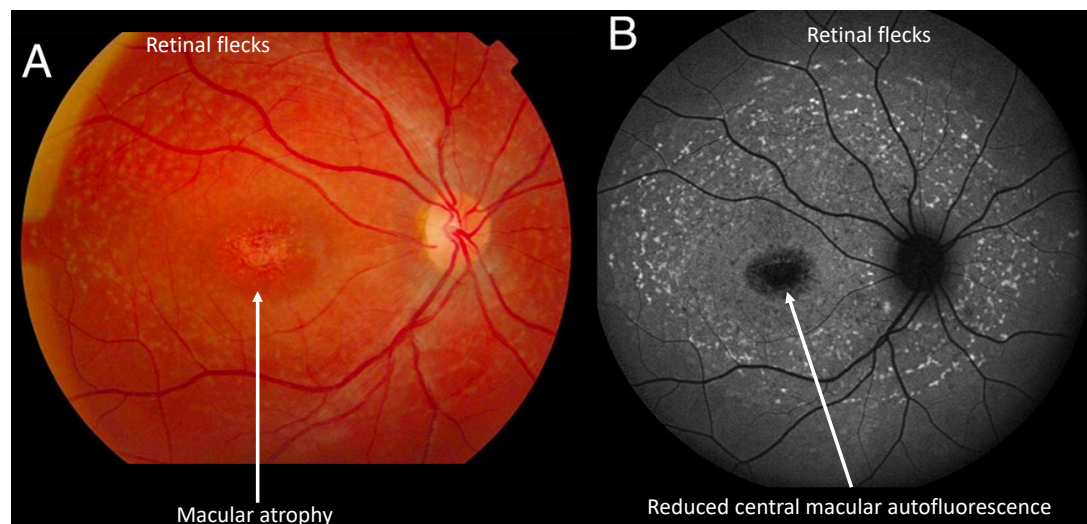
Figure 1.8 Fundus image of retina affected by STGD

A) Colour fundus image of STGD retina with visible macular atrophy and peripheral yellow/white retinal flecks. **B)** Autofluorescence fundus image of same retina with reduced central macular signal, and retinal flecks clearly visible. Adapted from (Tanna et al., 2017), under CC BY 4.0.

1.3.1.2 Function of ABCA4

ABCA4 has a crucial role in the visual cycle by transporting sequestered all-*trans*-retinal so it is available to RDH8. Pathogenic mutations in the *ABCA4* gene result in pathological dysfunction of this process. ATP-binding cassette transporters (ABC) are one of the oldest and most diverse protein super-families (Higgins, 1992). Broadly, ABC transporters use ATP to transport substrates across biological membranes. In mammals, there are seven ABC families, the ABCA group comprise the largest group of ABC transporters, with mutations in several resulting in disease in humans (Albrecht and Viturro, 2007).

ABCA4 is 2273 amino acids, however the protein structure is not well characterised (Tsybovsky and Palczewski, 2014). As with all ABC transporters, ABCA4 consists of two nucleotide binding domains and two transmembrane domains arranged into homologous halves (Tsybovsky and Palczewski, 2014). Much effort has been made to infer the properties of

ABCA4 by indirect predictions, and characterisation of subdomains and recombinant analogues (Ahn et al., 2000; Bungert et al., 2001; Ahn et al., 2003; Biswas-Fiss et al., 2012). More recently, Cryo-EM studies have been used to characterise the structure of active and inactive ABCA4, although the exact mechanism of ABCA4 activity is still not well-characterised (Liu et al., 2021; Scortecci et al., 2021).

The exact mechanism of ABCA4-related disease is still a subject of debate. One model which has been proposed is that dysfunction of ABCA4 reduces the efficiency of N-ret-PE clearance, causing it to form phosphatidylpyridinium bisretinoid (A2PE) (Lenis et al., 2018). When the photoreceptor outer segment is phagocytosed by the RPE, A2PE then forms N-retinylidene-N-retinylethanolamine (A2E), a major component of the lipofuscin which is characteristic of ABCA4 disease (Young and Bok, 1969; Mata et al., 2000). This then accumulates in the RPE, eventually leading to increased stress on the RPE, leading to photoreceptor death and retinal thinning. Reduced choroid vasculature at the macula has been theorized to increase susceptibility to biproduct build up and resulting damage, perhaps partly explaining why the macula is primarily affected in STGD (Provis et al., 2005).

Alternative mechanisms of ABCA4 disease have also been proposed. *ABCA4* has also been found to be expressed in the RPE at reduced abundance (Lenis et al., 2018). This has been proposed to indicate an additional mechanism where ABCA4 has an additional function in the RPE endolysosome (Farnoodian et al., 2022). It has been proposed that ABCA4 aids processing of phagocytosed outer segments. As in the photoreceptor, ABCA4 in the RPE aids N-ret-PE processing and prevents lipofuscin formation. While ABCA4 is present at a reduced prevalence in the RPE compared to the photoreceptor, ABCA4 function in the RPE has been suggested to be at least as important as in the photoreceptor (Lenis et al., 2018). Other mechanisms of *ABCA4*-disease have been proposed, such as overactivation of the complement cascade (Hu et al., 2020; Jabri et al., 2020). Further research is required to uncover the contribution of, and interplay between, different mechanisms of *ABCA4*-associated disease.

1.3.1.3 Genetics of ABCA4-related disease

ABCA4 protein was initially characterised in 1976 (Papermaster et al., 1976) and the *ABCA4* locus was subsequently mapped to 1p13 (Kaplan et al., 1993; Gerber et al., 1995; Anderson et al., 1995). The *ABCA4* gene was then cloned in 1997 (Allikmets et al., 1997) and the first pathogenic mutations in *ABCA4* were identified and linked to STGD (Allikmets et al., 1997).

As the single most common cause of IRDs, the genetics of *ABCA4*-related disease has been the subject of intense study. This in-depth study of *ABCA4* has revealed an extensive spectrum of disease-causing variants. Currently there are >2000 known pathogenic, and likely pathogenic, variants in *ABCA4* (www.lovd.nl/ABCA4, accessed 9/1/2023). As more individuals are screened, in particular from non-European cohorts, this is likely to expand greatly.

Of interest, it can be difficult to predict the consequences of mutations in *ABCA4*. For example, many reported missense alleles are deleterious while variant categories that are commonly deleterious such as stop gains and indels can result in a partial retention of function (Cremers et al., 2020). Because of this, significant effort has been made to comprehensively categorise *ABCA4* variants. Most disease causing variants are missense (50% of unique alleles), while 33% are protein truncating, 2% are deep intronic, 3% are structural and 5% are noncanonical splice site variants. The remaining 7% of unique alleles are complex alleles composed of more than one missense variant in *cis* (Cornelis et al., 2017; Cremers et al., 2020; Cornelis et al., 2022). Approximately 10% of *ABCA4* alleles are comprised of more than one variant in *cis*, including other variant classes such as deep-intronic variants and protein truncating variants (Cremers et al., 2020). Extremely hypomorphic 'mild' alleles have been identified in *ABCA4* with distinctive phenotypes. For example, the *ABCA4*:c.5882G>A, p.(Gly1961Glu) (NM_000350.3, hg19) variant has been associated with a slightly later age of onset and milder disease course than classical STGD (Zernant et al., 2017). Of interest, approximately 7% of total *ABCA4* alleles are composed of more than one variant, forming a single complex allele with an altered affect (Shroyer et al., 2001). These have been associated with the previously described 'mild' alleles, accounting for a significant proportion of the disease burden (Zernant et al., 2017). The most prevalent of these is c.[2588G>C;5603A>T], p.[Gly863Ala,Gly863del;Asn1868Ile]. The

c.2588G>C variant has been shown by midi-gene assays to result in Gly863Ala in approximately 50% of transcripts and Gly863del in the other 50% and as such is usually referenced with both protein affects (Sangermano et al., 2018). This complex allele was discovered after the significance of c.5603A>T was discovered. It was found that c.5603A>T was capable of causing disease only when in *trans* with a severe or moderately severe allele, and then at a low penetrance (Runhart et al., 2018). It had been previously observed that this variant was always found in *cis* with c.2588G>C in *ABCA4*-related disease (Maugeri et al., 1999). After the significance of c.5603A>T in isolation was observed, it was then observed that in cases with both variants in a complex allele, the complex allele resulted in a single, fully penetrant allele (Runhart et al., 2018). Similarly, c.[1622T>C;3113C>T], p.[Leu541Pro;Ala1038Val] constitutes up to 34% of complex alleles. In this case, c.1622T>C is considered to be a moderately severe variant in isolation and c.3113C>T is considered to be mild (Cornelis et al., 2017). When presenting as a single complex allele it is a deleterious allele which causes complete loss of function (Zernant et al., 2017).

As *ABCA4*-related disease is relatively clinically distinctive compared to other IRDs, it is possible to segregate cohorts of cases with *ABCA4* disease with relatively high confidence. This has allowed the development of a phenotype-genotype model where the severity of observed mutations is correlated with age of onset, rate of progression and prognosis. In particular, a number of genotypes have been associated with distinctive outcomes. For example, c.5603A>T in *trans* with a deleterious allele has been associated with very late onset, relatively mild disease with the onset of symptoms in the 4th decade and disease rarely progressing beyond the extramacular stage (Zernant et al., 2017). In contrast, two deleterious alleles in *trans* is associated with severe cone-rod dystrophy with onset of disease before the age of ten (although this may not be noticed until later) and progression to an advanced stage of disease by the 3rd decade (Cornelis et al., 2022).

1.3.1.4   Differential diagnosis of *ABCA4*-related disease

Despite its distinctive manifestation, numerous phenocopies of *ABCA4*-related disease are well known. Previously phenocopies were named as STGD2, 3 and 4. These have been referred to as, 'Stargardt-like' and 'Dominant Stargardt'. Mutations in a number of genes result in phenotypes

which share the clinical features listed above and include *PROM1* and *PRPH2* (Kniazeva et al., 1999; Cideciyan et al., 2005). A common way of distinguishing phenocopies of STGD is through careful examination of the associated inheritance pattern. A common thread among phenocopies of STGD is that while the dominant form of the disease is a phenocopy of STGD, the recessive form features additional clinical features which distinguish it. For example, dominant mutations in *PROM1* can result in a phenotype similar to disease caused by mutations in *ABCA4*, however the recessive form of the disease more closely resembles RP (Kniazeva et al., 1999; Wolock et al., 2019b). This can also be difficult to identify as *ABCA4* is commonly associated with pseudo-dominant inheritance (Cornelis et al., 2022).

In addition to phenocopies, maculopathies with a similar presentation to *ABCA4*-related disease can also be caused by environmental factors such as drug toxicity (in particular hydroxychloroquine) (Nõupuu et al., 2016). As such, it can be difficult to distinguish *ABCA4*-related disease from other causes of maculopathies, although careful clinical assessment can correctly identify *ABCA4*-related disease in approximately 90% of cases (Cremers et al., 2020).

1.4   Genetic characterisation of IRDs

While it is the most common gene associated with IRD, *ABCA4* is only one of over 300 genes which have been associated with IRD, presenting a challenge to genetic diagnosis. As discussed, IRDs display variable and overlapping clinical features and genetic characterisation is often required for an unambiguous diagnosis. An accurate diagnosis is crucial for patients, for informing family planning, indicating likely disease progression and for enrolment in a growing number of gene-specific gene therapy trials (Britten-Jones et al., 2022).

The genetic diagnosis of IRDs has been an area of transformative change in recent years (Chiang et al., 2015). Prior to the widespread use of next generation sequencing (NGS), IRDs were genetically diagnosed by Sanger sequencing of target genes. Sanger sequencing utilises sequencing by synthesis using radioactive-tagged chain terminating dideoxynucleotides (ddNTPs), with later iterations using fluorescently tagged ddNTPs (Sanger et al., 1977; Baudhuin et al., 2015). While highly accurate and still commonly

performed as a gold standard confirmation of results, Sanger sequencing is performed on a single target at a time, limiting its throughput. In contrast, the adoption of NGS sequencing platforms enables massively parallelised sequencing, greatly increasing throughput.

A number of NGS platforms have been developed, beginning with Roche's 454 platform in 2005 (Margulies et al., 2005). Illumina-based NGS has come to dominate this market (Canard and Sarfati, 1994; Smith et al., 2008; Bentley et al., 2008). During Illumina-based NGS, platform specific adaptors are ligated to the DNA to be sequenced (Figure 1.9). This is then introduced to the flowcell where the DNA fragment is amplified in place, forming clusters of each DNA fragment. These clusters are then sequenced by synthesis using fluorescently tagged dNTPs (Smith et al., 2008). The adoption of NGS has drastically reduced sequencing costs and increased output, from the Human Genome Project which cost approximately $3 billion (Lander et al., 2001) to, soon to be, commercially available $200 genomes (https://www.illumina.com/systems/sequencing-platforms/novaseq-x-plus.html, https://www.elementbiosciences.com/200-dollar-genome accessed 21/2/2023). NGS produces short reads of approximately 150-500 bases which are then mapped to a reference genome, producing a 'haploid' alignment with differences to the reference highlighted (Behjati and Tarpey, 2013).

A key development in the widespread use of NGS was the development of targeting methods. These function by enriching target sequences, typically genes associated with the disease of interest or the entire exome (known as whole exome sequencing or WES) (Ng et al., 2009; Choi et al., 2009). A number of target capture arrays are available which function by a variety of methods (Clark et al., 2011).
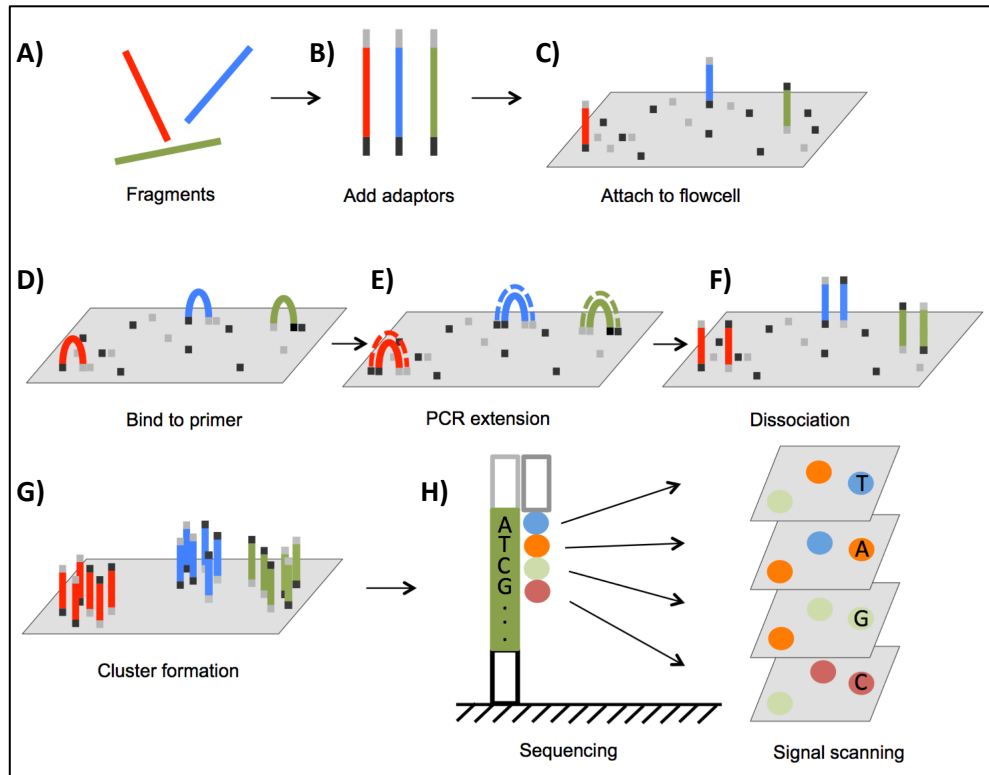
Figure 1.9 Illumina based NGS

A) The target DNA is fragmented to 150-500bp. **B)** Illumina sequencing adaptors are ligated to the exposed ends. **C)** The library is introduced to the flowcell where the adaptors bind to the floor of the cell. **D-F)** The target DNA is amplified by bridge amplification. The exposed adaptor binds to primers on the flowcell floor, the target sequence is amplified and then one adaptor disassociates. **G)** This results in clusters of amplified fragments originating from the same molecule. **H)** The clusters are then sequenced by the addition of fluorescently tagged nucleotides. The clusters allow for much stronger signal perception.

1.4.1 Next generation sequencing of IRDs

Initial efforts to genetically characterise IRDs using NGS were performed using targeted gene panels (Mamanova et al., 2010; Audo et al., 2012; Neveling et al., 2012), an approach which is still very commonly used (Ellingford, Barton, Bhaskar, O'Sullivan, et al., 2016). Targeted sequencing is used because it reduces costs compared to more widespread sequencing while still solving a majority of cases. It is commonly asserted that targeted NGS genetically solves approximately two-thirds of IRD cases, however this varies between 40-80% depending on a variety of factors including which IRD phenotypes are studied, the location of the study and the technology used (Britten-Jones et al., 2023).

While targeted sequencing typically results in high solve rates, the majority (71.2%) of these solved cases are due to mutations in approximately 20 genes (Pontikos et al., 2020). This supports using targeted sequencing as an initial approach, but limits the study of less well characterised genes. Reflecting this, 200 genes and loci were associated with IRDs in 2010 (Britten-Jones et al., 2023), compared to over 300 now (https://web.sph.uth.edu/RetNet/sum-dis.htm#D-graph, accessed 21/2/2023). Many of these cause extremely rare disease and are not commonly including in gene panels (Hanany et al., 2020).

1.4.1.1 Targeting approaches

A broad variety of targeted sequencing methods have been developed to allow for efficient target capture, many of which are either amplicon based or hybridization based (Hung et al., 2018; Bewicke-Copley et al., 2019). Amplification based methods typically use parallel PCR targeting regions of interest with modified PCR primers (Mertes et al., 2011; Coppieters et al., 2012), for example HaloPlex panels (Aparisi et al., 2014; Consugar et al., 2015; Perez-Carro et al., 2016). In contrast, hybridization techniques use tagged probes to capture target fragments. A common hybridization method is using biotinylated probes which are isolated using streptavidin coated magnetic beads (Samorodnitsky et al., 2015). These target panels can either be commercially available panels targeting common targets, or custom panels designed to capture a target of interest (Watson et al., 2014; Taylor et al., 2017).

1.4.1.2 smMIPs based target capture

A target capture approach which has been highly successful is the use of single molecule molecular inversion probes (smMIPs). Each smMIP is a single stranded DNA molecule composed of a 30 nucleotide 'linker' backbone, 16-24 nt extension and ligation probe arms, a 5 nt patient index, and two random 5 nt sequences which function as unique molecule identifiers (Figure 1.10).

The extension and ligation probes hybridize at target sequences flanking the 225 bp region of interest. The target sequence is captured by amplification, using the hybridized extension arm as a primer (Hiatt et al., 2013). The amplified target DNA is then ligated to the ligation arm, resulting in a circularised probe. This is typically done for many smMIPs in a single reaction

as the smMIPs should not interact during amplification. Linear DNA, including genomic DNA and smMIPs which failed to hybridize are then removed by an exonuclease. The smMIP with the captured target is then amplified using universal primers which hybridize to the smMIP backbone, amplification of the smMIP results in a linearised amplification product containing the target sequences (Hiatt et al., 2013). The universal primers used to amplify and linearise the probes contain unique patient tags, indexing the smMIP with a patient identifier. The linearized probes for each patient are then quantified and pooled. Multiple patient libraries are typically sequenced as part of a single sequencing run (Khan et al., 2020; Panneman et al., 2022).

smMIPs are the latest development in a series of molecular inversion probes. The first iteration of these were padlock probes used for tagging target sequences (Nilsson et al., 1994). This was then further developed, resulting in Molecular Inversion Probes (MIPs). Initial iterations of MIPs only captured a single base and were used for SNP analysis (Hardenbol et al., 2003). Later versions were capable of capturing 20 bases (Akhras et al., 2007), later expanding to over 100 bases (Weisschuh et al., 2018). MIPs were shown to be capable of multi-exon targeting (Porreca et al., 2007; Turner et al., 2009; Shen et al., 2011; O'Roak et al., 2012) and subsequently used to screen multiple conditions such as stillbirth, Parkinson's disease, various cancers and IRD cohorts (Jahromi et al., 2012; Rowe et al., 2013; Weisschuh et al., 2018; Qin et al., 2019).

MIPs screening facilitates cost effective capture of large loci, and the addition of single molecule tagging (resulting in smMIPs) has greatly increased accuracy. As MIPs capture is amplification based, any artefacts introduced during amplification of target sequences have the potential to confound variant detection, potentially resulting in false positive results. This can be greatly mitigated by removing PCR duplicates and creating a single consensus sequence of each probe. By including a single molecule tag, known as a Unique Molecule Identifier (UMI) in the MIP, it is possible to remove PCR duplicates and create a more accurate consensus sequence of each molecule of origin (Kivioja et al., 2011; Fu et al., 2011; Kinde et al., 2011; Casbon et al., 2011; Jabara et al., 2011; Shiroguchi et al., 2012). This has been shown to greatly reduce PCR artefacts and result in consensus sequences with a per base error rate as low as 2.6×10^{-5} (Hiatt et al., 2013).
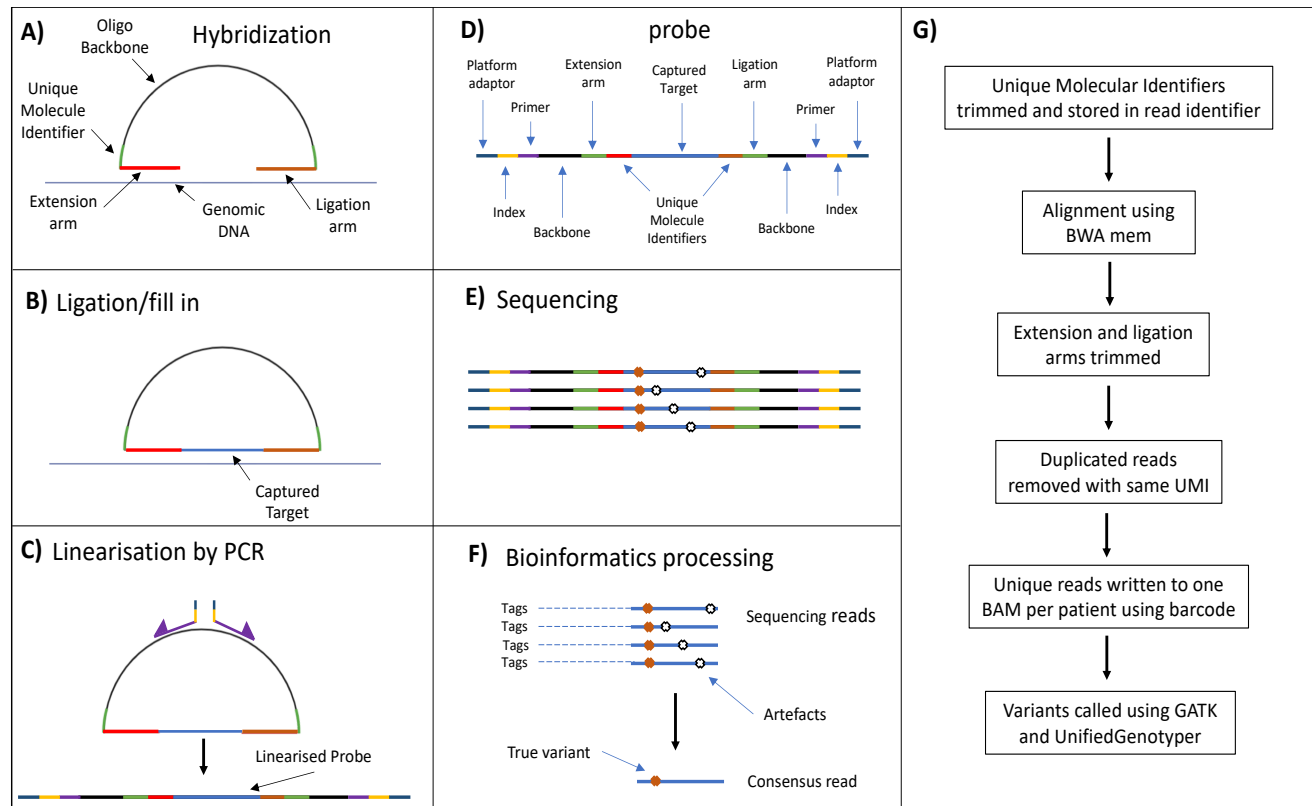
Figure 1.10 smMIPs sequencing workflow

A) Each smMIP consists of an oligo backbone (30 nt), unique molecular identifiers (5nt), an extension arm (16-20nt) and a ligation arm (20-24nt). This captures a 225 nt target. **B)** The target is then captured by polymerase extension of the extension arm using the target sequence as a template followed by ligation to the ligation arm. **C)** The resulting circularised probe is then amplified by PCR using universal primers containing a patient index. The resulting product is linear. **D)** The resulting linearized probe contains platform specific adaptors, patient indexes, smMIPs backbone sequence, the extension and ligation arm, two unique molecular identifiers (UMIs) and the captured target. **E)** This happens in parallel for many smMIPs per patient. Each resultant patient library is then quantitated, pooled and sequenced. **F)** The resulting reads are then processed using the tags to obtain a single consensus sequence per smMIP which is aligned. The resulting reads are separated by patient tag. **G)** Bioinformatic processing of the resulting reads.

1.4.1.3 Untargeted screening methods

If targeted panel sequencing is unsuccessful, a more widespread technique such as WES may be used. WES has been shown to improve solve rates compared to targeted sequencing but has higher associated costs (Britten-Jones et al., 2023). Similarly, whole genome sequencing (WGS) is also commonly used as a second line test and has been shown to improve solve rates compared to targeted sequencing methods but has even higher costs (Lionel et al., 2018; Schwarze et al., 2018).

WES and WGS have the significant advantage of being bias free compared to targeted sequencing panels. By sequencing the entire exome or genome, mutations in genes which are rarely implicated in IRD can be detected along with mutations in genes which have not previously been implicated. While the cost of WGS in particular has fallen in recent years, the associated data storage and analysis costs remain significant (Dockery et al., 2021).

1.4.1.4 Limitations of NGS based sequencing

Inherent limitations in the underlying sequencing technology continue to limit NGS screen results. The reads resulting from NGS are typically short in nature (150-500 bases), leading to regions which are generally intractable to NGS. This results in 'NGS dead zones' which typically have reduced coverage when sequenced (Mandelker et al., 2016). For example, repetitive regions cannot be unambiguously sequenced by NGS unless the read is larger than the repeat (Treangen and Salzberg, 2012). This leads to particular challenges for characterisation of repetitive regions which have been associated with IRDs. Up to 60% of pathogenic mutations in *RPGR* are harboured in *ORF15*, but as it is very repetitive and purine rich it is very difficult to sequence with NGS or Sanger sequencing and difficult to align to the resulting sequence against a reference (Ruddle et al., 2009).

Further, while NGS has been successfully used to detect and characterise copy number variants (CNVs) and structural variants affecting IRDs, this has been acknowledged as an area of missing heritability in many IRDs (Bernardis et al., 2016). Using NGS to characterise CNVs leads to particular challenges, likely leading to a significant number of missed variants during sequencing (Zampaglione et al., 2020).

The short nature of NGS reads also makes it difficult to establish the phase of distal variants (i.e. assigning variants as originating from the maternal or paternal chromosome). As the resulting reads are short, it is typically not possible to map reads to the homologous chromosome of origin and as such reads are typically represented as a haploid alignment (Levy et al., 2007; Ballouz et al., 2019). Phasing of clinically relevant variants is typically done by segregation analysis using direct family members. If this is not possible, it is generally very difficult to segregate variants into haplotypes. This is of importance in the study of IRDs for a number of reasons. As many IRDs are inherited in a recessive pattern, compound heterozygotes are very commonly observed (Hanany et al., 2020). If there are no available family members to segregate variants of interest, it can be difficult to confidently assign haplotypes to variants of interest and compound heterozygote variants are generally assumed to be in *trans*. Complex alleles are particularly common in *ABCA4* and can be difficult to characterise without phasing by segregation.

Complex alleles have been implicated in other genes in IRD and beyond. *IMPG2*: c.[3023-15T>A;3023G>A], p.[Gly1008Asp;?] (NM_016247.4, hg19) has been demonstrated to function as a complex allele resulting in adult onset vitelliform MD. The c.3023G>A variant was demonstrated to enhance the splicing effect of c.3023-15T>A, resulting in a single complex allele with an increased affect (Vázquez-Domínguez et al., 2022). Complex alleles are often implicated in cases with variable phenotypes resulting from the same pathogenic variant. For example, the *CFTR*: (c.2991G>C, p.Leu997Phe) (NM_000492.4, hg19) variant is a highly debated cause of cystic fibrosis associated with high clinical heterogeneity. In-depth examination of a series of cases revealed the presence of CFTR: (c.350G>T, p.Arg117Leu) in *cis* with p.Leu997Phe, which was associated with more severe disease than cases with p.Leu997Phe in isolation (Lucarelli et al., 2010). From this it is clear that complex alleles may be responsible for much of the unexplained variation in Mendelian diseases.

In response to these NGS specific issues, a number of sequencing platforms have been developed which produce significantly longer reads. These can either be synthetic long-read sequencers or single molecule long-read sequencers. Single molecule sequencers have generally dominated this

emerging field, in particular the Pacific Biosciences range of single molecule real time (SMRT) sequencers (hereafter referred to as PacBio sequencing or SMRT sequencing) and the Oxford Nanopore Technologies range of sequencers (hereafter referred to as nanopore sequencing).

Long-read sequencing has been known for reduced accuracy compared to NGS, with sequencing accuracy of 75-90% (Wang et al., 2021). This has largely limited their use for applications requiring very high accuracy like single nucleotide variant detection. In the last few years, significant increases in unpolished per base accuracy have seen greatly increased use of long-read sequencing.

1.4.2 PacBio sequencing

Pacific Biosciences released the first iteration of SMRT sequencing in 2011 and since that time, the technology has undergone multiple changes which have increased accuracy, read length and decreased cost (Eid et al., 2009; Travers et al., 2010). Like Illumina based NGS, PacBio sequencing also utilises 'sequencing by synthesis' during which bases are read by addition of fluorescently labelled nucleotides. In contrast to Illumina sequencing, PacBio sequencing uses a circular library in conjunction with a long-range polymerase. As such, there is no pause in the read process and long-read rates are achievable (Eid et al., 2009; Travers et al., 2010).

During library preparation, digested genomic DNA or enriched target DNA is circularised by ligating hairpin adaptors to the ends of the DNA molecule. This creates a closed, single stranded circular SMRTbell template. The SMRTbell is then loaded on to a flowcell consisting of many nanoscale wells known as zero mode waveguides (ZMWs). At the bottom of each ZMW there is a high fidelity polymerase which sequences the SMRTbell by synthesis using fluorescent dNTPs. As each base is incorporated, it produces a distinctive fluorescence which is captured and stored as a 'movie' which can be base called (Eid et al., 2009; Travers et al., 2010). As the ZMW is nanoscale, outside light is not captured thereby reducing noise (Eid et al., 2009; Travers et al., 2010). In addition, by also measuring the kinetics of each base addition, base modifications can be detected (Koren and Phillippy, 2015).

A key aspect of PacBio sequencing is that depending on the size of the target molecule, the molecule can either be sequenced in a single pass producing a single subread, or the circularised molecule can be sequenced multiple times to create a highly accurate consensus circular sequence, the most accurate of which commonly referred to as High Fidelity (HiFi) reads (Travers et al., 2010; Loomis et al., 2013). The read lengths achieved through PacBio sequencing have increased from mean 1500 bp by the first commercially available iteration of PacBio sequencing (Brown et al., 2014) to mean HiFi reads of up to 25 kb (Wenger et al., 2019).

PacBio sequencing has a reputation for increased accuracy compared to other long-read sequencing methods, in particular nanopore sequencing. Specifically, HiFi reads are known to be highly accurate with read accuracies exceeding 99.5% (Hon et al., 2020). Pacific biosciences have claimed to achieve >Q50 consensus accuracy with HiFi sequencing allowing for accuracy greater than 99.999% (https://www.pacb.com/technology/hifi-sequencing/how-it-works/ accessed 14/2/2023). Q refers to the Phred read quality score calculated by $Q = -10 \times \log10(P)$ where P is the measured error rate. Of particular interest, PacBio sequencing is largely free of the systemic errors known to effect nanopore sequencing. While this is far more accurate than nanopore sequencing, these super accurate reads require HiFi sequencing, limiting insert size during library preparation.

1.4.3 Nanopore sequencing

Nanopore sequencing is performed using a nanoscale pore referred to as a 'nanopore'. The Oxford Nanopore range of sequencers use a nanopore which is an engineered membrane protein that is inserted into a polymer membrane (Figure 1.11). In the flowcell, an electrical current is applied, producing a negatively charged side of the membrane (the *cis* side) and a positively charged *trans* side. DNA and RNA molecules are driven through the pore from the *cis* to the *trans* side, disrupting the electrical current in the 'sensing' region of the pore (Wang et al., 2021). Bases passing through the sensing region result in a distinctive disruption to the current which can then be base called.

Nanopore sequencing was first developed in the 1980s where engineered α-haemolysin from *Staphylococcus aureus* was used to distinguish base sequences (Deamer et al., 2016). Further iterations of the pore protein

allowed for more accurate distinguishing of single bases (Stoddart et al., 2009; Stoddart et al., 2010). The addition of processive enzymes, known as motor proteins, greatly increased accuracy. In particular the addition of phi29 DNA polymerase increased accuracy by regulating translocation and unwinding the DNA so that single stranded DNA passed through the pore (Cherf et al., 2012). Following this, Oxford Nanopore Technologies released the first commercially available nanopore sequencer in 2015 (Jain et al., 2016). Oxford Nanopore Technologies produces a range of flowcells with between 120 and 2,675 pores (https://nanoporetech.com/products/specifications, accessed 21/2/2023). Nanopore sequencing can be used to sequence DNA and directly sequence RNA, as well as detecting base modifications (Seki et al., 2019).
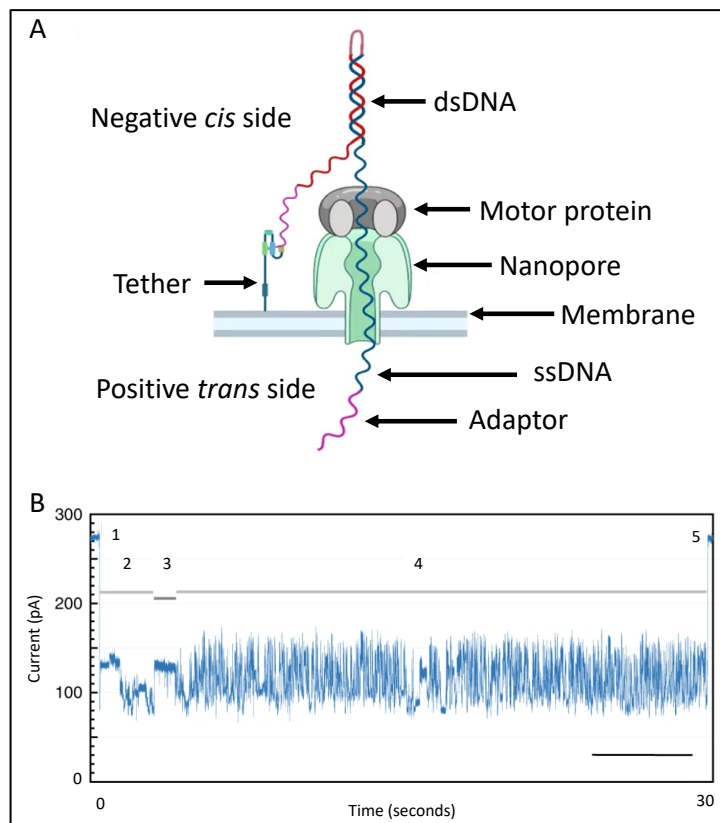
Figure 1.11 Schematic of nanopore

A) Schematic of nanopore sequencing with major features labelled. Double stranded DNA is ratcheted through the nanopore (mutant CsgG from *E. Coli*) by a motor protein. This disrupts the electrical current in the flowcell. Adapted from (Gao et al., 2023) under CC BY 4.0. **B)** The disruption to the electrical current results in 'the squiggle' which can be basecalled. Shown is an ionic current trace of RNA sequencing although the same principles apply to DNA sequencing. The parts of the squiggle have been labelled: 1) strand capture, 2) adaptor sequence, 3) polyA RNA tail, 4) mRNA sequence, 5) the strand exiting the pore. Adapted with permission from (Workman et al., 2019).

A number of iterations of the nanopore have been used since its initial release, with iterative increases in accuracy and translocation speed. The R7 pore had a per base accuracy rate of 64% and a translocation speed of 70 bases a second, in comparison the R9.4 pore (the pore used for all the experiments in this thesis) has a per base accuracy of 85-94% and a translocation speed of up to 450 bases per second (Ashton et al., 2015; Minei et al., 2018; Wick et al., 2019). The pore proteins and motor proteins used in nanopore sequencers are not disclosed in detail, the R9.4 pore uses a mutant Curlin sigma S-dependent growth subunit G (CsgG) from *Escherichia coli* and an undisclosed motor protein (Wang et al., 2021).

The major advantage of nanopore sequencing compared to other methods is that ultra-long-reads are achievable. In contrast to NGS and PacBio sequencing, there is no direct limit on read length during nanopore sequencing. The limiting factor for read length during nanopore sequencing is generally the preservation of high molecular weight DNA. With improvements to DNA extraction techniques, reads of up to 4.4Mb have been reported ([https://nanoporetech.com/products/minion)](https://nanoporetech.com/products/minion), accessed 8/2/2022).

While nanopore sequencing generally has reduced read accuracy compared to PacBio sequencing, this has been improved greatly in recent years. In addition to improved pore proteins, improvements to nanopore base callers and variant callers have increased accuracy (Helal et al., 2022). Although this will vary by variant caller, benchmarking using PEPPER-Margin-DeepVariant has shown that to achieve high quality variant calling (defined as 99.5% precision), 40-50 × coverage is required for single nucleotide variant calling using nanopore data compared to 35 × for PacBio HiFi data (Shafin et al., 2021).

Recently, Oxford Nanopore Technologies have released the R10.4.1 pore with updated sequencing kits and base calling workflows. These have been reported to achieve greater than Q20 read accuracy (>99%). In addition, nanopore sequencing is generally simplex, where each DNA molecule is read once. These kits and pores are capable of duplex sequencing, where each strand of a double stranded molecule is sequenced in succession. This has been reported by Oxford Nanopore Technologies to be capable of producing

Q30 reads of 260 kb in length (https://nanoporetech.com/accuracy, accessed 21/2/2023).

The development of long-read sequencers promises to allow the study of missing heritability in *ABCA4* and beyond by allowing robust sequencing of areas intractable to NGS, phasing of variants of interest, detection of CNVs and more.

1.5 Aims

The aim of this research was to further study the genetic causes of *ABCA4*-related disease and to establish methods to search for missing heritability. A method to robustly phase distal variants in *ABCA4* using nanopore long-read sequencing was developed. This was optimized and applied to a series of cases with biallelic *ABCA4* mutations. An assay to comprehensively characterise large deletions detected by targeted NGS was also developed using nanopore sequencing. This was used on a series of cases with large pathogenic deletions in IRD genes. Finally, a cohort of 57 cases were screened using an smMIPs based panel targeting genes associated with macular dystrophy.

2 Materials and methods

2.1 General buffers and solutions

- Tris- Ethylenediaminetetraacetic acid (EDTA) (TE) buffer
    - 10 mM Tris-HCl pH 7.5
    - 1 mM EDTA
- 50 × Tris-Acetate-EDTA (TAE)
    - 2 M Tris-HCl
    - 50 mM EDTA (pH 8.0)
    - 0.97 M Glacial acetic acid
- Nuclei Preparation Buffer (NPB)
    - 25 mM Tris-HCl, pH 8.0
    - 5 mM EDTA
    - 0.5% Triton X100
    - 500 mM sucrose
    - 4 mM spermidine (3HCl)
- Luria-Bertani (LB) Broth
    - 1% tryptone
    - 1% NaCl
    - 0.5% yeast extract
- Phosphate-buffered saline (PBS)
    - 137 mM NaCl
    - 2.7 mM KCl
    - 10 mM Na_2HPO_4
    - 1.8 mM KH_2PO_4

Where constituents allowed, solutions were autoclaved prior to use at 115 ºC, 0.6 kg/cm^{-3} for 30 minutes.

2.2 Patients

All patients were diagnosed and recruited by Ophthalmologists at St James's University Hospital. Blood samples were collected from patients and family members after obtaining informed consent. Ethical approval was provided by the Leeds East Research Ethics Committee (Project numbers 03/362 and

17/YH/003). For all experiments performed in Leeds, patient genomic DNA was extracted using standard protocols by the North East and Yorkshire Genomics Laboratory Hub, Central Laboratory, Leeds.

2.3 Polymerase Chain Reaction (PCR)

PCR primers were designed using Primer3 (http://primer3.ut.ee/) and validated using the UCSC *in silico* PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr). All primers were prepared using desalted purification and were purchased from Sigma-Aldrich. Primers were designed to be approximately 20 nt in length with a GC content of approximately 50%.

2.3.1 Standard PCR

Typically, PCR reactions were performed in a total reaction volume of 25 µl and comprised; 1 µl of 10 pmol/µl forward primer, 1 µl of 10 pmol/µl reverse primer, 1 µl of a 200µM mix of each dNTP, 0.5 µl or 0.75 µl MgCl$_2$ (Invitrogen) (1.0 mM or 1.5 mM), 1 × PCR Reaction buffer (Invitrogen), 1 unit of Taq Polymerase (Invitrogen) and 1 µl of genomic DNA (50 ng/µl), made up to 25 µl with nuclease-free water. Thermocycling was performed on a Veriti 96-well Thermal Cycler (Applied Biosystems) and consisted of a denaturation step at 96 °C for 3 minutes, then 35 cycles of 92 °C for 30 seconds, 55 °C - 65 °C for 30 seconds and 72 °C for 30 seconds then a final extension step at 72 °C for 10 minutes.

2.3.2 Long-range PCR

All long-range PCR thermocycling steps were performed in a Veriti 96-well Thermal Cycler (Applied Biosystems). Each 20 µl long-range PCR reaction consisted of ~50 ng of genomic DNA, 2 µl of 10 × SequalPrep^{TM} reaction buffer (Invitrogen), 0.3 µl of 5 U/µl SequalPrep^{TM} long polymerase (Invitrogen), 0.4 µl of dimethyl sulfoxide (DMSO) (Invitrogen), 2 µl of 10 × SequalPrep^{TM} Enhancer A (Invitrogen), and 0.5 µl 10pmol/µl forward primer and 0.5 µl 10pmol/µl reverse primer, made up to 20 µl with nuclease-free water. Long range PCRs were performed using a Hybaid HBPXE02110 PXE 0.2 Thermal Cycler (Thermo Scientific). Thermocycling consisted of an initial denaturing step of 94 °C for 2 minutes, followed by 10 cycles of 94 °C for 10 seconds, 50-60 °C for 30 seconds and 68 °C for 1 minute/kb of the target. This was

then followed by 15-25 cycles of 94 °C for 10 seconds, 50-60 °C for 30 seconds and 68 °C for 1 minute/kb of target (+20 seconds per cycle). And a final extension step of 72 °C for 10 minutes.

2.3.3   Gel electrophoresis

PCR products were sized by agarose gel (Thermo Fisher) electrophoresis. The gel was prepared by dissolving agarose (Sigma Aldritch) in 0.5 × TAE. This was done using 0.5% w/v to 2% w/v agarose depending on the size of the target product to be analysed. The agarose gel was pre-stained using 4-6 µl of Midori Green Advance (Geneflow) per 100 ml of agarose gel. Electrophoresis was performed in 0.5× TAE buffer in an electrophoresis rig (Thermo Fisher). Gels were separated using a PowerPac Basic Electrophoresis Power Supply (BioRad) at between 70-140 millivolts. DNA was loaded with 6× DNA Gel Loading Dye (Thermo Fisher). GeneRuler 1 kb Plus DNA ladder (Thermo Fisher), Quick-Load 1 kb Plus DNA ladder (New England Biolabs) and EasyLadder (BioLine) were used to size DNA fragments. To visualise the DNA, a Bio-Rad gel documentation system using ultraviolet illumination and Image Lab 1-D software were used.

2.4   Sanger sequencing

Amplified products were prepared for Sanger Sequencing by first removing unincorporated dNTPs and primers using ExoSAP-IT (Thermo Fisher) at a ratio of PCR product:ExoSAP-IT of 5:2. The mixture was then treated at 37 °C for 15 minutes and inactivated at 80 °C for 15 minutes.

Sequencing reactions were prepared consisting of 1.6 pmol of primer, 0.75 µl of BigDye Terminator (Thermo Fisher), 1.5 µl of 5× BigDye Terminator Sequencing Buffer (Thermo Fisher), 1 µl of ExoSAP treated DNA and dH_2O to a final volume of 10 µl. Thermocycling for Sanger sequencing was performed using a Hybaid HBPXE02110 PXE 0.2 Thermal Cycler (Thermo Scientific).

A sequencing reaction was then performed consisting of an initial denaturation step of 96 °C for 1 minute, followed by 25 cycles of 96 °C for 10 seconds, 50

$^{\circ}$C for 5 seconds and 60°C for 4 minutes using a Veriti 96-well Thermal Cycler (Applied Biosystems). The samples were precipitated by adding 5 µl of 125 mM EDTA and 60 µl of 100% ethanol. This was chilled for 15 minutes. The samples were next centrifuged at 3061 relative centrifugal force (g) for 30 minutes at room temperature. The supernatant was decanted and traces were removed by low speed centrifugation on an inverted plate. Following this, 60 µl of freshly prepared 70% ethanol was added and this was centrifuged at 805 g for 15 minutes at 4°C. Again, the supernatant was removed by centrifugation. The dried samples were then resuspended in 10 µl HiDi Formamide™ (Thermo Fisher) and sequenced in an ABI3130xl Genetic Analyser (Applied Biosystems). POP7 polymer, 3730 sequencing buffer and FragmentAnalysis36_pop7_1 module was used for all runs. The resulting data was analysed using 4Peaks v1.8 by Nucleobytes (https://nucleobytes.com/4peaks/).

2.5 HLS-CATCH

High molecular weight DNA Library System (HLS) - Cas9-assisted targeting of chromosome segments (CATCH) experiments were performed in accordance with protocols provided by Sage Sciences (CATCH Guide for SageHLS, March 2018) and published by Zhou et al., 2020.

2.5.1 Nuclei isolation

Fresh whole blood was collected in Acid Citrate Dextrose or sodium EDTA blood collection tubes. These were stored for up to 5 days at 4 °C. Nuclei were extracted according to instructions provided by Sage Science (Preparation of WBC or Cultured Cell Nuclei for SageHLS Workflows 20220422). Unless otherwise indicated, all reagents were provided in the HLS-CATCH kit HIT004 (Sage Sciences). Briefly this process consisted of lysing the red blood cells using the provided red blood cell lysis buffer and isolating the total white blood cells by high speed centrifugation.

The pelleted white blood cells were resuspended in chilled Mg^{2+}/Ca^{2+}-free PBS (approximately 0.5-1 ml for up to 1×10^9 mammalian cells is recommended by the manufacturer). In this case, 0.5 ml of PBS was used. The nuclei were then isolated using chilled Nuclei Preparation Buffer (NPB)

by addition of 9.5 mls of NPB (Section 2.1). The nuclei were pelleted by centrifuging for 3 minutes at 3,224 g. The supernatant was decanted and the nuclear pellet was resuspended in in 200 ml of NPB and quantified.

To quantify the nuclei, 10 µl of resuspended nuclei were transferred to a 1.5 ml tube. 190 µl of Sage Qubit lysis buffer was added to this. This was mixed by vigorous pipetting to avoid the DNA forming a clot, but the generation of bubbles was avoided. The tube was mixed at maximum speed for 1 minute. 500 µl of TE was added, and again the tube was vortexed for 1 minute. 5 µl of the lysed sample was then transferred to a Qubit assay tube (Thermo Fisher). 194 µl Qubit HS buffer (Thermo Fisher) and 1 µl of Qubit HS Reagent (Thermo Fisher) were used. DNA concentration was measured using a Qubit Fluorometer (Invitrogen). The DNA content of the nuclei suspension was calculated by:

[Qubit tube conc, ng/ml] × (800/10) × (200/5) = [DNA conc original cell suspension, ng/ml]

[Dilution factor of diluted lysate] × (Dilution factor Qubit HS assay)

This was done in triplicate for all samples.

2.5.2 HLS-CATCH workflow

HLS-CATCH was performed on a HLS2 HMW Library System with HLS-CATCH cassettes (Sage Sciences) following the manufacturer's instructions (CATCH Guide for SageHLS, March 2018). The workflow "CATsCH100-300 extr3h inj4m80V sep3h" was used for all experiments. All reagents for HLS-CATCH were supplied by Sage Sciences (cat. number HIT004).

Briefly this consisted of introducing 70 µl of the prepared nuclei (corresponding to approximately 375,000 nuclei) to each lane of the cassette and lysing the nuclei in the cassette using 230 µl of lysis buffer. This was done during a 3 hour extraction step. In this step, the nuclei are lysed and the genomic DNA is pulsed on to the side of the well. As the genomic DNA is very high molecular weight, it is too large to pass through the agarose gel and remains on the side of the well.

Following nuclei lysis, prepared Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR-associated (Cas) endonuclease 9 complexes (CRISPR/Cas9) (section 2.5.3) were introduced to the cassette. 80 µl of the prepared Cas9 reaction was added to each lane of the cassette. The CRISPR-Cas9 complexes were allowed to digest the target for 30 minutes. To stop the reaction, the reagent well contents were replaced with 230 µl of HLS Lysis Buffer. The cassette was then left to run for 4.5 hours. This consisted of a 3 hour separation step, followed by a 1.5 hour elution step. During this, the fragmented DNA is separated by size by gel electrophoresis and then migrates to the elution wells. Following this, the size separated DNA was transferred from the elution wells to separate 1.5 ml Eppendorf LoBind tubes using wide bore pipettes with an electronic pipette set to the slowest setting. While transferring, the volume of elute transferred to each tube was measured.

2.5.3 Cas9 enzyme assembly

RNA sequences (IDT) were provided by Dr Gavin Arno (UCL). These were designed using the IDT online Custom Alt-R CRISPR-Cas9 guide RNA design tool (https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE). Guides were optimised for cutting efficiency by Dr Gavin Arno by digestion of PCR products containing the target sequence.

The crRNA and tracrRNAs were dissolved at 100 µM concentration in Duplexing Buffer (IDT). For each well, a 200 µl reaction was prepared in a PCR tube containing 15.4 µl of duplexing buffer (IDT), and 4 µl of a pooled mixture of the crRNAs. Two crRNAs were used per experiment, one targeting 5′ of the target and one 3′. This was then mixed thoroughly by vortexing. 2.6 µl of tracRNA was added and the solution was mixed by vortexing. To anneal the guide RNAs, the 20 µl mixture was heated for 5 minutes at 95 °C in a thermocycler. The tube was then allowed to cool for five minutes at room temperature.

The Cas9 complex was assembled by adding, in order; 10 µl Sage HLS 4× Enzyme buffer, 22 µl annealed guides (from previous), 5.3 µl of molecular biology grade water, 2.7 µl of Cas9 (IDT). The mixture was then incubated at

37 °C for 10 minutes. The assembled Cas9 mixture was then diluted with 40 µl of HLS enzyme buffer and the diluted mixture was stored on ice.

2.5.4 qPCR

A qPCR was then performed on the products from each of the sample wells to verify which sample well was enriched for the gene of interest using a QuantStudio 7 Flex Real-time qPCR System (Thermo Fisher). A custom TaqMan qPCR assay (Thermo Life) was used. Probes were designed targeting RNaseP as a control gene and custom probes were designed targeting the gene of interest. This was performed using genomic DNA from a control specimen. All qPCRs were performed in triplicate on both the control DNA and the contents of the elution wells. 1 µl of each target well was compared to an input of 1 µl control DNA at 100 ng/µl, 50 ng/µl, 25 ng/µl, 12.5 ng/µl, 6.25 ng/µl, 3.12 ng/µl and 1.6 ng/µl.

2.5.5 HMW DNA clean-up and concentration

The product resulting from HLS-CATCH was cleaned and concentrated using Sage Hi-Bead reagents according to the manufacturer's instructions (CATCH Guide for SageHLS, March 2018). The Sage Hi-Bead reagents were first brought to room temperature. The volume of elute in each LoBind tube was defined as "one volume", generally between 70 µl and 80 µl. One volume of HLS Binding Buffer was added to the side of the tube above the level of the liquid. The tube was placed on its side on a rotator such that the liquid did not invert in the tube. This was rotated for 10 minutes at 4 rpm. The mixture was then incubated at room temperature for 5 minutes.

The HLS Hi-Bead Suspension was resuspended by inverting the mixture gently to avoid creating detergent foam. One volume of resuspended beads was then added to the tube containing the elute. This was mixed by gentle flicking. The tube was again placed on its side on a rotator so the liquid did not invert and rotated for 30 minutes at 4 rpm.

After 30 minutes the tubes were placed on a magnetic rack. Once the sample had cleared and the beads had bound to the magnet, the supernatant was removed and the sample was washed twice with 500 µl of freshly prepared 80% ethanol. The beads were then left to dry for ~1 minute. The beads were

then resuspended in 20 µl of 1× TE Buffer, after which the tubes were placed on their side on a rotator which was placed in an incubator or warm room set to 37 °C. This was rotated for 2 hours at 4 rpm. The tubes were then spun down and stored overnight at 4 °C. The next day, the tubes were transferred back to the magnetic rack and the supernatant was transferred to fresh 1.5 ml LoBind tubes using a wide bore pipette.

2.6   Nanopore sequencing

2.6.1   Library clean-up

Amplified samples were cleaned for use on an Flongle flowcell (Oxford Nanopore Technologies) using an AxyPrep Mag PCR Clean Up Kit (AppletonWoods). This removed dNTPs, salts and enzymes and PCR primer dimers. AxyPrep beads were added to the products to be sequenced in a Product:Beads 1:0.8 ratio and resuspended thoroughly. This was incubated on a Hula Mixer for 5 minutes. After 5 minutes, the tube was placed on a magnetic rack and the tube was left until the solution was clear. The supernatant was then removed by slow pipetting and the beads were washed twice with freshly prepared 80% ethanol. After washing, the ethanol was removed from the bottom of the tube using a P2 pipette. Following evaporation of the ethanol (a maximum of 1 minute exposure to air) the tube was removed from the rack and the beads were resuspended in a convenient volume (3-10 µl) of nuclease-free water (Thermo Fisher). The mixture was incubated at room temperature for 5 minutes, after which the tube was returned to the magnetic rack. Once the beads had migrated to the side of the tube, the supernatant was removed and retained.

2.6.2   Long-range PCR sequencing

For long range PCR products, nanopore-sequencing compatible DNA libraries were prepared using the LSK109 ligation sequencing kit (Oxford Nanopore Technologies [ONT]). Initially, end-repair and nickase treatment reaction was performed, this comprised 1.75 µl of Ultra™ II end prep reaction buffer (New England Biolabs [NEB]), 1.75 µl FFPE DNA repair buffer (NEB), 1.5 µl Ultra ™ II end prep enzyme mix (NEB), 1 µl FFPE DNA repair mix (NEB), 13.5 µl nuclease-free H_2O, and 10 µl of equimolar PCR products at (50 ng/ µl). The reaction was incubated at 20 °C for 5 minutes then 65 °C for 5 minutes. A further bead-based clean-up reaction was performed using AMPure XP

beads, after which sequencing adaptors were ligated to the treated PCR products. This reaction comprised 30 µl of treated PCR products, 12.5 µl of Ligation Buffer (ONT), 5 µl of Quick Ligase (NEB), and 5 µl of AMX Adapter Mix (ONT). The reaction was incubated at room temperature for 10 minutes, after which a final AMPure XP bead clean-up was performed with washes that used 250 µl Long Fragment Buffer (ONT) rather than ethanol. The final library was eluted in 6 µl of Elution Buffer (ONT) following a 10-minute incubation at 37 °C. Each library was sequenced using a Flongle flowcell (R.9.4.1) on a MinION (ONT). A 24-hour run was initiated using MinKNOW v. 3.6.5.

For combining multiple amplicons into a single sequencing run, equimolar concentrations of each amplicon were loaded onto the Flongle. These were normalised by quantification using a Qubit Fluorometer (Invitrogen) using the dsDNA BR assay kit (Invitrogen), aiming to pool 500 ng of amplification products for library preparation.

2.6.3   Amplification-free DNA sequencing

For CATCH products, a modified library preparation protocol using the LSK110 ligation sequencing kit was performed. If multiple elution wells were used, they were pooled and 49 µl of the pooled, cleaned sample was transferred to a clean 1.5 ml LoBind tube (Eppendorf). End-repair and A-tail buffer (NEB) (7 µl) and enzymes (NEB) (3 µl) were added to the tube. The solution was mixed by gently tapping and transferred to a 0.2 ml PCR tube. This was incubated at 20 °C for 30 minutes and then 65 °C for 30 minutes. The tube was then allowed to cool to room temperature and 25 µl of LNB buffer, 10 µl of T4 quick Ligase, and 5 µl of AMX-F were added in order and mixed by gentle tapping. This was left at room temperature for 10 minutes. The reaction was then purified using the AMPure Bead clean up protocol (Section 2.6.1) except that the tube was placed on its side on the rotator as during Section 2.5.6. The cleaned beads were dried for 30 seconds and the library was eluted into 12.5 µl of EB buffer. The prepared library was then sequenced on a MinION R.9.4.1 flowcell, using either a standalone MinION device or a MinION Mk1C (with integrated data acquisition hardware).

2.7 Bioinformatics

2.7.1 Sequence visualisation

Several software packages were used to visualise sequence data. 4Peaks (v.1.8) was used for interpretation of Sanger sequencing electropherogram traces (https://nucleobytes.com/4peaks/). The UCSC (University of California, Santa Cruz) Genome Browser (http://genome.ucsc.edu/) was used for the identification of genomic features, including the location of introns and exons, common SNPs, gene isoforms, and for BLAT localisation of identified sequences. The Integrative Genome Viewer (IGV) v.2.7.2 and v.2.14.1 was used to visualize aligned short-read (Illumina) and long-read (nanopore) datasets (Robinson et al., 2011). The IGV was also used for the interpretation of constructed phase blocks, visualization of called variants and interrogation of chimeric reads.

2.7.2 Nanopore analysis

Nanopore analysis was performed through a pipeline initially established by Dr Watson and colleagues (Watson et al., 2019). A complete list of commands can be viewed in Appendix 1.

Bioinformatics packages were downloaded and managed using Bioconda, v.4.10.3 (https://docs.conda.io/en/latest/).

Basecalling, to convert raw data from FAST5 to FastQ format was performed using Guppy v.5.0.16 (http://nanoporetech.com) on a Nvidia GPU. For CATCH experiments, basecalling was performed on the MinION Mk1C using the High Accuracy Calling (HAC) model.

Adapter sequences were trimmed from read ends using Porechop v.0.2.4 (https://github.com/rrwick/Porechop). NanoStat was used to generate run statistics and inform read length and quality filtering steps (https://github.com/wdecoster/nanostat) (De Coster et al., 2018). NanoFilt v.2.8.0 (https://github.com/wdecoster/nanofilt) (De Coster et al., 2018) was used to remove low-quality reads (-q 10) and select reads that were between 3.0-14.5 kb in length (this step also removed the first 75 bp from each read which are typically of lower quality). Reads were aligned to the human

reference genome (build hg19, and subsequently build GRCh38) using minimap2 v.2.22 (https://github.com/lh3/minimap2) (Li, 2018).

File manipulation, including SAM to BAM conversion and read sorting by genomic coordinate, was performed using SAMtools v.1.9 (http://www.htslib.org/) (Li et al., 2009).

A number of variant callers were examined. These were compared with a high quality, benchmark VCF. These were; Medaka, which uses neural networks applied to a pileup of sequencing reads to perform variant calling (https://github.com/nanoporetech/medaka). LongShot which uses the haplotype information present in long single molecule reads to perform variant calling (https://github.com/pjedge/longshot) (Edge and Bansal, 2019). NanoPolish which uses signal level information from fast5 files to perform variant calling v.0.13.2. (https://github.com/jts/nanopolish) (Simpson et al., 2017). Later Clair3, which combines both pileup calling and full-alignment for variant calling was also examined (https://github.com/HKU-BAL/Clair3) (Zheng et al., 2022). Comparison with the Platinum VCF was performed by measuring the number of variants called in both datasets (considered true positives), the number of variants called only in the nanopore dataset (considered as false positives) and only called in the Platinum VCF (considered as false negatives).

Haplotypes were constructed using WhatsHap v.1.1 (https://github.com/whatshap/whatshap) (Martin et al., 2016). WhatsHap phases variants in a VCF using associated long-read sequencing reads. This phased VCF is then used to tag individual reads belonging to either haplotype. To inspect haplotypes manually, read groups were isolated by segregating reads based on a single variant using the Jvarkit tool biostar214299 (http://lindenb.github.io/jvarkit/Biostar214299.html) (Lindenbaum, 2015). This tool allows for the selection of reads that were mapped across a specific genomic coordinate. The separated reads were used to verify haplotypes constructed by WhatsHap and to assess chimeric read formation.

The NA12878 case was used as a benchmark for testing the accuracy of base calling, variant calling and phase block formation (Dausset et al., 1990). A

whole genome NGS dataset was downloaded from (https://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/NA12878/NIST_NA12878_HG001_HiSeq_300x/RMNISTHS_30xdownsample.bam) This was used to visually compare nanopore data and NGS data through IGV. 'Platinum' VCFs was downloaded from the Genome in a Bottle Dataset (https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/NA12878/) (Zook et al., 2014).

For comparison of variant calling, VCFs generated using the NanoPolish variant caller, were compared to HG001_GR37.1_22v4.2.1_benchmark.vcf.gz from GIAB. For verification of phase blocks generated by WhatsHap, tagged VCFs and BAMs was compared to sp_v37.7.0_NA12878.vcf from GIAB. These VCFs were generated by high quality Illumina sequencing data, phasing information was generated by sequencing of 11 family members, allowing this to be used as a 'truth set'. Comparison was performed using VCFtools (v.0.1.16) (Danecek et al., 2011).

Variants located in target loci (the overlaps between adjacent long-range amplification products) were isolated using BEDTools (v.2.30) (https://bedtools.readthedocs.io/en/latest/).

Graphs were generated using RStudio (https://www.r-project.org/). The ggplot2, and ggrepel libraries were loaded (Villanueva and Chen, 2019).

Whole genome datasets for use as *in silico* controls were obtained from the Personal Genomes Project (https://www.personalgenomes.org.uk/).

2.7.3 Variant analysis

Population level variation was obtained from gnomAD v.2.1.1 (https://gnomad.broadinstitute.org/). This was created by exporting all variants in the region of interest present in more than 10% of the general population to a CSV file. This dataset was obtained from whole genome data only.

Potentially pathogenic variants of interest in IRD genes were examined using the Leiden Open Variation Database (LOVD, https://www.lovd.nl/) , ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) and MutationTaster (https://www.mutationtaster.org/) (Anon, 2015). LOVD is a repository of previously reported variants in IRD genes and ClinVar is a public repository with reported variant pathogenicity and supporting evidence. Previously identified copy number variants were examined using *D*atabas*e* of *C*hromosomal *I*mbalance and *P*henotype in *H*umans Using *E*nsembl *R*esources, DECIPHER (www.deciphergenomics.org) (Firth et al., 2009).

2.8 smMIPs sequencing

2.8.1 smMIPs sample preparation

Genomic DNA samples were quantified using a Qubit Fluorometer and diluted to a starting DNA concentration of 15-25 ng/µl. Sample quality was determined by gel electrophoresis of 100 ng of patient DNA sample using a 1 kb plus ladder (Invitrogen) and Lambda DNA Hind III digest marker (Thermo Fisher). Samples with sufficient DNA integrity of both low and high molecular weight proceeded to library preparation.

Sequencing of patients using smMIP panels was performed by Prof Frans Cremers and colleagues at Radboud MC using the High Input DNA Capture Kit, Chemistry 2.3.0H produced by Molecular Loop Biosciences Inc. Protocol version 2.4.1H was used. smMIPs were designed as part of a macular dystrophy panel to capture the coding regions of 105 genes and non-coding loci associated with MD and age-related macular dystrophy, totalling 17,934 probes. Sequencing was performed by paired-end sequencing on a NovaSeq 6000 platform (Illumina, California, USA) using two SP reagent kits v1.5 (300 cycles), each with a capacity of 1.3 – 1.6 billion paired end reads per run (Hitti-Malin et al., 2022).

Bioinformatic analysis was performed by collaborators in Radboud MC and the Radboud Genomics Technology Center (Maartje van de Vorst, Bart de Koning, Marlie Jacobs-Camps, Anita Roelofs, Saskia van der Velde-Visser, Christian Gilissen, Marcel Nelen). Details of this have been published by Khan et al., 2019, Khan et al., 2020, Hitti-Malin et al., 2022 and Panneman et al., 2022. Briefly this consisted of converting reads to fastq files using bcl2fastq

(https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) (Illumina). The tags were then separated from the quality controlled reads and this information was stored in the read header. The reads were then aligned using BWA-mem (https://github.com/lh3/bwa) (Li and Durbin, 2009). The unique molecular identifier was then used to remove any PCR duplicates and obtain a consensus sequence for each smMIP. The patient indexes were then used to separate the reads and write each patient reads into a separate BAM file. Variant calling and annotation was then performed using GATK and UnifiedGenotyper (https://gatk.broadinstitute.org/hc/en-us).

2.8.2   CNV analysis

CNV analysis was performed according to an excel script provided by Dr Rebekkah Hitti-Malin and Zelia Corradi (Radboud MC) (Khan et al., 2020; Hitti-Malin et al., 2022). Briefly this consisted of calculating the average per smMIP target coverage across all 384 patients in the run and comparing each target to the average coverage of each target. A value of 1 indicated that the target for that case had the same number of reads as the average in that run. A value of less than 0.65 was considered a potential heterozygous deletion, a value of less than 0.1 was considered a potential homozygous deletion. Conversely a value of 1.2 or greater was considered a potential heterozygous duplication and a value of 1.7 or greater was considered a potential homozygous duplication. The presence of six consecutive smMIPs indicating either a deletion or a duplication warranted visual inspection with IGV. Full details of this are detailed in Appendix 2.

2.8.3   Variant prioritization

Variants were prioritized according to an interpretation workflow provided by Dr Rebekkah Hitti-Malin and Zelia Corradi and detailed in Hitti-Malin et al., 2022. This protocol priortized variants based on predicted pathogenicity. Initially, variants previously found to be pathogenic were identified. Next, variants previously reported in LOVD and by the Radboud MC group were identified (https://databases.lovd.nl/shared/genes/ABCA4) (Cornelis et al., 2017; Zernant et al., 2017; Hitti-Malin et al., 2022). Next all of the called variants which were deemed too common to cause rare IRDs were excluded using gnomAD and a Radboud MC dataset consisting of 24,488 individuals who had previously been sequenced. Variants present in more than 10% of

probands in the total sequencing run were also excluded. Following this, rare variants with greater than 10 × coverage and an allele fraction between 35-80% were considered heterozygous and variants with greater than 10 × coverage above 80% base fraction were considered homozygous. Variants in genes associated with recessive disease with a frequency greater than 0.5% in the general population and variants in genes associated with dominant disease with a frequency greater than 0.1% were excluded (Hitti-Malin et al., 2022). Variants were then prioritised (from high to low priority) based on whether the variant; caused a frameshift, induced a stop gain or altered a canonical splice site; caused an in-frame deletion; variants overlapping known splice sites; was likely to be a pathogenic missense variant; and finally if the variant was silent. Missense variants were prioritised using the following *in silico* prediction tools: PhyloP (predicted pathogenic ≥2.7) (Pollard et al., 2010), CADD-PHRED (range: 1-99; predicted pathogenic ≥15) (Kircher et al., 2014) and Grantham (range: 0-215; predicted pathogenic ≥80) (Grantham, 1974). Variants that met all three criteria (allele frequencies, predicted protein effect and variant type) were prioritized, followed by those that met the threshold scores of any *in silico* tools used.

2.8.4 Variant classification

All candidate variants were manually inspected in IGV. Putative splice variants were assessed using SpliceAI. SpliceAI was used via the BROAD Institute web interface tool (https://spliceailookup.broadinstitute.org/#) to further predict splicing effects, using a 10,000 bp (5,000 bp upstream and 5000 bp downstream) window. Variants with a predicted delta score of ≥0.2 (range: 0–1) for at least one of the four predictions (acceptor loss, donor loss, acceptor gain, donor gain) were considered potential candidates.

All prioritised variants were classified according to ACMG guidelines (Richards et al., 2015) using the Franklin by Genoox platform (https://franklin.genoox.com/clinical-db/home). ACMG guidelines categorise variants into; class 1 (benign); class 2 (likely benign); class 3 (variant of uncertain significance i.e. VUS); class 4 (likely pathogenic) and class 5 (pathogenic) Benign and likely benign variants were not considered unless they were previously shown to be causative. Prioritised variants were also interrogated using the Leiden Open (source) Variation Database (LOVD; https://www.lovd.nl/), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and

MutationTaster (https://www.mutationtaster.org/) (Schwarz et al., 2014). As segregation analysis was not performed for most cases, cases were classified as 'unsolved', 'possibly solved' or 'very likely solved' according to the following criteria provided by Prof Frans Cremers, Dr Rebekkah Hitti-Malin and Zelia Corradi (Table 2.1).

Table 2.1 Variant interpretation for smMIPs screen

Recessive inheritance			
Zygosity	**Variant 1**	**Variant 2**	**Prediction**
Homozygous	VUS		Possibly solved
Homozygous	Likely Pathogenic		Very likely solved
Homozygous	Pathogenic		Very likely solved
Heterozygous	VUS	VUS	Possibly solved
Heterozygous	Likely Pathogenic	VUS	Possibly solved
Heterozygous	Pathogenic	VUS	Very likely solved
Heterozygous	Likely Pathogenic	Pathogenic	Very likely solved
Heterozygous	Likely Pathogenic	Likely Pathogenic	Very likely solved
Heterozygous	Pathogenic	Pathogenic	Very likely solved
Dominant inheritance			
Zygosity	**Variant**		**Prediction**
Heterozygous	VUS		Unsolved
Heterozygous	Likely Pathogenic		Very likely solved
Heterozygous	Pathogenic		Very likely solved

2.9   DNA isolation

All media were autoclaved at 115 ºC, 0.6 kg/cm^{-3} for 30 minutes and cooled prior to use. Antibiotics were added, after autoclaving, to cool (55 ºC) liquid cultures and plates, where required. For plates 1.5% agar was added to the appropriate broth before autoclaving.

2.9.1   BACs

Bacterial artificial chromosomes (BACs) (Osoegawa et al., 2001) were initially used to optimise the long-range PCR at the of the *ABCA4* locus. Human RPCI-11 BACs (Source Bioscience) were received in an E. *coli* glycerol stock. These were streaked and grown on agar plates. Then, single colonies were picked and inoculated in 5 mls of LB broth supplemented with

chloramphenicol. This was grown in a shaking incubator overnight at 225-300 rpm at 37 °C.

The BAC DNA was extracted using an adapted rapid alkaline lysis miniprep. The LB broth was centrifuged at 800 g for 10 minutes and the supernatant was discarded. The pelleted bacteria were then resuspended in 0.3 mls of P1 (50 mM Tris, pH 8, 10 mM EDTA, 100 μg/ml RNase A) and 0.3 mls of P2 (0.2M NaOH, 1% SDS) and incubated at room temperature for 5 minutes. Then 0.3mls of P3 (3M ammonium acetate, pH 5.5) was added. This resulted in precipitation of E. Coli protein and DNA. This was pelleted by centrifugation at 8000 × g for 10 minutes at 4 °C. The supernatant was then removed and transferred to a 1.5 ml Eppendorf containing 0.8 ml ice-cold isopropanol. This solution was then spun in a cold microfuge for 15 minutes. The supernatant was removed and two washes were performed with 0.5mls of 70% ethanol. The DNA pellet was then resuspended in 40 µl of TE.

2.9.2 Cell culture

Cells from patient NA12878 were purchased as a lymphoblastoid cell line (GM12878, Coriell Institute). The B-Lymphocytes donated by patient NA12878 were transformed with Epstein-Barr Virus as part of the CEPH project (Cann, 1992). Cells were propagated in Corning T25 or T75 flasks (Sigma Aldrich) at 37 °C with 5% CO_2 in Sanyo MCO 20AIC incubators. Cell culture work was performed in NuAire Labgard 437 ES Class II Biosafety Cabinets under sterile conditions. Lymphoblastoid cells were grown in RPMI 1640 + L-glutamine media with 10% foetal calf serum and and 100 U/ml penicillin and 100 mg/ml streptomycin (Sigma Aldrich). Cells were sub-cultivated by adding fresh medium or by splitting at a split ratio of 1:6 every 3-5 days. Fresh cultures were seeded with at least 200,000 viable cells/ml. Cells were stored in liquid nitrogen in foetal calf serum with 10% DMSO. Genomic DNA was extracted from the cultured cells using QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's guidelines.

3   Assay to phase *ABCA4* by long range sequencing

3.1   Introduction

While sequencing-by-synthesis NGS has revolutionised the study of IRDs, the resulting short read lengths limited the investigation of several clinically-relevant scenarios. In particular, currently it is not commonly possible to determine the phase of detected variants unless familial DNA is available or the variants are within the same sequenced fragment. For cases with compound heterozygous variants in genes associated with recessive disease, if no familial DNA is available the variants are commonly assumed to be in *trans*, raising the possibility that many 'solved cases' are not genetically explained by their known variants. This also limits the study of complex alleles, which are becoming increasingly important in the study of IRDs. Variants in *ABCA4* are the most common cause of IRDs and there is a growing need to phase variants in patients, especially with an increasing number of potential therapies under development. *ABCA4* is a suitable gene to develop a phasing assay as complex alleles in *ABCA4* are commonly observed in cases with *ABCA4*-related disease, comprising approximately 10% of alleles (Cremers et al., 2020). Additionally, as *ABCA4* is a large gene at ~130 kb, any effective phasing strategy could be applied to other genes.

Several methods have been developed to phase large loci, however these often rely on the creation of ultra-long reads or phasing by segregation using familial DNA. Long-read methods generally use either direct sequencing of genomic DNA or amplification-free enrichment methods, both of which require a large input of high molecular weight DNA to achieve sufficient read depth, limiting the number of samples they can be used on. The aim of this study was to develop a cost-effective strategy to reliably phase distant variants at the *ABCA4* locus in the majority of samples.

3.2 Results

3.2.1 Design and optimisation of long-range PCR assay to phase *ABCA4*

The first approach used to phase the 128.3 kb *ABCA4* locus was based on amplifying the locus in relatively small overlapping amplicons (each spanning 6-12 kb) and sequencing these on a MinION sequencer using a Flongle adaptor. This strategy was chosen as many of the samples to be phased were old, with reduced DNA quality and limited mass of available DNA. Phase blocks were constructed from heterozygous variants identified in the overlapping genomic intervals of adjacent amplification products (Figure 3.1).
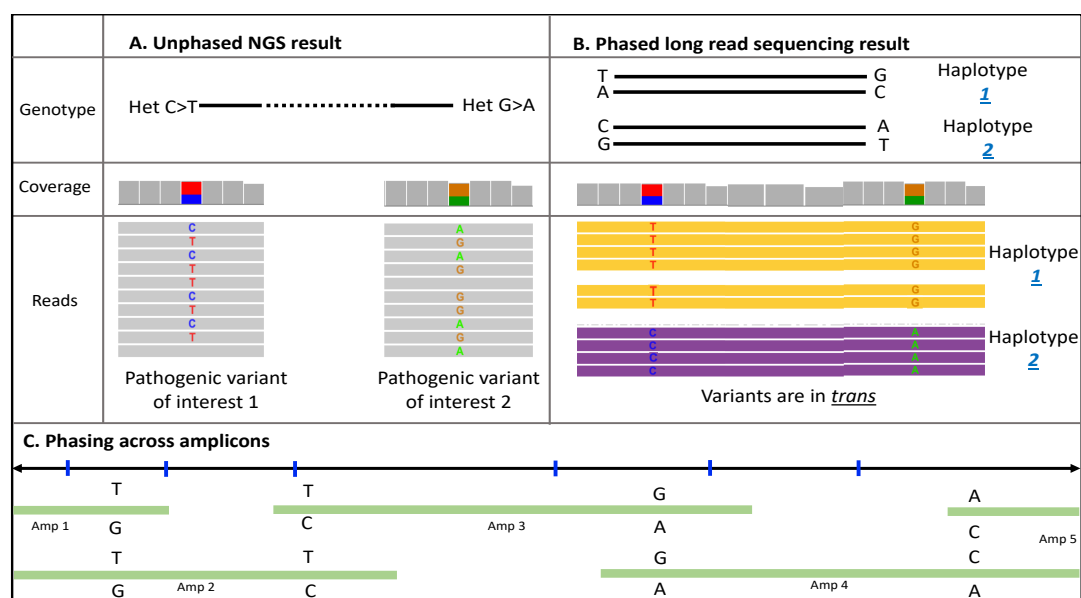
Figure 3.1 Phasing by long-range PCR and long-read sequencing.

> **A)** Unphased result from short-read NGS sequencing where two heterozygous pathogenic variants of interest have been identified but were not captured on the same read. Without family members to allow phasing by segregation analysis, these variants would be assumed to be in trans. **B)** Long-read sequencing of the same variants. Two haplotypes have been assigned, showing the variants of interest are in trans. **C)** Phased amplicons (green) can then be stitched together to create larger phase blocks using heterozygous variants in the overlaps between amplicons.

To assess the size of the amplicons and overlaps which would be required to capture sufficient variation to allow for phasing, the density of polymorphic variants at the 128.3 kb *ABCA4* locus was assessed using gnomAD v2.1 (Section 2.7.3). This yielded 385 variants with an allele frequency of >0.1 with

a theoretical distribution of 2.96 variants per kb. From this, a design strategy of tiled PCR amplicons between 7-10 kb in length with a minimum overlap of 2 kb was adopted. A total of 22 amplicons were designed spanning 132,260 bp. Details of primers reported in Appendix 3.

To facilitate optimisation of the 22 primer sets, DNA extracted from Bacterial Artificial Chromosomes (BACs) spanning *ABCA4* was used as a positive control (Osoegawa et al., 2001). BACs were initially used to develop a working set of primer conditions without off-target effects. BACs derived from library RPCI-11 were identified using UCSC (Section 2.7.1) (Appendix 4. The BACs were streak purified from glycerol stocks, grown overnight and DNA was extracted (Section 2.9.1).

Primers were optimized by varying the annealing temperature and volumes of enhancers supplied with the SequelPrep™ long-range PCR kit. Details of conditions are reported in Appendix 3 The optimized primers were then validated using human genomic DNA from a control. All long-range PCRs were performed at 35 cycles of amplification. An example of primer optimization using BACs is shown in Figure 3.2.

To initially test whether phasing using tiled long-range PCR products was feasible, a series of four test cases were selected. Two cases with previously identified *ABCA4* variant(s) (Probands 4777 and 4781) and two controls (1343 and NA12878) were selected. Case 4777 had previously identified biallelic *ABCA4* variants which required phasing to confirm their clinical relevance. Case 4781 had a single pathogenic *ABCA4* variant identified during targeted sequencing (*ABCA4*:c.1906G>A). This case was selected to attempt to identify and phase a second pathogenic variant in *ABCA4*.
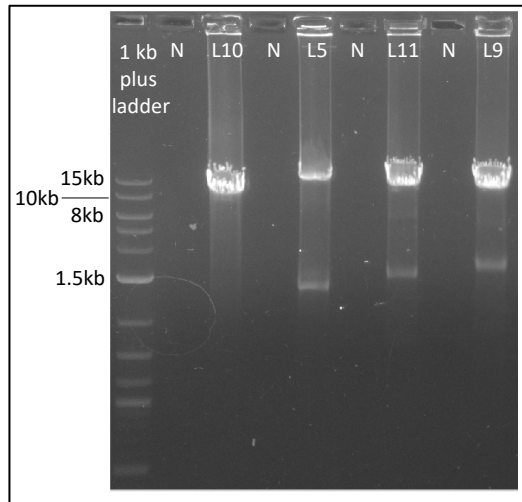
Figure 3.2 Example of optimization of long-range PCR products at *ABCA4* using BAC DNA

A sample of the PCR products amplified from control BAC DNA spanning *ABCA4* separated on a 0.5% agarose gel. Shown are a number of exemplar long-range PCR products amplified using an annealing temperature of 58 °C. Negative controls are indicated by an N. L10, L5, L11 and L9 indicate the name of the amplicon which was used to amplify the BAC DNA. A Thermofisher Scientific 1 kb plus ladder was used for fragment sizing. The expected size of the long-range PCR products was 10-13 kb. Non-specific amplification was also observed (visible bands of 1.5 kb).

The first control was Case 1343 which had a diagnosis of Late Onset Retinal Dystrophy (LORD) and had no molecular genetic diagnosis following multiple rounds of genetic testing. This case had undergone previous analysis by WGS as had several members of the same pedigree, as such this case was initially used as a benchmark. The second control was case NA12878 which was initially sampled as part of the CEPH project and has since become a widely used benchmark for sequencing strategies (Zook et al., 2014; Zook et al., 2016a; Choi et al., 2018; Helal et al., 2022). Lymphoblastoid cells were obtained and this was used as a source of control human genomic DNA (Section 2.9.2). The SequelPrep™ kit was used to amplify genomic DNA from all cases (Section 2.3.2), followed by size fractionation by gel electrophoresis using a 0.5% agarose gel with 0.5 × TAE buffer.

All 22 amplicons generated for each case were purified, quantitated and pooled before each case was sequenced on a single Flongle R.9.4.1 flowcell (Section 2.6). High read depths were obtained for each case, with a mean read depth of up to 3927 × at the target locus and between 94.2% and 99.7%

reads on target (Table 3.1). Variants were initially identified using Nanopolish with default settings. This test series revealed that while a set of working PCR amplicons could be established to span *ABCA4*, and that when sequenced this resulted in high read depth, for all cases there were recurring overlaps which did not have sufficient variation to enable phasing across *ABCA4*.

To further test the set of primers which had been used, a series of nine exemplar datasets were tested *in silico* to identify which overlaps required redesign. These consisted two in-house Familial Exudative Vitreoretinopathy (FEVR) cases who had undergone WGS (Cases F1268 and F1396) and seven cases from the Public Genomes Project (FR07961001, FR07961002, FR07961003, FR07961005, FR07961006, FR07961008, and FR07961009) (Section 2.7.2). This *in silico* testing was performed by isolating the previously called variants in the overlaps between each amplicon for each case. This indicated that three overlaps required redesign due to a consistent lack of variation (overlaps 13, 15 and 19, Figure 3.3).

Table 3.1 Sequencing statistics for long-range PCR test cases

All PCRs were performed at 35 cycles of amplification. Filtering was performed using NanoFilt with quality score Q7, excluding reads above 14,500 bases and smaller than 4,500 bases. The interval column refers to the distance between the indicated variants. Where no variants or a single variant is shown, the interval refers to the interval covered by the amplification product. All variants refer to NM_000350.3, hg19.

Local sample ID	Filtered read count	Reads on target	Proportion of reads on target (%)	Mean read depth at target locus (×)	Target variants			Target Amplicons (#)	Mean variants per overlap (#)	Overlaps without variants (#)
					Allele 1	Allele 2	Interval (b)			
NA12878	59842	58590	97.91	3137.26	N/A	N/A	132260	22	5.14	1
4781	71580	70222	98.10	3927.74	c.1906G>A	N/A	N/A	22	4.40	5
4777	48442	48298	99.70	2648.68	c.4469C>T	c.6112G>A	24039	22	3.04	5
1343	52049	49040	94.22	2747.32	N/A	N/A	132260	22	2.45	7
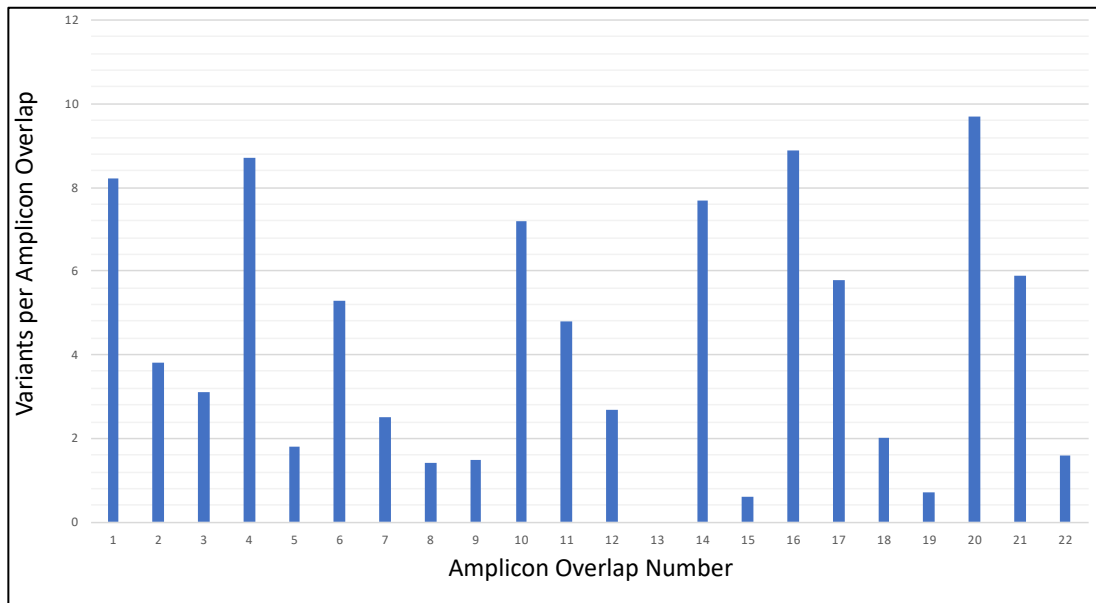
Figure 3.3 Mean number of variants per amplicon overlap calculated using exemplar datasets.

The mean number of variants in the overlaps between the designed amplicons across *ABCA4* was measured in nine exemplar cases *in silico*. This showed that several overlaps required redesigning to capture sufficient variation.

GnomAD was then used to inform the redesign of these amplicons to capture sufficient variation to allow robust phasing. Amplicons L17, S20, L14, L13 and L12 were redesigned to extend the identified overlap length between the amplicons (overlaps 13, 15 and 19) by at least 2 kb. Additionally, a locus was identified as amplifying poorly across the four test cases. This was supplemented with an additional amplicon (L8.2), resulting in 24 amplicons in total in the updated set of primers. The updated set of primers were then again tested *in silico* using the nine exemplar datasets. Across the exemplar datasets, all cases contained at least one variant per amplicon overlap. The average number of variants per amplicon overlap across the exemplar cases ranged from 1 to 9.44. It was also observed, utilizing the gnomAD v2.1.1 dataset, that there were an average of 8.9 variants in the redesigned overlaps. With this, the redesigned amplicons were predicted to capture sufficient variation to allow for robust inter-amplicon phasing. The NA12878 control and the control Case 1343 (with LORD) were then amplified using the updated set of primers and sequenced on separate Flongles. The resulting reads from Case NA12878 are visible in Figure 3.4. This resulted in complete coverage of *ABCA4* and high read depths across the locus.
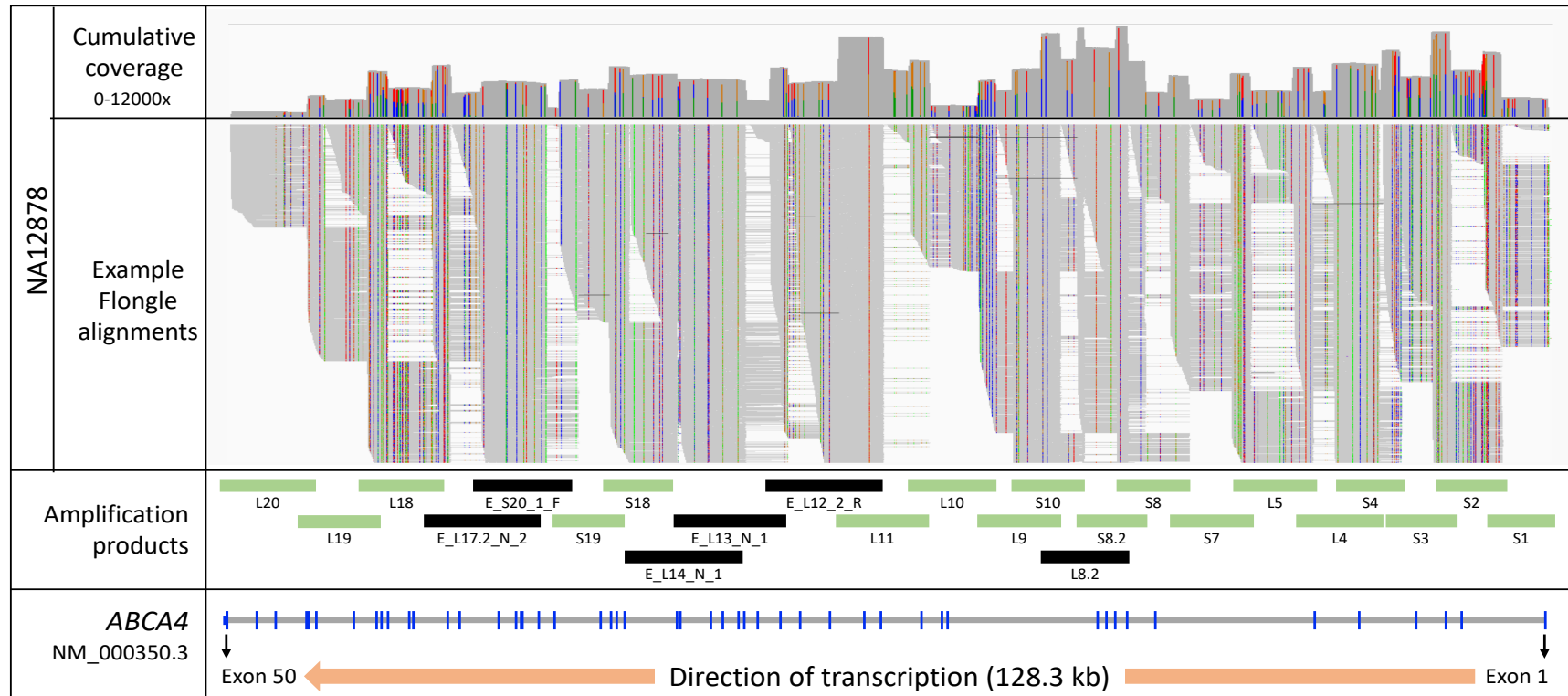
Figure 3.4 Aligned reads of Case NA12878 across *ABCA4* using updated long-range primer set.

The cumulative read depth across *ABCA4* is displayed in the top panel. The long reads spanning *ABCA4* are shown in the middle panel. The amplification products used are shown below this. In this panel, green bars indicate primers which were not redesigned, while black bars indicate primers which were redesigned. Only non-reference bases which are present in more than 33% of the reads are shown in the top two panels.

3.2.2 Development of bioinformatics pipeline to phase long-read sequencing data

3.2.2.1 Long-read variant caller selection

The initial test series of cases had been analysed using NanoFilt and Nanopolish for filtering and variant calling. This was done using default parameters. However, a number of variant callers were identified which may have been more suitable for use in this experiment. These variant callers were compared and this data was used to identify the most suitable filtering and variant calling parameters. This was initially done using control Case 1343 (LORD). The long-read sequencing data obtained using the redesigned primer set was compared to a previous WGS dataset which was used as a truth set for this case. During this analysis using Medaka and Nanopolish, it was discovered that this case had relatively few heterozygous variants in *ABCA4* (N=17). A case with a greater number of heterozygous variants was then sought to allow concurrent investigation of whether the variant callers would detect sufficient variation to allow for phasing. Due to this, the pipeline was then optimized using the NA12878 control, with the Platinum dataset from the Genome in a Bottle (GIAB) consortium as a truth set (Section 2.7.2). This is a well characterised dataset which has been widely used to benchmark sequencing strategies and variant callers (Zook et al., 2016b). Nanopolish, Medaka, and Longshot were all compared (Sections 2.7.2).

Initially, read filtering using NanoFilt was optimized to exclude reads of unexpected length. A minimum read length of 4,500 bp was adopted. Although various maximum read lengths were tested, a maximum read length of 14,500 bases was decided upon. Filtering by read quality was also tested. Full results from each variant caller can be viewed in Appendix 6, the best result obtained by each variant caller is listed in Table 3.2. A minimum read quality of Q10 was identified as the most suitable for this experiment. For Nanopolish, the base fraction (the proportion of reads containing the non-reference base) was varied. For Medaka, the SNPs filter argument was varied and the results compared. Longshot was then compared with default parameters. For all cases, the number of true positives (TN) (variants called in both the truth set and by the variant caller) were compared to the number of false negatives (FN) (variants present in the truth set but not called by the variant caller) and the false positives (FP) (variants called by the variant caller which were not in the truth set). Precision was calculated by: (TP/TP+FP) × 100 and Sensitivity was calculated by (TP/TP+FN) × 100.

This analysis indicated that Longshot was less sensitive than Medaka and Nanopolish (using the selected parameters), and further refinement was not performed. For all variant callers, comparison between calling indels and excluding indels showed that calling indels decreased precision and sensitivity.

Table 3.2 Precision and sensitivity of all tested nanopore base callers.

Three variant callers were tested, Medaka, Longshot and Nanopolish (highlighted in red text). For each, the filter at which variants are called was varied. Shown are the best results for each of the variant callers. Medaka was tested using Case 1343, and an in-house WGS VCF was used as a truth set. Longshot and Nanopolish were tested using the NA12878 control. This was compared to the GIAB Platinum VCF. For comparison with Nanopolish with indels not called, indels were removed from the GIAB Platinum VCF (GIAB VCF#). Nanopolish used to call only SNPs was shown to be the best variant caller for this experiment (highlighted in green). A number of variants were found to be miscalled indels, this is adjusted in the calculations of precision and sensitivity (*). TP, True positives, FN, False negatives, FP, False positives

Medaka								
Truth set	Case	Custom parameters		Observed			Results (%)	
		SNPs filter	Indels filter	TP	FN	FP	Sensitivity	Precision
1343 WGS	Case A. LORD	12	15	98	34	3	74.24	97.03
1343 WGS	Case A. LORD	Default	Default	111	20	11	84.73	90.98
Longshot								
Truth set	Case	Custom parameters		Observed			Results (%)	
		Parameters	Indels?	TP	FN	FP	Sensitivity	Precision
GIAB VCF	NA12878	Default	Yes	117	115	5	50.43	95.90
NanoPolish								
Truth set	Case	Custom parameters		Observed			Results (%)	
		Base Fraction	Indels?	TP	FN	FP	Sensitivity	Precision
GIAB VCF	NA12878	0.33	Yes	195	57	12	77.38	94.20
GIAB VCF	NA12878	0.2	Yes	231	19	63	92.40	78.57
GIAB VCF#	NA12878	0.2	No	219	33	28	86.90	88.66
GIAB VCF#	NA12878	0.33	No	204	20	7*	91.07	96.68

The variants called by each variant caller were compared to the variants called by a previously curated truth set of variants. This was initially done with Case 1343 (LORD) however this was deemed to be unsuitable for use as a

benchmark due to a lack of heterozygous variants at this locus, as such Case NA12878 was then used to compare the variant callers. The results showed that by comparison of NA12878 with the GIAB Platinum VCF, Nanopolish was the optimum variant caller. Using Nanopolish with a base fraction (the proportion of reads with the non-reference base) of 0.33 and excluding indels gave the highest precision (96.68%) and sensitivity (91.07%) (Figure 3.5). As detailed in Table 3.2, there were 231 total variants called by Nanopolish at these parameters and the GIAB dataset (highlighted in green). This consisted of 204 true positives called by both platforms, 20 variants called by the GIAB dataset only and seven variants only called by Nanopolish. However, closer examination using IGV of these seven variants revealed that four of them were actually in both datasets. As Nanopolish was directed to only call single nucleotide variants (SNPs), the GIAB dataset was filtered to exclude indels. However, this resulted in four variants which were actually indels being incorrectly called as SNPs. For example, in the GIAB dataset, g.94478782 CGC>TA was correctly called. However, for comparison with Nanopolish calling only SNVs, indels were filtered out of this dataset. Nanopolish then called this variant as G>T at g.94478783 and C>A at g.94478784. With these variants removed, these parameters gave a sensitivity of 91.07% and precision of 98.7%.
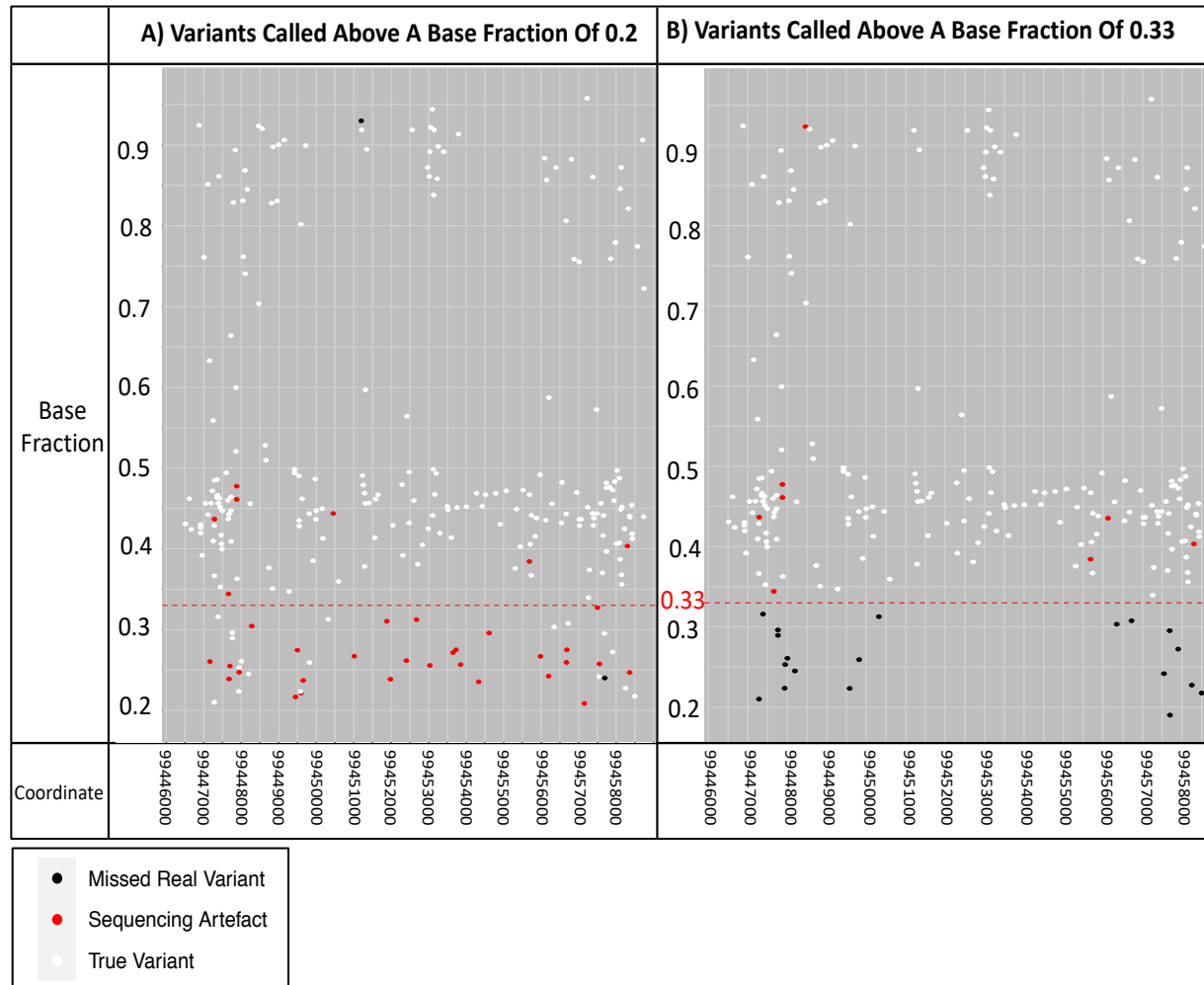
Figure 3.5 Concordance between variants called by Nanopolish from long-range PCR data and Genome in a Bottle benchmark for control NA12878

A) Variants called above a base fraction of 0.2, indicating a large number of false positive calls (red) compared to true positive calls (white). This indicated that a cut-off of 0.33 would yield the most true positives while reducing false negatives. B) Variants called above a base fraction of 0.33, showing very few (N=5) false positive calls. While there was an increased number of false negative calls (black), it was determined by comparison with the phased NA12878 truth set that this would not affect the accuracy of the constructed phase blocks.

Examination of the variation in the overlaps in isolation for Case NA12878 revealed that with these parameters, all bar one overlap (ES20-S19) would have sufficient variation to allow for phasing across the locus (Figure 3.6). In contrast, control Case 1343 was found to have seven overlaps with insufficient variation to phase with these parameters. It was determined that due to the relatively stringent filtering of these parameters, any overlaps with no heterozygous variants would be manually examined for potential missed variants. Phasing of Case NA12878 using WhatsHap confirmed that *ABCA4* could be phased into two distinct phase blocks due to a lack of variation in the overlap between ES20 and S19 (Figure 3.7). From this, it was determined that in a real-world case, sufficient variation was present to allow phasing of the majority (94.9 kb, 73%) of *ABCA4* and that the assay and bioinformatics pipeline were both sufficiently precise and sensitive to allow for scaling up to a series of ten cases.

Figure 3.6 Variants called in the overlap between long-range PCR amplicons spanning *ABCA4* in control NA12878

The variants called by Nanopolish were compared to the variants in the Genome in a Bottle truth set in the overlaps between the long-range amplicons spanning *ABCA4*. The variants in both datasets (True Variants) were compared to the variants only present in the nanopore dataset (Sequencing Artefact) and variants only present in the Genome in a bottle dataset (Missed Real Variant).

Figure 3.7 Phased long-read alignments spanning *ABCA4* for control NA12878 using updated long-range primer set

The created phase blocks are indicated in the top panel above the cumulative coverage. The next three panels show the aligned long reads, separated into Haplotype 1 (orange), Haplotype 2 (purple) and unphased reads (grey). A lack of variation in the overlap between S19 and E_S20_1_F (highlighted in red) has necessitated the creation of two phase blocks.

3.2.3   Chimeric read formation during long-range PCR

As the assay was deemed to be working robustly when sufficient variation was present, a series of eight cases (Cases 3514, 3856, 4682, 5755, 3947, 4672, 4793, and 5601) with previously identified biallelic *ABCA4* variants which could not be phased or where only one *ABCA4* variant allele had been identified were selected (Table 3.3). These cases were amplified using the updated primer set and a number of cases were sequenced (Probands 3515, 3856, 4682, and 5755). Initial inspection of these cases as they were sequenced showed that the approach was successful when sufficient variation was present and WhatsHap was successfully able to phase large sections of *ABCA4* in these cases. However, closer examination of these results using IGV revealed that four haplotypes were observed for a high proportion of the observed reads in each case (Figure 3.8). This was deemed to be a major confounding issue and the sequencing of the remaining cases was postponed while this was investigated using the NA12878 control. As WhatsHap did not identify this, this was done manually.

Table 3.3 Details of *ABCA4* cases sequenced with updated long-range primer set at 35 cycles of amplification

All PCRs were performed at 35 cycles of amplification using the updated set of primers. Filtering was performed using NanoFilt with quality score Q10, excluding reads greater than 14,500 bases and shorter than 4,500 bases in length. The interval refers to the distance between the indicated variants. Where multiple variants are present in a single allele, the interval refers to the distance between the most distal variants. Where no variants or a single variant are shown, the interval refers to the interval covered by the assay. Variants were called by Nanopolish at a base fraction of 0.33.

Local sample ID	Length and quality filtered read count	Reads on target	Proportion of reads on target (%)	Mean combined read depth across target locus (×)	Target variants			Number of target amplicons	Mean number of variants per overlap	Number of overlaps without variants
					Allele 1	Allele 2	Interval (bp)			
NA12878	89280	84266	94.38	4807.70	N/A	N/A	132260	24	5.14	1
1343 (LORD)	51260	50166	97.87	2859.08	N/A	N/A	N/A	24	2.50	7
3515	68309	57731	84.51	3201.80	c.1622T>C	c.1903C>A	635	24	4.9	3
3856	15284	11510	75.31	644.06	c.2160+1G>C	c.5316G>A	45849	24	3.74	3
4682	9776	7935	81.17	442.19	c.4556G>A	c.6320C>T	23964	24	5.14	0
5755	45023	40807	90.64	2251.89	c.[5461-10T>C; 5603A>T]	c.5603A>T	484	24	3.05	3
3947	Sequencing Not Performed	N/A	N/A	N/A	c.5585-1G>A	c.6079G>A	5421	N/A	N/A	N/A
4672		N/A	N/A	N/A	c.161C>T	c.2617A>T	59910	N/A	N/A	N/A
4793		N/A	N/A	N/A	c.[2588G>C; 5603A>T]	c.1906C>T	51697	N/A	N/A	N/A
5601		N/A	N/A	N/A	c.[5461-10T>C; 5603A>T]	c.6079C>T	5886	N/A	N/A	N/A

Figure 3.8 Chimeric read observation during sequencing of long-range PCR products.

A) Expected result from phasing by long-read sequencing. The two variants are in *trans* and segregate into two distinct haplotypes. B) Observed result from phasing using long-range PCR. The two heterozygous variants segregate into four distinct haplotypes (every possible allelic combination), preventing phasing.

Examination of the proportion of chimeric read formation was performed by manually separating reads based on which nucleotide was present in the read at a given coordinate as detailed in Section 2.7.2. This was done for each amplicon with a heterozygous variant at the 5′ end of the amplicon. By then measuring the proportion of reads represented by each nucleotide at the opposite end of the amplicon, the proportion of chimeric reads could be estimated. As such, all the positions across this amplicon should be concordant with one-another, and any discordance could be interpreted as reads which had formed a chimera with the opposite allele (Figure 3.9). Case NA12878 was selected to investigate chimeric read formation in detail.

A literature search was then undertaken to investigate if this was a previously encountered issue during long-range PCR. Recombination between PCR products resulting in chimeric molecules has been previously observed (Meyerhans et al., 1990; Cronn et al., 2002). A model of chimera formation has been proposed, when PCR is performed, if a product is not fully formed it can act as a primer for another allele, producing a chimeric fragment (Meyerhans et al., 1990; Potapov and Ong, 2017). A further search revealed that chimeric read formation has been cited as an issue when phasing using long-range PCR and long-read sequencing, however it was observed during this experiment that reducing the number of PCR cycles reduced chimeric read formation (Laver et al., 2016).

To assess the effects of PCR cycle count on chimeric read formation at this locus, control NA12878 was amplified again and the number of rounds of thermocycling was reduced from 35 × to 25 ×. Then the base fraction of the most distal allele for each amplification product was measured. The decrease in cycle count reduced the effects of chimerism as reported in Table 3.4 and illustrated in Figures 3.9 and 3.10. The mean proportion of reads which were observed as chimeric reduced from 26% at 35 cycles to 7% at 25 cycles. The maximum proportion of chimeric reads was 42% at 35 cycles and 13% at 25 cycles. However, this coincided with a drop in the molarity of available PCR products, from ~9 nM to ~1.3 nM per amplicon (Figure 3.11). This greatly reduced the robustness of previously optimised PCRs and necessitated multiple reactions to ensure sufficient product was obtained.
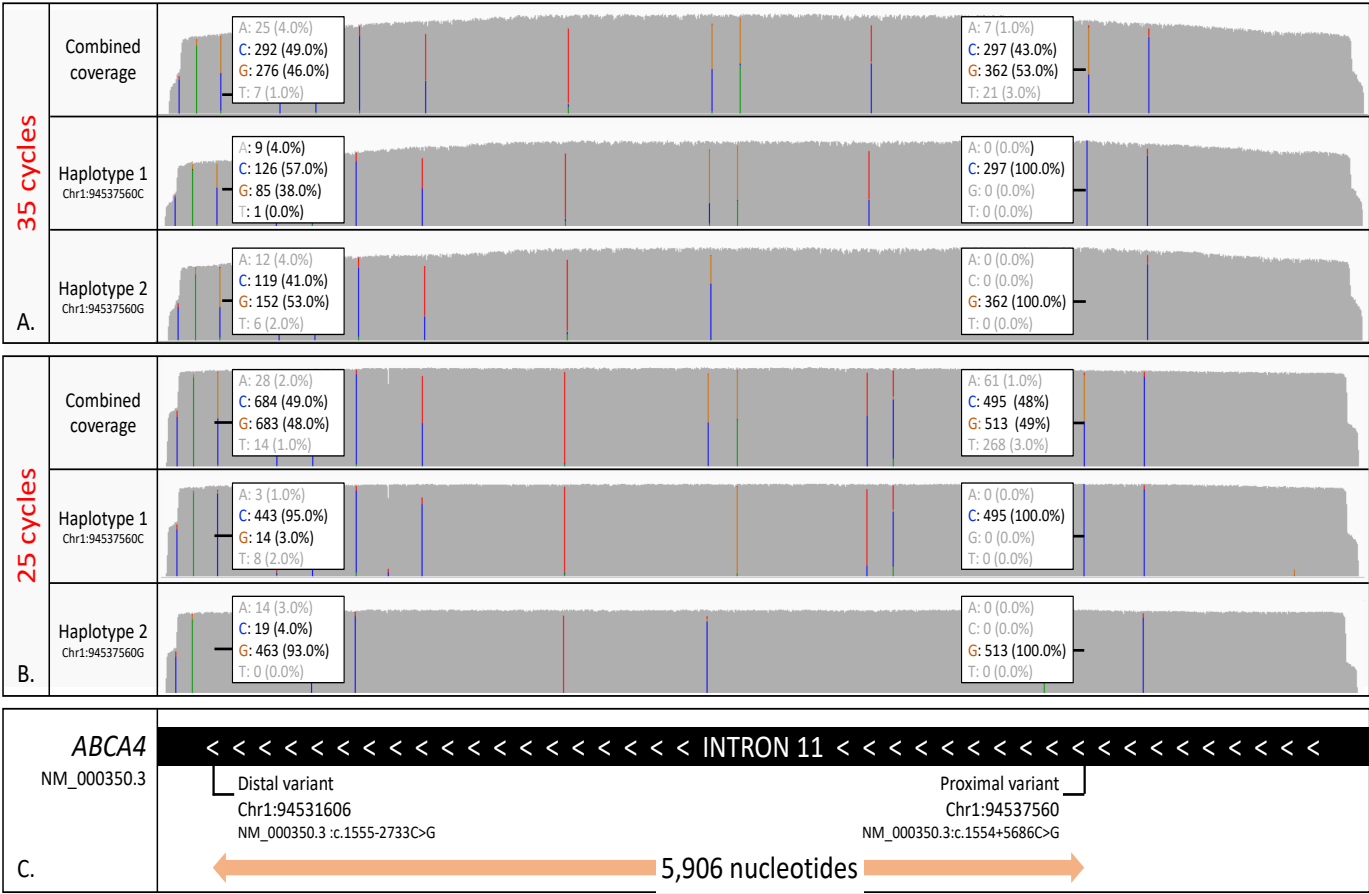
Figure 3.9 Measuring the formation of chimeric reads from long-range amplification products.

It was possible to measure the effect of chimeric read formation by isolating the reads with the most proximal reference or non-reference nucleotide. By then going to the most distal variant, the proportion chimeric reads could be measured by recording the proportion of reads with each nucleotide at the most distal variant. **A)** At 35 cycles, a high proportion of chimeric reads were observed, preventing phasing. **B)** At 25 cycles chimeric read formation was reduced allowing accurate phasing.

Table 3.4 Comparison of chimeric read formation in control NA12878 at 25 cycles and 35 cycles of amplification

High levels of chimeric reads were observed at 35 cycles of PCR. Reducing cycle count to 25 led to a sharp reduction in chimeric read formation but a corresponding drop in mass of product was also observed. All coordinates refer to chromosome 1 on genome build hg19. There were no distally located variants found in S8.2. "Too low" refers to when it was not possible to ascertain the DNA concentration using a Qubit Fluorometer due to very low DNA concentration.

Amplicon	Amplicon coordinates (start-stop)	Amplicon size (bp)	Proximal variant	Distal variant	Proportion of chimeric reads			Yield of amplification product	
					35× (%)	25× (%)	Reduction (%)	35× (nM)	25× (nM)
S1	94581079-94587651	6,572	94587362G>T	94581258T>A	31	2	29	47.41	0.87
S2	94576013-94583016	7,003	94582249T>G	94576360A>G	19	5	14	10.27	1.70
S3	94571091-94577994	6,903	94577423C>T	94572434C>G	17	12	5	13.24	1.55
S4	94566272-94573027	6,755	94572434C>T	94567223T>C	16	3	13	11.69	1.96
L4	94562394-94570893	8,499	94570625G>A	94562924C>T	31	9	22	8.11	0.57
L5	94556161-94564368	8,207	94562924C>T	94556894A>G	27	8	19	9.94	0.85
S7	94550053-94558218	8,165	94557434T>C	94550715G>A	24	7	17	10.62	0.38
S8	94544899-94552183	7,284	94550715G>A	94545160T>C	42	11	31	3.31	1.11
S8.2	94540974-94547825	6,851						25.74	2.60
L8.2	94537456-94546134	8,678	94545160C>T	94537560G>C	28	8	20	8.63	4.36
S10	94534599-94541790	7,191	94540069G>C	94534980G>C	24	1	23	24.53	2.21
L9	94531257-94539449	8,192	94537560G>C	94531606G>C	25	3	22	6.28	0.41
L10	94524472-94533111	8,639	94532013C>T	94524784A>G	19	4	15	11.50	0.20
L11	94517468-94526499	9,031	94526044A>G	94520451G>T	15	13	2	2.79	Too low
E_L12_2	94510646-94522068	11,422	94520451G>T	94512360G>A	23	11	12	3.77	5.15
E_L13	94501620-94512558	10,938	94512360G>A	94504545C>T	29	10	19	16.72	4.08
E_L14	94496875-94508372	11,497	94505971G>A	94498133G>A	33	12	21	5.70	Too low
S18	94494792-94501639	6,847	94501594A>C	94495487G>A	28	8	20	16.85	4.66
S19	94489852-94496894	7,042	94496253G>A	94492773C>T	19	4	15	30.10	0.50

E_S20	94482132-94491845	9,713	94488326T>A	94486355A>G	12	1	11	12.96	4.86
E_L17.2	94477294-94488723	11,429	94488326T>A	94480037C>T	40	7	33	4.29	7.25
L18	94470930-94479249	8,319	94478847G>A	94471154C>T	31	10	21	9.72	2.76
L19	94464973-94472982	8,009	94472909G>A	94465132G>T	37	11	26	59.19	1.80
L20	94457435-94466789	9,354	94466659A>G	94464553C>T	26	7	19	3.21	0.18
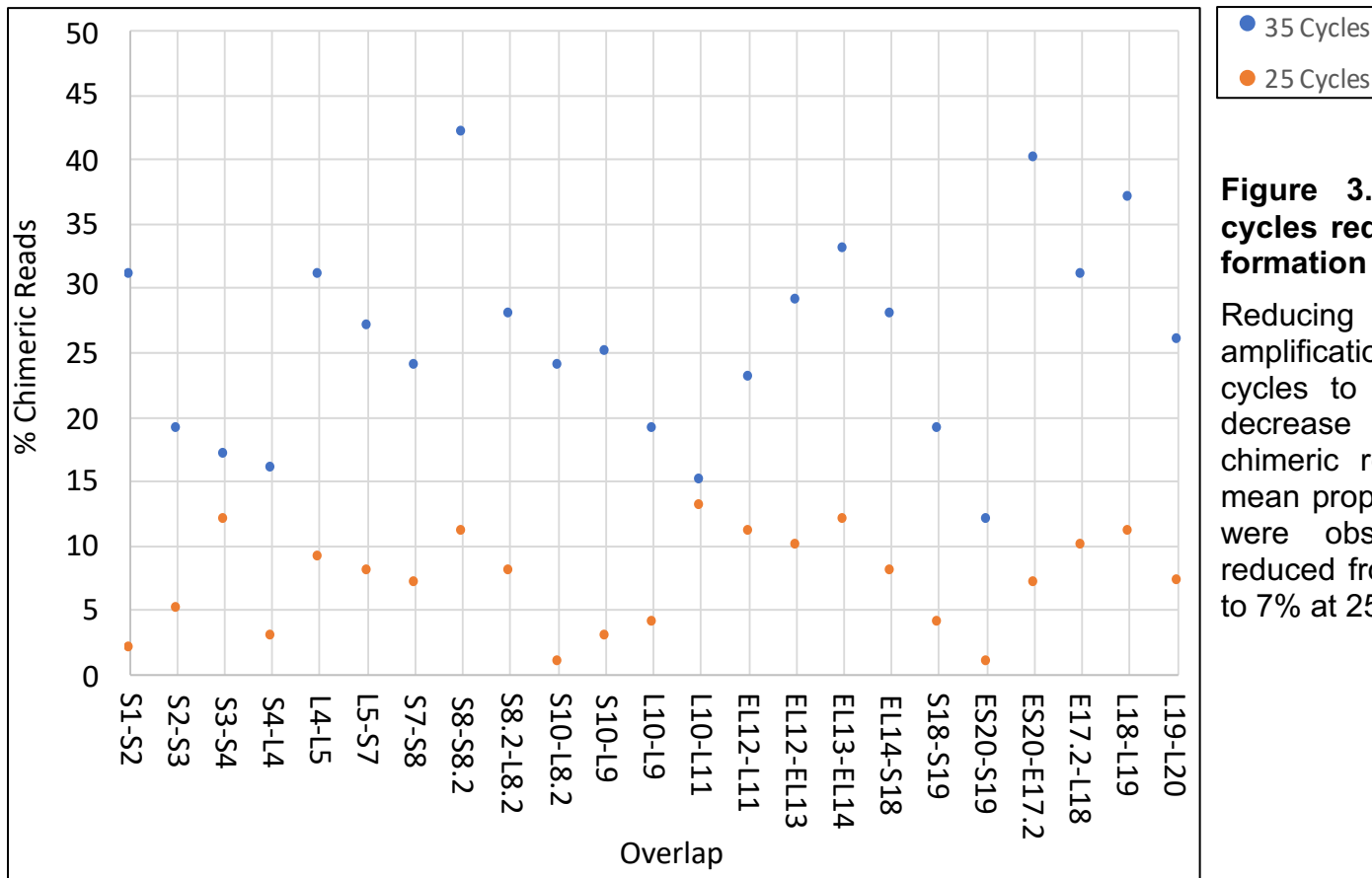
Figure 3.10 Reducing PCR cycles reduced chimeric read formation at *ABCA4*

Reducing the number of amplification cycles from 35 cycles to 25 cycles led to a decrease in the proportion of chimeric reads observed. The mean proportion of reads which were observed as chimeric reduced from 26% at 35 cycles to 7% at 25 cycles.
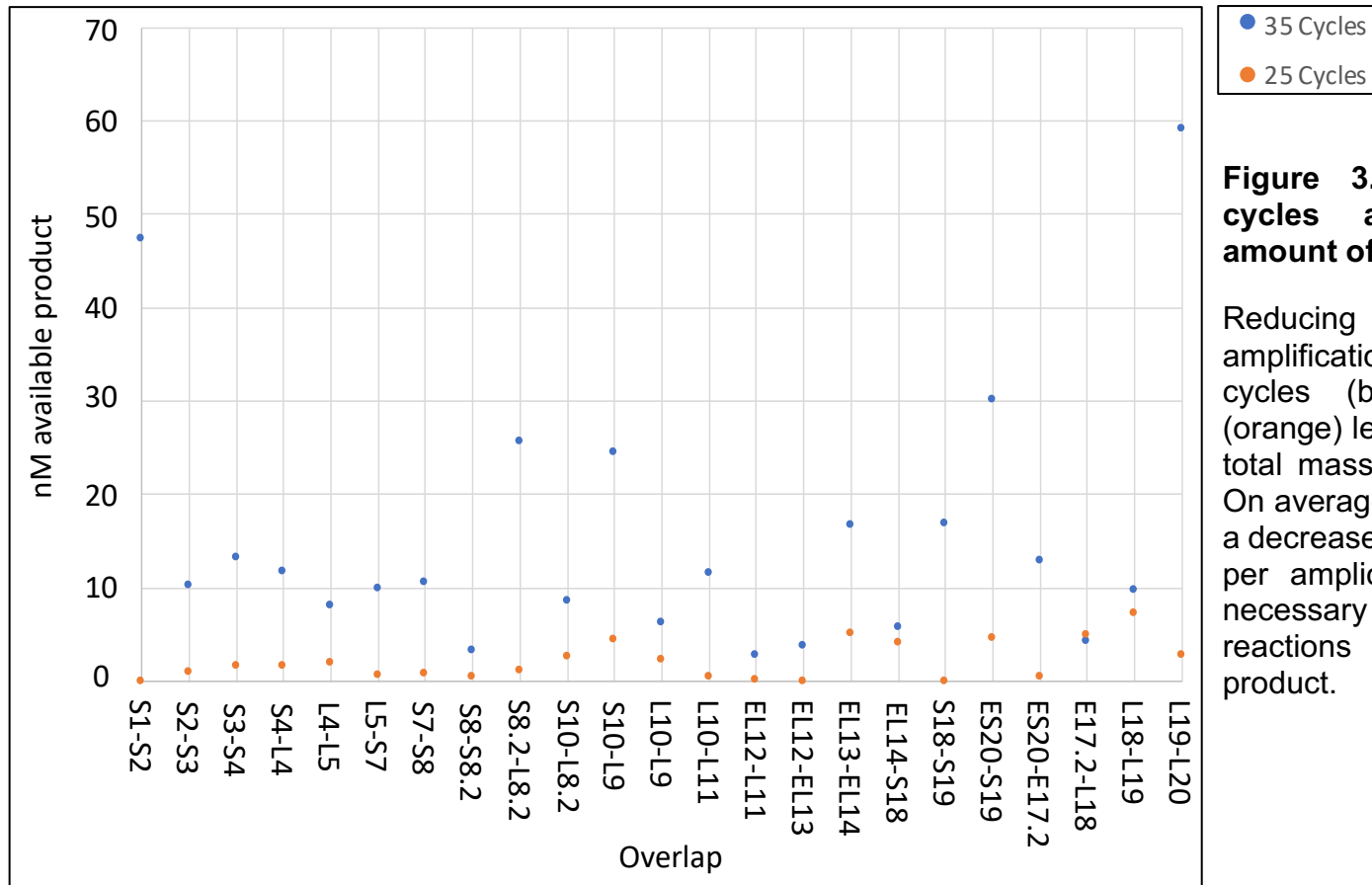
Figure 3.11 Reducing PCR cycles also reduced the amount of available product.

Reducing the number of amplification cycles from 35 cycles (blue) to 25 cycles (orange) led to a decrease in the total mass of the PCR product. On average this corresponded to a decrease from ~9nM to ~1.3nM per amplicon, as such it was necessary to pool multiple reactions to obtain sufficient product.

3.2.4 Concordance of long-range PCR phase blocks compared to Genome in a Bottle dataset

To verify the accuracy of the constructed phase blocks, the experimentally derived phased sequence for Case NA12878 (at 25 cycles of amplification) was compared with a Platinum phased dataset from the GIAB consortium as detailed in Section 2.7.2. This benchmark dataset is generated from a number of long-read and short-read sequencing technologies as well as segregation analysis using 11 family members and as such was considered a truth set to compare the experimentally derived data. Phasing this case by long-range PCR and WhatsHap had shown that the assay would result in two phase blocks due a lack of variation in one of the overlaps (Figure 3.7). The GIAB dataset confirmed that at 25 cycles the *ABCA4* locus could be resolved using the updated primer set into a minimum of two phase blocks due to a lack of variants as previously shown by phasing using WhatsHap. Within the constructed phase blocks, all experimentally derived phase blocks were concordant with those present in the phased dataset (Figure 3.12) (Appendix 5). This demonstrated that at 25 cycles, variant calling using Nanopolish at a base fraction of 0.33 and phasing using WhatsHap was sufficiently accurate to allow the creation of a concordant phase block from multiple PCR products.

To investigate whether high levels of chimerism would have significantly affected phasing using WhatHap, this analysis was repeated on the dataset generated using 35 cycles (for Case NA12878). Using 35 cycles, there was a switch in haplotypes compared to the Platinum dataset, resulting in 25.3 kb of the 97.9 kb phase block being called incorrectly. The switch in haplotype occurred at amplicon S8. At 35 cycles of amplification, 42% of the reads originating from this amplicon were chimeric. Therefore, it was concluded that at 35 cycles, sufficient chimeric reads were formed to confound phasing.

From this analysis, it was concluded that phasing by tiled PCR was accurate across long distances but only when 25 PCR cycles were performed so chimeric read formation was reduced sufficiently to no longer confound the formation of accurate phase blocks. As such it was decided that despite the reduction in available PCR product, it was necessary to perform future PCRs at 25 cycles to ensure accurate phasing. Other strategies to reduce chimeric read formation could have included the use of alternative polymerases or altering the cycle conditions, however this was not explored.
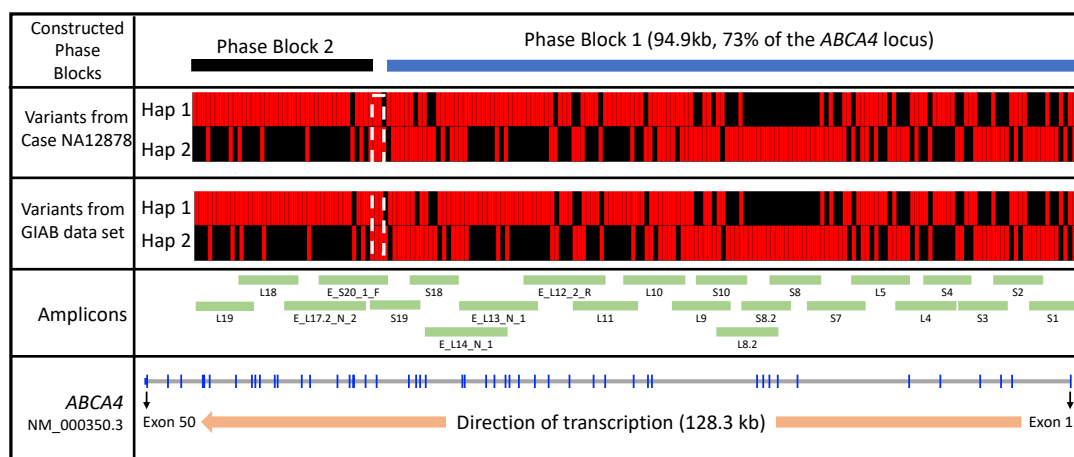
Figure 3.12 Comparison of haplotypes of case NA12878 from long-range PCR data at 25 cycles and GIAB dataset

> In the top panel, the phase blocks constructed by WhatsHap are shown. In the second panel, the variants across *ABCA4* are shown as called from the nanopore sequencing data. The black bars represent the reference allele and the red bars represent the non-reference allele. Haplotype 1 is shown on the top and haplotype 2 is shown on the bottom. In the third panel, the variants from the GIAB Platinum dataset are shown. False positive and false negative variants are not shown.

3.2.5 Implementation of long-range PCR method to phase *ABCA4* on a series of five cases

The assay was next applied to a series of five cases (Probands 3515, 5863, 5607, 3636 and 1337) with clinically relevant but unphased variants identified through smMIPs sequencing (Chapter 5) (Table 3.5). When sequencing the NA12878 case using 25 × cycles of amplification there was a considerable reduction in the available mass of amplification products. As such, PCRs were less robustly amplified and frequently multiple reactions had to be pooled to obtain sufficient sample. Because of this, a targeted approach was adopted to target only the relevant amplicons as opposed to the entire *ABCA4* locus as had been done previously. To validate this, an initial test case (3515) whose known pathogenic variants *ABCA4:*c.1622T>C, p.(L514P) (ACMG Pathogenic) and c.1903C>A, p.(Q635K) (ACMG Pathogenic), located within a single long-range PCR amplicon (amplicon L10, 9 kb) were sequenced. The reads were grouped based on the reference and non-reference nucleotide at position c.1622 (chr1:94528806, hg19). From this, the variants were confirmed to be in *trans* and the proportion of chimeric reads that were identified was low (17%) (Figure 3.13).

Table 3.5 Details of *ABCA4* cases sequenced with updated long-range primer set at 25 cycles of amplification

All PCRs were performed at 25 cycles of amplification using the updated primer set. Filtering was performed using NanoFilt with Q10, excluding reads above 14,500 bases and smaller than 4,500 bases. The interval refers to the distance between the indicated variants. Where no variants are shown, the interval refers to the distance covered by the assay. Variants were called by Nanopolish at a base fraction of 0.33. N/A indicates the variants were covered by a single amplicon and there were no overlaps as a result.

Local sample ID	Length and quality filtered read count	Reads on target	Proportion of reads on target (%)	Mean combined read depth across target locus (×)	Target variants			Number of target amplicons	Mean number of variants per overlap	Number of overlaps without variants
					Allele 1	Allele 2	Interval (bp)			
NA12878	64628	63116	97.66	3556.96	N/A	N/A	132,260	24	5.14	1
3515	9064	120	80.19	108.46	c.1622T>C	c.1903C>A	635	1	N/A	N/A
5863	18386	7081	38.51	1294.84	(c.2588G>C, c.5603A>T)	c.4537del	40,787	8	4.1	1
5607	7437	2767	37.21	547.86	c.3259G>A	c.6089G>A	37,331	7	6.0	1
3536	7036	1421	20.23	608.03	c.1906C>T	c.3113C>T	19,195	4	4.0	2
1337	39608	26012	65.67	11114.77	c.5603A>T, c.5819T>C	c.6817-2A>C	17,667	3	4.3	1
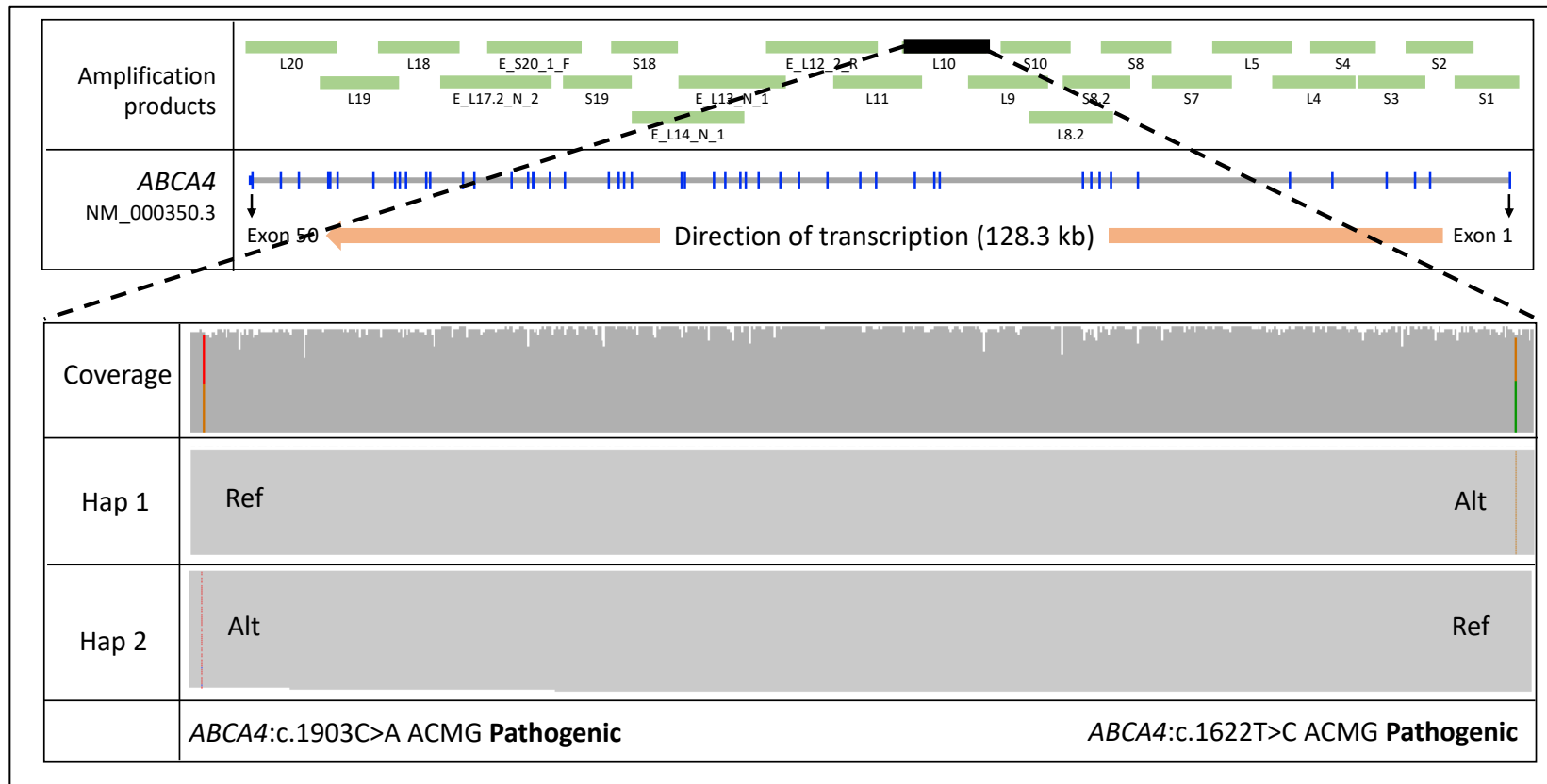
Figure 3.13 Phasing of pathogenic *ABCA4* variants in Case 3515 at 25 cycles of amplification

A case was selected whose pathogenic variants were contained in a single long-range PCR amplicon. This demonstrated that at 25 cycles in a real-world case phasing of the variants was successful with little chimeric read formation (chimeric reads not shown). This was concordant with data generated using WhatsHap.

Having established the assay was accurate at 25 cycles of amplification, probands 5863, 5607, 3636 and 1337 were then amplified and sequenced. These cases were also amplified using a targeted approach which required between 3 and 8 amplification products per case. All of the target variants were identifiable, however for each case at least one (maximum two) overlap did not contain any heterozygous variants, preventing phasing of the target variants (Table 3.6). Chimeric read formation was measured at a sampling of amplicons across the five probands. At all examined amplicons, chimeric reads were observed at a low proportion of total reads.

At this point it was clear that while it was possible to phase large sections of *ABCA4* using long-range PCR target enrichment combined with nanopore sequencing, a lack of variation prevented robust phasing across multiple individuals. This approach had been designed to target relatively short fragments as the assay was intended to be used on screening of samples which had been repeatedly freeze thawed and therefore may have a short fragment profile. However, this necessitated a tiled approach which led to issues with phasing using variation. While the primers had been designed using population level variation and tested *in silico* on a series of cases, it is difficult to predict an individual's variant profile prior to sequencing. A single set of primers is evidently not suitable for screening unrelated individuals; however this method may be more suitable for use on cases where there is short-read sequence information available so that the primers can be designed to capture known variation in the overlaps.

Table 3.6 Phase blocks created by long-range PCR of amplicons spanning *ABCA4* variants of interest

As the PCRs were less efficient at 25 cycles compared to 35 cycles, only the amplicons of interest, (rather than the entire gene) were sequenced. The coloured cells show the created phase blocks for each case. A change in colour indicates a new phase block was created due to a lack of variation in the overlaps between the indicated amplicons. For each case, there was at least one overlap where there was not sufficient variation to allow phasing and these overlaps were not recurrent between any of the cases, showing that redesigned primers were unlikely to be robust moving forward. All coordinates refer to chromosome 1 (build hg19).

Amplicon Start	Amplicon End	Amplicon Name	Amplicon Size (bp)	3515	5863	5607	3636	1337
94524472	94533111	L10	8,640	blue			blue	
94517468	94526499	L11	9,032				blue	
94510646	94522068	Extend_L12_2	11,423		blue		black	
94501620	94512558	E_L13_N_1	10,939			blue	red	
94496875	94508372	E_L14_N_1	11,498		blue	black		
94494792	94501639	S18	6,848		blue	black		
94489852	94496894	S19	7,043		blue	black		
94482132	94491845	E_S20_1	9,714		black	black		
94477294	94488723	E_L17.2_2	11,585		black	black		
94470930	94479249	L18	8,320		black			blue
94464973	94472982	L19	8,010					blue
94457435	94466789	L20	9335					black

3.2.6 Phasing *ABCA4* by amplification-free enrichment

3.2.6.1 HLS-CATCH

Given the outlined issues with phasing by tiled long-range PCR an amplification-free, ultra-long-read sequencing method was explored. In collaboration with Dr Gavin Arno (University College London, UCL) during a secondment, four experiments using Cas9 Assisted Targeting of Chromosome Segments (CATCH) were performed, using the Sage High

molecular weight DNA Library System (Sage-HLS) platform as described in Section 2.5.

During Sage-HLS CATCH, a cell suspension or previously isolated nuclei were lysed in an agarose plug, referred to below as a cassette. A CRISPR-Cas9 complex was then added which targeted loci upstream and downstream of the region of interest. This introduced double strand breaks at these loci. The fragmented DNA was then size fractionated by electrophoresis *in situ*. As DNA extraction and target enrichment were performed in the same cassette, very large loci could be targeted and ultra-long fragments isolated. The isolated target was then sequenced using a MinION.

In this experiment, two guides were used to target sites upstream and downstream of *ABCA4* respectively. The upstream guide cut at chr1:94121375 (hg38). The downstream guide cut at chr1:93976930 (hg38). This targeted a 132,260 kb locus. These guides had previously been designed and optimized by Dr Gavin Arno as part of an ongoing series of experiments targeting *ABCA4*.

3.2.6.2 HLS-CATCH on control case

An initial run was performed on a control sample (CATCH 1) using nuclei extracted from fresh blood (Section 2.6.3) and sequenced using a MinION flowcell. Details of all cases enriched using CATCH are reported in Table 3.7. This run generated a read depth up to 87× with seven reads spanning the *ABCA4* locus (Figure 3.14). This resulted in ultra-long reads with an N50 (the length of the shortest read for which longer and equal length read cover at least 50% of the assembly) of 55,100 bp at this locus.

Table 3.7 Summary run metrics of cases sequenced using HLS-CATCH enrichment

Three CATCH experiments were performed. Reads were filtered on a quality score of Q7 and a minimum read length of 5 kb. The interval refers to the distance between the variants of interest, or where there are no variants of interest, the distance between the two guides. Mean read depth calculated by aligned bases at target locus/size of target locus

Case	Target variants			Filtered read count	Reads on target	Reads on target (%)	Mean Read depth (×)	N50 at locus	Reads spanning *ABCA4*
	Allele 1	Allele 2	Interval (b)						
CATCH 1	N/A	N/A	132260	14335	268	1.87	66.38	55,100	7
CATCH 2	c.1335C>G	c.4256 T>C	48087	27044	44	0.16	12.63	95,142	4
CATCH 3	c.1938-289 G>T	c.5882 G>A	52797	11559	99	0.86	33.60	82,284	10
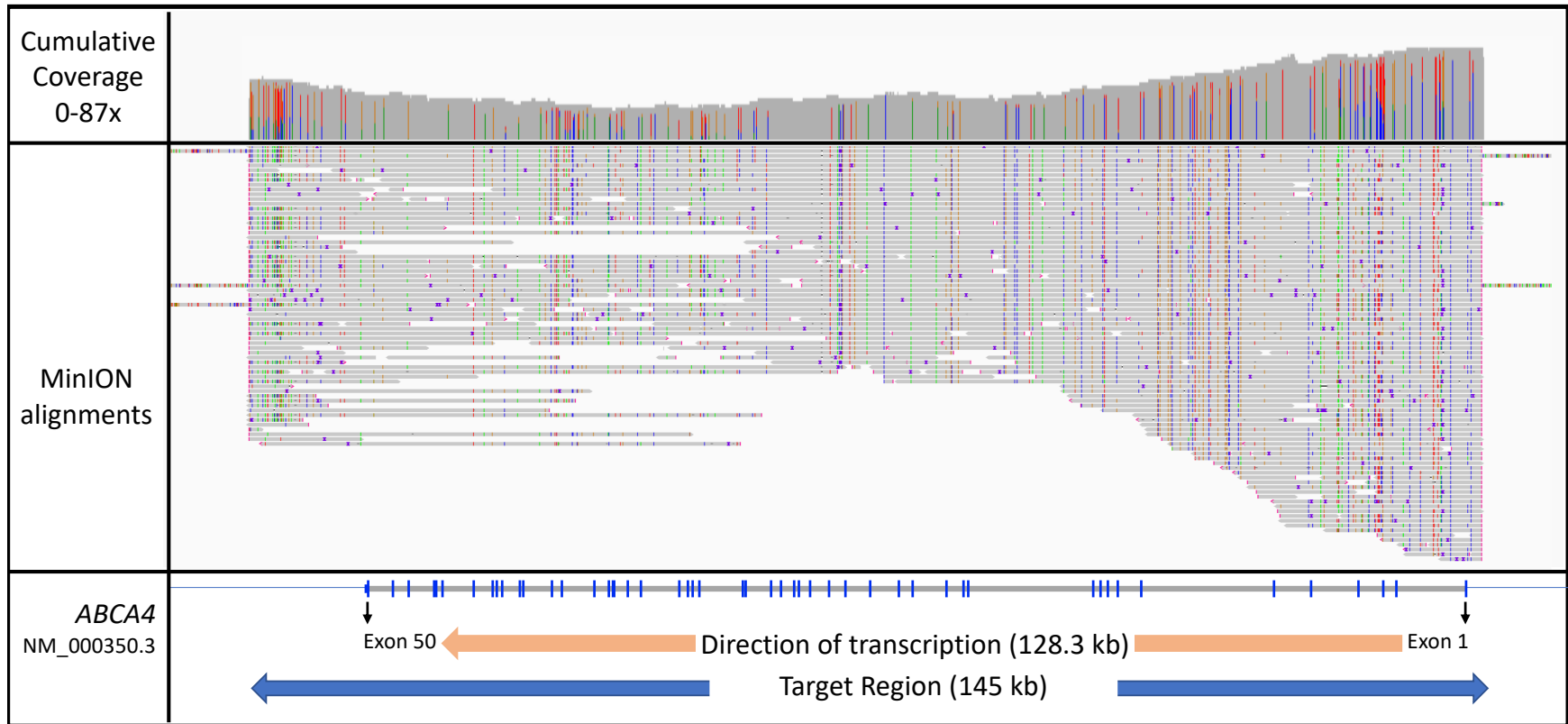
Figure 3.14 Sequencing of an *ABCA4* control case (CATCH 1) enriched by CATCH using the Sage-HLS platform.

Ultra-long-read alignment shows that this run was generated a maximum read depth of 87 reads, corresponding to a 653-fold enrichment of the human genome. Reads aligned against human reference genome build hg38.

The bioinformatics pipeline which had been developed for analysis of the long-range PCR datasets was adapted for use on these data. Reads shorter than 5 kb were removed and reads with a quality score of less than Q10 were filtered out. There was no upper limit on read length imposed. Due to an update in the MinION operating software MinKNOW, resulting in different compression of the output files, at the time of these experiments it was not possible to use Nanopolish to perform variant calling. As such an alternate variant caller, Clair3 was identified which was suitable for use with the updated compression system. Clair3 was internally validated by re-calling all the variants in Case NA12878 amplified at 25 cycles and comparing to the GIAB dataset (Section 3.2). This showed that Clair3 was slightly more accurate and sensitive than Nanopolish when calling SNPs only (97.37% sensitivity, 96.94% precision) (Table 3.8).

Table 3.8 Comparison of variant calling using Nanopolish and Clair3.

Clair3 and Nanopolish were compared using Case NA12878 amplified using 25 cycles. As before, both variant callers were compared to the Genome in a bottle Platinum dataset (GIAB VCF). This comparison was also made with the Genome in a Bottle dataset with no indels (GIAB VCF#). TP=true positive, FN=false negative, F=false positive

Nanopolish							
Truth Set	**Custom Parameters**		**Observed**			**Results (%)**	
	Parameters	**Indels?**	**TP**	**FN**	**FP**	**Sensitivity**	**Precision**
GIAB VCF#	0.33	No	204	20	7*	91.07	96.68
Clair3							
Truth Set	**Custom Parameters**		**Observed**			**Results (%)**	
	Parameters	**Indels?**	**TP**	**FN**	**FP**	**Sensitivity**	**Precision**
GIAB VCF	Default	Yes	245	15	18	94.23	93.16
GIAB VCF#	Default	No	222	6	6*	97.37	96.94

This run resulted in ultra-long reads, there were seven reads spanning the whole *ABCA4* locus. It was possible to establish the phase of distal variants at low read depth using just these reads. The ultra-long reads were phased and segregated into haplotypes using heterozygous SNPs by WhatsHap to create larger phase blocks . As there were reads spanning the entire locus, it

was possible to segregate the majority of the reads as reported in Figure 3.15, thus increasing confidence in the phasing of distal variants.
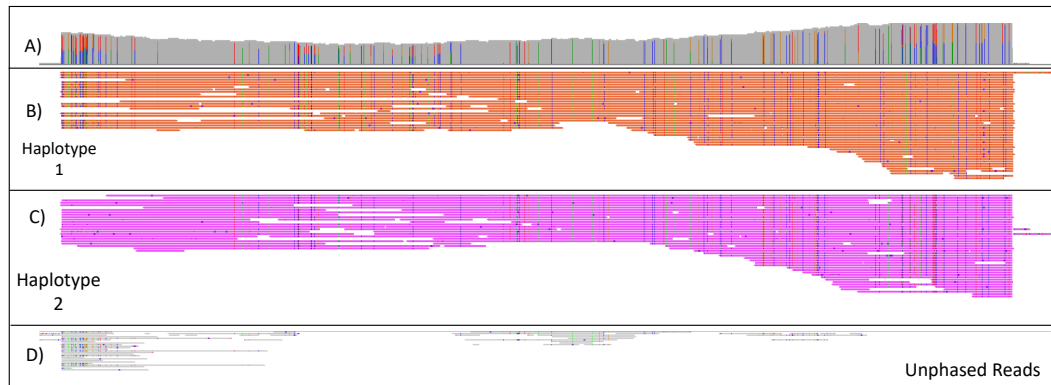
Figure 3.15 Phased reads spanning *ABCA4* from HLS-CATCH control case (CATCH 1) using WhatsHap

The cumulative coverage for this run is illustrated in panel A. This run achieved high coverage, with a maximum read depth of 87 × (hg38). WhatsHap was able to phase the entire *ABCA4* locus into a single phase block with two haplotypes (panels B and C) and a small number of unphased reads (panel D).

As the assay was quite expensive when performed using a MinION flowcell and only a proportion of the sequencing capacity of the flowcell is used, it was attempted to sequence excess product from the previous CATCH run on a Flongle flowcell. This yielded no usable data (data not shown).

3.2.6.3 Phasing variants of interest in *ABCA4* disease cases using HLS-CATCH

While the successful CATCH run had been performed on freshly prepared nuclei, the protocol provided by Sage Sciences provided details on storing nuclei at -80 °C in Bambanker; a freezing medium typically used for preservation of cells. As such, a number of samples had been collected and stored in Bambanker at -80 °C.

One of these was selected for sequencing as part of this experiment. This case (CATCH 2) had a diagnosis of group 1 Stargardt with foveal sparing and imaging revealed mild macular dystrophy. This case had undergone previous WES. Analysis by Dr Gavin Arno identified biallelic *ABCA4* variants,

c.1335C>G (ACMG: Pathogenic) and c.4256T>C (ACMG: Likely Benign). Previous reports identified c.4256T>C as a VUS (Cornelis et al., 2017) and it was sought to establish the phase of this variant to help determine its clinical relevance.

CATCH for this case was performed on previously extracted frozen nuclei. The nuclei were frozen in Bambanker and then resuspended in Nuclei Preparation Buffer (NPB). However, during resuspension, the nuclei aggregated; this affected sequencing yield. Later investigation revealed this to be a known issue with resuspension of frozen nuclei. A cumulative read depth of maximum 17 × (mean 12.63 ×) was nevertheless achieved (Figure 3.16). WhatsHap was used to phase *ABCA4* confirming the variants of interest were in *trans*.

Further investigation of c.4256T>C revealed that it was unlikely to be contributing to the disease in this case. Later reports had identified this variant as benign (Cornelis et al., 2022) and this variant was listed as benign or likely benign in seven entries in ClinVar (VCV000298239.12). As the entire *ABCA4* locus had been sequenced in this experiment, it was possible to search for variants which had not been found during initial sequencing. This identified an ultra-rare variant in exon 6 of *ABCA4* (chr1:94098819 A>G hg38, NM_000350.3, c.743T>C, p.(Val248Ala), ACMG: VUS)). This variant has been observed in only one other individual in gnomADv2 (chr1:94564375 A-G, hg19 allele frequency 0.000003977) and had not been observed in gnomADv3 (hg38). This variant has also not been listed in LOVD or ClinVar. *In silico* analysis of this variant revealed that it may be relevant in this case. This variant had a CADD-Phred score of 22.7 (top 1% of pathogenic variants), and was deemed Deleterious by MutationTaster (53|47 Deleterious|Benign). This variant was *in trans* with c.1335C>G (ACMG: Pathogenic), prioritising it for further investigation.
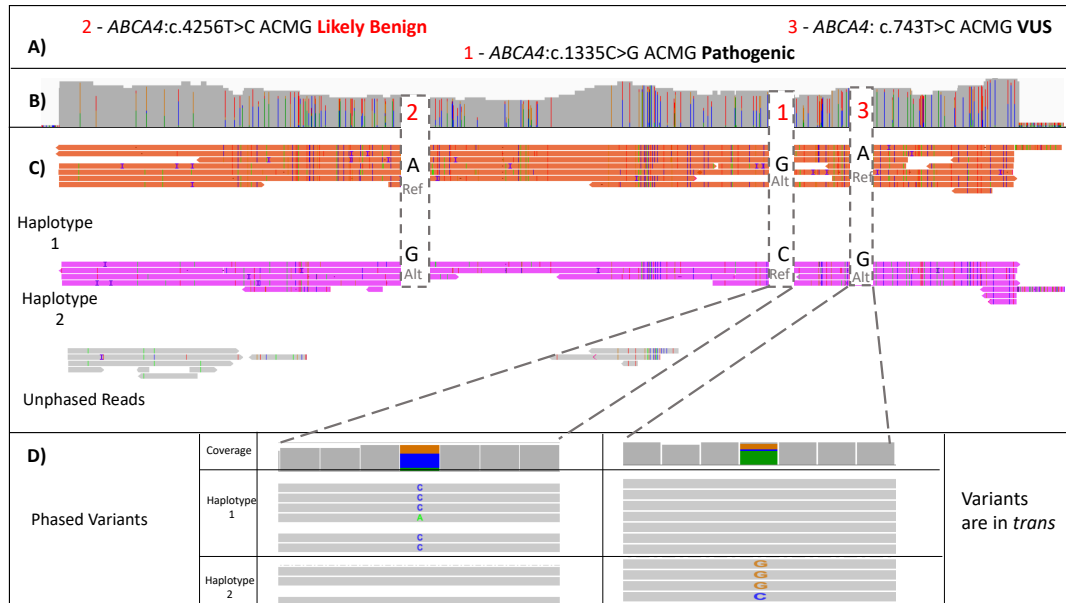
Figure 3.16 Phasing *ABCA4* variants in a patient case (CATCH 2) using HLS-CATCH

A) Analysis by Dr Gavin Arno of WGS data identified a single pathogenic variant (1) and a rare variant of interest that had been reported as Likely Benign (2). Later analysis revealed an additional variant of interest (3) which was reported as a VUS. B) Enrichment by CATCH resulted in a maximum read depth of 17 × and an enrichment of 97-fold compared to the rest of the genome. C) The resulting reads were phased using WhatsHap. D) Variants 1 and 3 are arranged in trans, prioritising variant 3 for further investigation.

The thawing of previously frozen nuclei was identified as a major issue in this and other experiments performed in parallel targeting the opsin array (data not shown). As such, numerous experiments were performed to optimise this process. Repeated nuclei extractions on control blood, attempts at freezing in Bambanker and suspending in Bambanker without freezing, showed that the nuclei irreversibly clumped upon initial suspension in Bambanker. It was therefore determined that the frozen nuclei present were unlikely to result in successful experiments regardless of how they were thawed.

Therefore, a fresh aliquot of blood was obtained from a patient who had previously had nuclei extracted and frozen, and a CATCH run on *ABCA4* was then performed on the fresh sample (CATCH 3). This case had group 1 STGD with only macular involvement at 45 years old. Sequencing in the 100,000 Genomes Project and analysis by Dr Gavin Arno had revealed a single pathogenic variant *ABCA4*:c.5882G>A and three deep intronic variants of

unknown significance (*ABCA4*:c.5196+415T>C, *ABCA4*:c.1938-289G>T, *ABCA4*:c.443-699G>A) which required phasing to establish their pathogenicity.

CATCH of this case resulted in good enrichment of the target locus. qPCR of the product from this experiment showed 68.8-fold enrichment of *ABCA4* compared to *RNAse P* in the most successful elution well (Figure 3.17).
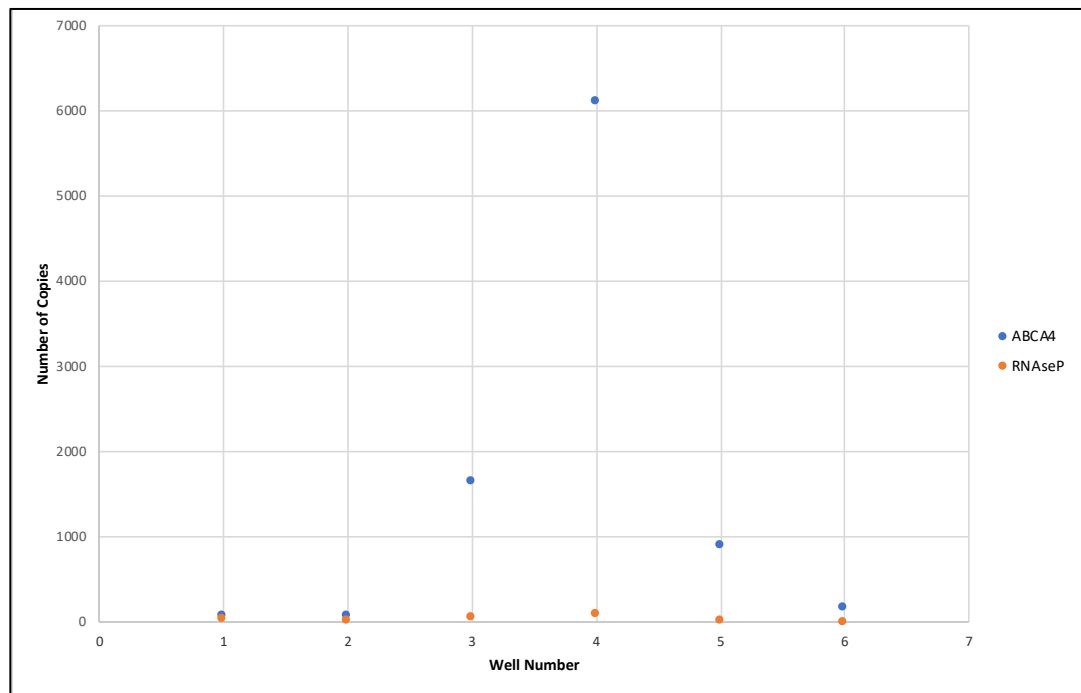
Figure 3.17 qPCR result of *ABCA4*-related disease case (CATCH 3) shows high enrichment for target gene

qPCR was used to identify which of the six elution wells in the HLS-CATCH cassette contained the target gene. This showed that *ABCA4* was enriched 68.8-fold compared to *RNAse P* in the most successful well.

As illustrated in Figure 3.18, sequencing of this case resulted in a read depth of 39 reads with a read length N50 at *ABCA4* of over 82 kb. This confirmed that the *ABCA4*:c.1938-289G>T variant was in *trans* and the other two variants were in *cis* with the variant of interest, prioritizing the variant in *trans* for further investigation.
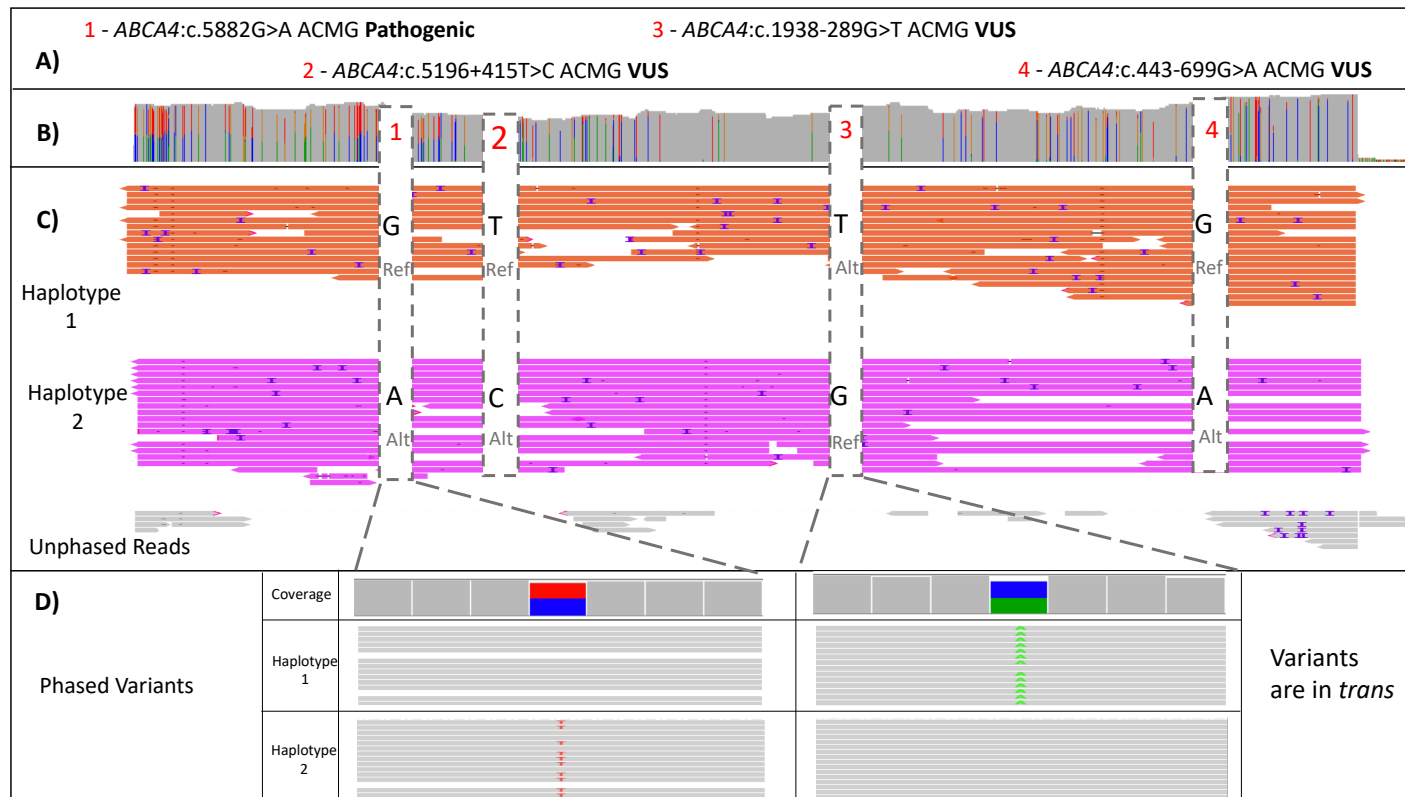
Figure 3.18 Phasing variants of interest in *ABCA4*-related disease case (CATCH 3) using HLS-CATCH

A) Analysis by Dr Gavin Arno uncovered a single pathogenic variant (1) and three novel deep intronic variants (2-4). B) Enrichment by CATCH resulted in a maximum read depth of 39× and an enrichment of 286-fold compared to the rest of the genome. C) The resulting reads were phased by WhatsHap. D) Only variant 3 was in *trans* with the pathogenic variant of interest, indicating this variant for further investigation.

A further issue which was identified during these experiments was the generation of excess amounts of foam in the HLS cassette (Figure 3.19). This disrupted the separation and elution of DNA fragments. On occasion, so much foam was generated that the current was disrupted (Figure 3.19), disrupting separation of target fragments. Additionally, the elution wells were flooded on several occasions, resulting in reduced product.

While it was possible to obtain high quality, ultra-long reads using this technique, practical issues such as foam generation and difficulties thawing frozen nuclei necessitated a number of repeated experiments. Further optimization is required to increase the practical utility of this method.
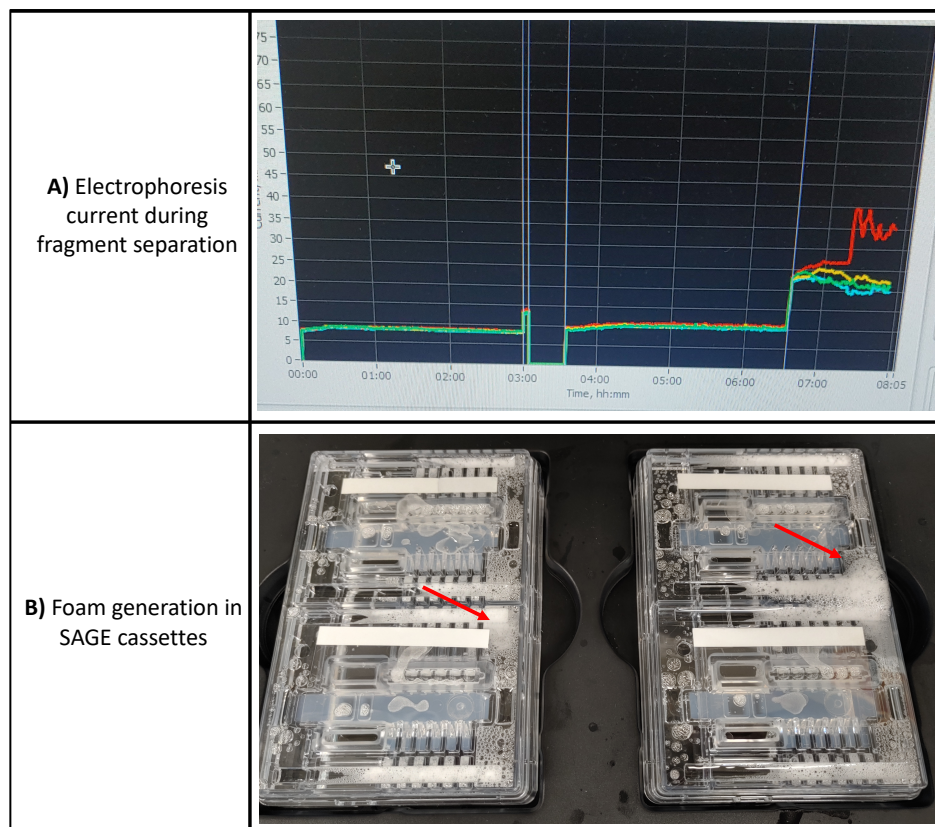
Figure 3.19 Foam generation during HLS-CATCH

> Panel A shows a photograph of the electrical current measurement during fragment separation. The current passing through each section of the cassette is indicated by a colored line. The current to the cassette indicated by the red line (top right in panel B) has been disrupted. Panel B shows a photograph of four Sage cassettes following separation. As can be seen, large amounts of foam have been generated as indicated by the red arrows.

3.3 Discussion

This research project was undertaken as part of a large consortium of 14 PhD projects across Europe established to study *ABCA4* in detail. The aim of this consortium was to identify missing heritability in *ABCA4*-related disease, further uncover the mechanism of disease and, develop novel treatments. The purpose of the experiments detailed in this chapter was to develop an assay to phase *ABCA4*.

In this experiment, two assays were used to phase large sections of *ABCA4*. The first of these, a long-range PCR assay, was designed for use on samples which had been repeatedly freeze-thawed and likely had reduced genomic DNA fragment size as a result. As such, a tiled approach was used to allow targeting of relatively low molecular weight DNA. An amplification-based approach was adopted as many older samples have a limited amount of DNA available for experiments. This method was used to develop and validate a bioinformatics pipeline for analysis of the generated nanopore reads. While this method allowed for the creation of accurate phase blocks (as validated with a phased truth dataset), when used on patient cases, a lack of variation prevented the formation of full-length phase blocks. In contrast, the second method, HLS-CATCH, used DNA extracted in agarose cassettes from fresh blood to allow isolation of high molecular weight DNA. This method was used to phase variants of clinical interest in two cases with *ABCA4*-related disease. This method resulted in ultra-long reads, with multiple reads in each case spanning the entire *ABCA4* locus.

3.3.1 Phasing in human genomics

3.3.1.1 *ABCA4* as a target gene for establishing a phasing assay

ABCA4 variants are the most common cause of Mendelian retinal disease (Cremers et al., 2020). *ABCA4*-related disease represents a range of IRDs with variable severity, progression and penetrance (Section 1.3). Despite this, many cases with *ABCA4*-associated disease show a set of clinical features which, when all are present, are highly indicative that defective ABCA4 is the underlying cause.

There are a number of genes which when mutated result in a similar phenotype to *ABCA4*-related disease including *PROM1*, *PRPH2* and *ELOVL4*

(Kniazeva et al., 1999; Bernstein et al., 2001; Cideciyan et al., 2005) (Section 1.3.1.4). Despite this, it is often possible to correctly identify cases of *ABCA4*-associated disease through careful clinical examination. Because of this, and its role as a frequent cause of disease, *ABCA4* has been the subject of intense study. As cases with *ABCA4*-related disease can relatively easily be isolated (in comparison to other genes with far less distinct presentations), it has been possible to study cohorts of *ABCA4* cases in great detail (Khan et al., 2020; Cornelis et al., 2022). In particular, when cases have a single pathogenic *ABCA4* allele, there may be a second undiscovered pathogenic allele.

This has allowed the identification of complex alleles in ABCA4 which comprise ~10% of the alleles in ABCA4 (Runhart et al., 2018; Cremers et al., 2020) As such, *ABCA4* is an excellent target to establish a robust phasing assay to facilitate phasing of these complex alleles in individuals with no available familial DNA. In addition, as *ABCA4* is a relatively large gene, any successful method is likely to be effective for use on other genes. Finally, a number of therapies are under investigation for *ABCA4*-related disease, including by members of this consortium, and accurate genetic diagnosis is crucial for participation in these studies.

3.3.1.2 Current strategies to establish phase

As reads produced by NGS are short in nature, it is difficult to unambiguously map them to regions of high sequence homology. In the context of Mendelian disorders such as *ABCA4*-related disease, it is often required to phase detected variants by segregating them based on the homologous chromosome of origin.

Current methods to establish phase in diagnostic settings generally rely on segregation of homologous chromosomes by genotyping direct family members. However, if no familial DNA is available, pathogenic variants are generally assumed to be in *trans* (Danecek et al., 2011; Stephenson et al., 2021; Hitti-Malin et al., 2022; Panneman et al., 2022). This limits the study of complex alleles to small cohorts with available familial DNA. A wide variety of other methods have been established to determine phase. Many of these have been developed to establish phase across entire chromosomes or to establish haplotype relationships between distal loci. This includes probabilistic analysis of large cohorts using a reference genome to infer the

phase of variants (Browning and Browning, 2007; O'Connell et al., 2016; Loh et al., 2016). While this is useful for inferring relationships between alleles, it is not suitable in a diagnostic setting for phasing of an individual. Similarly, methods such as SNP arrays and ultra-low coverage WGS have been used to estimate phase of loci of interest, but are not suitable for phasing in a diagnostic setting (Rubinacci et al., 2021). Other methods such as Hi-C are extremely useful for establishing distal relationships, but are designed to phase distal loci and lack the resolution to establish phase within a gene (Beitel et al., 2014). Methods have also been established to allow for single cell sequencing of a single homologous chromosome, allowing for accurate phasing of large loci (Falconer et al., 2012; Sanders et al., 2017). However, this results in extremely low read depths and is again not suitable for phasing in a diagnostic setting.

As such, a relatively low-cost, high-throughput, accurate method with high resolution is required to fill this diagnostic gap. Ideally, this method would result in a 'phased haploid' output, allowing integration with existing pipelines.

3.3.2 Long-read sequencing

There a number of long-read sequencing platforms which have recently been developed. These are generally either synthetic long-read sequencers or single molecule long-read sequencers.

3.3.2.1 Synthetic long-read sequencing

Over the past decade, a number of synthetic long-read platforms have been developed to construct synthetic long-reads from short reads. These typically function by tagging fragments of native genomic DNA and then amplifying the region. The resulting fragments contain tags which link to the molecule of origin (Mandelker et al., 2016). The fragments are sequenced using NGS and the resulting short reads can then be mapped back to the fragment of origin, allowing synthetic longer reads to be constructed (White et al., 2016; Marks et al., 2019). While numerous platforms have been developed, such as 10X Genomics' Linked Reads (Zheng et al., 2016), Illumina's Moleculo (White et al., 2016; Da Fonseca et al., 2020), LoopSeq (Callahan et al., 2021) and more recently Illumina Complete Long-Read (https://www.illumina.com/science/technology/next-generation-

sequencing/long-read-sequencing.html, accessed 3/4/2023), these all function very similarly, differing in how the molecules of origin are labelled and amplified.

Synthetic long-read sequencers benefit from the increased accuracy associated with NGS while allowing applications previously intractable to NGS due to read length. Synthetic long-read sequencers have seen successfully used for human *de novo* genome assembly with diploid contig assembly, generation of improved reference sequences, and to phase regions of interest resulting in 'phased haploid' reads (Kuleshov et al., 2014; Zheng et al., 2016; White et al., 2016; Weisenfeld et al., 2017; Da Fonseca et al., 2020; Chen et al., 2020; Meleshko et al., 2022). Moleculo and 10X Genomics Linked Reads have since both been discontinued. Recently (Launching Q1 2023), Illumina have released an updated synthetic long-read platform 'Illumina Complete Long-Read'.

While these technologies are capable of generating longer reads compared to NGS, it was decided that single molecule sequencing was more suitable for use in these experiments. Synthetic long-read sequencing methods typically have a comparatively short read length compared to single molecule sequencing. Illumina Complete reads have a reported N50 of 6-7 kb but benefit from the increased read quality associated with Illumina sequencing (https://www.illumina.com/science/technology/next-generation-sequencing/long-read-sequencing.html accessed 9/1/2023), whereas Nanopore reads of over 4 Mb have been reported (https://nanoporetech.com/products/minion, accessed 8/2/2023). Due to the reduced read length obtained when using synthetic long-read sequencers, these platforms would have encountered the same issues surrounding creating a larger phase block using heterozygous variants with no possibility of using an approach which generated ultra-long reads.

Additionally, workflows such as the 10X Linked Reads feature amplification steps during their workflow. While modified amplification such as 10X Linked Reads gel-beads in emulsion (GEMs) reduce the likelihood of PCR artefacts, as demonstrated in this chapter PCR can significantly confound accurate sequencing and phasing. Additionally, there are not enough tags for all of the

sequenced molecules, introducing the possibility (although very unlikely), that both haplotypes at the same locus will share a tag (Zook et al., 2016a).

Finally, in comparison with single molecule sequencing, it would have been difficult to directly observe any artefacts generated using synthetic long-read sequencing which may have confounded phasing. In this experiment, it was possible to measure chimeric read formation as each read originated from a single DNA fragment. As a result it was possible to determine that the chimeras were formed during long-range PCR and adjust this step to reduce chimera formation. In contrast, if a synthetic long-read platform had been used, it would have been far more difficult to investigate any artefacts which may have confounded phasing. Similarly, when phasing using HLS-CATCH, although reduced read depth was obtained the phased reads originated from genomic DNA and spanned the entire locus. As such, the observed haplotypes were likely accurate.

3.3.2.2 Single molecule sequencing

Single molecule sequencers such as the PacBio Single Molecule Real Time (SMRT) sequencing platform (Section 1.4.2) and the Oxford Nanopore Technologies nanopore platform (section 1.4.3) now dominate the emerging long-read sequencing market. Single molecule sequencing is characterised by direct sequencing of individual DNA or RNA molecules and sequencing has been used to improve applications which have been understudied due to the size of NGS reads, such as comprehensive RNA sequencing and structural variant identification (Lavrichenko et al., 2021; Nurk et al., 2022).

Single molecule sequencing has also been used to establish large haplotypes at high accuracy and resolution. This has been done to improve references, including complete human genomes and telomere to telomere assemblies (Miga et al., 2020; Nurk et al., 2022). Single molecule sequencing has also been used for diploid *de novo* assembly of human genomes (Shafin et al., 2020). A wide variety of methods have also been developed to establish 'phased haplotypes' of variants of interest using both PacBio SMRT sequencing and Oxford Nanopore sequencing. Of particular interest here, due to the long reads obtained from single molecule sequencing, it is possible to directly differentiate sequences deriving from homologous sequences including sequences derived from homologous chromosomes.

PacBio sequencing offered an attractive alternate sequencing platform at the beginning of this study. As outlined in Section 1.4.2, PacBio SMRT sequencing utilises parallelised single molecule sequencing by synthesis (Eid et al., 2009). PacBio SMRT sequencing's main advantage compared to comparative methods is the very high accuracy of circular consensus sequencing. SMRT sequencing is highly accurate, although it is worth noting that the often quoted figures for accuracy are from circular consensus sequencing, also known as HiFi sequencing, During HiFi sequencing, the circular SMRT bell template is sequenced multiple times. By sequencing the same molecule multiple times, a highly accurate consensus sequence can be obtained (Travers et al., 2010; Loomis et al., 2013).

In spite of the increased accuracy of PacBio sequencing, ONT sequencing has several significant advantages which led to its selection for this study. As sequencing is not performed by synthesis, there is functionally no limit on read length. Reads of >2.2 Mb have been reported (Payne et al., 2019) and ONT have claimed reads of over 4 Mb in internal datasets (https://nanoporetech.com/about-us/news/blog-kilobases-whales-short-history-ultra-long-reads-and-high-throughput-genome, accessed 11/1/2023). In contrast, HiFi sequencing is generally only possible for targets smaller than ~13.5 kb (Wenger et al., 2019).

ONT sequencing has also grown increasingly accurate with the development of new library preparation kits, basecalling softwares and updated variant calling softwares (Wang et al., 2021). PacBio claim to have achieved greater than Q50 predicted accuracy (>99.999%) consensus accuracy (https://www.pacb.com/technology/hifi-sequencing/how-it-works/ accessed 10/1/2023). While this is impressive, and often advertised, the per base accuracy is closer to 99.9% for circular consensus reads and lower for single pass reads (https://www.pacb.com/blog/understanding-accuracy-in-dna-sequencing/, accessed 11/1/2023). This is still greater than Nanopore sequencing, however consensus accuracy was shown not to affect sequencing using stringent filtering parameters (Section 3.2.2).

Finally, the (at the time) newly released Flongle was a major factor in the selection of nanopore sequencing. The Flongle is an adaptor for the MinION

flowcell which offers a reduced output at a reduced cost. The theoretical maximum output of a Flongle flowcell is 2.8 Gb (https://nanoporetech.com/products/Flongle, accessed 9/1/2023) compared to 50 Gb for a MinION flowcell (https://nanoporetech.com/products/minion, accessed 9/1/2023). This allowed for bench-top sequencing at a comparatively low per-run cost. Because of this, it was cost effective to run one sample per Flongle without multiplexing, which reduced experiment and analysis complexity. In contrast, the current range of PacBio sequencers are large, expensive machines generally intended for use in a central sequencing facility. Additionally, PacBio Sequel IIe instruments use SMRT cells at a cost of ~$1,300 per cell. For variant detection in a ~3 Gb human genome, two cells are recommended at a total cost of ~$2,600 (excluding other reagents) (https://www.pacb.com/wp-content/uploads/Application-Brochure-What-Can-You-Do-with-One-SMRT-Cell.pdf, accessed 11/1/2023). This is cost effective for many pooled samples, approximately 1,000 amplicons of 1-10 kb can be sequenced for a per sample cost of $1-2. While this is cost effective at scale, only ~525 amplicons were sequenced on Flongles throughout this experiment. Using Flongles, it was possible to optimise the assay cost effectively, a process which would have been far more expensive using SMRT sequencing.

3.3.3   Long-read enrichment

Whichever long-read sequencing platform is selected for phasing, appropriate enrichment to maintain fragment length of target sequences is important to properly use the advantages of these platforms. These methods aim to either capture the variants of interest on a single read, allowing unambigious phasing, or to use heterozyous variants to create larger phase blocks from tiled overlapping reads.

3.3.3.1   Amplification-free sequencing

3.3.3.1.1   Long-read whole genome sequencing

Depending on experimental design, it may be possible to forgo enrichment and perform long-read WGS. While this allows for robust phasing and long-read lengths to be achieved, this is a very expensive method for routine use and is more expensive than whole genome NGS. If WGS is performed on a MinION an average coverage of ~3 × is typically achieved. This is generally not considered sufficient for many purposes and is typically either

supplemented with NGS, additional MinION flowcells or the larger PromethION flowcells. Simiarly while PacBio sequencing is more accurate, PacBio recommends using two SMRT cells to sequence a single genome at a cost of ~$2,600. Additionally as with Illumina WGS, long-read WGS requires increased data storage and data analysis, increasing experimental costs and preventing its use in many circumstances. Long-read WGS has been used effectively on smaller genomes such as multidrug resistant bacterial strains and SARS-CoV-2 (Luo et al., 2022; Vereecke et al., 2023). For the phasing of *ABCA4* long-read WGS would have been prohibitively expensive and required too much DNA for many of the samples intended for sequencing as part of this study.

Target enrichment methods are commonly employed to maximize limited sequencing capacity for human samples. Enrichment methods for single molecule sequencing can broadly be divided into amplification-free methods, where native genomic DNA is enriched and directly sequenced, or amplification based methods such as long-range PCR. More latterly, adaptive sampling for the nanopore has also been developed.

3.3.3.1.2   CRISPR-Cas9 amplification-free enrichment

An amplification-free method which has been used succesfully in conjunction with single molecule sequencing is CRISPR/Cas9 mediated enrichment. The CRISPR/Cas9 system is a programmable endonuclease which allows for efficient targeting of double strand breaks (Barrangou et al., 2007; Marraffini and Sontheimer, 2008; Barrangou et al., 2012; Nachmanson et al., 2018).

This system has been widely used for amplification-free enrichment of high molecular weight DNA fragments. Various strategies have been developed for both nanopore and SMRT sequencing. For example, selection of fragments generated by CRISPR-Cas9 complexes targeting sequences flanking the region of interest (Nachmanson et al., 2018). Other methods utilise the exonuclease activity of Cas9 in an altered library preparation for nanopore and SMRT sequencing (Tsai et al., 2017; Watson et al., 2019). Broadly, by targeting regions of interest with CRISPR-Cas9 complexes, exposed ends are made available to which adaptors can be ligated. These systems are powerful because they allow for native targeting of regions which are known to cause

disease but are intractable to amplification, reduce PCR artefacts and allow for targeted study of DNA modifications.

Amplification-free enrichment was developed for SMRT sequencing by Tsai et al., 2017. A SMRT bell library is prepared as in standard SMRT sequencing library preparation (Section 1.4.2). CRISPR-Cas9 complexes are added targeting the region of interest, resulting in the excision of one of the hairpin adaptors. Capture adaptors are then ligated to the open end. Magnetic beads are introduced which bind to the capture adaptors. A magnet is then used to pull down the magnetic beads with the attached SMRT bell molecules, allowing the rest of the library to be washed away (Höijer et al., 2018; Hafford-Tear et al., 2019).The resulting enriched library is then sequenced on a SMRT cell.

Comparable methods have been developed using the nanopore (Fourquet et al., 2018; Gilpatrick et al., 2020). One such method uses CRISPR-Cas9 complexes to cut at the target region and guide adaptor ligation (Watson et al., 2019). In this method, the genomic DNA is dephosphorylated. CRISPR-Cas9 complexes are then introduced targeting the region of interest. The exonuclease activity of Cas9 exposes the 5′-phosphate at the target locus. The other direction is not sequenced as the Cas9 remains bound to the side of the cut proximal to the PAM sequence. Because of this, the 5′-phosphate on this strand will be sterically hindered, effectively only exposing one 5′-phosphate. A dA tail is then ligated to this tail. The terminal phosphate is required for dA tail ligation so only the region of interest is eligible to have the adaptor sequence bound. The prepared library is then sequenced on a MinION flowcell. This method has been used to succesfully characterise genomic duplications (Watson et al., 2019).

Amplification-free methods were initially explored for use during this study in conjuction with nanopore sequencing. However this was decided against for several reasons. As there is no amplification during targeting or library preparation, a large mass of input DNA would be required to ensure sufficient target copies are enriched. This would have precluded sequencing of many samples which have relatively little DNA available for sequencing. Additionally, relatively high molecular weight DNA would be required to ensure sufficient read length, limiting the number of samples this technique can be

performed on. Due to the read lengths which were expected to be obtained from older samples, a tiled approach spanning *ABCA4* was expected to be required. This would have required a prohibitive number of guides. Even with large inputs of high quality DNA, low read depths are typically achieved. Finally, as these methods require a MinION flowcell as opposed to a Flongle flowcell, they would have been approximately 10-fold more expensive than the PCR method which was used.

3.3.3.1.3   Sage-HLS CATCH

When the issue of chimeric read formation was identified, enrichment by CRISPR/Cas9 was explored again for a more limited series of cases, however it was decided that a tiling approach would still be required. This would have required a prohibitive number of guides and would likely have encountered the same practical challenges surrounding a lack of variation preventing phasing.

Instead, Dr. Gavin Arno facilitated the use of a modified CRISPR/Cas9 based method of enrichment, CATCH using the the Sage-HLS system (Jiang et al., 2015; Shin et al., 2017; Zhou et al., 2020). This system has several advantages compared to other methods of CRISPR/Cas9 enrichment which made it more feasible. As DNA extraction and targeting are performed *in situ* in an agarose gel plug, very large fragments of DNA can be targeted. In this experiment, reads of over 143 kb were obtained with an N50 (the length of the shortest contig for which longer and equal length contigs cover at least 50% of the assembly) at this locus of up to 95 kb. Other studies using this technique have succesfully completely characterised deletions at 22q11.2 and 16p11.2 (Zhou et al., 2020). In these experiments, reads of over 405 kb were obtained. The results presented in this chapter compares favourably to those of Zhou and colleagues as the maximum read depth in this experiment was 87 × at *ABCA4* compared to 30-40 × at 16p11.2. Additionally, the longest N50 obtained in that experiment set was 64.51 kb compared to 95 kb in this experiment (Zhou et al., 2020). Finally, target enrichment in the Zhou et al paper was 30-fold compared to up to 68.6-fold achieved in this project (calculated by qPCR of *ABCA4* and *RNase P*, Figure 3.17). Opimization by Dr Gavin Arno and colleagues has improved targeting and fragment length preservation.

This method is still relatively early in development and further optimisation is required before its general use. Firstly, the protocol provided by Sage Biosciences included details on freezing freshly extracted nuclei in Bambanker (a medium typically used for freezing cell cultures). However when thawing the nuclei, they formed an irreversible clump which could not be resuspended. This reduced the sequencing yield, as can be seen by comparing experiments performed on freshly extracted nuclei with experiments performed on frozen nuclei. This was explored by extracting control nuclei and resuspending in Bambanker without freezing, which also caused the nuclei to irreversibly clump. Resuspending in greater volumes of Bambanker did not decrease this. Additionally, numerous trials on the opsin array (Dr Gavin Arno, personal communication) on frozen nuclei resulted in very low read depths. As such, Case CATCH 3 was asked to return to clinic and give a fresh blood sample so that freshly isolated nuclei could be used instead of previously frozen nuclei. As it is possible to only do this assay on freshly extracted samples, this increases the logistical challenges associated with this experiment. Nuclei must be extracted from fresh blood within 4-5 days of extraction and the CATCH experiment performed on the nuclei as quickly as possible. A method to successfully freeze nuclei would greatly ease the wider adoption of this technique.

Further, generation of excess foam reduced output in some experiments. This disrupted the electrical current through the cassette and in some cases flooded the elution wells leading to reduced product. It was determined by Sage Biosciences that this was due to a high concentration of SDS in the buffer used to stop the CRISPR-Cas9 reaction. By replacing the supplied 3% SDS Lysis buffer with a 1% SDS lysis buffer less foam was generated for Case CATCH 3 compared to other experiments. Additionally, replacing the running buffer after each user interaction also helped to mitigate, but not eliminate foam. Further improvements to the SAGE-HLS cassette and reagents may help mitigate this in future.

While CATCH is capable of generating far longer reads than other enrichment methods, it must be sequenced on a MinION flowcell as opposed to the lower throughput (and comparatively cheaper) Flongle flowcell. Even with a MinION flowcell, a relatively low yield is obtained. It was hypothesised that using the products of a CATCH run on a Flongle flowcell would result in a comparable amount of data at a reduced cost, however this experiment did not generate

any usable data. Nanopore recommends between 2-20 fmol of DNA is added to the Flongle flowcell and 100-200 fmol of DNA is added to the MinION flowcell (Oxford Nanopore Technologies). If insufficient DNA is added to the flowcell, there are a larger proportion of inactive pores which quickly become unusable. Amplification-free methods such as CATCH isolate very few copies of DNA. It was theorised that using the low molarity CATCH elution product on a Flongle would increase the proportion of active pores, thereby resulting in a successful sequencing run, in reality too many pores died too quickly to generate any meaningful data.

As such, the more expensive MinION flowcell is required for each CATCH experiment. At a cost of ~£1000 a flowcell, this is an expensive experiment to perform routinely. Additionally, there were increased costs associated with the enrichment compared to amplification based methods, such as the SAGE-HLS machine, cassettes and CRISPR/Cas9 reagents. As has been noted previously, this experiment only uses 20-30% of the sequencing capacity of a MinION flowcell (Zhou et al., 2020). While there is theoretically excess sequencing capacity in the MinION and Flongle to allow for multiplexing of CATCH and long-range PCR respectively, considerations of data quality prevented their use in these experiments.

3.3.3.1.4 Adaptive sampling

A recently developed amplification-free enrichment method is adaptive sampling. Adaptive sampling or Read Until, allows for enrichment of genomic DNA samples by real-time basecalling and molecule selection (Loose et al., 2016). Prior to sequencing, regions of interest are defined *in silico*. During sequencing of sheared genomic DNA, the resulting sequence is basecalled in real time. If the molecule passing through the pore is the target sequence, it continues to pass through the pore. However, if the molecule is a different sequence the electric current is temporarily reversed, pushing the molecule back out of the pore (Payne et al., 2021). This method is highly promising and has been demonstrated to enable over 40 × coverage of target sequence corresponding to a 5.7-fold enrichment (Payne et al., 2021). This method was currently being optimised in the lab as this project completed and so was not available for use in this experiment, but promises to be a powerful tool capable of targeting of panels of genes of interest with no additional laboratory steps. This technique has been succesfully leveraged for sequencing of microbial

communities, tumors, repeat expansions and for the detection of a myriad of other variation in human genomes (Miller et al., 2021; Stevanovski et al., 2022; Marquet et al., 2022; Patel et al., 2022). Further developments of this platform have led to programmes such as BOSS-RUNS, which use adaptive sampling to dynamically update targets. This has been used to characterise microbial communities and, during sequencing, select against highly abundance species (Weilguny et al., 2023).

While this promises to be a very powerful technique, the genomic DNA is fragmented to approximately 10 kb prior to sequencing (Payne et al., 2021). To phase across a large locus, reads would be segregated into larger phase blocks by heterozygous variants which overlap multiple reads. As the reads are randomly fragmented, this would not rely on variation at specific loci, however if a region is encountered with no reads spanning between heterozygous variants, a new phase block would be created as happened during this experiment. A future increase in fragment length which can be sequenced would greatly mitigate this.

3.3.3.2 Amplification based enrichment

Finally, other enrichment methods rely on amplification of target sequences. Major advantages of enrichment by PCR include; cost effectiveness, low input DNA and very high enrichment of target sequences. However, relatively low target size, low throughput and the introduction of PCR artefacts remain potent limitations to be addressed.

While phasing by long-range PCR was demonstrated to be possible, if a reduced cycle count was used and if sufficient variation was present to allow stitching together of tiled amplicons as validated with the NA12878 dataset, there were practical challenges which limited its utility. A tiled approach was adopted so that previously extracted, freeze thawed samples could be sequenced, however this led to challenges stitching together larger phase blocks.

Additionally, chimeric read formation limited the number of PCR cycles which could be used, increasing the practical difficulty of long-range PCR. It has been estimated that up to 40% of PCR products from mixed populations are

chimeric and that this can occur between homologous chromosomes (Meyerhans et al., 1990; Lahr and Katz, 2009). While it has been previously observed that chimeric product formation is reduced by reducing cycle number, it has also been reported that chimeras can occur after a single round of amplification (Odelberg et al., 1995). It has been proposed that chimeras form when an incomplete PCR product binds to a homologous locus during amplification and acts as a primer. Studies using single molecule sequencing have supported this and have found that the average rate of recombination occurs at 1.1×10^{-4} per base per amplification cycle for *Taq* DNA polymerase (Potapov and Ong, 2017). An alternative model has been proposed where the polymerase switches template during an amplification cycle, supported by the observation that chimeras can form after a single round of amplification (Odelberg et al., 1995). While the rate of chimeric product formation in this experiment was found to vary between amplicons, Potapov and colleagues observed that across all observed loci, the rate of chimera formation was found to be evenly distributed. It is likely that the amplicons which were observed to have high chimera formation in this experiment were less likely to be successfully amplified during a single round of PCR. Alternate primer pairs spanning the same region may have resulted in reduced chimeric read formation, however this was not investigated. In addition to reducing cycle number, it is plausible that alternate polymerases will have reduced rates of chimera formation.

A previous report described the presence of chimeric amplification products after amplification of ~9 kb products (Laver et al., 2016). In that report, initial sequencing was carried out on the MinION and the presence of the chimeric products was confirmed with PacBio sequencing. As observed in this thesis, Laver and colleagues observed that at high cycle number (39 cycles in their case) there were three observed haplotypes at similar proportions and a fourth haplotype which was at a reduced proportion. The proportion of reads belonging to the final haplotype was judged to be reduced due to reference alignment bias. Laver and colleagues observed that by reducing cycle number to 29, they were able to significantly reduce chimeric read formation. However, they concluded that: "*PCR carried out with fewer than 29 cycles did not yield enough product for sequencing, this means that a reduction of cycles alone will not be able to entirely eliminate chimera formation.*". Laver and colleagues concluded that this was a major pitfall preventing robust phasing using nanopore sequencing. The workflow used in the current experiment, which

benefits from the significantly increased read depth offered by Flongle flowcells, and from robust amplification using the SequelPrep™ Long PCR kit has enabled this issue to be addressed. The increased read depth compared to that obtained by Laver et al (maximum 111 nanopore reads aligned to both variants of interest, in comparison to a maximum read depth of >12,000 × achieved on Case NA12878) gives greater confidence in the construction of larger phase blocks and called variants. Further, since the previous report, there has been a general improvement in nanopore sequencing through greater throughput and increased accuracy (Wang et al., 2021). The reference alignment bias observed by Laver and colleagues was not observed as a confounding issue during the experiments performed in this study. Generally, proportions of all four observed haplotypes were similar at 35 cycles, possibly due to improvements in aligners since the previous report.

By comparison with the publicly available Case NA12878, the constructed phase blocks were accurate at 25 cycles and the proportion of chimeric reads was low enough to not affect phasing. While it was possible to significantly reduce chimeric read formation by reducing cycle number to 25 cycles, this also greatly reduced PCR efficiency. Previously optimized PCRs required re-optimization and frequently multiple PCR products of the same amplicons would need to be pooled to ensure sufficient molarity. Further, this necessitated a reduction in annealing temperature, reducing the specificity of the primers and further reducing the moles of available PCR product (Figure 3.20).

In this experiment a tiled approach was used, relying on heterozygous variants situated in the overlaps between amplicons to construct larger phase blocks. To maximize the amount of variation captured in the overlaps between the amplicons, public genomic datasets were interrogated. This was done to map areas of low variation and avoid placing overlaps in these. While this was relatively successful, and it was possible to use a series of ten cases to redesign low variation overlaps, these datasets are generally centered around populations of European descent (Ballouz et al., 2019). The controls and *in silico* exemplar cases which were used in this study were also of European descent. It was predicted that the majority of cases sequenced using this method would also be of European descent, and as such, the amplicons which were designed using population variation should have captured sufficient variation as predicted. However, this experiment was performed on individuals

from a diverse cohort, and as the individuals were pseudo-anonymized, their ethnicity was not considered at the time of sequencing. While this assay was biased towards individuals of European descent, it is evidently difficult to predict common variation in individuals using population data even from individuals of the 'target' population. Further, it would not be feasible to design genetic assays in diverse cohorts to target specific populations. This would require segregation of patients by ethnicity and would necessitate attaching ethnicity data to pseudo-anonymized genetic testing, introducing potential ethical and patient privacy issues.
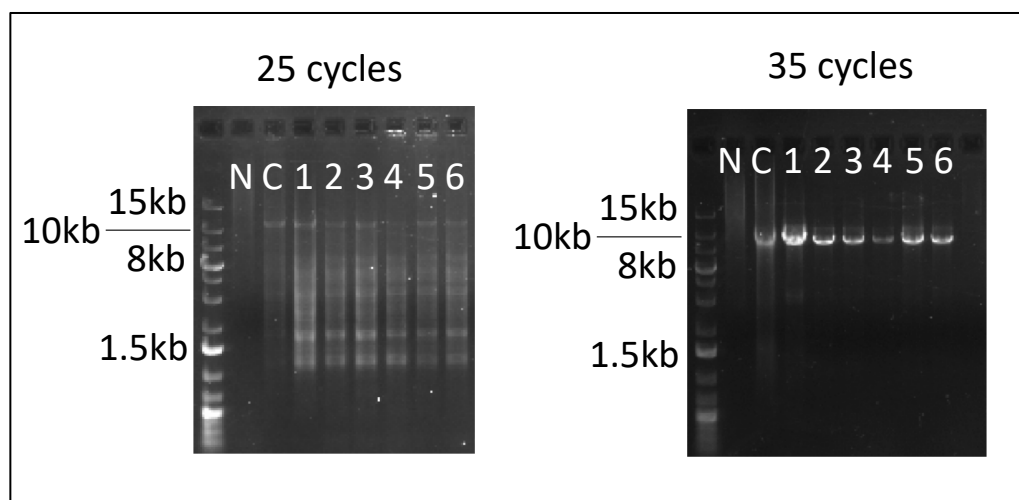
Figure 3.20 Long-range PCR at *ABCA4* became less efficient at 25 cycles of amplification compared to 35 cycles

In both panels, amplification products of amplicon L10 are shown. In an attempt to increase the yield of product at 25 cycles, the annealing temperature was reduced, however this resulted in a corresponding drop in specificity. N indicates a no template control, C indicates a positive control, in both cases the NA12878 case was used as a positive control. The numbers indicate cases which were amplified. Products separated on a 0.5% agarose gel, fragment size shown using an NEB 1 kb plus ladder.

Further, variation can vary greatly between individuals at the same locus. For example, the control which was used throughout this experiment, NA12878, was selected as this individual is a commonly used reference for benchmarking studies. Case NA12878 has 228 single nucleotide variants spanning *ABCA4*, of which 169 were heterozygous. Prior to the selection of NA12878 as a benchmark, an in-house case who had undergone WGS was selected (1343). This case had a diagnosis of Late Onset Retinal Dystrophy and sequencing of several members of this pedigree have so far, not revealed the cause of disease. In contrast to Case NA12878, this case had 106 variants

spanning *ABCA4* of which only 17 were heterozygous. As such, this case had seven overlaps with insufficient variation to allow phasing. While difficult to predict, individuals with low levels of variation at the target locus are unsuitable for methods such as the one used here which rely on variation to phase. Notwithstanding the lack of variation within the overlaps between amplicons, there is no reason why significantly larger phase blocks than 94.9 kb could not be constructed. In particular this method is well suited for phasing of variants which are relatively close together and can be captured in a single amplicon, or for cases who have had prior NGS. In these cases, it may be beneficial to design bespoke primers to target variation in the overlaps between amplicons.

Ultimately, during long-read sequencing enrichment, a trade off must be made between desired fragment length and sample quality. Techniques which require ultra-long reads require specialist DNA extraction and library preparation methods to preserve fragment length. Because of this, they are not possible on most samples, in particular freeze thawed samples. To phase lower quality samples, enrichment strategies which produce shorter fragments are required and a tiling approach such as the one adopted in this study are necessary. As demonstrated here, this can lead to practical challenges. As such, targeted long-range PCR presents a cost effective method for cases where the variants are relatively close together, or for cases which have undergone previous screening and as such overlaps can be targeted to where there is variation present. This presents as a viable assay to quickly screen targeted cohorts of patients, however a single set of amplicons (as attempted here) is unlikely to be able to reliably phase large loci.

As access to sequencing continues to expand, the ability to cost effectively phase loci of interest will be of increasing importance. While this work focused primarily on establishing an assay to study complex alleles in *ABCA4*, a method to reliably phase large loci is also important to allow for the study of new genes thought to be implicated in disease, to examine the relationship between functional variants when multiple functional variants are found in the same gene, and to establish accurate genotypes of patients for inclusion in gene therapy trials. As such, a method which is cost effective, technically feasible and replicable, is necessary to allow for the routine phasing of relevant variants in eye disease and beyond.

Previously established methods to phase variants, such as cloning of target genes, are accurate but do not allow for sufficient throughput compared to other techniques. Similarly, while phasing by segregation analysis is routinely used, many patients do not have available familial DNA and so cannot be studied. As a result, it is likely that long-read sequencing technologies will be commonly used going forward to phase variants of interest.

While phasing by long-range PCR is technically feasible as shown here, it is practically challenging. A major limitation of this approach was the formation of chimeric reads, although this was addressed by reducing PCR cycle number. A limitation of this approach which could not be overcome was the lack of heterozygous variants in the overlaps between amplicons. As a result it is possible that this method will be of some use in the near future, however this will likely be restricted to phasing smaller loci or for targeted phasing of cases with prior sequencing. In the more distant future it is likely that an alternate technique such as adaptive sampling or HLS-CATCH will be used to routinely phase variants. Adaptive sampling was recently used to phase variants of interest in *USH2A* (Nakamichi et al., 2023), however it is not possible to use this method to sequence highly repetitive regions such as the opsin array, as the long-reads achieved are still too small to uniquely align. In contrast, phasing by HLS-CATCH is very effective and it is possible to sequence repetitive loci such as the opsin array, although HLS-CATCH is significantly more expensive compared to adaptive sampling. As such, it is likely that in the future adaptive sampling will be used to sequence genes of interest and phase resulting variants in the majority of cases, while a method such as HLS-CATCH will be used to target more challenging loci. Finally, as the cost of long-read WGS continues to fall it is possible that long-read WGS will be used to routinely screen patients, allowing for high quality sequencing of target loci as well as phasing of variants of interest in a single step.

3.4   Summary

In this study, phasing by long-range PCR in conjunction with nanopore sequencing was shown to be a cost effective strategy to phase large sections of *ABCA4*. Further, the resulting phase blocks were shown to be accurate by comparison with a well-characterised benchmark sample. Phasing was

hampered by a lack of variation in the overlaps between amplicons, however where sufficient variation was present, this strategy represents a robust, albeit practically challenging method of phasing. This was supplemented by an amplification-free method which proved to be effective and presented a method of ambiguity-free phasing, however it required fresh samples and was significantly more expensive than the PCR based method.

4   Characterisation of deletions using nanopore sequencing

4.1   Introduction

Copy Number Variants (CNVs) have long been recognised as an important contributor to genetic disease and methods to detect and characterise these variants at greater sensitivity and resolution is an ongoing area of development, with the potential for significant clinical impact (Girirajan et al., 2011). Next generation sequencing (NGS) has allowed for improved detection of CNVs and has been an important addition to existing methods. However, substantial challenges remain when using NGS to detect CNVs, in particular when using targeted sequencing methods. Numerous bioinformatics approaches, which typically use comparative read depth or paired end mapping information have been developed (Guo et al., 2013). However, these methods commonly characterise deletions to the resolution of the nearest exon, and it can be challenging to characterise breakpoints at single-nucleotide resolution. For example, when targeted sequencing is performed, areas with lower coverage may be putatively "called" as containing a deletion, even though the breakpoints are not captured. To characterise the breakpoints of the putative deletion, numerous standard PCRs will be attempted to span the variant and sequence the breakpoint by Sanger sequencing, a time-consuming process known as PCR walking. In this chapter, a novel method to quickly and cost effectively characterise previously detected deletions using long-range PCR and nanopore sequencing is described.

4.2   Results

A methodology was developed using a long-range PCR spanning the putative deletion (Section 2.3.2), followed by sequencing on a Flongle flowcell (Section 2.6.3). Using long-range PCR, it would be possible to span the putative deletion with a single amplification product and the primers could be positioned to avoid repetitive sequences. Nanopore sequencing allowed the breakpoints of the deletion to be un-ambiguously mapped. This could then be verified by standard PCR (Section 2.3.1) and Sanger sequencing (Section 2.4).

4.2.1 Selection of cases with large deletions

This approach was applied to a series of IRD cases with large deletions as a proof of concept. An initial cohort of five cases who had undergone targeted screening which had resulted in putative large deletions being identified were selected. Details of all cases can be viewed in Table 4.1. This cohort comprised of one case with a heterozygous deletion in *PRPF31* and three cases with homozygous deletions in *EYS, CNGB1, CNGA1* and one case with a hemizygous deletion in *NDP*.

Table 4.1 Details of cases with suspected large deletions

Proband ID	Case number	Gene	Sex	Inheritance pattern	Phenotype	Detection method	Identified by
4755	Case 1	*CNGA1*	M	Sporadic	RP	ExomeDepth on MIPs data	J.P.
2249	Case 2	*CNGB1*	M	Recessive	RP	ExomeDepth on MIPs data	J.P.
644	Case 3	*EYS*	F	Sporadic	RP	ExomeDepth on MIPs data	J.P.
769	Case 4	*PRPF31*	M	Dominant	RP	ExomeDepth on WES data	C.S.
F1265	Case 5	*NDP*	M	Sporadic	FEVR	ExomeDepth on WES data	E.P.

J.P.: James Poulter. C.S.: Claire Smith. E.P.: Evangelia Panagiotou. RP: Retinitis pigmentosa. FEVR: Familial Exudative Vitreoretinopathy.

4.2.2 Nanopore sequencing of putative deletion-variants in IRD cases

4.2.2.1 Characterisation of *CNGA1* deletion

Case 1 (Proband 4755) was a singleton case with a diagnosis of RP. Prior to this study, ExomeDepth (Plagnol et al., 2012) was used on MIPs screening data targeting exonic sequences in 100 IRD genes (method detailed in Weisschuh et al., 2018) (Table 4.1). This indicated a homozygous deletion spanning exons 6-10 of *CNGA1*. From this, the minimum size of the deletion was estimated to be 7.5 kb and the maximum size of the deletion was estimated at 56 kb. The deletion was confirmed to be present by standard PCR targeting exon 7 (Forward: CACAAGCCACAACCTGAGAC, Reverse: ACATCGAAGGGGTAAGGCAA, Target: chr4:47941522-47942022 hg19, 500bp) which generated no PCR product (data not shown). Following this, the deleted allele was amplified by long-range PCR and sized by gel electrophoresis.

Two pairs of primers were designed to flank the deletion. Primers were placed in the next loci which had been successfully sequenced, or in the intervening intron. Long-range PCR primer sequences are recorded in Table 4.3. Primer pair 2 generated an estimated 9 kb amplification product. This was significantly smaller than the 25 kb product expected from the reference sequence (Figure 4.1).

Figure 4.1 Gel electrophoresis of long-range PCR product spanning putative *CNGA1* deletion

Long-range PCR was performed using the SequelPrep™ kit at an annealing temperature of 58 °C. Amplification of primer set 1 using a control genomic DNA (C) and patient DNA (P) sample did not result in a visible amplification product. Amplification using primer set 2 resulted in a product only in the patient DNA of approximately 9 kb (P*). Both primer sets were used in the no template control (N). The size of the products is indicated by a Thermo Scientific GeneRuler 1 kb plus ladder.

Nanopore sequencing on a Flongle flowcell of this product allowed characterisation of the breakpoints of the deletion at a nucleotide resolution (Figure 4.2). These data defined a novel 14.8 kb homozygous deletion spanning exons 6-10 of CNGA1, (NM_001142564.1) g.47931965_47946798del (hg19). More than 71% of the encoded amino acids in the indicated transcript were deleted. Further, this deletion affects the 3′ end of the gene and removes the stop codon, likely resulting in nonsense-mediated or non-stop decay (Vasudevan et al., 2002) and resulting in degradation of the transcript. Sequencing of this case yielded high read depths at the target locus (Table 4.2).
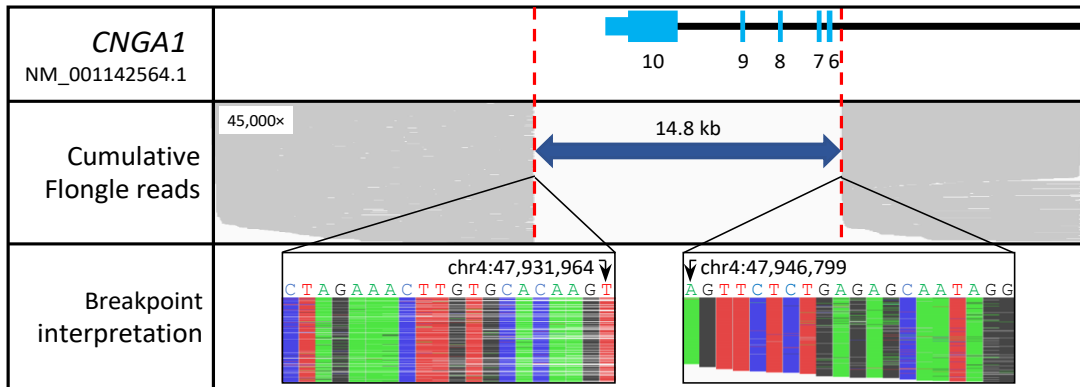
Figure 4.2 Nucleotide resolution of the breakpoint of large deletion in *CNGA1*.

The top track displays a schematic representation of the deletion at the target locus. The middle track displays the nanopore reads as viewed in IGV. The displayed reads are a 1% downsample of the total yield. The bottom panel displays the nanopore data providing nucleotide resolution of the breakpoint. *CNGA1* is encoded on the minus strand. Exons are numbered according to transcript NM_001142564.1 Genomic coordinates are provided according to hg19.

Table 4.2 Sequencing metrics for nanopore sequencing of PCR products spanning large deletions in IRD cases

Proband	Case number	Gene	Reads generated (thousands)	Median Q score *	Read length N50	Proportion of reads on target (%)	Maximum read depth at target	Read Depth at target
4755	Case 1	*CNGA1*	269.91	11.4	9,458	73.5	44,936	23,810
2249	Case 2	*CNGB1*	325.54	12.7	4,432	99.8	273,742	202,575
644	Case 3	*EYS*	327.21	13.8	4,203	99.9	250,203	224,065
769	Case 4	*PRPF31*	378.39	14.1	5,799	62.2	66,304	44,365
F1265	Case 5	*NDP*	197.66	11.9	2,993	0	0	0

*Of reads following length and quality filtering. Filtering on a quality score greater than 10 was used. All reads less than 1 kb were removed.

Table 4.3 Long-range primers used to amplify deletions in IRD cases

Proband	Case	Gene	Primer Set	Forward Sequence	Reverse Sequence	Target	Reference size (bp)
4755	*Case 1*	*CNGA1*	1	CCTAAACAGGGAAAGGGCTC	CCACCTGCATAGCCACTGTA	chr4:47925704-47951144	25441
			2	**TCGCTGAGTGAGAAAACAAACA**	**ACCACCTGCATAGCCACT**	**chr4:47925813-47951145**	**25333**
2249	Case 2	*CNGB1*	**1**	**CCTCCAGCTCAGTTCCTTGA**	**CACTGGCATCTTCGTGTCTG**	**chr16:57935182-57949096**	**13915**
			2	CCAGACATACATACCGGCCA	GAATGCTCACCGTCCTGTTC	chr16:57935577-57947025	11449
644	Case 3	*EYS*	1	GTGATGTGGTTGCTGTTGGT	CCCCAAAGCAGAGTCAACAG	chr6:65526313-65640828	114516
			2	**GCCCATCCAGCCCTAACATA**	**AGGTTTGCTCCCAGATGACT**	**chr6:65528846-65640635**	**111790**
769	Case 4	*PRPF31*	**1**	**CACCGAATCCCACTCTTCCT**	**CTTGGCGATGGTCTTGACTG**	**chr19:54613265-54621817**	**8553**
			2	CCTCTCCTCCAACACCGAAT	GCAGGAGAGACAGGAGATGG	chr19:54613253-54622060	8808
F1265	Case 5	*NDP*	1	AGGCTTGCCCTCCACCATTA	AGCCCTCAAACTTGAATGGGTC	chrX:43806190-43812786	6597
			2	**GCTTGCCCTCCACCATTATAGA**	**AACTTGAATGGGTCCAGAGAC**	**chrX:43806192-43812778**	**6587**
			3	AGGAAGTCTGTGAGTTGGCA	ATGTGTTGGGTGGACTAGCA	chrX:43805928-43813100	7173
			4	ACATGGTATCAGGAGGCAGG	ATAGGTTTCAGTGGTGCCCA	chrX:43805790-43813066	7277

Bold indicates PCR products which were sequenced on Flongle flowcells. All coordinates refer to hg19.

4.2.2.2 Characterizing a deletion in *CNGB1*

Case 2 (Proband 2249) had a diagnosis of recessive RP. A homozygous deletion of exons 25-27 of *CNGB1* had been identified in a previous study using ExomeDepth (Plagnol et al., 2012) on MIPs screening data (Weisschuh et al., 2018) (Table 4.1). From these data, the minimum size of this deletion was estimated to be 8 kb and the maximum size was estimated to be 11 kb. The deletion was confirmed using standard PCR targeting exon 26 (Forward: CAGTCTAGTGCCCTTTCCCA, Reverse: GCCTGTGATGACCCATGTTC, Target: chr16:57938402-57938874 hg19, 472bp) which resulted in no PCR product when compared to a control sample (data not shown). Following this, the deleted allele was amplified by long-range PCR. Two pairs of primers were designed to flank the deletion. Primer pair 1 successfully yielded an estimated 5 kb product. This was significantly smaller than the 14 kb product expected from the reference sequence (Figure 4.3).
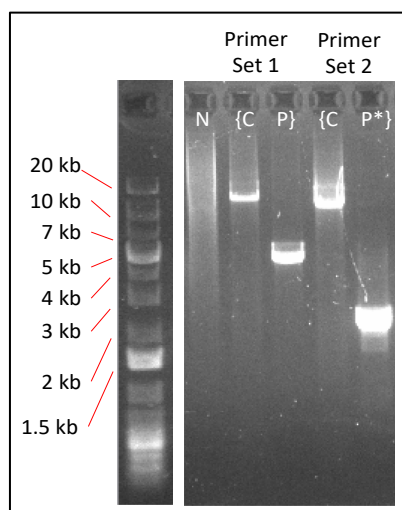
Figure 4.3 Gel electrophoresis of long-range PCR product spanning putative *CNGB1* deletion.

Long-range PCR using the SequelPrep™ kit at an annealing temperature of 58 °C resulted in an approximately 14 kb product from amplification of human genomic control (C) using both primer sets. Use of primer set 1 with the patient DNA resulted in a 5 kb product (P*) which was sequenced and a 2.5 kb product (P) which was not. Both primer sets were used in the no template control (N). The size of the products is indicated by a Thermo Scientific GeneRuler 1 kb plus ladder.

Nanopore sequencing on this product revealed a novel 9.4 kb deletion encompassing exons 25-27 of *CNGB1,* (NM_001286130.2) g.57937451_57946811del (hg19). (Figure 4.4). This is predicted to be an out-

of-frame deletion in *CNGB1*. The new reading frame introduces a premature stop codon 23 amino acids after the deletion, in exon 28, likely resulting in nonsense mediated decay (Baker and Parker, 2004; Sharma et al., 2020).
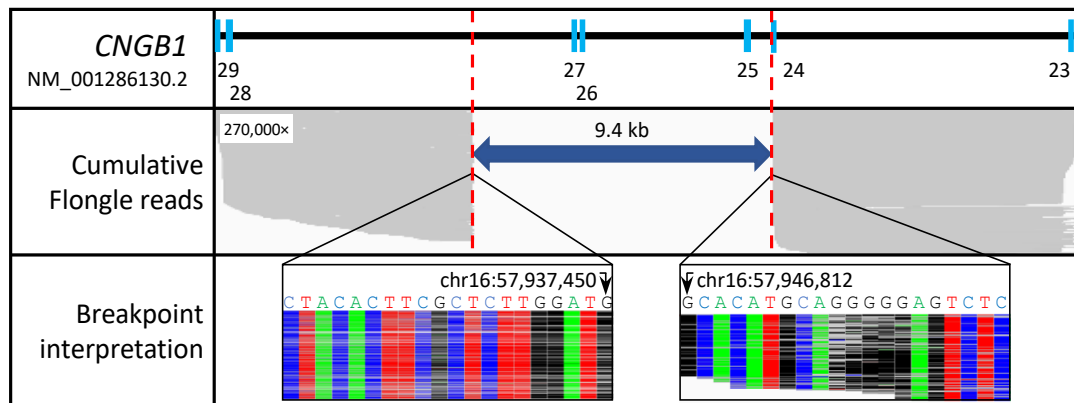
Figure 4.4 Nucleotide resolution of the breakpoint of large deletion in *CNGB1*.

Long-read data characterising a deletion in *CNGB1*. The top track shows a representation of the deletion at the target locus. The middle track shows the nanopore reads as viewed in IGV. The displayed reads are a 1% downsample of the total read count. The bottom panel shows the nanopore data providing nucleotide resolution of the breakpoint.

4.2.2.3 Characterizing a deletion in *EYS*

Case 3 (Proband 644) also had a diagnosis of sporadic RP. As with Cases 1 and 2, previously performed ExomeDepth indicated a homozygous deletion of exons 16-21 of *EYS*. The minimum size of the deletion was estimated to be 91 kb and the maximum size of the deletion was estimated to be 137 kb. The deletion was confirmed to be present by standard PCR targeting the intron which was indicated to be affected by the deletion (F: ATGGTGGGAGATGAAAGCCA, R: GCATTTGAGAAGCCACCACA, Target: chr6:65635027-65635338 hg19, 311bp). This confirmed the presence of the deletion and informed long range primer design. Following confirmation, long-range PCR primers were designed to flank the deletion. Primer pair 2 yielded a 5 kb product which is significantly smaller than the 112 kb predicted from reference data (Figure 4.5).
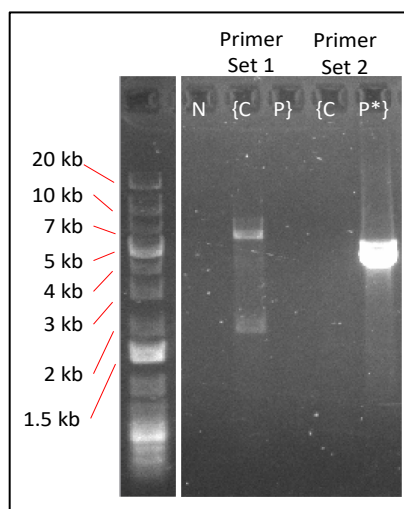
Figure 4.5 Gel electrophoresis of long-range PCR product spanning putative *EYS* deletion.

Long-range PCR was performed using the SequelPrep™ kit at an annealing temperature of 58 °C. Amplification of primer set 1 using control genomic DNA (C) resulted in a 7 kb product and no product from amplification of the patient DNA (P). Amplification using primer set 2 resulted in a 5 kb product (P*) from the patient DNA, which was sequenced on a Flongle flowcell, and no product from the control. Both primer sets were used in the no template control (N). The size of the products is indicated by a Thermo Scientific GeneRuler 1 kb plus ladder.

Nanopore sequencing of this product resulted in characterisation of a novel 107 kb deletion encompassing exons 16 to 21 of *EYS,* (NM_001292009.1) g.65529289_65636754del (hg19) (Figure 4.6). This is also predicted to be an out of frame deletion. The new reading frame in *EYS* introduces a premature stop codon four amino acids after the deletion in exon 22, likely resulting in nonsense mediated decay and no functional protein (due to transcript degradation). Additionally, this deletion overlaps with similar previously identified pathogenic deletions in *EYS* (Abd El-Aziz et al., 2008; Itsara et al., 2009; Cooper et al., 2011; Collins et al., 2020).
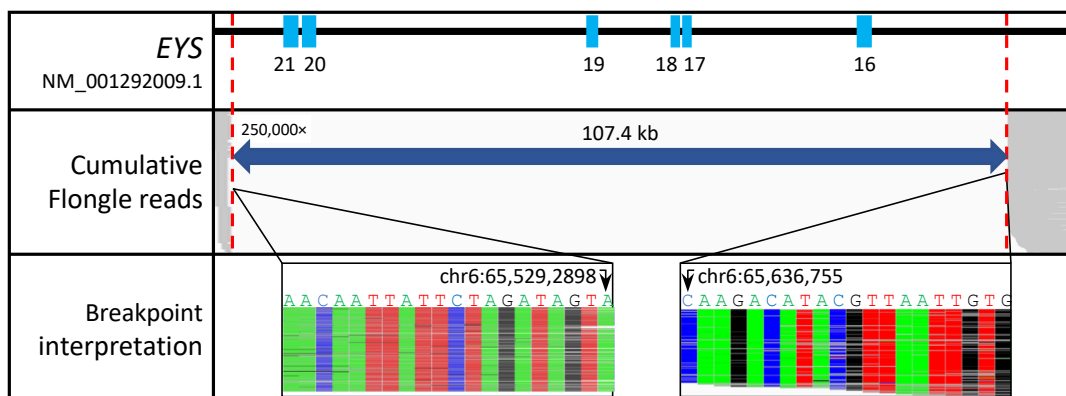
Figure 4.6 Nucleotide resolution of the breakpoint of large deletion in *EYS*.

Long-read data characterising a deletion in *CNGB1*. The top track shows a representation of the deletion at the target locus. The middle track shows the reads as viewed in IGV. The displayed reads are a 1% downsample of the total yield. The bottom panel shows the nanopore data providing nucleotide resolution of the breakpoint.

4.2.2.4 Characterizing a heterozygous deletion in *PRPF31*

Case 4 (Proband 769) had a diagnosis of autosomal dominant RP. Prior to this study, ExomeDepth was used on WES data which identified a putative heterozygous deletion of the first non-coding exon of *PRPF31* (Table 4.1). The deletion was estimated to be a maximum of 8.6 kb based on available sequence data. The deletion was confirmed by long-range PCR across the suspected locus. A band of alternate size was visible in the affected heterozygous case and an affected family member compared to the control using primer pair 1 (Figure 4.7).
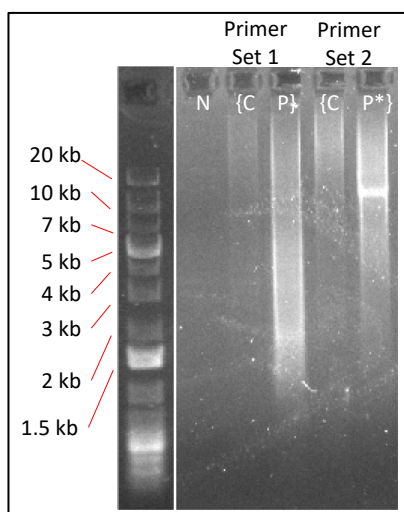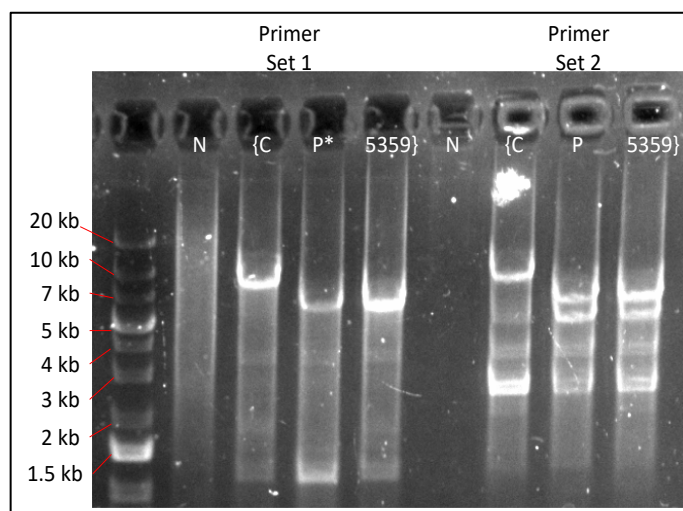
Figure 4.7 Gel electrophoresis of long-range PCR product spanning putative *PRPF31* deletion.

Long-range PCR was performed using the SequelPrep™ kit at an annealing temperature of 58 °C. In this case, the proband (P, Proband ID 769) and an affected family member (5359) were both tested. Amplification using primer set 1 resulted in an approximately 10 kb product in the control (C) compared to an approximately 7 kb product in the affected patients. Amplification using primer set 2 resulted in multiple products from both the control and the patients. The amplification product from the proband using primer set 1 (P*) was sequenced using a Flongle flowcell. No products were visible in the no template control (N). The size of the products is indicated by a Thermo Scientific 1 kb plus ladder.

Nanopore sequencing of the PCR products characterised a 2.4 kb heterozygous deletion encompassing exon 1 of *PRPF31* (NM_015629.4) as well as exons 1 and 2 of the overlapping gene *TFPT,* (NM_013342.4) g.54617206_54619550delinsG (hg19) (Figure 4.8). As the deleted allele is smaller than the wild type allele, it will preferentially amplify and sequence. As such, approximately 90% of the reads at this location originated from the deleted allele. Because of this no wild type product was visible during gel electrophoresis. The deletion appears to be novel although pathogenic deletions encompassing the first non-coding exon of *PRPF31* have recently been reported (Ruberto et al., 2021). This has been predicted to result in reduced transcription of *PRPF31*, resulting in pathogenic haploinsufficiency (Ruberto et al., 2021).
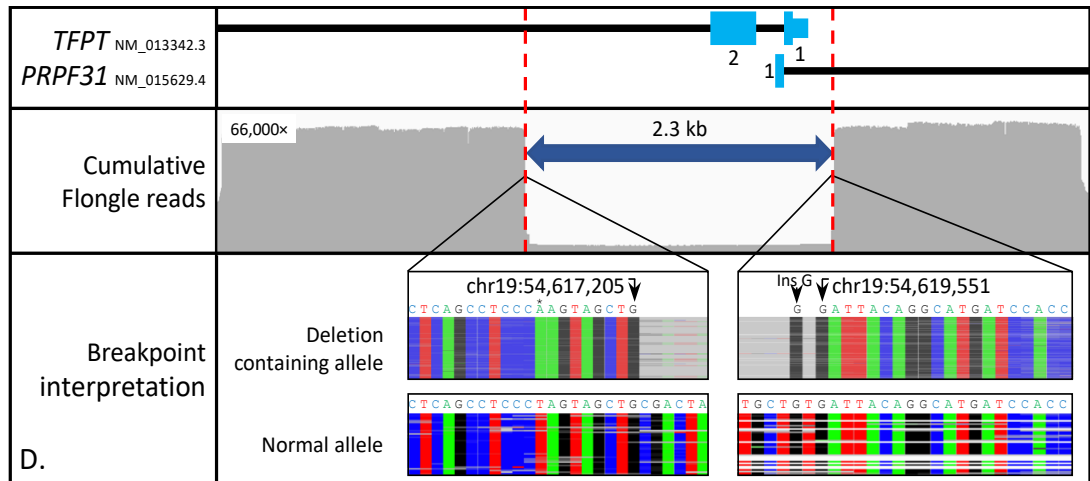
Figure 4.8 Nucleotide resolution of the breakpoint of large deletion in *PRPF31*.

Long-read data characterising deletion *PRPF31*. The top track shows a representation of the deletion at the target locus. The middle track shows the cumulative reads as viewed in IGV. The displayed reads are a 1% down sample of the total yield. The bottom panel shows the nanopore data providing single nucleotide resolution of the breakpoint. The variant containing allele and unaffected allele have been separated. A point mutation on the same allele is indicated by a *. An insertion of a G is indicated by ins G.

A long-range PCR assay was then used to segregate this deletion. A primer set was designed to more robustly capture both alleles (F: GAGACAAAGGGGTGACAGGA, R: AAACTTGCCCCGACAAAGTC, 7153bp in wild type, 5464bp expected in deleted allele). This was used to segregate the deletion in available family members. During segregation, it was observed that this pedigree shared a rare second name and phenotype with a seemingly unrelated pedigree. The segregation assay confirmed that this pedigree shared the same deletion and are likely to be related (Figure 4.9).
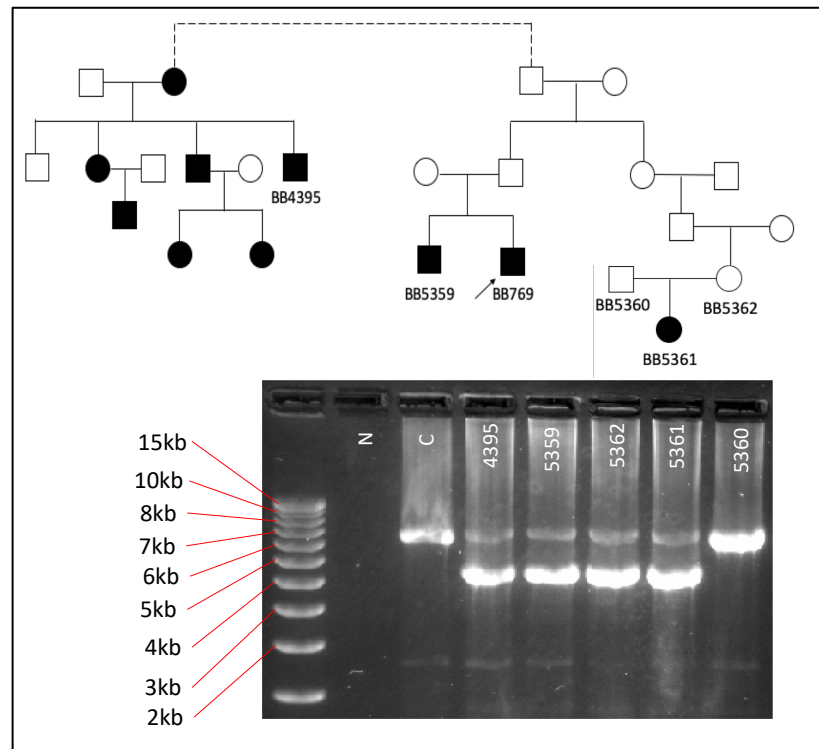
Figure 4.9 Segregation analysis of *PRPF31* deletion.

A long-range PCR assay was developed to segregate the heterozygous deletion in *PRPF31*. This assay consisted of a long-range PCR spanning the characterised deletion. Amplification of this resulted in an approximately 7 kb product amplified from the wild type allele and an approximately 5.5 kb product amplified from the affected allele. This confirmed the deletion was present in all affected cases, and was present in a single unaffected case, demonstrating the reduced penetrance of *PRPF31* associated RP. An additional pedigree was identified with the same second name and a similar phenotype. Segregation analysis of BB4395 demonstrated the deletion was also present in this case. No template control is indicated by N, positive control is indicated by C, fragment size indicated by an NEB 1 kb plus ladder. The proband is indicated by an arrow (BB769).

4.2.2.5 Confirming the deletions by Sanger sequencing

All of the characterised deletions were confirmed by Sanger sequencing of the breakpoint in the proband. This was performed by reamplifying the long-range PCR products sequenced on the nanopore using the primers listed in Table 4.4. Sanger sequencing was also performed on standard PCR products of the breakpoint in a control. This confirmed that in all cases, the nucleotide resolution of the breakpoint obtained through nanopore sequencing was accurate (Figure 4.10).

Table 4.4 Primers used for Sanger sequence confirmation of breakpoint characterisation

The primers which were used for Sanger sequencing are highlighted in bold. All coordinates refer to hg19.

Case ID	Case	Target Gene	Breakpoint	Forward	Reverse	Location	Reference size (bp)	Expected size with deletion (bp)
			Deletion breakpoint	**AATGGGGTCAAGCATCAGGT**	TCATGGGTCTTGGCTTCTCC	chr4:47931825-47947354	15530	454
			Wild type Upstream	**AATGGGGTCAAGCATCAGGT**	AGGGGACATGGGACTCTAGT	chr4:47931825-47931976	472	N/A
4755	Case 1	CNGA1	Wild type Downstream	AGAGGGCTCAAAGGATGCAA	**AAGCCACCATGCCCTACTAG**	chr4:47946670-47947115	445	N/A
			Deletion breakpoint	**TTCGGCCTCACACCTGTAAT**	ATGGAGTTTCTGGCCCCTAC	chr16:57937123-57947074	9952	478
			Wild type Upstream	**TTCGGCCTCACACCTGTAAT**	TGTCTCTGAGGTCCCGAGTA	chr16:57937123-57937629	507	N/A
2249	Case 2	CNGB1	Wild type Downstream	TTCGCTGCAATGCTGGTAAG	CCTCCTCTCCCTCCAGTACA	chr16:57946260-57946914	446	N/A
			Deletion breakpoint	**GCATTGCACCATAGAGAAAGC**	TGCAAAGCCATACATCAAGACA	chr6:65529027-65636958	107932	466
			Wild type Upstream	**GCATTGCACCATAGAGAAAGC**	CACACCGCTTACCTAGATCCT	chr6:65529027-65529968	963	N/A
644	Case 3	EYS	Wild type Downstream	N/A	N/A	N/A	N/A	N/A
			Deletion breakpoint	**AGACCTGATGCTGTTCCTCC**	AGCCTGGCCAACATAGTGAA	chr19:54616933-54619799	2867	473
			Wild type Upstream	**AGACCTGATGCTGTTCCTCC**	ACCCTTCTCCTCTCCTCCAT	chr19:54616933-54617466	543	N/A
769	Case 4	PRPF31	Wild type Downstream	TGGGTCTGGGAGAAGAAGTG	**ACTTTTGGGTAAACAGGCTGT**	chr19:54619220-54619910	691	N/A
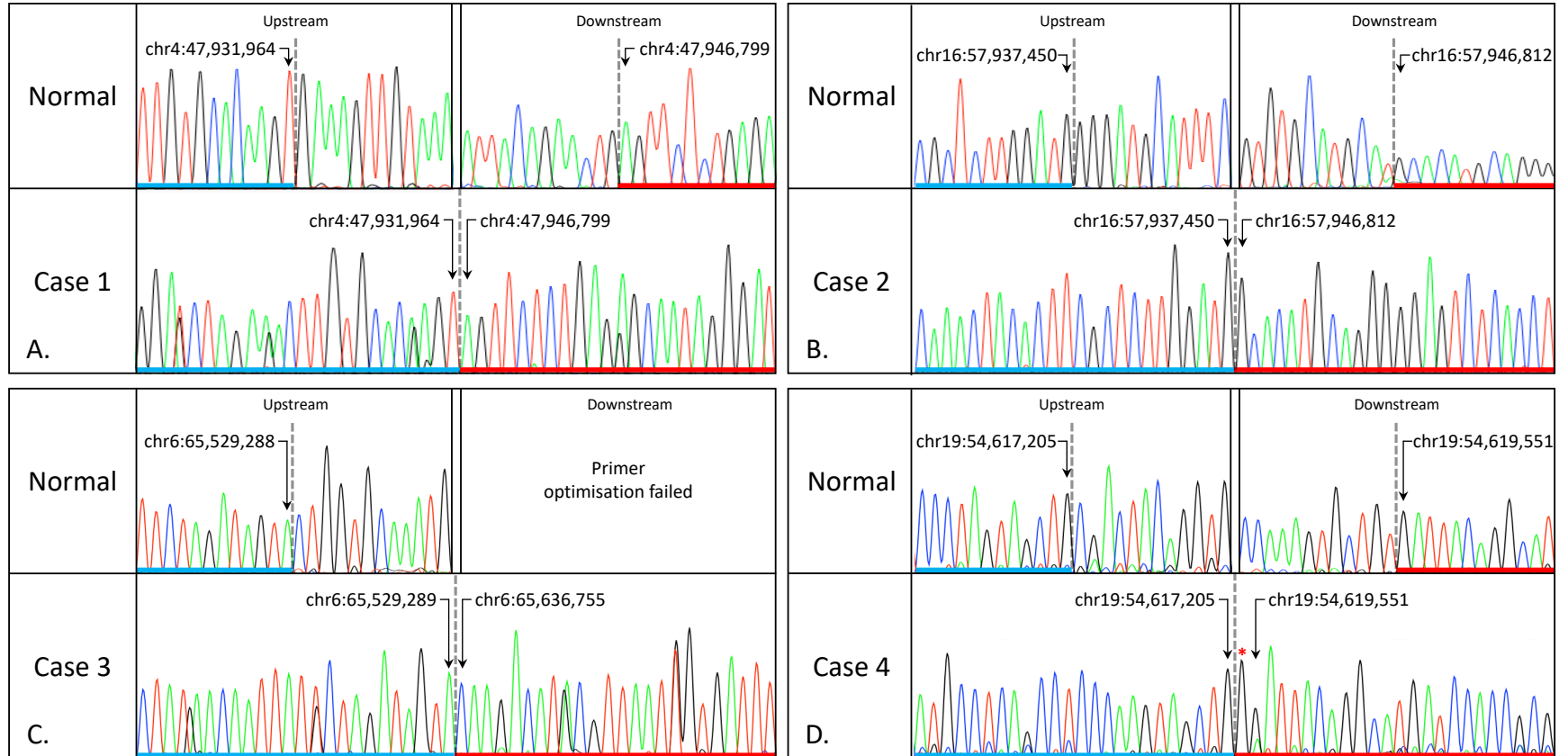
Figure 4.10 Sanger sequence verification of breakpoints characterised by nanopore sequencing.

For all cases, the breakpoints were verified using nested PCR and Sanger sequencing. This confirmed the nucleotide resolution obtained by nanopore sequencing. An inserted G is indicated by *.

4.2.2.6 Attempts to characterise a deletion in *NDP*

Despite this success, this assay was not successful in all examined cases. Case 5 (Proband F1265) had a diagnosis of Familial Exudative Vitreoretinopathy (FEVR). ExomeDepth analysis previously performed on WES data indicated a hemizygous deletion affecting exon 3 of *NDP* (Table 4.1). Prior attempts at characterising this deletion by PCR and Sanger sequencing were unsuccessful (Figure 4.11). The deletion was estimated to be a maximum of 55 kb based on available sequence data. The deletion was confirmed by standard PCR of exon 3 (Forward: CCAAACACTGACAGCCTGAC, Reverse: CCGGTACATCCTCTCCTGTC, chrX:43808688-43809094 hg19, 406bp).
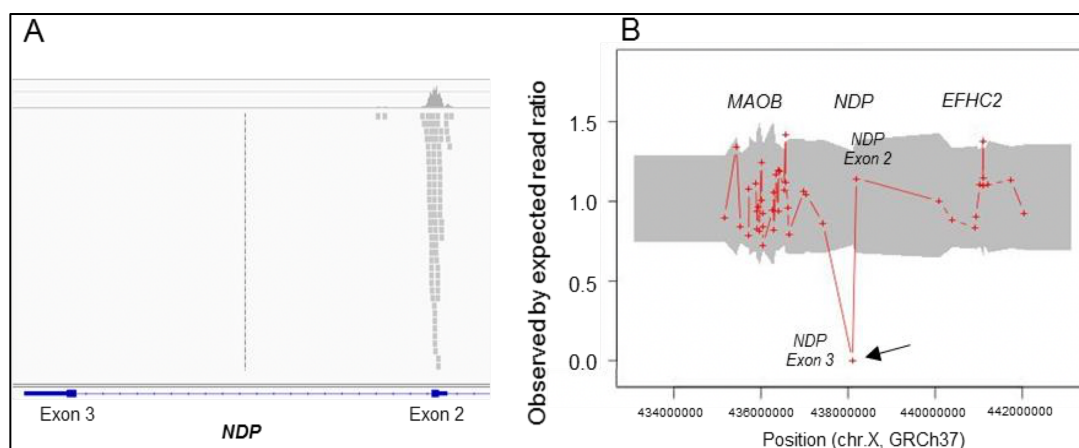
Figure 4.11 Exome depth analysis of *NDP* deletion

> The deletion in *NDP* was discovered by prior ExomeDepth analysis of WES sequence data. Reproduced with permission from (Panagiotou, 2018)

A number of standard PCRs were performed in the inter-genic sequence between exon 3 of *NDP* and *MAOB* to inform the design of long range PCR primers (Figure 4.12)*.* Full primer details available in Appendix 3 This indicated the maximum size of the deletion was 9 kb. Two long-range PCR primers were designed spanning the putative deletion Table 4.2). Two products of approximately 3 kb were obtained using primer set 2, in comparison to the 7 kb product obtained in the control case (Figure 4.13).
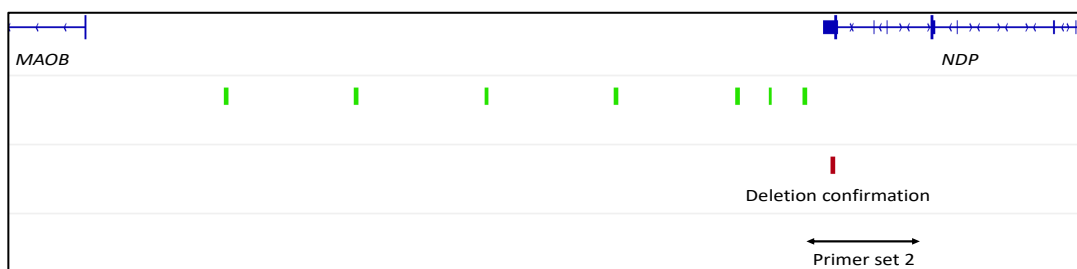
Figure 4.12 Standard PCRs to confirm deletion in *NDP*

A standard PCR was performed to confirm the presence of the deletion (red bar, bottom track) in *NDP* (blue, top track). Following this, seven standard PCRs were performed to inform long-range primer design (green bars indicate PCRs which were successful in Case 5). This indicated the maximum size of the deletion was 9 kb.
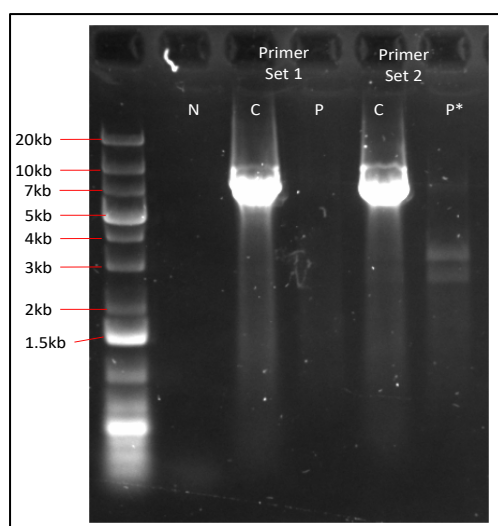
Figure 4.13 Off target long-range PCR of *NDP* deletion

Re-optimization of the primers (annealing temperature 54 °C) designed to span the deletion in *NDP* resulted in two products of approximately 3 kb, smaller than the 7 kb reference allele. However, nanopore sequencing revealed that this product was off target. No template control is indicated by N. Positive control is indicated by C. PCR products resulting from amplification of the case are indicated by P, products which were sequenced are indicated by P*. Fragment size indicated by a ThermoFisher GeneRuler 1 kb plus ladder.

This case was sequenced on a Flongle, however no reads were obtained at the region of interest. By using Samtools Depth (Section 2.7.2) it was found that high numbers of off target reads were obtained at *ITPK1:* NM013216.6, chr14:93406068-93582263 (maximum coverage of 39,669 ×), and 1.5 kb downstream of *LRRC4C*: NM_020929.2, chr11:40135751-41481186 (maximum coverage of 45,478 ×). Further attempts at PCR design and optimization did not yield a unique product (Table 4.2).

4.2.3 Repeat elements and deletion formation in IRD cases

The nucleotide resolution characterisation of the breakpoints in the successful cases allowed for association of the deletions with nearby repetitive elements through UCSC and IGV (section 2.7.1). Examination of the deletion in Case 1 found that the breakpoints in *CNGA1* overlapped two common repeats, a long interspersed nuclear element (LINE) and a large long terminal repeat (LTR) (Figure 4.14a). Additionally, the LINE element was flanked by smaller LTRs. While the repeat elements on either side of the breakpoint shared regions of microhomology, examination using BLAST did not reveal any large homologous regions making it unlikely non-allelic homologous recombination was responsible for the deletion in this case.
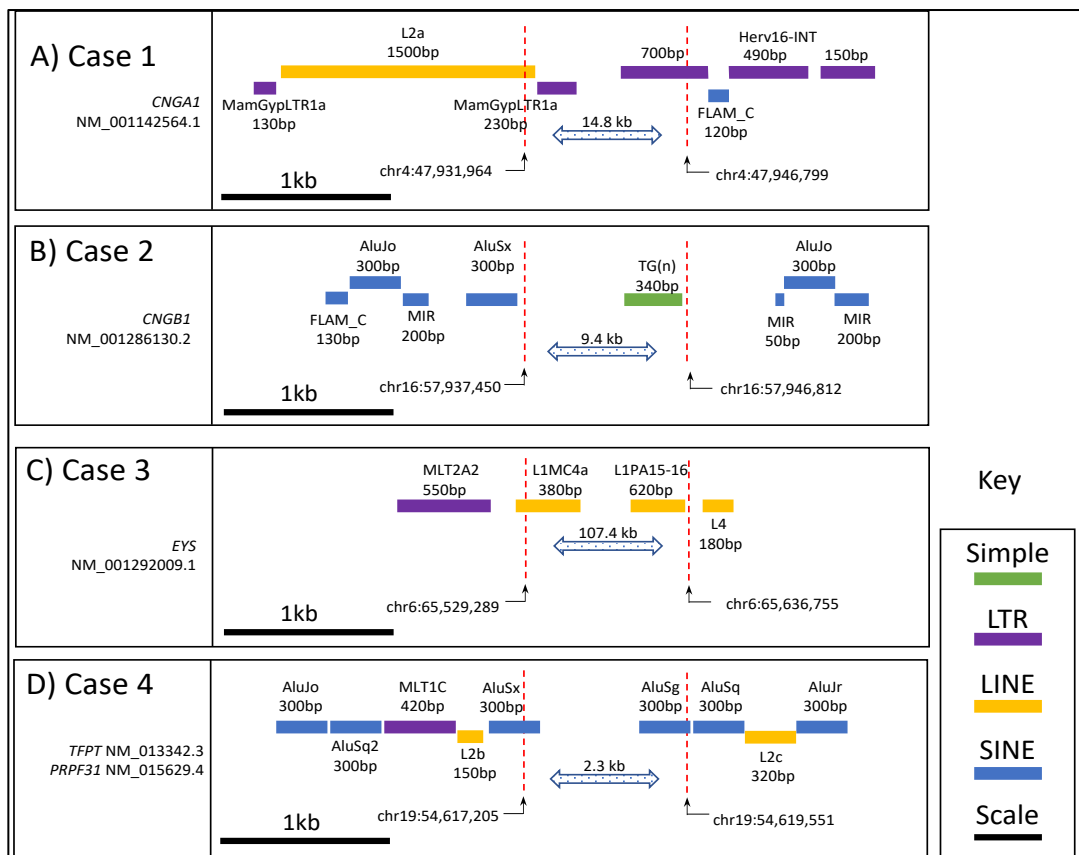
Figure 4.14 High copy repeats and repetitive elements at the breakpoints of characterised deletions in IRD cases.

The nucleotide resolution of the breakpoints in Cases 1-4 allowed for characterisation of the breakpoints in these cases. All breakpoints (indicated by red dashed lines) were situated in or near to repeats and repetitive elements (coloured bars). All indicated distances are approximate.

Analysis of the deleted region in DECIPHER found that the region overlaps with a number of large pathogenic copy number variants, however no CNVs of similar size were found at this location (Figure 4.15) (www.deciphergenomics.org, accessed 7/3/2023). It can be speculated that the repeat elements at this locus formed secondary structures, or that there was a double stranded break at this locus which was repaired using an error prone mechanism due to the homology present.
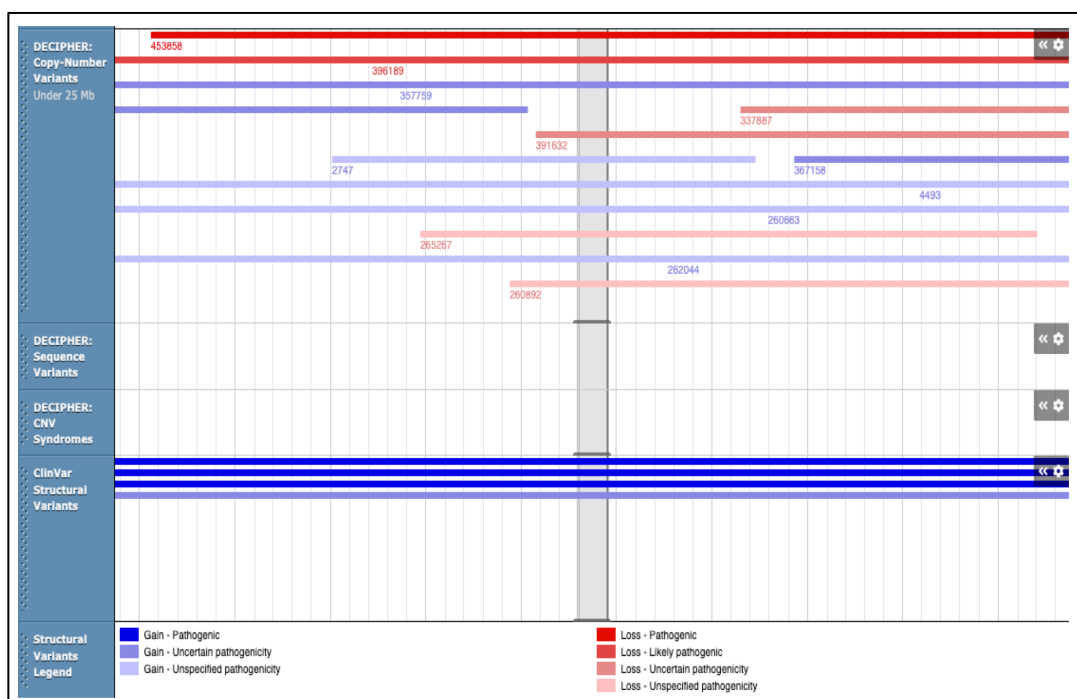
Figure 4.15 Known CNVs at deletion in *CNGA1* through DECIPHER

Examination of DECIPHER revealed a number of pathogenic and likely pathogenic deletions (red bars) in *CNGA1*, however none of these were of a similar size to the one observed in this experiment (indicated by the grey vertical bar).

Examination of repeat elements in proximity to the deletion in Case 2 found that the breakpoints in this deletion were situated in a region of *CNGB1* composed of multiple high copy repeats such as Alus and mammalian-wide interspersed repeat (MIR) elements (Figure 4.14b). There were AluJo elements on either side of the deletion which shared 72.89% sequence identity. Additionally, both AluJo elements were flanked by other repeat elements. It is possible that the Alu elements are responsible for the deletion in this case, either by Alu mediated recombination, interaction during retrotransposition of these elements or interaction during DNA repair.

Interrogation of DECIPHER found that a number of pathogenic CNVs were found at this location, but none were smaller than 1 Mb. However, a deletion was found in gnomAD originating from a single East Asian individual, DEL_16_154684 (gnomAD, accessed through UCSC 7/3/2023). Further examination of this showed that the downstream breakpoint in this deletion was 304 bases downstream of the deletion characterised in this experiment (Figure 4.16). It is possible that the common repeats at this location were responsible for the formation of the deletion in both these cases.
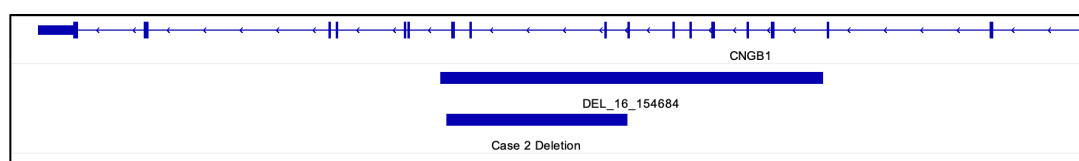
Figure 4.16 IGV image showing position of *CNGB1* deletion relative to gnomAD variant

Interrogation of gnomAD revealed a 19,826 bp heterozygous deletion in an individual in gnomAD. This overlapped the deletion found in Case 2, with downstream breakpoints 304 bases apart.

Examination of the deletion in Case 3 found that the breakpoints were in close proximity to multiple high copy repeats in *EYS* (Figure 4.14c). The numerous large high copy repeats may have reduced genomic stability at this locus, resulting in the observed deletion. *EYS* is the largest gene expressed in the human eye and genomic instability resulting in pathogenic CNVs has previously been observed (Abd El-Aziz et al., 2008) (Figure 4.17). A large deletion of exon 15-19 has been observed in a homozygous state (Abd El-Aziz et al., 2008). Additionally, a large deletion encompassing exon 12 was observed in a compound heterozygous state (Abd El-Aziz et al., 2008). Consistent with this, this deletion overlaps many other observed CNVs (Database of Genomic Variants, accessed through UCSC 21/1/22). The breakpoints observed in this case are close to breakpoints observed in other deletions in *EYS*. (Itsara et al., 2009; Cooper et al., 2011; Collins et al., 2020). This strongly indicates that the high copy repeats observed at the breakpoints in this case are likely to greatly reduce genomic stability and prompt CNV formation.
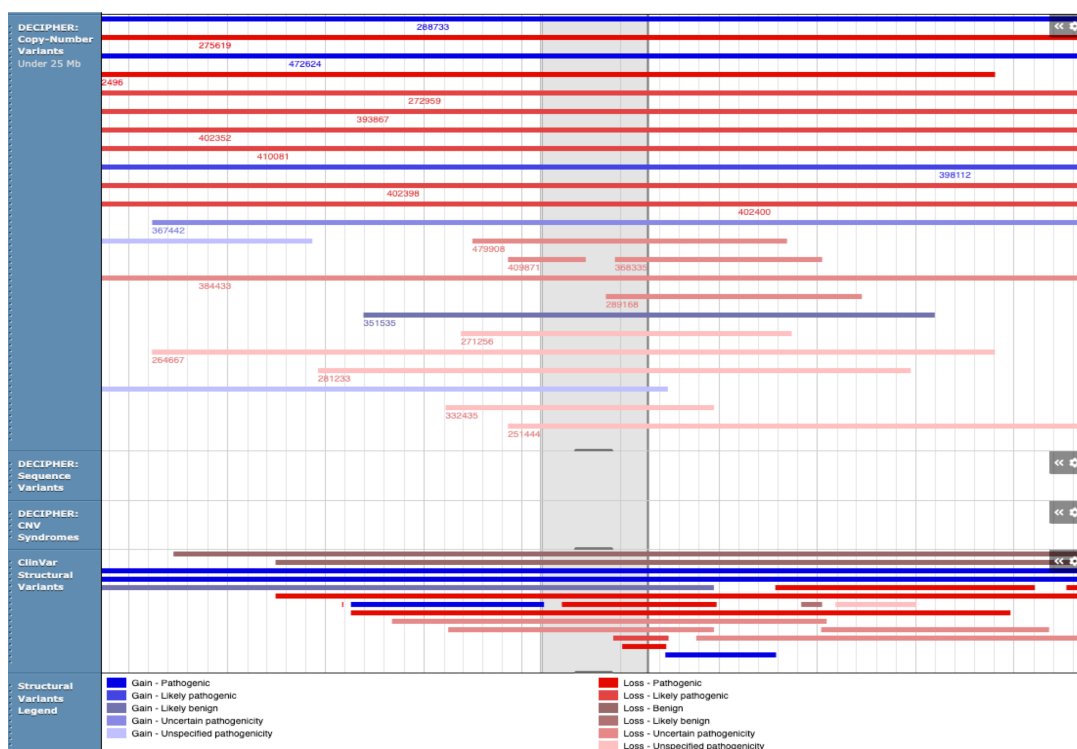
Figure 4.17 DECIPHER output showing a large number of deletions at
EYS

Interrogation of DECIPHER revealed a large number of pathogenic deletions (red bars) overlapping the deletion observed in Case 3 (grey vertical bar). Many of the breakpoints in these deletions were close to the breakpoints observed in Case 3, highlighting this as a region of reduced genomic stability.

The breakpoints of the deletion in Case 4 were situated in a region with many high copy repeats (Figure 4.14d). It is likely that the presence and orientation of these elements were the drivers of the CNV formation in this pedigree. Deletions of the first non-coding exon of *PRPF31* (as observed here) have previously been attested to cause disease by disruption of promoter sequences, thereby reducing mRNA levels and causing disease by haploinsufficiency (Ruberto et al., 2021). The presence of high copy repeats in this locus have been proposed to cause genomic instability, resulting in deletion of all or part of *PRPF31* (Sullivan et al., 2006; Ruberto et al., 2021).

4.3 Discussion

Here, a method to quickly and accurately characterise putative deletion variants detected by targeted NGS has been described. The method uses long-range PCR combined with long-read nanopore sequencing to enable single-nucleotide characterisation of the deletion breakpoints. This method was used to characterise the breakpoints of pathogenic deletions in *PRPF31*, *EYS*, *CNGA1* and *CNGB1*. Attempts to characterise a deletion in *NDP* were unsuccessful. For the cases which were successfully characterised, high read depth and unambiguous mapping of long-reads allowed nucleotide resolution of the breakpoints in all cases, even in repetitive regions and of complex breakpoints. This data was then verified by Sanger sequencing in all cases. This allowed correlation with nearby repetitive elements and assessment of the cause of the deletions. This method promises to be a useful approach to quickly characterise deletions detected by targeted sequencing.

4.3.1 Copy number variants

CNVs were first recognised during Drosophila studies, before the structure of DNA had been characterised (Tice, 1914; Bridges, 1936). Later, CNVs were also recognised as an important cause of genetic disease in humans (Jacobson et al., 1964). For a long period, CNVs were assumed to be rare and generally disease causing, however it is now known that CNVs are a frequently occurring feature of the human genome, and many CNVs represent benign polymorphisms (Sebat et al., 2004; Iafrate et al., 2004). It has been estimated that on average between 4.8% and 9.5% of the genome is affected by CNVs (Zarrei et al., 2015). This results in CNVs being responsible for 3-10 times more genomic difference between individuals than SNPs (Pang et al., 2010; Alkan et al., 2011; Sudmant et al., 2015).

4.3.1.1 CNV Formation

The mechanisms of CNV formation are typically inferred by studying the sequences at the breakpoints of CNVs. This has revealed a variety of mechanisms, many relating to DNA recombination, DNA replication and DNA repair processes, prompted by structural features at the breakpoints of the CNV (Lupski, 1998). The most common cause of CNVs are recombination events, in particular non-allelic homologous recombination (NAHR) (Shaikh et al., 2000; Stankiewicz and Lupski, 2002; Erdogan et al., 2006).

CNVs can occur during replication due to replication stress induced by metabolic conditions, DNA damage or secondary DNA structures. Interaction between sequence motifs has been shown to result in the formation of secondary DNA structures (Chou et al., 2003). These structures have been associated with errors during replication resulting in CNVs (Bacolla et al., 2001; Bacolla et al., 2004). Additionally, low copy repeats have been associated with non-B DNA structures (in particular cruciform structures) such as those bordering 22q11, indicating a possible source of CNVs at this locus (Richardson and Jasin, 2000; Edelmann et al., 2001; Kurahashi and Emanuel, 2001).

CNVs also commonly form through DNA repair processes. Double stranded breaks are common events which have a variety of causes, such as from mechanical stress placed on the cell during mitosis and from environmental stresses including ionizing radiation and reactive oxygen species (Chance et al., 1979; Fouladi et al., 2000; Lo et al., 2002; Riballo et al., 2004). When a double stranded break is encountered, a number of repair mechanisms are utilized, generally dependent on the availability of homologous DNA (Lieber, 2010). If no homology is detected, non-homologous end joining is used (Lobban and Kaiser, 1973). This frequently results in addition or loss of nucleotides, resulting in small indels and rearrangements (Lieber et al., 2003). Depending on the level of homology present at the breakage site, microhomology mediated end joining (Zhong et al., 2002) or homologous recombination can be used to repair the CNV (Smithies et al., 1971; Smith, 1972; Hanawalt, 1972).

Microhomology mediated end joining is error prone, often resulting in insertions, deletions and point mutations at the breakage (Arana et al., 2008; Mateos-Gomez et al., 2015). Microhomology mediated end joining has been linked to complex breakpoints in diseases such as cancer (Lee et al., 2007; Zhang et al., 2013; Willis et al., 2015). Regardless of the repair mechanism used, double stranded breaks can interact, resulting in pathological rearrangements. Breaks which are spatially linked but genomically distant can interact, resulting in CNVs and rearrangements (Zhang et al., 2012; Roukos et al., 2013). CNVs resulting from double strand break repair have been linked

to pathological outcomes such as cancer and chromosomal abnormalities (Willis et al., 2015).

4.3.1.2 Repeat elements and CNV formation

CNVs can be separated into recurrent CNVs and non-recurrent CNVs. Recurrent CNVs are found in unrelated individuals and across individuals are generally a similar size, share similar breakpoints and genomic content (Smajlagić et al., 2020). Recurrent CNVs arise in regions of the genome which are less stable due to the surrounding genomic architecture (Stankiewicz and Lupski, 2002). Recurrent CNVs have been associated with a range of rare disorders such as Williams-Beuren syndrome, Prader–Willi/Angelman syndrome, 16p11.2 microdeletion- and microduplication syndrome, and 22q11 deletion syndrome (Cooper et al., 2011).

In contrast, non-recurrent CNVs (such as those characterised as part of this study) typically differ in size and breakpoints but may overlap with other non-recurrent CNVs. The breakpoints of non-recurrent CNVs typically overlap high copy repeat motifs such as mobile elements. Mobile elements such as retrotransposons are 'selfish elements' which replicate throughout the genome, either autonomously or non-autonomously (McClintock, 1950). They constitute approximately 45% of the human genome and have been implicated in the promotion of NAHR resulting in non-recurrent CNVs (Beck et al., 2011; Deininger, 2011; Robberecht et al., 2013). Mobile elements have been associated with a number of human diseases. They have been shown to do this both by inserting into loci such as coding sequences, as well as by reducing genomic stability (Kazazian et al., 1988).

Long interspersed nuclear elements (LINEs) are common mobile elements in the human genome (Dombroski et al., 1991; Kazazian and Moran, 1998; de Koning et al., 2011). It has been proposed that LINEs can stimulate NAHR if they have a common homologous sequence of at least 1 kb and share 96% sequence identity in their homologous regions (Startek et al., 2015). LINEs have been implicated in pathogenic CNVs resulting from NAHR in diseases such as kinase deficiency and Mesomelia-synostoses syndrome (Burwinkel and Kilimann, 1998; Robberecht et al., 2013; Kohmoto et al., 2017).

The most common class of Short Interspersed Elements (SINEs) are Alus which constitute approximately 10% of the human genome (Batzer and Deininger, 2002; Deininger, 2011). Alus have been demonstrated to be an important driver of CNV formation. Alu elements have been reported to prompt NAHR resulting in CNV formation, resulting in a variety of diseases such as hypercholesterolemia and Smith-Magenis syndrome (Lehrman et al., 1985; Shaw and Lupski, 2005). Alus have been linked to more than 500 deletions and CNVs mediated by unequal crossover events prompted by Alus have been demonstrated in more than 70 diseases (Sen et al., 2006; Cordaux and Batzer, 2009; Konkel and Batzer, 2010; Haig H. Kazazian and Moran, 2017).

The influence of high copy repeats on non-recurrent CNV formation can be seen in the deletions characterised in this experiment. All of the deletions studied are characteristically bordered by high copy repeats such as LINEs and Alus, implicating the involvement of these elements in the deletion. As such, it is important to fully characterise detected CNVs to accurately associate the breakpoint with nearby genomic architecture and to use this to elucidate the cause of the CNV formation.

4.3.2   CNVs and disease

While CNVs are common in the genome and frequently benign, they are recognised as an important, albeit difficult-to-characterise cause of disease. Reflecting this, ACMG guidelines for categorising CNVs were published in 2020 (Riggs et al., 2020). CNVs are an important driver of human disease in many categories including susceptibility to disease, Mendelian traits and complex genetic traits (Gonzalez et al., 2005). CNV formation has been shown to cause disease through a variety of mechanisms. As can be seen in the deletions characterised in this experiment, CNV formation can directly affect the coding sequence of genes, either by affecting the entire gene, multiple exons or single exons.

Case 1 (4755) and Case 2 (2249) had deletions involving the human cGMP-gated cation channel which is a vital part of the phototransduction cascade (Section 1.2.1.3). It is heterotetrameric, consisting of an A subunit (*CNGA1*) (Dhallan et al., 1992; Dryja et al., 1995) and a B subunit (*CNGB1*) (Ardell et al., 1995; Ardell et al., 1996). Mutations in *CNGA1* have been associated with autosomal recessive RP, with characteristic progressive photoreceptor death

(Pittlers et al., 1992). The deletion observed in Case 1 encompassed 71% of the coding sequence, likely indicating a total loss of function and resulting in non-stop decay as the stop codon was deleted (Vasudevan et al., 2002).

Mutations in *CNGB1* as observed in Case 2 have been associated with autosomal recessive RP (Bareil et al., 2001; Kondo et al., 2004; Fu et al., 2013). A similar deletion was found in one individual in gnomAD (DEL_16_154684). The deletion observed in this experiment is likely a total loss of function as nonsense mediated decay is likely to occur due to an introduced stop codon shortly after the deletion.

Mutations in *EYS* as observed in Case 3 have been associated with autosomal recessive RP (Abd El-Aziz et al., 2008). Additionally, large CNVs in *EYS* as observed in this experiment have been observed to cause pathogenic total loss of function (Abd El-Aziz et al., 2008). Similarly to *CNGB1*, the deletion observed in this experiment likely results in nonsense mediated decay as a stop codon is introduced.

CNVs can also cause disease by disrupting gene regulation through a variety of mechanisms. CNVs have been observed to effect the genomic conformation of a region, leading to novel regulatory interactions. For example, deletions have been observed creating a novel topologically associated domain (TAD), resulting in pathogenic mis-regulation of GDPD1 (de Bruijn et al., 2020). CNVs have also been observed to pathologically effect regulatory RNAs. CNVs affecting small non-coding RNAs, miRNAs, piRNAs, tRNAs and long-non coding RNAs have been shown to affect regulation, including altering tissue specific expression (Sahoo et al., 2008; Marcinkowska et al., 2011; Gould et al., 2012; Iben and Maraia, 2014; Kumaran et al., 2018; Liu et al., 2019).

CNVs can also directly affect the regulatory elements of a gene. Deletions altering distal enhancers have been observed to pathogenically disrupt regulation (Ghiasvand et al., 2011). As discussed, repetitive elements have been demonstrated to form non-B DNA conformations such as cruciform, triplexes, slipped hairpins, tetraplexes and Z conformations. The formation of non-B DNA structures has been shown to cause CNVs, however these

elements have also been demonstrated to be enriched in the promotor sequences of genes, indicating secondary structures play an important regulatory role as well as prompting CNV formation (Conrad et al., 2009).

The role of regulatory elements in prompting CNV formation has been previously observed at the *PRPF31* locus as can be seen in Case 4 (Sullivan et al., 2006; Ruberto et al., 2021). *PRPF31* is a vital component of the spliceosome and is ubiquitously expressed. Complete loss of function of PRPF31 is not compatible with life, however heterozygous mutations have been associated with autosomal dominant RP with incomplete penetrance (Weidenhammer et al., 1996; Vithana et al., 2001). Incomplete disease penetrance was also observed in the pedigree examined in this experiment. Disease penetrance and severity is mediated by expression levels of the unmutated copy of *PRPF31* (Vithana et al., 2003; Rivolta et al., 2006).

4.3.3 Methods to detect CNVs

From this it is clear that CNVs play an important role in human disease and that their detection and characterisation is important to increase diagnostic rates in the study of IRDs. It is also clear that methods to detect CNVs must be able to do so at high resolution, as the characterisation of the breakpoints of the CNV is crucial for understanding their formation and role in disease and cascade testing of family members.

4.3.3.1 Cytogenetic and hybridization techniques

The development of increasingly sensitive methods to detect CNVs has reduced the size of CNVs which can be detected and characterised. Initial methods to detect CNVs were primarily chromosome banding techniques such as Q banding and G banding (Caspersson et al., 1968; Seabright, 1971). This limited the size of CNV which could be detected to large scale genomic rearrangements. While large rearrangements could be detected at high sensitivity, this was still limited to approximately 5 Mb (Haraksingh et al., 2017).

The development of hybridization techniques such as Fluorescent In Situ Hybridization (FISH) and Comparative Genomic Hybridization (CGH) greatly increased sensitivity and resolution (Langer-Safer et al., 1982; Kallioniemi et

al., 1992). These techniques use hybridized probes to compare fluorescent signals from chromosomes derived from a case of interest compared to a control to identify CNVs. A further development of CGH, array-CGH, has become a standard for CNV screening (Guo et al., 2013). During array-CGH, DNA fragments from the case and the control are fluorescently tagged and co-hybridized to an array of fixed probes. Fluorescence is then measured and compared, allowing rapid and accurate genome wide CNV detection (Solinas-Toldo et al., 1997). However, array-CGH is relatively expensive, can only detect unbalanced structural variants, and can only detect CNVs of sequences present in the reference which was used to design the probes. CNVs undetected in the reference and novel sequences inserted into the case of interest will not be detected by array CGH (Kidd et al., 2010).

Modified PCR-based methods have also been developed for the detection of CNVs. These include qPCR, where the number of copies of product generated during amplification is measured, QF-PCR where the signal strength of the targets are quantified using capillary electrophoresis and multiplex ligation-dependent probe amplification (Higuchi et al., 1993; Yau et al., 1996; Heid et al., 1996; Schouten et al., 2002).

While the methods described above are effective and are commonly used for CNV detection, they are ill suited for high resolution detection of small CNVs and also don't provide molecular level breakpoint information. They are therefore being replaced by NGS-based approaches.

4.3.3.2 Next generation sequencing and CNVs

The development of NGS has significantly altered CNV discovery pipelines. While the methods described above are still frequently used as a high throughput screening tool, particularly for larger genomic rearrangements, it is becoming increasingly common to perform NGS as a first line test. NGS based techniques are typically more sensitive than comparative methods and provide additional sequence information. It is becoming common practice to perform NGS and then use a method such as MLPA to validate potential CNVs (Moreno-Cabrera et al., 2020). Typically, when NGS is performed, it is desirable to detect small CNVs affecting a portion of a gene or a single exon. Methods to detect CNVs from NGS data can typically be divided into methods which use read pairs, methods which compare read depth such as

ExomeDepth (Plagnol et al., 2012), methods which analyse split reads and methods which assemble contigs (Alkan et al., 2011; Zhou et al., 2018). It is common practice to use a combination of methods to increase sensitivity (Coutelier et al., 2021). Methods which compare read depth can only detect unbalanced CNVs. NGS based methods which use read pairs and split reads can in theory detect all types of CNVs however they require high quality data covering the entire region of interest to do so (Singh et al., 2021)

As discussed previously, NGS typically is performed as targeted sequencing, WES or WGS. WGS is very effective at CNV detection as the entire genome will be covered. As a consequence of this, it is common to capture the breakpoints of the CNV, allowing nucleotide or near nucleotide resolution of the breakpoint. A number of tools have been developed to great effect to detect CNVs from WGS datasets using all of the above methods, such as SVRare, CoNVaDING and CODEX2 (Ye et al., 2009; Chen et al., 2009; Handsaker et al., 2011; Abyzov et al., 2011; Rausch et al., 2012; Layer et al., 2014; Yu et al., 2022). These tools have been highly successful at increasing sensitivity and throughput, however they commonly result in high numbers of putative CNVs which must be parsed and validated (Sedlazeck et al., 2018). In spite of this it has been estimated that approximately 83% of insertions are missed by even WGS (Chaisson et al., 2019).

To decrease costs, NGS is typically performed on targeted regions of the genome as an initial test. This leads to unique challenges for CNV detection. The methods developed for WGS datasets are typically accurate but are not as suitable for targeted sequencing data. While many tools have been developed to do this on targeted data (typically using comparative read depths), this is still a significant challenge (Talevich et al., 2016; Johansson et al., 2016; Povysil et al., 2017; Chiang et al., 2019). Typically, the breakpoints of a CNV are not captured by targeted sequencing, reducing the resolution of the putative CNV to exon level.

Additionally, when CNVs are detected through NGS, large amounts of putative variants are typically identified, generally due to areas of reduced coverage or quality. As such, it is common practice to verify the existence of putative CNVs through an independent method, commonly independent PCR and Sanger Sequencing. This can present its own set of challenges. For CNVs

detected through WGS, the breakpoints of the CNV may not be captured. As the breakpoints of CNVs frequently overlap repetitive regions, it can be impossible to unambiguously map short reads to these regions. As such, Sanger sequencing is generally performed to verify the existence of the CNV and to fully characterise the breakpoints. Further, the breakpoints of CNVs detected during targeted sequencing are not commonly captured, as was the case with all of the deletions characterised in this experiment. In these cases, it is generally necessary to attempt to produce a PCR product spanning the breakpoint which can then be sequenced using Sanger sequencing to allow breakpoint characterisation. As it is common for there to be no available sequence data in the introns, this typically results in a large locus which must be characterised.

For some CNVs, such as homozygous deletions, it is possible to use multiple PCRs to map the breakpoint of the deletion, a process known as PCR-walking. For other types of CNVs, such as heterozygous CNVs, it is instead necessary to attempt to blindly span the breakpoint of the CNV. This is frequently a laborious, time consuming process which is frequently unsuccessful.

4.3.3.3 Long-read sequencing and CNVs

Long reads have significant advantages over NGS for CNV detection. As the resulting reads are long, it is not uncommon to capture smaller CNVs in a single read. This is particularly the case for deletions, where the affected allele will be shorter than the unaffected allele. For example, in the method developed in this experiment, all of the characterised deletions were encompassed by single long-reads at high read depth, allowing nucleotide resolution of their breakpoints.

Long-reads are significantly better at sequencing repetitive and duplicated regions of the genome such as Alus and pseudogenes. Regions such as these are generally intractable to NGS (Mandelker et al., 2016). As has been discussed, these highly repetitive, low complexity regions are more likely to promote CNV formation. Long-reads are superior at correctly characterising CNVs involving these features. For example long-read WGS of a series of cases who were negative following WGS revealed an additional 54,900 small

CNVs on average per case, 73% of which were in regions poorly covered by NGS (Sanford Kobayashi et al., 2022).

In humans, tissue-specific genes have been shown to vary in copy number more frequently than ubiquitously expressed genes, leading to the CNV-mediated formation of tissue specific gene families such as the opsin array, haemoglobin and myoglobin and olfactory genes (Gilad et al., 2004; Young et al., 2008; Storz, 2016). Pathogenic CNVs are common at these gene clusters, however they are extremely difficult to characterise by NGS due to their highly homologous nature. For example, the middle-wavelength and long-wavelength opsins (*OPN1MW* and *OPN1LW*) are located in close proximity in the opsin array (Nathans et al., 1986; Vollrath et al., 1988; Wang et al., 1992; Wang et al., 1999; Smallwood et al., 2002). These genes share 98% sequence homology and are generally extremely difficult to sequence using NGS (Mandelker et al., 2016; Nash et al., 2022a; Haer-Wigman et al., 2022). Mutations in the opsin array, and particularly CNVs due to recombination events, cause a wide spectrum of natural variation in vision as well as a number of vision deficiencies. As genetic diagnosis is difficult, these frequently remain underdiagnosed. Long-read sequencing has been used to accurately sequence the opsin array and characterise a spectrum of pathogenic recombination (Haer-Wigman et al., 2022).

Further, methods have been developed using the properties of specific long-read sequencing platforms. For example, nanopore sequencing allows base calling in real time. As the reads generated are distributed randomly throughout the genome, it is possible to use this to infer the existence of CNVs as the run is ongoing. This has been used to detect CNVs at a similar resolution to array based methodologies after 6-12 hours of sequencing (Magi et al., 2019).

4.3.3.4 Comparison of characterising a deletion by Sanger sequencing and nanopore sequencing

The method presented here represents one component of an emerging field of long-read sequencing methods which will allow for higher throughput and more thorough detection and characterisation of CNVs. This method has a number of significant advantages over comparative techniques to characterise detected CNVs. This can best be discussed by comparison with

the methodology used by Ruberto and colleagues to characterise a novel *PRPF31* deletion (Ruberto et al., 2021). While they were able to thoroughly characterise this deletion and investigate its effects, the workflow available for use in their study was less efficient than the method presented here.

In their study, Ruberto and colleagues investigated two unrelated pedigrees with autosomal dominant RP. The probands of these pedigrees had undergone WES which was subsequently negative. To further investigate these cases, array-CGH targeting the exons of *PRPF31* (on one pedigree) and MLPA targeting *PRPF31* (on the other) were used. Array-CGH indicated a putative heterozygous deletion of exon 1 and possibly other regions of the gene. MLPA also indicated a putative heterozygous deletion of exon 1. To further characterise the putative heterozygous deletion, a series of 12 real-time PCRs were performed to map the extent of the deletion. When the extent of the deletion was mapped, a long-range PCR was performed to span the deletion. This PCR product was then sequenced by Sanger sequencing using 16 different primer sets spanning the fragment. This resulted in nucleotide resolution of the breakpoint. While this was successful and allowed for detection and thorough characterisation of the deletion in *PRPF31*, it represents a much more laborious workflow compared to that which was used here. In comparison, the deletion in *PRPF31* characterised in this experiment was initially identified through WES. Primers were then situated at the next locus with sequence which was identified as not affected by the deletion (exon 2 of *PRPF31* and exon 3 of *TPFT*). This amplicon was then sequenced on a Flongle, allowing nucleotide resolution of this deletion in a single step.

This method also has a number of other advantages compared to other methods. As the long-range PCR is generally robust, it is generally possible to situate the primers to avoid repetitive regions which are common at the breakpoints of CNVs. It is also not necessary to generate a unique PCR product as demonstrated here. During nanopore sequencing any off target products which are generated by PCR will map to different parts of the genome and so can be sequenced along with the target band. In contrast if Sanger sequencing is performed, it is necessary to obtain unique PCR products by further optimization or cutting the band of interest out of the gel.

Finally, this is a cost effective method with comparatively simple library preparation and reduced laboratory time compared to other CNV detection methods. Each case was done using a single Flongle. At a cost of approximately £70 per Flongle (https://nanoporetech.com/products/Flongle accessed 3/3/2023), this is lower than multiple rounds of qPCR and Sanger sequencing. Additionally, this method could be performed at an even lower cost per sample if multiple amplicons were sequenced on a single Flongle. While it was not tested as part of this study, if amplicons targeting different loci were used there is no reason that this could not be done without barcoding.

However, while this method is effective at characterising deletions, it is not effective at characterising insertions. While characterising a deletion, the deleted allele is preferentially amplified, resulting in high coverage of the deleted allele. In this experiment, this resulted in the wild type allele having 90% reduced read depth compared to the allele containing the deletion. In contrast, when characterising an insertion, the unaffected allele will preferentially amplify resulting in low read depth of the affected allele.

While this method was successfully used to characterise four deletions, it was not possible to amplify the putative deletion in *NDP*. While the presence of a CNV was confirmed by a number of standard PCRs, it is possible that this is not a deletion as originally thought and is instead a more complex rearrangement which was intractable to amplification. To characterise the CNV in this case, it would be useful to perform an alternate method such as WGS. However, if the CNV is a complex rearrangement, it is possible that NGS would not be suitable for characterising this. A long-read method such as ultra-long read nanopore sequencing (Section 3.2) or long-read WGS would be perhaps more appropriate for the characterisation of this variant (Bowden et al., 2019; Martignano et al., 2021).

As demonstrated, this is an effective method to characterise previously detected deletions. It is both faster and provides higher resolution than previously used methods such as aCGH or MLPA. It is also less expensive than performing WGS to fully characterise a detected deletion. In particular, this method is more efficient than PCR walking and Sanger sequencing, as can be seen by comparison with the workflow used by Ruberto and

colleagues. However, this method is primarily effective for deletions detected through targeted sequencing. As the cost of WGS continues to fall, this method will become less relevant as nucleotide resolution of the breakpoints is obtained during initial sequencing. In particular, if the cost of long-read WGS falls sufficiently for routine screening, then very high resolution characterisation will be routinely possible. In the interim, this method is a cost effective and efficient tool to characterise deletions.

4.4 Summary

In this chapter, a novel method to cost-effectively characterise deletions to a nucleotide resolution has been described. This was used to characterise four pathogenic deletions in IRD-related genes, however an additional case was unsuccessful. Nucleotide resolution of the four characterised deletions allowed accurate associations with nearby high copy repeats in all cases, allowing for speculation as to the mechanism of deletion in these cases. This method represents a high throughput, cost effective method to characterise deletions previously detected through targeted NGS due to fast library preparation and potential for scaling to higher throughputs.

5   smMIPs Screening of Macular Dystrophy Cases

5.1   Introduction

Studies aiming to screen large cohorts of patients with IRDs have made use of increasingly advanced enrichment techniques, sequencing technologies and improved standards in variant analysis to increase solve rates. In particular, smMIPs based panels have been an effective strategy to screen large cohorts of patients (Hiatt et al., 2013) (Section 1.4.1.1). Each smMIP is a probe which can amplify a single target locus, and many smMIPs probes can be used in parallel to screen either a single large area or multiple target areas at sites around the genome. smMIPs panels can therefore efficiently screen large cohorts of patients for mutations in a panel of candidate genes implicated in a given inherited disease.

The laboratory of Prof Frans Cremers in Radboud MC have used a series of smMIPs based panels to screen candidate genes or gene regions in large cohorts of patients with IRDs (Khan et al., 2019; Khan et al., 2020). The Radboud MC group had recently developed an improved panel targeting 107 loci associated with Mendelian forms of MD and age related macular dystrophy (AMD). This panel had previously been tested on a cohort of 379 probands with a variety of MDs and was found to facilitate robust genetic screen of this patient group, confidently solving 38% of cases and highlighting candidate variants in a number of others (Hitti-Malin et al., 2022).

Cases in the Leeds retinal dystrophy cohort with phenotypes suggesting *ABCA4* disease had not been actively worked on for a number of years, as researchers in the Leeds group focused primarily on genes and phenotypes other than those relevant to *ABCA4*. As a result, there was a cohort of cases with macular phenotypes which had either not been previously screened or had had only limited screening. In collaboration with Cremers and colleagues, a series of 57 UK cases with macular dystrophies were therefore sequenced using the Nijmegen 107 gene MD smMIPs panel as part of a larger screen.

The Leeds cases to be screened were selected, DNA was obtained and DNA quality checked by the author. smMIPs library preparation, sequencing and much of the bioinformatics analysis was performed by the Nijmegen group

(Khan et al., 2019). Read depth and CNV analysis for the Leeds cohort was carried out by the author. The clinical relevance of the variants was then interpreted by the author and a selection of variants were verified and, where possible, segregated by independent sequencing. The results of this screen provided confirmation of the efficacy of this screening approach, revealed a much higher "solved" rate of 63% in the Leeds cohort, and were used to confirm correlation of specific genotypes with phenotypes in a model of *ABCA4* disease progression put forward by the Nijmegen group and others (Cremers et al., 2020). These findings were published as a collaboration between the Leeds and Nijmegen groups (Mc Clinton et al., 2023).

5.2   Results

5.2.1   Case Selection

A cohort of 57 unsolved probands were selected from the Leeds Retinal dystrophy cohort of over 4000 IRD cases, for participation in this study (Appendix 8). All cases had been diagnosed with either STGD, a STGD-like phenotype, MD or retinal dystrophy, characterized by primarily macular or cone involvement, and were diagnosed and recruited by either Mr Martin McKibbin (St James's University Hospital) or Mr David Steel (Sunderland Eye Infirmary).

5.2.2   smMIPs sequencing

As described in Section 2.8.1, genomic DNA samples were quantified by Qubit Fluorometer and diluted to a DNA concentration of 15-25 ng/µl prior to library preparation. DNA quality and quantity for all samples was verified by gel electrophoresis and comparison to lambda bacteriophage DNA (Figure 5.1). All samples that were of sufficient DNA quantity and did not appear severely degraded on gel electrophoresis were sent to the Nijmegen group for inclusion in their smMIPs screening programme.
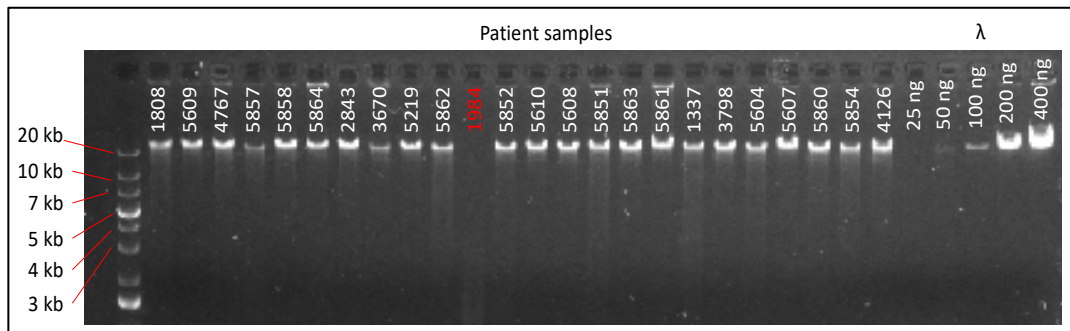
Figure 5.1 Gel electrophoresis verification of samples for smMIPs screening

> DNA quality was verified by gel electrophoresis of patient samples (50 ng of undigested genomic DNA, lanes 2-25) compared to lambda bacteriophage genomic DNA (λ, lanes 26-30 with increasing concentration) and a Thermo Scientific 1 kb plus Gene Ruler (0.5 µg, lane 1). This was done for all samples using a 0.5% w/v agarose gel. DNA quality for sample 1984 was deemed to be too low to proceed to library preparation.

Target capture was performed in the laboratory of Prof Frans Cremers (Radboud MC) using Molecular Loop Biosciences smMIPs (Section 1.4.1.2). This panel was optimized using a cohort screened prior to this study (Hitti-Malin et al., 2022). The panel used on average eight smMIPs targeting each nucleotide, resulting in in 16,973 smMIPs targeting 107 genomic loci, totalling 436,017 nucleotides of target sequence (Hitti-Malin et al., 2022). A full target list is detailed in Appendix 7 This panel also targeted loci associated with AMD, but these loci were not considered during variant analysis and interpretation carried out for this study.

This smMIP panel targeting MD loci generated excellent coverage, as calculated by the author. The 16,973 smMIP panel, used across the 57 MD samples, generated a mean of 59.01 reads per smMIP in each patient. However, each target nucleotide is targeted by up to eight smMIPs, increasing the per-base coverage and the robustness of the panel. There were 1,086 probes which yielded less than mean 1 reads across the 57 cases, but these were generally compensated for by other probes. For example, a probe targeting the 5′ region of *ABCA4* exon 18 generated an average of less than 1 read per case across the 57 cases. However due to 7 other probes targeting this locus, the mean coverage at this locus was 830 × (Figure 5.2). Despite this, even with multiple probes targeting each locus, there was uneven coverage between some targets. As an example, across the 57 cases there

were on average 115 reads per smMIP captured by the 587 smMIPs targeting *ABCA4* in each case (i.e. on average 115 reads per smMIP per sample across the 57 cases). In contrast, the 46 smMIPs targeting the highly repetitive, purine-rich ORF15 region of *RPGR* only generated mean 44.84 reads per smMIP across the 57 samples.
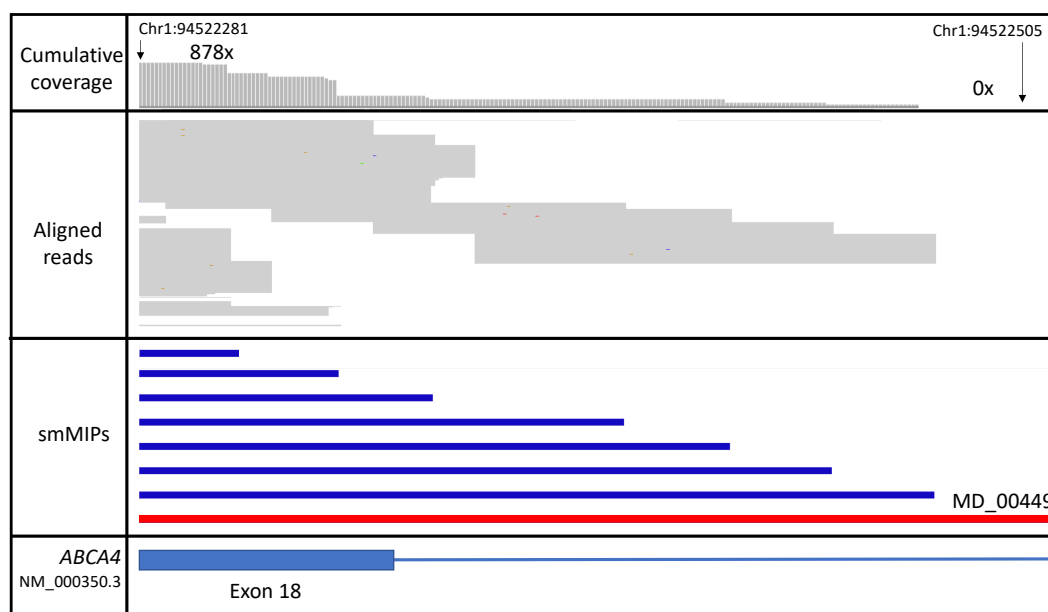
Figure 5.2 Multiple tiled smMIPs increased target coverage

> The top panel shows the cumulative reads in a sample case (5863). The middle panels shows the aligned reads for each case, and the smMIPs targeted at this locus. The bottom panel shows the location in relation to *ABCA4.* smMIP MD_00449 (highlighted in red) targeted chr1:94522281-94533505 (hg19) but generated no reads. However, the 3′ end of this target was covered by seven other probes, resulting in 878 × reads at this locus in this case.

There were an average of 1,003,127.25 reads generated by all of the smMIPs for each sample. Each smMIP targeted an average of 223.59 nt, resulting in mean 224,289,223 sequenced bases per sample. This panel targeted 436,017 target nucleotides in total, meaning the average per base coverage across the 57 samples was 514.40 ×. This corresponded to high coverage of called variants. There were 84 variants identified in this screen with a mean coverage of 721.72 × (Table 5.1).

Table 5.1 Coverage per sample in smMIPs screen

Cases with less than 400,000 reads are highlighted in bold text.

Case	Total reads	Reads per smMIP
Mean	1003127.254	59.101
01337	1446134.5	85.202
1437	1399335	82.445
1746	832798.5	49.066
1808	1482013	87.316
2190	830941	48.957
2196	1415789	83.414
2215	808847.5	47.655
2272	1227208	72.304
2469	1343871.5	79.177
2729	796678	46.938
2843	679346	40.025
3528	1003243.5	59.108
3536	925926.5	54.553
3615	1636096.5	96.394
3616	899043.5	52.969
3654	1166656.5	68.736
3656	1147160	67.587
3670	1008370	59.410
3798	905824	53.369
3941	726352.5	42.795
4030	1090890	64.272
4126	894851.5	52.722
4168	1148468	67.664
4169	801945	47.248
4172	756734.5	44.585
4181	1065906	62.800
4187	679034	40.007
4203	448584	26.429
4756	**138518**	**8.161**
4759	**377359.5**	**22.233**
4767	1002152.5	59.044
5152	873897	51.487
5202	1117740	65.854
5219	897309.5	52.867
5225	**177901**	**10.481**

5258	833166	49.088
5270	1017196.5	59.930
5274	**20504.5**	**1.208**
5349	1266184.5	74.600
5604	1768658.5	104.204
5607	1285556	75.741
5608	828047	48.786
5609	1040233	61.288
5610	524399	30.896
5851	1130620	66.613
5852	1095314	64.533
5853	1360706	80.169
5854	1443879.5	85.069
5855	1322080.5	77.893
5857	1179173	69.473
5859	1068216.5	62.936
5860	1099182.5	64.761
5861	1224102.5	72.121
5862	1393082	82.076
5863	807874.5	47.598
5864	1298420.5	76.499
5865	1018731.5	60.021

5.2.3   Variant analysis

5.2.3.1   CNV analysis

CNV analysis was performed using an excel script provided by Dr Rebekkah Hitti-Malin and Zelia Corradi (Radboud MC) (Khan et al., 2020; Hitti-Malin et al., 2022). Briefly this consisted of calculating the average per smMIP target coverage across all patients in the run and comparing each target in a patient to the average coverage of that target (Section 1.8.2) (Appendix 2). This was performed for all samples in this screen. However no CNVs were identified (Figure 5.3). There were four cases with less than 400,000 reads (probands 5225, 5274, 4756 and 4759). Due to the generally reduced coverage compared to the rest of the samples it was not possible to search for CNVs by comparing coverage in these cases.
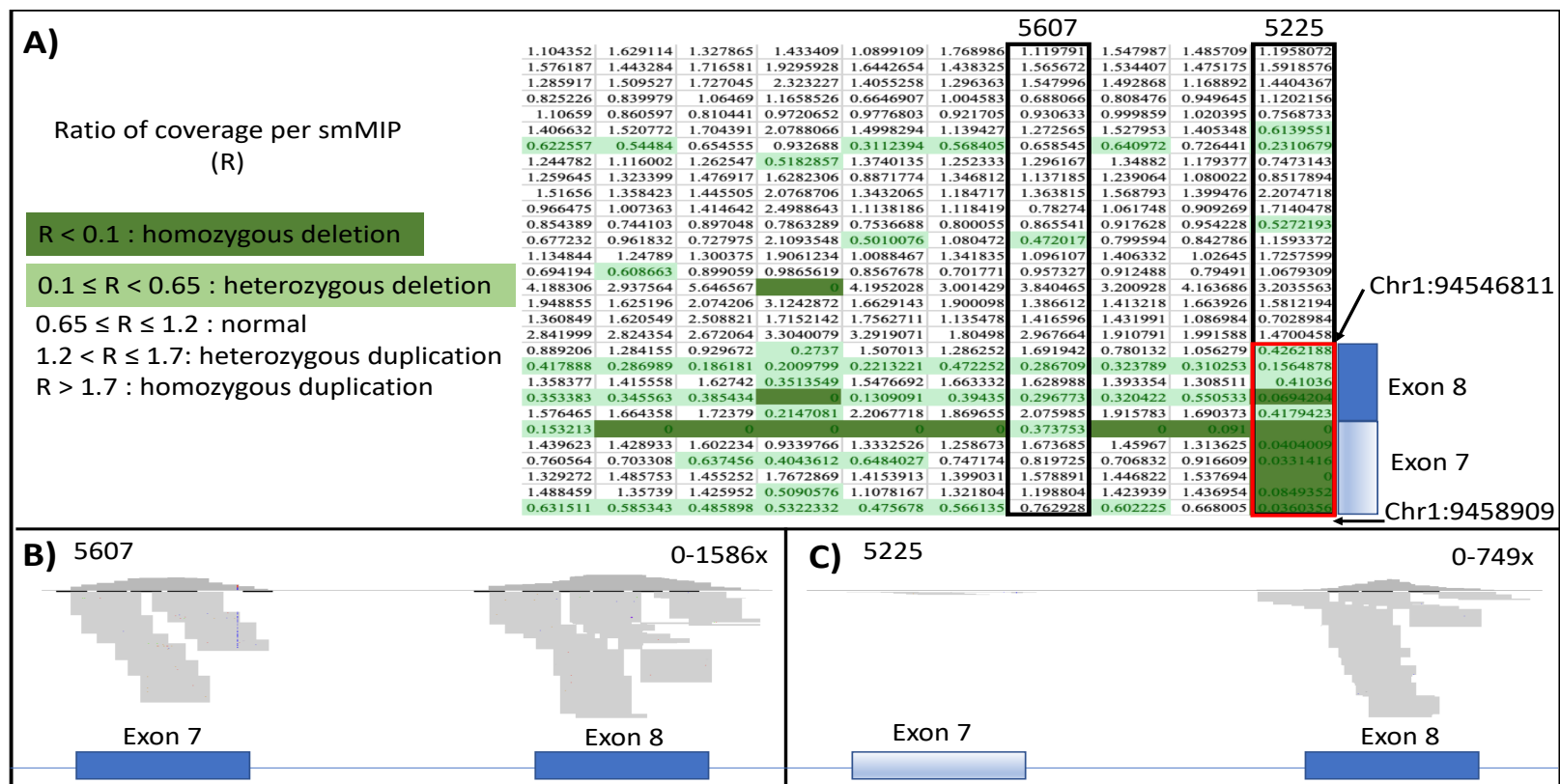
Figure 5.3 Putative deletion found through smMIPs sequencing in a case with reduced reads

A) The coverage for 30 smMIPs (rows) for 10 cases (columns) is shown. Each cell shows the ratio (R) of the observed coverage compared to the expected coverage from the rest of the run, where a value of 1 indicates the smMIP generated as many reads in that case as expected from the rest of the sequencing run. Potential duplications were examined in a separate analysis. A putative deletion of exon 7 of *ABCA4* was identified in case 5225, however this case had 177,901 reads in total compared to a mean of 1,003,127 reads per sample for the rest of the cases. At this locus, case 5607 had a max read depth of 1586 reads, compared to 749 reads for case 5225. **B)** The putative deletion was compared to a control (5607) with 1,285,556 reads. **C)** While there are very few reads aligned to exon 7 of case 5225, this case had generally amplified poorly and as such it was not possible to determine whether there was a deletion by comparing coverage.

5.2.3.2 Variant confirmation and segregation

To independently ensure the quality of the results, variants in seven cases (16/84 variants, 20% of variants) were confirmed by independent PCR amplification and sequencing (Table 5.2). Three of these cases were sequenced by long-range PCR and nanopore sequencing as part of an assay to phase the variants of interest (Section 3.2.5). The remaining four cases were sequenced by standard PCR and Sanger sequencing (Appendix 9). Examples of variants characterised by each of these approaches is illustrated in Figure 5.4. In all cases, the variants identified in smMIPs screening were confirmed by Sanger or nanopore sequencing.
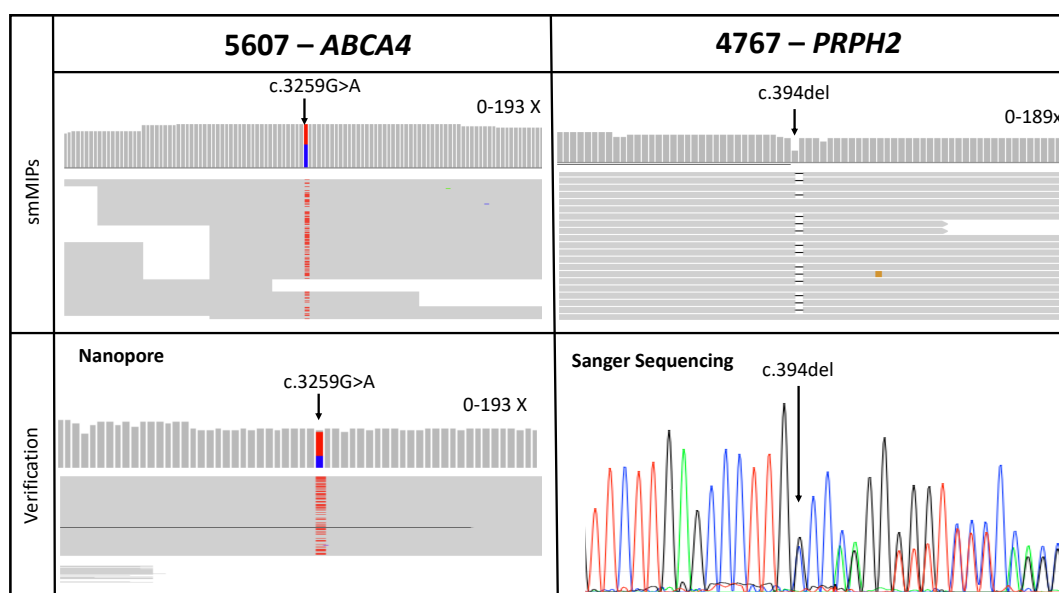
Figure 5.4 Verification of smMIPs results

Pathogenic *ABCA4* variants were found in case 5607 by smMIPs screening **(top left)**. These were confirmed to be present using long-range PCR and nanopore sequencing **(bottom left)**. A pathogenic variant was found in *PRPH2* in case 4767 **(top right)**. This was confirmed by independent PCR and Sanger sequencing **(bottom right)**.

Table 5.2 Variants confirmed by independent sequencing

Case	Gene	Allele 1 (cDNA)	Allele 2 (cDNA)	Sequencing method
1746	*ABCA4*	c.[2588G>C(;)c.5603A>T]	c.5603A>T(;)c.5461-10T>C	Sanger
3536	*ABCA4*	c.3113C>T	c.1906C>T	Nanopore
4767	*PRPH2*	c.394del	--	Sanger
5607	*ABCA4*	c.3259G>A	c.6089G>A	Nanopore
5857	*ABCA4*	c.6229C>T	c.6229C>T	Sanger
5863	*ABCA4*	c.[2588G>C(;)c.5603A>T]	c.4537del	Nanopore
3656	*ABCA4*	c.4139C>T	c.5882G>A	Sanger

Segregation analysis was not possible for most cases due to a lack of available familial DNA. For cases with a mutation in a gene associated with dominant disease, this prevented confident association of the variant in question with the disease phenotype. This also limited variant interpretation for cases with compound heterozygous variants in genes associated with recessive disease as it was not possible to phase the variants in these cases. Efforts to phase variants in *ABCA4* by long-read sequencing were unsuccessful (Section 3.2.5). It was possible to segregate the pathogenic variants in three pedigrees (Appendix 9). An example pedigree with a variant in *PRPH2* is reported in Figure 5.5.

As phasing of compound heterozygous variants was not possible, compound heterozygous variants were assumed to be in *trans* except when known complex alleles were present. As this precluded assignment of haplotypes, classified cases were classed as "Very Likely Solved", "Possibly Solved" or "Unsolved" (Table 2.1). This was also used for variants which were segregated for consistency.
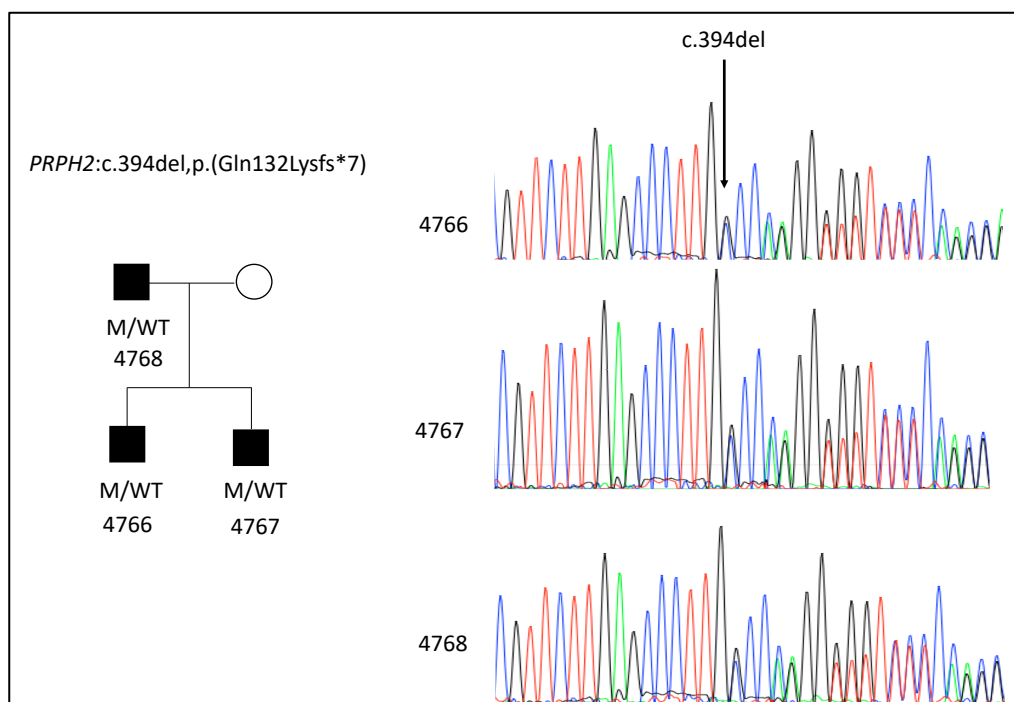
Figure 5.5 Segregation analysis of *PRPH2* variant

A pathogenic *PRPH2* variant was found in proband 4767 which segregated with the phenotype. The phenotype was diagnosed by clinical examination by Mr Martin McKibbin as autosomal dominant STGD.

5.2.4 Screen results

All variants were assessed using ACMG guidelines through Franklin. For cases with variants in *ABCA4*, variant grading published by Cremers and colleagues was used in addition (Cornelis et al., 2021) (Section 2.8.4).. This screen resulted in high solve rates, with thirty-five cases (60%) considered very likely solved, one considered possibly solved and 21 remaining genetically unsolved. All four cases with reduced coverage were unsolved. This resulted in an overall diagnostic yield of 63.2%. Solved cases with their identified variants are listed in Table 5.3.

Table 5.3 Identified variants considered to very likely or possibly solve probands

Novel variants are highlighted in bold text. As segregation analysis was not possible for most cases, variants were distributed on two alleles from 5′ to 3′ considering proximity. For cases with three or more variants, the variants which were located closer together were considered as one allele. When more than one allele is listed, the allele which is 5′ is listed as Allele 1. When variants c.5461-10T>C and c.5603A>T, or variant c.2588G>C and c.5603A>T in *ABCA4* were found in the same patient, they are reported as complex alleles, as they are found in these combinations in >95% of reported STGD cases. ACMG classifications were obtained using Franklin by Genoox. P = Pathogenic, LP = Likely pathogenic, VUS = Variant of uncertain significance, LB = likely benign. # = Segregation analysis was performed.

Case ID	Gene	Allele 1			Allele 2		
		cDNA	Protein	ACMG	cDNA	Protein	ACMG
1337	*ABCA4*	c.5603A>T(;)5819T>C	p.(Asn1868Ile)(;)(Leu1940Pro)	VUS, LP	c.6817-2A>C	p.(?)	P
1746#	*ABCA4*	c.[2588G>C;5603A>T]	p.[Gly863Ala,Gly863de];(Asn1868Ile)]	P	c.[5461-10T>C;5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	P
1808	*ABCA4*	c.[5461-10T>C; 5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	P	c.5882G>A	p.(Gly1961Glu)	P
2469	*ABCA4*	c.[2588G>C; 5603A>T]	p.[[Gly863Ala,Gly863del];(Asn1868Ile)]	P	c.[5461-10T>C;5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	P
2843	*ABCA4*	c.4016G>A(;)**5313-1_5313del**	p.(Cys1339Tyr)**(;)(?)**	VUS, P	c.6088C>T	p.(Arg2030*)	P
3536#	*ABCA4*	c.3113C>T	p.(Ala1038Val)	P	c.1906C>T	p.(Gln636*)	P
3616	*ABCA4*	c.4469G>A	p.(Cys1490Tyr)	P	c.5603A>T	p.(Asn1868Ile)	VUS
	PRPH2	c.623G>A	p.(Gly208Asp)	P	--	--	--
3656#	*ABCA4*	c.4139C>T	p.(Pro1380Leu)	P	c.5882G>A	p.(Gly1961Glu)	P
4126	*ABCA4*	c.4774-27T>C(;)5196+1137G>A	p.[?] (;)[=;Met1733Glufs*78]	LB, P	c.[5461-10T>C;5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	P
5219	*ABCA4*	c.634C>T	p.(Arg212Cys)	P	c.[5461-10T>C;5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	P
5270	*ABCA4*	c.1906C>T	p.(Gln636*)	P	c.[2588G>C;5603A>T]	p.[[Gly863Ala,Gly863del];(Asn1868Ile)]	P
5349	*ABCA4*	c.1906C>T	p.(Gln636*)	P	c.5603A>T	p.(Asn1868Ile)	VUS
5604	*ABCA4*	c.4577C>T(;)4469G>A	p.(Thr1526Met)(;)(Cys1490Tyr)	P, P	c.5603A>T	p.(Asn1868Ile)	VUS
5607#	*ABCA4*	c.3259G>A	p.(Glu1087Lys)	P	c.6089G>A	p.(Arg2030Gln)	P
5608	*ABCA4*	c.[2588G>C;5603A>T]	p.[[Gly863Ala,Gly863del];(Asn1868Ile)]	P	c.[5461-10T>C;5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	P
5609	*ABCA4*	c.4139C>T	p.(Pro1380Leu)	P	c.4139C>T	p.(Pro1380Leu)	P

5851	ABCA4	c.1282del	p.(Val428Serfs*7)	P	c.6320G>A	p.(Arg2107His)	P
5852	ABCA4	c.[2588G>C;5603A>T]	p.[[Gly863Ala,Gly863del];(Asn1868Ile)]	P	c.[5461-10T>C;5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	P
5853	ABCA4	c.365_366insCA	p.(Gly123Metfs*32)	P	c.5560G>T(;)5603A>T(;)5882G>A	p.(Val1854Leu)(;)(Asn1868Ile)(;)(Gly1961Glu)	LP, VUS, P
5854	ABCA4	c.4195G>A	p.(Glu1399Lys)	LP	c.5318C>T	p.(Ala1773Val)	P
5857#	ABCA4	c.6229C>T	p.(Arg2077Trp)	P	c.6229C>T	p.(Arg2077Trp)	P
5860	ABCA4	c.4326C>A	p.(Asn1442Lys)	LP	c.5882G>A	p.(Gly1961Glu)	P
5861	ABCA4	c.5714+5G>A	p.[=,Glu1863Leufs*33]	P	c.[5461-10T>C;5603A>T]	p.[[Thr1821Aspfs*6,Thr1821Valfs*13];(Asn1868Ile)]	VUS, P
5862	ABCA4	c.1906C>T	p.(Gln636*)	P	c.4577C>T	p.(Thr1526Met)	P
5863#	ABCA4	c.[2588G>C;5603A>T]	p.[[Gly863Ala,Gly863del];(Asn1868Ile)]	P	c.4537del	p.(Gln1513Argfs*13)	P
5864	ABCA4	c.1253T>C	p.(Phe418Ser)	P	c.1317G>A	p.(Trp439*)	P
5865	ABCA4	c.1906C>T	p.(Gln636*)	P	c.5603A>T	p.(Asn1868Ile)	VUS
3670	BEST1	c.728C>T	p.(Ala243Val)	P	--	--	--
3654	BEST1	c.889C>T	p.(Pro297Ser)	P	--	--	--
4030	C1QTNF5	c.489C>G	p.(Ser163Arg)	P	--	--	--
5258	CRB1	c.249T>A	p.(Tyr83*)	LP	c.2506C>A	p.(Pro836Thr)	LP
3615	PROM1	c.1117C>T	p.(Arg373Cys)	P	--	--	--
5610	PROM1	c.1117C>T	p.(Arg373Cys)	P	--	--	--
3798	PRPH2	c.638G>A	p.(Cys213Tyr)	P	--	--	--
4767#	PRPH2	c.394del	p.(Gln132Lysfs*7)	P	--	--	--
5855	PRPH2	c.291G>A	p.(Trp97*)	LP	--	--	--

The majority of cases were genetically explained by variants in *ABCA4* (N=27, 75%). Variants in other genes represented 25% (n=9) of the solved cases: *PRPH2*, *BEST1*, *PROM1*, *CRB1* and *C1QTNF5* were represented. Of the variants implicated in the cases considered solved or very likely solved, 56 (77.78%) were considered pathogenic by ACMG guidelines, eight were VUS, and seven variants were likely pathogenic. One variant was considered likely benign by ACMG guidelines: *ABCA4*: c.4774-27T>C. This variant was also highlighted in ClinVar as benign although the evidence for this is weak (RCV000253178.1). Further, this variant was not predicted by SpliceAI to alter splicing (score 0.03). However this variant is extremely rare in the general population (gnomAD: 4/280514 alleles, 0.00001426 allele frequency) and the variant was found in a case (4126) with three other *ABCA4* variants. As such it was reported as it may have an impact on the disease progression.

Four novel variants (not found in gnomAD Aggregate, ClinVar or LOVD) were identified in this screen, two of which were found in *ABCA4* (Table 5.4). The first of these was a 2 bp deletion identified at the start of exon 38 in case 2843 (chr1:94480246_94480247delTC hg19, NM_000350.3, c.5313-1_5313del, p.(Trp1772fs*6)), found in 0/251486 alleles in gnomAD aggregate. This mutation was found as a compound heterozygote assumed to be in *trans* with a known pathogenic premature stop codon. It deletes the in-variant splice acceptor site and was considered to be likely pathogenic by prediction using Franklin. Examination using SpliceAI predicted this variant was likely to affect splicing (score 0.91). This variant is predicted to be deleterious by Mutation Taster (196|4 Deleterious|Benign). This variant is likely to result in a complete loss of function as the resulting transcript is predicted to undergo NMD (MutationTaster). Deletion of the splice acceptor site likely causes skipping of exon 38 (Anna and Monika, 2018), resulting in a frameshift which leads to a premature stop codon three codons in to exon 29.

The second of these *ABCA4* variants was the deletion of a cytosine in exon 10, identified in case 5851 (chr1:94544220del hg19, NM_000350.3, c.1282del, p.(Val428Serfs*7)), again as a compound heterozygous variant assumed to be in *trans* with a known pathogenic missense variant. This variant was found in 0/251486 alleles in gnomAD aggregate. This variant was considered to be likely pathogenic by prediction using Franklin. Prediction using MutationTaster found this variant was considered deleterious (200|0).

This variant is also predicted to cause NMD (prediction by MutationTaster), resulting in a total loss of function.

Additionally, an ultra-rare *ABCA4* variant in exon 4 variant was found in case 5853 (chr1:94574209_94574210insTG hg 19, NM_000350.3, c.365_366insCA, p.(Gly123Metfs*32)). This variant was found in only one other individual in gnomAD (allele frequency 0.000003976). It was considered to be likely pathogenic by prediction using Franklin, while MutationTaster considered it to be deleterious (196|4). This variant is also predicted to cause NMD in the indicated transcript (prediction by MutationTaster), resulting in a total loss of function.

A novel nonsense variant was also found in case 5258 in exon 8 of *CRB1* (chr1:197391064 T>A hg19, NM_201253.3 c.2106T>A p.(Tyr702*)). This variant was found in a putative compound heterozygous state and was assumed to be in *trans* with a *CRB1* known pathogenic missense variant. It is absent from gnomAD aggregate (0/249110 alleles). This variant was considered to be likely pathogenic by prediction using Franklin and deleterious using Mutation Taster (200|0). It introduces a premature stop, likely resulting in NMD (prediction by MutationTaster).

The final novel variant was a heterozygous nonsense variant found in exon 1 of *PRPH2* in case 5852 (chr6:42689782 C>T hg19, NM_000322.5, c.291G>A, p.(Trp97*), found in 0/251302 individuals in gnomAD aggregate. This variant was considered to be likely pathogenic by Franklin and deleterious by MutationTaster (194|6) and, being a nonsense variant, is likely to result in complete loss of function.

5.2.4.1 Complex alleles in *ABCA4*

As detailed in Table 3, of the 27 solved cases with causative variants in *ABCA4*, 13 had two pathogenic *ABCA4* variants, eight had three pathogenic *ABCA4* variants (probands 1337, 1808, 2843, 5219, 5270, 5604, 5861, 5863) and, six cases had four pathogenic *ABCA4* variants (probands 1746, 2469, 4126, 5608, 5852, 5853). For cases with more than two pathogenic variants in *ABCA4*, the variants which were located closer together were considered

as a single allele. This cannot be determined with certainty without establishing the phase of the variants.

However, some known *ABCA4* variants have previously been demonstrated to almost always occur in *cis* as a single complex allele. These variants are in linkage disequilibrium and are more likely than would otherwise be expected to appear together on the same allele. The common *ABCA4*:c.5603A>T, p.(Asn1868Ile) variant has been found to be a part of two of these complex alleles; c.[2588G>C;5603A>T] (Zernant et al., 2017) and c.[c.5603A>T;5461-10T>C] (Khan et al., 2020). In cases in this screen with these variants, it was assumed that the variants previously shown to be found as a single complex allele were in *cis*. The complex allele c.[2588G>C;5603A>T] was observed seven times in this screen and c.[c.5603A>T;5461-10T>C] was observed nine times. Four probands shared the genotype; *ABCA4*: c.[2588G>A;5603A>T];[5461-10T>C;5603A>T]. Neither c.2588G>C or c.5461-10T>C were observed in a case without c.5603A>T also present. As a result, it is highly likely that in these cases the variants are arranged as complex alleles, as assumed and as previously published.

The *ABCA4*:c.5603A>T variant was present in a heterozygous state in 12/27 of the cases genetically explained by *ABCA4* variants and in a homozygous state in four of these cases. This variant was assumed to be part of a complex allele in 10/16 of the solved cases in which it was found. Additionally, there were seven unsolved cases with one *ABCA4* variant (Table 5), *ABCA4*:c.5603A>T was present in five of these cases. In total, *ABCA4*:c.5603A>T was present in 20 of 57 cases in this cohort, representing an enrichment compared to the general population (allele frequency 0.351 in this cohort, 0.04219 in gnomAD Aggregated). Based on this frequency, this variant would be expected in 2.4 cases in this cohort. This represents an 8.33-fold enrichment compared to the expected (hypergeometric p=2.01e-14). The CNV analysis was repeated for all unsolved cases with single variants, however this did not yield any additional findings.

Table 5.4 Novel pathogenic variants identified in this study

MutationTaster Tree vote Deleterious|Benign shows the weighting that a variant is considered to be deleterious or benign. CADD Phred scores ≥10 represent variants which are predicted to be top the 10% of pathogenic variants, ≥20 are the top 1% and ≥30 are the top 0.1% (Kircher et al., 2014). Coordinates refer to hg19.

Case	Gene	Reference Transcript	Coordinates	cDNA	Protein	Exon	Frequency in gnomAD	Reads total/Reads with variant	MutationTaster (Tree vote)	CADD Phred score
2843	*ABCA4*	NM_000350.3	chr1:94480246_ 94480247 delTC	c.5313-1_5313del	p.Trp1772fs*6	38	0/251486	237/127	Deleterious 198|2	33
5851	*ABCA4*	NM_000350.3	chr1:94544220 del	c.1282del	p.Val428Serfs*	10	0/251486	357/186	Deleterious 200|0	13.87
5853	*ABCA4*	NM_000350.3	chr1:94574209_ 94574210 insTG	c.365_366insCA	p.Gly123fs*32	4	1/251486	997/479	Deleterious 196|4	8.649
5258	*CRB1*	NM_201253.3	chr1:197391064 T>A	c.2106T>A	p.Tyr702*	6	0/249110	500/263	Deleterious 200|0	24.8
5852	*PRPH2*	NM_000322.5	chr6:42689782 C>T	c.291G>A	p.Trp97*	1	0/251302	326/158	Deleterious 194|6	37

Table 5.5 Findings in unsolved probands and incidental findings in solved probands from smMIPS analysis

(A) Non-ABCA4 variants identified in unsolved probands. (B) *ABCA4* frequent mild variants identified in unsolved probands. (C) Additional sequencing findings in solved probands.

	Patient ID	Gender	Status	Gene	Transcript	Inheritance Pattern for MD	Allele		Zygosity	ACMG	Frequency (gnomAD Aggregated)
							cDNA	Protein			
A	4168	M	Unsolved	*CLN3*	NM_001042432.2	AR	c.1141C>T	p.(Arg381Trp)	Het	P	0.000067
	4756	M	Unsolved	*TIMP3*	NM_000362.5	AD	c.484G>A	p.(Glu162Lys)	Het	VUS	0
	5152	F	Unsolved	*RP1L1*	NM_178857	AD, AR	c.190C>T	p.(Leu64Phe)	Het	VUS	0.0000351
	5202	F	Unsolved	*IFT81*	NM_014055.4	AR	c.899_900insT	p.(Glu301Argfs*8)	Het	P	0.000011
	5225	M	Unsolved	*EHMT2*	NM_006709.5	AR	c.38C>A	p.(Ala13Asp)	Hom	VUS	0
B	2272	M	Unsolved	*ABCA4*	NM_000350.3	AR	c.5603A>T	p.(Asn1868Ile)	Het	VUS	0.042191
	3528	F	Unsolved	*ABCA4*	NM_000350.3	AR	c.5603A>T	p.(Asn1868Ile)	Het	VUS	0.042191
	4181	F	Unsolved	*ABCA4*	NM_000350.3	AR	c.5603A>T	p.(Asn1868Ile)	Het	VUS	0.042191
	5859	F	Unsolved	*ABCA4*	NM_000350.3	AR	c.5882G>A	p.(Gly1961Glu)	Het	P	0.004564
C	3670	F	Solved by *BEST1* variants	*ABCA4*	NM_000350.3	AR	c.5603A>T	p.(Asn1868Ile)	Het	VUS	0.042191
	5219	M	Solved by *ABCA4* variants	*IMPG1*	NM_001563.4	AD, AR	c.1157C>A	p.(Ala386Asp)	Het	LP	0.000078
	5258	F	Solved by *CRB1* variants	*RP1*	NM_006269.2	AR	c.4603C>G	p.(Pro1535Ala)	Het	VUS	0.000028
	5607	F	Solved by *ABCA4* variants	*FBN2*	NM_001999.4	AD	c.1742A>T	p.(Gln581Leu)	Het	VUS	0
	5608	F	Solved by *ABCA4* variants	*BEST1*	NM_004183.4	AR, AD	c.1175G>A	p.(Arg392His)	Het	VUS	0.000064
	5857	M	Solved by *ABCA4* variants	*CNGA3*	NM_001298.3	AR	c.1279C>T	p.(Arg427Cys)	Het	P	0.00039
	3656	F	Solved by *ABCA4* variants	*CACNA1F*	NM_001256789.3	X-Linked recessive	c.5585G>A	p.(Arg1862His)	Hom	VUS	0.00003

5.2.5 Genotype-phenotype correlation

To determine whether this cohort of patients included cases with phenocopies of *ABCA4*-related disease, the phenotypes of the 27 cases genetically explained by variants in *ABCA4*-related disease were contrasted with the phenotypes of the nine cases with central retinal disease caused by variants in genes other than *ABCA4*. All 27 cases with pathogenic variants in *ABCA4* were correctly identified by clinical examination as *ABCA4*-related disease (i.e. STGD1). For nine cases that were genetically solved by variants in other genes, only two were erroneously identified as STGD (probands 4767 and 5855). These cases had heterozygous variants in *PRPH2*, a well-known phenocopy of *ABCA4* (Kniazeva et al., 1999; Cideciyan et al., 2005). Furthermore, proband 4767 had a diagnosis of autosomal dominant STGD, which is commonly attributed to *PRPH2*. Of the unsolved cases, there were diagnoses of STGD (N=4), MD (N=6), cone-rod dystrophy (N=10) and one diagnosis of North Carolina macular dystrophy. These cases will be taken forward for WES or WGS. Alternatively, cases with phenotypes indicating *ABCA4* disease may be taken forward for an smMIPs panel covering the entirety of *ABCA4* (Khan et al., 2020).

For cases with pathogenic variants in *ABCA4*, it was possible to compare the phenotype with a model of *ABCA4*-related disease proposed by Cremers and colleagues (Cremers et al., 2020). This was done in collaboration with Zelia Corradi (Radboud MC) and Mr Martin McKibbin. The previously calculated severity of *ABCA4* variants (Cornelis et al., 2022) was compared with age at diagnosis, peripheral involvement and progression of disease, in light of the model proposed by Cremers and colleagues (Figure 5.6).
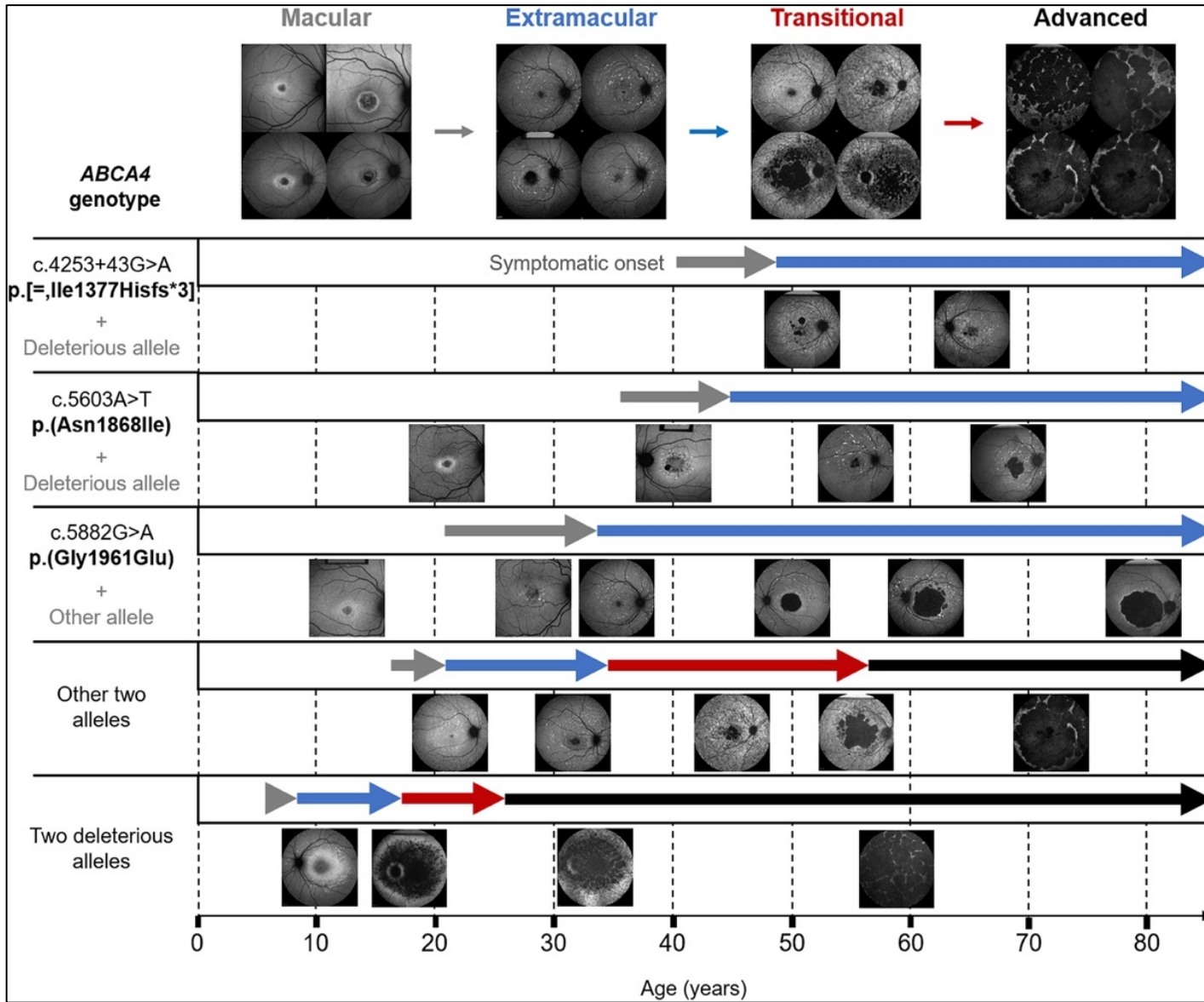
Figure 5.6 Model of _ABCA4_ disease progression proposed by Cremers and colleagues.

The cases with _ABCA4_-related disease were compared to the pictured model, according to grading performed by Mr Martin McKibbin. Reproduced with permission from (Cremers et al., 2020).

This analysis was possible for 25 of 27 cases which were genetically explained by biallelic variants in *ABCA4*, while for the remaining two there was insufficient diagnostic information. It was found that for 18 of these 25 cases, the genotype-phenotype relationship was as predicted by the scoring model. Three exemplar cases are displayed in Figure 5.7A-C. Of the 18 cases that fit with the model, eight had more than two pathogenic variants in *ABCA4*. However, each of these cases included variants that are known to coexist in complex alleles composed of two or more variants that have been observed previously in *cis*, allowing the phase of these cases to be assumed with a high degree of confidence. This is highlighted in Figure 5.7D, where assuming the presence of the known complex alleles, a good genotype-phenotype correlation was obtained as proposed by the model. There were five cases where more than two pathogenic variants were identified but phase could not be established or inferred from previous studies, preventing accurate assessment using the model (1337, 2843, 4126, 5604, 5853). This is shown in the case displayed in Figure 5.7E. Here it was not possible to assume the phase of these variants and therefore the degree of correlation with the model could not be assessed. Additionally, there were two cases (3536, 5609), where the phenotype-genotype model did not fit with the clinically examined grade (Table 5.6). Therefore, in summary, of the 25 biallelic *ABCA4* cases with sufficient clinical information available, 18 correlated with the model and two did not, while the remaining five could not be interpreted due to a lack of phasing information.
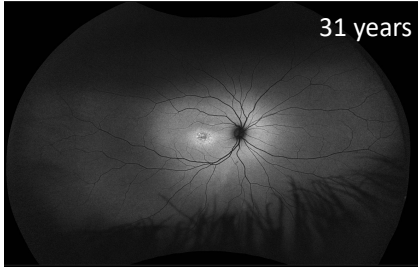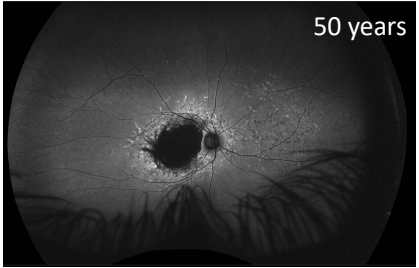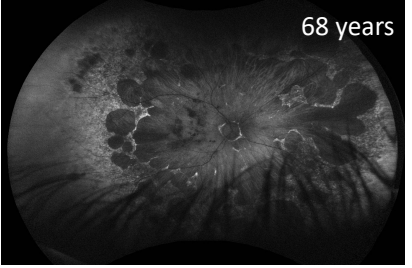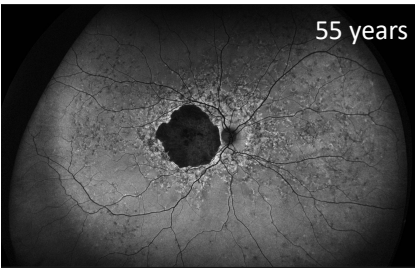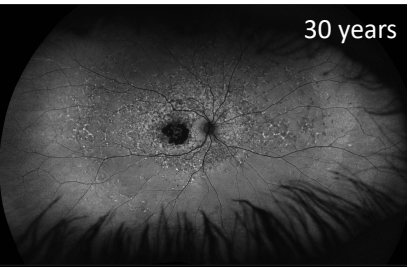
A) 5860	c.1253T>C c.1317G>A	Severe Severe	B) 5851	c.1282del c.6320G>A	Severe Mild	C) 5864	c.4326C>A c.5882G>A	Severe Mild

Grade: Macular	Predicted: Macular	Grade: Transitional	Predicted: Transitional	Grade: Advanced	Predicted: Advanced

D) 1746	c.[2588G>C;5603A>T]* c.[5461-10T>C;5603A>T]*	Severe Mild	E) 5853	c.365_366insCA c.5603A>T(;)c.5560G>T(;)c.5882G>A **

Key:
*Presumed complex alleles
**Phase unknown
Uncertain: model prediction could not be performed as the phase of the variants is unknown.

Grade: Transitional	Predicted: Transitional	Grade: Macular	Predicted: *Uncertain*

Figure 5.7 Exemplar genotype-phenotype correlations for *ABCA4*-associated disease with associated Ultra-Wide Field Autofluorescence images

A, B and **C** show three exemplar cases which fit with the model proposed by Cremers and colleagues (Cremers et al., 2020), based on gradings assigned by Mr Martin McKibbin, guided by definitions from Cremers and colleagues. **D** and **E** show two cases which show the importance of accurate phasing information for predicting the result of the mutations. Case 1746 has two known complex alleles, allowing the phase to be predicted with a relatively high degree of confidence. In contrast case 5853 has four variants which could not be phased, preventing accurate prediction (Images provided by Mr Martin McKibbin). Variant severity scores were obtained form Cornelis et al., 2022 and were calculated from the observed phenotype in affected patients.

Table 5.6 Genotype-phenotype correlation for *ABCA4*-associated disease.

Grading and model comparison from Cremers et al. 2020. Identified alleles are reported in protein notation. Severity score
for the variants was obtained from Cornelis et al. 2022. In "Diagnosis" columns, when not otherwise specified, STGD
refers to intermediate/classical Stargardt disease.

| Case | Clinical data | | | Genetic data | | | | | Match to model |
	Age at grading	Grade	Diagnosis	Allele 1	Severity	Allele 2	Severity	Diagnosis	
1337	59	3	STGD	c.5603A>T(;)5819T>C	MildLP Severe	c.6817-2A>C	Severe	STGD	Uncertain
1746	55	3	STGD	c.[2588G>C; 5603A>T]	Mild	c.[5461-10T>C; 5603A>T]	Severe	STGD	Yes
1808	34	1	STGD	c.[5461-10T>C;5603A>T]	Severe	c.5882G>A	Mild	STGD	Yes
2469	36	3	STGD	c.[2588G>C; 5603A>T]	Mild	c.[5461-10T>C; 5603A>T]	Severe	STGD	Yes
2843	22	3	CRD	c.4016G>A(;)5313-1_5313del	Unknown Severe	c.6088C>T	Severe	STGD	Uncertain
3536	60	½	Late onset STGD	c.3113C>T	Mild	c.1906C>T	Severe	STGD	No
4126	13	½	STGD	c.4774-27T>C(;)5196+1137G>A	Benign Mild	c.[5461-10T>C; 5603A>T]	Severe	STGD	Uncertain
5219	68	4	Early onset STGD	c.634C>T	Severe	c.[5461-10T>C; 5603A>T]	Severe	Early onset STGD	Yes
5349	40	2	STGD	c.1906C>T	Severe	c.5603A>T	MildLP	STGD	Yes
5604	54	4	STGD	c.4577C>T(;)4469G>A	Moderate Severe	c.5603A>T	MildLP	STGD	Uncertain
5607	33	2	STGD	c.3259G>A	Severe	c.6089G>A	Mild	STGD	Yes
5608	45	3	STGD	c.[2588G>C; 5603A>T]	Mild	c.[5461-10T>C; 5603A>T]	Severe	STGD	Yes
5609	37	4	STGD	c.4139C>T	Moderate	c.4139C>T	Moderate	STGD	No
5851	50	3	STGD	c.1282del	Severe	c.6320G>A	Mild	STGD	Yes
5852	66	3	STGD	c.[2588G>C; 5603A>T]	Mild	c.[5461-10T>C; 5603A>T]	Severe	STGD	Yes
5853	30	1	STGD	c.365_366insCA	Severe	c.5560G>T(;)5603A>T(;)5882G>A	Severe, MildLP, Mild	STGD	Uncertain
5854	36	2	STGD	c.4195G>A	Mild	c.5318C>T	Severe	STGD	Yes
5857	20	3	Early onset STGD	c.6229C>T	Severe	c.6229C>T	Severe	Early onset STGD	Yes

5860	31	1	STGD	c.4326C>A	Severe	c.5882G>A	Mild	STGD	Yes
5861	37	3	STGD	c.5714+5G>A	Moderate	c.[5461-10T>C; 5603A>T]	Severe	STGD	Yes
5862	17	3	Early onset STGD	c.1906C>T	Severe	c.4577C>T	Moderate	Early onset STGD	Yes
5863	42	2	STGD	c.[2588G>C; 5603A>T]	Mild	c.4537del	Severe	STGD	Yes
5864	68	4	STGD	c.1253T>C	Severe	c.1317G>A	Severe	STGD	Yes
5865	49	2	STGD	c.1906C>T	Severe	c.5603A>T	MildLP	STGD	Yes
3656	42	1	Occult MD	c.4139C>T	Moderate	c.5882G>A	Mild	STGD	Yes

Early onset, <10 years old at onset; Intermediate, 10-40 years old at onset; late-onset, >40 year old at onset.

In "Grade" column: 1 = Macular, 2 = Extramacular, 3 = Transitional, 4 = Advanced. STGD, Stargardt disease; CRD, Cone-rod dystrophy; MD, Macular dystrophy. MildLP, Mild, Low Penetrance

5.3   Discussion

In this chapter, a cohort of 57 UK cases with a variety of MDs were sequenced using a previously established smMIPs panel targeting 71 MD loci and 36 AMD loci. Loci associated with AMD were not analysed as part of this study. The application of this smMIPs screening reagent in cases from the Leeds IRD cohort provided further confirmation of the efficacy of the reagent and screening approach. This yielded high solve rates (36/57, 63.2%), consistent with other targeted IRD screens (Britten-Jones et al., 2023). A high proportion of these cases were genetically explained by variants in *ABCA4*. In part this reflects the fact that *ABCA4* is the most common cause of MD. In addition, cases sent to the Leeds IRD research group with phenotypes suggestive of *ABCA4*-associated disease had not previously been prioritised for sequencing, leading to a backlog of such cases. Genetic characterisation of cases with pathogenic variants in *ABCA4* allowed for grading of these cases in line with a model proposed by Cremers and colleagues. This model was found to be largely accurate and clinically useful, but a lack of phasing information in a number of cases limited the utility of the model.

5.3.1   *ABCA4* genotype-phenotype correlation

The genotype-phenotype model discussed by Cremers and colleagues was examined (Cremers et al., 2020). Cases with confirmed *ABCA4* disease underwent comprehensive clinical examination by Mr Martin McKibbin and Mr David Steel in light of the grading system used in the model. Following examination of the genotype-phenotype relationship in these cases, it was concluded that the model was accurate and clinically useful.

It was not possible to phase the variants observed in *ABCA4* which limited the application of the model in a number of cases. There were 15 cases in this analysis with more than two variants. However the majority (N=11) had a known complex allele. It was possible to grade these cases using the model by assuming the observed variants were in *cis* in a single complex allele. All cases with presumed complex alleles fit the model as proposed by Cremers and colleagues, both supporting the presumed phase of these alleles and underscoring the accuracy of the model.

It was valuable to be able to presume the phase of these complex alleles as they are commonly observed in *ABCA4* and result in an altered affect compared to considering the variants in isolation. *ABCA4*: c.[2588G>A;5603A>T] has been observed to be present in up to 50% of all *ABCA4* complex alleles, an observation which was replicated in this cohort (six of 11 individuals with presumed complex alleles) (Zernant et al., 2017; Runhart et al., 2018). The c.2588G>A variant has been proposed to act as a modifier of the c.5603A>T as opposed to acting as a pathogenic variant in isolation (Zernant et al., 2017; Runhart et al., 2018). The c.2588G>A variant was not observed without c.5603A>T in this screen, supporting previous observations that these variants are in linkage disequilibrium. Of further interest, there were four probands who shared the genotype; *ABCA4*: c.[2588G>A;5603A>T];[5461-10T>C;5603A>T]. In all cases in which it was observed in this screen, this allele combination of a mild-complete penetrant allele and severe allele resulted in 'classic' STGD, progressing to an advanced disease stage by the 6th decade, as observed previously and predicted by the model (Runhart et al., 2018). It was possible to use the model to assess the genotype-phenotype relationship in 25 cases in this screen, of which five cases could not be interpreted. If these variants had not been treated as a single complex allele, it would not have been possible to employ the model in a further 11 cases, severely limiting its utility.

There were four cases with four variants in *ABCA4* (4126, 5608, 5852, and 5853). It was possible to use the model in two of the cases (5608 and 5852), both of which had *ABCA4*:c.[2588G>C;5603A>T];[5461-10T>C;5603A>T]. In both of these cases, the phenotype was as predicted by the model. For the other two cases, it was not possible to use the model as the haplotypes could not be assigned. However it is possible to use the model to speculate as to the haplotypes in these cases.

Case 4126 exhibited a macular phenotype at age 13, indicating that there were two deleterious alleles present. This case contained a benign allele (c.4774-27T>C), a mild allele (c.5196+1137G>A) and a severe allele (c.[5461-10T>C;5603A>T]). This phenotype is more severe than would be expected from the combination of a mild allele and a severe allele. As such it is possible that the benign variant is in *cis* with c.5196+1137G>A and the resulting allele has a stronger affect than would otherwise be expected, or that other undiscovered modifiers may be playing a role in this case.

Case 5853 exhibited a macular phenotype at age 30. This case contained two severe alleles (c.365_366insCA, and c.5560G>T), a mild allele associated with low penetrance (c.5603A>T) and a mild allele (c.5882G>A). While it is not possible to assign these variants to haplotypes, the relatively mild phenotype likely indicates that the severe alleles may be on the same allele, with one or more mild alleles on the other allele. The phenotype-genotype model proposed by Cremers and colleagues suggests that genotypes with c.5882G>A as one of the alleles exhibit a macular phenotype at 30 years old, as was observed in this case (Zernant et al., 2017; Cremers et al., 2020). As such it is possible that the haplotypes are arranged as c.[5882G>A];[365_366insCA;5560G>T;5603A>T]. While these arrangements are possible considering the available information, phasing of the variants in both cases is required to definitively discriminate these haplotypes.

In addition to pathogenic variants in *ABCA4*, six cases harboured additional variants in MD genes (3616: *PRPH2*, 5219: *IMPG1,* 5857: *CNGA3*, 5607: *FBN2*, 5608: *BEST1,* and 3656: *CACNA1F*). These variants were considered when deciding whether it was appropriate to apply the genotype-phenotype model.

Of interest, a single case, 3616, had pathogenic variants which could be considered causal in both *ABCA4* (c.[4469G>A];[5603A>T]) and *PRPH2* (c.623G>A, p.(Gly208Asp)). This was a sporadic case so it was not possible to determine whether there was a dominant or recessive inheritance pattern. However, this case had a diagnosis of 'classical STGD'. *ABCA4:*c.4469G>A is predicted to alter splicing (SpliceAI 0.8) and has been shown to result in mis-splicing in minigene assays (Heidi L. et al., 2017). *In vivo* experiments have found that this variant results in mis-localisation of ABCA4 and was also found to result in reduced ABCA4 activity *in vitro* (Sun et al., 2000; Wiszniewski et al., 2005). This variant is considered a severe variant (Cornelis et al., 2021). In this case, this variant was assumed to be in *trans* with c.5603A>T, resulting in relatively mild disease (Zernant et al., 2017; Cremers et al., 2020). When only considering the *ABCA4* variants, this case fits within the genotype-phenotype model as proposed by Cremers and colleagues, exhibiting a macular phenotype at age 40. However this case was not included in the analysis using the genotype-phenotype model as the presence of the pathogenic *PRPH2* variant will have an unknown effect on disease progression in future.

It has previously been observed that variants in *PRPH2* can modify the disease progression of *ABCA4*-associated disease. A *PRPH2* haplotype has been observed acting as a *trans* acting modifier, increasing the penetrance of *ABCA4* variants (Zernant et al., 2022). It has also previously been observed that *ABCA4* mutations can affect the phenotype of *PRPH2*-associated disease caused by the common *PRPH2:*p.R172W variant (Poloschek et al., 2010). Further, the *PRPH2* variant observed in this case, *PRPH2*:c.623G>A, is a variant which is considered pathogenic by ACMG standards, and which has been observed in cases with a variety of IRDs, in particular cone dystrophy, and autosomal dominant RP (Manes et al., 2015; da Palma et al., 2019). This variant has also been associated with dominant cone rod dystrophy with incomplete penetrance (Soucy et al., 2023). As a result, it is difficult to use the model to grade this case.

It has previously been observed that the *ABCA4*:c.5603A>T variant when in *trans* with a severe allele causes disease with reduced penetrance (Zernant et al., 2017). It is likely that the presence of pathogenic variants in both *PRPH2* and *ABCA4* increased the penetrance of the disease phenotype. If DNA from other family members became available in the future, it would be of interest to segregate these variants and observe whether the *PRPH2* variant increases the penetrance of the *ABCA4* phenotype and *vice versa.* In addition it would be of interest to follow-up this case in the future and determine whether future disease progression is more severe than expected due to the presence of the *PRPH2* variant.

Case 5219 had a diagnosis of STGD and analysis with the *ABCA4* genotype-phenotype model found that the genotype (*ABCA4*:c.[634C>T];[5461-10T>C;5603A>T]) caused a phenotype consistent with prediction from the model. This case additionally had a likely pathogenic variant in *IMPG1*: c.1157C>A. Mutations in *IMPG1* have been found in cases with vitelliform MD and RP (Carss et al., 2017; Olivier et al., 2021). While it is possible that this variant will modify the disease course in the future, *IMPG1* has not been reported as a *trans* acting modifier of *ABCA4* variants in the past. An IRD case has been observed with single pathogenic variants in *ABCA4, IMPG1* and other genes associated with MD (Lofaro et al., 2021). As *IMPG1* is not a reported modifier of *ABCA4*-associated disease and dysfunction in the two genes results in distinct phenotypes, it was decided that this case could be graded using the model.

Case 5857 had a diagnosis of early onset STGD, consistent with the prediction from the genotype-phenotype model for a homozygous severe *ABCA4* variant, c.6229C>T. Additionally, this case was found to harbour a variant in *CNGA3* (*CNGA3*: c.1279C>T, p.(Arg427Cys)). This variant was deemed unlikely to affect the progression of the *ABCA4*-associated phenotype in this case. This variant is primarily associated with achromatopsia (Wissinger et al., 2001; Doucette et al., 2013; Yousaf et al., 2022). Additionally, while this variant is pathogenic by ACMG standards, mutations in *CNGA3* cause disease in an autosomal recessive inheritance pattern and only a single *CNGA3* variant was found. As such this case was deemed to be suitable for grading with the *ABCA4* phenotype-genotype model.

Cases 5607, 5608 and 3656 were found to harbour additional mutations in *FBN2*, *BEST1* and *CACNA1F* respectively. Each of the variants in these genes were found to be VUS by ACMG guidelines. As a result, it was decided that these variants were unlikely to affect the progression of the *ABCA4*-associated phenotype and that these cases could be graded using the genotype-phenotype model. In all three cases, the phenotype was as predicted by the model, further indicating that these variants were unlikely to have an effect.

There were seven cases that did not definitively fit within the model. Five of these are due to the presence of more than two variants which could not be phased (1337, 2843, 4126, 5604, 5853). Two cases, 3536 and 5609 did not fit within the model. Case 3536 exhibited a far milder phenotype than expected, while case 5609 exhibited a more severe phenotype. Further sequencing of these case and others may uncover previously undiscovered relationships between variants which may explain this. In addition, further screening of cases may allow for the discovery of additional complex alleles, permitting relatively confident segregation of variants into known complex alleles. In the cases with more than two variants in this screen, phasing of the variants found in these cases may allow for association of the resulting genotypes with their phenotypes. Finally, further screening of more patients, including screening intronic sequences, may uncover further deep-intronic variants and modifier alleles, allowing further refinement of the genotype-phenotype model.

5.3.2 Genes found to be mutated by smMIPs sequencing

Many cases with IRDs are solved by variants in a relatively small number of genes reflecting both increased rates of disease associated with those genes, and better characterisation of those targets. As a result most screens of IRD cohorts use targeted sequencing panels, reducing costs while solving many cases (Britten-Jones et al., 2023). In one large IRD screen, 71.2% of cases were solved by variants in 20 genes, while in another, 68% cases were solved by mutations in just 10 genes (Goetz et al., 2020; Pontikos et al., 2020). The genes most commonly found to be mutated in cases with MD include *ABCA4*, *PRPH2*, *CRB1*, *PROM1*, and *BEST1* (Pontikos et al., 2020).

This screen was designed to target a large number of loci associated with macular disease. As many genes were targeted, this allowed for detection of variants which are less commonly implicated in MD (for example *C1QTNF5*). Additionally, it was possible to search for *trans* acting modifiers to the disease phenotype and to have increased confidence in findings in the cases which were solved. Although 71 loci associated with MD were targeted, only a small number of genes yielded causative variants Figure 5.8.
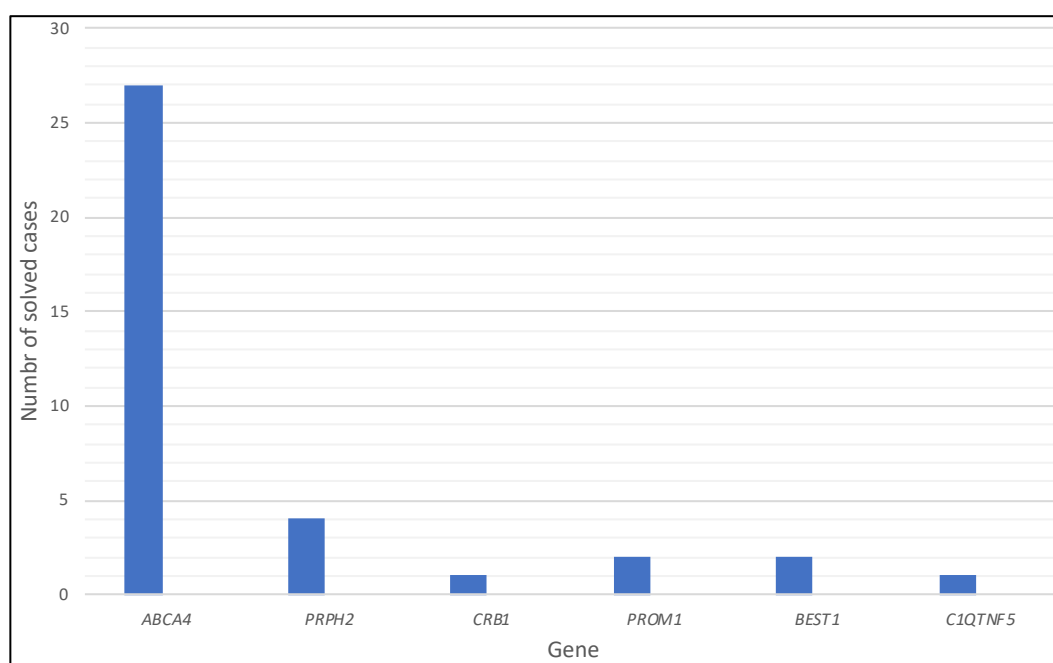
Figure 5.8 Genes implicated in solved cases in smMIPs screen

The majority of cases were solved by mutations in a small number of genes, in particular *ABCA4* (75% of solved cases, N=27). *PRPH2* (N=4), *CRB1* (N=1), *PROM1* (N=2), *BEST1* (N=2), *C1QTNF5* (N=1) were also implicated.

5.3.2.1 ABCA4

The majority (75%) of cases were solved by mutations in *ABCA4*. *ABCA4* is the most commonly mutated gene in IRDs and pathogenic mutations in *ABCA4* result in a variety of phenotypes which primarily affect the macula (Section 1.3.1.1). The probands selected for inclusion in this screen were selected based on a diagnosis of a macular dystrophy and it is to be expected that many would have pathogenic *ABCA4* mutations. In addition, cases with a phenotype which was likely due to mutations in *ABCA4* had not been prioritised for screening prior to this study. Because of this, the unscreened cases with macular phenotypes were biased towards *ABCA4*. In total 26/57 cases in this screen had a diagnosis of STGD following clinical examination and 18 of these cases were solved by pathogenic variants in *ABCA4*.

It has previously been observed that 10-15% of cases clinically diagnosed with a phenotype likely caused by *ABCA4*-associated disease are caused by mutations in genes other than *ABCA4* (Cremers et al., 2020). In this screen, four cases with a diagnosis of STGD were found to harbour pathogenic mutations in genes other than *ABCA4* (namely *PRPH2* and *C1QTNF5*). *PRPH2* has been previously identified as the most commonly mutated gene other than ABCA4 in cases with a clinical diagnosis of STGD (Wolock et al., 2019a), an observation which was also made in this screen.

ABCA4 is a flippase with a crucial role in the visual cycle. Pathogenic variants in *ABCA4* result in characteristic build-up of lipofuscin, resulting in damage to the macula (Section 1.3.1.2). Over 2000 known variants are listed in the LOVD entry for *ABCA4*, reflecting the effort of many research groups to comprehensively catalogue known variants in *ABCA4*. Previously proposed categorizations of the severity of known variants in *ABCA4* (Cornelis et al., 2022) were used for grading *ABCA4* variants in this experiment. *ABCA4* has been extensively studied in the past, and this allowed for more confident assessment of the variants observed in this screen.

5.3.2.2 PRPH2

The gene next most commonly carrying pathogenic variants in cases in this cohort was *PRPH2* (OMIM 179605), with four cases explained by mutations in this gene. These were 3616 (*PRPH2*:c.623G>A, p.(Gly208Asp)), 3798

(*PRPH2*:c.638G>A, p.(Cys213Tyr)), 4767 (*PRPH2*:c.394del, p.(Gln132Lysfs*7)) and 5855 (*PRPH2*: c.291G>A, p.(Trp97*)). Mutations in *PRPH2* predominantly cause autosomal dominant MD (Peeters et al., 2021), and accordingly all of the *PRPH2* variants observed in this screen were heterozygous and predicted to be pathogenic or likely pathogenic.

PRPH2:c.623G>A is a common cause of *PRPH2* associated disease, with 31 entries on LOVD and 9 submissions to ClinVar (VCV000437965.13), where it has been listed as causing STGD, MD and pigmentary retinal dystrophy. It is the most common pathogenic *PRPH2* variant in gnomAD (Reeves et al., 2020). While this variant has been reported in ClinVar as a cause of STGD (VCV000437965.13), this case (3616) also had pathogenic *ABCA4* variants. *PRPH2*:c.638G>A is also a relatively common cause of *PRPH2* associated disease. It has been reported 13 times in LOVD (www.lovd.nl, accessed 10/4/2021) and is reported as Pathogenic/Likely Pathogenic in ClinVar. With 6 submissions (VCV000098692.11), where it has been listed as causing STGD, pattern MD and retinal dystrophy. *PRPH2*:c.394del is also a relatively common cause of *PRPH2* disease. Previous studies estimated it to be among the 10 most common pathogenic *PRPH2* variants (Ba-Abbad et al., 2014; Reeves et al., 2020). It has previously been listed in LOVD (6 entries) and ClinVar, with 4 submissions (VCV000684461.10), where it has been listed as a cause of STGD and RP. Finally, *PRPH2*: c.291G>A is a novel variant which was found in this screen. Prediction using multiple prediction tools found that it is likely to be pathogenic, resulting in NMD of the transcript.

PRPH2 is a tetraspanin which localises to the outer segments of rods and cones. PRPH2 interacts with itself or with ROM1 to form complexes which are required for outer segment formation and organisation (Tebbe et al., 2020a). Mutations in *PRPH2* have been demonstrated to result in disorganized outer segment discs (Tebbe et al., 2020a), leading to both autosomal dominant (Farrar et al., 1991; Kajiwara et al., 1991) and autosomal recessive MD (Wells et al., 1993; Boon et al., 2007). PRPH2 also has an important role in the maintenance and cycling of outer segments, dysfunction of which has been associated with progressive retinal damage (Tebbe et al., 2020a).

5.3.2.3 PROM1

Two cases were explained by mutations in *PROM1* (OMIM 604365) both of which had *PROM1*: c.1117C>T, p.(Arg373Cys). Mutations in *PROM1* have been observed to cause autosomal dominant and recessive MD (Yang et al., 2008b) as well as autosomal recessive RP (Zhang et al., 2007). This variant is the most common cause of *PROM1*-associated disease, with 115 entries on LOVD and 17 entries on ClinVar (VCV000005610.33) where it listed as causing STGD4, MD and cone-rod dystrophy. Probands 3615 and 5610 had diagnoses of Best disease and cone rod dystrophy respectively. This variant has previously been found to result in a bull's eye phenotype, characteristic of STGD4 (Michaelides et al., 2010; Kim et al., 2017). While this variant is the most common cause of *PROM1*-associated disease, it is possible the two cases are related. Visual examination of *PROM1* did not reveal variants which could be used to discriminate whether the samples were related.

PROM1 is a penta-span transmembrane protein which localises to microvilli, with important functions in stem cells and a number of progenitor cells (Bahn and Ko, 2023). *PROM1* is expressed in a variety of tissues including the brain, colon, liver and lung (Yu et al., 2002). In the retina, the exact mechanism of *PROM1*-associated disease remains under investigation. Drosophila with mutant *prom* do not develop functional microvilli in the photoreceptor, leading to photoreceptor degeneration (Gurudev et al., 2014). In humans, disrupted disc morphogenesis due to mutations in *PROM1* has been associated with progressive thinning of the retina (Zacchigna et al., 2009; Kobayashi et al., 2021). Mutations in *PROM1* result in a similar phenotype to *ABCA4*-associated disease and a pedigree has been reported in the literature with both *PROM1* and *ABCA4* mutations, resulting in an altered phenotype (Imani et al., 2017; W. Lee et al., 2019). *PROM1* deficient mice show downregulation of *ABCA4*, presumably causing downstream dysregulation of the visual cycle and suggesting a possible mechanism for the similarity between the two phenotypes (W. Lee et al., 2019).

5.3.2.4 BEST1

Two cases were also solved by heterozygous mutations in *BEST1* (OMIM 607854), 3654 (*BEST1*:c.889C>T, p.(Pro297Ser)) and 3670 (*BEST1*:c.728C>T, p.(Ala243Val)). Both cases had diagnoses of MD. Mutations in *BEST1* cause a range of Bestrophinopathies (Ramsden et al.,

2012) which can primarily affect the macula, such as Best disease (Petrukhin et al., 1998), or the peripheral retina such as RP (Grewal et al., 2021). *BEST1*: c.889C>T is a previously reported cause of *BEST1*-associated disease, with 18 entries on LOVD and 8 entries on ClinVar (VCV000099766.20). Previous reports have also found that this mutation causes MD (Haer-Wigman et al., 2017; Weisschuh et al., 2020), as was the case in this study. *BEST1*: c.728C>T is a common cause of *BEST1*-associated disease, with 56 entries on LOVD and 10 submissions to ClinVar (VCV000002737.26). This variant has previously been reported to cause a range of phenotypes, including MD. Of interest, this mutation has previously been linked to a characteristic pattern dystrophy with a later age of onset and has also been hypothesised to have a milder effect on channel activity (Yu et al., 2006; Khan et al., 2018).

In the eye, *BEST1* is exclusively expressed in the RPE (Esumi et al., 2009). *BEST1* encodes a Ca^{2+} responsive chloride channel (Sun et al., 2002; Marmorstein et al., 2015). BEST1 channels have been observed to be permeable to other molecules, suggesting additional roles in homeostasis and neurotransmitter release (Grewal et al., 2021). The role of *BEST1* mutations in disease is still a subject of investigation although it has been suggested that that altered channel permeability leads to progressive retinal damage (Wittström et al., 2011).

5.3.2.5 *C1QTNF5*

A single case was solved by a pathogenic variant in *C1QTNF5* (OMIM 608752), 4030 (*C1QTNF5*: c.489C>G, p.(Ser163Arg)). Mutations in *C1QTNF5* have been associated with dominant, late onset retinal degeneration (LORD) (Hayward et al., 2003a; Xu et al., 2022). This variant is the most common pathogenic variant in *C1QTNF5* (Hayward et al., 2003a; Stanton et al., 2017), and has 15 entries in LOVD and 7 entries on ClinVar (VCV000002126.12). C1QTNF5 encodes a multimeric protein which has been suggested to form an extracellular complex with MFRP and facilitate adhesion of the RPE (Hayward et al., 2003b; Wong et al., 2008). *In vitro* experiments have found a dominant negative mechanism resulting in reduced stability of wild type C1QTNF5 (Stanton et al., 2017) and dysfunction of the visual cycle, resulting in extracellular deposit formation (Papastavrou et al., 2015).

5.3.2.6 CRB1

A single case (5258) was solved by the identification of pathogenic variants in *CRB1* (OMIM 604210), (*CRB1*:c.2106T>A p.(Tyr702*); c.2506C>A, p.(Pro836Thr) NM_201253.3). This case had a diagnosis of cone-rod dystrophy with an unknown inheritance pattern. Mutations in *CRB1* are commonly associated with RP and Leber congenital amaurosis (LCA) (Den Hollander et al., 2004; Henderson et al., 2011; Bujakowska et al., 2012) although macular phenotypes associated with *CRB1* mutations have also been reported (Henderson et al., 2011; Oh et al., 2021). Mutations in *CRB1* are associated with significant phenotypic variability, even within pedigrees (Bujakowska et al., 2012). CRB1 is a member of the Crumbs homologue family (Pellissier et al., 2013). CRB1, CRB2 and CRB3 form a complex which is involved in cell adhesion, polarity and communication and has been theorised to be crucial for morphogenesis and subsequent function (Cho et al., 2012; Alves, Pellissier and Wijnholds, 2014; Alves, Pellissier, Vos, et al., 2014). In particular, CRB1 has been observed to be vital for integrity of the junctions between Muller glia cells and photoreceptors (Quinn et al., 2017).

CRB1: c.2106T>A p.(Tyr702*) is a novel variant which is likely to result in a complete loss of function, and is located in exon 6 of *CRB1*. Other null mutations in this exon have been reported (LOVD), including *CRB1*: c.2107G>T p.(Glu703*) which has previously been reported as the pathogenic variant in a pedigree with LCA which displayed phenotypic variability (Saberi et al., 2019). *CRB1*:c.2506 C>A p.(Pro836Thr) has previously been reported 19 times on LOVD. This mutation has previously been associated with cone rod dystrophy in a homozygous case (Saberi et al., 2019) and with early onset retinal dystrophy in a compound heterozygous case (Henderson et al., 2011). This mutation affects the AG laminin like domain of CRB1 (Bujakowska et al., 2012) and is located in exon 7, which has previously been observed to be the most frequently mutated exon in screens of *CRB1* (Bujakowska et al., 2012) . Of interest, it has been suggested that the canonical isoform of *CRB1* (NM_201253.3) is not the major transcript in human, and that a previously undiscovered isoform is crucial for junction functioning (Ray et al., 2020).

5.3.3 Unsolved cases

Variants were identified in MD associated genes in six unsolved cases (4168: *CLN3*, 4181: *RP1L1*, 4756: *TIMP3*, 5152: *RP1L1*, 5202: *IFT81*, 5225:

EHMT2). The variants found in *EHMT2*, *TIMP3* and *RP1L1* were all considered VUS by ACMG standards. These variants were reported as it is possible that they may in future be found to be clinically relevant in these cases, but at present they are considered likely to be polymorphisms.

Pathogenic variants in genes associated with autosomal recessive disease were found in 4168 (*CLN3*:c.1141C>T, p.(Arg381Trp)) and 5202 (*IFT81*: c.899_900insT, p.(Glu301Argfs*8)). Case 4168 had a family history which suggested autosomal dominant disease and mutations in *CLN3* are primarily associated with recessive disease (Smirnov et al., 2021). As such it is unlikely that *CLN3* is the cause of disease in this case. Mutations in *CLN3* may also cause syndromic neurological disruptions which can appear later in life (Mirza et al., 2019). If these do manifest in this case, this would be more indicative of *CLN3*-associated disease.

For case 5202, it is possible that there is a second, undiscovered mutation in *IFT81*. Mutations in *IFT81* may also cause syndromic or non-syndromic IRD (Dharmat et al., 2017). In the case of syndromic disease, cases present with short-rib polydactyly syndrome (Duran et al., 2016). It is possible that this case had the syndromic form of disease and this information was not available, although this is unlikely.

In this screen there were four unsolved cases with frequent mild *ABCA4* variants (2272, 3528, 4181, and 5859). In these cases, no additional pathogenic *ABCA4* variants were discovered, indicating that these variants are likely polymorphisms. However, again it is possible that undiscovered deep intronic or other missed variants may be present on the second *ABCA4* allele in these cases. Further sequencing, in particular of the entire *ABCA4* locus may uncover the second variant in these cases.

For all unsolved cases, further sequencing may prove fruitful. This screen largely targeted the exons of genes implicated in MD. As such it is possible that causative mutations in the introns of screened genes or in intragenic sequences were not captured. Variants in the introns of genes may create non-canonical splice sites, leading to pathogenic mis-splicing (Sangermano et al., 2018). Variants in the regulatory elements of genes may pathogenically

disrupt their expression. For example, one case had a diagnosis of North Carolina MD (3528). North Carolina MD is caused by mutations upstream of *PRDM13* which result in mis-regulation (Green et al., 2021). WGS has been estimated to increase solve rates by 29%, but at a significantly increased per sample cost (Ellingford et al., 2016). In addition, it is possible that unsolved cases in this screen may harbour pathogenic variants in novel IRD associated genes. More widespread sequencing, in particular WGS, may uncover variants in the unsolved cases which were not targeted by this smMIPs screen.

Further, a number of loci were less well covered by this panel. These areas are generally not well covered by Illumina based NGS and include purine rich regions such as ORF15 of *RPGR* (Chiang et al., 2018) and the opsin array (Nash et al., 2022a). Although these regions were targeted in this panel, the reduced coverage and mapping quality of reads at these regions reduced confidence in the observed variants and it is possible that some of the unsolved cases in this screen harbour causative mutations in loci which were relatively poorly covered by this screen. For unsolved cases, assays which specifically target these hard to sequence areas may be of use (Chiang et al., 2018; Khateb et al., 2022).

Additionally, targeted sequencing panels struggle to detect CNVs and structural variants (Section 4.3.3.2). Although CNV analysis was performed, no CNVs were identified in this screen. CNVs have been previously estimated to be a cause of missing variants in 9% of IRD cases (Zampaglione et al., 2020). A major advantage of using WGS as opposed to targeted sequencing is that CNVs which do not directly affect the coding sequencing of the gene may also be detected. CNVs in the non-coding sequence can result in mis-regulation and have previously been implicated in the creation of pathogenic novel topologically associated domains (TADs) (de Bruijn et al., 2020). More widespread sequencing, in particular WGS may detect pathogenic classes of variants such as CNVs which were undetected in this screen.

5.4 Summary

In this chapter, a cohort of 57 cases with macular dystrophies were sequenced by a targeted panel approach using smMIPs based enrichment. 36 (63.2%)

cases were found to carry likely causative mutations in MD genes after screening at a very low per sample cost, supporting the use of smMIPs based targeted sequencing as an initial screening tool. This also allowed for verification of the smMIPs panel used by the Nijmegen group. A proportion of the variants identified were retested by an independent sequencing method and all were confirmed.

75% (27/36) of the solved cases were genetically explained by variants in *ABCA4* and two novel pathogenic variants in *ABCA4* were found during this screen. Additionally, a novel variant was found in *PRPH2* and *CRB1* respectively. Thorough clinical examination allowed comparison of the genotype-phenotype relationship of cases with *ABCA4*-related disease to a model proposed by Cremers and colleagues. This comparison found that this model is largely accurate and clinically useful, providing a likely prognosis that is accurate in the majority of cases. A number of cases had more than two variants which could not be accurately phased, limiting the implementation of the model on these cases. Importantly, the results of screen also provide diagnostic information to 36 individuals affected by MD who previously did not have a genetic diagnosis, although these results are a research result and should be verified by an accredited diagnostic laboratory.

6   General Discussion

The aims of this project were to study *ABCA4*-related disease and associated disorders, and to develop new methods to increase molecular diagnostic rates. The advent of long-read sequencing has opened up new avenues of research previously intractable to NGS-based sequencing. In Chapter 3, two methods to phase distal variants in *ABCA4* using long-read sequencing were developed. In Chapter 4, a method to quickly and effectively confirm and characterise deletions found through exon-based technologies in IRD genes using long-read sequencing was established and tested on a series of cases. In Chapter 5, a cohort of cases with macular dystrophies were screened using an smMIPs-based macular disease panel to genetically diagnose these cases.

6.1   Future of genetic testing

Rare diseases such as IRDs collectively represent a significant disease burden. The lifetime risk of a rare disease is 1 in 17 in the UK, and over 80% of rare conditions are estimated to have a genetic cause (Brittain et al., 2017). Obtaining an accurate genetic diagnosis is of great importance for those affected. A genetic diagnosis has been shown to help patients with IRDs by helping them adjust to the condition, providing a clearer prognosis and aiding decision making and family planning (Hayeems et al., 2005; Combs et al., 2013; Britten-Jones et al., 2022). While it is difficult to adjust to non-actionable results (results indicating a genetic disease with no available therapy), a common theme is that "*if I know, it might be difficult to deal with, but at least I'll know*" (McVeigh et al., 2019). In addition, there are a growing number of clinical trials for potential treatments of IRDs and an accurate genetic diagnosis is often a prerequisite for inclusion (https://clinicaltrials.gov/, accessed 27/4/2023). As such, improving diagnostic rates of genetic testing is an area of active research. Following genetic screening, only approximately two-thirds of IRD patients will receive a genetic diagnosis (Britten-Jones et al., 2023). This research can be divided into improving sequencing technologies to capture more potentially pathogenic variants, and classifying the variants which are found during sequencing.

6.1.1 Improved sequencing technologies

6.1.1.1 The falling cost of NGS

At present, for individuals undergoing genetic testing, targeted sequencing is typically performed as a frontline test for many genetic diseases, followed by more widespread screening such as WES or WGS for cases which are negative. Targeted sequencing panels will likely remain an important part of genetic testing in the future, in particular in geographic regions with less developed genetic screening programmes. An advantage of targeted panels such as smMIPs (Chapter 5) is that they can be adapted to new target loci as they are identified. As the cost of sequencing and targeting technologies continues to fall, the shift to targeting the entire locus of genes which are commonly mutated in IRDs, known as whole gene sequencing (GS) may improve solve rates in panel-based approaches (Khan et al., 2020; Dockery et al., 2021). This improvement is caused by allowing for the detection and study of non-coding variants such as deep intronic variants which are not commonly covered in targeted panels and remain difficult to interpret when detected. Deep intronic variants have been estimated to be ~2% of pathogenic variants in *ABCA4* but are not commonly covered in targeted screens at present (Cremers et al., 2020). Four unsolved cases which were screened using an smMIPs panel (Table 5.5) were found to harbour a single variant in *ABCA4*. If future targeted panels covered the entire *ABCA4* locus, it would have been possible to search for a second, non-exonic variant in these cases or to exclude the single variants as coincidental findings which are not related to disease in these individuals.

As the cost of NGS continues to fall, there is a continuing shift towards screening more of the genome as a front line test. At present, soon to be released instruments may allow for $200 whole genome short-read sequencing (https://www.illumina.com/systems/sequencing-platforms/novaseq-x-plus.html, https://www.elementbiosciences.com/200-dollar-genome accessed 21/2/2023). It has been suggested previously that WGS may be a more cost effective front line test compared to targeted sequencing as the cost of WGS falls (Lionel et al., 2018). This is reflected in an ongoing shift towards more routine use of WGS as standard, in particular in the UK NHS and through programmes such as the 100,000 Genomes Project and elsewhere (Turro et al., 2020; Costain et al., 2021; Trotman et al., 2022). However, the very large size of WGS files and increased data analysis requirements remain a significant issue. To compensate for this, it is common

practice to sequence the entire genome but only analyse panels of relevant genes (Martin et al., 2019). This is a good compromise as many IRD cases are genetically explained by mutations in a small number of genes while reducing the amount of incidental findings (Stone et al., 2017). As the entire genome is covered, unsolved cases following panel analysis can then be analysed for causative variants in other genes.

A possible future workflow for cases undergoing screening may be that cases with a phenotype highly suggestive of variants in a known gene, for example *ABCA4*-associated disease, will undergo targeted sequencing. This targeted sequencing will cover the entire gene of interest, allowing for high confidence examination of these loci at relatively low cost. Following analysis, any unsolved cases will then undergo WGS. In parallel, any cases with a less determinate phenotype will undergo WGS as a first test. When possible, direct relatives will also undergo WGS concurrently, allowing for the segregation of variants found during initial testing (French et al., 2022). This possible future parallel workflow may allow for cost effective diagnosis for cases with variants in well characterised genes, and effective examination in other cases.

6.1.1.2 Long-read sequencing

As the cost of long-read sequencing falls, the increased use of these technologies will increase solve rates still further by allowing for sequencing of regions currently intractable to short-read NGS. It has been estimated that 4264 exons in 619 clinically relevant genes have 100% homology to other loci and as a result it is very challenging to align the resulting NGS reads (Mandelker et al., 2016). Using long-read sequencing, it is now possible to sequence the entire chromosome from telomere to telomere, although this remains challenging (Nurk et al., 2022). Future screening using long-read sequencing will likely allow for sequencing of the entire genome, allowing for all potentially relevant loci to be covered.

Similarly, some categories of variants are challenging to detect and categorise by NGS but are more tractable to long-read sequencing, in particular CNVs and structural re-arrangements (Section 4.3.3.2). As an example of this, a study examining high quality WGS datasets found 4,442 structural variants per genome (Abel et al., 2020), however a similar study using both short read and long-read sequencing data found 27,622 structural variants per genome

(Chaisson et al., 2019). It has been estimated that structural variants account for up to 17.2% of deleterious alleles and that these may represent a significant and understudied group of disease causing variants in both Mendelian and complex disorders (Chiang et al., 2017; Chaisson et al., 2019; Abel et al., 2020). More routine use of long-read sequencing will allow for more reliable CNV detection which will allow for the full impact of these variant types in IRDs to be studied.

While long-read sequencing technologies are powerful and promise to significantly advance genetic testing, currently they are significantly more expensive than comparable methods (Section 1.4.2-1.4.3). Further, to obtain long read lengths using these platforms, specialist DNA preparation methods are required, often requiring expensive equipment. For example, while it was possible to obtain reads from unamplified genomic DNA which spanned the length of *ABCA4* using CATCH (Section 3.2.6), this required a fresh blood extraction and a HLS-CATCH machine. These limitations are likely to be addressed as the technology continues to develop. While currently non-standard DNA extraction and handling is required for use with long-read sequencing platforms, if these technologies become more common place then the processes surrounding their use will also likely be improved. If the cost of long-read sequencing continues to fall, it is possible that in the future long-read WGS will be used routinely, allowing for even more effective sequencing of the genome.

In particular, the increased use of single-molecule long-read sequencers will allow for the phasing of variants of interest. At present, this remains challenging as the high cost of these technologies require enrichment of target sequences. In Chapter 3, targeting ABCA4 using long-range PCR introduced artefacts and practical difficulties, however there were no significant confounding issues introduced by the long-read sequencer when amplification free enrichment was used. As the cost of long-read sequencing continues to fall, adapted DNA handling processes and routine amplification free long-read WGS using large sequencing platforms such as the PromethION will likely allow for routine phasing of large genes such as *ABCA4*. As this occurs, it is likely that it will no longer be sufficient to assume the phase of variants of interest.

6.1.2 Interpreting found variants

Screening more of the genome through more routine WGS, and future use of long-read WGS, will likely increase molecular diagnostic rates for patients with IRDs by capturing clinically relevant variants which are not captured by targeted sequencing panels. However undoubtedly the largest bottleneck to obtaining a genetic diagnosis for an IRD is the classification of the large numbers of variants which are found during screening.

As more of the genome is sequenced, more variants are found which must be examined. For example, in Chapter 5 a targeted panel covering 107 loci associated with MD and AMD was used. This sequencing resulted in 955 variants per case found prior to filtering. In contrast, following WGS as many as 3.1×10^6 SNVs per individual have been reported (Vidal et al., 2019). This represents a massive challenge to search for clinically relevant variants when WGS is used.

Most variants which are found during genetic screening are benign polymorphisms with no impact on the genetic disease in question. Following filtering and examination, many of the variants which are still considered are of uncertain significance. In ClinVar, 42% of variants are categorised as VUS (Iancu et al., 2021). This directly impedes the genetic diagnosis of cases with IRDs. In one IRD cohort, 17% of the cohort were unsolved pending classification of at least one VUS (Iancu et al., 2021). Similarly, both cases in Chapter 3 which underwent CATCH had at least one VUS in *ABCA4* and in Chapter 5, six unsolved cases had a VUS in genes relevant to MD.

In silico prediction tools have been developed which are effective at preliminary classification of variants, such as CADD (Kircher et al., 2014) and MutationTaster (Schwarz et al., 2014). However, these tools often contradict when interpreting variants with a less clear effect. Variant interpretation standards are also important for consistent interpretation and reporting of found variants. The most commonly used variant interpretation guideline is the ACMG standards (Richards et al., 2015). These standards were used to assess the variants assayed in Chapter 3 and to parse the variants found in Chapter 5. A systemic review of NGS screens of IRD cohorts found that of 80 studies published after 2015, an increased diagnostic yield was found in screens which adhered to ACMG guidelines (Britten-Jones et al., 2023).

Although *in silico* prediction tools and variant interpretation guidelines are important, even when these are used many found variants are considered VUS.

When variants in coding sequences are interpreted, they are generally considered in the context of the canonical transcript of the gene. However, novel transcripts for genes relevant to IRDs are continuously being discovered (Aísa-Marín et al., 2021). A recent analysis found ~23 transcripts per gene associated with IRDs, of which ~15 were novel (Ruiz-Ceja et al., 2023). This analysis discovered novel transcripts in genes examined in this study, including *ABCA4* and *PRPH2*. In addition, long-read sequencing may be more effective at sequencing RNA than NGS, resulting in novel transcripts which were previously missed. A recent study using PacBio sequencing reported a novel isoform of *CRB1* which is the most abundantly expressed in the retina (Ray et al., 2020). Further, where the effected transcript is expressed is an important consideration. Variants which have a minor effect on the canonical transcript, may have a pathogenic effect on a retina specific transcript (Vig et al., 2020). It has previously been suggested that IRD causing mutations may affect unannotated transcripts and the full impact of this remains to be seen (Aísa-Marín et al., 2021).

Variants which effect the non-coding genome are a particular challenge to classify. It has been estimated that 9% of IRD causing mutations affect splicing (Ruiz-Ceja et al., 2023). Many variants in the intronic sequences of genes associated with IRD will have an unclear effect on splicing, and it can be challenging to classify these variants. It is also likely that synonymous variants which are currently not considered to be pathogenic have an unforeseen effect on splicing. The mechanisms which regulate splicing, such as exonic and intronic splice enhancers and inhibitors are still under investigation and disruption of these mechanisms may be an under-appreciated cause of disease in some cases (Qian et al., 2021; Mauro-Herrera et al., 2021). As a result the effect of variants in the non-coding sequences of genes is generally more difficult to estimate than genes which effect the coding sequence.

Beyond variants in genes, the mechanism and impact of intra-genic variation is only beginning to be understood. Structural variants may affect the genomic

conformation of a locus, resulting in pathogenic mis-regulation of IRD associated genes (de Bruijn et al., 2020). Over 80% of transcripts are non-coding RNAs and variants affecting these are increasingly recognised as an important cause of IRDs (Sun et al., 2020). Pathogenic mutations in a diverse range of non-coding RNAs including microRNAs, long noncoding RNAs, and circular RNAs have all been associated with damage to the retina (Conte et al., 2015; Chen et al., 2017; Cao et al., 2019). Pathogenic variants may affect the *cis*-regulatory elements controlling gene expression (Green et al., 2021). The full impact of these variant classes remains to be seen and continuing research to annotate the non-coding genome will help to uncover the role of these variants in IRD (Cheung et al., 2019; French and Edwards, 2020; Waldern et al., 2022).

6.1.2.1   Incomplete penetrance

Phenotypes with reduced penetrance remain a confounding issue for classifying variants found during sequencing. As many as 1 in 6 genes are associated with variable penetrance of the associated phenotype (Green et al., 2020). Many of the genes found to be mutated in this study, including *ABCA4*, *PRPH2*, *BEST1* and *PRPF31,* have been associated with variable penetrance (Green et al., 2020). This makes it more challenging to categorise variants in these genes as they may be present in non-affected individuals in the examined pedigrees. For example, the deletion affecting *PRPF31* which was characterised in Chapter 4 was found to be present in an unaffected individual in the proband's pedigree.

An incomplete understanding of the mechanisms of incomplete penetrance impedes family planning for affected individuals. The exact mechanism of incomplete penetrance is likely to vary between affected genes. Variable gene expression levels have been implicated as one potential mechanism of variable penetrance (Green et al., 2020), for example variable expression of *PRPF31* has been linked to variable penetrance of the associated phenotype (Rose et al., 2016).

Modifier alleles have been proposed as another mechanism of incomplete penetrance. *Trans* acting modifiers have been associated with increased penetrance of the disease phenotype. For example, variants in *PRPH2* have been observed acting as a *trans* acting modifier, increasing the penetrance of

an *ABCA4*-associated phenotype (Section 5.3.1). It is frequently only possible to identify these interactions by comprehensive examination of pedigrees, as well as screening of large cohorts of patients (Zernant et al., 2017). However, as IRDs are individually rare, it can be challenging to identify suitable pedigrees which can be studied. For example, the case identified in Chapter 5 with pathogenic variants in both *PRPH2* and *ABCA4* was a singleton case. As such, it was not possible to examine whether these variants affected the penetrance of the disease phenotype. Large cohorts are more powerful for uncovering these variant interactions, for example the smMIPs based panel which was used in Chapter 5 has been used to screen large cohorts of IRD patients (Khan et al., 2019; Khan et al., 2020; Hitti-Malin et al., 2022; Panneman et al., 2022). Further study of the mechanisms of incomplete penetrance will likely have a significant impact on variant categorisation in the future.

6.1.2.2 Functional assays

Ideally functional assays would be performed to infer the clinical relevance of all identified variants of uncertain significance. While these assays are time consuming, they are essential for the investigation of variants that are too rare to obtain genetic proof of their involvement in IRDs.

The most comprehensive method to assay the effect of mutations in a gene implicated in IRDs is the use of animal models. While mouse models have been used extensively, this is complicated by the fact that mice do not have a macula, limiting modelling of macular diseases (Molday et al., 2018). As such, other animal models have also been used including canine models and zebrafish (Mäkeläinen et al., 2019; Perkins, 2022). Animal models have been invaluable for studying the effect of specific mutations and the role of the affected protein (Tebbe et al., 2020b; Tebbe et al., 2023). Animal models have also been used to estimate the effect of modifier alleles on the disease phenotype (Meyer and Anderson, 2017). While animal models are useful, there are significant differences in the genomic architecture and anatomy between the models and humans which can limit the translation of findings in animal models to humans (Slijkerman et al., 2015).

As a result, another approach is using cultured retinal and RPE organoids to model disease mechanisms. Patient derived organoids from induced

pluripotent stem cells have been used to model the effect of specific variants on the affected cells (Buskin et al., 2018; Watson and Lako, 2022; Su et al., 2022). Organoids are advantageous compared to animal models as they share the genomic architecture of the patient in question.

For variants which are thought to affect splicing, a variety of splicing assays have been developed such as amplification of RNA obtained from biopsies for clinical accessible tissue and non-accessible tissues (Aicher et al., 2020; Nash et al., 2022b). *In vitro* assays have also been used extensively, in particular mini-gene assays. In these assays, a plasmid containing the variant containing exon and surrounding introns is transfected into a cell line and the effect on splicing can then be examined (Desviat et al., 2012). These have been used to investigate variants thought to alter splicing in genes found to be mutated in this project such as *BEST1*, *PRPH2* and *PROM1* (Becirovic et al., 2016; Weisschuh et al., 2021; Rodriguez-Muñoz et al., 2022). This has also been done extensively in *ABCA4* using midi-genes, constructs which contain more of the gene of interest (Sangermano et al., 2018). The impact of the variants which were phased using CATCH (Section 3.2.6) on splicing will be investigated using a midi-gene assay.

While these assays are valuable tools to assess the effect on proper splicing, the splice environment in the HEK293T cells which are commonly used does not reflect the retina and the effect of these variants may be different when examined in photoreceptors derived from induced pluripotent stem cells (Aísa-Marín et al., 2021).

Other functional experiments include assays to indirectly measure the activity of the affected protein, for example by measuring transport activity and ATPase activity to indirectly assay residual *ABCA4* function (Zhang et al., 2015). While these assays are valuable tools, their results do not always reflect observations made of the resulting phenotypes in patients (Cremers et al., 2020). This may reflect undiscovered modifiers in the patients which are affecting the phenotype, or that the result of the assays do not fully reflect the biological situation. Although these assays have their limitations, they are an invaluable tool to assay the impact of VUS and to investigate the function of the affected proteins.

6.1.2.3 Re-analysing datasets

A valuable tool to increase solve rates is the re-analysis of previously sequenced cases. Updates to base callers, aligners and variant callers can improve data quality, and increased knowledge of disease mechanisms can lead to increased solve rates (Sarmady and Abou Tayoun, 2018; Dai et al., 2022). However this process is labour and time intensive. A meta-review of studies which re-examined previous NGS found that this yielded an additional solve rate of approximately 10% (Dai et al., 2022). Performing CNV analysis on targeted sequencing data can also improve solve rates by around 10% in IRD screens (Zampaglione et al., 2020). Reanalysing previously sequenced datasets also allows reinterpretation of variants previously considered VUS. The common *ABCA4*:c.5603A>T variant until relatively recently was considered a VUS. Reclassification of this variant as pathogenic solved more than 50% of cases with a single known pathogenic variant in *ABCA4* at the time (Zernant et al., 2017). If this variant had not been reclassified, it would have limited genetic characterisation in 15 cases sequenced using the smMIPs panel (Section 5.2.5). As such, periodic reanalysis of previously screened cases is an important tool as the understanding of the genetics of IRDs continues to advance. In addition, combining previously performed sequencing with novel methods to detect challenging variant classes such as CNVs has been used to increase diagnostic rates for IRD cohorts (de Bruijn et al., 2023).

Automated pipelines and AI have been proposed as solutions to improving variant interpretation, reporting and cohort re-examination (Ji et al., 2021). NGS data is generally inconsistent and of variable quality, with heterogenous data sources, screening methods and reporting standards. This introduces a challenge for the creation of automated analysis programmes. Despite this, a number of automated pipelines have been used to make diagnoses (Baker et al., 2019; Lassmann et al., 2020) and these have already been used to achieve similar solve rates to humans when re-examining NGS datasets (Dai et al., 2022). Exhaustive characterisation of model genes such as *ABCA4* may allow training of machine learning pipelines for application to other genes. This may be of particular value for the detection and classification of non-coding variants.

While automation and AI-based tools have the power to aid analysis, these programmes are not at the stage where they can be used to replace manual analysis. All examined variants in this study were classified using ACMG guidelines automatically generated by Franklin (https://franklin.genoox.com/clinical-db/home, accessed 25/5/2023). While this was useful for initial categorisation of the queried variants, all variants were also examined using independent sources such as ClinVar, LOVD and additional *in silico* prediction tools. While automated tools such as Franklin aided the analysis of variants in this study, they were used to supplement manual analysis rather than replace it.

6.1.2.4 Improving reference sequences

Another area which has the potential to improve solve rates is the improvement of reference sequences. A major drawback of the reference genome is that it is constructed using a relatively small number of well characterised samples. For example, 93% of GRCh38 is derived from just 11 individuals and 70% of the primary assembly is from a single individual (RPC1-11) (https://www.ncbi.nlm.nih.gov/grc/help/faq/#human-reference-genome-individuals, accessed 30/3/2023). This presents a source of potential bias in a number of ways. Firstly, the reference may not be representative of the typical genome, for example the reference genome was found to be at a higher risk for diabetes than a healthy control (Chen and Butte, 2011). As the reference is largely from a small number of sources, this can result in issues such as reference alignment bias (Barbitoff et al., 2017; Ballouz et al., 2019). Additionally, the reference genome may struggle to represent the full scope of variation. In particular, the reference genome is not as long as other genomes which have been sequenced and structural variation may be challenging to represent or interpret using the reference (Ballouz et al., 2019). Further, while presenting the reference as a haploid assembly is convenient and simplifies analysis and interpretation, it is not representative of the diploid genome. This leads to challenges when the haploid reference genome is not representative of the biological situation (Weisenfeld et al., 2017). Finally, reference genomes and genomic data bases are biased towards individuals of European descent which confounds estimation of the frequency of variants of interest (Popejoy and Fullerton, 2016). As variant detection is largely by comparison to a reference, this can confound identification of novel structural variants and SNVs in understudied populations. This also significantly limits

the study of population specific variants which are common in *ABCA4* in particular (Cremers et al., 2020).

An alternate approach which may reduce these issues is pangenomics. In this approach, rather than aligning sequences of interest to a single reference, the entire genetic diversity from the sequenced samples is captured (typically in a pangenome graph), reducing bias and capturing sequences which may not be present in the reference genome (Liao et al., 2022). Pangenome assembly requires high quality long reads, but if this challenge is addressed it promises to be a significant improvement compared to using a single reference (Eizenga et al., 2020). Pangenome graphs have already been demonstrated to be superior at CNV detection and can be stored at 3-6 Gb without decreasing the quality of the data (Eizenga et al., 2020; Rautiainen et al., 2023; Jiang et al., 2023). As long-read WGS is used more routinely, it is possible that there may be a migration to pangenomic based alignments. This has the potential to increase solve rates by aiding the analysis of interactions such as *trans* and *cis* acting modifiers.

6.2   Therapies for IRD

Continuing to improve our understanding of the genetic mechanisms of disease is key for the design and implementation of therapies to treat IRDs. Screening large cohorts of patients, for example by smMIPs based sequencing, allows for the identification of common variants which may be suitable targets and to identify cases with specific variants who may benefit from inclusion in ongoing trials. For example, following screening with smMIPs, there were eight cases identified with *ABCA4*: c.5461-10T>C (Section 5.2.4). This is a target for antisense oligonucleotide (AON) therapy (see below) which is under active investigation (Kaltak et al., 2023) and these cases may benefit in the future from this.

It is also important for the clinical relevance of pathogenic variants to be understood prior to targeting. Phasing of the associated variants by assays such as long-read sequencing allows for accurate haplotypes to be observed Section 3.2). If cases are included in a trial with incorrectly inferred haplotypes, this may result in the treatment incorrectly appearing to be ineffective (for example if the variants which are thought to be causative are in *cis* and there

is a third undiscovered variant in *trans*), which may lead to effective therapies being wrongly rejected.

As such, in addition to providing direct benefit to patients by informing family planning and genetic counselling, an accurate molecular diagnosis is key for patients to partake in a growing number of clinical trials. Until very recently, there were no therapies available to slow vision loss caused by IRDs (Britten-Jones et al., 2022). However, therapies targeting IRDs are an area of very active research and a broad spectrum of approaches are under development (Vázquez-Domínguez et al., 2019; Britten-Jones et al., 2022). In particular, gene therapies are an area of active research and are a promising avenue to treat IRDs. The retina has a number of advantages for targeting with gene therapies. It has comparatively non-invasive examination routes and administration generally does not result in systemic effects compared to other tissues. (J.H. Lee et al., 2019). Patients often notice changes in vision early compared to other symptoms, allowing for diagnosis before widespread cell death. The retina is also a relatively small environment, requiring only small doses of vector (J.H. Lee et al., 2019).

A number of delivery vectors for gene therapies have been explored including the use of viral and non-viral vectors. The first adeno-associated virus (AAV) gene therapy was approved for an IRD. Luxterna is an AAV-based gene-therapy delivered to the RPE, where it delivers functional *RPE65* cDNA (Russell et al., 2017). A variety of non-viral vectors such as nanoparticles have also been developed (Sun et al., 2022). A number of gene replacement therapies for IRDs are under investigation with trials targeting *CHM* (Dimopoulos et al., 2018; Fischer et al., 2019), *ND4* (Guy et al., 2017; Fischer et al., 2019), *CNGA3* (Fischer et al., 2020), and *RPGR* (Cehajic-Kapetanovic et al., 2020).

Similarly, a number of approaches have been developed which seek to directly edit genomic DNA *in vivo*, commonly using CRISPR-Cas9 complexes (Pulman et al., 2022). This approach has been used successfully in IRDs, with clinical trials underway to rescue *CEP290* associated LCA (Maeder et al., 2019). These also commonly use AAV vectors for CRISPR-Cas9 complex delivery. However this results in sustained Cas9 expression with potential negative effects for wider genome stability (Hanlon et al., 2019).

Other avenues include efforts to indirectly correct the pathogenic variant by modulating splicing, for example using antisense oligonucleotides (AONs) which can rescue splice defects by binding to pre-mRNA and altering splicing or degrading pathogenic transcripts (Hammond and Wood, 2011). AONs were first approved for use on Duchenne muscular dystrophy to skip an exon and restore the reading frame (Niks and Aartsma-Rus, 2016). These have shown promise *in vitro* and *in vivo* for correcting splice defects in IRD genes (Garanto et al., 2016; Duijkers et al., 2018; Tomkiewicz et al., 2021; Suárez-Herrera et al., 2022; Kaltak et al., 2023).

In parallel, gene agnostic approaches are also under development. These also have the added benefit that they can be useful for patients with significant damage to the retina. These include strategies seeking to replace or augment the affected cells, which may be particularly effective for cases with widespread damage to of loss of the affected cells (Battu et al., 2022; John et al., 2023). Other approaches include bypassing the affected cells, for example by introducing photoreceptive proteins to downstream neurons in a retina where photoreceptors have been lost through disease (Simunovic et al., 2019; Prosseda et al., 2022). Another way to treat complete photoreceptor loss is through the use of retinal implants, although these remain early in development and current versions can only restore limited visual capability (Suaning et al., 2014; Küçükoğlu et al., 2022).

6.2.1 Therapies for *ABCA4* disease

As variants in *ABCA4* are a common cause of disease, it is no surprise that therapies targeting *ABCA4* are an active area of research. However, *ABCA4* represents a challenging therapeutic target and therapies lag behind those developed for other genes (Piotter et al., 2021).

Gene replacement therapies targeting *ABCA4* have been the subject of intense research. While AAV vectors are very commonly used, AAVs only have a carrying capacity of ~4.7 kb (Grieger and Samulski, 2005) while the *ABCA4* gene encodes a 6.8 kb transcript (Piotter et al., 2022). Alternative viral and non-viral vectors have been explored, including using lentivirus, nanoparticles, equine infectious anaemia virus, and dual AAV vectors have

been developed for *ABCA4* (Parker et al., 2016; Dyka et al., 2019; Parker et al., 2022; Sun et al., 2022).

Gene editing approaches have also been explored to rescue *ABCA4* mutations. A large proportion (~63%) of pathogenic *ABCA4* mutations are theoretically amenable to gene editing and 35-47% of these are within range of a PAM site which is required for targeting with CRISPR/Cas9 (Piotter et al., 2022). Methods to edit single bases such as base editing and prime editing are also being explored (Komor et al., 2016; Anzalone et al., 2019). *ABCA4* exhibits a very broad range of pathogenic mutations (>2000 LOVD, https://www.lovd.nl/ accessed 30/3/2023), potentially requiring targeting of a large number of variants. However it has been estimated that correction of the 5 most common pathogenic *ABCA4* variants would be beneficial in up to 21% of cases (Piotter et al., 2022).

In addition, AONs targeting *ABCA4* have been used to rescue splice defects *in vivo* (Tomkiewicz et al., 2021; Suárez-Herrera et al., 2022) and proposed as a potential therapy for the common *ABCA4:* c.5461-10T>C variant (Kaltak et al., 2022). A number of small molecule therapies have also been tested, aiming to modulate the visual cycle or reduce oxidative stress, but these have reported mixed results (Issa et al., 2015; MacDonald and Sieving, 2018; Piccardi et al., 2019; Piotter et al., 2021).

Cell replacement therapies have also been trialled for *ABCA4* disease, aiming to deliver functional RPE cells to the retina (Schwartz et al., 2012; Song et al., 2015). In isolation this approach may only temporarily alleviate disease progression as *ABCA4* dysfunction in the photoreceptors will still result in cumulative damage.

6.3   Future work

In addition to the topics already discussed, there are other research areas which remain under explored in the study of *ABCA4* and IRDs more generally. The frequency of *ABCA4*-associated disease remains difficult to estimate. While a frequency of 1 in 6,578 has been proposed (Hanany et al., 2020), this remains an area of debate. As there is significant phenotype variability, it is

likely that many individuals with mild disease are not screened and as a result it is difficult to accurately estimate the frequency of *ABCA4*-associated disease. Further, the population carrier rate is also difficult to estimate and likely varies significantly between populations with a high number of founder mutations in *ABCA4* (Cremers et al., 2020). Continuing to screen more populations is valuable for the identification of potential therapy targets, for example for variants which are rare in the general population but more common in a subpopulation.

Further, the interaction with *cis* and *trans* acting modifiers, as well as environmental affects remains to be fully explored. Some of these modifiers have already been identified, for example the role of *PRPH2* variants (Poloschek et al., 2010; Zernant et al., 2022). The interaction of these modifiers may go some way to explaining the phenotypic variability which is observed in pedigrees affected by *ABCA4*-associated disease. The identification of modifiers of *ABCA4*-disease requires both the study of large cohorts, and the in depth examination of large pedigrees. The continued uptake of long-read sequencing technologies will greatly aid this process.

The methods used in Chapter 3, long-range PCR with nanopore sequencing and CATCH with long-range sequencing, (and future use of long-read WGS) promise to be an effective method for phasing variants of interest in *ABCA4*. This will be useful for the identification of complex alleles and modifier variants moving forwards. However, these methods did have limitations which impacted their use. While long-range PCR was a cost effective method designed for use on samples which had been repeatedly freeze thawed, the relatively short read lengths introduced difficulties in creating larger phase blocks, which limited the utility of this method. In contrast, CATCH allowed for high confidence phasing of distal variants in *ABCA4*, but required a fresh blood extraction and specialist equipment. There remains a trade-off between the read length which is obtained and the practical utility of the workflow for methods which use long-read sequencing. When fresh samples can be obtained, the full potential of long-read sequencing methods such as CATCH is clear. In future, more routine long-read sequencing, in particular long-read WGS, will require updated best practices for DNA extraction, handling and storage.

The work which has been undertaken to elucidate the mechanism of *ABCA4*-related disease, including in this study, will contribute to the implementation of personalised medicine for people affected by *ABCA4*-associated disease. The benefits of personalised medicine have been proposed frequently in the past and cost effective sequencing has started to allow for its implementation (Harvey et al., 2012; Brittain et al., 2017; Vicente et al., 2020). This can range from personalised lifestyle alterations to personalised therapies and treatment regimens (Brittain et al., 2017; Essadi et al., 2023). While the promise of personalised medicine has potential widespread and significant positive impacts on human health worldwide, significant ethical challenges remain to ensure that this continues. In the past, wilfully misunderstood and mischaracterised genetic concepts have been used to justify racism, discrimination and ethnic cleansing (Farber, 2008). It is worth remembering that many human genetics department began as eugenics departments (Rutherford, 2023), and safeguards are required to continue to protect patients. It is also important that patients are able to undergo genetic testing without future financial penalty, for example from increased insurance costs. Increasingly sophisticated and personalised genetic screening increases the vulnerability of the resulting data. It is also important that the befits of genetic research are accessible to as many people as possible. While the cost of genetic screening continues to fall, comprehensive screening, in particular long-read WGS is very expensive. In addition, personalised treatments such as gene therapies are often extremely expensive, leading to debates on whether these should be funded at patient expense. It would be a poor outcome if, from all of the work which has gone into genetic research, comprehensive genetic screening and personalised therapies were available only to a small number of people.

This project was undertaken as part of the StarT consortium which was established to investigate *ABCA4*-associated disease in detail. In addition to the results of this study, the outcomes of the other projects in the consortium will add to the growing body of work which will characterise *ABCA4*-associated disease and allow *ABCA4* to be used as a model gene for other genes implicated in IRDs. As mutations in *ABCA4* are the most common cause of IRD, by drawing together all of the samples available to the members of the consortium into a single large cohort, it was possible to investigate disease mechanisms which would have been difficult using smaller cohorts. The other projects in the consortium focused on diverse topics such as the

regulation of *ABCA4* expression, identification of novel transcripts of *ABCA4*, the identification of novel pathogenic variants in *ABCA4*, the development of potential therapies for *ABCA4*-associated disease and novel models of *ABCA4* (Corradi et al., 2022; De Angeli et al., 2022; Kaltak et al., 2023; Ruiz-Ceja et al., 2023). This work will further characterise *ABCA4*, elucidate the mechanisms of *ABCA4*-associated disease and will also indicate areas of research which are likely to be relevant for other genes associated with recessive IRDs.

In conclusion, although further research is required, the work in this thesis promises to be of direct benefit to patients with IRDs. The development of long-read sequencing methods to phase distal variants in IRD-related genes such as *ABCA4* will allow for the study of complex alleles and modifier alleles, although further research is required to allow for the widespread use of these methods. In addition, long-read sequencing allowed for the comprehensive characterisation of causative deletions in IRD genes, facilitating characterisation of these variants at high resolution. Finally a cost effective panel was used to provide a genetic diagnosis to 36 patients (following confirmation by an accredited clinical laboratory), providing a direct benefits to the patients and their families. This also resulted in the discovery of four novel, clinically relevant variants which add to a growing body of evidence which can be used to infer the mechanism of disease in the affected genes.

7 References

Abd El-Aziz, M.M., Barragan, I., O'Driscoll, C.A., Goodstadt, L., Prigmore, E., Borrego, S., Mena, M., Pieras, J.I., El-Ashry, M.F., Safieh, L.A., Shah, A., Cheetham, M.E., Carter, N.P., Chakarova, C., Ponting, C.P., Bhattacharya, S.S. and Antinolo, G. 2008. EYS, encoding an ortholog of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa. *Nature genetics.* **40**(11), pp.1285–1287.

Abel, H.J., Larson, D.E., Regier, A.A., Chiang, C., Das, I., Kanchi, K.L., Layer, R.M., Neale, B.M., Salerno, W.J., Reeves, C., Buyske, S., Abecasis, G.R., Appelbaum, E., Baker, J., Banks, E., Bernier, R.A., Bloom, T., Boehnke, M., Boerwinkle, E., Bottinger, E.P., Brant, S.R., Burchard, E.G., Bustamante, C.D., Chen, L., Cho, J.H., Chowdhury, R., Christ, R., Cook, L., Cordes, M., Courtney, L., Cutler, M.J., Daly, M.J., Damrauer, S.M., Darnell, R.B., Deluca, T., Dinh, H., Doddapaneni, H., Eichler, E.E., Ellinor, P.T., Estrada, A.M., Farjoun, Y., Felsenfeld, A., Foroud, T., Freimer, N.B., Fronick, C., Fulton, L., Fulton, R., Gabriel, S., Ganel, L., Gargeya, S., Germer, G., Geschwind, D.H., Gibbs, R.A., Goldstein, D.B., Grove, M.L., Gupta, N., Haiman, C.A., Han, Y., Howrigan, D., Hu, J., Hutter, C., Iossifov, I., Ji, B., Jorde, L.B., Jun, G., Kane, J., Kang, C.J., Kang, H.M., Kathiresan, S., Kenny, E.E., Khaira, L., Khan, Z., Khera, A., Kooperberg, C., Krasheninina, O., Kraus, W.E., Kugathasan, S., Laakso, M., Lappalainen, T., Locke, A.E., Loos, R.J.F., Ly, A., Maier, R., Maniatis, T., Le Marchand, L., Marcus, G.M., Mayeux, R.P., McGovern, D.P.B., Mendoza, K.S., Menon, V., Metcalf, G.A., Momin, Z., Narzisi, G., Nelson, J., Nessner, C., Newberry, R.D., North, K.E., Palotie, A., Peters, U., Ponce, J., Pullinger, C., Quinlan, A., Rader, D.J., Rich, S.S., Ripatti, S., Roden, D.M., Salomaa, V., Santibanez, J., Shah, S.H., Shoemaker, M.B., Sofia, H., Stephan, T., Stevens, C., Targan, S.R., Taskinen, M.R., Tibbetts, K., Tolonen, C., Turner, T., De Vries, P., Waligorski, J., Walker, K., Wang, V.O., Wigler, M., Wilson, R.K., Winterkorn, L., Wojcik, G., Xing, J., Young, E., Yu, B., Zhang, Y., Matise, T.C., Muzny, D.M., Zody, M.C., Lander, E.S., Dutcher, S.K., Stitziel, N.O. and Hall, I.M. 2020. Mapping and characterization of structural variation in 17,795 human genomes. *Nature 2020 583:7814.* **583**(7814), pp.83–89.

Abyzov, A., Urban, A.E., Snyder, M. and Gerstein, M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome research.* **21**(6), pp.974–984.

Adams, N.A., Awadein, A. and Toma, H.S. 2007. The retinal ciliopathies. *Ophthalmic genetics.* **28**(3), pp.113–125.

Agte, S., Junek, S., Matthias, S., Ulbricht, E., Erdmann, I., Wurm, A., Schild, D., Käs, J.A. and Reichenbach, A. 2011. Müller Glial Cell-Provided Cellular Light Guidance through the Vital Guinea-Pig Retina. *Biophysical Journal.* **101**(11), pp.2611–2619.

Ahn, J., Beharry, S., Molday, L.L. and Molday, R.S. 2003. Functional interaction between the two halves of the photoreceptor-specific ATP binding cassette protein ABCR (ABCA4). Evidence for a non-exchangeable ADP in the first nucleotide binding domain. *The Journal of*

biological chemistry. **278**(41), pp.39600–39608.

Ahn, J., Wong, J.T. and Molday, R.S. 2000. The effect of lipid environment and retinoids on the ATPase activity of ABCR, the photoreceptor ABC transporter responsible for Stargardt macular dystrophy. *The Journal of biological chemistry*. **275**(27), pp.20399–20405.

Aicher, J.K., Jewell, P., Vaquero-Garcia, J., Barash, Y. and Bhoj, E.J. 2020. Mapping RNA splicing variations in clinically-accessible and non-accessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genetics in medicine : official journal of the American College of Medical Genetics*. **22**(7), p.1181.

Aísa-Marín, I., García-Arroyo, R., Mirra, S. and Marfany, G. 2021. The Alter Retina: Alternative Splicing of Retinal Genes in Health and Disease. *International Journal of Molecular Sciences 2021, Vol. 22, Page 1855*. **22**(4), p.1855.

Akhras, M.S., Unemo, M., Thiyagarajan, S., Nyrén, P., Davis, R.W., Fire, A.Z. and Pourmand, N. 2007. Connector Inversion Probe Technology: A Powerful One-Primer Multiplex DNA Amplification System for Numerous Scientific Applications. *PLoS ONE*. **2**(9).

Ala-Laurila, P., Greschner, M., Chichilnisky, E.J. and Rieke, F. 2011. Cone photoreceptor contributions to noise and correlations in the retinal output. *Nature neuroscience*. **14**(10), p.1309.

Ala-Laurila, P. and Rieke, F. 2014. Coincidence Detection of Single-Photon Responses in the Inner Retina at the Sensitivity Limit of Vision. *Current Biology*. **24**(24), pp.2888–2898.

Albrecht, C. and Viturro, E. 2007. The ABCA subfamily--gene and protein structures, functions and associated hereditary diseases. *Pflugers Archiv : European journal of physiology*. **453**(5), pp.581–589.

Alkan, C., Coe, B.P. and Eichler, E.E. 2011. Genome structural variation discovery and genotyping. *Nature reviews. Genetics*. **12**(5), pp.363–376.

Allikmets, R., Singh, N., Sun, H., Shroyer, N.F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A., Rattner, A., Smallwood, P., Li, Y., Anderson, K.L., Lewis, R.A., Nathans, J., Leppert, M., Dean, M. and Lupski, J.R. 1997. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Starqardt macular dystrophy. *Nature Genetics 1997 15:3*. **15**(3), pp.236–246.

Allikmets, R., Zernant, J. and Lee, W. 2018. Penetrance of the ABCA4 p.Asn1868Ile Allele in Stargardt Disease. . **59**(13).

Alves, C.H., Pellissier, L.P., Vos, R.M., Garrido, M.G., Sothilingam, V., Seide, C., Beck, S.C., Klooster, J., Furukawa, T., Flannery, J.G., Verhaagen, J., Seeliger, M.W. and Wijnholds, J. 2014. Targeted ablation of Crb2 in photoreceptor cells induces retinitis pigmentosa. *Human Molecular Genetics*. **23**(13), pp.3384–3401.

Alves, C.H., Pellissier, L.P. and Wijnholds, J. 2014. The CRB1 and adherens junction complex proteins in retinal development and maintenance. *Progress in Retinal and Eye Research*. **40**, pp.35–52.

Anderson, K.L., Baird, L., Lewis, R.A., Chinault, A.C., Otterud, B., Leppert, M. and Lupski, J.R. 1995. A YAC contig encompassing the recessive Stargardt disease gene (STGD) on chromosome 1p. *American Journal of Human Genetics.* **57**(6), p.1351.

Anderson, R., Rotenstreich, Y., Fishman, G.A. and Anderson, R.J. 2003. *Visual acuity loss and clinical observations in a large series of patients with Stargardt disease* [Online]. Ophthalmology. [Accessed 30 May 2020]. Available from: https://pubmed.ncbi.nlm.nih.gov/12799240/.

Anderson, R.E. and Maude, M.B. 1970. Phospholipids of Bovine Rod Outer Segments. *Biochemistry.* **9**(18), pp.3624–3628.

De Angeli, P., Reuter, P., Hauser, S., Schöls, L., Stingl, K., Wissinger, B. and Kohl, S. 2022. Effective splicing restoration of a deep-intronic ABCA4 variant in cone photoreceptor precursor cells by CRISPR/ Sp Cas9 approaches. *Molecular therapy. Nucleic acids.* **29**, pp.511–524.

Anna, A. and Monika, G. 2018. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics 2018 59:3.* **59**(3), pp.253–268.

Anon 2015. Genome in a bottle—a human DNA standard. *Nature Biotechnology.* **33**(7), pp.675–675.

Anzalone, A. V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. and Liu, D.R. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature 2019 576:7785.* **576**(7785), pp.149–157.

Aparisi, M.J., Aller, E., Fuster-García, C., García-García, G., Rodrigo, R., Vázquez-Manrique, R.P., Blanco-Kelly, F., Ayuso, C., Roux, A.F., Jaijo, T. and Millán, J.M. 2014. Targeted next generation sequencing for molecular diagnosis of Usher syndrome. *Orphanet Journal of Rare Diseases.* **9**(1), pp.1–15.

Arana, M.E., Seki, M., Wood, R.D., Rogozin, I.B. and Kunkel, T.A. 2008. Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic acids research.* **36**(11), pp.3847–3856.

Ardell, M.D., Aragon, I., Oliveira, L., Porche, G.E., Burke, E. and Pittler, S.J. 1996. The beta subunit of human rod photoreceptor cGMP-gated cation channel is generated from a complex transcription unit. *FEBS letters.* **389**(2), pp.213–218.

Ardell, M.D., Makhija, A.K., Oliveira, L., Miniou, P., Viegas-Pequignot, E. and Pittler, S.J. 1995. cDNA, gene structure, and chromosomal localization of human GAR1 (CNCG3L), a homolog of the third subunit of bovine photoreceptor cGMP-gated channel. *Genomics.* **28**(1), pp.32–38.

Arikawa, K., Molday, L.L., Molday, R.S. and Williams, D.S. 1992. Localization of peripherin/rds in the disk membranes of cone and rod photoreceptors: relationship to disk membrane morphogenesis and retinal degeneration. *The Journal of cell biology.* **116**(3), pp.659–667.

Ashton, P.P.M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J. and O'Grady, J. 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance

island. . **33**(3).

Audo, I., Bujakowska, K.M., Léveillard, T., Mohand-Sad, S., Lancelot, M.E., Germain, A., Antonio, A., Michiels, C., Saraiva, J.P., Letexier, M., Sahel, J.A., Bhattacharya, S.S. and Zeitz, C. 2012. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet Journal of Rare Diseases*. **7**(1).

Ba-Abbad, R., Robson, A.G., Yap, Y.C., Moore, A.T., Webster, A.R. and Holder, G.E. 2014. Prph2 mutations as a cause of electronegative ERG. *Retina (Philadelphia, Pa.)*. **34**(6), pp.1235–1243.

Bacolla, A., Jaworski, A., Connors, T.D. and Wells, R.D. 2001. Pkd1 unusual DNA conformations are recognized by nucleotide excision repair. *The Journal of biological chemistry*. **276**(21), pp.18597–18604.

Bacolla, A., Jaworski, A., Larson, J.E., Jakupciak, J.P., Chuzhanova, N., Abeysinghe, S.S., O'Connell, C.D., Cooper, D.N. and Wells, R.D. 2004. Breakpoints of gross deletions coincide with non-B DNA conformations. *Proceedings of the National Academy of Sciences of the United States of America*. **101**(39), pp.14162–14167.

Bahn, M.S. and Ko, Y.G. 2023. PROM1-mediated cell signal transduction in cancer stem cells and hepatocytes. *BMB Reports*. **56**(2), p.65.

Baker, K.E. and Parker, R. 2004. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current Opinion in Cell Biology*. **16**(3), pp.293–299.

Baker, S.W., Murrell, J.R., Nesbitt, A.I., Pechter, K.B., Balciuniene, J., Zhao, X., Yu, Z., Denenberg, E.H., DeChene, E.T., Wilkens, A.B., Bhoj, E.J., Guan, Q., Dulik, M.C., Conlin, L.K., Abou Tayoun, A.N., Luo, M., Wu, C., Cao, K., Sarmady, M., Bedoukian, E.C., Tarpinian, J., Medne, L., Skraban, C.M., Deardorff, M.A., Krantz, I.D., Krock, B.L. and Santani, A.B. 2019. Automated Clinical Exome Reanalysis Reveals Novel Diagnoses. *The Journal of Molecular Diagnostics*. **21**(1), pp.38–48.

Ballouz, S., Dobin, A. and Gillis, J.A. 2019. Is it time to change the reference genome? *Genome Biology*. **20**(1), pp.1–9.

Barbitoff, Y.A., Bezdvornykh, I. V., Polev, D.E., Serebryakova, E.A., Glotov, A.S., Glotov, O.S. and Predeus, A. V. 2017. Catching hidden variation: systematic correction of reference minor allele annotation in clinical variant calling. *Genetics in Medicine 2018 20:3*. **20**(3), pp.360–364.

Bareil, C., Hamel, C.P., Delague, V., Arnaud, B., Demaille, J. and Claustres, M. 2001. Segregation of a mutation in CNGB1 encoding the beta-subunit of the rod cGMP-gated channel in a family with autosomal recessive retinitis pigmentosa. *Human genetics*. **108**(4), pp.328–334.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**(5819), pp.1709–1712.

Barrangou, R., Jinek, M., Wiedenheft, B., Zhou, K., Doudna, J.A. and Charpentier, E. 2012. CRISPR Provides Acquired Resistance Against

Viruses in Prokaryotes. *Science*. **315**(5819), pp.1709–1712.

Barret, D.C.A., Kaupp, U.B. and Marino, J. 2022. The structure of cyclic nucleotide-gated channels in rod and cone photoreceptors. *Trends in Neurosciences*. **45**(10), pp.763–776.

Battu, R., Ratra, D. and Gopal, L. 2022. Newer therapeutic options for inherited retinal diseases: Gene and cell replacement therapy. *Indian journal of ophthalmology*. **70**(7), pp.2316–2325.

Batzer, M.A. and Deininger, P.L. 2002. Alu repeats and human genomic diversity. *Nature reviews. Genetics*. **3**(5), pp.370–379.

Baudhuin, L.M., Lagerstedt, S.A., Klee, E.W., Fadra, N., Oglesbee, D. and Ferber, M.J. 2015. Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. *The Journal of Molecular Diagnostics*. **17**(4), pp.456–461.

Becirovic, E., Böhm, S., Nguyen, O.N.P., Riedmayr, L.M., Koch, M.A., Schulze, E., Kohl, S., Borsch, O., Santos-Ferreira, T., Ader, M., Michalakis, S. and Biel, M. 2016. In Vivo Analysis of Disease-Associated Point Mutations Unveils Profound Differences in mRNA Splicing of Peripherin-2 in Rod and Cone Photoreceptors. *PLoS Genetics*. **12**(1).

Beck, C.R., Garcia-Perez, J.L., Badge, R.M. and Moran, J. V. 2011. LINE-1 elements in structural variation and disease. *Annual review of genomics and human genetics*. **12**, pp.187–215.

Behjati, S. and Tarpey, P.S. 2013. What is next generation sequencing? *Archives of Disease in Childhood - Education and Practice*. **98**(6), pp.236–238.

Beitel, C.W., Froenicke, L., Lang, J.M., Korf, I.F., Michelmore, R.W., Eisen, J.A. and Darling, A.E. 2014. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ*. **2**(1).

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Catenazzi, M.C.E., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer,

N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D. V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C., Pike, A.C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, J., Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R. and Smith, A.J. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature 2008 456:7218*. **456**(7218), pp.53–59.

Bernardis, I., Chiesi, L., Tenedini, E., Artuso, L., Percesepe, A., Artusi, V., Simone, M.L., Manfredini, R., Camparini, M., Rinaldi, C., Ciardella, A., Graziano, C., Balducci, N., Tranchina, A., Cavallini, G.M., Pietrangelo, A., Marigo, V. and Tagliafico, E. 2016. Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing: Added Value of Targeted Next-Generation Sequencing. *BioMed Research International*. **2016**.

Bernstein, P.S., Tammur, J., Singh, N., Hutchinson, A., Dixon, M., Pappas, C.M., Zabriskie, N.A., Zhang, K., Petrukhin, K., Leppert, M. and Allikmets, R. 2001. Diverse macular dystrophy phenotype caused by a novel complex mutation in the ELOVL4 gene. *Investigative Ophthalmology & Visual Science*. **42**(13), pp.3331–3336.

Berntson, A., Smith, R.G. and Taylor, W.R. 2016. Transmission of single photon signals through a binary synapse in the mammalian retina. *Visual Neuroscience*. **21**(5), pp.693–702.

Besharse, J.C. and Horst, C.J. 1990. The Photoreceptor Connecting Cilium A Model for the Transition Zone. *Ciliary and Flagellar Membranes*., pp.389–417.

Bewicke-Copley, F., Arjun Kumar, E., Palladino, G., Korfi, K. and Wang, J. 2019. Applications and analysis of targeted genomic sequencing in cancer studies. *Computational and Structural Biotechnology Journal*. **17**, pp.1348–1359.

Bhartiya, S., Sharma, N., Verma, S., Dada, T., Gagrani, M., Chauhan, N. and Satpute, K. 2022. Ocular and Systemic Factors Associated with Glaucoma. *Journal of current glaucoma practice*. **16**(3), pp.179–191.

Biswas-Fiss, E.E., Affet, S., Ha, M. and Biswas, S.B. 2012. Retinoid binding

properties of nucleotide binding domain 1 of the Stargardt disease-associated ATP binding cassette (ABC) transporter, ABCA4. *The Journal of biological chemistry*. **287**(53), pp.44097–44107.

Boon, C.J.F., Van Schooneveld, M.J., Den Hollander, A.I., Van Lith-Verhoeven, J.J.C., Zonneveld-Vrieling, M.N., Theelen, T., Cremers, F.P.M., Hoyng, C.B. and Klevering, B.J. 2007. Mutations in the peripherin/RDS gene are an important cause of multifocal pattern dystrophy simulating STGD1/fundus flavimaculatus. *British Journal of Ophthalmology*. **91**(11), pp.1504–1511.

Bowden, R., Davies, R.W., Heger, A., Pagnamenta, A.T., Cesare, M. de, Oikkonen, L.E., Parkes, D., Freeman, C., Dhalla, F., Patel, S.Y., Popitsch, N., Ip, C.L.C., Roberts, H.E., Salatino, S., Lockstone, H., Lunter, G., Taylor, J.C., Buck, D., Simpson, M.A. and Donnelly, P. 2019. Sequencing of human genomes with nanopore technology. *Nature Communications*. **10**.

Bowmaker, J.K. and Hunt, D.M. 2006. Evolution of vertebrate visual pigments. *Current biology : CB*. **16**(13).

Bridges, C.B. 1936. The Bar 'gene' a duplication. *Science*. **83**(2148), pp.210–211.

Brittain, H.K., Scott, R. and Thomas, E. 2017. The rise of the genome and personalised medicine. *Clinical Medicine*. **17**(6), p.545.

Britten-Jones, A.C., Gocuk, S.A., Goh, K.L., Huq, A., Edwards, T.L. and Ayton, L.N. 2023. The Diagnostic Yield of Next Generation Sequencing in Inherited Retinal Diseases: A Systematic Review and Meta-analysis. *American Journal of Ophthalmology*. **249**, pp.57–73.

Britten-Jones, A.C., Jin, R., Gocuk, S.A., Cichello, E., O'Hare, F., Hickey, D.G., Edwards, T.L. and Ayton, L.N. 2022. The safety and efficacy of gene therapy treatment for monogenic retinal and optic nerve diseases: A systematic review. *Genetics in Medicine*. **24**(3), pp.521–534.

Brown, S.D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., Land, M.L., Dassanayake, A. and Köpke, M. 2014. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of Clostridium autoethanogenum and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnology for biofuels*. **7**(1).

Browning, S.R. and Browning, B.L. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics*. **81**(5), pp.1084–1097.

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., Arno, G., Roberts, L., Greenberg, J., Albert, S., Gilissen, C., Aben, M., Rebello, G., Mead, S., Raymond, F.L., Corominas, J., Smith, C.E.L., Kremer, H., Downes, S., Black, G.C., Webster, A.R., Inglehearn, C.F., van den Born, L.I., Koenekoop, R.K., Michaelides, M., Ramesar, R.S., Hoyng, C.B., Mundlos, S., Mhlanga, M.M., Cremers, F.P.M., Cheetham, M.E., Roosing, S. and Hardcastle, A.J. 2020. Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene

Contact in Dominant Retinitis Pigmentosa. *American journal of human genetics*. **107**(5), pp.802–814.

de Bruijn, S.E., Rodenburg, K., Corominas, J., Ben-Yosef, T., Reurink, J., Kremer, H., Whelan, L., Plomp, A.S., Berger, W., Farrar, G.J., Ferenc Kovács, Á., Fajardy, I., Hitti-Malin, R.J., Weisschuh, N., Weener, M.E., Sharon, D., Pennings, R.J.E., Haer-Wigman, L., Hoyng, C.B., Nelen, M.R., Vissers, L.E.L.M., van den Born, L.I., Gilissen, C., Cremers, F.P.M., Hoischen, A., Neveling, K. and Roosing, S. 2023. Optical genome mapping and revisiting short-read genome sequencing data reveal previously overlooked structural variants disrupting retinal disease−associated genes. *Genetics in Medicine*. **25**(3), p.100345.

Bujakowska, K., Audo, I., Mohand-Säid, S., Lancelot, M.E., Antonio, A., Germain, A., Eveillard, T.Ĺ., Letexier, Ḿelanie, Saraiva, J.P., Lonjou, C., Carpentier, W., Sahel, J.A., Bhattacharya, S.S. and Zeitz, C. 2012. CRB1 mutations in inherited retinal dystrophies. . **33**(2), pp.306–315.

Bungert, S., Laurie L., M. and Robert S., M. 2001. *Membrane topology of the ATP binding cassette transporter ABCR and its relationship to ABC1 and related ABCA transporters: identification of N-linked glycosylation sites* [Online]. J Biol Chem. [Accessed 30 May 2020]. Available from: https://pubmed.ncbi.nlm.nih.gov/11320094/.

Burwinkel, B. and Kilimann, M.W. 1998. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *Journal of molecular biology*. **277**(3), pp.513–517.

Busck, A. and OSTERBER, G. 1935. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica*.

Buskin, A., Zhu, L., Chichagova, V., Basu, B., Mozaffari-Jovin, S., Dolan, D., Droop, A., Collin, J., Bronstein, R., Mehrotra, S., Farkas, M., Hilgen, G., White, K., Pan, K.T., Treumann, A., Hallam, D., Bialas, K., Chung, G., Mellough, C., Ding, Y., Krasnogor, N., Przyborski, S., Zwolinski, S., Al-Aama, J., Alharthi, S., Xu, Y., Wheway, G., Szymanska, K., McKibbin, M., Inglehearn, C.F., Elliott, D.J., Lindsay, S., Ali, R.R., Steel, D.H., Armstrong, L., Sernagor, E., Urlaub, H., Pierce, E., Lührmann, R., Grellscheid, S.N., Johnson, C.A. and Lako, M. 2018. Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. *Nature communications*. **9**(1).

Callahan, B.J., Grinevich, D., Thakur, S., Balamotis, M.A. and Yehezkel, T. Ben 2021. Ultra-accurate microbial amplicon sequencing with synthetic long reads. *Microbiome*. **9**(1), pp.1–13.

Canard, B. and Sarfati, R.S. 1994. DNA polymerase fluorescent substrates with reversible 3'-tags. *Gene*. **148**(1), pp.1–6.

Cao, M., Zhang, L., Wang, J.H., Zeng, H., Peng, Y., Zou, J., Shi, J., Zhang, L., Li, Y., Yoshida, S., Tang, L. and Zhou, Y. 2019. Identifying circRNA-associated-ceRNA networks in retinal neovascularization in mice. *International journal of medical sciences*. **16**(10), pp.1356–1365.

Carss, K., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S., Megy, K., Grozeva, D., Dewhurst, E., Malka, S., Plagnol, V., Penkett, C.J., Stirrups, K., Rizzo, R., Wright, G., Josifova, D., Bitner-Glindzicz, M.,

Scott, R., Clement, E., Allen, L., Armstrong, R., Brady, A.F., Carmichael, J., Chitre, M., Henderson, R., Hurst, J., MacLaren, R., Murphy, E., Paterson, J., Rosser, E., Thompson, D., Wakeling, E., Ouwehand, W.H., Michaelides, M., Moore, A., Aitman, T., Alachkar, H., Ali, S., Allsup, D., Ambegaonkar, G., Anderson, J., Antrobus, R., Arumugakani, G., Ashford, S., Astle, W., Attwood, A., Austin, S., Bacchelli, C., Bakchoul, T., Bariana, T.K., Baxendale, H., Bennett, D., Bethune, C., Bibi, S., Bleda, M., Boggard, H., Bolton-Maggs, P., Booth, C., Bradley, J.R., Brady, A., Brown, M., Browning, M., Bryson, C., Burns, S., Calleja, P., Canham, N., Caulfield, M., Chalmers, E., Chandra, A., Chinnery, P., Church, C., Clements-Brod, N., Clowes, V., Coghlan, G., Collins, P., Cooper, N., Creaser-Myers, A., DaCosta, R., Daugherty, L., Davies, S., Davis, J., De Vries, M., Deegan, P., Deevi, S.V.V., Deshpande, C., Devlin, L., Doffinger, R., Dormand, N., Drewe, E., Edgar, D., Egner, W., Erber, W.N., Everington, T., Favier, R., Firth, H., Fletcher, D., Flinter, F., Fox, J.C., Frary, A., Freson, K., Furie, B., Furnell, A., Gale, D., Gardham, A., Gattens, M., Ghali, N., Ghataorhe, P.K., Ghurye, R., Gibbs, S., Gilmour, K., Gissen, P., Goddard, S., Gomez, K., Gordins, P., Gräf, S., Greene, D., Greenhalgh, A., Greinacher, A., Grigoriadou, S., Hackett, S., Hadinnapola, C., Hague, R., Haimel, M., Halmagyi, C., Hammerton, T., Hart, D., Hayman, G., Heemskerk, J.W.M., Hensiek, A., Henskens, Y., Herwadkar, A., Holden, S., Holder, M., Holder, S., Hu, F., Huissoon, A., Humbert, M., James, R., Jolles, S., Kazmi, R., Keeling, D., Kelleher, P., Kelly, A.M., Kennedy, F., Kiely, D., Kingston, N., Koziell, A., Krishnakumar, D., Kuijpers, T.W., Kumararatne, D., Kurian, M., Laffan, M.A., Lambert, M.P., Allen, H.L., Lawrie, A., Lear, S., Lees, M., Lentaigne, C., Liesner, R., Linger, R., Longhurst, H., Lorenzo, L., Machado, R., Mackenzie, R., Maher, E., Maimaris, J., Mangles, S., Manson, A., Mapeta, R., Markus, H.S., Martin, J., Masati, L., Mathias, M., Matser, V., Maw, A., McDermott, E., McJannet, C., Meacham, S., Meehan, S., Mehta, S., Millar, C.M., Moledina, S., Morrell, N., Mumford, A., Murng, S., Nejentsev, S., Noorani, S., Nurden, P., Oksenhendler, E., Papadia, S., Park, S.M., Parker, A., Pasi, J., Patch, C., Payne, J., Peacock, A., Peerlinck, K., Pepke-Zaba, J., Perry, D.J., Pollock, V., Polwarth, G., Ponsford, M., Qasim, W., Quinti, I., Rankin, S., Rankin, J., Raymond, F.L., Rehnstrom, K., Reid, E., Rhodes, C.J., Richards, M., Richardson, S., Richter, A., Roberts, I., Rondina, M., Roughley, C., Rue-Albrecht, K., Webster, A., Sandford, R., Santra, S., Sargur, R., Savic, S., Schulman, S., Schulze, H., Scully, M., Seneviratne, S., Sewell, C., Shamardina, O., Shipley, D., Simeoni, I., Sivapalaratnam, S., Smith, K., Sohal, A., Southgate, L., Staines, S., Staples, E., Stauss, H., Stein, P., Stock, S., Suntharalingam, J., Tait, R.C., Talks, K., Tan, Y., Thachil, J., Thaventhiran, J., Thomas, E., Thomas, M., Thrasher, A., Tischkowitz, M., Titterton, C., Toh, C.H., Toshner, M., Treacy, C., Trembath, R., Tuna, S., Turek, W., Turro, E., Van Geet, C., Veltman, M., Vogt, J., von Ziegenweldt, J., Noordegraaf, A.V., Wanjiku, I., Warner, T.Q., Wassmer, E., Watkins, H., Welch, S., Westbury, S., Wharton, J., Whitehorn, D., Wilkins, M., Willcocks, L., Williamson, C., Woods, G., Wort, J., Yeatman, N., Yong, P., Young, T. and Yu, P. 2017. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *American Journal of Human*

Genetics. **100**(1), p.75.

Casbon, J.A., Osborne, R.J., Brenner, S. and Lichtenstein, C.P. 2011. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic acids research*. **39**(12).

Caspersson, T., Farber, S., Foley, G.E., Kudynowski, J., Modest, E.J., Simonsson, E., Wagh, U. and Zech, L. 1968. Chemical differentiation along metaphase chromosomes. *Experimental Cell Research*. **49**(1), pp.219–222.

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., Jolly, J.K., Luo, E., Lujan, B.J., Ong, T., Girach, A., Black, G.C.M., Gregori, N.Z., Davis, J.L., Rosa, P.R., Lotery, A.J., Lam, B.L., Stanga, P.E. and MacLaren, R.E. 2020. Initial results from a first-in-human gene therapy trial on X-linked retinitis pigmentosa caused by mutations in RPGR. *Nature Medicine 2020 26:3*. **26**(3), pp.354–359.

Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., Fan, X., Wen, J., Handsaker, R.E., Fairley, S., Kronenberg, Z.N., Kong, X., Hormozdiari, F., Lee, D., Wenger, A.M., Hastie, A.R., Antaki, D., Anantharaman, T., Audano, P.A., Brand, H., Cantsilieris, S., Cao, H., Cerveira, E., Chen, C., Chen, X., Chin, C.S., Chong, Z., Chuang, N.T., Lambert, C.C., Church, D.M., Clarke, L., Farrell, A., Flores, J., Galeev, T., Gorkin, D.U., Gujral, M., Guryev, V., Heaton, W.H., Korlach, J., Kumar, S., Kwon, J.Y., Lam, E.T., Lee, J.J.E., Lee, J.J.E., Lee, W.P., Lee, S.P., Li, S., Marks, P., Viaud-Martinez, K., Meiers, S., Munson, K.M., Navarro, F.C.P., Nelson, B.J., Nodzak, C., Noor, A., Kyriazopoulou-Panagiotopoulou, S., Pang, A.W.C., Qiu, Y., Rosanio, G., Ryan, M., Stütz, A., Spierings, D.C.J., Ward, A., Welch, A.M.E., Xiao, M., Xu, W., Zhang, C., Zhu, Q., Zheng-Bradley, X., Lowy, E., Yakneen, S., McCarroll, S., Jun, G., Ding, L., Koh, C.L., Ren, B., Flicek, P., Chen, K., Gerstein, M.B., Kwok, P.Y., Lansdorp, P.M., Marth, G.T., Sebat, J., Shi, X., Bashir, A., Ye, K., Devine, S.E., Talkowski, M.E., Mills, R.E., Marschall, T., Korbel, J.O., Eichler, E.E. and Lee, C. 2019. Multi-platform discovery of haplotype-resolved structural variation in human genomes. . **10**(1).

Chance, B., Sies, H. and Boveris, A. 1979. Hydroperoxide metabolism in mammalian organs. *Physiological reviews*. **59**(3), pp.527–605.

Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., Shi, X., Fulton, R.S., Ley, T.J., Wilson, R.K., Ding, L. and Mardis, E.R. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature methods*. **6**(9), pp.677–681.

Chen, L., Lee, W., de Carvalho, J.R.L., Chang, S., Tsang, S.H., Allikmets, R. and Sparrow, J.R. 2019. Multi-platform imaging in ABCA4-Associated Disease. *Scientific reports*. **9**(1).

Chen, R. and Butte, A.J. 2011. The reference human genome demonstrates high risk of type 1 diabetes and other disorders. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing.*, pp.231–242.

Chen, X., Jiang, C., Qin, B., Liu, G., Ji, J., Sun, X., Xu, M., Ding, S., Zhu, M., Huang, G., Yan, B. and Zhao, C. 2017. LncRNA ZNF503-AS1 promotes RPE differentiation by downregulating ZNF503 expression. *Cell death & disease*. **8**(9).

Chen, Z., Pham, L., Wu, T.C., Mo, G., Xia, Y., Chan, P.L., Porter, D., Phan, T., Che, H., Tran, H., Bansal, V., Shaffer, J., Belda-Ferre, P., Humphrey, G., Knight, R., Pevzner, P., Pham, S., Wang, Y. and Lei, M. 2020. Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. *Genome research*. **30**(6), pp.898–909.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K. and Akeson, M. 2012. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nature biotechnology*. **30**(4), pp.344–348.

Cheung, R., Insigne, K.D., Yao, D., Burghard, C.P., Wang, J., Hsiao, Y.H.E., Jones, E.M., Goodman, D.B., Xiao, X. and Kosuri, S. 2019. A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. *Molecular cell*. **73**(1), p.183–194.e8.

Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., Montgomery, S.B., Battle, A., Conrad, D.F. and Hall, I.M. 2017. The impact of structural variation on human gene expression. *Nature genetics*. **49**(5), p.692.

Chiang, J.P.W., Lamey, T., McLaren, T., Thompson, J.A., Montgomery, H. and De Roach, J. 2015. Progress and prospects of next-generation sequencing testing for inherited retinal dystrophy. *http://dx.doi.org/10.1586/14737159.2015.1081057*. **15**(10), pp.1269–1275.

Chiang, J.P.W., Lamey, T.M., Wang, N.K., Duan, J., Zhou, W., McLaren, T.L., Thompson, J.A., Ruddle, J. and 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**(11), pp.4434–4440.

Chiang, T., Liu, X., Wu, T.J., Hu, J., Sedlazeck, F.J., White, S., Schaid, D., Andrade, M. de, Jarvik, G.P., Crosslin, D., Stanaway, I., Carrell, D.S., Connolly, J.J., Hakonarson, H., Groopman, E.E., Gharavi, A.G., Fedotov, A., Bi, W., Leduc, M.S., Murdock, D.R., Jiang, Y., Meng, L., Eng, C.M., Wen, S., Yang, Y., Muzny, D.M., Boerwinkle, E., Salerno, W., Venner, E. and Gibbs, R.A. 2019. Atlas-CNV: a validated approach to call single-exon CNVs in the eMERGESeq gene panel. *Genetics in Medicine 2019 21:9*. **21**(9), pp.2135–2144.

Cho, S.H., Kim, J.Y., Simons, D.L., Song, J.Y., Le, J.H., Swindell, E.C., Jamrich, M., Wu, S.M. and Kim, S. 2012. Genetic ablation of Pals1 in retinal progenitor cells models the retinal pathology of Leber congenital amaurosis. *Human Molecular Genetics*. **21**(12), pp.2663–2676.

Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Özen, S., Sanjad, S., Nelson-Williams, C., Farhi, A.,

Mane, S. and Lifton, R.P. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(45), pp.19096–19101.

Choi, Y., Chan, A.P., Kirkness, E., Telenti, A. and Schork, N.J. 2018. Comparison of phasing strategies for whole human genomes. *PLOS Genetics*. **14**(4), p.e1007308.

Chou, S.H., Chin, K.H. and Wang, A.H.J. 2003. Unusual DNA duplex and hairpin motifs. *Nucleic acids research*. **31**(10), pp.2461–2474.

Cideciyan, A. V., Swider, M., Aleman, T.S., Sumaroka, A., Schwartz, S.B., Roman, M.I., Milam, A.H., Bennett, J., Stone, E.M. and Jacobson, S.G. 2005. ABCA4-associated retinal degenerations spare structure and function of the human parapapillary retina. *Investigative ophthalmology & visual science*. **46**(12), pp.4739–4746.

Clark, M.J., Chen, R., Lam, H.Y.K., Karczewski, K.J., Chen, R., Euskirchen, G., Butte, A.J. and Snyder, M. 2011. Performance comparison of exome DNA sequencing technologies. *Nature biotechnology*. **29**(10), pp.908–916.

Cohen, A.I. 1961. The fine structure of the extrafoveal receptors of the rhesus monkey. *Experimental Eye Research*. **1**(2), pp.128-IN16.

Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A. V., Lowther, C., Gauthier, L.D., Wang, H., Watts, N.A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C.W., Huang, Y., Brookings, T., Sharpe, T., Stone, M.R., Valkanas, E., Fu, J., Tiao, G., Laricchia, K.M., Ruano-Rubio, V., Stevens, C., Gupta, N., Cusick, C., Margolin, L., Alföldi, J., Armean, I.M., Banks, E., Bergelson, L., Cibulskis, K., Collins, R.L., Connolly, K.M., Covarrubias, M., Cummings, B., Daly, M.J., Donnelly, S., Farjoun, Y., Ferriera, S., Francioli, L., Gabriel, S., Gauthier, L.D., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Karczewski, K.J., Laricchia, K.M., Llanwarne, C., Minikel, E. V., Munshi, R., Neale, B.M., Novod, S., O'Donnell-Luria, A.H., Petrillo, N., Poterba, T., Roazen, D., Ruano-Rubio, V., Saltzman, A., Samocha, K.E., Schleicher, M., Seed, C., Solomonson, M., Soto, J., Tiao, G., Tibbetts, K., Tolonen, C., Vittal, C., Wade, G., Wang, A., Wang, Q., Ware, J.S., Watts, N.A., Weisburd, B., Whiffin, N., Salinas, C.A.A., Ahmad, T., Albert, C.M., Ardissino, D., Atzmon, G., Barnard, J., Beaugerie, L., Benjamin, E.J., Boehnke, M., Bonnycastle, L.L., Bottinger, E.P., Bowden, D.W., Bown, M.J., Chambers, J.C., Chan, J.C., Chasman, D., Cho, J., Chung, M.K., Cohen, B., Correa, A., Dabelea, D., Daly, M.J., Darbar, D., Duggirala, R., Dupuis, J., Ellinor, P.T., Elosua, R., Erdmann, J., Esko, T., Färkkilä, M., Florez, J., Franke, A., Getz, G., Glaser, B., Glatt, S.J., Goldstein, D., Gonzalez, C., Groop, L., Haiman, C., Hanis, C., Harms, M., Hiltunen, M., Holi, M.M., Hultman, C.M., Kallela, M., Kaprio, J., Kathiresan, S., Kim, B.J., Kim, Y.J., Kirov, G., Kooner, J., Koskinen, S., Krumholz, H.M., Kugathasan, S., Kwak, S.H., Laakso, M., Lehtimäki, T., Loos, R.J.F., Lubitz, S.A., Ma, R.C.W., MacArthur, D.G., Marrugat, J., Mattila, K.M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Meigs, J.B., Melander, O., Metspalu, A., Neale, B.M., Nilsson, P.M., O'Donovan, M.C., Ongur, D., Orozco, L., Owen, M.J., Palmer, C.N.A.,

Palotie, A., Park, K.S., Pato, C., Pulver, A.E., Rahman, N., Remes, A.M., Rioux, J.D., Ripatti, S., Roden, D.M., Saleheen, D., Salomaa, V., Samani, N.J., Scharf, J., Schunkert, H., Shoemaker, M.B., Sklar, P., Soininen, H., Sokol, H., Spector, T., Sullivan, P.F., Suvisaari, J., Tai, E.S., Teo, Y.Y., Tiinamaija, T., Tsuang, M., Turner, D., Tusie-Luna, T., Vartiainen, E., Watkins, H., Weersma, R.K., Wessman, M., Wilson, J.G., Xavier, R.J., Taylor, K.D., Lin, H.J., Rich, S.S., Post, W.S., Chen, Y.D.I., Rotter, J.I., Nusbaum, C., Philippakis, A., Lander, E., Gabriel, S., Neale, B.M., Kathiresan, S., Daly, M.J., Banks, E. and Talkowski, M.E. 2020. A structural variation reference for medical and population genetics. *Nature 2020 581:7809*. **581**(7809), pp.444–451.

Combs, R., Hall, G., Payne, K., Lowndes, J., Devery, S., Downes, S.M., Moore, A.T., Ramsden, S., Black, G.C.M. and McAllister, M. 2013. Understanding the expectations of patients with inherited retinal dystrophies. *The British journal of ophthalmology*. **97**(8), pp.1057–1061.

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.C., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W. and Hurles, M.E. 2009. Origins and functional impact of copy number variation in the human genome. *Nature 2009 464:7289*. **464**(7289), pp.704–712.

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., Sims, K.B., Sweetser, D.A., Fulton, A.B., Liu, Q., Wiggs, J.L., Gai, X. and Pierce, E.A. 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**(4), pp.253–261.

Conte, I., Hadfield, K.D., Barbato, S., Carrella, S., Pizzo, M., Bhat, R.S., Carissimo, A., Karali, M., Porter, L.F., Urquhart, J., Hateley, S., O'Sullivan, J., Manson, F.D.C., Neuhauss, S.C.F., Banfi, S. and Black, G.C.M. 2015. MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proceedings of the National Academy of Sciences of the United States of America*. **112**(25), pp.E3236–E3245.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., Abdel-Hamid, H., Bader, P., McCracken, E., Niyazov, D., Leppig, K., Thiese, H., Hummel, M., Alexander, N., Gorski, J., Kussmann, J., Shashi, V., Johnson, K., Rehder, C., Ballif, B.C., Shaffer, L.G. and Eichler, E.E. 2011. A copy number variation morbidity map of developmental delay. *Nature genetics*. **43**(9), pp.838–846.

Coppieters, F., De Wilde, B., Lefever, S., De Meester, E., De Rocker, N., Van Cauwenbergh, C., Pattyn, F., Meire, F., Leroy, B.P., Hellemans, J., Vandesompele, J. and De Baere, E. 2012. Massively parallel sequencing for early molecular diagnosis in Leber congenital amaurosis. *Genetics in Medicine*. **14**(6), pp.576–585.

Cordaux, R. and Batzer, M.A. 2009. The impact of retrotransposons on human

genome evolution. *Nature reviews. Genetics*. **10**(10), pp.691–703.

Cornelis, S.S., Bax, N.M., Zernant, J., Allikmets, R., Fritsche, L.G., den Dunnen, J.T., Ajmal, M., Hoyng, C.B. and Cremers, F.P.M. 2017. In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Human Mutation*. **38**(4), pp.400–408.

Cornelis, S.S., Runhart, E.H., Bauwens, M., Corradi, Z., De Baere, E., Roosing, S., Haer-Wigman, L., Dhaenens, C.-M., Vulto-Van Silfhout, A.T., Cremers, F.P.M. and Affiliations,) 2021. Genetic risk estimates for offspring of patients with Stargardt disease. *medRxiv*., 2021.08.11.21261888.

Cornelis, S.S., Runhart, E.H., Bauwens, M., Corradi, Z., De Baere, E., Roosing, S., Haer-Wigman, L., Dhaenens, C.M., Vulto-van Silfhout, A.T. and Cremers, F.P.M. 2022. Personalized genetic counseling for Stargardt disease: Offspring risk estimates based on variant severity. . **109**(3), pp.498–507.

Corradi, Z., Salameh, M., Khan, M., Héon, E., Mishra, K., Hitti-Malin, R.J., AlSwaiti, Y., Aslanian, A., Banin, E., Brooks, B.P., Zein, W.M., Hufnagel, R.B., Roosing, S., Dhaenens, C.M., Sharon, D., Cremers, F.P.M. and AlTalbishi, A. 2022. ABCA4 c.859-25A>G, a Frequent Palestinian Founder Mutation Affecting the Intron 7 Branchpoint, Is Associated With Early-Onset Stargardt Disease. *Investigative ophthalmology & visual science*. **63**(4).

Costain, G., Cohn, R.D., Scherer, S.W. and Marshall, C.R. 2021. Genome sequencing as a diagnostic test. *CMAJ : Canadian Medical Association Journal*. **193**(42), p.E1626.

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M. and Van Broeckhoven, C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. **34**(15), pp.2666–2669.

Coutelier, M., Holtgrewe, M., Jäger, M., Flöttman, R., Mensah, M.A., Spielmann, M., Krawitz, P., Horn, D., Beule, D. and Mundlos, S. 2021. Combining callers improves the detection of copy number variants from whole-genome sequencing. *European Journal of Human Genetics 2021 30:2*. **30**(2), pp.178–186.

Cremers, F.P.M., Lee, W., Collin, R.W.J. and Allikmets, R. 2020. Clinical spectrum, genetic complexity and therapeutic approaches forretinal disease caused by ABCA4 mutations. *Progress in retinal and eye research*. **79**, p.100861.

Cronn, R., Cedroni, M., Haselkorn, T., Grover, C. and Wendel, J.F. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*. **104**(2–3), pp.482–489.

Cukras, C.A., Wong, W.T., Caruso, R., Cunningham, D., Zein, W. and Sieving, P.A. 2012. Centrifugal expansion of fundus autofluorescence patterns in Stargardt disease over time. *Archives of ophthalmology (Chicago, Ill. : 1960)*. **130**(2), pp.171–179.

Curcio, C.A. and Allen, K.A. 1990. Topography of ganglion cells in human

retina. *The Journal of comparative neurology*. **300**(1), pp.5–25.

Curcio, C.A., Sloan, K.R., Kalina, R.E. and Hendrickson, A.E. 1990. Human photoreceptor topography. *The Journal of comparative neurology*. **292**(4), pp.497–523.

Dai, P., Honda, A., Ewans, L., McGaughran, J., Burnett, L., Law, M. and Phan, T.G. 2022. Recommendations for next generation sequencing data reanalysis of unsolved cases with suspected Mendelian disorders: A systematic review and meta-analysis. *Genetics in Medicine*. **24**(8), pp.1618–1629.

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., McVean, G. and Durbin, R. 2011. The variant call format and VCFtools. *Bioinformatics*. **27**(15), pp.2156–2158.

Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.M. and White, R. 1990. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics*. **6**(3), pp.575–577.

Deamer, D., Akeson, M. and Branton, D. 2016. Three decades of nanopore sequencing. *Nature biotechnology*. **34**(5), p.518.

Deininger, P. 2011. Alu elements: know the SINEs. *Genome biology*. **12**(12).

Derwent, J.J.K., Derlacki, D.J., Hetling, J.R., Fishman, G.A., Birch, D.G., Grover, S., Stone, E.M., Pepperberg, D.R., Kang Derwent JJ, Derlacki DJ and Hetling JR 2004. Dark adaptation of rod photoreceptors in normal subjects, and in patients with Stargardt disease and an ABCA4 mutation. *Investigative ophthalmology & visual science*. **45**(7), pp.2447–2456.

Desviat, L.R., Pérez, B. and Ugarte, M. 2012. Minigenes to confirm exon skipping mutations. *Methods in molecular biology (Clifton, N.J.)*. **867**, pp.37–47.

Dhallan, R.S., Macke, J.P., Eddy, R.L., Shows, T.B., Reed, R.R., Yau, K.W. and Nathans, J. 1992. Human rod photoreceptor cGMP-gated channel: amino acid sequence, gene structure, and functional expression. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. **12**(8), pp.3248–3256.

Dharmat, R., Liu, W., Ge, Z., Sun, Z., Yang, L., Li, Y., Wang, K., Thomas, K., Sui, R. and Chen, R. 2017. IFT81 as a Candidate Gene for Nonsyndromic Retinal Degeneration. *Investigative ophthalmology & visual science*. **58**(5), pp.2483–2490.

Dimopoulos, I.S., Hoang, S.C., Radziwon, A., Binczyk, N.M., Seabra, M.C., MacLaren, R.E., Somani, R., Tennant, M.T.S. and MacDonald, I.M. 2018. Two-Year Results After AAV2-Mediated Gene Therapy for Choroideremia: The Alberta Experience. *American Journal of Ophthalmology*. **193**, pp.130–142.

Dockery, A., Whelan, L., Humphries, P. and Jane Farrar, G. 2021. Next-Generation Sequencing Applications for Inherited Retinal Diseases. *International Journal of Molecular Sciences*. **22**(11).

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F. and Kazazian,

H.H. 1991. Isolation of an active human transposable element. *Science (New York, N.Y.)*. **254**(5039), pp.1805–1808.

Doucette, L., Green, J., Black, C., Schwartzentruber, J., Johnson, G.J., Galutira, D. and Young, T.L. 2013. Molecular genetics of achromatopsia in Newfoundland reveal genetic heterogeneity, founder effects and the first cases of Jalili syndrome in North America. *Ophthalmic genetics*. **34**(3), pp.119–129.

Dryja, T.P., Finn, J.T., Peng, Y.W., Mcgee, T.L., Berson, E.L. and Yau, K.W. 1995. Mutations in the gene encoding the alpha subunit of the rod cGMP-gated channel in autosomal recessive retinitis pigmentosa. *Proceedings of the National Academy of Sciences of the United States of America*. **92**(22), pp.10177–10181.

Duijkers, L., Van Den Born, L.I., Neidhardt, J., Bax, N.M., Pierrache, L.H.M., Klevering, B.J., Collin, R.W.J. and 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**(3).

Duncker, T., Marsiglia, M., Lee, W., Zernant, J., Tsang, S.H., Allikmets, R., Greenstein, V.C. and Sparrow, J.R. 2014. Correlations among near-infrared and short-wavelength autofluorescence and spectral-domain optical coherence tomography in recessive Stargardt disease. *Investigative ophthalmology & visual science*. **55**(12), pp.8134–8143.

Duran, I., Taylor, S.P., Zhang, W., Martin, J., Forlenza, K.N., Spiro, R.P., Nickerson, D.A., Bamshad, M., Cohn, D.H. and Krakow, D. 2016. Destabilization of the IFT-B cilia core complex due to mutations in IFT81 causes a Spectrum of Short-Rib Polydactyly Syndrome. *Scientific reports*. **6**.

Dyka, F.M., Molday, L.L., Chiodo, V.A., Molday, R.S. and Hauswirth, W.W. 2019. Dual ABCA4-AAV Vector Treatment Reduces Pathogenic Retinal A2E Accumulation in a Mouse Model of Autosomal Recessive Stargardt Disease. *Human gene therapy*. **30**(11), pp.1361–1370.

Edelmann, L., Spiteri, E., Koren, K., Pulijaal, V., Bialer, M.G., Shanske, A., Goldberg, R. and Morrow, B.E. 2001. AT-rich palindromes mediate the constitutional t(11;22) translocation. *American journal of human genetics*. **68**(1), pp.1–13.

Edge, P. and Bansal, V. 2019. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nature Communications 2019 10:1*. **10**(1), pp.1–10.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D.,

Zhao, P., Zhong, F., Korlach, J. and Turner, S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)*. **323**(5910), pp.133–138.

Eizenga, J.M., Novak, A.M., Sibbesen, J.A., Heumos, S., Ghaffaari, A., Hickey, G., Chang, X., Seaman, J.D., Rounthwaite, R., Ebler, J., Rautiainen, M., Garg, S., Paten, B., Marschall, T., Sireacuten, J. and Garrison, E. 2020. Pangenome graphs. *Annual review of genomics and human genetics*. **21**, p.139.

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., Hall, G., Gale, T., Christopher Lloyd, I., Bishop, P.N., Ramsden, S.C. and Black, G.C.M. 2016. Molecular findings from 537 individuals with inherited retinal disease. *Journal of Medical Genetics*. **53**(11), p.761.

Ellingford, J.M., Barton, S., Bhaskar, S., Williams, S.G., Sergouniotis, P.I., O'Sullivan, J., Lamb, J.A., Perveen, R., Hall, G., Newman, W.G., Bishop, P.N., Roberts, S.A., Leach, R., Tearle, R., Bayliss, S., Ramsden, S.C., Nemeth, A.H. and Black, G.C.M. 2016. Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology*. **123**(5), pp.1143–1150.

Erdogan, F., Chen, W., Kirchhoff, M., Kalscheuer, V.M., Hultschig, C., Müller, I., Schulz, R., Menzel, C., Bryndorf, T., Ropers, H.H. and Ullmann, R. 2006. Impact of low copy repeats on the generation of balanced and unbalanced chromosomal aberrations in mental retardation. *Cytogenetic and genome research*. **115**(3–4), pp.247–253.

Essadi, I., Benbrahim, Z., Kaakoua, M., Reverdy, T., Corbaux, P. and Freyer, G. 2023. HER2-Positive Metastatic Breast Cancer: Available Treatments and Current Developments. *Cancers*. **15**(6), p.1738.

Esumi, N., Kachi, S., Hackler, L., Masuda, T., Yang, Z., Campochiaro, P.A. and Zack, D.J. 2009. BEST1 expression in the retinal pigment epithelium is modulated by OTX family members. *Human molecular genetics*. **18**(1), pp.128–141.

Euler, T., Haverkamp, S., Schubert, T. and Baden, T. 2014. Retinal bipolar cells: elementary building blocks of vision. *Nature Reviews Neuroscience 2014 15:8*. **15**(8), pp.507–519.

Fain, G. and Sampath, A.P. 2018. Rod and cone interactions in the retina. *F1000Research*. **7**.

Fain, G.L., Hardie, R. and Laughlin, S.B. 2010. Phototransduction and the Evolution of Photoreceptors. *Current Biology*. **20**(3), pp.R114–R124.

Falconer, E., Hills, M., Naumann, U., Poon, S.S.S., Chavez, E.A., Sanders, A.D., Zhao, Y., Hirst, M. and Lansdorp, P.M. 2012. DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. *Nature Methods 2012 9:11*. **9**(11), pp.1107–1112.

Farber, S.A. 2008. U.S. Scientists' Role in the Eugenics Movement (1907–1939): A Contemporary Biologist's Perspective. *Zebrafish*. **5**(4), p.243.

Farnoodian, M., Bose, D., Khristov, V., Susaimanickam, P.J., Maddileti, S.,

Mariappan, I., Abu-Asab, M., Campos, M., Villasmil, R., Wan, Q., Maminishkis, A., McGaughey, D., Barone, F., Gundry, R.L., Riordon, D.R., Boheler, K.R., Sharma, R. and Bharti, K. 2022. Cell-autonomous lipid-handling defects in Stargardt iPSC-derived retinal pigment epithelium cells. *Stem Cell Reports*. **17**(11), pp.2438–2450.

Farrar, G.J., Kenna, P., Jordan, S.A., Kumar-Singh, R., Humphries, M.M., Sharp, E.M., Sheils, D.M. and Humphries, P. 1991. A three-base-pair deletion in the peripherin–RDS gene in one form of retinitis pigmentosa. *Nature 1991 354:6353*. **354**(6353), pp.478–480.

Firth, H. V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Vooren, S. Van, Moreau, Y., Pettett, R.M. and Carter, N.P. 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics*. **84**(4), p.524.

Fischer, M.D., Michalakis, S., Wilhelm, B., Zobor, D., Muehlfriedel, R., Kohl, S., Weisschuh, N., Ochakovski, G.A., Klein, R., Schoen, C., Sothilingam, V., Garcia-Garrido, M., Kuehlewein, L., Kahle, N., Werner, A., Dauletbekov, D., Paquet-Durand, F., Tsang, S., Martus, P., Peters, T., Seeliger, M., Bartz-Schmidt, K.U., Ueffing, M., Zrenner, E., Biel, M. and Wissinger, B. 2020. Safety and Vision Outcomes of Subretinal Gene Therapy Targeting Cone Photoreceptors in Achromatopsia: A Nonrandomized Controlled Trial. *JAMA Ophthalmology*. **138**(6), pp.643–651.

Fischer, M.D., Ochakovski, G.A., Beier, B., Seitz, I.P., Vaheb, Y., Kortuem, C., Reichel, F.F.L., Kuehlewein, L., Kahle, N.A., Peters, T., Girach, A., Zrenner, E., Ueffing, M., MacLaren, R.E., Bartz-Schmidt, K.U. and Wilhelm, B. 2019. Efficacy and Safety of Retinal Gene Therapy Using Adeno-Associated Virus Vector for Patients With Choroideremia: A Randomized Clinical Trial. *JAMA Ophthalmology*. **137**(11), pp.1247–1254.

Fishman, G.A., Stone, E.M., Grover, S., Derlacki, D.J., Haines, H.L. and Hockey, R.R. 1999. Variation of clinical expression in patients with Stargardt dystrophy and sequence variations in the ABCR gene. *Archives of ophthalmology (Chicago, Ill. : 1960)*. **117**(4), pp.504–510.

Da Fonseca, R.R., Couto, A., Machado, A.M., Brejova, B., Albertin, C.B., Silva, F., Gardner, P., Baril, T., Hayward, A., Campos, A., Ribeiro, Â.M., Barrio-Hernandez, I., Hoving, H.J., Tafur-Jimenez, R., Chu, C., Frazão, B., Petersen, B., Peñaloza, F., Musacchia, F., Alexander, G.C., Osório, H., Winkelmann, I., Simakov, O., Rasmussen, S., Rahman, M.Z., Pisani, D., Vinther, J., Jarvis, E., Zhang, G., Strugnell, J.M., Castro, L.F.C., Fedrigo, O., Patricio, M., Li, Q., Rocha, S., Antunes, A., Wu, Y., Ma, B., Sanges, R., Vinar, T., Blagoev, B., Sicheritz-Ponten, T., Nielsen, R. and Gilbert, M.T.P. 2020. A draft genome sequence of the elusive giant squid, Architeuthis dux. *GigaScience*. **9**(1).

Fouladi, B., Sabatier, L., Miller, D., Pottier, G. and Murnane, J.P. 2000. The Relationship Between Spontaneous Telomere Loss and Chromosome Instability in a Human Tumor Cell Line. *Neoplasia*. **2**(6), pp.540–554.

Fourquet, J., Noirot, C., Klopp, C., Pinton, P., Combes, S., Hoede, C., Pascal,

G., Chaubet, A., Chiapello, H., Gaspin, C., Haenni, M., Klopp, C., Lupo, A., Mainguy, J., Noirot, C., Rochegue, T., Zytnicki, M., Ferry, T. and Hoede, C. 2018. Whole metagenome analysis with metagWGS. *Clinical Cancer Research*. **24**(15), pp.3539–3549.

French, C.E., Dolling, H., Mégy, K., Sanchis-Juan, A., Kumar, A., Delon, I., Wakeling, M., Mallin, L., Agrawal, S., Austin, T., Walston, F., Park, S.M., Parker, A., Piyasena, C., Bradbury, K., Ellard, S., Rowitch, D.H. and Raymond, F.L. 2022. Refinements and considerations for trio whole-genome sequence analysis when investigating Mendelian diseases presenting in early childhood. *HGG advances*. **3**(3).

French, J.D. and Edwards, S.L. 2020. The Role of Noncoding Variants in Heritable Disease. *Trends in Genetics*. **36**(11), pp.880–891.

Fu, G.K., Hu, J., Wang, P.H. and Fodor, S.P.A. 2011. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(22), pp.9026–9031.

Fu, Q., Wang, F., Wang, H., Xu, F., Zaneveld, J.E., Ren, H., Keser, V., Lopez, I., Tuan, H.F., Salvo, J.S., Wang, X., Zhao, L., Wang, K., Li, Y., Koenekoop, R.K., Chen, R. and Sui, R. 2013. Next-generation sequencing-based molecular diagnosis of a Chinese patient cohort with autosomal recessive retinitis pigmentosa. *Investigative ophthalmology & visual science*. **54**(6), pp.4158–4166.

Fujinami, K., Sergouniotis, P.I., Davidson, A.E., Wright, G., Chana, R.K., Tsunoda, K., Tsubota, K., Egan, C.A., Robson, A.G., Moore, A.T., Holder, G.E., Michaelides, M. and Webster, A.R. 2013. Clinical and molecular analysis of Stargardt disease with preserved foveal structure and function. *American journal of ophthalmology*. **156**(3).

Gao, F.J., Tian, G.H., Hu, F.Y., Wang, D.D., Li, J.K., Chang, Q., Chen, F., Xu, G.Z., Liu, W. and Wu, J.H. 2020. Next-generation sequencing-based clinical diagnosis of choroideremia and comprehensive mutational and clinical analyses. *BMC Ophthalmology*. **20**(1).

Gao, L., Xu, W., Xin, T. and Song, J. 2023. Application of third-generation sequencing to herbal genomics. *Frontiers in Plant Science*. **14**.

Garanto, A., Chung, D.C., Duijkers, L., Corral-Serrano, J.C., Messchaert, M., Xiao, R., Bennett, J., Vandenberghe, L.H. and Collin, R.W.J. 2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Human Molecular Genetics*. **25**(12), p.2552.

Gerber, S., Rozet, J.M., Bonneau, D., Souied, E., Camuzat, A., Dufier, J.L., Amalric, P., Weissenbach, J., Munnich, A. and Kaplan, J. 1995. A gene for late-onset fundus flavimaculatus with macular dystrophy maps to chromosome 1p13. *American Journal of Human Genetics*. **56**(2), p.396.

Ghiasvand, N.M., Rudolph, D.D., Mashayekhi, M., Brzezinski, J.A., Goldman, D. and Glaser, T. 2011. Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nature neuroscience*. **14**(5), pp.578–588.

Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D. and Pääbo, S. 2004. Loss of Olfactory Receptor Genes Coincides with the Acquisition of Full Trichromatic Vision in Primates. *PLOS Biology*. **2**(1), p.e5.

Gilpatrick, T., Lee, I., Graham, J.E., Raimondeau, E., Bowen, R., Heron, A., Downs, B., Sukumar, S., Sedlazeck, F.J. and Timp, W. 2020. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology 2020 38:4*. **38**(4), pp.433–438.

Girirajan, S., Campbell, C.D. and Eichler, E.E. 2011. Human copy number variation and complex genetic disease. *Annual review of genetics*. **45**, pp.203–226.

Goetz, K.E., Reeves, M.J., Gagadam, S., Blain, D., Bender, C., Lwin, C., Naik, A., Tumminia, S.J. and Hufnagel, R.B. 2020. Genetic testing for inherited eye conditions in over 6,000 individuals through the eyeGENE network. *American journal of medical genetics. Part C, Seminars in medical genetics*. **184**(3), pp.828–837.

Goldberg, A.F.X., Moritz, O.L. and Williams, D.S. 2016. Molecular basis for photoreceptor outer segment architecture. *Progress in Retinal and Eye Research*. **55**, pp.52–81.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., Murthy, K.K., Rovin, B.H., Bradley, W., Clark, R.A., Anderson, S.A., O'Connell, R.J., Agan, B.K., Ahuja, S.S., Bologna, R., Sen, L., Dolan, M.J. and Ahuja, S.K. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. **307**(5714), pp.1434–1440.

Gould, D.W., Lukic, S. and Chen, K.C. 2012. Selective Constraint on Copy Number Variation in Human Piwi-Interacting RNA Loci. *PLOS ONE*. **7**(10), p.e46611.

Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. *Science (New York, N.Y.)*. **185**(4154), pp.862–864.

Green, D.J., Lenassi, E., Manning, C.S., McGaughey, D., Sharma, V., Black, G.C., Ellingford, J.M. and Sergouniotis, P.I. 2021. North Carolina Macular Dystrophy: Phenotypic Variability and Computational Analysis of Disease-Associated Noncoding Variants. *Investigative Ophthalmology & Visual Science*. **62**(7), pp.16–16.

Green, D.J., Sallah, S.R., Ellingford, J.M., Lovell, S.C. and Sergouniotis, P.I. 2020. Variability in Gene Expression is Associated with Incomplete Penetrance in Inherited Eye Disorders. *Genes*. **11**(2).

Grewal, S.S., Smith, J.J. and Carr, A.-J.F. 2021. Bestrophinopathies: perspectives on clinical disease, Bestrophin-1 function and developing therapies. *Therapeutic Advances in Ophthalmology*. **13**, p.251584142199719.

Grieger, J.C. and Samulski, R.J. 2005. Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. *Journal of virology*. **79**(15), pp.9933–9944.

Guo, Y., Sheng, Q., Samuels, D.C., Lehmann, B., Bauer, J.A., Pietenpol, J.

and Shyr, Y. 2013. Comparative Study of Exome Copy Number Variation Estimation Tools Using Array Comparative Genomic Hybridization as Control. *BioMed Research International*. **2013**.

Gurudev, N., Yuan, M. and Knust, E. 2014. chaoptin, prominin, eyes shut and crumbs form a genetic network controlling the apical compartment of Drosophila photoreceptor cells. *Biology Open*. **3**(5), pp.332–341.

Guy, J., Feuer, W.J., Davis, J.L., Porciatti, V., Gonzalez, P.J., Koilkonda, R.D., Yuan, H., Hauswirth, W.W. and Lam, B.L. 2017. Gene Therapy for Leber Hereditary Optic Neuropathy: Low- and Medium-Dose Visual Results. *Ophthalmology*. **124**(11), pp.1621–1634.

Haer-Wigman, L., den Ouden, A., van Genderen, M.M., Kroes, H.Y., Verheij, J., Smailhodzic, D., Hoekstra, A.S., Vijzelaar, R., Blom, J., Derks, R., Tjon-Pon-Fong, M., Yntema, H.G., Nelen, M.R., Vissers, L.E.L.M., Lugtenberg, D. and Neveling, K. 2022. Diagnostic analysis of the highly complex OPN1LW/OPN1MW gene cluster using long-read sequencing and MLPA. *npj Genomic Medicine 2022 7:1*. **7**(1), pp.1–10.

Haer-Wigman, L., Van Zelst-Stams, W.A.G., Pfundt, R., Van Den Born, L.I., Klaver, C.C.W., Verheij, J.B.G.M., Hoyng, C.B., Breuning, M.H., Boon, C.J.F., Kievit, A.J., Verhoeven, V.J.M., Pott, J.W.R., Sallevelt, S.C.E.H., Van Hagen, J.M., Plomp, A.S., Kroes, H.Y., Lelieveld, S.H., Hehir-Kwa, J.Y., Castelein, S., Nelen, M., Scheffer, H., Lugtenberg, D., Cremers, F.P.M., Hoefsloot, L. and Yntema, H.G. 2017. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *European Journal of Human Genetics*. **25**(5).

Hafford-Tear, N.J., Tsai, Y.-C., Sadan, A.N., Sanchez-Pintado, B., Zarouchlioti, C., Maher, G.J., Liskova, P., Tuft, S.J., Hardcastle, A.J., Clark, T.A. and Davidson, A.E. 2019. CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy–associated TCF4 triplet repeat. *Genetics in Medicine 2019 21:9*. **21**(9), pp.2092–2102.

Haig H. Kazazian, J. and Moran, J. V. 2017. Mobile DNA in Health and Disease. *The New England journal of medicine*. **377**(4), p.361.

Haim, M. 2002. Epidemiology of retinitis pigmentosa in Denmark. *Acta Ophthalmologica Scandinavica. Supplement*. **80**(233), pp.1–34.

Hammond, S.M. and Wood, M.J.A. 2011. Genetic therapies for RNA mis-splicing diseases. *Trends in genetics : TIG*. **27**(5), pp.196–205.

Hanany, M., Rivolta, C. and Sharon, D. 2020. Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases. *Proceedings of the National Academy of Sciences of the United States of America*. **117**(5), pp.2710–2716.

Hanawalt, P.C. 1972. Repair of genetic material in living cells. *Endeavour*. **31**(113), pp.83–87.

Handsaker, R.E., Korn, J.M., Nemesh, J. and McCarroll, S.A. 2011. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nature genetics*. **43**(3), pp.269–276.

Hanlon, K.S., Kleinstiver, B.P., Garcia, S.P., Zaborowski, M.P., Volak, A.,

Spirig, S.E., Muller, A., Sousa, A.A., Tsai, S.Q., Bengtsson, N.E., Lööv, C., Ingelsson, M., Chamberlain, J.S., Corey, D.P., Aryee, M.J., Joung, J.K., Breakefield, X.O., Maguire, C.A. and György, B. 2019. High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nature communications*. **10**(1).

Haraksingh, R.R., Abyzov, A. and Urban, A.E. 2017. Comprehensive performance comparison of high-resolution array platforms for genome-wide Copy Number Variation (CNV) analysis in humans. *BMC Genomics*. **18**(1).

Hardenbol, P., Banér, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U. and Davis, R.W. 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. . **21**(6), pp.673–678.

Hartzell, C., Qu, Z., Putzier, I., Artinian, L., Chien, L.T. and Cui, Y. 2005. Looking chloride channels straight in the eye: bestrophins, lipofuscinosis, and retinal degeneration. *Physiology (Bethesda, Md.)*. **20**(5), pp.292–302.

Harvey, A., Brand, A., Holgate, S.T., Kristiansen, L. V., Lehrach, H., Palotie, A. and Prainsack, B. 2012. The future of technologies for personalised medicine. *New Biotechnology*. **29**(6), pp.625–633.

Hayeems, R.Z., Geller, G., Finkelstein, D. and Faden, R.R. 2005. How patients experience progressive loss of visual function: a model of adjustment using qualitative methods. *The British journal of ophthalmology*. **89**(5), pp.615–620.

Hayward, C., Shu, X., Cideciyan, A. V., Lennon, A., Barran, P., Zareparsi, S., Sawyer, L., Hendry, G., Dhillon, B., Milam, A.H., Luthert, P.J., Swaroop, A., Hastie, N.D., Jacobson, S.G. and Wright, A.F. 2003a. Mutation in a short-chain collagen gene, CTRP5, results in extracellular deposit formation in late-onset retinal degeneration: a genetic model for age-related macular degeneration. *Human molecular genetics*. **12**(20), pp.2657–2667.

Hayward, C., Shu, X., Cideciyan, A. V., Lennon, A., Barran, P., Zareparsi, S., Sawyer, L., Hendry, G., Dhillon, B., Milam, A.H., Luthert, P.J., Swaroop, A., Hastie, N.D., Jacobson, S.G. and Wright, A.F. 2003b. Mutation in a short-chain collagen gene, CTRP5, results in extracellular deposit formation in late-onset retinal degeneration: a genetic model for age-related macular degeneration. *Human molecular genetics*. **12**(20), pp.2657–2667.

Heath Jeffery, R.C., Mukhtar, S.A., McAllister, I.L., Morgan, W.H., Mackey, D.A. and Chen, F.K. 2021. Inherited retinal diseases are the most common cause of blindness in the working-age population in Australia. *Ophthalmic Genetics*. **42**(4), p.431.

Heid, C.A., Stevens, J., Livak, K.J. and Williams, P.M. 1996. Real time quantitative PCR. *Genome Research*. **6**(10), pp.986–994.

Helal, A.A., Saad, B.T., Saad, M.T., Mosaad, G.S. and Aboshanab, K.M. 2022. Evaluation of the Available Variant Calling Tools for Oxford Nanopore Sequencing in Breast Cancer. *Genes*. **13**(9).

Henderson, R.H., Mackay, D.S., Li, Z., Moradi, P., Sergouniotis, P., Russell-Eggitt, I., Thompson, D.A., Robson, A.G., Holder, G.E., Webster, A.R. and Moore, A.T. 2011. Phenotypic variability in patients with retinal dystrophies due to mutations in CRB1. *The British journal of ophthalmology*. **95**(6), pp.811–817.

Hendrickson, A.E. 1994. Primate foveal development: a microcosm of current questions in neurobiology. *Investigative Ophthalmology & Visual Science*. **35**(8), pp.3129–3133.

Hiatt, J.B., Pritchard, C.C., Salipante, S.J., O'Roak, B.J. and Shendure, J. 2013. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Research*. **23**(5).

Higgins, C.F. 1992. ABC transporters: from microorganisms to man. *Annual review of cell biology*. **8**, pp.67–113.

Higuchi, R., Fockler, C., Dollinger, G. and Watson, R. 1993. Kinetic PCR Analysis: Real-time Monitoring of DNA Amplification Reactions. *Bio/Technology 1993 11:9*. **11**(9), pp.1026–1030.

Hitti-Malin, R.J., Dhaenens, C., Panneman, D.M., Corradi, Z., Khan, M., den Hollander, A.I., Farrar, G.J., Gilissen, C., Hoischen, A., van de Vorst, M., Bults, F., Boonen, E.G.M., Saunders, P., Roosing, S. and Cremers, F.P.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. *Human mutation*.

HL Heidi L., S., Felix, G., Ulrich, K., Georg, S., Klaus, R., Herbert, J., Karsten, H., Philipp, R., Cord, H., Maria J., B., Bernhard H.F., W. and Heidi, S. 2017. Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort—Impact of Selected Deep Intronic Variants and Common SNPs. . **58**(1), p.394.

Höijer, I., Tsai, Y., Clark, T.A., Kotturi, P., Dahl, N., Stattin, E., Bondeson, M., Feuk, L., Gyllensten, U. and Ameur, A. 2018. Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing. *Human Mutation*. **39**(9), p.1262.

Den Hollander, A.I., Davis, J., Van Der Velde-Visser, S.D., Zonneveld, M.N., Pierrottet, C.O., Koenekoop, R.K., Kellner, U., Van Den Born, L.I., Heckenlively, J.R., Hoyng, C.B., Handford, P.A., Roepman, R. and Cremers, F.P.M. 2004. CRB1 mutation spectrum in inherited retinal dystrophies. *Human mutation*. **24**(5), pp.355–369.

Hon, T., Mars, K., Young, G., Tsai, Y.C., Karalius, J.W., Landolin, J.M., Maurer, N., Kudrna, D., Hardigan, M.A., Steiner, C.C., Knapp, S.J., Ware, D., Shapiro, B., Peluso, P. and Rank, D.R. 2020. Highly accurate long-read HiFi sequencing data for five complex genomes. *Scientific Data 2020 7:1*. **7**(1), pp.1–11.

Hoon, M., Okawa, H., Della Santina, L. and Wong, R.O.L. 2014. Functional architecture of the retina: Development and disease. *Progress in Retinal and Eye Research*. **42**, pp.44–84.

Hu, J. and Bok, D. 2010. Culture of highly differentiated human retinal pigment epithelium for analysis of the polarized uptake, processing, and secretion

of retinoids. *Methods in molecular biology (Clifton, N.J.)*. **652**, pp.55–73.

Hu, J., Pauer, G.J., Hagstrom, S.A., Bok, D., DeBenedictis, M.J., Bonilha, V.L., Hollyfield, J.G. and Radu, R.A. 2020. Evidence of complement dysregulation in outer retina of Stargardt disease donor eyes. *Redox Biology*. **37**, p.101787.

Huang, L., Xiao, X., Li, S., Jia, X., Wang, P., Sun, W., Xu, Y., Xin, W., Guo, X. and Zhang, Q. 2016. Molecular genetics of cone-rod dystrophy in Chinese patients: New data from 61 probands and mutation overview of 163 probands. *Experimental Eye Research*. **146**, pp.252–258.

HUBBARD, R. and WALD, G. 1952. CIS-TRANS ISOMERS OF VITAMIN A AND RETINENE IN THE RHODOPSIN SYSTEM. *Journal of General Physiology*. **36**(2), pp.269–315.

van Huet, R.A.C., Bax, N.M., Westeneng-Van Haaften, S.C., Muhamad, M., Zonneveld-Vrieling, M.N., Hoefsloot, L.H., Cremers, F.P.M., Boon, C.J.F., Klevering, B.J. and Hoyng, C.B. 2014. Foveal sparing in Stargardt disease. *Investigative ophthalmology & visual science*. **55**(11), pp.7467–7478.

Humphries, M.M., Rancourt, D., Jane Farrar, G., Kennal, P., HazeP, M., Bush, R.A., Sieving, P.A., Sheils, D.M., McNally, N., Creighton, P., Erven, A., Boros, A., Gulya, K., Capecchi, M.R., Humphries, P., Ndel BamHI -, B. and -, E. 1997. Retinopathy induced in mice by targeted disruption of the rhodopsin gene. *Nature Genetics 1997 15:2*. **15**(2), pp.216–219.

Hung, S.S., Meissner, B., Chavez, E.A., Ben-Neriah, S., Ennishi, D., Jones, M.R., Shulha, H.P., Chan, F.C., Boyle, M., Kridel, R., Gascoyne, R.D., Mungall, A.J., Marra, M.A., Scott, D.W., Connors, J.M. and Steidl, C. 2018. Assessment of Capture and Amplicon-Based Approaches for the Development of a Targeted Next-Generation Sequencing Pipeline to Personalize Lymphoma Management. *The Journal of Molecular Diagnostics*. **20**(2), pp.203–214.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. and Lee, C. 2004. Detection of large-scale variation in the human genome. *Nature Genetics 2004 36:9*. **36**(9), pp.949–951.

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. and Ayuso, C. 2021. Prioritizing variants of uncertain significance for reclassification using a rule-based algorithm in inherited retinal dystrophies. *NPJ Genomic Medicine*. **6**(1).

Iben, J.R. and Maraia, R.J. 2014. tRNA gene copy number variation in humans. *Gene*. **536**(2), pp.376–384.

Imani, S., Cheng, J., Shasaltaneh, M.D., Wei, C., Yang, L., Fu, S., Zou, H., Khan, M.A., Zhang, X., Chen, H., Zhang, D., Duan, C., Lv, H., Li, Y., Chen, R. and Fu, J. 2017. Genetic identification and molecular modeling characterization reveal a novel PROM1 mutation in Stargardt4-like macular dystrophy. *Oncotarget*. **9**(1), pp.122–141.

Issa, P.C., Barnard, A.R., Herrmann, P., Washington, I. and MacLaren, R.E.

2015. Rescue of the Stargardt phenotype in Abca4 knockout mice through inhibition of vitamin A dimerization. *Proceedings of the National Academy of Sciences of the United States of America*. **112**(27), pp.8415–8420.

Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., Mefford, H., Ying, P., Nickerson, D.A. and Eichler, E.E. 2009. Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease. . **84**(2), pp.148–161.

Jabara, C.B., Jones, C.D., Roach, J., Anderson, J.A. and Swanstrom, R. 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(50), pp.20166–20171.

Jabri, Y., Biber, J., Diaz-Lezama, N., Grosche, A. and Pauly, D. 2020. Cell-Type-Specific Complement Profiling in the ABCA4-/- Mouse Model of Stargardt Disease. *International journal of molecular sciences*. **21**(22), pp.1–16.

JACOBSON, T.S., TISCHLER, B. and MILLER, J.R. 1964. ENLARGED CHROMOSOMAL SATELLITES ASSOCIATED WITH MENTAL RETARDATION AND DIGITAL ARCHES IN THREE GENERATIONS. *Annals of human genetics*. **28**(1–3), pp.21–26.

Jahromi, M.S., Putnam, A.R., Druzgal, C., Wright, J., Spraker-Perlman, H., Kinsey, M., Zhou, H., Boucher, K.M., Randall, R.L., Jones, K.B., Lucas, D., Rosenberg, A., Thomas, D., Lessnick, S.L. and Schiffman, J.D. 2012. Molecular inversion probe analysis detects novel copy number alterations in Ewing sarcoma. *Cancer genetics*. **205**(7–8), pp.391–404.

Jain, M., Olsen, H.E., Paten, B. and Akeson, M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology 2016 17:1*. **17**(1), pp.1–11.

Jeon, Y.H., Heo, Y.S., Kim, C.M., Hyun, Y.L., Lee, T.G., Ro, S. and Cho, J.M. 2005. Phosphodiesterase: Overview of protein structures, potential therapeutic applications and recent progress in drug development. *Cellular and Molecular Life Sciences*. **62**(11), pp.1198–1220.

Ji, J., Leung, M.L., Baker, S., Deignan, J.L. and Santani, A. 2021. Clinical Exome Reanalysis: Current Practice and Beyond. *Molecular Diagnosis and Therapy*. **25**(5), pp.529–536.

Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y. and Zhu, T.F. 2015. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nature Communications 2015 6:1*. **6**(1), pp.1–8.

Jiang, Y.F., Wang, S., Wang, C.L., Xu, R.H., Wang, W.W., Jiang, Y., Wang, M.S., Jiang, L., Dai, L.H., Wang, J.R., Chu, X.H., Zeng, Y.Q., Fang, L.Z., Wu, D.D., Zhang, Q. and Ding, X.D. 2023. Pangenome obtained by long-read sequencing of 11 genomes reveal hidden functional structural variants in pigs. *iScience*. **26**(3).

Johansson, L.F., van Dijk, F., de Boer, E.N., van Dijk-Bos, K.K., Jongbloed,

J.D.H., van der Hout, A.H., Westers, H., Sinke, R.J., Swertz, M.A., Sijmons, R.H. and Sikkema-Raddatz, B. 2016. CoNVaDING: Single Exon Variation Detection in Targeted NGS Data. *Human Mutation*. **37**(5), pp.457–464.

John, M.C., Quinn, J., Hu, M.L., Cehajic-Kapetanovic, J. and Xue, K. 2023. Gene-agnostic therapeutic approaches for inherited retinal degenerations. *Frontiers in molecular neuroscience*. **15**.

Kajiwara, K., Hahn, L.B., Mukai, S., Travis, G.H., Berson, E.L. and Dryja, T.P. 1991. Mutations in the human retinal degeneration slow gene in autosomal dominant retinitis pigmentosa. *Nature 1991 354:6353*. **354**(6353), pp.480–483.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F. and Pinkel, D. 1992. Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors. *Science*. **258**(5083), pp.818–821.

Kaltak, M., Bruijn, P. de, Piccolo, D., Lee, S.-E., Dulla, K., Hoogenboezem, T., Beumer, W., Webster, A.R., Collin, R.W.J., Cheetham, M.E., Platenburg, G. and Swildens, J. 2022. Antisense oligonucleotide therapy for the common Stargardt disease type 1-causing variant in ABCA4. *bioRxiv*., 2022.08.12.503728.

Kaltak, M., de Bruijn, P., Piccolo, D., Lee, S.E., Dulla, K., Hoogenboezem, T., Beumer, W., Webster, A.R., Collin, R.W.J., Cheetham, M.E., Platenburg, G. and Swildens, J. 2023. Antisense oligonucleotide therapy corrects splicing in the common Stargardt disease type 1-causing variant ABCA4 c.5461-10T>C. *Molecular Therapy - Nucleic Acids*. **31**, pp.674–688.

Kaplan, J., Gerber, S., Larget-Piet, D., Rozet, J.M., Dollfus, H., Dufier, J.L., Odent, S., Postel-Vinay, A., Janin, N., Briard, M.L., Frézal, J. and Munnich, A. 1993. A gene for Stargardt's disease (fundus flavimaculatus) maps to the short arm of chromosome 1. *Nature genetics*. **5**(3), pp.308–311.

Kaylor, J.J., Xu, T., Ingram, N.T., Tsan, A., Hakobyan, H., Fain, G.L. and Travis, G.H. 2017. Blue light regenerates functional visual pigments in mammals through a retinyl-phospholipid intermediate. *Nature Communications 2017 8:1*. **8**(1), pp.1–10.

Kazazian, H.H. and Moran, J. V. 1998. The impact of L1 retrotransposons on the human genome. *Nature genetics*. **19**(1), pp.19–24.

Kazazian, H.H., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G. and Antonarakis, S.E. 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*. **332**(6160), pp.164–166.

Khan, K.N., Islam, F., Moore, A.T. and Michaelides, M. 2018. THE FUNDUS PHENOTYPE ASSOCIATED WITH THE p.Ala243Val BEST1 MUTATION. *Retina (Philadelphia, Pa.)*. **38**(3), pp.606–613.

Khan, M., Cornelis, S.S., Khan, M.I., Elmelik, D., Manders, E., Bakker, S., Derks, R., Neveling, K., Vorst, M. van de, Gilissen, C., Meunier, I., Defoort, S., Puech, B., Devos, A., Schulz, H.L., Stöhr, H., Grassmann, F.,

Weber, B.H.F., Dhaenens, C. and Cremers, F.P.M. 2019. Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. *Human Mutation*. **40**(10), pp.1749–1759.

Khan, M., Cornelis, S.S., Pozo-Valero, M. Del, Whelan, L., Runhart, E.H., Mishra, K., Bults, F., AlSwaiti, Y., AlTalbishi, A., Baere, E. De, Banfi, S., Banin, E., Bauwens, M., Ben-Yosef, T., Boon, C.J.F., Born, L.I. van den, Defoort, S., Devos, A., Dockery, A., Dudakova, L., Fakin, A., Farrar, G.J., Sallum, J.M.F., Fujinami, K., Gilissen, C., Glavač, D., Gorin, M.B., Greenberg, J., Hayashi, T., Hettinga, Y.M., Hoischen, A., Hoyng, C.B., Hufendiek, K., Jägle, H., Kamakari, S., Karali, M., Kellner, U., Klaver, C.C.W., Kousal, B., Lamey, T.M., MacDonald, I.M., Matynia, A., McLaren, T.L., Mena, M.D., Meunier, I., Miller, R., Newman, H., Ntozini, B., Oldak, M., Pieterse, M., Podhajcer, O.L., Puech, B., Ramesar, R., Rüther, K., Salameh, M., Salles, M.V., Sharon, D., Simonelli, F., Spital, G., Steehouwer, M., Szaflik, J.P., Thompson, J.A., Thuillier, C., Tracewska, A.M., Zweeden, M. van, Vincent, A.L., Zanlonghi, X., Liskova, P., Stöhr, H., Roach, J.N. De, Ayuso, C., Roberts, L., Weber, B.H.F., Dhaenens, C. and Cremers, F.P.M. 2020. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genetics in Medicine 2020*., pp.1–12.

Khateb, S., Shemesh, A., Offenheim, A., Sheffer, R., Ben-Yosef, T., Chowers, I., Leibu, R., Baumann, B., Wissinger, B., Kohl, S., Banin, E. and Sharon, D. 2022. Relatively mild blue cone monochromacy phenotype caused by various haplotypes in the L- and M-cone opsin genes. *Molecular Vision*. **28**, p.21.

Kidd, J.M., Sampas, N., Antonacci, F., Graves, T., Fulton, R., Hayden, H.S., Alkan, C., Malig, M., Ventura, M., Giannuzzi, G., Kallicki, J., Anderson, P., Tsalenko, A., Yamada, N.A., Tsang, P., Kaul, R., Wilson, R.K., Bruhn, L. and Eichler, E.E. 2010. Characterization of missing human genome sequences and copy-number polymorphic insertions. *Nature Methods 2010 7:5*. **7**(5), pp.365–371.

Kim, J.M., Lee, C., Lee, G.I., Kim, N.K.D., Ki, C.S., Park, W.Y., Kim, B.J. and Kim, S.J. 2017. Identification of the PROM1 Mutation p.R373C in a Korean Patient With Autosomal Dominant Stargardt-like Macular Dystrophy. *Annals of laboratory medicine*. **37**(6), pp.536–539.

Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K.W. and Vogelstein, B. 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(23), pp.9530–9535.

Kircher, M., Witten, D.M., Jain, P., O'roak, B.J., Cooper, G.M. and Shendure, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics 2014 46:3*. **46**(3), pp.310–315.

Kiser, P.D., Golczak, M., Maeda, A. and Palczewski, K. 2012. Key enzymes of the retinoid (visual) cycle in vertebrate retina. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. **1821**(1), pp.137–151.

Kiser, P.D. and Palczewski, K. 2016. Retinoids and Retinal Diseases. *https://doi.org/10.1146/annurev-vision-111815-114407*. **2**, pp.197–234.

Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S. and Taipale, J. 2011. Counting absolute numbers of molecules using unique molecular identifiers. *Nature methods*. **9**(1), pp.72–74.

Kniazeva, M., Chiang, M.F., Morgan, B., Anduze, A.L., Zack, D.J., Han, M. and Zhang, K. 1999. A New Locus for Autosomal Dominant Stargardt-Like Disease Maps to Chromosome 4. . **64**(5).

Kobayashi, Y., Watanabe, S., Ong, A.L.C., Shirai, M., Yamashiro, C., Ogata, T., Higashijima, F., Yoshimoto, T., Hayano, T., Asai, Y., Sasai, N. and Kimura, K. 2021. Early manifestations and differential gene expression associated with photoreceptor degeneration in Prom1-deficient retina. *Disease Models & Mechanisms*. **14**(11).

Kocaoglu, O.P., Liu, Z., Zhang, F., Kurokawa, K., Jonnal, R.S. and Miller, D.T. 2016. Photoreceptor disc shedding in the living human eye. *Biomedical Optics Express*. **7**(11), p.4554.

Kohmoto, T., Naruto, T., Watanabe, M., Fujita, Y., Ujiro, S., Okamoto, N., Horikawa, H., Masuda, K. and Imoto, I. 2017. A 590 kb deletion caused by non-allelic homologous recombination between two LINE-1 elements in a patient with mesomelia-synostosis syndrome. *American journal of medical genetics. Part A*. **173**(4), pp.1082–1086.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature 2015 533:7603*. **533**(7603), pp.420–424.

Kondo, H., Qin, M., Mizota, A., Kondo, M., Hayashi, H., Hayashi, K., Oshima, K., Tahira, T. and Hayashi, K. 2004. A homozygosity-based search for mutations in patients with autosomal recessive retinitis pigmentosa, using microsatellite markers. *Investigative ophthalmology & visual science*. **45**(12), pp.4433–4439.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A. and Pollock, D.D. 2011. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genetics*. **7**(12), p.1002384.

Konkel, M.K. and Batzer, M.A. 2010. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Seminars in cancer biology*. **20**(4), pp.211–221.

Koren, S. and Phillippy, A.M. 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current opinion in microbiology*. **23**, pp.110–120.

Küçükoğlu, B., Rueckauer, B., Ahmad, N., Steveninck, J. de R. van, Güçlü, U. and Gerven, M. van 2022. Optimization of Neuroprosthetic Vision via End-to-end Deep Reinforcement Learning. *bioRxiv.*, 2022.02.25.482017.

Kuleshov, V., Dan, X., Rui, C., Dmitry, P., Zhihai, M., Tim, B., Michael, K. and Michael, S. 2014. Whole-genome haplotyping using long reads and statistical methods. *Nature biotechnology*. **32**(3), p.261.

Kumaran, M., Krishnan, P., Cass, C.E., Hubaux, R., Lam, W., Yasui, Y. and Damaraju, S. 2018. Breast cancer associated germline structural variants harboring small noncoding RNAs impact post-transcriptional gene regulation. *Scientific Reports 2018 8:1*. **8**(1), pp.1–11.

Kurahashi, H. and Emanuel, B.S. 2001. Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Human molecular genetics*. **10**(23), pp.2605–2617.

Lahr, D.J.G. and Katz, L.A. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *BioTechniques*. **47**(4), pp.857–866.

Lamb, T.D. 2015. Why rods and cones? *Eye 2016 30:2*. **30**(2), pp.179–185.

Lamb, T.D., Corless, R.M. and Pananos, A.D. 2015. The kinetics of regeneration of rhodopsin under enzyme-limited availability of 11-cis retinoid. *Vision Research*. **110**, pp.23–33.

Lamb, T.D. and Curtin, J. 2022. Photoreceptor physiology and evolution: cellular and molecular basis of rod and cone phototransduction. *The Journal of Physiology*. **600**(21), pp.4585–4601.

Lamb, T.D. and Pugh, E.N. 2004. Dark adaptation and the retinoid cycle of vision. *Progress in Retinal and Eye Research*. **23**(3), pp.307–380.

Lambertus, S., Lindner, M., Bax, N.M., Mauschitz, M.M., Nadal, J., Schmid, M., Schmitz-Valckenberg, S., den Hollander, A.I., Weber, B.H.F.F., Holz, F.G., van der Wilt, G.J., Fleckenstein, M., Hoyng, C.B., Hollander, A.I. den, Weber, B.H.F.F., Holz, F.G., Wilt, G.J. van der, Fleckenstein, M. and Hoyng, C.B. 2016. Progression of Late-Onset Stargardt Disease. . **57**(13), pp.5186–5191.

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., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock,

K., Hong, M.L., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., De La Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A. and Morgan, M.J. 2001. Initial sequencing and analysis of the human genome. *Nature 2001 409:6822*. **409**(6822), pp.860–921.

Lang, W. 1885. Central choroiditis with disseminated patches in remainder of fundus. *Trans. Ophthalmol. Soc. U. K.* **5**, pp.140–141.

Langer-Safer, P.R., Levine, M. and Ward, D.C. 1982. Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences*. **79**(14), pp.4381–4385.

Lassmann, T., Francis, R.W., Weeks, A., Tang, D., Jamieson, S.E., Broley, S., Dawkins, H.J.S., Dreyer, L., Goldblatt, J., Groza, T., Kamien, B., Kiraly-Borri, C., McKenzie, F., Murphy, L., Pachter, N., Pathak, G., Poulton, C., Samanek, A., Skoss, R., Slee, J., Townshend, S., Ward, M., Baynam, G.S. and Blackwell, J.M. 2020. A flexible computational pipeline for research analyses of unsolved clinical exome cases. *npj Genomic Medicine 2020 5:1*. **5**(1), pp.1–11.

Laver, T.W., Caswell, R.C., Moore, K.A., Poschmann, J., Johnson, M.B., Owens, M.M., Ellard, S., Paszkiewicz, K.H. and Weedon, M.N. 2016. Pitfalls of haplotype phasing from amplicon-based long-read sequencing. *Scientific Reports 2016 6:1*. **6**(1), pp.1–6.

Lavrichenko, K., Johansson, S. and Jonassen, I. 2021. Comprehensive characterization of copy number variation (CNV) called from array, long- and short-read data. *BMC Genomics*. **22**(1), pp.1–15.

Layer, R.M., Chiang, C., Quinlan, A.R. and Hall, I.M. 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology*. **15**(6), p.R84.

Lee, J.A., Carvalho, C.M.B. and Lupski, J.R. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. **131**(7), pp.1235–1247.

Lee, J.H., Wang, J.H., Chen, J., Li, F., Edwards, T.L., Hewitt, A.W. and Liu,

G.S. 2019. Gene therapy for visual loss: Opportunities and concerns. *Progress in Retinal and Eye Research*. **68**, pp.31–53.

Lee, W., Paavo, M., Zernant, J., Stong, N., Laurente, Z., Bearelly, S., Nagasaki, T., Tsang, S.H., Goldstein, D.B. and Allikmets, R. 2019. Modification of the PROM1 disease phenotype by a mutation in ABCA4. *Ophthalmic genetics*. **40**(4), pp.369–375.

Lee, W., Zernant, J., Nagasaki, T., Tsang, S.H. and Allikmets, R. 2018. Deep Scleral Exposure: A Degenerative Outcome of End-Stage Stargardt Disease. *American journal of ophthalmology*. **195**, pp.16–25.

Lehrman, M.A., Schneider, W.J., Südhof, T.C., Brown, M.S., Goldstein, J.L. and Russell, D.W. 1985. Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science (New York, N.Y.)*. **227**(4683), pp.140–146.

Leitch, C.C., Zaghloul, N.A., Davis, E.E., Stoetzel, C., Diaz-Font, A., Rix, S., Alfadhel, M., Lewis, R.A., Eyaid, W., Banin, E., Dollfus, H., Beales, P.L., Badano, J.L. and Katsanis, N. 2008. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. *Nature genetics*. **40**(4), pp.443–448.

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., Michael Redmond, T., Travis, G.H. and 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 of the United States of America*. **115**(47), pp.E11120–E11127.

Leskov, I.B., Klenchin, V.A., Handy, J.W., Whitlock, G.G., Govardovskii, V.I., Bownds, M.D., Lamb, T.D., Pugh, E.N. and Arshavsky, V.Y. 2000. The gain of rod phototransduction: Reconcillation of biochemical and electrophysiological measurements. *Neuron*. **27**(3), pp.525–537.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W.C., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W., Strausberg, R.L. and Venter, J.C. 2007. The Diploid Genome Sequence of an Individual Human. *PLOS Biology*. **5**(10), p.e254.

Li, H. 2018. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics (Oxford, England)*. **34**(18).

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*. **25**(14), pp.1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**(16), p.2078.

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., Tomlinson, C., Villani, F., Vollger, M.R., Consortium, H.P.R., Bourque, G., Chaisson, M.J., Flicek, P., Phillippy, A.M., Zook, J.M., Eichler, E.E., Haussler, D., Jarvis, E.D., Miga, K.H., Wang, T., Garrison, E., Marschall, T., Hall, I., Li, H. and Paten, B. 2022. A Draft Human Pangenome Reference. *bioRxiv.*, 2022.07.09.499321.

Lieber, M.R. 2010. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annual review of biochemistry*. **79**, p.181.

Lieber, M.R., Ma, Y., Pannicke, U. and Schwarz, K. 2003. Mechanism and regulation of human non-homologous DNA end-joining. *Nature reviews. Molecular cell biology*. **4**(9), pp.712–720.

Liew, G., Michaelides, M. and Bunce, C. 2014. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ Open*. **4**(2), p.e004015.

Lindenbaum, P. 2015. JVarkit: java-based utilities for Bioinformatics.

Lionel, A.C., Costain, G., Monfared, N., Walker, S., Reuter, M.S., Hosseini, S.M., Thiruvahindrapuram, B., Merico, D., Jobling, R., Nalpathamkalam, T., Pellecchia, G., Sung, W.W.L., Wang, Z., Bikangaga, P., Boelman, C., Carter, M.T., Cordeiro, D., Cytrynbaum, C., Dell, S.D., Dhir, P., Dowling, J.J., Heon, E., Hewson, S., Hiraki, L., Inbar-Feigenberg, M., Klatt, R., Kronick, J., Laxer, R.M., Licht, C., MacDonald, H., Mercimek-Andrews, S., Mendoza-Londono, R., Piscione, T., Schneider, R., Schulze, A., Silverman, E., Siriwardena, K., Snead, O.C., Sondheimer, N., Sutherland, J., Vincent, A., Wasserman, J.D., Weksberg, R., Shuman, C., Carew, C., Szego, M.J., Hayeems, R.Z., Basran, R., Stavropoulos, D.J., Ray, P.N., Bowdin, S., Meyn, M.S., Cohn, R.D., Scherer, S.W. and Marshall, C.R. 2018. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genetics in Medicine*. **20**(4), p.435.

Liu, F., Lee, J. and Chen, J. 2021. Molecular structures of the eukaryotic retinal importer abca4. *eLife*. **10**, pp.1–30.

Liu, H., Gu, X., Wang, G., Huang, Y., Ju, S., Huang, J. and Wang, X. 2019. Copy number variations primed lncRNAs deregulation contribute to poor prognosis in colorectal cancer. *Aging (Albany NY)*. **11**(16), p.6089.

Lo, A.W.I., Sabatier, L., Fouladi, B., Pottiert, G., Ricoul, M. and Murnane, J.P. 2002. DNA Amplification by Breakage/Fusion/Bridge Cycles Initiated by Spontaneous Telomere Loss in a Human Cancer Cell Line. *Neoplasia*. **4**(6), pp.531–538.

Lobban, P.E. and Kaiser, A.D. 1973. Enzymatic end-to end joining of DNA molecules. *Journal of molecular biology*. **78**(3).

Lofaro, F.D., Mucciolo, D.P., Murro, V., Pavese, L., Quaglino, D. and Boraldi, F. 2021. A Case Report of Pseudoxanthoma Elasticum with Rare

Sequence Variants in Genes Related to Inherited Retinal Diseases. *Diagnostics*. **11**(10).

Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Reshef, Y.A., Finucane, H.K., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., Durbin, R. and Price, A.L. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature genetics*. **48**(11), pp.1443–1448.

Loomis, E.W., Eid, J.S., Peluso, P., Yin, J., Hickey, L., Rank, D., McCalmon, S., Hagerman, R.J., Tassone, F. and Hagerman, P.J. 2013. Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene. *Genome Research*. **23**(1).

Loose, M., Malla, S. and Stout, M. 2016. Real-time selective sequencing using nanopore technology. *Nature Methods 2016 13:9*. **13**(9), pp.751–754.

Lucarelli, M., Narzi, L., Pierandrei, S., Bruno, S.M., Stamato, A., D'avanzo, M., Strom, R. and Quattrucci, S. 2010. A new complex allele of the CFTR gene partially explains the variable phenotype of the L997F mutation. . **12**(9), pp.548–555.

Lujan, B.J., Roorda, A., Knighton, R.W. and Carroll, J. 2011. Revealing Henle's fiber layer using spectral domain optical coherence tomography. *Investigative Ophthalmology and Visual Science*. **52**(3), pp.1486–1492.

Luo, D.G., Xue, T. and Yau, K.W. 2008. How vision begins: An odyssey. *Proceedings of the National Academy of Sciences*. **105**(29), pp.9855–9862.

Luo, J., Meng, Z., Xu, X., Wang, L., Zhao, K., Zhu, X., Qiao, Q., Ge, Y., Mao, L. and Cui, L. 2022. Corrigendum: Systematic benchmarking of nanopore Q20+ kit in SARS-CoV-2 whole genome sequencing. *Frontiers in microbiology*. **13**.

Lupski, J.R. 1998. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in Genetics*. **14**(10), pp.417–422.

MacDonald, I.M. and Sieving, P.A. 2018. Investigation of the effect of dietary docosahexaenoic acid (DHA) supplementation on macular function in subjects with autosomal recessive Stargardt macular dystrophy. *Ophthalmic genetics*. **39**(4), pp.477–486.

Maeda, A., Maeda, T., Golczak, M. and Palczewski, K. 2008. Retinopathy in mice induced by disrupted all-trans-retinal clearance. *Journal of Biological Chemistry*. **283**(39), pp.26684–26693.

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., Dass, A., Dhanapal, V., Fennell, T.J., Friedland, A.E., Giannoukos, G., Gloskowski, S.W., Glucksmann, A., Gotta, G.M., Jayaram, H., Haskett, S.J., Hopkins, B., Horng, J.E., Joshi, S., Marco, E., Mepani, R., Reyon, D., Ta, T., Tabbaa, D.G., Samuelsson, S.J., Shen, S., Skor, M.N., Stetkiewicz, P., Wang, T., Yudkoff, C., Myer, V.E., Albright, C.F. and Jiang, H. 2019. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nature medicine*. **25**(2),

pp.229–233.

Magi, A., Bolognini, D., Bartalucci, N., Mingrino, A., Semeraro, R., Giovannini, L., Bonifacio, S., Parrini, D., Pelo, E., Mannelli, F., Guglielmelli, P. and Maria Vannucchi, A. 2019. Nano-GLADIATOR: real-time detection of copy number alterations from nanopore sequencing data. *Bioinformatics*. **35**(21), pp.4213–4221.

Mäkeläinen, S., Gòdia, M., Hellsand, M., Viluma, A., Hahn, D., Makdoumi, K., Zeiss, C.J., Mellersh, C., Ricketts, S.L., Narfström, K., Hallböök, F., Ekesten, B., Andersson, G. and Bergström, T.F. 2019. An ABCA4 loss-of-function mutation causes a canine form of Stargardt disease. *PLoS Genetics*. **15**(3).

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. and Turner, D.J. 2010. Target-enrichment strategies for next-generation sequencing. *Nature methods*. **7**(2), pp.111–118.

Mandelker, D., Schmidt, R.J., Ankala, A., McDonald Gibson, K., Bowser, M., Sharma, H., Duffy, E., Hegde, M., Santani, A., Lebo, M. and Funke, B. 2016. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. . **18**(12), pp.1282–1289.

Manes, G., Guillaumie, T., Vos, W.L., Devos, A., Audo, I., Zeitz, C., Marquette, V., Zanlonghi, X., Defoort-Dhellemmes, S., Puech, B., Said, S.M., Sahel, J.A., Odent, S., Dollfus, H., Kaplan, J., Dufier, J.L., Le Meur, G., Weber, M., Faivre, L., Cohen, F.B., Béroud, C., Picot, M.C., Verdier, C., Sénéchal, A., Baudoin, C., Bocquet, B., Findlay, J.B., Meunier, I., Dhaenens, C.M. and Hamel, C.P. 2015. High Prevalence of PRPH2 in Autosomal Dominant Retinitis Pigmentosa in France and Characterization of Biochemical and Clinical Features. *American Journal of Ophthalmology*. **159**(2), pp.302–314.

Marcinkowska, M., Szymanski, M., Krzyzosiak, W.J. and Kozlowski, P. 2011. Copy number variation of microRNA genes in the human genome. *BMC Genomics*. **12**(1), pp.1–9.

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., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X. V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. and Rothberg, J.M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature 2005 437:7057*. **437**(7057), pp.376–380.

Marks, P., Garcia, S., Barrio, A.M., Belhocine, K., Bernate, J., Bharadwaj, R., Bjornson, K., Catalanotti, C., Delaney, J., Fehr, A., Fiddes, I.T., Galvin, B., Heaton, H., Herschleb, J., Hindson, C., Holt, E., Jabara, C.B., Jett, S., Keivanfar, N., Kyriazopoulou-Panagiotopoulou, S., Lek, M., Lin, B., Lowe,

A., Mahamdallie, S., Maheshwari, S., Makarewicz, T., Marshall, J., Meschi, F., O'Keefe, C.J., Ordonez, H., Patel, P., Price, A., Royall, A., Ruark, E., Seal, S., Schnall-Levin, M., Shah, P., Stafford, D., Williams, S., Wu, I., Xu, A.W., Rahman, N., MacArthur, D. and Church, D.M. 2019. Resolving the full spectrum of human genome variation using Linked-Reads. *Genome research*. **29**(4), pp.635–645.

Marmorstein, A.D., Kinnick, T.R., Stanton, J.B., Johnson, A.A., Lynch, R.M. and Marmorstein, L.Y. 2015. Bestrophin-1 influences transepithelial electrical properties and Ca2+ signaling in human retinal pigment epithelium. *Molecular Vision*. **21**, p.347.

Marquet, M., Zöllkau, J., Pastuschek, J., Viehweger, A., Schleußner, E., Makarewicz, O., Pletz, M.W., Ehricht, R. and Brandt, C. 2022. Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. *Scientific Reports 2022 12:1*. **12**(1), pp.1–10.

Marraffini, L.A. and Sontheimer, E.J. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*. **322**(5909), pp.1843–1845.

Martignano, F., Munagala, U., Crucitta, S., Mingrino, A., Semeraro, R., Del Re, M., Petrini, I., Magi, A. and Conticello, S.G. 2021. Nanopore sequencing from liquid biopsy: analysis of copy number variations from cell-free DNA of lung cancer patients. *Molecular Cancer*. **20**(1), pp.1–6.

Martin, A.R., Williams, E., Foulger, R.E., Leigh, S., Daugherty, L.C., Niblock, O., Leong, I.U.S., Smith, K.R., Gerasimenko, O., Haraldsdottir, E., Thomas, E., Scott, R.H., Baple, E., Tucci, A., Brittain, H., de Burca, A., Ibañez, K., Kasperaviciute, D., Smedley, D., Caulfield, M., Rendon, A. and McDonagh, E.M. 2019. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature genetics*. **51**(11), pp.1560–1565.

Martin, M., Patterson, M., Garg, S., Fischer, S.O., Pisanti, N., Klau, G.W., Schöenhuth, A. and Marschall, T. 2016. WhatsHap: fast and accurate read-based phasing. *bioRxiv*., p.085050.

Mata, N.L., Weng, J. and Travis, G.H. 2000. Biosynthesis of a major lipofuscin fluorophore in mice and humans with ABCR-mediated retinal and macular degeneration. *Proceedings of the National Academy of Sciences of the United States of America*. **97**(13), pp.7154–7159.

Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E. and Sfeir, A. 2015. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature*. **518**(7538), pp.254–257.

Maugeri, A., Van Driel, M.A., Van De Pol, D.J.R., Klevering, B.J., Van Haren, F.J.J., Tijmes, N., Bergen, A.A.B., Rohrschneider, K., Blankenagel, A., Pinckers, A.J.L.G., Dahl, N., Brunner, H.G., Deutman, A.F., Hoyng, C.B. and Cremers, F.P.M. 1999. The 2588G-->C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with Stargardt disease. *American journal of human genetics*. **64**(4), pp.1024–1035.

Mauro-Herrera, M., Chiang, J., Radojevic, B. and Bennett, L.D. 2021.

Functional evaluation of splicing for variants of uncertain significance in patients with inherited retinal diseases. *Genes*. **12**(7), p.993.

Mc Clinton, B., Corradi, Z., McKibbin, M., Panneman, D.M., Roosing, S., Boonen, E.G.M., Ali, M., Watson, C.M., Steel, D.H., Cremers, F.P.M., Inglehearn, C.F., Hitti-Malin, R.J. and Toomes, C. 2023. Effective smMIPs-Based Sequencing of Maculopathy-Associated Genes in Stargardt Disease Cases and Allied Maculopathies from the UK. *Genes*. **14**(1), p.191.

McCLINTOCK, B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*. **36**(6), pp.344–355.

McVeigh, E., Jones, H., Black, G. and Hall, G. 2019. The psychosocial and service delivery impact of genomic testing for inherited retinal dystrophies. *Journal of Community Genetics*. **10**(3), p.425.

Meleshko, D., Yang, R., Marks, P., Williams, S. and Hajirasouliha, I. 2022. Efficient detection and assembly of non-reference DNA sequences with synthetic long reads. . **50**(18), p.e108.

Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J.M.L.M., van der Zaag, P.J., Franke, A., Nilsson, M., Lehrach, H. and Brookes, A.J. 2011. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics*. **10**(6), pp.374–386.

Meyer, K.J. and Anderson, M.G. 2017. Genetic modifiers as relevant biological variables of eye disorders. *Human Molecular Genetics*. **26**(R1), pp.R58–R67.

Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S. 1990. DNA recombination during PCR. *Nucleic acids research*. **18**(7), pp.1687–1691.

Michaelides, M., Gaillard, M.C., Escher, P., Tiab, L., Bedell, M., Borruat, F.X., Barthelmes, D., Carmona, R., Zhang, K., White, E., McClements, M., Robson, A.G., Holder, G.E., Bradshaw, K., Hunt, D.M., Webster, A.R., Moore, A.T., Schorderet, D.F. and Munier, F.L. 2010. The PROM1 mutation p.R373C causes an autosomal dominant bull's eye maculopathy associated with rod, rod-cone, and macular dystrophy. *Investigative ophthalmology & visual science*. **51**(9), pp.4771–4780.

Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., Schneider, V.A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., Markovic, C., Maduro, V., Dutra, A., Bouffard, G.G., Chang, A.M., Hansen, N.F., Wilfert, A.B., Thibaud-Nissen, F., Schmitt, A.D., Belton, J.M., Selvaraj, S., Dennis, M.Y., Soto, D.C., Sahasrabudhe, R., Kaya, G., Quick, J., Loman, N.J., Holmes, N., Loose, M., Surti, U., Risques, R. ana, Graves Lindsay, T.A., Fulton, R., Hall, I., Paten, B., Howe, K., Timp, W., Young, A., Mullikin, J.C., Pevzner, P.A., Gerton, J.L., Sullivan, B.A., Eichler, E.E. and Phillippy, A.M. 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature 2020 585:7823*. **585**(7823), pp.79–84.

Miller, D.E., Sulovari, A., Wang, T., Loucks, H., Hoekzema, K., Munson, K.M.,

Lewis, A.P., Almanza Fuerte, E.P., Paschal, C.R., Walsh, T., Thies, J., Bennett, J.T., Glass, I., Dipple, K.M., Patterson, K., Bonkowski, E.S., Nelson, Z., Squire, A., Sikes, M., Beckman, E., Bennett, R.L., Earl, D., Lee, W., Allikmets, R., Perlman, S.J., Chow, P., Hing, A. V, Wenger, T.L., Adam, M.P., Sun, A., Lam, C., Chang, I., Zou, X., Austin, S.L., Huggins, E., Safi, A., Iyengar, A.K., Reddy, T.E., Majoros, W.H., Allen, A.S., Crawford, G.E., Kishnani, P.S., King, M.-C., Cherry, T., Chong, J.X., Bamshad, M.J., Nickerson, D.A., Mefford, H.C., Doherty, D. and Eichler, E.E. 2021. Targeted long-read sequencing identifies missing disease-causing variation.

Minei, R., Hoshina, R. and Ogura, A. 2018. De novo assembly of middle-sized genome using MinION and Illumina sequencers 06 Biological Sciences 0604 Genetics. *BMC Genomics*. **19**(1), pp.1–12.

Mirza, M., Vainshtein, A., DiRonza, A., Chandrachud, U., Haslett, L.J., Palmieri, M., Storch, S., Groh, J., Dobzinski, N., Napolitano, G., Schmidtke, C. and Kerkovich, D.M. 2019. The CLN3 gene and protein: What we know. *Molecular genetics & genomic medicine*. **7**(12).

Mockel, A., Perdomo, Y., Stutzmann, F., Letsch, J., Marion, V. and Dollfus, H. 2011. Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies. *Progress in retinal and eye research*. **30**(4), pp.258–274.

Moiseyev, G., Chen, Y., Takahashi, Y., Wu, B.X. and Ma, J.X. 2005. RPE65 is the isomerohydrolase in the retinoid visual cycle. *Proceedings of the National Academy of Sciences of the United States of America*. **102**(35), pp.12413–12418.

Molday, L.L., Wahl, D., Sarunic, M. V. and Molday, R.S. 2018. Localization and functional characterization of the p.Asn965Ser (N965S) ABCA4 variant in mice reveal pathogenic mechanisms underlying Stargardt macular degeneration. *Human Molecular Genetics*. **27**(2), p.295.

Molday, R.S., Garces, F.A., Scortecci, J.F. and Molday, L.L. 2022. Structure and function of ABCA4 and its role in the visual cycle and Stargardt macular degeneration. *Progress in Retinal and Eye Research*. **89**, p.101036.

Molin, A., Benoist, G., Jeanne-Pasquier, C., Elkartoufi, N., Litzer, J., Decamp, M., Gruchy, N., Durand-Malbruny, M., Begorre, M., Attie-Bitach, T. and Leporrier, N. 2013. 12q21 Microdeletion in a fetus with Meckel syndrome involving CEP290/MKS4. *European journal of medical genetics*. **56**(10), pp.580–583.

Molina-Ramírez, L.P., Lenassi, E., Ellingford, J.M., Sergouniotis, P.I., Ramsden, S.C., Bruce, I.A. and Black, G.C.M. 2020. Establishing Genotype-phenotype Correlation in USH2A-related Disorders to Personalize Audiological Surveillance and Rehabilitation. *Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*. **41**(4), pp.431–437.

Moreno-Cabrera, J.M., del Valle, J., Castellanos, E., Feliubadaló, L., Pineda, M., Brunet, J., Serra, E., Capellà, G., Lázaro, C. and Gel, B. 2020. Evaluation of CNV detection tools for NGS panel data in genetic

diagnostics. *European Journal of Human Genetics 2020 28:12*. **28**(12), pp.1645–1655.

Motta, F.L., Martin, R.P., Filippelli-Silva, R., Salles, M.V. and Sallum, J.M.F. 2018. Relative frequency of inherited retinal dystrophies in Brazil. *Scientific Reports 2018 8:1*. **8**(1), pp.1–9.

Müller, L.J., Marfurt, C.F., Kruse, F. and Tervo, T.M.T. 2003. Corneal nerves: Structure, contents and function. *Experimental Eye Research*. **76**(5), pp.521–542.

Muniz, A., Betts, B.S., Trevino, A.R., Buddavarapu, K., Roman, R., Ma, J.X. and Tsin, A.T.C. 2009. Evidence for two retinoid cycles in the cone-dominated chicken eye. *Biochemistry*. **48**(29), pp.6854–6863.

Mustafi, D., Engel, A.H. and Palczewski, K. 2009. Structure of cone photoreceptors. *Progress in Retinal and Eye Research*. **28**(4), pp.289–302.

Nachmanson, D., Lian, S., Schmidt, E.K., Hipp, M.J., Baker, K.T., Zhang, Y., Tretiakova, M., Loubet-Senear, K., Kohrn, B.F., Salk, J.J., Kennedy, S.R. and Risques, R.A. 2018. Targeted genome fragmentation with CRISPR/Cas9 enables fast and efficient enrichment of small genomic regions and ultra-accurate sequencing with low DNA input (CRISPR-DS). *Genome Research*. **28**(10), p.1589.

Nakamichi, K., Van Gelder, R.N., Chao, J.R. and Mustafi, D. 2023. Targeted adaptive long-read sequencing for discovery of complex phased variants in inherited retinal disease patients. *Scientific Reports 2023 13:1*. **13**(1), pp.1–9.

Nakao, T., Tsujikawa, M., Sawa, M., Gomi, F. and Nishida, K. 2012. Foveal sparing in patients with Japanese Stargardt's disease and good visual acuity. *Japanese journal of ophthalmology*. **56**(6), pp.584–588.

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., Heaps, L.S., Wright, D., Dinger, M.E., Bennetts, B., Grigg, J.R. and Jamieson, R. V. 2022a. Whole Genome Sequencing, Focused Assays and Functional Studies Increasing Understanding in Cryptic Inherited Retinal Dystrophies. *International Journal of Molecular Sciences*. **23**(7), p.3905.

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., Heaps, L.S., Wright, D., Dinger, M.E., Bennetts, B., Grigg, J.R. and Jamieson, R. V. 2022b. Whole Genome Sequencing, Focused Assays and Functional Studies Increasing Understanding in Cryptic Inherited Retinal Dystrophies. *International Journal of Molecular Sciences*. **23**(7).

Nathans, J. and Hogness, D.S. 1984. Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America*. **81**(15), pp.4851–4855.

Nathans, J., Thomas, D. and Hogness, D.S. 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science (New York, N.Y.)*. **232**(4747), pp.196–202.

Neveling, K., Collin, R.W.J., Gilissen, C., Van Huet, R.A.C., Visser, L., Kwint, M.P., Gijsen, S.J., Zonneveld, M.N., Wieskamp, N., De Ligt, J., Siemiatkowska, A.M., Hoefsloot, L.H., Buckley, M.F., Kellner, U., Branham, K.E., den Hollander, A.I., Hoischen, A., Hoyng, C., Klevering, B.J., Van den Born, L.I., Veltman, J.A., Cremers, F.P.M. and Scheffer, H. 2012. Next-generation genetic testing for retinitis pigmentosa. *Human Mutation*. **33**(6), pp.963–972.

Newton, F. and Megaw, R. 2020. Mechanisms of Photoreceptor Death in Retinitis Pigmentosa. *Genes*. **11**(10), pp.1–29.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., Bamshad, M., Nickerson, D.A. and Shendure, J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. **461**(7261), pp.272–276.

Niks, E.H. and Aartsma-Rus, A. 2016. Exon skipping: a first in class strategy for Duchenne muscular dystrophy. *https://doi.org/10.1080/14712598.2017.1271872*. **17**(2), pp.225–236.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B.P. and Landegren, U. 1994. Padlock Probes: Circularizing Oligonucleotides for Localized DNA Detection. *Science*. **265**(5181), pp.2085–2088.

Nõupuu, K., Lee, W., Zernant, J., Greenstein, V.C., Tsang, S. and Allikmets, R. 2016. Recessive Stargardt disease phenocopying hydroxychloroquine retinopathy. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie*. **254**(5), pp.865–872.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G. V., Chen, N.C., Cheng, H., Chin, C.S., Chow, W., de Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V. V., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sovi, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'Neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H. and Phillippy, A.M. 2022. The complete sequence of a human genome. . **376**(6588), pp.44–53.

O'Connell, J., Sharp, K., Shrine, N., Wain, L., Hall, I., Tobin, M., Zagury, J.F., Delaneau, O. and Marchini, J. 2016. Haplotype estimation for biobank-

scale data sets. *Nature genetics*. **48**(7), pp.817–820.

O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., Munson, J., Hiatt, J.B., Turner, E.H., Levy, R., O'Day, D.R., Krumm, N., Coe, B.P., Martin, B.K., Borenstein, E., Nickerson, D.A., Mefford, H.C., Doherty, D., Akey, J.M., Bernier, R., Eichler, E.E. and Shendure, J. 2012. Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders. *Science (New York, N.Y.)*. **338**(6114), p.1619.

Odelberg, S.J., Weiss, R.B., Hata, A. and White, R. 1995. Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I. *Nucleic acids research*. **23**(11), pp.2049–2057.

Oh, D.J., Daily, M.J. and Grassi, M.A. 2021. CRB1 maculopathy presenting as fenestrated sheen macular dystrophy with 15-year follow-up. *Documenta ophthalmologica. Advances in ophthalmology*. **142**(3), pp.381–388.

Olivier, G., Corton, M., Intartaglia, D., Verbakel, S.K., Sergouniotis, P.I., Le Meur, G., Dhaenens, C.M., Naacke, H., Avila-Fernández, A., Hoyng, C.B., Klevering, J., Bocquet, B., Roubertie, A., Sénéchal, A., Banfi, S., Muller, A., Hamel, C.L., Black, G.C., Conte, I., Roosing, S., Zanlonghi, X., Ayuso, C., Meunier, I. and Manes, G. 2021. Pathogenic variants in IMPG1 cause autosomal dominant and autosomal recessive retinitis pigmentosa. *Journal of Medical Genetics*. **58**(8), pp.570–578.

Ortega, J.T. and Jastrzebska, B. 2022. Rhodopsin as a Molecular Target to Mitigate Retinitis Pigmentosa. *Advances in experimental medicine and biology*. **1371**, pp.61–77.

Osoegawa, K., Mammoser, A.G., Wu, C., Frengen, E., Zeng, C., Catanese, J.J. and De Jong, P.J. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome research*. **11**(3), pp.483–496.

Packer, O., Hendrickson, A.E. and Curcio, C.A. 1989. Photoreceptor topography of the retina in the adult pigtail macaque (Macaca nemestrina). *Journal of Comparative Neurology*. **288**(1), pp.165–183.

Pahlberg, J. and Sampath, A.P. 2011. Visual threshold is set by linear and nonlinear mechanisms in the retina that mitigate noise: How neural circuits in the retina improve the single-to-noise ratio of the single-photon response. *BioEssays : news and reviews in molecular, cellular and developmental biology*. **33**(6), p.438.

Palczewski, K., Jäger, S., Buczyłko, J., Crouch, R.K., Bredberg, L., Hofmann, K.P., Asson-Batres, M.A. and Saari, J.C. 1994. Rod Outer Segment Retinol Dehydrogenase: Substrate Specificity and Role in Phototransduction. *Biochemistry*. **33**(46), pp.13741–13750.

da Palma, M.M., Martin, D., Salles, M.V., Teixeira Motta, F.L., Abujamra, S. and Sallum, J.M.F. 2019. Retinal dystrophies and variants in PRPH2. *Arquivos brasileiros de oftalmologia*. **82**(2), pp.158–160.

Panagiotou, E. 2018. *Using NGS to identify new genes and modifiers underlying Familial Exudative Vitreoretinopathy (FEVR)*.[Online]

University of Leeds. [Accessed 3 June 2020]. Available from: http://etheses.whiterose.ac.uk/21578/.

Panagiotou, E.S., Sanjurjo Soriano, C., Poulter, J.A., Lord, E.C., Dzulova, D., Kondo, H., Hiyoshi, A., Chung, B.H.Y., Chu, Y.W.Y., Lai, C.H.Y., Tafoya, M.E., Karjosukarso, D., Collin, R.W.J., Topping, J., Downey, L.M., Ali, M., Inglehearn, C.F. and Toomes, C. 2017. Defects in the Cell Signaling Mediator β-Catenin Cause the Retinal Vascular Condition FEVR. *American journal of human genetics*. **100**(6), pp.960–968.

Pang, A.W., MacDonald, J.R., Pinto, D., Wei, J., Rafiq, M.A., Conrad, D.F., Park, H., Hurles, M.E., Lee, C., Venter, J.C., Kirkness, E.F., Levy, S., Feuk, L. and Scherer, S.W. 2010. Towards a comprehensive structural variation map of an individual human genome. *Genome biology*. **11**(5).

Panneman, D.M., Hitti-Malin, R.J., Holtes, L.K., Bruijn, S.E. de, Reurink, J., Boonen, E.G.M., Khan, M.I., Ali, M., Andréasson, S., Baere, E. De, Banfi, S., Bauwens, M., Ben-Yosef, T., Bocquet, B., Bruyne, M. De, Cerda, B. de la, Coppieters, F., Farinelli, P., Guignard, T., Inglehearn, C.F., Karali, M., Kjellström, U., Koenekoop, R., Koning, B. de, Leroy, B.P., McKibbin, M., Meunier, I., Nikopoulos, K., Nishiguchi, K.M., Poulter, J.A., Rivolta, C., Rúa, E.R. de la, Saunders, P., Simonelli, F., Tatour, Y., Testa, F., Thiadens, A.A.H.J., Toomes, C., Tracewska-Siemiatkowska, A.M., Tran, H.V., Ushida, H., Vaclavik, V., Verhoeven, V.J.M., Vorst, M. van de, Gilissen, C., Hoischen, A., Cremers, F.P.M. and Roosing, S. 2022. Cost-effective sequence analysis of 113 genes in 1,192 probands with retinitis pigmentosa and Leber congenital amaurosis. *medRxiv*., 2022.11.24.22282656.

Papastavrou, V.T., Bradshaw, K.R., Aye, K.H., Turney, C. and Browning, A.C. 2015. Improvement of retinal function in L-ORD after prolonged dark adaptation. *Canadian journal of ophthalmology. Journal canadien d'ophtalmologie*. **50**(2), pp.112–118.

Papermaster, D.S., Converse, C.A. and Zorn, M. 1976. Biosynthetic and immunochemical characterization of large protein in frog and cattle rod outer segment membranes. *Experimental eye research*. **23**(2), pp.105–115.

Parker, M.A., Choi, D., Erker, L.R., Pennesi, M.E., Yang, P., Chegarnov, E.N., Steinkamp, P.N., Schlechter, C.L., Dhaenens, C.M., Mohand-Said, S., Audo, I., Sahel, J., Weleber, R.G. and Wilson, D.J. 2016. Test–Retest Variability of Functional and Structural Parameters in Patients with Stargardt Disease Participating in the SAR422459 Gene Therapy Trial. *Translational Vision Science & Technology*. **5**(5).

Parker, M.A., Erker, L.R., Audo, I., Choi, D., Mohand-Said, S., Sestakauskas, K., Benoit, P., Appelqvist, T., Krahmer, M., Ségaut-Prévost, C., Lujan, B.J., Faridi, A., Chegarnov, E.N., Steinkamp, P.N., Ku, C., da Palma, M.M., Barale, P.O., Ayelo-Scheer, S., Lauer, A., Stout, T., Wilson, D.J., Weleber, R.G., Pennesi, M.E., Sahel, J.A. and Yang, P. 2022. Three-Year Safety Results of SAR422459 (EIAV-ABCA4) Gene Therapy in Patients With ABCA4-Associated Stargardt Disease: An Open-Label Dose-Escalation Phase I/IIa Clinical Trial, Cohorts 1-5. *American Journal of Ophthalmology*. **240**, pp.285–301.

Parker, R.O. and Crouch, R.K. 2010. The interphotoreceptor retinoid binding (IRBP) is essential for normal retinoid processing in cone photoreceptors. *Advances in Experimental Medicine and Biology*. **664**, pp.141–149.

Patel, A., Dogan, H., Payne, A., Krause, E., Sievers, P., Schoebe, N., Schrimpf, D., Blume, C., Stichel, D., Holmes, N., Euskirchen, P., Hench, J., Frank, S., Rosenstiel-Goidts, V., Ratliff, M., Etminan, N., Unterberg, A., Dieterich, C., Herold-Mende, C., Pfister, S.M., Wick, W., Loose, M., von Deimling, A., Sill, M., Jones, D.T.W., Schlesner, M. and Sahm, F. 2022. Rapid-CNS2: rapid comprehensive adaptive nanopore-sequencing of CNS tumors, a proof-of-concept study. *Acta Neuropathologica*. **143**(5), pp.609–612.

Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B.J. and Loose, M. 2021. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nature biotechnology*. **39**(4), p.442.

Payne, A., Holmes, N., Rakyan, V. and Loose, M. 2019. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics (Oxford, England)*. **35**(13), pp.2193–2198.

Peeters, M.H.C.A., Khan, M., Rooijakkers, A.A.M.B., Mulders, T., Haer-Wigman, L., Boon, C.J.F., Klaver, C.C.W., van den Born, L.I., Hoyng, C.B., Cremers, F.P.M., den Hollander, A.I., Dhaenens, C.M. and Collin, R.W.J. 2021. PRPH2 mutation update: In silico assessment of 245 reported and 7 novel variants in patients with retinal disease. *Human Mutation*. **42**(12), p.1521.

Pellissier, L.P., Alves, C.H., Quinn, P.M., Vos, R.M., Tanimoto, N., Lundvig, D.M.S., Dudok, J.J., Hooibrink, B., Richard, F., Beck, S.C., Huber, G., Sothilingam, V., Garcia Garrido, M., Le Bivic, A., Seeliger, M.W. and Wijnholds, J. 2013. Targeted Ablation of Crb1 and Crb2 in Retinal Progenitor Cells Mimics Leber Congenital Amaurosis. *PLoS Genetics*. **9**(12).

Perez-Carro, R., Corton, M., Sánchez-Navarro, I., Zurita, O., Sanchez-Bolivar, N., Sánchez-Alcudia, R., Lelieveld, S.H., Aller, E., Lopez-Martinez, M.A., López-Molina, M.I., Fernandez-San Jose, P., Blanco-Kelly, F., Riveiro-Alvarez, R., Gilissen, C., Millan, J.M., Avila-Fernandez, A. and Ayuso, C. 2016. Panel-based NGS Reveals Novel Pathogenic Mutations in Autosomal Recessive Retinitis Pigmentosa. . **6**(1), pp.1–10.

Perkins, B.D. 2022. Zebrafish models of inherited retinal dystrophies. *Journal of Translational Genetics and Genomics*. **6**(1), pp.95–110.

Petrukhin, K., Koisti, M.J., Bakall, B., Li, W., Xie, G., Marknell, T., Sandgren, O., Forsman, K., Holmgren, G., Andreasson, S., Vujic, M., Bergen, A.A.B., McGarty-Dugan, V., Figueroa, D., Austin, C.P., Metzker, M.L., Caskey, C.T. and Wadelius, C. 1998. Identification of the gene responsible for Best macular dystrophy. *Nature genetics*. **19**(3), pp.241–247.

Piccardi, M., Fadda, A., Martelli, F., Marangoni, D., Magli, A., Minnella, A.M., Bertelli, M., Di Marco, S., Bisti, S. and Falsini, B. 2019. Antioxidant Saffron and Central Retinal Function in ABCA4-Related Stargardt Macular Dystrophy. *Nutrients*. **11**(10).

Piotter, E., McClements, M.E. and Maclaren, R.E. 2021. Therapy Approaches for Stargardt Disease. *Biomolecules*. **11**(8).

Piotter, E., McClements, M.E. and MacLaren, R.E. 2022. The Scope of Pathogenic ABCA4 Mutations Targetable by CRISPR DNA Base Editing Systems—A Systematic Review. *Frontiers in Genetics*. **12**.

Pittlers, S.J., Lee, A.K., Altherrq, M.R., Howardn, T.A., Seldinq, M.F., Hurwitzii, R.L., Wasmuthe, J.J. and Baehr, W. 1992. THE JOURNAL OF BIOLOGICAL CHEMISTRY Primary Structure and Chromosomal Localization of Human and Mouse Rod Photoreceptor cGMP-gated Cation Channel*. . **267**(9), pp.6257–6262.

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. and Nejentsev, S. 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. **28**(21), pp.2747–2754.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*. **20**(1), pp.110–121.

Poloschek, C.M., Bach, M., Lagrèze, W.A., Glaus, E., Lemke, J.R., Berger, W. and Neidhardt, J. 2010. ABCA4 and ROM1: Implications for Modification of the PRPH2-Associated Macular Dystrophy Phenotype. *Investigative Ophthalmology & Visual Science*. **51**(8), pp.4253–4265.

Pontikos, N., Arno, G., Jurkute, N., Schiff, E., Ba-Abbad, R., Malka, S., Gimenez, A., Georgiou, M., Wright, G., Armengol, M., Knight, H., Katz, M., Moosajee, M., Yu-Wai-Man, P., Moore, A.T., Michaelides, M., Webster, A.R. and Mahroo, O.A. 2020. Genetic Basis of Inherited Retinal Disease in a Molecularly Characterized Cohort of More Than 3000 Families from the United Kingdom. *Ophthalmology*. **127**(10).

Popejoy, A.B. and Fullerton, S.M. 2016. Genomics is failing on diversity. *Nature*. **538**(7624), p.161.

Porreca, G.J., Zhang, K., Li, J.B., Xie, B., Austin, D., Vassallo, S.L., LeProust, E.M., Peck, B.J., Emig, C.J., Dahl, F., Gao, Y., Church, G.M. and Shendure, J. 2007. Multiplex amplification of large sets of human exons. *Nature methods*. **4**(11), pp.931–936.

Potapov, V. and Ong, J.L. 2017. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLOS ONE*. **12**(1), p.e0169774.

Povysil, G., Tzika, A., Vogt, J., Haunschmid, V., Messiaen, L., Zschocke, J., Klambauer, G., Hochreiter, S. and Wimmer, K. 2017. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Human Mutation*. **38**(7), pp.889–897.

Prosseda, P.P., Tran, M., Kowal, T., Wang, B. and Sun, Y. 2022. Advances in Ophthalmic Optogenetics: Approaches and Applications. *Biomolecules*. **12**(2).

Provis, J.M., Penfold, P.L., Cornish, E.E., Sandercoe, T.M. and Madigan, M.C. 2005. Anatomy and development of the macula: specialisation and the vulnerability to macular degeneration. *Clinical and Experimental*

Optometry. **88**(5), pp.269–281.

Pugh, E.N. and Lamb, T.D. 1993. Amplification and kinetics of the activation steps in phototransduction. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* **1141**(2–3), pp.111–149.

Pulman, J., Sahel, J.A. and Dalkara, D. 2022. New Editing Tools for Gene Therapy in Inherited Retinal Dystrophies. *The CRISPR Journal.* **5**(3), p.377.

Purves, D., Augustine, G.J., Fitzpatrick, D., Katz, L.C., LaMantia, A.-S., McNamara, J.O. and Williams, S.M. 2001. Anatomical Distribution of Rods and Cones *In*: *Neuroscience. 2nd edition* [Online]. Sinauer Associates. [Accessed 21 February 2023]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK10848/.

Qian, X., Wang, J., Wang, M., Igelman, A.D., Jones, K.D., Li, Y., Wang, K., Goetz, K.E., Birch, D.G., Yang, P., Pennesi, M.E. and Chen, R. 2021. Identification of Deep-Intronic Splice Mutations in a Large Cohort of Patients With Inherited Retinal Diseases. *Frontiers in Genetics.* **12**, p.276.

Qin, L., Shu, L., Zhong, J., Pan, H., Guo, J., Sun, Q., Yan, X., Tang, B. and Xu, Q. 2019. Association of HIF1A and Parkinson's disease in a Han Chinese population demonstrated by molecular inversion probe analysis. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology.* **40**(9), pp.1927–1931.

Quazi, F., Lenevich, S. and Molday, R.S. 2012. ABCA4 is an N-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer. *Nature Communications 2012 3:1.* **3**(1), pp.1–9.

Quazi, F. and Molday, R.S. 2014. ATP-binding cassette transporter ABCA4 and chemical isomerization protect photoreceptor cells from the toxic accumulation of excess 11-cis-retinal. *Proceedings of the National Academy of Sciences of the United States of America.* **111**(13), pp.5024–5029.

Quinn, P.M., Pellissier, L.P. and Wijnholds, J. 2017. The CRB1 Complex: Following the Trail of Crumbs to a Feasible Gene Therapy Strategy. *Frontiers in Neuroscience.* **11**(APR), p.175.

Ramsden, S.C., Davidson, A.E., Leroy, B.P., Moore, A.T., Webster, A.R., Black, G.C.M. and Manson, F.D.C. 2012. Clinical utility gene card for: BEST1-related dystrophies (Bestrophinopathies). *European journal of human genetics : EJHG.* **20**(5), p.592.

Rattner, A., Smallwood, P.M. and 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**(15), pp.11034–11043.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V. and Korbel, J.O. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxford, England).* **28**(18).

Rautiainen, M., Nurk, S., Walenz, B.P., Logsdon, G.A., Porubsky, D., Rhie, A., Eichler, E.E., Phillippy, A.M. and Koren, S. 2023. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nature biotechnology*.

Ray, T.A., Cochran, K., Kozlowski, C., Wang, J., Alexander, G., Cady, M.A., Spencer, W.J., Ruzycki, P.A., Clark, B.S., Laeremans, A., He, M.X., Wang, X., Park, E., Hao, Y., Iannaccone, A., Hu, G., Fedrigo, O., Skiba, N.P., Arshavsky, V.Y. and Kay, J.N. 2020. Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. *Nature Communications*. **11**(1), pp.1–20.

Reeves, M.J., Goetz, K.E., Guan, B., Ullah, E., Blain, D., Zein, W.M., Tumminia, S.J. and Hufnagel, R.B. 2020. Genotype-phenotype associations in a large PRPH2-related retinopathy cohort. *Human mutation*. **41**(9), p.1528.

Riballo, E., Kühne, M., Rief, N., Doherty, A., Smith, G.C.M., Recio, M.J., Reis, C., Dahm, K., Fricke, A., Krempler, A., Parker, A.R., Jackson, S.P., Gennery, A., Jeggo, P.A. and Löbrich, M. 2004. A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci. *Molecular cell*. **16**(5), pp.715–724.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K. and Rehm, H.L. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics*. **17**(5), pp.405–424.

Richardson, C. and Jasin, M. 2000. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature*. **405**(6787), pp.697–700.

Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., Pineda-Alvarez, D., Aradhya, S., Martin, C.L. and on behalf of the ACMG 2020. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genetics in Medicine*. **22**(2), pp.245–257.

Rivolta, C., McGee, T.L., Frio, T.R., Jensen, R. V., Berson, E.L. and Dryja, T.P. 2006. Variation in retinitis pigmentosa-11 (PRPF31 or RP11) gene expression between symptomatic and asymptomatic patients with dominant RP11 mutations. *Human mutation*. **27**(7), pp.644–653.

Robberecht, C., Voet, T., Esteki, M.Z., Nowakowska, B.A. and Vermeesch, J.R. 2013. Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations. *Genome Research*. **23**(3), p.411.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. 2011. Integrative genomics viewer. . **29**(1).

Rodriguez-Muñoz, A., Liquori, A., García-Bohorquez, B., Jaijo, T., Aller, E.,

Millán, J.M. and García-García, G. 2022. Functional assays of non-canonical splice-site variants in inherited retinal dystrophies genes. *Scientific Reports 2022 12:1*. **12**(1), pp.1–8.

Röhlich, P. 1975. The sensory cilium of retinal rods is analogous to the transitional zone of motile cilia. *Cell and Tissue Research*. **161**(3), pp.421–430.

Rose, A.M., Shah, A.Z., Venturini, G., Krishna, A., Chakravarti, A., Rivolta, C. and Bhattacharya, S.S. 2016. Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. *Scientific reports*. **6**.

Roukos, V., Voss, T.C., Schmidt, C.K., Lee, S., Wangsa, D. and Misteli, T. 2013. Spatial dynamics of chromosome translocations in living cells. *Science (New York, N.Y.)*. **341**(6146), pp.660–664.

Rowe, L.R., Thaker, H.M., Opitz, J.M., Schiffman, J.D., Haddadin, Z.M., Erickson, L.K. and South, S.T. 2013. Molecular inversion probe array for the genetic evaluation of stillbirth using formalin-fixed, paraffin-embedded tissue. *The Journal of molecular diagnostics : JMD*. **15**(4), pp.466–472.

Ruberto, F.P., Balzano, S., Namburi, P., Kimchi, A., Pescini-Gobert, R., Obolensky, A., Banin, E., Ben-Yosef, T., Sharon, D. and Rivolta, C. 2021. Heterozygous deletions of noncoding parts of the PRPF31 gene cause retinitis pigmentosa via reduced gene expression. *Molecular Vision*. **27**, p.107.

Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J. and Delaneau, O. 2021. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics 2021 53:1*. **53**(1), pp.120–126.

Ruddle, J.B., Ebenezer, N.D., Kearns, L.S., Mulhall, L.E., Mackey, D.A. and Hardcastle, A.J. 2009. RPGR ORF15 genotype and clinical variability of retinal degeneration in an Australian population. *The British journal of ophthalmology*. **93**(9), pp.1151–1154.

Ruiz-Ceja, K.A., Capasso, D., Pinelli, M., Del Prete, E., Carrella, D., di Bernardo, D. and Banfi, S. 2023. Definition of the transcriptional units of inherited retinal disease genes by meta-analysis of human retinal transcriptome data. *BMC Genomics 2023 24:1*. **24**(1), pp.1–11.

Runhart, E.H., Sangermano, R., Cornelis, S.S., Verheij, J.B.G.M., Plomp, A.S., Boon, C.J.F., Lugtenberg, D., Roosing, S., Bax, N.M., Blokland, E.A.W., Jacobs-Camps, M.H.M., van der Velde-Visser, S.D., Pott, J.W.R., Rohrschneider, K., Thiadens, A.A.H.J., Klaver, C.C.W., van den Born, L.I., Hoyng, C.B. and Cremers, F.P.M. 2018. The common ABCA4 variant p.Asn1868ile shows nonpenetrance and variable expression of stargardt disease when present in trans with severe variants. *Investigative Ophthalmology and Visual Science*. **59**(8), pp.3220–3231.

Runhart, E.H., Valkenburg, D., Cornelis, S.S., Khan, M., Sangermano, R., Albert, S., Bax, N.M., Astuti, G.D.N., Gilissen, C., Pott, J.W.R., Verheij, J.B.G.M., Blokland, E.A.W., Cremers, F.P.M., Ingeborgh van den Born, L. and Hoyng, C.B. 2019. Late-Onset Stargardt Disease Due to Mild, Deep-Intronic ABCA4 Alleles. *Investigative ophthalmology & visual science*. **60**(13), pp.4249–4256.

Russell, S., Bennett, J., Wellman, J.A., Chung, D.C., Yu, Z.F., Tillman, A., Wittes, J., Pappas, J., Elci, O., McCague, S., Cross, D., Marshall, K.A., Walshire, J., Kehoe, T.L., Reichert, H., Davis, M., Raffini, L., George, L.A., Hudson, F.P., Dingfield, L., Zhu, X., Haller, J.A., Sohn, E.H., Mahajan, V.B., Pfeifer, W., Weckmann, M., Johnson, C., Gewaily, D., Drack, A., Stone, E., Wachtel, K., Simonelli, F., Leroy, B.P., Wright, J.F., High, K.A. and Maguire, A.M. 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**(10097), pp.849–860.

Rutherford, A. 2023. *CONTROL : the dark history and troubling present of eugenics.* [Online]. Orion Publishing Group. [Accessed 13 March 2023]. Available from: https://books.google.com/books/about/Control.html?id=kqUEzwEACAAJ.

Saari, J.C. 2012. Vitamin A metabolism in rod and cone visual cycles. *Annual review of nutrition*. **32**, pp.125–145.

Saberi, M., Golchehre, Z., Karamzade, A., Entezam, M., Eshaghkhani, Y., Alavinejad, E., Jafari, H.K. and Keramatipour, M. 2019. CRB1-Related Leber Congenital Amaurosis: Reporting Novel Pathogenic Variants and a Brief Review on Mutations Spectrum. *Iranian Biomedical Journal*. **23**(5), p.362.

Sahoo, T., Del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W. and Beaudet, A.L. 2008. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nature Genetics 2008 40:6*. **40**(6), pp.719–721.

Salvatore, S., Fishman, G.A., McAnany, J.J. and Genead, M.A. 2014. Association of dark-adapted visual function with retinal structural changes in patients with Stargardt disease. *Retina (Philadelphia, Pa.)*. **34**(5), pp.989–995.

Samorodnitsky, E., Jewell, B.M., Hagopian, R., Miya, J., Wing, M.R., Lyon, E., Damodaran, S., Bhatt, D., Reeser, J.W., Datta, J. and Roychowdhury, S. 2015. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Human Mutation*. **36**(9), pp.903–914.

Sanders, A.D., Falconer, E., Hills, M., Spierings, D.C.J. and Lansdorp, P.M. 2017. Single-cell template strand sequencing by Strand-seq enables the characterization of individual homologs. *Nature Protocols 2017 12:6*. **12**(6), pp.1151–1176.

Sanford Kobayashi, E., Batalov, S., Wenger, A.M., Lambert, C., Dhillon, H., Hall, R.J., Baybayan, P., Ding, Y., Rego, S., Wigby, K., Friedman, J., Hobbs, C. and Bainbridge, M.N. 2022. Approaches to long-read sequencing in a clinical setting to improve diagnostic rate. *Scientific Reports 2022 12:1*. **12**(1), pp.1–7.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. **74**(12), p.5463.

Sangermano, R., Khan, M., Cornelis, S.S., Richelle, V., Albert, S., Garanto, A., Elmelik, D., Qamar, R., Lugtenberg, D., Ingeborgh van den Born, L., Collin, R.W.J. and Cremers, F.P.M. 2018. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Research*. **28**(1), pp.100–110.

Sarmady, M. and Abou Tayoun, A. 2018. Need for Automated Interactive Genomic Interpretation and Ongoing Reanalysis. *JAMA Pediatrics*. **172**(12), pp.1113–1114.

Scheidecker, S., Etard, C., Pierce, N.W., Geoffroy, V., Schaefer, E., Muller, J., Chennen, K., Flori, E., Pelletier, V., Poch, O., Marion, V., Stoetzel, C., Strähle, U., Nachury, M. V. and Dollfus, H. 2014. Exome sequencing of Bardet-Biedl syndrome patient identifies a null mutation in the BBSome subunit BBIP1 (BBS18). *Journal of medical genetics*. **51**(2), pp.132–136.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F. and Pals, G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*. **30**(12), pp.e57–e57.

Schreiber, R., Taschler, U., Preiss-Landl, K., Wongsiriroj, N., Zimmermann, R. and Lass, A. 2012. Retinyl ester hydrolases and their roles in vitamin A homeostasis. *Biochimica et Biophysica Acta*. **1821**(1), p.113.

Schwartz, S.D., Hubschman, J.P., Heilwell, G., Franco-Cardenas, V., Pan, C.K., Ostrick, R.M., Mickunas, E., Gay, R., Klimanskaya, I. and Lanza, R. 2012. Embryonic stem cell trials for macular degeneration: a preliminary report. *Lancet (London, England)*. **379**(9817), pp.713–720.

Schwarz, J.M., Cooper, D.N., Schuelke, M. and Seelow, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods 2014 11:4*. **11**(4), pp.361–362.

Schwarze, K., Buchanan, J., Taylor, J.C. and Wordsworth, S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in medicine : official journal of the American College of Medical Genetics*. **20**(10), pp.1122–1130.

Scortecci, J.F., Molday, L.L., Curtis, S.B., Garces, F.A., Panwar, P., Van Petegem, F. and Molday, R.S. 2021. Cryo-EM structures of the ABCA4 importer reveal mechanisms underlying substrate binding and Stargardt disease. *Nature Communications 2021 12:1*. **12**(1), pp.1–13.

Seabright, M. 1971. A RAPID BANDING TECHNIQUE FOR HUMAN CHROMOSOMES. *The Lancet*. **298**(7731), pp.971–972.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A. and Wigler, M. 2004. Large-scale copy number polymorphism in the human genome. *Science*. **305**(5683), pp.525–528.

Sedlazeck, F.J., Lee, H., Darby, C.A. and Schatz, M.C. 2018. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics 2018 19:6*. **19**(6), pp.329–346.

Seki, M., Katsumata, E., Suzuki, A., Sereewattanawoot, S., Sakamoto, Y., Mizushima-Sugano, J., Sugano, S., Kohno, T., Frith, M.C., Tsuchihara, K. and Suzuki, Y. 2019. Evaluation and application of RNA-Seq by MinION. *DNA research : an international journal for rapid publication of reports on genes and genomes*. **26**(1), pp.55–65.

Sen, S.K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P.A., Dyer, M., Cordaux, R., Liang, P. and Batzer, M.A. 2006. Human genomic deletions mediated by recombination between Alu elements. *American journal of human genetics*. **79**(1), pp.41–53.

Shafin, K., Pesout, T., Chang, P.C., Nattestad, M., Kolesnikov, A., Goel, S., Baid, G., Kolmogorov, M., Eizenga, J.M., Miga, K.H., Carnevali, P., Jain, M., Carroll, A. and Paten, B. 2021. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nature Methods 2021 18:11*. **18**(11), pp.1322–1332.

Shafin, K., Pesout, T., Lorig-Roach, R., Haukness, M., Olsen, H.E., Bosworth, C., Armstrong, J., Tigyi, K., Maurer, N., Koren, S., Sedlazeck, F.J., Marschall, T., Mayes, S., Costa, V., Zook, J.M., Liu, K.J., Kilburn, D., Sorensen, M., Munson, K.M., Vollger, M.R., Monlong, J., Garrison, E., Eichler, E.E., Salama, S., Haussler, D., Green, R.E., Akeson, M., Phillippy, A., Miga, K.H., Carnevali, P., Jain, M. and Paten, B. 2020. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nature Biotechnology 2020 38:9*. **38**(9), pp.1044–1053.

Shaikh, T.H., Kurahashi, H., Saitta, S.C., O'Hare, A.M., Hu, P., Roe, B.A., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budarf, M.L. and Emanuel, B.S. 2000. Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Human Molecular Genetics*. **9**(4), pp.489–501.

Sharma, J., Keeling, K.M. and Rowe, S.M. 2020. PHARMACOLOGICAL APPROACHES FOR TARGETING CYSTIC FIBROSIS NONSENSE MUTATIONS. *European journal of medicinal chemistry*. **200**, p.112436.

Shaw, C.J. and Lupski, J.R. 2005. Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Human genetics*. **116**(1–2), pp.1–7.

Sheck, L.H.N., Esposti, S.D., Mahroo, O.A., Arno, G., Pontikos, N., Wright, G., Webster, A.R., Khan, K.N. and Michaelides, M. 2021. Panel-based genetic testing for inherited retinal disease screening 176 genes. *Molecular genetics & genomic medicine*. **9**(12).

Shen, P., Wang, W., Krishnakumar, S., Palm, C., Chi, A.K., Enns, G.M., Davis, R.W., Speed, T.P., Mindrinos, M.N. and Scharfe, C. 2011. High-quality DNA sequence capture of 524 disease candidate genes. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(16), pp.6549–6554.

Shichida, Y. and Imai, H. 1998. Visual pigment: G-protein-coupled receptor for light signals. *Cellular and molecular life sciences : CMLS*. **54**(12), pp.1299–1315.

Shin, G., Grimes, S.M., Lee, H., Lau, B.T., Xia, L.C. and Ji, H.P. 2017.

CRISPR–Cas9-targeted fragmentation and selective sequencing enable massively parallel microsatellite analysis. *Nature Communications 2017 8:1*. **8**(1), pp.1–13.

Shiroguchi, K., Jia, T.Z., Sims, P.A. and Xie, X.S. 2012. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(4), pp.1347–1352.

Shroyer, N.F., Lewis, R.A., Yatsenko, A.N. and Lupski, J.R. 2001. Null missense ABCR (ABCA4) mutations in a family with stargardt disease and retinitis pigmentosa. *Investigative Ophthalmology & Visual Science*. **42**(12), pp.2757–2761.

Simó, R., Villarroel, M., Corraliza, L., Hernández, C. and Garcia-Ramírez, M. 2010. The retinal pigment epithelium: Something more than a constituent of the blood-retinal barrier-implications for the pathogenesis of diabetic retinopathy. *Journal of Biomedicine and Biotechnology*. **2010**.

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J. and Timp, W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods 2017 14:4*. **14**(4), pp.407–410.

Simunovic, M.P., Shen, W., Lin, J.Y., Protti, D.A., Lisowski, L. and Gillies, M.C. 2019. Optogenetic approaches to vision restoration. *Experimental eye research*. **178**, pp.15–26.

Singh, A.K., Olsen, M.F., Lavik, L.A.S., Vold, T., Drabløs, F. and Sjursen, W. 2021. Detecting copy number variation in next generation sequencing data from diagnostic gene panels. *BMC Medical Genomics*. **14**(1), pp.1–12.

SJOSTRAND, F.S. 1953. The ultrastructure of the outer segments of rods and cones of the eye as revealed by the electron microscope. *Journal of Cellular and Comparative Physiology*. **42**(1), pp.15–44.

Slijkerman, R.W.N., Song, F., Astuti, G.D.N., Huynen, M.A., van Wijk, E., Stieger, K. and Collin, R.W.J. 2015. The pros and cons of vertebrate animal models for functional and therapeutic research on inherited retinal dystrophies. *Progress in retinal and eye research*. **48**, pp.137–159.

Smajlagić, D., Lavrichenko, K., Berland, S., Helgeland, Ø., Knudsen, G.P., Vaudel, M., Haavik, J., Knappskog, P.M., Njølstad, P.R., Houge, G. and Johansson, S. 2020. Population prevalence and inheritance pattern of recurrent CNVs associated with neurodevelopmental disorders in 12,252 newborns and their parents. *European Journal of Human Genetics 2020 29:1*. **29**(1), pp.205–215.

Smallwood, P.M., Wang, Y. and Nathans, J. 2002. Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *Proceedings of the National Academy of Sciences of the United States of America*. **99**(2), pp.1008–1011.

Smirnov, V.M., Nassisi, M., Solis Hernandez, C., Méjécase, C., El Shamieh, S., Condroyer, C., Antonio, A., Meunier, I., Andrieu, C., Defoort-Dhellemmes, S., Mohand-Said, S., Sahel, J.A., Audo, I. and Zeitz, C.

2021. Retinal Phenotype of Patients With Isolated Retinal Degeneration Due to CLN3 Pathogenic Variants in a French Retinitis Pigmentosa Cohort. *JAMA ophthalmology*. **139**(3), pp.278–291.

Smith, D.R., Quinlan, A.R., Peckham, H.E., Makowsky, K., Tao, W., Woolf, B., Shen, L., Donahue, W.F., Tusneem, N., Stromberg, M.P., Stewart, D.A., Zhang, L., Ranade, S.S., Warner, J.B., Lee, C.C., Coleman, B.E., Zhang, Z., McLaughlin, S.F., Malek, J.A., Sorenson, J.M., Blanchard, A.P., Chapman, J., Hillman, D., Chen, F., Rokhsar, D.S., McKernan, K.J., Jeffries, T.W., Marth, G.T. and Richardson, P.M. 2008. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Research*. **18**(10), pp.1638–1642.

Smith, K.C. 1972. Dark repair of DNA damage. *Research progress in organic, biological and medicinal chemistry*. **3 Pt 1**, pp.356–382.

Smithies, O., Gibson, D.M., Fanning, E.M., Percy, M.E., Parr, D.M. and Connell, G.E. 1971. Deletions in immunoglobulin polypeptide chains as evidence for breakage and repair in DNA. *Science (New York, N.Y.)*. **172**(3983), pp.574–577.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dö, H., Cremer, T. and Lichter, P. 1997. Matrix-Based Comparative Genomic Hybridization: Biochips to Screen for Genomic Imbalances. *Genes Chromosomes Cancer*. **20**, pp.399–407.

Song, H., Rossi, E.A., Latchney, L., Bessette, A., Stone, E., Hunter, J.J., Williams, D.R. and Chung, M. 2015. Cone and rod loss in Stargardt disease revealed by adaptive optics scanning light ophthalmoscopy. *JAMA ophthalmology*. **133**(10), pp.1198–1203.

Song, W.K., Park, K.M., Kim, H.J., Lee, J.H., Choi, J., Chong, S.Y., Shim, S.H., Del Priore, L. V. and Lanza, R. 2015. Treatment of macular degeneration using embryonic stem cell-derived retinal pigment epithelium: preliminary results in Asian patients. *Stem cell reports*. **4**(5), pp.860–872.

Soucy, M., Kolesnikova, M., Kim, A.H. and Tsang, S.H. 2023. Phenotypic variability in PRPH2 as demonstrated by a family with incomplete penetrance of autosomal dominant cone-rod dystrophy. *Documenta Ophthalmologica*., pp.1–6.

Stankiewicz, P. and Lupski, J.R. 2002. Genome architecture, rearrangements and genomic disorders. *Trends in Genetics*. **18**(2), pp.74–82.

Stanton, C.M., Borooah, S., Drake, C., Marsh, J.A., Campbell, S., Lennon, A., Soares, D.C., Vallabh, N.A., Sahni, J., Cideciyan, A. V., Dhillon, B., Vitart, V., Jacobson, S.G., Wright, A.F. and Hayward, C. 2017. Novel pathogenic mutations in C1QTNF5 support a dominant negative disease mechanism in late-onset retinal degeneration. *Scientific reports*. **7**(1).

Stargardt, K. 1909. Über familiäre, progressive Degeneration in der Maculagegend des Auges. *Albrecht von Græfe's Archiv für Ophthalmologie*. **71**(3), pp.534–550.

Startek, M., Szafranski, P., Gambin, T., Campbell, I.M., Hixson, P., Shaw, C.A., Stankiewicz, P. and Gambin, A. 2015. Genome-wide analyses of

LINE-LINE-mediated nonallelic homologous recombination. *Nucleic acids research*. **43**(4), pp.2188–2198.

Stephenson, K.A.J., Zhu, J., Wynne, N., Dockery, A., Cairns, R.M., Duignan, E., Whelan, L., Malone, C.P., Dempsey, H., Collins, K., Routledge, S., Pandey, R., Crossan, E., Turner, J., O'Byrne, J.J., Brady, L., Silvestri, G., Kenna, P.F., Farrar, G.J. and Keegan, D.J. 2021. Target 5000: a standardized all-Ireland pathway for the diagnosis and management of inherited retinal degenerations. *Orphanet Journal of Rare Diseases*. **16**(1), pp.1–8.

Stevanovski, I., Chintalaphani, S.R., Gamaarachchi, H., Ferguson, J.M., Pineda, S.S., Scriba, C.K., Tchan, M., Fung, V., Ng, K., Cortese, A., Houlden, H., Dobson-Stone, C., Fitzpatrick, L., Halliday, G., Ravenscroft, G., Davis, M.R., Laing, N.G., Fellner, A., Kennerson, M., Kumar, K.R. and Deveson, I.W. 2022. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Science Advances*. **8**(9), p.17.

Stoddart, D., Heron, A.J., Klingelhoefer, J., Mikhailova, E., Maglia, G. and Bayley, H. 2010. Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. *Nano letters*. **10**(9), pp.3633–3637.

Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G. and Bayley, H. 2009. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(19), pp.7702–7707.

Stone, E.M., Andorf, J.L., Whitmore, S.S., DeLuca, A.P., Giacalone, J.C., Streb, L.M., Braun, T.A., Mullins, R.F., Scheetz, T.E., Sheffield, V.C. and Tucker, B.A. 2017. Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. *Ophthalmology*. **124**(9), pp.1314–1331.

Storz, J.F. 2016. Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. *Physiology*. **31**(3), pp.223–232.

Strauss, O. 2005. The retinal pigment epithelium in visual function. *Physiological reviews*. **85**(3), pp.845–881.

Su, P.Y., Lee, W., Zernant, J., Tsang, S.H., Nagasaki, T., Corneo, B. and Allikmets, R. 2022. Establishment of the iPSC line CUIMCi005-A from a patient with Stargardt disease for retinal organoid culture. *Stem Cell Research*. **65**, p.102973.

Suaning, G.J., Lovell, N.H. and Lehmann, T. 2014. Neuromodulation of the retina from the suprachoroidal space: The Phoenix 99 implant. *IEEE 2014 Biomedical Circuits and Systems Conference, BioCAS 2014 - Proceedings*., pp.256–259.

Suárez-Herrera, N., Riswick, I.B., Vázquez-Domínguez, I., Duijkers, L., Piccolo, D., Cheetham, M.E., Garanto, A. and Collin, R.W. 2022. Combined AON-U7snRNA therapy decreases aberrant splicing caused by multiple deep-intronic ABCA4 variants in vitro. *Investigative Ophthalmology & Visual Science*. **63**(7), pp.64 – A0037-64 – A0037.

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H.Y., Konkel, M.K., Malhotra, A., Stütz, A.M., Shi, X., Casale, F.P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M.J.P., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H.Y.K., Mu, X.J., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J.M., Kong, Y., Lameijer, E.W., McCarthy, S., Flicek, P., Gibbs, R.A., Marth, G., Mason, C.E., Menelaou, A., Muzny, D.M., Nelson, B.J., Noor, A., Parrish, N.F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E.E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalin, A.A., Untergasser, A., Walker, J.A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebat, J., Batzer, M.A., McCarroll, S.A., Mills, R.E., Gerstein, M.B., Bashir, A., Stegle, O., Devine, S.E., Lee, C., Eichler, E.E. and Korbel, J.O. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature*. **526**(7571), pp.75–81.

Sullivan, L.S., Bowne, S.J., Seaman, C.R., Blanton, S.H., Lewis, R.A., Heckenlively, J.R., Birch, D.G., Hughbanks-Wheaton, D. and Daiger, S.P. 2006. Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. *Investigative ophthalmology & visual science*. **47**(10), pp.4579–4588.

Sun, D., Sun, W., Gao, S.Q., Lehrer, J., Naderi, A., Wei, C., Lee, S., Schilb, A.L., Scheidt, J., Hall, R.C., Traboulsi, E.I., Palczewski, K. and Lu, Z.R. 2022. Effective gene therapy of Stargardt disease with PEG-ECO/pGRK1-ABCA4-S/MAR nanoparticles. *Molecular Therapy - Nucleic Acids*. **29**, pp.823–835.

Sun, H., Smallwood, P.M. and Nathans, J. 2000. Biochemical defects in ABCR protein variants associated with human retinopathies. *Nature genetics*. **26**(2), pp.242–246.

Sun, H., Tsunenari, T., Yau, K.W. and Nathans, J. 2002. The vitelliform macular dystrophy protein defines a new family of chloride channels. *Proceedings of the National Academy of Sciences of the United States of America*. **99**(6), pp.4008–4013.

Sun, L.F., Chen, X.J. and Jin, Z.B. 2020. Emerging roles of non-coding RNAs in retinal diseases: A review. *Clinical & experimental ophthalmology*. **48**(8), pp.1085–1101.

Szikra, T., Trenholm, S., Drinnenberg, A., Jüttner, J., Raics, Z., Farrow, K., Biel, M., Awatramani, G., Clark, D.A., Sahel, J.A., Da Silveira, R.A. and Roska, B. 2014. Rods in daylight act as relay cells for cone-driven horizontal cell-mediated surround inhibition. *Nature neuroscience*. **17**(12), pp.1728–1735.

Talevich, E., Shain, A.H., Botton, T. and Bastian, B.C. 2016. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Computational Biology*. **12**(4), p.e1004873.

Tanaka, K., Lee, W., Zernant, J., Schuerch, K., Ciccone, L., Tsang, S.H., Sparrow, J.R. and Allikmets, R. 2018. The Rapid-Onset Chorioretinopathy Phenotype of ABCA4 Disease. *Ophthalmology*. **125**(1), pp.89–99.

Tanna, P., Strauss, R.W., Fujinami, K. and Michaelides, M. 2017. *Stargardt disease: clinical features, molecular genetics, animal models and therapeutic options* [Online]. BMJ Publishing Group Ltd. [Accessed 30 May 2020]. Available from: https://bjo.bmj.com/content/101/1/25.

Taylor, R.L., Parry, N.R.A., Barton, S.J., Campbell, C., Delaney, C.M., Ellingford, J.M., Hall, G., Hardcastle, C., Morarji, J., Nichol, E.J., Williams, L.C., Douzgou, S., Clayton-Smith, J., Ramsden, S.C., Sharma, V., Biswas, S., Lloyd, I.C., Ashworth, J.L., Black, G.C. and Sergouniotis, P.I. 2017. Panel-Based Clinical Genetic Testing in 85 Children with Inherited Retinal Disease. *Ophthalmology*. **124**(7), pp.985–991.

Tebbe, L., Kakakhel, M., Makia, M.S., Al-Ubaidi, M.R. and Naash, M.I. 2020a. The Interplay between Peripherin 2 Complex Formation and Degenerative Retinal Diseases. *Cells*. **9**(3).

Tebbe, L., Kakakhel, M., Makia, M.S., Al-Ubaidi, M.R. and Naash, M.I. 2020b. The Interplay between Peripherin 2 Complex Formation and Degenerative Retinal Diseases. *Cells*. **9**(3).

Tebbe, L., Mwoyosvi, M.L., Crane, R., Makia, M.S., Kakakhel, M., Cosgrove, D., Al-Ubaidi, M.R. and Naash, M.I. 2023. The usherin mutation c.2299delG leads to its mislocalization and disrupts interactions with whirlin and VLGR1. *Nature Communications 2023 14:1*. **14**(1), pp.1–20.

Terakita, A. 2005. The opsins. *Genome Biology*. **6**(3), p.213.

TICE, S.C. 1914. A NEW SEX-LINKED CHARACTER IN DROSOPHILA. *https://doi.org/10.2307/1536196*. **26**(4), pp.221–230.

Tomkiewicz, T.Z., Suárez-Herrera, N., Cremers, F.P.M., Collin, R.W.J. and Garanto, A. 2021. Antisense Oligonucleotide-Based Rescue of Aberrant Splicing Defects Caused by 15 Pathogenic Variants in ABCA4. *International journal of molecular sciences*. **22**(9).

Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S. and Turner, S.W. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Research*. **38**(15).

Treangen, T.J. and Salzberg, S.L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews. Genetics*. **13**(1), p.36.

Trotman, J., Armstrong, R., Firth, H., Trayers, C., Watkins, J., Allinson, K., Jacques, T.S., Nicholson, J.C., Burke, G.A.A., Ambrose, J.C., Arumugam, P., Bevers, R., Bleda, M., Boardman-Pretty, F., Boustred, C.R., Brittain, H., Caulfield, M.J., Chan, G.C., 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, F.J., Maleady-Crowe, F., McEntagart, M., Minneci, F., Moutsianas, L., Mueller, M., Murugaesu, N., Need, A.C., O'Donovan, P., Odhams, C.A., Patch, C., Perez-Gil, D., Pereira, M.B., 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., Behjati, S., Murray, M.J., Hook, C.E. and Tarpey, P. 2022. The NHS

England 100,000 Genomes Project: feasibility and utility of centralised genome sequencing for children with cancer. *British Journal of Cancer 2022 127:1*. **127**(1), pp.137–144.

Tsai, Y.-C., Greenberg, D., Powell, J., Höijer, I., Ameur, A., Strahl, M., Ellis, E., Jonasson, I., Pinto, R.M., Wheeler, V.C., Smith, M.L., Gyllensten, U., Sebra, R., Korlach, J. and Clark, T.A. 2017. Amplification-free, CRISPR-Cas9 Targeted Enrichment and SMRT Sequencing of Repeat-Expansion Disease Causative Genomic Regions. *bioRxiv.*, p.203919.

Tsin, A., Betts-Obregon, B. and Grigsby, J. 2018. Visual cycle proteins: Structure, function, and roles in human retinal disease. *Journal of Biological Chemistry*. **293**(34), pp.13016–13021.

Tsybovsky, Y. and Palczewski, K. 2014. Expression, purification and structural properties of ABC transporter ABCA4 and its individual domains. *Protein expression and purification*. **97**, pp.50–60.

Turner, E.H., Lee, C., Ng, S.B., Nickerson, D.A. and Shendure, J. 2009. Massively parallel exon capture and library-free resequencing across 16 genomes. . **6**(5).

Turro, E., Astle, W.J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H.L., Sanchis-Juan, A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O.S., Downes, K., Haimel, M., Tuna, S., Deevi, S.V.V., Aitman, T.J., Bennett, D.L., Calleja, P., Carss, K., Caulfield, M.J., Chinnery, P.F., Dixon, P.H., Gale, D.P., James, R., Koziell, A., Laffan, M.A., Levine, A.P., Maher, E.R., Markus, H.S., Morales, J., Morrell, N.W., Mumford, A.D., Ormondroyd, E., Rankin, S., Rendon, A., Richardson, S., Roberts, I., Roy, N.B.A., Saleem, M.A., Smith, K.G.C., Stark, H., Tan, R.Y.Y., Themistocleous, A.C., Thrasher, A.J., Watkins, H., Webster, A.R., Wilkins, M.R., Williamson, C., Whitworth, J., Humphray, S., Bentley, D.R., Abbs, S., Abulhoul, L., Adlard, J., Ahmed, M., Alachkar, H., Allsup, D.J., Almeida-King, J., Ancliff, P., Antrobus, R., Armstrong, R., Arno, G., Ashford, S., Attwood, A., Aurora, P., Babbs, C., Bacchelli, C., Bakchoul, T., Banka, S., Bariana, T., Barwell, J., Batista, J., Baxendale, H.E., Beales, P.L., Bentley, D.R., Bierzynska, A., Biss, T., Bitner-Glindzicz, M.A.K., Black, G.C., Bleda, M., Blesneac, I., Bockenhauer, D., Bogaard, H., Bourne, C.J., Boyce, S., Bradley, J.R., Bragin, E., Breen, G., Brennan, P., Brewer, C., Brown, M., Browning, A.C., Browning, M.J., Buchan, R.J., Buckland, M.S., Bueser, T., Diz, C.B., Burn, J., Burns, S.O., Burren, O.S., Burrows, N., Campbell, C., Carr-White, G., Carss, K., Casey, R., Chambers, J., Chambers, J., Chan, M.M.Y., Cheah, C., Cheng, F., Chinnery, P.F., Chitre, M., Christian, M.T., Church, C., Clayton-Smith, J., Cleary, M., Brod, N.C., Coghlan, G., Colby, E., Cole, T.R.P., Collins, J., Collins, P.W., Colombo, C., Compton, C.J., Condliffe, R., Cook, S., Cook, H.T., Cooper, N., Corris, P.A.A., Furnell, A., Cunningham, F., Curry, N.S., Cutler, A.J., Daniels, M.J., Dattani, M., Daugherty, L.C., Davis, J., De Soyza, A., Deevi, S.V.V., Dent, T., Deshpande, C., Dewhurst, E.F., Dixon, P.H., Douzgou, S., Downes, K., Drazyk, A.M., Drewe, E., Duarte, D., Dutt, T., Edgar, J.D.M., Edwards, K., Egner, W., Ekani, M.N., Elliott, P., Erber, W.N., Erwood, M., Estiu, M.C., Evans, D.G., Evans, G., Everington, T., Eyries, M., Fassihi, H., Favier, R., Findhammer, J., Fletcher, D., Flinter, F.A., Floto, R.A., Fowler, T., Fox, J., Frary, A.J.,

French, C.E., Freson, K., Frontini, M., Gale, D.P., Gall, H., Ganesan, V., Gattens, M., Geoghegan, C., Gerighty, T.S.A., Gharavi, A.G., Ghio, S., Ghofrani, H.A., Gibbs, J.S.R., Gibson, K., Gilmour, K.C., Girerd, B., Gleadall, N.S., Goddard, S., Goldstein, D.B., Gomez, K., Gordins, P., Gosal, D., Gräf, S., Graham, J., Grassi, L., Greene, D., Greenhalgh, L., Greinacher, A., Gresele, P., Griffiths, P., Grigoriadou, S., Grocock, R.J., Grozeva, D., Gurnell, M., Hackett, S., Hadinnapola, C., Hague, W.M., Hague, R., Haimel, M., Hall, M., Hanson, H.L., Haque, E., Harkness, K., Harper, A.R., Harris, C.L.L., Hart, D., Hassan, A., Hayman, G., Henderson, A., Herwadkar, A., Hoffman, J., Holden, S., Horvath, R., Houlden, H., Houweling, A.C.C., Howard, L.S., Hu, F., Hudson, G., Hughes, J., Huissoon, A.P., Humbert, M., Humphray, S., Hunter, S., Hurles, M., Irving, M., Izatt, L., Johnson, S.A., Jolles, S., Jolley, J., Josifova, D., Jurkute, N., Karten, T., Karten, J., Kasanicki, M.A., Kazkaz, H., Kazmi, R., Kelleher, P., Kelly, A.M., Kelsall, W., Kempster, C., Kiely, D.G., Kingston, N., Klima, R., Koelling, N., Kostadima, M., Kovacs, G., Koziell, A., Kreuzhuber, R., Kuijpers, T.W., Kumar, A., Kumararatne, D., Kurian, M.A., Laffan, M.A., Lalloo, F., Lambert, M., Lawrie, A., Layton, D.M., Lench, N., Lentaigne, C., Lester, T., Levine, A.P., Linger, R., Longhurst, H., Lorenzo, L.E., Louka, E., Lyons, P.A., Machado, R.D., MacKenzie Ross, R. V., Madan, B., Maher, E.R., Maimaris, J., Malka, S., Mangles, S., Mapeta, R., Marchbank, K.J., Marks, S., Marschall, H.U., Marshall, A., Martin, J., Mathias, M., Matthews, E., Maxwell, H., McAlinden, P., McCarthy, M.I., McKinney, H., McMahon, A., Meacham, S., Mead, A.J., Castello, I.M., Megy, K., Mehta, S.G.G., Michaelides, M., Millar, C., Mohammed, S.N., Moledina, S., Montani, D., Moore, A.T., Morales, J., Morrell, N.W., Mozere, M., Muir, K.W., Mumford, A.D., Nemeth, A.H., Newman, W.G., Newnham, M., Noorani, S., Nurden, P., O'Sullivan, J., Obaji, S., Odhams, C., Okoli, S., Olschewski, A., Olschewski, H., Ong, K.R., Oram, S.H., Ormondroyd, E., Ouwehand, W.H., Palles, C., Papadia, S., Park, S.M., Parry, D., Patel, S., Paterson, J., Peacock, A., Pearce, S.H.H., Peden, J., Peerlinck, K., Penkett, C.J., Pepke-Zaba, J., Petersen, R., Pilkington, C., Poole, K.E.S., Prathalingam, R., Psaila, B., Pyle, A., Quinton, R., Rahman, S., Rankin, S., Rao, A., Raymond, F.L., Rayner-Matthews, P.J., Rees, C., Renton, T., Rhodes, C.J., Rice, A.S.C., Richardson, S., Richter, A., Robert, L., Roberts, I., Rogers, A., Rose, S.J., Ross-Russell, R., Roughley, C., Roy, N.B.A., Ruddy, D.M., Sadeghi-Alavijeh, O., Saleem, M.A., Samani, N., Samarghitean, C., Sanchis-Juan, A., Sargur, R.B., Sarkany, R.N., Satchell, S., Savic, S., Sayer, J.A., Sayer, G., Scelsi, L., Schaefer, A.M., Schulman, S., Scott, R., Scully, M., Searle, C., Seeger, W., Sen, A., Sewell, W.A.C., Seyres, D., Shah, N., Shamardina, O., Shapiro, S.E., Shaw, A.C., Short, P.J., Sibson, K., Side, L., Simeoni, I., Simpson, M.A.A., Sims, M.C., Sivapalaratnam, S., Smedley, D., Smith, K.R., Snape, K., Soranzo, N., Soubrier, F., Southgate, L., Spasic-Boskovic, O., Staines, S., Staples, E., Stark, H., Stephens, J., Steward, C., Stirrups, K.E., Stuckey, A., Suntharalingam, J., Swietlik, E.M., Syrris, P., Tait, R.C., Talks, K., Tan, R.Y.Y., Tate, K., Taylor, J.M., Taylor, J.C., Thaventhiran, J.E., Themistocleous, A.C., Thomas, E., Thomas, D., Thomas, M.J., Thomas, P., Thomson, K., Thrasher, A.J., Threadgold, G., Thys, C., Tilly, T., Tischkowitz, M., Titterton, C., Todd, J.A., Toh, C.H., Tolhuis, B.,

Tomlinson, I.P., Toshner, M., Traylor, M., Treacy, C., Treadaway, P., Trembath, R., Tuna, S., Turek, W., Turro, E., Twiss, P., Vale, T., Geet, C. Van, Zuydam, N. van, Vandekuilen, M., Vandersteen, A.M., Vazquez-Lopez, M., von Ziegenweidt, J., Noordegraaf, A.V., Wagner, A., Waisfisz, Q., Walker, S.M., Walker, N., Walter, K., Ware, J.S., Watkins, H., Watt, C., Webster, A.R., Wedderburn, L., Wei, W., Welch, S.B., Wessels, J., Westbury, S.K., Westwood, J.P., Wharton, J., Whitehorn, D., Whitworth, J., Wilkie, A.O.M., Wilkins, M.R., Williamson, C., Wilson, B.T., Wong, E.K.S., Wood, N., Wood, Y., Woods, C.G., Woodward, E.R.R., Wort, S.J., Worth, A., Wright, M., Yates, K., Yong, P.F.K., Young, T., Yu, P., Yu-Wai-Man, P., Zlamalova, E., Kingston, N., Walker, N., Penkett, C.J., Freson, K., Stirrups, K.E. and Raymond, F.L. 2020. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. **583**(7814), pp.96–102.

Vasudevan, S., Peltz, S.W. and Wilusz, C.J. 2002. Non-stop decay—a new mRNA surveillance pathway. *BioEssays*. **24**(9), pp.785–788.

Vázquez-Domínguez, I., Garanto, A. and Collin, R.W.J. 2019. Molecular Therapies for Inherited Retinal Diseases-Current Standing, Opportunities and Challenges. *Genes*. **10**(9).

Vázquez-Domínguez, I., Li, C.H.Z., Fadaie, Z., Haer-Wigman, L., Cremers, F.P.M., Garanto, A., Hoyng, C.B. and Roosing, S. 2022. Identification of a Complex Allele in IMPG2 as a Cause of Adult-Onset Vitelliform Macular Dystrophy. *Investigative Ophthalmology & Visual Science*. **63**(5), pp.27–27.

Vereecke, N., Botteldoorn, N., Brossé, C., Bonckaert, C., Nauwynck, H., Haesebrouck, F., Boyen, F., Maes, D. and Theuns, S. 2023. Predictive Power of Long-Read Whole-Genome Sequencing for Rapid Diagnostics of Multidrug-Resistant Brachyspira hyodysenteriae Strains A. S. Rogovskyy, ed. *Microbiology spectrum*.

Vicente, A.M., Ballensiefen, W. and Jönsson, J.I. 2020. How personalised medicine will transform healthcare by 2030: The ICPerMed vision. *Journal of Translational Medicine*. **18**(1), pp.1–4.

Vidal, E.A., Moyano, T.C., Bustos, B.I., Pérez-Palma, E., Moraga, C., Riveras, E., Montecinos, A., Azócar, L., Soto, D.C., Vidal, M., Di Genova, A., Puschel, K., Nürnberg, P., Buch, S., Hampe, J., Allende, M.L., Cambiazo, V., González, M., Hodar, C., Montecino, M., Muñoz-Espinoza, C., Orellana, A., Reyes-Jara, A., Travisany, D., Vizoso, P., Moraga, M., Eyheramendy, S., Maass, A., De Ferrari, G. V., Miquel, J.F. and Gutiérrez, R.A. 2019. Whole Genome Sequence, Variant Discovery and Annotation in Mapuche-Huilliche Native South Americans. *Scientific Reports 2019 9:1*. **9**(1), pp.1–11.

Vig, A., Poulter, J.A., Ottaviani, D., Tavares, E., Toropova, K., Tracewska, A.M., Mollica, A., Kang, J., Kehelwathugoda, O., Paton, T., Maynes, J.T., Wheway, G., Arno, G., Ambrose, J.C., Arumugam, P., Baple, E.L., Bleda, M., Boardman-Pretty, F., Boissiere, J.M., Boustred, C.R., Brittain, H., Caulfield, M.J., Chan, G.C., Craig, C.E.H., Daugherty, L.C., de Burca, A., Devereau, A., Elgar, G., Foulger, R.E., Fowler, T., Furió-Tarí, P., Hackett, J.M., Halai, D., Hamblin, A., Henderson, S., Holman, J.E., Hubbard,

T.J.P., Ibáñez, K., Jackson, R., Jones, L.J., Kasperaviciute, D., Kayikci, M., Lahnstein, L., Lawson, K., Leigh, S.E.A., Leong, I.U.S., Lopez, F.J., Maleady-Crowe, F., Mason, J., McDonagh, E.M., Moutsianas, L., Mueller, M., Murugaesu, N., Need, A.C., Odhams, C.A., Patch, C., Perez-Gil, D., Polychronopoulos, D., Pullinger, J., Rahim, T., Rendon, A., Riesgo-Ferreiro, P., Rogers, T., Ryten, M., Savage, K., Sawant, K., Scott, R.H., Siddiq, A., Sieghart, A., Smedley, D., Smith, K.R., Sosinsky, A., Spooner, W., Stevens, H.E., Stuckey, A., Sultana, R., Thomas, E.R.A., Thompson, S.R., Tregidgo, C., Tucci, A., Walsh, E., Watters, S.A., Welland, M.J., Williams, E., Witkowska, K., Wood, S.M., Zarowiecki, M., Khan, K.N., McKibbin, M., Toomes, C., Ali, M., Di Scipio, M., Li, S., Ellingford, J., Black, G., Webster, A., Rydzanicz, M., Stawiński, P., Płoski, R., Vincent, A., Cheetham, M.E., Inglehearn, C.F., Roberts, A. and Heon, E. 2020. DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. *Genetics in medicine : official journal of the American College of Medical Genetics*. **22**(12), pp.2041–2051.

Villanueva, R.A.M. and Chen, Z.J. 2019. ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *https://doi.org/10.1080/15366367.2019.1565254*. **17**(3), pp.160–167.

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., Bird, A.C., Hunt, D.M. and Bhattacharya, S.S. 2001. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Molecular cell*. **8**(2), pp.375–381.

Vithana, E.N., Abu-Safieh, L., Pelosini, L., Winchester, E., Hornan, D., Bird, A.C., Hunt, D.M., Bustin, S.A. and 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**(10), pp.4204–4209.

Völgyi, B., Deans, M.R., Paul, D.L. and Bloomfield, S.A. 2004. Convergence and segregation of the multiple rod pathways in mammalian retina. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. **24**(49), pp.11182–11192.

Vollrath, D., Nathans, J. and Davis, R.W. 1988. Tandem array of human visual pigment genes at Xq28. *Science (New York, N.Y.)*. **240**(4859), pp.1669–1672.

Wald, G. 1951. The photochemical basis of rod vision. *Journal of the Optical Society of America*. **41**(12), pp.949–956.

Wald, G. and Brown, P.K. 1956. SYNTHESIS AND BLEACHING OF RHODOPSIN.

Waldern, J.M., Kumar, J. and Laederach, A. 2022. Disease-associated human genetic variation through the lens of precursor and mature RNA structure. *Human Genetics*. **141**(10), pp.1659–1672.

Wang, J.S. and Kefalov, V.J. 2009. An alternative pathway mediates the mouse and human cone visual cycle. *Current biology : CB*. **19**(19), pp.1665–1669.

Wang, J.S., Nymark, S., Frederiksen, R., Estevez, M.E., Shen, S.Q., Corbo, J.C., Cornwall, M.C. and Kefalov, V.J. 2014. Chromophore Supply Rate-Limits Mammalian Photoreceptor Dark Adaptation. *The Journal of Neuroscience*. **34**(34), p.11212.

Wang, Y., Macke, J.P., Merbs, S.L., Zack, D.J., Klaunberg, B., Bennett, J., Gearhart, J. and Nathans, J. 1992. A locus control region adjacent to the human red and green visual pigment genes. *Neuron*. **9**(3), pp.429–440.

Wang, Y., Smallwood, P.M., Cowan, M., Blesh, D., Lawler, A. and Nathans, J. 1999. Mutually exclusive expression of human red and green visual pigment-reporter transgenes occurs at high frequency in murine cone photoreceptors. *Proceedings of the National Academy of Sciences of the United States of America*. **96**(9), pp.5251–5256.

Wang, Y., Zhao, Y., Bollas, A., Wang, Y. and Au, K.F. 2021. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology 2021 39:11*. **39**(11), pp.1348–1365.

Wässle, H. 2004. Parallel processing in the mammalian retina. *Nature reviews. Neuroscience*. **5**(10), pp.747–757.

Wässle, H., Yamashita, M., Greferath, U., GrüNert, U. and MüLler, F. 1991. The rod bipolar cell of the mammalian retina. *Visual neuroscience*. **7**(1–2), pp.99–112.

Watson, A. and Lako, M. 2022. Retinal organoids provide unique insights into molecular signatures of inherited retinal disease throughout retinogenesis. *Journal of Anatomy*. **00**, pp.1–18.

Watson, C.M., Crinnion, L.A., Hewitt, S., Bates, J., Robinson, R., Carr, I.M., Sheridan, E., Adlard, J. and Bonthron, D.T. 2019. Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications. *Laboratory Investigation 2019 100:1*. **100**(1), pp.135–146.

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., Toomes, C., McKibbin, M., Inglehearn, C.F. and Ali, M. 2014. Mutation screening of retinal dystrophy patients by targeted capture from tagged pooled DNAs and next generation sequencing. . **9**(8), p.e104281.

Weidenhammer, E.M., Singh, M., Ruiz-Noriega, M. and Woolford, J.L. 1996. The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae. *Nucleic acids research*. **24**(6), pp.1164–1170.

Weilguny, L., De Maio, N., Munro, R., Manser, C., Birney, E., Loose, M. and Goldman, N. 2023. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nature Biotechnology 2023*., pp.1–8.

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M. and Jaffe, D.B. 2017. Direct determination of diploid genome sequences. *Genome Research*. **27**(5), pp.757–767.

Weisschuh, N., Buena-Atienza, E. and Wissinger, B. 2021. Splicing mutations in inherited retinal diseases. *Progress in Retinal and Eye Research*. **80**,

p.100874.

Weisschuh, N., Feldhaus, B., Khan, M.I., Cremers, F.P.M., Kohl, S., Wissinger, B. and Zobor, D. 2018. Molecular and clinical analysis of 27 German patients with Leber congenital amaurosis. *PLoS ONE.* **13**(12).

Weisschuh, N., Obermaier, C.D., Battke, F., Bernd, A., Kuehlewein, L., Nasser, F., Zobor, D., Zrenner, E., Weber, E., Wissinger, B., Biskup, S., Stingl, K. and Kohl, S. 2020. 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**(9), pp.1514–1527.

Wells, J., Wroblewski, J., Keen, J., Inglehearn, C., Jubb, C., Eckstein, A., Jay, M., Arden, G., Bhattacharya, S., Fitzke, F. and Bird, A. 1993. Mutations in the human retinal degeneration slow (RDS) gene can cause either retinitis pigmentosa or macular dystrophy. *Nature Genetics 1993 3:3.* **3**(3), pp.213–218.

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., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R. and Hunkapiller, M.W. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature biotechnology.* **37**(10), p.1155.

Werdich, X.Q., Place, E.M. and Pierce, E.A. 2014. Systemic Diseases Associated with Retinal Dystrophies. *http://dx.doi.org/10.3109/08820538.2014.959202.* **29**(5–6), pp.319–328.

White, R.A., Bottos, E.M., Roy Chowdhury, T., Zucker, J.D., Brislawn, C.J., Nicora, C.D., Fansler, S.J., Glaesemann, K.R., Glass, K. and Jansson, J.K. 2016. Moleculo Long-Read Sequencing Facilitates Assembly and Genomic Binning from Complex Soil Metagenomes. *mSystems.* **1**(3).

Wick, R.R., Judd, L.M. and Holt, K.E. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology.* **20**(1), pp.1–10.

Willis, N.A., Rass, E. and Scully, R. 2015. Deciphering the Code of the Cancer Genome: Mechanisms of Chromosome Rearrangement. *Trends in cancer.* **1**(4), pp.217–230.

Willoughby, C.E., Ponzin, D., Ferrari, S., Lobo, A., Landau, K. and Omidi, Y. 2010. Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a review. *Clinical & Experimental Ophthalmology.* **38**(SUPPL. 1), pp.2–11.

Wissinger, B., Gamer, D., Jägle, H., Giorda, R., Marx, T., Mayer, S., Tippmann, S., Broghammer, M., Jurklies, B., Rosenberg, T., Jacobson, S.G., Sener, E.C., Tatlipinar, S., Hoyng, C.B., Castellan, C., Bitoun, P., Andreasson, S., Rudolph, G., Kellner, U., Lorenz, B., Wolff, G., Verellen-Dumoulin, C., Schwartz, M., Cremers, F.P.M., Apfelstedt-Sylla, E., Zrenner, E., Salati, R., Sharpe, L.T. and Kohl, S. 2001. CNGA3 Mutations in Hereditary Cone Photoreceptor Disorders. *American Journal of Human Genetics.* **69**(4), p.722.

Wiszniewski, W., Zaremba, C.M., Yatsenko, A.N., Jamrich, M., Wensel, T.G., Lewis, R.A. and Lupski, J.R. 2005. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. *Human molecular genetics*. **14**(19), pp.2769–2778.

Wittström, E., Ekvall, S., Schatz, P., Bondeson, M.L., Ponjavic, V. and Andréasson, S. 2011. Morphological and functional changes in multifocal vitelliform retinopathy and biallelic mutations in BEST1. *Ophthalmic genetics*. **32**(2), pp.83–96.

Wolock, C.J., Stong, N., Ma, C.J., Nagasaki, T., Lee, W., Tsang, S.H., Kamalakaran, S., Goldstein, D.B. and Allikmets, R. 2019a. A case-control collapsing analysis identifies retinal dystrophy genes associated with ophthalmic disease in patients with no pathogenic ABCA4 variants. *Genetics in medicine : official journal of the American College of Medical Genetics*. **21**(10), pp.2336–2344.

Wolock, C.J., Stong, N., Ma, C.J., Nagasaki, T., Lee, W., Tsang, S.H., Kamalakaran, S., Goldstein, D.B. and Allikmets, R. 2019b. A case–control collapsing analysis identifies retinaldystrophy genes associated with ophthalmic disease in patients with nopathogenic ABCA4 variants. *Genetics in medicine : official journal of the American College of Medical Genetics*. **21**(10), p.2336.

Wong, G.W., Krawczyk, S.A., Kitidis-Mitrokostas, C., Revett, T., Gimeno, R. and Lodish, H.F. 2008. Molecular, biochemical and functional characterizations of C1q/TNF family members: adipose-tissue-selective expression patterns, regulation by PPAR-gamma agonist, cysteine-mediated oligomerizations, combinatorial associations and metabolic functions. *The Biochemical journal*. **416**(2), pp.161–177.

Workman, R.E., Tang, A.D., Tang, P.S., Jain, M., Tyson, J.R., Razaghi, R., Zuzarte, P.C., Gilpatrick, T., Payne, A., Quick, J., Sadowski, N., Holmes, N., de Jesus, J.G., Jones, K.L., Soulette, C.M., Snutch, T.P., Loman, N., Paten, B., Loose, M., Simpson, J.T., Olsen, H.E., Brooks, A.N., Akeson, M. and Timp, W. 2019. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods 2019 16:12*. **16**(12), pp.1297–1305.

Xu, L., Ruddick, W.N., Bolch, S.N., Klingeborn, M., Dyka, F.M., Kulkarni, M.M., Simpson, C.P., Beltran, W.A., Bowes Rickman, C., Smith, W.C. and Dinculescu, A. 2022. Distinct Phenotypic Consequences of Pathogenic Mutants Associated with Late-Onset Retinal Degeneration. *The American journal of pathology*.

Yang, Z., Chen, Y., Lillo, C., Chien, J., Yu, Z., Michaelides, M., Klein, M., Howes, K.A., Li, Y., Kaminoh, Y., Chen, H., Zhao, C., Chen, Y., Al-Sheikh, Y.T., Karan, G., Corbeil, D., Escher, P., Kamaya, S., Li, C., Johnson, S., Frederick, J.M., Zhao, Y., Wang, C., Cameron, D.J., Huttner, W.B., Schorderet, D.F., Munier, F.L., Moore, A.T., Birch, D.G., Baehr, W., Hunt, D.M., Williams, D.S. and Zhang, K. 2008a. Mutant prominin 1 found in patients with macular degeneration disrupts photoreceptor disk morphogenesis in mice. *The Journal of clinical investigation*. **118**(8), pp.2908–2916.

Yang, Z., Chen, Y., Lillo, C., Chien, J., Yu, Z., Michaelides, M., Klein, M.,

Howes, K.A., Li, Y., Kaminoh, Y., Chen, H., Zhao, C., Chen, Y., Al-Sheikh, Y.T., Karan, G., Corbeil, D., Escher, P., Kamaya, S., Li, C., Johnson, S., Frederick, J.M., Zhao, Y., Wang, C., Cameron, D.J., Huttner, W.B., Schorderet, D.F., Munier, F.L., Moore, A.T., Birch, D.G., Baehr, W., Hunt, D.M., Williams, D.S. and Zhang, K. 2008b. Mutant prominin 1 found in patients with macular degeneration disrupts photoreceptor disk morphogenesis in mice. *The Journal of clinical investigation*. **118**(8), pp.2908–2916.

Yau, S.C., Bobrow, M., Mathew, C.G., Abbs, S.J., Yau M Bobrow, S.C., Mathew S J Abbs, C.G. and address, P. 1996. Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. *Journal of Medical Genetics*. **33**(7), pp.550–558.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R. and Ning, Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics (Oxford, England)*. **25**(21), pp.2865–2871.

Young, J.M., Endicott, R.L.M., Parghi, S.S., Walker, M., Kidd, J.M. and Trask, B.J. 2008. Extensive Copy-Number Variation of the Human Olfactory Receptor Gene Family. *The American Journal of Human Genetics*. **83**(2), pp.228–242.

Young, R.W. and Bok, D. 1969. Participation of the retinal pigment epithelium in the rod outer segment renewal process. *The Journal of cell biology*. **42**(2), pp.392–403.

Yousaf, S., Tariq, N., Sajid, Z., Sheikh, S.A., Kausar, T., Waryah, Y.M., Shaikh, R.S., Waryah, A.M., Sethna, S., Riazuddin, S. and Ahmed, Z.M. 2022. Delineating the Molecular and Phenotypic Spectrum of the CNGA3-Related Cone Photoreceptor Disorder in Pakistani Families. *Genes*. **13**(4).

Yu-Wai-Man, P., Griffiths, P.G., Hudson, G. and Chinnery, P.F. 2009. Inherited mitochondrial optic neuropathies. *Journal of medical genetics*. **46**(3), pp.145–158.

Yu, J., Szabo, A., Pagnamenta, A.T., Shalaby, A., Giacopuzzi, E., Taylor, J., Shears, D., Pontikos, N., Wright, G., Michaelides, M., Halford, S., Downes, S. and Consortium, G.E.R. 2022. SVRare: discovering disease-causing structural variants in the 100K Genomes Project. *medRxiv*., 2021.10.15.21265069.

Yu, K., Cui, Y. and Hartzell, H.C. 2006. The bestrophin mutation A243V, linked to adult-onset vitelliform macular dystrophy, impairs its chloride channel function. *Investigative ophthalmology & visual science*. **47**(11), pp.4956–4961.

Yu, Y., Flint, A., Dvorin, E.L. and Bischoff, J. 2002. AC133-2, a novel isoform of human AC133 stem cell antigen. *The Journal of biological chemistry*. **277**(23), pp.20711–20716.

Zacchigna, S., Oh, H., Wilsch-Bräuninger, M., Missol-Kolka, E., Jászai, J., Jansen, S., Tanimoto, N., Tonagel, F., Seeliger, M., Huttner, W.B., Corbeil, D., Dewerchin, M., Vinckier, S., Moons, L. and Carmeliet, P.

2009. Loss of the cholesterol-binding protein prominin-1/CD133 causes disk dysmorphogenesis and photoreceptor degeneration. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. **29**(7), pp.2297–2308.

Zampaglione, E., Kinde, B., Place, E.M., Navarro-Gomez, D., Maher, M., Jamshidi, F., Nassiri, S., Mazzone, J.A., Finn, C., Schlegel, D., Comander, J., Pierce, E.A. and Bujakowska, K.M. 2020. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. . **22**(6), pp.1079–1087.

Zarrei, M., MacDonald, J.R., Merico, D. and Scherer, S.W. 2015. A Copy Number Variation Map of the Human Genome. *Nature Reviews Genetics 2015 16:3*. **16**(3), pp.172–183.

Zernant, J., Lee, W., Collison, F.T., Fishman, G.A., Sergeev, Y. V., Schuerch, K., Sparrow, J.R., Tsang, S.H. and 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**(6).

Zernant, J., Lee, W., Wang, J., Goetz, K., Ullah, E., Nagasaki, T., Su, P.Y., Fishman, G.A., Tsang, S.H., Tumminia, S.J., Brooks, B.P., Hufnagel, R.B., Chen, R. and Allikmets, R. 2022. Rare and common variants in ROM1 and PRPH2 genes trans-modify Stargardt/ABCA4 disease. *PLoS genetics*. **18**(3).

Zhang, C.Z., Leibowitz, M.L. and Pellman, D. 2013. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes & development*. **27**(23), pp.2513–2530.

Zhang, N., Tsybovsky, Y., Kolesnikov, A. V., Rozanowska, M., Swider, M., Schwartz, S.B., Stone, E.M., Palczewska, G., Maeda, A., Kefalov, V.J., Jacobson, S.G., Cideciyan, A. V. and Palczewski, K. 2015. Protein misfolding and the pathogenesis of ABCA4-associated retinal degenerations. *Human molecular genetics*. **24**(11), pp.3220–3237.

Zhang, Q., Zulfiqar, F., Xiao, X., Riazuddin, S.A., Ahmad, Z., Caruso, R., MacDonald, I., Sieving, P., Riazuddin, S.A. and Hejtmancik, J.F. 2007. Severe retinitis pigmentosa mapped to 4p15 and associated with a novel mutation in the PROM1 gene. *Human genetics*. **122**(3–4), pp.293–299.

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W. and Dekker, J. 2012. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*. **148**(5), pp.908–921.

Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M., Mudivarti, P.A., Wyatt, P.W., Bharadwaj, R., Makarewicz, A.J., Li, Y., Belgrader, P., Price, A.D., Lowe, A.J., Marks, P., Vurens, G.M., Hardenbol, P., Montesclaros, L., Luo, M., Greenfield, L., Wong, A., Birch, D.E., Short, S.W., Bjornson, K.P., Patel, P., Hopmans, E.S., Wood, C., Kaur, S., Lockwood, G.K., Stafford, D., Delaney, J.P., Wu, I., Ordonez, H.S., Grimes, S.M., Greer, S., Lee, J.Y., Belhocine, K., Giorda, K.M., Heaton, W.H., McDermott, G.P., Bent, Z.W., Meschi, F., Kondov,

N.O., Wilson, R., Bernate, J.A., Gauby, S., Kindwall, A., Bermejo, C., Fehr, A.N., Chan, A., Saxonov, S., Ness, K.D., Hindson, B.J. and Ji, H.P. 2016. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature biotechnology*. **34**(3), pp.303–311.

Zheng, Z., Li, S., Su, J., Leung, A.W.S., Lam, T.W. and Luo, R. 2022. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nature Computational Science 2022 2:12*. **2**(12), pp.797–803.

Zhong, Q., Chen, C.F., Chen, P.L. and Lee, W.H. 2002. BRCA1 facilitates microhomology-mediated end joining of DNA double strand breaks. *The Journal of biological chemistry*. **277**(32), pp.28641–28647.

Zhou, B., Ho, S.S., Zhang, X., Pattni, R., Haraksingh, R.R. and Urban, A.E. 2018. Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. *Journal of medical genetics*. **55**(11), p.735.

Zhou, B., Shin, G., Greer, S.U., Vervoort, L., Huang, Y., Pattni, R., Ho, M., Wong, W.H., Vermeesch, J.R., Ji, H.P. and Urban, A.E. 2020. Complete and haplotype-specific sequence assembly of segmental duplication-mediated genome rearrangements using targeted CRISPR-targeted ultra-long read sequencing (CTLR-Seq). *bioRxiv*., 2020.10.23.349621.

Zook, J.M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., Weng, Z., Liu, Y., Mason, C.E., Alexander, N., Henaff, E., McIntyre, A.B.R., Chandramohan, D., Chen, F., Jaeger, E., Moshrefi, A., Pham, K., Stedman, W., Liang, T., Saghbini, M., Dzakula, Z., Hastie, A., Cao, H., Deikus, G., Schadt, E., Sebra, R., Bashir, A., Truty, R.M., Chang, C.C., Gulbahce, N., Zhao, K., Ghosh, S., Hyland, F., Fu, Y., Chaisson, M., Xiao, C., Trow, J., Sherry, S.T., Zaranek, A.W., Ball, M., Bobe, J., Estep, P., Church, G.M., Marks, P., Kyriazopoulou-Panagiotopoulou, S., Zheng, G.X.Y., Schnall-Levin, M., Ordonez, H.S., Mudivarti, P.A., Giorda, K., Sheng, Y., Rypdal, K.B. and Salit, M. 2016a. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific Data 2016 3:1*. **3**(1), pp.1–26.

Zook, J.M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., Weng, Z., Liu, Y., Mason, C.E., Alexander, N., Henaff, E., McIntyre, A.B.R., Chandramohan, D., Chen, F., Jaeger, E., Moshrefi, A., Pham, K., Stedman, W., Liang, T., Saghbini, M., Dzakula, Z., Hastie, A., Cao, H., Deikus, G., Schadt, E., Sebra, R., Bashir, A., Truty, R.M., Chang, C.C., Gulbahce, N., Zhao, K., Ghosh, S., Hyland, F., Fu, Y., Chaisson, M., Xiao, C., Trow, J., Sherry, S.T., Zaranek, A.W., Ball, M., Bobe, J., Estep, P., Church, G.M., Marks, P., Kyriazopoulou-Panagiotopoulou, S., Zheng, G.X.Y., Schnall-Levin, M., Ordonez, H.S., Mudivarti, P.A., Giorda, K., Sheng, Y., Rypdal, K.B. and Salit, M. 2016b. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific Data 2016 3:1*. **3**(1), pp.1–26.

Zook, J.M., Chapman, B., Wang, J., Mittelman, D., Hofmann, O., Hide, W. and Salit, M. 2014. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nature Biotechnology 2014 32:3*. **32**(3), pp.246–251.

8 Appendices

8.1 Appendix 1 - Bioinformatic commands

8.1.1 Nanopore sequence analysis

Basecalling:

Request Nvidia node

#$ -l coproc_v100=1

$module load singularity

$module load cuda/10.1.168

$singularity run --nv --bind /nobackup:/nobackup
/nobackup/containers/guppy-gpu-5.0.16.simg guppy_basecaller -x "cuda:0" -
i /nobackup/medbmccb/LR_PCR/2249_CNGB1/fast5 -s
/nobackup/medbmccb/LR_PCR/2249_CNGB1 --flowcell FLO-MIN106 --kit
SQK-LSK109

Define Variables:

SAMPLE_DIR_PATH=$(pwd)

SAMPLE_DIR=$SAMPLE_DIR_PATH/${PATIENT}

FASTQ_SUBDIR="fastq"

FAST5_SUBDIR="fast5"

FASTQ=${SAMPLE_DIR}/${FASTQ_SUBDIR}

FAST5=${SAMPLE_DIR}/${FAST5_SUBDIR}

Compress and merge fastq files:

$gzip ${FASTQ}/*.fastq

$cat ${FASTQ}/*.fastq.gz >> ${FASTQ}/${PATIENT}.concat.fastq.gz

Porechop:

$porechop --discard_middle -i ${FASTQ}/${PATIENT}.concat.fastq -o
${FASTQ}/${PATIENT}.concat.adapt.trim.fastq >>

${FASTQ}/${PATIENT}.adapt.trim.log.std.out >>
${FASTQ}/${PATIENT}.adapt.trim.log.std.err

Generate run statistics and filter reads:

$NanoStat --fastq ${FASTQ}/${PATIENT}.concat.adapt.trim.fastq

> ${FASTQ}/${PATIENT}.unfiltered.nanostat

$NanoFilt ${FASTQ}/${PATIENT}.concat.adapt.trim.fastq -q ${Q}

 --headcrop ${HEADCROP} --length ${LENGTH} --maxlength
${MAXLENGTH} | gzip

> ${FASTQ}/${PATIENT}.concat.adapt.trim.filt.fastq.gz

NanoStat --fastq ${FASTQ}/${PATIENT}.concat.adapt.trim.filt.fastq.gz >
${FASTQ}/${PATIENT}.filtered.nanostat

Alignment:

$minimap2 -ax map-ont -t 3 /nobackup/medbmccb/references/golden_path/

hg19.fa ${PATIENT}.concat.adapt.trim.filt.fastq.gz 2>>
${PATIENT}.minimap.std.err >> ${PATIENT}.concat.sam

Sort and index SAM:

$samtools view -bh ${PATIENT}.concat.sam > ${PATIENT}.concat.bam

$samtools sort -o ${PATIENT}.concat.sorted.bam ${PATIENT}.concat.bam

$samtools index ${PATIENT}.concat.sorted.bam

Downsample SAM:

$samtools view -q 60 -bh -s 0.01 ${PATIENT}.concat.sorted.bam

> ${PATIENT}.concat.sorted.0.01.bam

$samtools index ${PATIENT}.concat.sorted.0.01.bam

Variant Calling with Nanopolish:

```
$gunzip ${FASTQ}/${PATIENT}.concat.adapt.trim.filt.fastq.gz

$nanopolish index -d /nobackup/medbmccb/LR_PCR/${PATIENT}
/fast5 ${FASTQ}/${PATIENT}.concat.adapt.trim.filt.fastq


nanopolish variants --faster --snps -m ${M2} --window chr1:94431793-
94603221 -t 10 --ploidy 2 --verbose --outfile ${FASTQ}/${PATIENT}_above-
${M2}_snps.vcf --reads ${FASTQ}/${PATIENT}.concat.adapt.trim.filt.fastq --
bam ${FASTQ}/${PATIENT}.concat.sorted.bam --genome
/nobackup/medbmccb/references/golden_path/hg19.fa
2>>${FASTQ}/${PATIENT}_above-${M2}.nanopolish.log
```

Variant calling with Longshot:

```
$longshot --strand_bias_pvalue_cutoff 0.01 -S -r chr1:94441353-94600611 -
E 0.25 -C 100000 --bam ceph-new-primers.concat.sorted.bam --ref
/nobackup/medbmccb/references/golden_path/hg19.fa  --out ceph-new-
primers_strand0.01_E0.33_longshot.vcf 2>>longshot.log
```

Variant calling with Medaka:

```
$medaka_variant -i 1343.concat.sorted.bam -r chr1:94456209-94590257 -o .
-f /nobackup/medbmccb/references/golden_path/hg19.fa -N 15 -P 12
```

Filter Medaka VCF:

```
$grep PASS round_1.vcf > filtered_round1.vcf
```

Variant calling with Clair3, without phasing:

```
MODEL_NAME=r941_prom_sup_g5014

SAMPLE=CATCH_ABCA4_pass_only

GENOME=hg38


$run_clair3.sh --bam_fn=${SAMPLE}.concat.sorted.bam --
ref_fn=/nobackup/medbmccb/references/golden_path/${GENOME}.fa --
```

threads=3 --platform="ont" --
model_path="/home/home02/medbmccb/miniconda3/envs/clair3/bin/models/
${MODEL_NAME}" --output=.

Variant calling with clair3, with phasing:

$run_clair3.sh --bam_fn=${SAMPLE}.concat.sorted.bam --
bed_fn=/nobackup//medbmccb/references/ABCA4_bed_files/ABCA4_${GE
NOME}.bed --enable_phasing --
whatshap=/home/home02/medbmccb/miniconda3/envs/clair3/bin/whatshap-
-ref_fn=/nobackup/medbmccb/references/golden_path/${GENOME}.fa --
threads=3 --platform="ont" --
model_path="/home/home02/medbmccb/miniconda3/envs
/clair3/bin/models/${MODEL_NAME}" --output=.

Change read group in bam to 'sample':

$picard AddOrReplaceReadGroups

i= ${FASTQ}/${PATIENT}.concat.sorted.bam
o=${FASTQ}/${PATIENT}.concat.sorted.withrg.bam RGID=1 RGLB=lib1
RGPL=ont RGPU=unit1 RGSM=sample

$samtools index ${FASTQ}/${PATIENT}.concat.sorted.withrg.bam

Phase reads:

$tabix ${FASTQ}/${PATIENT}_above-${M2}_snps.vcf

$whatshap phase ${FASTQ}/${PATIENT}_above-${M2}_snps.vcf
${FASTQ}/${PATIENT}.concat.sorted.withrg.bam --ignore-read-groups -o
${FASTQ}/${PATIENT}_above-${M2}_snps.phased.vcf -r
/nobackup/medbmccb/references/golden_path/hg19.fa
2>>${FASTQ}/${PATIENT}.phasing.log

$bgzip ${FASTQ}/${PATIENT}_above-${M2}_snps.phased.vcf

$tabix ${FASTQ}/${PATIENT}_above-${M2}_snps.phased.vcf.gz

Haplotag reads:

```
$whatshap haplotag ${FASTQ}/${PATIENT}_above-
${M2}_snps.phased.vcf.gz ${FASTQ}/${PATIENT}.concat.sorted.withrg.bam
-o ${FASTQ}/${PATIENT}.phased.bam --reference
/nobackup/medbmccb/references/golden_path/hg19.fa 2>>haplotag.log
```

Manually check variants per overlap:

```
$bedtools intersect -a ${FASTQ}/${PATIENT}_above-${M}snps.vcf -b
/nobackup/medbmccb/references/ABCA4_bed_files/ABCA4_LR_PCR_exten
sions_overlaps.bed -wo > ${FASTQ}/${PATIENT}._above-
${M}_snps_in_extended_overlaps.vcf
```

1.1.1 Nanopore sequence accuracy assay

Change headers in Platinum VCFs:

```
$awk '{if($0 !~ /^#/) print "chr"$0; else print $0}'
HG001_GRCh37_1_22_v4.2.1_benchmark.vcf >
HG001_GRCh37_1_22_v4.2.1_benchmark_headers.vcf
```

Isolate variants in *ABCA4*:

```
$bedtools intersect -a
HG001_GRCh37_1_22_v4.2.1_benchmark_headers.vcf -b
/nobackup/medbmccb/references/ABCA4_bed_files/ABCA4_hg19.bed -wo >
HG001_GRCh37_1_22_v4.2.1_benchmark_headers_ABCA4.vcf
```

Remove indels:

```
$Vcftools --vcf
HG001_GRCh37_1_22_v4.2.1_benchmark_headers_ABCA4.vcf --remove-
indels --recode --recode-INFO-all --out
HG001_GRCh37_1_22_v4.2.1_benchmark_headers_ABCA4_snps.vcf
```

Compare VCFs:

```
$vcftools --diff CEPH/vcfs/CEPH_nanopolish.0.33.vcf --vcf
1343.WGS_ABCA4.vcf --diff-site --not-chr chrM --out
HG001_GRCh37_1_22_v4.2.1_benchmark_headers_ABCA4_snps.vcf
```

Visualise using R:

ggplot(ceph25_0.33snps_overlaps_organised,aes(a,b,color=c))+geom_point
(size=1) + labs(title="Concordance NA12878",
subtitle=expression(paste("Variants Across ", italic("ABCA4 "), "Compared to
benchmark VCF (m=0.33)")), y = "Base Fraction", color="Key",
x="Overlapping Amplicons")
+scale_y_continuous(breaks=scales::pretty_breaks(n=10))+
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
scale_color_manual(values=c("black", "red", "lightskyblue")) +
theme(panel.background = element_rect(fill = 'white', colour = 'white'))
+theme(panel.grid.major=element_line(colour="grey75"),panel.grid.minor=el
ement_line(colour="white")) + geom_hline(yintercept=0.33,
linetype="dashed", color = "red")

1.1.2   Chimerism analysis

Set variables:

POS=94532013

AMP=L10

Tag reads with variant:

$java -jar ~/jvarkit/dist/biostar214299.jar -p ${POS}.tsv ${AMP}.bam >
ceph_${POS}_${AMP}_tagged.sam

Separate reads into separate BAMs by variant:

$samtools view -bh -r HapT ceph_${POS}_${AMP}_tagged.sam >
ceph_${POS}_${AMP}_tagged_HapT.bam

$samtools view -bh -r HapA ceph_${POS}_${AMP}_tagged.sam >
ceph_${POS}_${AMP}_tagged_HapA.bam

$samtools view -bh -r HapC ceph_${POS}_${AMP}_tagged.sam >
ceph_${POS}_${AMP}_tagged_HapC.bam

$samtools view -bh -r HapG ceph_${POS}_${AMP}_tagged.sam > ceph_${POS}_${AMP}_tagged_HapG.bam

$samtools index ceph_${POS}_${AMP}_tagged_HapG.bam

$samtools index ceph_${POS}_${AMP}_tagged_HapC.bam

$samtools index ceph_${POS}_${AMP}_tagged_HapA.bam

$samtools index ceph_${POS}_${AMP}_tagged_HapT.bam

8.2   Appendix 2 - smMIPs CNV analysis

- Calculate the sum of all mapped reads for each patient
- Calculate the average coverage per smMIP target across all patients in the run
  - o The number of reads per target divided by the total read count per proband
  - o Remove samples with less than 400,000 total reads
- Normalize the coverage in each sample
  - o Calculate the average coverage per smMIP target across all patients in the run
  - o Results in the Average coverage of each target of a proband divided by the average coverage of each target in all probands
- **Compare the patient to the group:** ratio R = (average coverage of each target of a patient)/(average coverage of each target in all patients). This number should range from 0 to 2 (approximately).
  - o Use a color code to visualize deletions and duplications
- **Thresholds used for interpretation**
  - o R < 0.1 : homozygous deletion - dark green
  - o ≤ R < 0.65 : heterozygous deletion - light green
  - o 0.65 ≤ R ≤ 1.2 : normal
  - o < R ≤ 1.7: heterozygous duplication - light red
  - o R > 1.7 : homozygous duplication - dark red

Calculate the sum of all reads for each patient

Remove samples with less than 400,000 total reads

Calculate the average coverage per smMIP target in each patient

Calculate the average coverage per smMIP target across all patients in the run

Normalize the coverage in each sample

(Read count per smMIP/Total number of reads per patient)

Calculate R

(Average coverage of each target of a patient/Average coverage of each target in all patients)

R

R > 1.7 : homozygous duplication

1.2 < R ≤ 1.7: heterozygous duplication

R < 0.1 : homozygous deletion

0.1 ≤ R < 0.65 : Heterozygous deletion

0.65 ≤ R ≤ 1.2 : normal

8.3 Appendix 3 - PCR Primers

8.3.1 Long-range Primers to phase *ABCA4*

Primers introduced following redesign are highlighted in red text. All coordinates refer to genome build hg19.

Amplicon name	Amplicon coordinates (chromosome:start-stop)	Amplicon size (bp)	Forward (5'-3')	Reverse (5'-3')	Annealing Temperature (ºC)	
					35 cycles	25 cycles
S1	chr1:94581079-94587651	6572	tatgggccctcatcacactc	gggttcctatccagcctgtt	58	54
S2	chr1:94576013-94583016	7003	tgtcagctcatcctccacag	accaggtatctcagttttcacc	58	54
S3	chr1:94571091-94577994	6903	tgtttgctttgttgttgctagt	ctgggagcctgaggaaagaa	58	54
S4	chr1:94566272-94573027	6755	tcacgtgtgtagtaggctcc	taagaccccaggcagaaagg	58	54
L4	chr1:94562394-94570893	8499	cccttctccccagcaagata	gacgagaggattgcacgaac	58	54
L5	chr1:94556161-94564368	8207	CTTCAAGCTCTTCCGTGTGg	gcccctcctagctattcctg	58	54
S7	chr1:94550053-94558218	8165	ttcacatacccaagagccca	gtctttgcccctctcctctt	58	54
S8	chr1:94544899-94552183	7284	tggggaagtttgcagagaca	TGCAGGTGAATCAGGAGTGT	58	54
S8.2	chr1:94540974-94547825	6851	acaccatttgacctgagaaca	cgcttagtagttgcctgcag	58	54
L8.2	chr1:94537456-94546134	8678	GGTGCTCTCCTTCAACTGGT	Acgatcccctgatttctccc	58	54
S10	chr1:94534599-94541790	7191	ctgttggtctggctcctgta	gttggttgttttgctggcac	58	54
L9	chr1:94531257-94539449	8192	gctctgttgcttcatttgtgc	tatgggcagaggaacacagg	58	54
L10	chr1:94524472-94533111	8639	gcaaggaagacaaagcacct	gacaagggaaaggtgggaga	58	54
L11	chr1:94517468-94526499	9031	tgcttcagggctaacatgga	agcaacagcaagaccaataatct	58	54
L12	chr1:94510646-94519610	8964	TGGTCTCCAAGTCAGCCAAT	AAAGGTGTTCGGTGCATGAC	58	N/A
E_L12_2	chr1:94510646-94522068	11422	TGGTCTCCAAGTCAGCCAAT	GGGTCCACTTTGCCTTTGAC	55	51
L13	chr1:94503740-94512558	8818	TGTGGACCGTCTGAACATCA	tgcggtgcgtatcatagact	58	N/A
E_L13	chr1:94501620-94512558	10938	AACCAGATCACCGCATTCCT	ACAGAGATGGATCAGAGGCC	55	51
L14	chr1:94497123-94506013	8890	ttcctatgtccatccacggg	ttctttgctaggggaggtcc	58	N/A
E_L14	chr1:94496875-94508372	11497	GGTGGACCCTTACTCGAGAC	TGGTGGCCACTGAAGAATCT	55	51

S18	chr1:94494792-94501639	6847	gacagctggcagagtcctta	tgaagcttttggtgagtggc	58	54
S19	chr1:94489852-94496894	7042	agattcttcagtggccacca	cttgttctgatacgggctgc	58	54
E_S20	chr1:94482132-94491845	9713	agttgggagtgagagaagcc	cttgttctgatacgggctgc	58	54
S20	chr1:94484945-94491845	6900	tggcaggagagagagaaaca	ttcgggctcctagttctgtg	58	N/A
E_L17.2	chr1:94477294-94488723	11429	CATTCCCTGATGTGGCGTTT	CCGAGGCTAACTGTGAGGAA	54	51
L17.2	chr1:94477231-94486075	8844	tcgtgtgtctccttccagtc	ttcccctttgtctagccagg	58	N/A
L18	chr1:94470930-94479249	8319	ccagagacagatggccaaac	tgcactctcatgaaacaggc	58	54
L19	chr1:94464973-94472982	8009	tcctctgttgggatggattct	gagtggtggtaagggctgat	58	54
L20	chr1:94457435-94466789	9354	aagtgaggggagtgactgtg	acgacgaagcccatctctag	54	51

8.3.2 Primers for PCR walking of *NDP* deletion

All coordinates refer to genome build hg19.

Amplicon name	Amplicon coordinates (chromosome:start-stop)	Amplicon size (bp)	Forward (5′-3′)	Reverse (5′-3′)	Annealing Temperature (ºC)
NDP_intergenic_1	chrX:43754040-43754446	406	TGCTCCTAGAAAGACTGATGTGA	GGTGAAGCCCCATCTCTACT	58
NDP_intergenic_2	chrX:43765720-43766126	406	GCACACCTATCCCCTAACCT	CCAGTTAGGGTGGTCAGATACA	58
NDP_intergenic_3	chrX:43777485-43777891	406	ACTGAACATGGTGGCAATCA	ACCATTTCATTGTTTAGGCACAT	55
NDP_intergenic_4	chrX:43789146-43789552	406	GGCTGGCTGGTCTAGAATGA	CTACCCAGGAACCCTCACTG	58
NDP_intergenic_5	chrX:43800109-43800515	406	TCTGCCCACTGTTTCTGAGT	ACTCTTCCTCCCTCTTCCCT	58
NDP_intergenic_6	chrX:43803091-43803406	315	CTAGCCCCAGAGTTCATGCT	AAGCCACAAGAAACCACAGC	55
NDP_intergenic_7	chrX:43806152-43806602	450	CAGTTTACAGGAGGCTTGCC	TGGCGAGACAGGGTAAAGTT	58

8.4 Appendix 4 - IGV screenshot of BACs used to optimize long-range ABCA4 primers
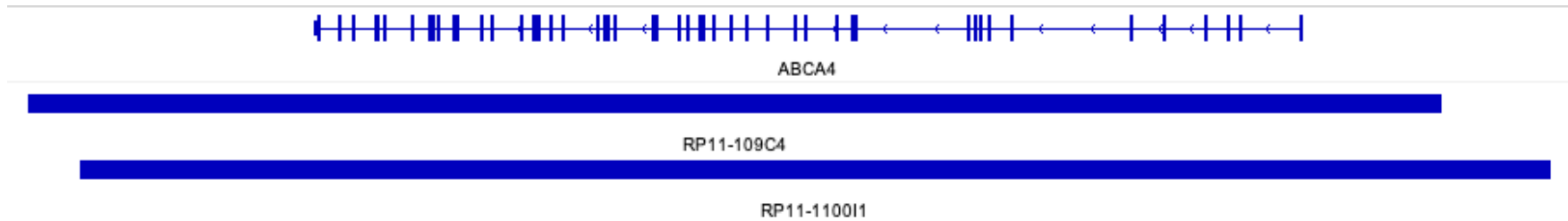
8.5 Appendix 5 - Full list of variants used to compare concordance of haplotypes from nanopore data and platinum VCF

Position (hg19)	Publicly available dataset		Phased nanopore dataset	
	Case 1 (NA12878)		Case 1 (NA12878)	
	Haplotype 1	Haplotype 2	Haplotype 1	Haplotype 2
94465132	T	G	G	T
94466154	T	C	C	T
94466659	A	G	G	A
94468808	T	T	T	T
94469140	A	T	T	A
94469168	T	A	A	T
94469175	T	C	C	T
94469573	G	T	T	G
94470047	A	A	A	A
94470426	T	A	A	T
94471075	G	G	G	G
94471154	C	T	T	C
94471519	C	T	T	C

94471948	C	A	A	C
94472275	G	A	A	G
94472489	C	C	C	C
94472520	A	G	G	A
94472814	G	A	A	G
94472820	T	C	C	T
94472845	G	C		
94472862			C	T
94472909	G	A	A	G
94473479	C	T	T	C
94473557	C	T	T	C
94473605	A	G	G	A
94473845	T	C		
94473864	C	T	T	C
94473896	G	T	T	G
94474020	A	A	A	A
94474185	A	C	C	A
94474328	T	C	C	T
94474452	T	G	G	T
94474647	G	C	C	G
94474727	A	G	G	A
94474872	G	A	A	G
94474961	T	C	C	T
94476035	C	T	T	C
94476388	C	G	G	C
94476555	A	G	G	A
94476605			C	T
94476695	C	T	T	C

94477232	G	G	G	G
94477278	C	T	T	C
94477634	G	A		
94477649	C	G		
94477893	C	C	C	C
94478425	T	T	T	T
94478562	C	C	C	C
94478573	T	C	C	T
94478847	G	A	A	G
94479338	C	T		
94479468	C	T		
94480037	C	T		
94480439	C	C	C	C
94480529	C	C	C	C
94480991	C	C	C	C
94481068	G	G	G	G
94481596	G	G	G	G
94481929	C	T		
94482354	T	C	C	T
94484608			A	A
94484705	C	C	C	C
94485600	C	C	C	C
94486355	G	A	A	G
94486587	T	G	G	T
94487354	G	A	A	G
94488103	C	C	C	C
94488326	T	A	A	T
94488497	A	A	A	A

94489553	T	T	T	T
94489975	C	C	C	C
94491468	A	A	A	A
94492773	T	C	T	C
94494223	T	G	T	G
94494231	A	G	A	G
94495407	C	T	C	T
94495417	C	T	C	T
94495487	G	A	G	A
94495816	T	C		
94495930	C	C	C	C
94496253	G	A	G	A
94497178	A	A	A	A
94498133	G	A	G	A
94498203	G	A		
94499133	T	C	T	C
94499732	C	T	C	T
94499876	C	G	C	G
94500012	T	C	T	C
94501594	A	C	A	C
94501799	C	T	C	T
94503197	A	C		
94504545	C	T	C	T
94505971	G	A	G	A
94512124	G	G	G	G
94512360	A	G	A	G
94512565	C	T	C	T
94512893	C	T	C	T

94512911	G	C	G	C
94513101	T	C	T	C
94513142	A	G	A	G
94513493	G	G	G	G
94514070	A	G	A	G
94515630	A	G	A	G
94515870	G	C	G	C
94516474	T	C	T	C
94520451	G	T	G	T
94522925	A	G	A	G
94523113	G	A	G	A
94524186	A	G	A	G
94524784	A	G	A	G
94524825	A	G	A	G
94525623	G	G	G	G
94526044	A	G	A	G
94527088	T	C	T	C
94528363	T	C	T	C
94529743	G	G	G	G
94529895	C	T	C	T
94530163	A	A	A	A
94530337	G	G	G	G
94530518	C	C	C	C
94531013	C	C	C	C
94531081	C	T	C	T
94531192	G	G	G	G
94531324	C	C	C	C
94531439	A	A	A	A

94531606	C	G	C	G
94532013	T	C	T	C
94532262	C	C	C	C
94532562	C	C	C	C
94533014	C	T	C	T
94533990	T	T	T	T
94534980	G	C	G	C
94535174	G	A	G	A
94536067	T	C	T	C
94537560	C	G	C	G
94537977	C	C	C	C
94540069	C	G	C	G
94542770	T	G	T	G
94544276	A	G	A	G
94545160	T	C	T	C
94547889	G	A	G	A
94549083	C	T	C	T
94550715			A	G
94553092	C	T	C	T
94553438	T	C	T	C
94554453	A	G	A	G
94555166	T	C	T	C
94556861			G	A
94556894	G	A	G	A
94557106	G	A	G	A
94557357	C	T	C	T
94557434	C	T	C	T
94558425	C	T	C	T

94558774	G	C	G	C
94559715	A	G	A	G
94560938	A	A	A	A
94561272	C	A	C	A
94561489	G	G	G	G
94562084	T	C	T	C
94562924	T	C	T	C
94563476	G	T		
94563916	C	C	C	C
94565430	C	G	C	G
94565577	A	G	A	G
94566683			C	C
94567223	T	C	G	A
94568104	A	A	A	A
94568822	C	C	C	C
94569504	G	A	G	A
94570154	A	A	A	A
94570189	C	G	C	G
94570234	G	T	G	T
94570625	A	G	A	G
94572270	G	G	G	G
94572434	T	C	T	C
94572614	T	C	T	C
94572890	A	C	A	C
94573722	G	T	G	T
94573798	C	C	C	C
94573920	A	G	A	G
94574053	C	A	C	A

94574808	A	G	A	G
94574915	C	T	C	T
94575440	A	G		
94575978	T	C	T	C
94576360	A	G	A	G
94576524	G	C	G	C
94576664	A	G	A	G
94576893	G	A		
94576968	T	C		
94577410	T	C	T	C
94577423	T	C	T	C
94577462	G	T	G	T
94578317	C	A	C	A
94578548	C	C	C	C
94579039	A	G	A	G
94579053	G	A		
94579426	C	T	C	T
94579672	A	G	A	G
94579829	C	C	C	C
94580059			C	T
94580233	T	C	T	C
94580874	C	T	C	T
94580906	T	C	T	C
94581119	T	C	T	C
94581125	A	A	A	A
94581258	T	A	T	A
94581348	T	T	T	T
94581456	C	G	C	G

94581529	G	A	G	A
94581540	T	C	T	C
94582249	G	T	G	T
94582492	C	T		
94583038			G	A
94583296	C	C	C	C
94583589	A	G	A	G
94583698	T	G	T	G
94584172	C	T	C	T
94584290	C	T	C	T
94584362	C	G	C	G
94585009	T	C		
94585736	C	C	C	C

8.6   Appendix 6 - Full results of variant callers

						Nanopolish				
Truth Set	**Case**	**Filtering**		**Custom Parameters**		**Concordance**			**Results (%)**	
		Q	**Max Length**	**Base Fraction**	**Indels?**	**True Positive**	**False Negative**	**False Positive**	**Sensitivity**	**Precision**
1343 WGS	Case A. LORD	10	9000	0.35	Yes	100	33	2	75.19	98.04
1343 WGS	Case A. LORD	10	9000	0.2	Yes	104	27	41	79.39	71.72
1343 WGS	Case A. LORD	7	10000	0.2	Yes	104	27	48	79.39	68.42
1343 WGS	Case A. LORD	7	10000	0.3	Yes	100	31	7	76.34	93.46
1343 WGS	Case A. LORD	7	10000	0.35	No	92	42	0	68.66	100.00
1343 WGS	Case A. LORD	7	10000	0.2	No	93	39	15	70.45	86.11

1343 WGS	Case A. LORD	10	9000	0.32	Yes	100	31	6	76.34	94.34
1343 WGS	Case A. LORD	10	9000	0.33	Yes	100	32	4	75.76	96.15
1343 WGS	Case A. LORD	10	9000	0.3	Yes	100	31	6	76.34	94.34

Medaka										
Truth Set	**Case**	**Filtering**		**Custom Parameters**		**Concordance**			**Results (%)**	
		Q	**Max Length**	**SNPs filter**	**Indels filter**	**True Positive**	**False Negative**	**False Positive**	**Sensitivity**	**Precision**
1343 WGS	Case A. LORD	7	10000	12	15	98	34	3	74.24	97.03
1343 WGS	Case A. LORD	7	10000	Default	Default	111	20	11	84.73	90.98
1343 WGS	Case A. LORD	10	9000	10	13	98	34	4	74.24	96.08
1343 WGS	Case A. LORD	10	9000	9	11	109	20	127	84.50	46.19
1343 WGS	Case A. LORD	7	10000	9	11	102	30	3	77.27	97.14
1343 WGS	Case A. LORD	10	9000	13	16	103	28	88	78.63	53.93
1343 WGS	Case A. LORD	10	9000	18	20	95	36	55	72.52	63.33

NanoPolish										
Truth Set	**Case**	**Filtering**		**Custom Parameters**		**Concordance**			**Results (%)**	
		Q	**Max Length**	**Base Fraction**	**Indels?**	**True Positive**	**False Negative**	**False Positive**	**Sensitivity**	**Precision**
GIAB VCF	Case A. NA12878	10	10000	0.33	Yes	195	57	12	77.38	94.20
GIAB VCF	Case A. NA12878	10	10000	0.2	Yes	231	19	63	92.40	78.57
GIAB VCF	Case A. NA12878	10	10000	0.2	No	217	35	29	86.11	88.21
GIAB VCF	Case A. NA12878	10	10000	0.3	Yes	209	43	17	82.94	92.48
GIAB VCF	Case A. NA12878	10	10000	0.3	No	202	50	10	80.16	95.28
GIAB VCF	Case A. NA12878	10	10000	0.25	No	212	40	18	84.13	92.17
GIAB VCF#	Case A. NA12878	10	10000	0.2	No	217	6	32	97.31	87.15

Truth Set	Case	Q	Max Length	Parameters	Indels?	True Positive	False Negative	False Positive	Sensitivity	Precision
GIAB VCF#	Case A. NA12878	10	10000	0.25	No	212	11	21	95.07	90.99
GIAB VCF#	Case A. NA12878	10	10000	0.27	No	208	44	16	82.54	92.86
GIAB VCF	C. NA12878	10	14500	0.33	Yes	213	39	15	84.52	93.42
GIAB VCF	C. NA12878	10	14500	0.2	No	219	33	28	86.90	88.66
GIAB VCF#	C. NA12878	10	14500	0.3	No	209	15	9	93.30	95.87
GIAB VCF#	C. NA12878	10	14500	0.35	No	200	24	7	89.29	96.62
GIAB VCF	C. NA12878	10	14500	0.3	Yes	221	34	19	86.67	92.08
GIAB VCF	C. NA12878	10	14500	0.35	Yes	212	43	11	83.14	95.07
GIAB VCF#	C. NA12878	10	14500	0.25	No	215	9	25	95.98	89.58
GIAB VCF#	C. NA12878	10	14500	0.33	No	204	20	7*		91.07	96.68

Longshot

Truth Set	Case	Filtering		Custom Parameters		Concordance			Results (%)	
		Q	Max Length	Parameters	Indels?	True Positive	False Negative	False Positive	Sensitivity	Precision
GIAB VCF	C. NA12878	10	14500	Default	Yes	114	132	5	46.34	95.80
GIAB VCF	C. NA12878	10	14500	Default	Yes	117	115	5	50.43	95.90

Clair3

Truth Set	Case	Filtering		Custom Parameters		Concordance			Results (%)	
		Q	Max Length	Parameters	Indels?	True Positive	False Negative	False Positive	Sensitivity	Precision
GIAB VCF	C. NA12878	10	14500	Default	Yes	245	15	18	94.23	93.16
GIAB VCF#	C. NA12878	10	14500	Default	No	222	6	6*	97.37	96.94

8.7 Appendix 7 - Target list of smMIPs panel

Gene/Locus	Associated disorder
ABCA1	AMD
ABCA4	MD/RP-LCA
ADAM19	AMD
ADAM9	MD
AIPL1	MD/RP-LCA
APOE	AMD
ARHGEF18	MD/RP-LCA
ARMS2	AMD
ATF6	MD
BCAR1	AMD
BEST1	MD/RP-LCA
C1QTNF5	MD
C2	AMD
C21orf2 (CFAP410)	MD
C2ORF71 (PCARE)	MD/RP-LCA
C3	AMD
C8ORF37	MD/RP-LCA
C9	AMD
CABP4	MD/RP-LCA
CACNA1F	MD
CACNA2D4	MD
CD46	AMD
CD55	AMD
CDH3	MD
CDHR1	MD
CEP250	MD
CEP78	MD
CERKL	MD/RP-LCA
CETP	AMD
CFB	AMD
CFH	MD/AMD
CFHR1	AMD
CFHR2	AMD
CFHR3	AMD
CFHR4	AMD

Gene/Locus	Associated disorder
FBN2	MD/AMD
GNAT2	MD
GUCA1A	MD
GUCY2D	MD/RP-LCA
HK1	MD/RP-LCA
HMCN1	MD
HTRA1	AMD
IFT81	MD
IMPG1	MD
IMPG2	MD/RP-LCA
IRX1/MCDR3 locus	MD
KCNV2	MD
KIR2DL4	AMD
LIPC	AMD
MFSD8	MD
NMNAT1	MD/RP-LCA
OPN1LW	MD
OPN1MW	MD
OPN1SW	MD
OTX2	MD/RP-LCA
PDE6C	MD
PDE6H	MD
PHF12	AMD
PITPNM3	MD
POC1B	MD
PRDM13/MCDR1 locus	MD
PROM1	MD/RP-LCA
PRPH2	MD/RP-LCA
PUS7	AMD
RAB28	MD
RAX2	MD
RDH12	MD/RP-LCA
RDH5	MD
RIMS1	MD
RLBP1	AMD/RP-LCA

CFHR5	AMD		*ROM1*	MD/RP-LCA
CFI	AMD		*RP1*	MD/RP-LCA
CLN3	MD/RP-LCA		*RP1L1*	MD/RP-LCA
CNGA3	MD		*RPGR*	MD/RP-LCA
CNGB3	MD		*RPGRIP1*	MD/RP-LCA
CNNM4	MD		*RPGRorf15*	MD
COL4A5	AMD		*RS1*	MD
COL8A1	AMD		*RXFP2*	AMD
CRB1	MD/RP-LCA		*SCN10A*	AMD
CRX	MD/RP-LCA		*SLC16A8*	AMD
CTNNA1	MD		*SPEF2*	AMD
CYMD locus	MD		*TACC2*	AMD
DRAM2	MD		*TIMP3*	AMD/RP-LCA
EFEMP1	MD		*TLCD3B*	MD
ELOVL4	MD		*TTLL5*	MD
ESRRA	AMD		*TULP1*	MD/RP-LCA
EYS	MD/RP-LCA		*UNC119*	MD
			VPS13B	AMD

8.8 Appendix 8 - Full cohort screened using smMIPs panel

STGD=Stargardt, CRD=Cone-rod dystrophy, RP=Retinitis Pigmentosa, MD, Macular Dystrophy, AD=Autosomal Dominant, AR=Automosomal Recessive

Case	Gender	Disease	Inheritance Pattern
127	M	STGD	?
1337	F	STGD	?
1437	M	CRD	?
1746	M	STGD	AR
1808	F	STGD	?
2190	F	CRD	?
2196	F	CRD	?
2215	F	CRD	?
2272	M	CRD	?
2469	M	STGD	Sporadic
2729	M	CRD	?
2843	M	Inverse RP	?
3528	F	North Carolina MD	AD
3536	F	MD	?
3615	M	Best Disease	?

| 3616	M	STGD	?
3654	F	MD	?
3656	F	Occult MD	?
3670	F	MD	?
3787	M	Pattern dystrophy	?
3798	F	STGD	?
3941	M	MD	?
3962	M	MD	?
4030	M	STGD	AD
4048	M	CRD	?
4126	F	STGD	?
4168	M	MD	?
4172	M	CRD	?
4181	F	MD	?
4187	F	MD	?
4203	M	MD	?
4756	M	CRD	?
4767	F	STGD	AD
5152	F	MD	AR
5202	F	STGD	?
5219	M	MD	Sporadic
5225	M	MD	?
5258	F	CRD	Sporadic
5270	M	CRD	AD
5349	M	MD	?
5604	M	STGD	?
5607	F	STGD	?
5608	F	STGD	?
5609	F	STGD	?
5610	F	CRD	AR
5851	M	STGD	?
5852	F	STGD	?
5853	F	STGD	?
5854	M	STGD	?
5855	M	STGD	?
5857	M	MD	?
5859	F	STGD	?
5860	F	STGD	?
5861	M	STGD	?
5862	M	STGD	Sporadic
5863	F	STGD	?
5864	M	STGD	?

| 5865 | M | MD | ? |

8.9 Appendix 9 - Sanger sequence verification of smMIPs results

8.9.1 Verification and segregation of *ABCA4* variants in proband 1746
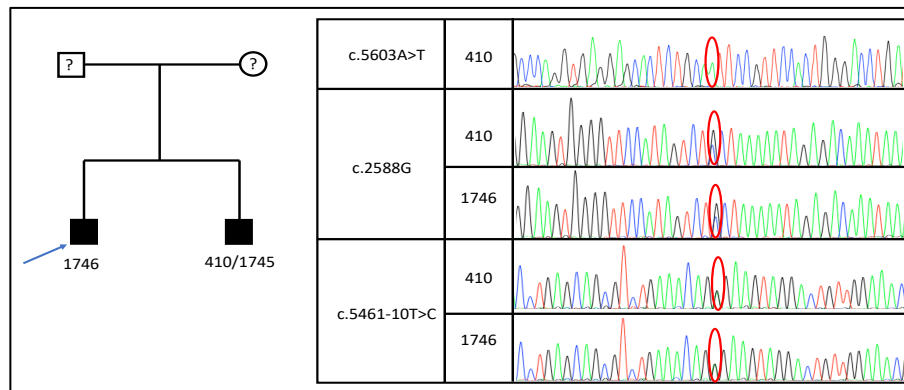
8.9.2 Verification of *ABCA4* variants found in proband 3656
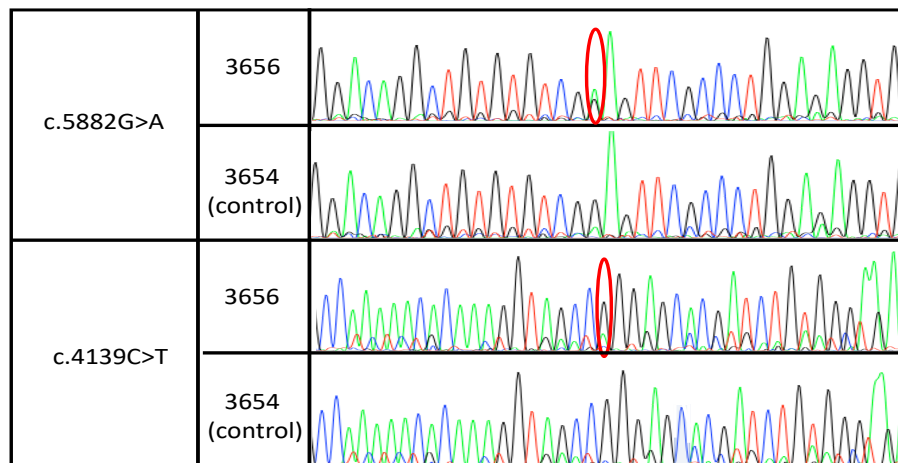
8.9.3 Verification and segregation of homozygous *ABCA4* variant found in case 5857